比特币去匿名化技术研究综述

程杰[1]，金伟[2]，夏清[3]，李淼[4]，戴韡[5]，张亚丰[3]，戴蓬[1]，李玉成[3]

（1.中国人民公安大学侦查学院，北京 100038；2.中国信息通信研究院安全研究所，北京 100191；
3.中国科学院软件研究所并行软件与计算科学实验室，北京 100190；4.中国人民公安大学法学院，北京 100038；
5.中央财经大学金融学院，北京 100081）

摘　要：比特币系统基于区块链技术，具备去中心化、无国界、匿名性等特点，受到产学研界广泛关注。然而，比特币系统在为用户提供隐私保护的同时，也为不法分子开展非法活动提供便利。因此，去匿名化技术研究持续进行并取得系列成果。现有综述多关注隐私保护方案，缺乏去匿名化技术系统梳理。基于此，从用户身份识别、关联地址识别、资金链路追踪3个维度分析现有去匿名化技术及其效果，总结发展现状和难点，并指出未来研究方向。

关键词：比特币；去匿名化；身份识别；关联地址识别；资金链路追踪

中图分类号：TP39

文献标志码：A

DOI：10.11959/j.issn.1000–436x.2024184

Survey of Bitcoin de-anonymization technology

CHENG Jie[1], JIN Wei[2], XIA Qing[3], LI Miao[4], DAI Wei[5], ZHANG Yafeng[3], DAI Peng[1], LI Yucheng[3]

1. School of Criminal Investigation, People's Public Security University of China, Beijing 100038, China
2. Security Research Institute of China Academy of Information and Communications Technology, Beijing 100191, China
3. Laboratory of Parallel Software and Computing Science, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China
4. Law School of People's Public Security University of China, Beijing 100038, China
5. School of Finance, Central University of Finance and Economics, Beijing 100081, China

Abstract: The Bitcoin system, based on blockchain technology, features decentralization, borderlessness, and anonymity, drawing widespread attention from academia and industry. While providing privacy protection for users, the Bitcoin system also facilitates illegal activities. Consequently, de-anonymization techniques have been actively researched, yielding a series of research results. Existing surveys primarily focus on privacy protection schemes, lacking a systematic analysis of de-anonymization techniques. The existing de-anonymization techniques and their effectiveness were analyzed from three dimensions, user identity identification, associated address recognition, and asset flow tracing. The current development status and challenges are summarized, and future research directions are pointed out.

Keywords: Bitcoin, de-anonymization, identity recognition, associated address recognition, asset flow tracing

0　引言

比特币系统是一种基于区块链技术的去中心化虚拟货币交易系统[1]。自2009年年初运行以来，比特币系统交易规模迅速增加，市值不断攀升，催生出一系列面向不同应用场景的虚拟货币。目前，比特币及其衍生的虚拟货币在全球范围内得到广泛应

收稿日期：2024–08–07；**修回日期：**2024–10–08

通信作者：夏清，xiaqing2018@iscas.ac.cn

基金项目：国家重点研发计划基金资助项目（No.2023YFB3106303）；中央高校基本科研业务费专项资金资助项目（No.2023JKF02ZK13）

Foundation Items: The National Key Research and Development Program of China (No.2023YFB3106303), The Fundamental Research Funds for the Central Universities (No.2023JKF02ZK13)

用，并逐渐成为一种支付手段。数据显示，截至2024年9月，全球共有32 627家商家在加密货币地图服务商coinmap上注册接受比特币支付。

比特币系统是一种公开无许可的区块链系统，具有去中心化、无国界、匿名性等显著特点。任何用户均可用假名参与交易，无须经过中心机构的身份验证或合规审查。这些特性为不法分子开展非法活动提供了便利[2]。近年来，我国及欧美监管机构多次将打击比特币等数字资产列上议题。将“比特币”“虚拟货币”作为关键词在裁判文书网上搜索，可以发现，2014—2024年，我国涉比特币案件高达1 431件，在涉虚拟货币的3 698件案件中占比约38.7%。

为应对上述挑战，国内外研究人员针对虚拟货币的匿名性和去匿名化进行了大量研究，部分学者对这些研究成果进行了总结[3-6]。表1对现有相关综述文献进行了归纳。可以看出，现有综述主要侧重于分析区块链系统[3-5]的隐私保护方案，较少涉及去匿名化技术研究。文献[6]分析了部分比特币去匿名化技术，但主要涉及关联地址识别，缺乏对比特币去匿名化技术及实践效果的系统梳理。本文从用户身份识别、关联地址识别、资金链路追踪3个方面对比特币去匿名化技术及实践效果进行总结。用户身份识别技术旨在揭示隐藏在匿名地址后的真实用户身份。关联地址识别技术通过分析用户交易模式等特征，致力于关联由同一用户控制的多个匿名地址。资金链路追踪则针对比特币交易多输入输出的特征，揭示复杂交易结构中的真实资金流动。通过系统性梳理相关工作，本文为研究人员和执法人员了解比特币去匿名化技术的理论与实践现状提供参考，具有较强的现实意义。

1　比特币交易流程

比特币是一种基于区块链技术的去中心化虚拟货币系统，不依赖任何中心化管理机构，通过遍布全球的点对点网络完成交易处理。本节将比特币交易的生命周期分为交易创建、交易传播与验证、交易确认3个阶段，并对各个阶段进行详细论述。通过分析比特币交易生命周期，总结比特币匿名化定义及对应的去匿名化技术。

1.1　交易创建

用户在发起比特币交易前，需要创建符合标准的公私钥对。公钥生成的比特币地址作为链上身份标识，用于接收比特币，私钥由用户保管，用于生成交易签名，以花费对应地址中的比特币。与常见的银行账户体系不同，比特币系统采用特有的未花费交易输出（UTXO, unspent transaction output）作为交易的基本花费单元。比特币交易执行全流程如图1所示，某用户钱包中包含3对密钥，其中，地址1接收3 BTC和1 BTC的UTXO，地址3接收4 BTC和1 BTC的UTXO。用户创建交易时，会从所有UTXO中选取符合条件的UTXO作为交易输入，并生成新的交易输出。例如，用户向第三方转账6 BTC，会将地址1中的3 BTC和地址3中的4 BTC同时作为交易输入，生成6 BTC和1 BTC的交易输出，前者用于完成转账，后者将剩余资金转回给自己。

比特币采用基于堆栈的脚本语言实现交易，利用锁定脚本和解锁脚本验证交易合法性。锁定脚本附加在交易输出上，包含交易接收方的公钥信息和用于验签的操作码，指定了花费该UTXO的条件。解锁脚本附加在交易输入上，包含交易发送方的签名和公钥信息，用于证明发送方对UTXO的所有权。针对不同场景，比特币开发者

表 1　本文与已有综述文献的对比

文献	主要工作	侧重
文献[3]	从可兼容的隐私保护方案（如混币、离链支付）和不可兼容的隐私保护方案（如隐蔽地址技术、环签名、零知识证明、同态加密）2个方面梳理了比特币隐私保护增强技术	比特币隐私保护
文献[4]	从地址混淆、信息隐藏、通道隔离3个方面分析区块链隐私保护机制	区块链隐私保护
文献[5]	从网络层、交易层和应用层分析了区块链系统的隐私保护机制及缺陷，并分析了交易溯源和账户聚类2类隐私攻击方法	区块链隐私保护
文献[6]	从保护与对抗2个维度讨论数字货币的匿名性	数字货币匿名性保护与对抗
本文	从用户身份识别、关联地址识别、资金链路追踪3个方面对比特币去匿名化技术及实践效果进行分析	比特币去匿名化技术实践

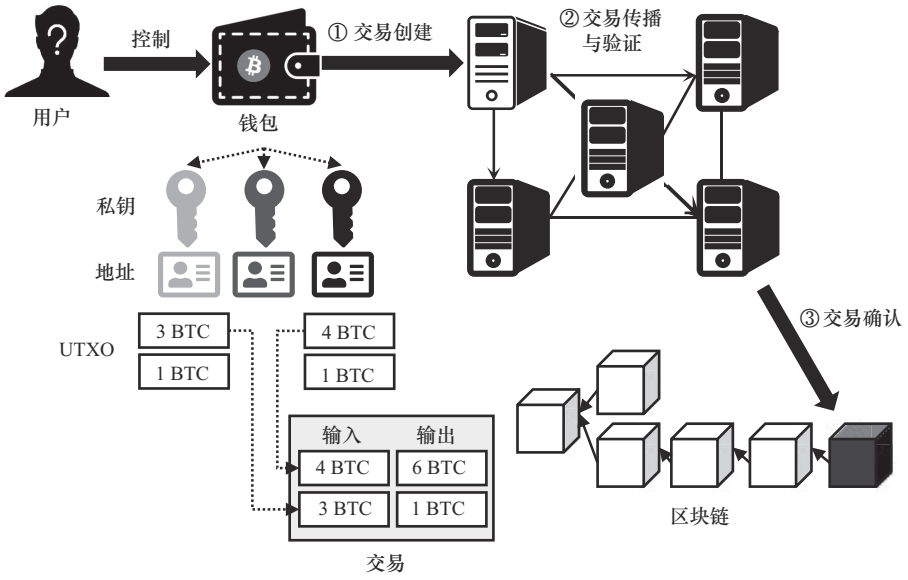


图1　比特币交易执行全流程

设计了多种脚本，包括付款到公钥哈希（P2PKH, pay-to-pubkey-hash）、付款到脚本哈希（P2SH, pay-to-script-hash）、多重签名（Multisig）、付款到隔离见证公钥哈希（P2WPKH, pay-to-witness-pub-key-hash）等[7]。

1.2　交易传播与验证

比特币系统采用点对点（P2P, peer-to-peer）网络完成消息传播。如图1所示，用户通过钱包创建一笔交易后，钱包首先将这笔交易传播给已建立网络连接的节点。节点收到交易后，会进行一系列验证确保交易合法性，包括交易格式是否符合协议规范、UTXO是否未被花费、数字签名是否正确等。验证通过后，节点将该交易存入本地交易缓存池，随后以泛洪的方式广播给其邻居节点。每个比特币节点维护一个邻居节点IP地址列表，并定期更新。如果某个邻居节点长期不响应，节点会移除该地址。此外，为了抵御拒绝服务（DoS, denial of service）攻击，节点对传播错误消息的节点设定惩罚分数，在分数达到阈值后拒绝接收该节点的消息。通过这种机制，比特币网络在高效传播消息的同时抵御潜在攻击。

1.3　交易确认

比特币系统采用基于工作量证明的概率性共识机制[8]对交易和区块进行确认。每个节点从本地交易缓存池中选取部分交易进行打包，尝试生成新区块，这需要节点解决一个复杂的工作量证明难题，即找到一个满足条件的区块哈希值。由于哈希值生成的随机性，节点只能通过计算设备不断尝试，这一过程也被称为“挖矿”。

节点成功生成新区块后，将该区块广播给其他节点，共同更新区块链账本。由于工作量证明机制的随机性，可能出现区块链分叉情形，比特币系统采用最长链机制解决分叉问题，将最长链视为主链，不在最长链上的区块被作为孤块丢弃。孤块中的交易被重新放回交易缓存池中等待重新打包。

1.4　比特币匿名性及去匿名化技术

通过分析比特币交易流程，本文对文献[6,9]总结的3个维度的比特币匿名性，开展对应的去匿名化技术研究。比特币匿名性包括用户身份不可标识、相关地址不可关联、交易资金难以追踪。在身份不可标识方面，用户利用符合标准的密码学算法生成假名地址，就可以参与比特币交易，而假名地址不包含任何用户在物理世界中的真实身份信息。在相关地址不可关联方面，用户可生成任意数量、不存在显式关联关系的假名地址，并在不同交易中使用不同地址。在交易资金难以追踪方面，随着混币等匿名性保护技术的引入，多个用户将交易输入和输出合并到一笔交易中，模糊了输入和输出的对应关系，使交易资金流向难以追踪。针对上述3个方面的比特币匿名性问题，本文开展对应的去匿名化技术研究，即用户身份识别技术、关联地址识别技术和资金链路追踪技术。比特币去匿名化技术概览如图2所示，用户身份识别技术旨在揭示比特币假名地址背后的真实用户身份；关联地址识别技术

致力于关联多个由同一用户控制的地址；资金链路追踪技术则旨在揭示交易中的资金流向。

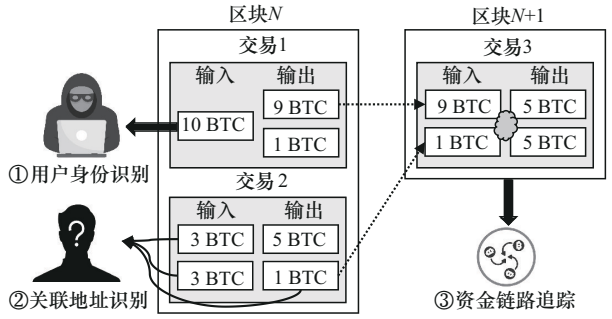


图2　比特币去匿名化技术概览

2　比特币用户身份识别技术

比特币用户身份识别技术旨在将比特币地址和拥有该地址所有权的用户身份进行关联。用户身份不仅包括姓名、联系方式、身份证件等直接身份信息[10-12]，还包括IP地址、行业类别等辅助身份信息[13-14]。后者虽然不能直接揭示用户身份，但可通过与其他数据源交叉关联，帮助推断用户真实身份。本文根据表2所展示的不同维度的身份识别技术对相关工作进行梳理。

2.1　真实身份识别

用户真实身份可通过公开披露、链上交易关联、外部信息关联等多种方式被识别。部分用户出于捐赠等需求，主动将自己的比特币地址公开发布到社交媒体上[10,27]。基于已经识别出身份的部分比特币地址，研究人员可利用链上交易扩充关联更多的比特币地址。此外，支持比特币的购物网站也可能会暴露用户真实身份。

Al Jawaheri等[10]利用社交媒体的公开信息揭露了多个比特币用户的交易信息和真实身份。研究人员观察到在匿名网络运行的隐藏服务公开披露了比特币接收地址，而用户在社交媒体上披露了个人比特币地址，当用户想要购买隐藏服务时，将发起从个人地址到隐藏服务提供商的比特币交易。研究人员从1 500个隐藏服务提供商的网页上收集到88个比特币接收地址，从50亿条Twitter和100万个Bitcointalk论坛页面上分别收集到4 100和4.1万个比特币地址，通过应用启发式聚类规则[28]消除了归属权不确定的地址后，将比特币地址扩展到1 980万个，随后，通过在比特币交易中匹配个人地址和隐藏服务提供商地址，研究人员关联了125个用户与17个隐藏服务发生的交易，识别出用户的姓名、性别、年龄等。

除用户主动公开地址外，用户使用比特币与现实世界商家交易留下的信息也可能暴露身份。Portnoff等[11]通过分析曾经全球最大的广告发布网站Backpage上的成人广告，将部分付费成人广告交易与比特币交易进行映射。Backpage允许用户免费发布广告，并提供付费置顶和推送广告的功能。多名研究人员发现，犯罪分子使用Backpage发布与受害者相关的成人广告[29]。通过对Backpage运营机制进行分析，文献[11]发现用户在发布付费成人广告时需要用比特币进行支付，且这些比特币交易由第三方支付公司GoCoin进行处理。一旦比特币交易在比特币节点的交易缓存池中被检测到，付费广告将立即生效，而无须等待交易被打包进区块链。基于这些观察，研究人员利用比特币地址标签、交易时间和交易价格将付费广告和比特币交易进行关联。实验显示，在4周内发生的54 799笔付费广告交易中，精确匹配到了5 310笔比特币交易。

与文献[11]类似，Goldfeder等[12]通过分析购物网站缓存，将用户的购物交易和比特币交易进行关联，从而推测比特币用户身份。具体来讲，大部分购物网站为了解用户群体，使用由第三方公司提供的嵌入式追踪器，收集用户的设备信息、地理位置、浏览行为等。然而，除了用户基本信息外，研究人员发现许多支持比特币支付的购物网站还向第三方公司提供的嵌入式追踪器共享了用户交易隐私信息，包括商品信息、付款时间、以比特币或法币衡量的付款金额等，甚至包括比特币地址。研究人员使用开源的网络隐私测量工具OpenWPM[30]对130个比特币购物网站进行分析后，发现53个商家向追踪器提供了交易信息。其中，17个商家共享了交易涉及的比特币地址或比特币金额，43个商家共享了以法币衡量的交易金额，28个商家共享了用户的购物车信息。此外，49个商家在购物过程中搜集了用户证件号等隐私信息。基于OpenWPM获取的交易信息，研究人员以76%的概率成功关联购物交易和链上交易。

2.2　IP地址识别

在辅助身份信息识别中，文献[13-15,31]关注比特币用户IP地址识别。这是由于比特币使用去中心化的点对点网络，研究者无须特殊权限即可获

表 2 比特币用户身份识别技术比较

识别维度	文献	利用外部信息	采用方法	实验效果	数据来源
真实身份	文献[10]	网站、论坛、社交媒体	匹配比特币交易的个人地址和隐藏服务提供商地址	关联125个用户和17个隐藏服务发生的交易	bitcointalk、Twitter
	文献[11]	广告网站	根据比特币地址标签、交易时间、交易价格等信息将付费广告和比特币交易进行关联	针对54 799笔付费广告，精确匹配对应的5 310笔比特币交易	—
	文献[12]	购物网站	利用购物交易缓存信息推测地址对应的用户身份	以76%的成功概率关联130家网站的购物交易和链上交易	—
IP地址	文献[13]	网络流量	根据异常中继模式推测比特币交易创建者的IP地址	5个月内识别1 162个账户地址对应的IP	—
	文献[14]		根据入口节点区分共享相同IP的不同比特币客户端	识别11%~60%的账户地址对应的客户端	—
	文献[15]		根据向比特币客户端注入“地址cookie”追踪使用Tor网络的客户端IP地址	控制1%~3%的Tor出口中继节点带宽和1 000~1 500个比特币客户端节点可以搜集全网交易	—
	文献[16]		根据接收的交易次序推测比特币交易创建者的IP地址	69.9%的比特币服务器节点适用于这种溯源机制，获得召回率50%、准确率31.25%的溯源精度	—
多行业类别	文献[17]	地址身份类型标签	利用深度神经网络识别地址身份类型	将相似性阈值设置为50%时，验证率达到86.9%	blockchain.com、bitcoin-whoswho
	文献[18]		利用基于图邻域特征的决策树模型识别地址身份类型	以92%的准确率对30 331 700个比特币地址进行5种类型的识别	walletexplorer
	文献[19]		利用多种有监督机器学习算法识别地址身份类型	梯度提升分类器准确率达到77%	Chainalysis
	文献[20]		利用多种有监督机器学习算法对比特币地址进行身份类型识别	梯度提升分类器准确率达到80%	Chainalysis
	文献[21-22]		基于时序网络的神经网络模型对比特币地址进行身份类别识别	分类准确率达到92%	WalletExplorer、Ethonym
	文献[23]		发布Elliptic数据集，用多种机器学习和深度学习方法进行非法交易二分类	随机森林召回率最高，达到67%	Elliptic数据集
	文献[24]		提出基于自监督深度图和图同构网络的图神经网络框架，并在Elliptic数据集上评估	随机森林的准确率和召回率分别为97.2%和72.1%	Elliptic数据集
	文献[25]		提出与线性层交织的图卷积网络二分类算法并在Elliptic数据集上评估	准确率和召回率分别为97%和67%	Elliptic数据集
	文献[26]		发布Elliptic++数据集，用多种机器学习和深度学习方法进行非法交易二分类	随机森林性能最好，准确率和召回率分别为98.6%和72.7%	Elliptic++数据集

取节点IP地址和网络通信流量，随后通过分析交易传播路径识别发起比特币交易的用户IP地址。进一步地，通过和网络服务提供商关联，IP地址能为用户真实身份识别提供重要线索。

Koshy等[13]部署比特币节点监听全网范围网络流量，通过对交易传播路径进行分析，总结基于网络中继模式的3种启发式规则，如图3所示，以识别交易创建者的IP地址。当一笔交易以规则1中的正常中继模式被传播时，中继链上的首个IP地址被认为是交易创建者对应的IP，规则2和规则3总

结了2种异常中继模式，其中异常IP地址被认为是交易创建者IP。实验显示以上启发式规则在5个月时间内以较高置信度映射了1 162个比特币地址对应的IP，然而，该方法存在无法应对交易创建者使用Tor匿名网络隐藏IP地址的问题。

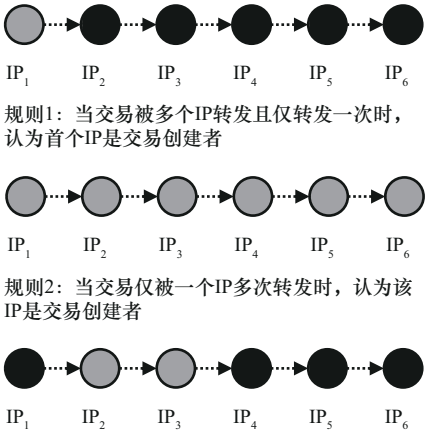


图3　基于网络中继模式的3种启发式规则

为此，Biryukov等[14]提出了一种应对Tor匿名网络的比特币地址与IP地址的映射方法。首先，针对交易发送方可使用Tor网络隐藏IP地址的问题，研究人员利用比特币客户端的抗DoS攻击机制，故意通过Tor节点传播空区块等错误格式信息，使Tor节点被禁用24 h，降低交易发送方使用Tor的可能性。随后利用与比特币客户端建立连接的入口节点对客户端进行唯一标识。如图4所示，客户端A和客户端B位于同一局域网中，具有相同的公共IP地址，分别与灰色节点和黑色节点建立传输连接。研究人员首先部署监测节点监听网络流量，以建立客户端和入口节点的映射关系。由于客户端加入网络时会通过addr消息广播IP地址，且入口节点最先收到addr消息，监测节点认为最早传递addr消息的节点是该客户端的入口节点。实验结果显示，当监测节点和入口节点数量变化时，该方法能够有效区分共享相同IP地址的不同比特币客户端，对11%~60%的比特币地址进行去匿名化。文献[14]利用比特币客户端的DoS保护机制限制用户使用Tor网络，但无法识别Tor后的比特币客户端IP地址。为此，Biryukov等[15]利用向比特币客户端注入“地址cookie”来追踪客户端IP地址。如图5所示，追踪者作为出口节点加入Tor网络，并和Tor出口节点建立连接。为提高成为出口节点的概率，追踪者向其他Tor节点发送错误格式消息使其被禁用。建立连接后，追踪者通过Tor网络的逆向路径向客户端发送经过特殊构造的addr消息，包含多个伪造的比特币节点地址。由于addr消息包含的地址较多，客户端不会转发该消息，而将这些“地址cookie”保存在本地地址库。随后，只要被标识的客户端不经过Tor通信时，追踪者便可发送getaddr消息查看该客户端的“地址cookie”，从而识别Tor网络后隐藏的IP地址。实验结果显示，当追踪者控制1%~3%的Tor出口节点带宽和1 000~1 500个比特币客户端节点后，可以追踪所有比特币客户端通过Tor网络发送的交易。

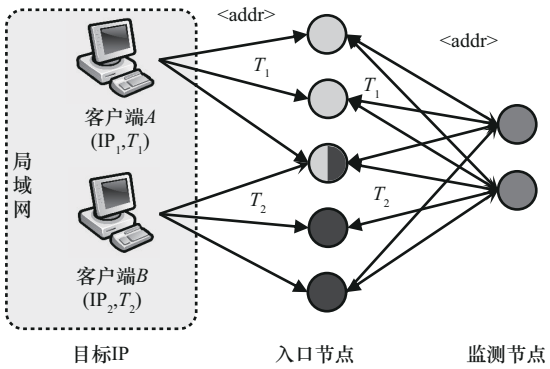


图4　利用入口节点区分共享相同IP的不同比特币客户端

针对文献[14]追踪IP地址时需要持续向比特币网络所有节点发送信息的问题，高峰等[16]提出了一种基于探针节点的轻量级溯源机制，利用探针节点和众多比特币节点建立网络连接，搜集从不同节点转发到探针节点的交易次序，推测目标节点的网络拓扑结构，从而推测待监测节点的始发交易。由于探针节点只接收消息、不转发消息，可以和较大规模的比特币节点同时保持连接，同时也不会对比特币网络的运行造成干扰。实验结果显示，69.9%的比特币服务器节点适用于这种溯源机制，能够获得召回率50%，准确率31.25%的溯源精度。

2.3　行业类别识别

行业类别识别是辅助身份识别的另一重要研究方向，这得益于近年来已知行业比特币地址的标注，包括矿池、钱包、交易所、捐赠地址、攻击地址等。文献[32-33]表明，同一行业的比特币地址在交易模式上表现出类似特征。因此，利用标记的比

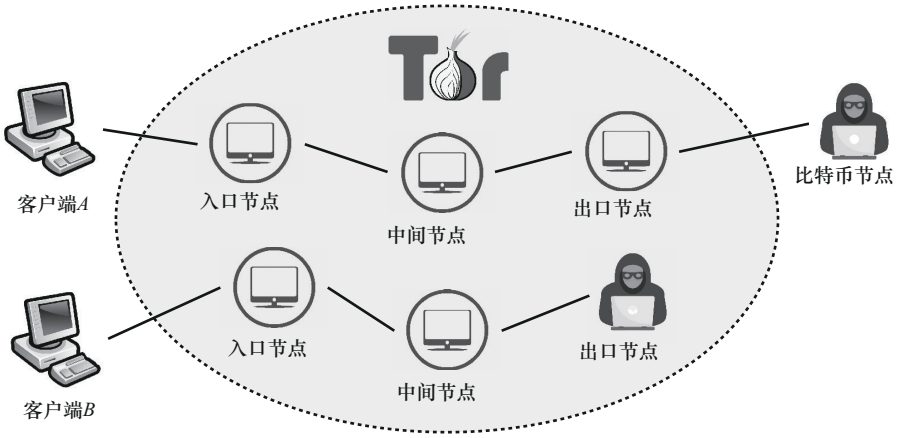
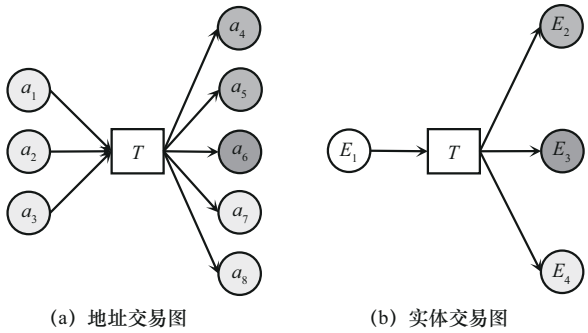


图5　追踪者与采用Tor通信的客户端建立的2种网络连接

特币地址，结合机器学习、深度学习等方法识别未知地址的行业类别成为研究热点。

Shao等[17]采用结合地址特征和交易特征的深度神经网络进行比特币地址行业识别。通过将交易版本号、锁定时间、交易输入和输出等交易特征转换为向量矩阵，利用递归神经网络提取固定长度的交易特征向量，并将其和归一化的地址余额等地址统计特征进行拼接，为每个地址生成了173维的特征向量，并将特征向量输入包含激活函数、线性函数、加性边距软最大值（AM-Softmax, additive margin softmax）成的3层模型进行训练。实验显示，从Blockchain.info和Bitcoin-WhosWho.com收集的包含66个分类标签的8,986个标记地址中，当相似性阈值设置为50%时，成功判断2个地址属于同一类别的准确率达到86.9%。

Jourdan等[18]除了考虑地址特征和交易特征外，还利用图邻域特征识别地址对应的行业类别。如图6所示，研究人员首先将比特币交易建模成地址交易图(A,T,L)，其中A代表地址节点，T代表交易节点，L代表地址和交易的关联边。随后，根据共同输入聚类和传递闭包聚类规则（第3.1节），将多个相关地址聚类成一个实体类别，从而将地址交易图转换为实体交易图。通过对实体交易图进行分析，研究人员观察到不同行业的实体展示出不同的一阶、二阶及三阶子图特征。例如，大部分非循环的一阶子图都涉及交易实体。基于这些观察，研究人员总结出涉及地址特征、实体特征、时序特征、中心度特征、一阶子图、二阶子图、三阶子图共7个方面的315种特征。实验结果显示，在walletExplorer搜集到的5个行业类别相关的30 331 700个地址上，按70%和30%的比例划分训练集和测试集，利用参数优化的决策树模型可达到92%的分类准确率。



（a）地址交易图　　（b）实体交易图

图6　地址交易图和实体交易图示例

Harlev等[19]评估了多种有监督机器学习算法在比特币地址行业类别识别上的效果。研究人员从比特币分析公司Chainalysis获得已被标注为10个行业类别的434个实体地址，包括交易所、托管钱包、商家服务、矿池、混币、勒索软件、诈骗、Tor市场等，共涉及19.8亿笔比特币交易。针对每个类别，研究人员提取包括交易信息、实体信息、集群信息在内的76维特征，分别用k-近邻、随机森林、极度随机树、AdaBoost、决策树、Bagging分类器和梯度提升共7种有监督机器学习分类器进行训练，实验结果显示梯度提升分类器性能最佳，准确率达到77%、F1分数为0.75。Yin等[20]在此基础上进一步扩展行业类别数据集，从Chainalysis公司获取到被标注为12个行业类别的957个实体地址，共涉及3.85亿笔交易，并采用与之相同的有监督机器学习分类器进行训练。实验显示梯度提升分类器性能最佳，达到80.42%的准确率和0.79的F1

分数，相比起来均有所提升。

Han 等[21-22]观察到上述研究工作均未考虑用户所处行业类别会随时间动态变化。例如，23% 的比特币用户在一周时间内从事多种行业活动。为此，Han 等[21-22]提出基于时序网络的多标签分类模型。具体来讲，研究人员首先根据谷歌搜索趋势选取了 5 个热度最高的行业事件，包括 SatoshiDice 非法平台发布、Liberty Reserve 投资平台发布、丝绸之路网站关闭、Mt.Gox 交易所关闭、BTCGuild 矿池关闭。随后从比特币交易图中抽取这些事件前后的交易子图作为时序网络，利用图嵌入算法 GraphSAGE[34]提取用户交易行为特征，并利用多层感知机模型进行行业类别识别。研究人员从 WalletExplorer 和 Ethonym 收集到涵盖 382 个实体的 1.5 亿个标记地址数据集，涉及暗网、交易所、投资和矿工等行业，实验显示该分类模型的平均准确率达到 92%。

Lee 等[35]使用随机森林和神经网络算法检测比特币非法交易的二分类算法，从 WalletExplorer 和 Blokcchain Explorer 收集了合法和非法的比特币交易，提取交易金额、费用等特征进行有监督训练。结果表明，人工神经网络和遗传算法的 F1 分数较高，分别为 89% 和 98%。Zhang 等[36]针对地址缺乏显著特征导致难以准确分类的问题提出比特币地址多分类算法。利用地址与实体之间的映射关系设计基于联合多模型预测的地址分类方案。具体来说，在获得整体特征后，分别进行地址分类和实体聚类任务，使最终结果满足尽可能多的具有相似行为的实体约束，完成最大化共识。使用文献[37]提供的超 2.6 万个比特币标注地址进行评估，发现该方法准确率达到 77.4%。

Weber 等[23]发布了 Elliptic 数据集，包括 20 多万个比特币交易和 23.4 万条资金流边，每笔交易被标记为合法、非法或未知。交易包含 166 维特征，前 94 维特征包含交易金额、费用、输入输出数量等交易信息，后 72 维特征是聚合前后跳交易的聚合特征。研究人员使用逻辑回归、随机森林、多层感知器、图卷积网络和基于动态时空图神经演化的图卷积网络（EvolveGCN, evolving graph convolutional network）方法检测非法交易，发现随机森林的召回率最高，达到 67%。

Elliptic 数据集发布后，研究人员开展了一系列的算法优化工作。文献[24]提出了一种基于自监督深度图和图同构网络的图神经网络框架 Inspection-L，并结合有监督学习算法随机森林进行比特币非法交易的二分类检测。通过在 Elliptic 数据集上评估，发现将节点嵌入和交易原始特征相结合后输入随机森林的分类效果最好，准确率和召回率分别达到 97.2% 和 72.1%。文献[38]提出基于自适应堆叠极端梯度提升的非法交易检测框架，在 Elliptic 数据集上提高了召回率。文献[25]提出使用与线性层交织的图卷积网络对 Elliptic 数据集进行非法交易检测，分类准确率和召回率分别为 97% 和 67%。

近期，Elmougy 等[26]基于 Elliptic 数据集发布了 Elliptic++ 数据集等，包含 82.2 万个账户和 20.3 万笔交易，每个账户和交易都被标注为非法、合法、未知 3 类。根据账户和交易的属性、统计特征等构建出 183 维的交易特征和 56 维的账户特征。应用数据集对随机森林、多层感知机、长短期记忆网络、极端梯度提升 4 种机器学习算法评估欺诈检测效果，并将逻辑回归算法作为基准，结果显示，使用特征细化的随机森林算法性能最优，取得了 98.6% 的准确率和 72.7% 的召回率。

2.4　身份识别技术研究现状总结

现有比特币用户身份识别研究主要集中在识别真实身份、IP 地址及行业类别。此外，还有文献识别比特币用户所处地理时区[39]。在文献梳理过程中，本文发现以下研究点。

1) 用户和商家交互时留下的个人信息有助于识别真实身份。因此，监管机构与比特币商家开展合作有助于追踪用户身份。目前，美国、欧盟、日韩等国家或地区都为比特币支付商家制定了了解你的客户（KYC, know your customer）、税务申报等合规性要求。

2) 识别比特币 IP 地址通常需要部署超级节点，通过多渠道的网络传播数据恢复真实的网络拓扑，从而推测比特币用户 IP 地址。然而，超级节点需要建立大量网络连接并消耗大量存储资源，在消耗较少资源的前提下追踪网络传播是一个重要挑战。

2) 佐治亚理工学院发布的 Elliptic 数据集是目前规模最大的、可公开使用的比特币标注数据集，已有部分学者基于该数据集开展工作。然而，比特币交易网络是一个大规模且动态变化的复杂网络，如何自动更新网络结构、交易行为等多维度特征并进行实时身份识别是研究难点之一。

3 比特币关联地址识别技术

比特币关联地址识别技术是一种通过规则或算法对用户控制的多个地址进行聚类的方法，旨在揭示用户地址和交易的内在联系。比特币用户可以使用多个不同地址参与交易，以保持其身份匿名性。然而，通过分析比特币区块链中的交易记录，包括交易金额、资金流向、交易行为等关键属性，发现可以利用一系列关联地址识别方法挖掘地址间的潜在关联性。关联地址识别技术比较如表3所示，本节将从交易输入、交易输出、交易行为3个维度对相关研究进行梳理。

3.1 基于交易输入的关联地址识别

比特币系统独特的未花费交易输出数据结构为研究人员提供了一种基于交易输入的关联地址识别方法。在诸多基于交易输入的关联规则中，多输入启发式算法（也称共同花费启发式算法）[27,33]最早被提出并得到广泛应用。多输入启发式算法实例如图7所示，该算法基于一笔交易中的多个输入地址

表3 关联地址识别技术比较

分类	文献	采用方法	实验效果
交易输入	文献[12]	提出一种基于集群交叉攻击的策略，如果能够识别出同一用户的至少2笔混币交易，就可以识别钱包中的所有地址和交易	以98%的准确率识别出用户钱包的地址集群
	文献[18]	基于多输入启发式规则，通过传递闭包扩展地址集合	将30 331 700个地址归类到272个用户实体
	文献[22]	先排除识别出的混币交易地址，再应用多输入启发式算法	准确揭示实体控制地址的同时，减少了过度聚类
	文献[40]	采用多输入启发式算法进行跨链地址的聚类分析	在shapeShift平台进行实验，共识别出2 895 445个节点和2 244 459条边，其中，最大的集群由12 868个地址组成
交易输出	文献[28]	新定义3个交易特征识别找零地址	产生383 904个不同的地址聚类，其中2 197个被标记，包括180万个地址
	文献[41]	利用地址脚本类型识别找零地址	在4.89亿笔交易中识别出1.22亿笔交易中的找零地址
	文献[42]	提出消费者启发式算法和最佳找零启发式算法	消费者启发式算法识别出69.26%的地址，最佳找零启发式算法识别出69.34%的地址
	文献[43]	结合阈值投票算法和随机森林分类器预测交易的找零输出	在3.1亿个标准交易中，使用99%的保守概率，识别出56亿个找零输出，占交易的50.24%
	文献[44]	提出了一种基于多条件识别的方法。如果满足特定条件，则将其识别为找零地址	识别出更多类型的找零地址，平均识别率提高了至少12.3%，并将地址聚类性能提高了5.7%
交易行为	文献[22]	提出识别剥离链交易中发送方的找零地址规则和识别连续锁时交易中的关联地址规则	识别95.81%的已知实体，同时减少20.59%的过度聚类现象
	文献[45]	引入剥离链交易、分发交易、中继交易等交易模式，识别比特币关联地址	对于不同集群选择的地址对，应用交易启发式后，这些地址对出现在相同交易中的概率增加了34%
	文献[46]	提出获得交易链集合的算法，提出了验证启发式算法和扩展启发式算法	验证启发式算法将一些嵌套服务的活动与交易所本身的活动区分开来；扩展启发式算法达到124.46倍的扩展因子
	文献[47]	用多输入启发式算法和找零启发式算法进行地址聚类，然后通过基于行为的聚类技术进行分析	当用户在每笔交易中使用新地址，研究人员能够识别40%的用户
	文献[48]	提出了一种多层启发式算法，通过结合区块链层和应用层信息，揭示隐藏的交易关系	大约31.68%的暗网市场评论数据与真实的比特币交易相匹配，并发现了与Silk Road 4相关的122个隐藏集群
	文献[49]	提出了基于多输入交易地址增量聚类算法，利用Petri网对网络进行建模，基于广度优先搜索（BFS）对新来地址进行增量聚类	在聚类效率方面，随着区块和地址数量的增加，基于多输入交易地址增量聚类（AICMT）算法的运行时间显著低于传统方法
	文献[50]	提出了替换交易启发式、一对一链启发式、梭形链启发式3种新的聚类启发式应用于未确认交易，进一步挖掘地址之间关联性	新的聚类启发式可以进一步减少聚类结果中实体的数量，在多输入启发式算法应用后，实体数量减少了9.8%

由同一用户控制的假设。本节根据表 3 对这些基于交易输入的关联地址识别工作进行梳理。

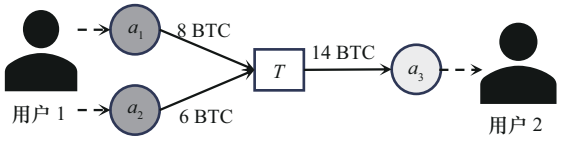


图 7　多输入启发式算法实例

最初的多输入启发式算法假设一笔交易的多个输入地址由同一用户控制。如图 8 所示，地址 a、b 和地址 b、c 分别出现在交易 T_1 和 T_2 中，传统的多输入启发式算法无法关联地址 a 和 c，这种局限性导致了地址关联的不完整。为此，Jourdan 等[18]提出了传递闭包的概念，其核心思想在于通过传递地址间的关联关系扩大关联范围。研究人员在 WalletExplorer 提供的标记数据集上应用传递闭包，成功将 30 331 70 个地址归类到 272 个不同的用户实体中。

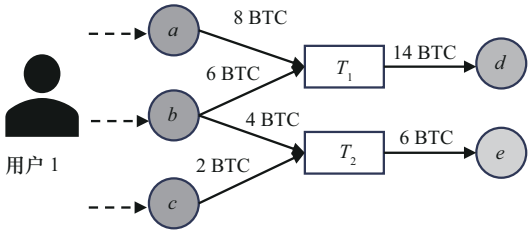


图 8　传递闭包

此外，混币交易[51]也是一种提高比特币交易匿名性的技术，得到了广泛使用。根据是否存在第三方管理机构，混币交易可分为中心化混币（如 Mixcoin[51]、Bitcoin Fog、和 Blindcoin）和去中心化混币（如 CoinJoin、CoinShuffle）。无论采用哪种混币交易，其核心思想都是将多个用户的交易合并到一笔交易中，从而破坏输入地址和输出地址，多输入地址和同一用户的对应关系，如图 9 所示，交易 1 和交易 2 分别由用户 1 和用户 2 发起，可以观察到交易输入和输出的对应关系，然而，当 CoinJoin 混币交易将两笔交易混淆在一笔交易中时，地址关联关系被破坏，在交易 TcoinJoin 中，无法确定输入和输出的对应关系。针对该问题，研究人员提出可以先进行混币交易检测，在排除混币交易的基础上再进行地址关联。例如，文献[22]提出了一种优化的方法，该方法在应用多输入启发式算法之前，会先利用简单的启发式规则排除来自 CoinJoinMess 和 JoinMarket 这 2 个平台的混币交易，对包含 382 个已知实体的比特币数据集进行评估，实验结果显示这一方法不仅更准确地揭示了同一实体控制的关联地址，还有效减少了过度聚类的现象。类似地，BlockSci[41]这一开源区块链分析平台在进行地址聚类时，会先检测 CoinJoin 交易，将相关地址排除，该平台运用多种启发式算法查找关联地址，包括多输入启发式、基于客户端软件或用户行为识别找零地址等技术。此外，文献[12]提出了一种基于集群交叉攻击的策略，即使采用了混币交易，只要观测到至少 2 笔由同一用户发起的交易，就可以通过分析交易模式和关联性来扩展识别出由该用户控制的一系列相关地址。在实验环节，使用集群交叉攻击策略识别准确率达到 98%。

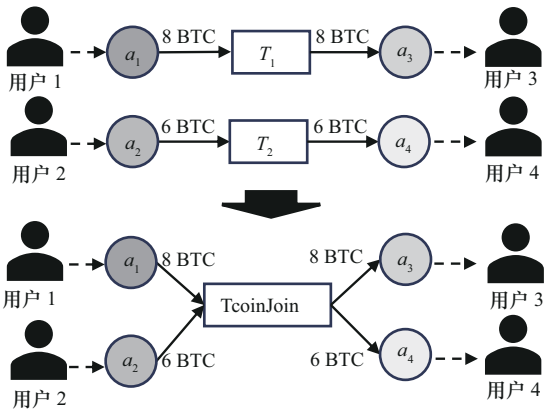


图 9　CoinJoin 交易实例

跨链技术的发展也为比特币地址关联分析带来挑战。跨链技术允许不同区块链的资产和信息交换，使得比特币可以与其他加密货币进行交互，这种交互扩大了地址关联分析的范围，增加了分析的复杂性。在跨链交易中，用户可将比特币转移到其他区块链上，再转回比特币系统。这种跨链转移行为使得地址关联分析不能仅局限于比特币区块链，还需考虑其他区块链上的地址和交易信息。当比特币通过跨链交易流入其他区块链时，其流向信息可能被隐藏或混淆。即便该笔资金回流到比特币系统，跨链前后的比特币地址也难以被关联识别。为解决此问题，文献[40]提出了基于共同关系启发式的跨链地址聚类方法。该方法通过识别出向同一地址发送资产或接收资产的多个地址，推断出这些地址具有某种社会关系。这种关系可能包括同一用户将来自不同区块链地址的资金集中到一个地址、不同用户向同一服务提供

商（如交易所）发送资金、2个用户具有资金往来等。研究人员通过这种方法在shapeShift平台进行实验，共识别出2 895 445个节点和2 244 459条边，其中，最大的集群由12 868个地址组成，集群中心地址来自CoinPayments.net。

3.2　基于交易输出的关联地址识别

通过识别交易输出中的找零地址识别关联地址是一种常用方法。在比特币中，交易输入与交易输出金额完全相等。用户进行交易时，其持有的UTXO金额往往无法精确匹配交易金额，此时需要选取多个UXTO作为输入，并将多余部分以找零的形式返回给自己。找零地址和交易输入地址可能属于同一用户。找零地址启发式算法实例如图10所示，用户1使用含有18 BTC的输入地址a_1发起交易，向用户2的地址a_3发送16 BTC后，剩余的2 BTC作为找零返回至用户1的新地址a_2。通过识别找零地址，可以推断地址a_1和地址a_2由同一用户控制。此外，找零地址启发式算法可以和基于交易输入的关联地址算法结合使用，以扩大地址聚类范围[52]。

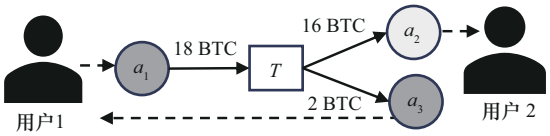


图10　找零地址启发式算法实例

Meiklejohn等[28]提出找零地址具备的4个特征，包括该地址所在交易不是造币交易、该地址在交易中首次出现、该地址未同时出现在交易输入和输出中、交易输出中除了该地址还有其他地址。在实验中，研究人员首先标记了涉及矿池、交易所、钱包服务等7种服务提供商的344笔交易。随后，基于这些被标记地址，应用上述规则识别出3 540 831个找零地址。

除上述方法外，研究人员还发现一种隐蔽的找零地址识别方法，通过分析交易输入和输出中的脚本特征来识别找零地址。文献[41]认为，找零地址通常是交易输出中唯一与所有输入地址的脚本类型相匹配的地址。如果交易输出中出现了用户从未用过的其他脚本，根据用户使用习惯推断，新类型脚本的地址更可能是第三方用户的地址，而非交易发起者的地址。例如，如果交易中的输入地址使用的是P2PKH脚本，而2个输出地址分别使用了P2PKH和P2SH这2种脚本，推断使用P2PKH脚本的地址很可能是找零地址。实验结果显示，基于脚本类型的找零地址识别方法对于一些比特币早期钱包效果明显，通过这一方法，研究人员在4.89亿笔交易中识别出1.22亿笔交易中的找零地址。

此外，研究人员发现交易金额也成为识别找零地址的重要特征。文献[42]提出如果一笔交易中的某个输出金额小于所有输入金额，则该输出很可能用于找零。例如，假设一笔交易的2个输入分别是2 BTC和3 BTC，2个输出分别是4 BTC和1 BTC。如果1 BTC是交易发起方支付给第三方的金额，那么3 BTC的输入就是不必要的，因为仅使用2 BTC的输入就能完成交易。因此，可以推断出4 BTC是支付给接收方的输出，而1 BTC是返回给发送方的找零。实验中，研究人员首先通过比特币钱包软件的漏洞收集了37 585个真实钱包及其对应地址。针对BitcoinCore、Electrum、MultiBit等较新的钱包软件，研究人员应用多输入启发式算法能够准确关联其中68.59%的地址。通过组合应用改进的找零地址识别算法，研究人员可以关联其中70.94%的地址。

文献[43]结合使用多种识别找零地址的启发式算法，如果超过一定数量的算法同时识别出某个地址为找零地址，则认为该地址为找零地址。实验室结果显示，当设定阈值为7，即至少有7个启发式算法同时识别出找零地址时，识别准确率达到94%。研究人员进一步对随机森林分类器和启发式算法的性能进行对比，发现在低误报率的情况下，随机森林分类器能够正确识别更多交易中的找零地址。此外，文中还使用了2个独立的数据集评估随机森林模型的性能。第一个数据集包含了与Locky和Cerber勒索软件相关的16 764笔交易，第二个数据集使用了GraphSense提供的273个不同实体的地址，随机森林模型在这2个数据集上的曲线下面积（AUC, area under curve）分别是99.6%和97.6%。

由于识别找零地址的识别率低、误报率高，Liu等[44]提出了多条件一次性找零地址识别方法。该方法首先排除Coinbase交易和自找零交易，对新地址从脚本类型、金额转换为美元情况以及金额多条件识别，如果满足特定条件，则将其识别为找零地址，并将其与其对应的输入地址聚为同一实体。

3.3　基于交易行为的关联地址识别

比特币交易存在多种交易模式，如剥离链交易、中继交易、造币交易、锁定时间交易等，现有的关联地址识别技术较少考虑交易模式上的区别。针对该问题，文献[45]提出了一种交易模式启发式算法，该算法在多输入启发式算法和找零启发式算法的基础上，引入区块链捕获的交易模式，以增强比特币地址关联的识别。文中还提出2种衡量比特币所有权识别性能的方法：一是通过基尼不纯性度量对已知身份的比特币地址进行聚类，确定聚类结果的纯度；二是通过比较“训练子集”和“测试子集”中的聚类对，确定在同一交易中再次出现的聚类对的比例。研究人员收集了4 600万比特币地址和4 650万笔交易进行验证，实验结果表明，基于聚类的基尼不纯性度量和基于标签的基尼不纯性度量具有一致性，并且在应用交易模式启发式后，数据集的结构发生了变化，同一实体的比特币地址变得更加统一。此外，对于不同集群选择的地址对，应用交易启发式后，这些地址对出现在相同交易中的概率增加了34%。

文献[22]通过深入分析剥离链交易与锁定时间交易的特点，提出了2种额外的启发式规则以优化地址聚类性能。第一种规则旨在识别剥离链交易中发送方产生的找零地址，其特征包括：首先，该找零地址接收的比特币数量比其他输出地址多；其次，该地址接收的比特币数额比其他输出地址多3个小数位；最后，该地址是首次出现的新地址。第二种规则针对具有相同有效时间设置的连续锁时交易中的关联比特币地址进行识别。当2个连续交易的输入地址满足以下条件时，它们被认为属于同一用户：首先，交易的所有输出已被花费；其次，2笔交易以完全相同的方式设定有效时间，即指定区块高度或时间戳。实验显示，在已知的382个实体所持有的比特币地址上，基于交易模式的关联地址识别算法[22]能够关联更多已知实体，占总数的95.81%，同时降低了过度聚类现象，平均减少20.59%。

除了文献[22,45]提出的启发式规则，研究人员进一步展开对剥离链交易模式的研究，增强用户地址的关联识别。剥离链交易通常从一个包含大额输入的交易开始，产生2笔输出：一笔用于实际交易的小额输出，另一笔作为找零的大额输出。如图11所示，剥离链通常起始于一个包含大额输入的交易（a_1），该交易将产生2笔输出，一笔小额输出用于完成和第三方的实际交易（a_2），另一笔大额输出作为找零返回给用户（a_3）。该大额找零将作为下一笔交易的输入并不断重复此过程。用户通过剥离链将资金不断分散到多个地址，提高了追踪难度。

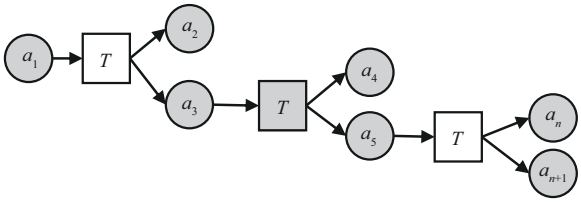


图11　剥离链示例

为应对这一挑战，Kappos等[46]提出了findNext和findPrev算法，基于用户行为和地址特征来识别同一用户控制的交易地址集合。还提出了followFwd与followBkwd算法，分别用于正向和反向追踪剥离链。这些算法综合考虑交易序列号、锁定时间、交易版本、是否采用隔离见证等交易特征，以及10种常见的比特币脚本类型等地址特征。实验显示，在区块链分析机构Chainalysis提供的241个真实用户地址上，该方法的假阳率仅为0.02%。

比特币关联地址识别技术和数据挖掘中的聚类算法可以结合使用[45]。聚类算法用于将数据集中的对象划分为不同的组或簇，使得组内的对象相似度较高，而组间的相似度较低。Androulaki等[47]模拟了大学环境中的比特币使用情况，首先使用多输入启发式算法和找零启发式算法进行地址关联，并将关联出的不同地址集群作为聚类算法的标准结果，随后利用K-Means和层次聚合聚类（HAC, hierarchical agglomerative clustering）算法对比特币交易进行组别划分，并将划分结果和多输入启发式算法进行比较。实验结果表明，即使用户每次使用新的地址进行交易，也有近40%的概率被关联。

为解决现有比特币地址聚类算法中的高误报率的问题，Kim等[48]提出了一种多层启发式算法，通过结合区块链层和应用层信息，揭示隐藏的交易关系。该算法首先利用暗网市场评论数据中的商品价值、发货时间和区块链交易数据，识别与评论数据对应的真实比特币地址，即匹配的地址（MA, matched address）。然后，使用多输入启发式算法

和找零启发式算法对MA进行聚类，得到多个初始集群。最后，分析集群随时间的变化，将来自同一评论数据的集群合并，揭示隐藏的交易关系。实验结果显示，大约31.68%的暗网市场评论数据与真实的比特币交易相匹配，并发现了与Silk Road相关的122个隐藏集群。这表明，多层启发式算法可以补充现有的聚类方法，并将误报率显著降低高达91.7%。

此外，文献[49]提出了AICMT算法。该算法首先利用Petri网对交易网络进行建模，将交易、币、地址等信息转化为Petri网中的变迁、库所和映射关系。然后，AICMT算法读取已有的地址聚类信息，并对新到来的地址进行增量聚类。通过遍历待聚类地址集合，判断每个地址是否已被聚类，并对未被聚类的地址进行BFS，找到与其相关的地址并进行聚类。最后，AICMT算法将更新后的聚类结果保存到存储中。用比特币网络前391 134个区块的数据进行实验，结果显示，与传统的地址聚类算法相比，AICMT算法在聚类精度方面表现相当。在聚类效率方面，随着区块和地址数量的增加，AICMT算法的运行时间显著低于传统算法。

不同于上述文献对比特币地址聚类的方法的研究，文献[53]提出一个新的指标——聚类率，通过聚类算法后，最终得到的簇数量与原始实体数量的比例，量化聚类后实体数量相对于初始实体数量的减少程度，从而衡量启发式算法减少实体数量的效果。通过对6种启发式算法的聚类率和时间演化进行分析，多输入启发式算法的效率最高，可以将实体数量减半，其余启发式算法可以将实体数量减少5%~15%。如果将多种启发算法组合使用可以进一步提高效率。例如，将共同输入启发式算法和找零启发式算法以及其他的启发式算法组合使用，可以将实体数量减少到70%。

上述地址聚类方法仅考虑比特币的已确认交易，文献[50]发现未确认交易也会显著影响聚类结果。对此，提出了替换交易启发式、一对一链启发式和梭形链启发式3种新的聚类启发式算法应用于未确认交易。结果表明，这3种新的启发式算法能够有效地挖掘地址之间的关联性，进一步减少聚类结果中实体的数量，在多输入启发式算法应用后，实体数量减少了9.8%。文献[54]提出一种基于社区检测的比特币用户重识别方法，通过构建身份提示网络并应用社区检测算法，能够更有效识别属于同一用户的多个地址。实验结果表明，与现有的算法相比，其召回率高达90%，F1分数为0.7。

3.4 关联地址识别技术研究现状总结

1) 比特币的关联地址识别技术和UTXO模型密切相关，无法应用于以太坊等基于账户模型的区块链系统，后者主要通过分析用户交易行为进行地址关联[55-56]。然而，比特币中观察到的有区分度的通用特征可为以太坊地址关联提供参考，如交易手续费、交易时间锁等。

2) 随着混币技术的引入，基于交易输入的关联地址识别算法可靠性下降。为此，研究人员通常先去除识别出的混币交易，再应用基于交易输入的关联地址算法。目前，尚无有效方法区分混币交易的多个输入。

3) 缺乏一个可供测试和效果比对的比特币真实地址标注数据集。研究人员一般通过小范围实验或典型事件案例获取标注地址，但大多数标注地址也未公开。

4 比特币资金链路追踪技术

比特币系统的UXTO模型只记录了交易发送方、交易接收方、交易金额，对于涉及多个发送方和接收方的交易，无法观察到特定发送方和接收方的具体资金流向。此外，混币、跨链等匿名性保护技术旨在进一步隐藏发送方和接收方的资金流动关系。本节从普通类型交易和特殊类型交易的资金链路追踪技术对现有文献进行梳理。

4.1 针对普通类型交易的资金链路追踪

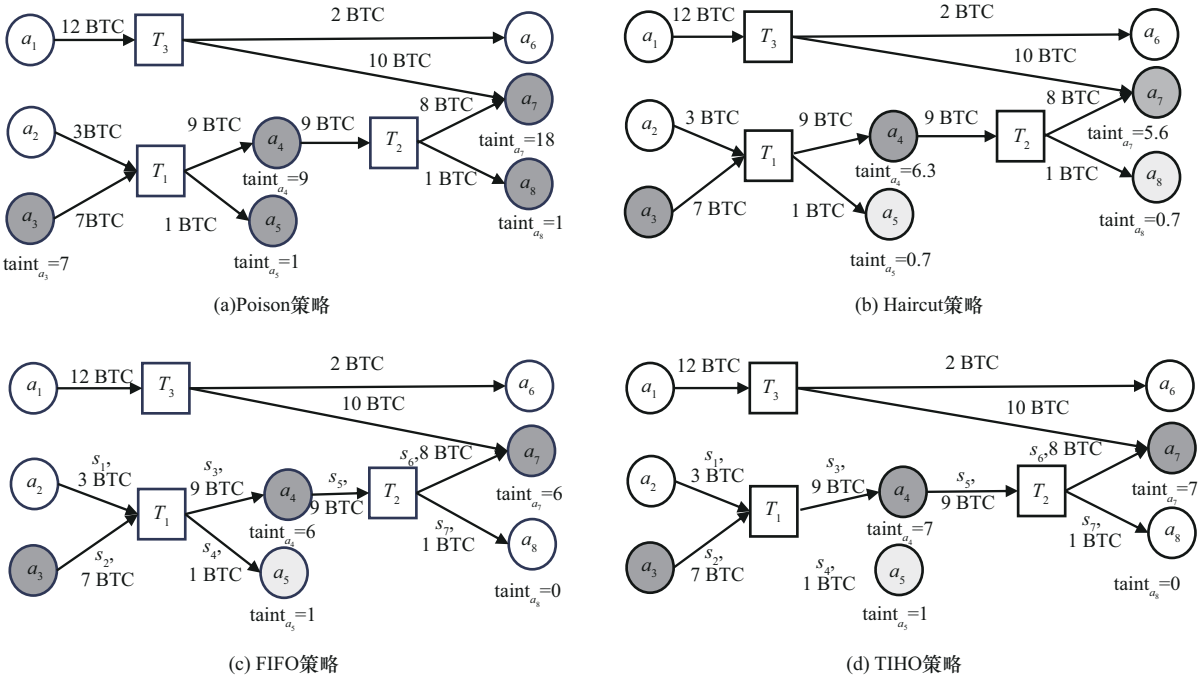
污点地址是与勒索软件、暗网等非法活动相关的地址，污点分析法通过追踪污点地址的交易路径来揭示资金流动关系。现有的比特币污染策略可大致分为4种：毒药（Poison）策略、理发（Haircut）策略、先进先出（FIFO, first in first out）策略和污点进最高出（TIHO, taint in highest out）策略。下面详细阐述这4种策略。

Poison策略是最简单的污染策略，由Möser等[57]提出，其核心思想是只要交易包含至少一个受污染的发送方地址，该笔交易中接收方收到的所有资金都是受污染资金。如图12(a)所示，圆圈表示比特币地址，方块表示比特币交易，箭头代表地址和交易的资金流向，黑色、灰色、白色分别表示

受污染、受部分污染、未受污染的比特币。图12(a)描述了Poison策略的核心思想。当起始受污染的7个比特币经过交易1后，地址4收到的9个比特币、地址5收到的1个比特币都被视为受污染比特币，随后，经过交易2和交易3后，地址7和地址6共收到19个受污染比特币。可以看出，Poison策略没有考虑交易输入中比特币的受污染程度，导致大量未受污染的比特币被错误归类为受污染比特币。基于此，在以太坊资金链路追踪中，Wu等[58]设计了截断Poison策略，层层追踪下游的交易和地址。

不同于Poison策略，Haircut策略[57]考虑了交易输入中比特币的受污染程度，不会将大量未受污染比特币错误归类。如图12(b)所示，受污染的7个比特币经过交易1后，地址4收到6.3($7\times\frac{9}{10}=6.3$)个受污染比特币，地址5收到受污染比特币为0.7($7\times\frac{1}{10}=0.7$)个，经过交易2后，地址8中受污染比特币为0.7（$\frac{6.3}{9}\times1=0.7$）个，地址7中受污染比特币为5.6（$\frac{6.3}{9}\times8=5.6$）个，由于交易3没有额外的受污染比特币作为输入，因此受污染比特币总额仍是7个。可以看出，Haircut策略中受污染比特币与未受污染比特币反复组合后，受污染比例在降低而污染总额不变。进一步地，文献[59]在原始Haircut策略的基础上引入了纯度的概念[60]，用于衡量一笔交易中包含受污染比特币的比例。当纯度低于某一特定阈值时，输出追踪链条所涉交易序列并停止追踪。研究人员将Haircut策略用于追踪2013—2016年由11个矿池开采出的新比特币，发现涉及相同矿池的追踪路径表现出高度相似性。

尽管Haircut策略追踪的受污染比特币总量不变，但仍会引入大量受污染地址及可疑链路，增加追踪工作量。为解决该问题，Anderson等[61]从1816年的克莱顿案件[62]中受到启发，提出针对比特币资金追踪的FIFO策略。如图12(c)所示，根据交易的先后顺序以及单笔交易内包含的输入输出的相对顺序，给每笔交易的UTXO分配一个序列号s_i用来标注相对顺序，因此，tx_1和tx_2中UTXO的相对顺序为$<s_1,s_2,s_3,s_4,s_5,s_6,s_7>$。根据FIFO策略，地址4优先收到来自地址2的未受污染的3个比特币，再收到来自地址3的受污染的6个比特币，地址5收到剩下的1个受污染比特币。同理，经过交易2后，地址7收到6个受污染比特币，地址8未收到受污染比特币。研究人员将Haircut策略和FIFO策略应用于Linode盗窃案和Flexcion攻击案件中进行评估比较。实验结果显示，在Linode盗窃案中，Haircut策略和FIFO策略分别追踪到2 694 051和371 544个受污染地址，在Flexcoin攻击事件中，两者分别追



(a)Poison策略　(b) Haircut策略

(c) FIFO策略　(d) TIHO策略

图12　4种污染策略示例

踪到1 429 794和18 208个受污染地址，FIFO策略较大减少了分析可疑追踪路径的工作量。此外，Ahmed等[63]实现了基于FIFO策略的比特币资金流前后向追踪系统，以一组被报告的涉非法活动比特币地址为起点，将追踪过程进行可视化。

Tironsakkul等[64]在FIFO策略的基础上提出了TIHO策略，既考虑交易输入输出顺序，也考虑输入输出金额。TIHO策略假设，交易中输出数额较大的地址是接收受污染比特币的主要目标，因此将受污染比特币优先分配给该地址。如图12(d)所示，在交易1中，由于地址4的输出金额大于地址5，来自地址3的7个受污染比特币优先分配给地址4，最终，地址4收到所有的7个受污染比特币。同理，经过交易2后，地址7收到所有的受污染比特币。研究人员利用2015年的Bter攻击案例[65]对Poison、Haircut、FIFO、TIHO这4种策略进行评估，结果显示，Poison和Haircut策略追踪到的交易和地址分别为1 256和55 099个，FIFO策略追踪到60笔交易和105个地址，TIHO策略追踪到44笔交易和83个地址，进一步减少了受污染的交易和地址数。实验结果显示FIFO和TIHO策略极大减少了可疑路径。

上述策略旨在寻找可疑地址和可疑交易路径，以揭示不同交易资金流路径的审计优先级。然而，找到的可疑路径需要追踪人员介入分析，无法自动输出从污染资金源地址到目标地址的资金流[66]。与污点分析不同，BFS算法通过遍历被污染比特币的资金流动，自动输出从污染源头到目的地址的资金流。BFS算法原理如图13所示，BFS算法以一笔受污染比特币作为起点，沿着资金流逐层扩展搜索，先访问受污染比特币的相邻节点，再依次访问这些相邻节点的相邻节点，以此类推，直到遍历完所有节点。Zhao等[67]对2014年2月3日到2014年2月7日的比特币交易构建了有向图，以Mt.Gox交易所被盗事件中涉及的14个地址作为起始节点，采用BFS算法追踪资金流。实验显示，遍历到10层相邻节点后，节点地址急剧下降，资金流最终汇集到单个节点。BFS算法目前在比特币追踪工具sky-trace和coinholmes上得到应用。

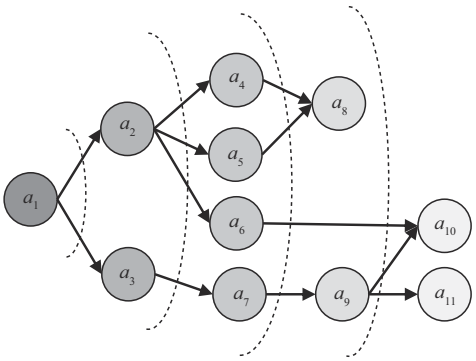
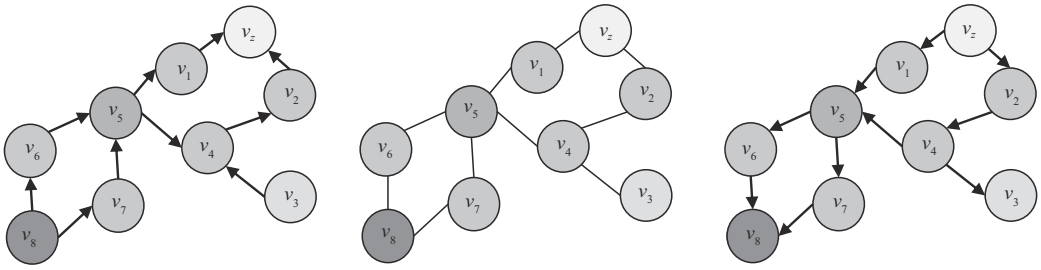


图13　BFS算法原理

文献[67]将BFS算法应用在比特币资金流上，但无法识别追踪路径上的可疑节点。针对此问题，文献[68]提出了基于翻转点的BFS算法，以检测追踪链路上的可疑节点，其中，图14(a)为针对比特币交易构建的有向图，图14(b)为对应的无向图，图14(c)是图14(a)的翻转图，图14中的资金流向均与图14(a)相反。研究人员首先在无向图中查找2个可疑地址的最短路径，例如地址v_1和v_2，随后查看该最短路径是否出现在有向图中，由于图14(a)不存在v_1到v_2的最短路径，说明在该最短路径存在资金流翻转，将资金流翻转的中心点视为翻转点，例如地址v_z。随后，针对每一个翻转点在翻转图上进行BFS，输出被多次遍历到的节点，这些节点被视为和翻转点具有较强协作关系的可疑节点，例如v_5和v_9。研究人员针对2015年的某事件中涉及的22个可疑节点构建资金流图，利用基于翻转点的BFS算



(a) 有向图　　(b) 无向图　　(c) 翻转图

图14　跨链交易模式

法进一步查找到23个可疑地址，并找到可疑地址间的102条资金链路。随后，该算法被集成到比特币可视化分析工具Biva中[69]。类似Biva，Sun等[70]设计了比特币交互可视化系统BitAnalysis，清晰地展示比特币在不同钱包间的流动。

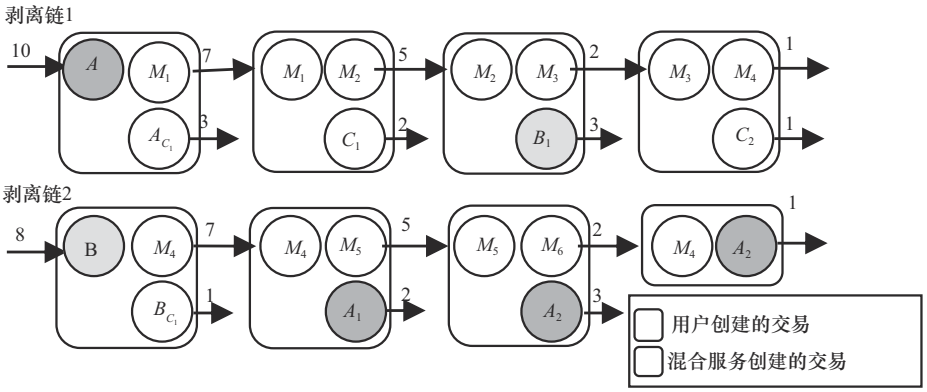
上述的资金追踪方法多采用正向搜索，文献[71]提出了一种“双向探索”的新型自动化比特币交易追踪技术，用于识别网络非法活动的资金关系。“双向探索”技术执行的是一个迭代过程，从一组属于网络活动反追的种子地址出发，执行如下3个步骤：一是通过多输入聚类以及从区块链提取的存取款交易来获取上下文信息。二是使用多输入标签数据库和交易所地址分类器来分类地址，识别出未知的交易所地址，防止图爆炸。三是用识别的交易地址更新交易图，并重复上述操作，直至结束，最后输出一个交易图。“双向探索”技术探索正向和反向的路径，从而可以返现仅通过正向探索无法发现的网络关系；此外，它结合了标签数据库和机器学习分类器，识别交易所地址，有效防止交易图爆炸。

除关注通用的资金追踪方法外，文献[72]重点探讨了在特定场景下的资金追踪，例如勒索软件赎金追踪。研究人员观察发现，Cerber和Locky等勒索软件通过每次使用新地址接收赎金来躲避追踪这些地址与其他比特币交易缺乏直接关联，使得传统多输入启发式方法难以应用。为应对这一难题，文献[72]提出了一种创新的微支付策略，通过向受害者赎金地址发送极小额的比特币，如0.001 BTC，并监测后续资金流动来追踪勒索软件资金流向。这一策略成功识别出一个包含7 093个地址的Locky勒索软件集群和一个包含8 526个地址的Cerber勒索软件集群，证实了即使在勒索软件采用新建赎金地址的策略下，通过微支付和资金追踪仍能有效实现地址聚类。

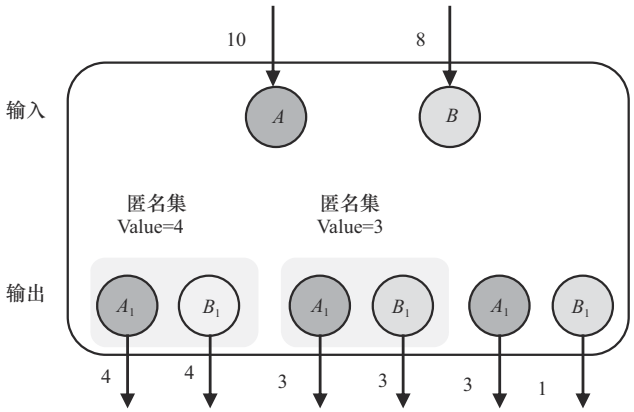
4.2　针对特殊类型交易的资金链路追踪

为进一步增强比特币交易的匿名性，各种混币、跨链等匿名性保护方案应运而生，意图隐藏交易发送方和接收方的关系。将涉混币、跨链等交易称为特殊类型交易，开展涉及特殊类型交易的资金追踪成为当前的研究热点和难点。由于用户在使用混币服务前后的资金总额保持不变，利用金额匹配的思想关联使用混币服务的用户交易地址成为可行方式。文献[73]提出了基于交易时间和交易数量的混币交易匹配方法。针对付款方和收款方只发生一笔交易的情况，提出基于概率模型的交易匹配方法。具体来讲，通过将混币交易按照时间先后顺序构成交易链条，研究人员可以利用最大似然估计方法将每个付款方与相应收款方配对，找出具有最高概率的配对方式。在实际应用中，用户为了混淆资金追踪链路，通常会采用多个付款方和多个收款方地址。针对该问题，研究人员提出基数模型，该模型的基本假设是付款方可以进行多次交易，但大多数交易是由几个固定的接收方完成的。具体来说，该模型涉及多个付款人和多个收款人，每个付款人和收款人分别有多笔交易，在匹配交易时，确保每个付款人的交易与多个不同的接收者匹配。该问题可以被制定为混合整数线性规划问题，可以通过Kannan算法等来求解。

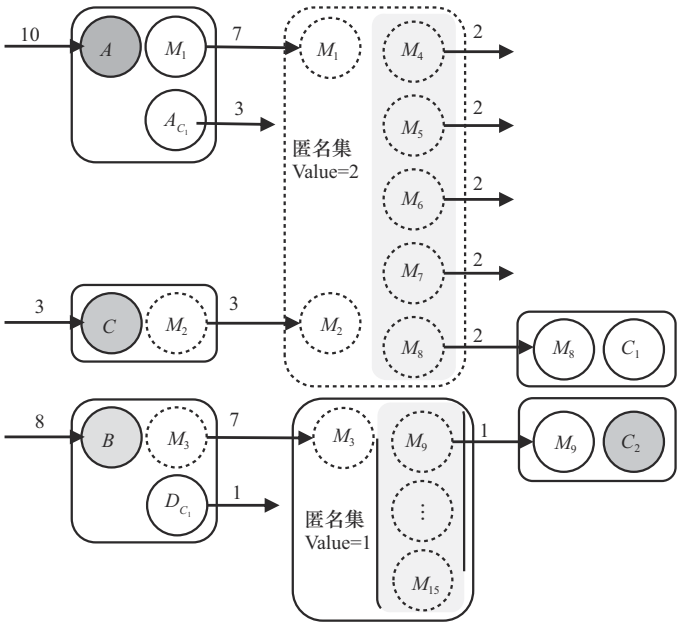
文献[74]通过对现有混币服务提供商开展实证研究，将现有的混币机制划分为交换机制和混淆机制2类。在交换机制中，每个混币交易通常只有2个输出，这些输出构成一条剥离链。混币服务提供商通过将不同用户的资金混淆在不同剥离链中，以此躲避地址关联。如图15(a)所示，地址M_1~M_6由混币服务提供商控制。用户A发起了剥离链1的起始交易，想把10个比特币转移到地址A_1、A_2和A_{C_1}中，其中，地址A_1和A_2分别持有4个和3个比特币。然而，混币服务提供商并未直接从地址A发起资金转移，而是利用用户B的交易（即剥离链2）将相应资金转移给地址A_1和A_2。可以看出，混币服务提供商通过交换不同用户的输入输出切断地址间的资金关联。与交换机制不同，混淆机制通过在交易中故意生成相同金额的多个输出实现多个用户的资金混淆。根据交易数量的不同，混淆机制进一步分为使用单笔交易的混淆机制和使用多笔交易的混淆机制。如图15(b)所示，用户A和用户B同时输入到一笔混币交易中，该交易产生了数额相同的2个匿名集，每个匿名集内的用户身份难以区分。具体来讲，尽管可以通过交易金额确定地址A_{C_1}由用户A控制，但是无法区分匿名集中地址的控制权。图15(c)显示了更复杂的使用多笔交易的混淆机制，通过两笔混淆交易将用户C的3个比特币转移到地址C_1和C_2中。在实验环节，研究人员通过与混币服务提供商进行交互以及调用公共API，搜集了来自Chipmixer、WasabiWallet、ShapeShift和Bitmix.biz 4个混币服务提供商的混币交易样本，应用针对混



(a) 交换机制



(b) 使用单笔交易的混淆机制



(c) 使用多笔交易的混淆机制

图15　交换机制和混淆机制示例

币机制的启发式识别算法，能够识别超过92%的混币交易。此外，研究人员对BinanceMayHack进行案例分析，通过追踪被盗比特币资金流揭示非法行为，追踪出攻击者共窃取了7 074 BTC，并使用Chipmixer进行混币行为，识别出的157笔混币交易总价值为4 797 BTC。

现有研究通常集中在单个区块链系统内的资金追踪。然而，用户可以通过跨链交易平台在不同区块链系统之间进行资金兑换和转移。跨链交易涉及多条链上的多个地址，这些地址之间没有直接的关联关系，因此跨链交易的检测和追踪非常困难。针对 shapeshift跨链交易平台运行机制如图16所示，在源链上，用户发起一笔向shapeshift账户转账的交易，期望按照一定汇率兑换成目标链上的加密货币，shapeshift收到用户转账后，在目标链上将一定数量的其他类型加密货币转给该用户。可以看出，跨链交易由2个独立区块链平台的交易配合实现，涉及用户和跨链交易平台的多个链上地址。针对该问题，Yousaf等[40]利用shapeshift跨链交易平台API获取跨链交易信息，包括源链币种、目标链币种、源链交易金额和交易时间。随后，利用交易金额和交易时间信息去源链上匹配源链交易，获取交易接收方地址，该地址即为shapeshift使用的地址。以shapeshift地址作为参数，研究人员再次利用shapeshift的交易接口获取该地址参与的跨链交易，找到该跨链交易的目标链交易哈希，从而成功匹配源链交易和目标链交易。实验结果显示，shapeshift在2017年11月至2018年12月共帮助用户完成280万笔跨链交易，涉及比特币、以太币、莱特币、Zcash等8种主流加密货币，该方法成功追踪到130万笔跨链交易。

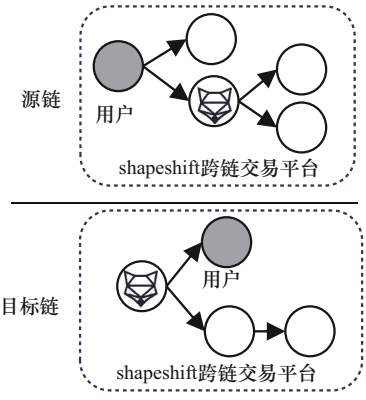


图16　针对shapeshift跨链交易平台运行机制

跨链交易的用户行为可以归纳为3种模式：穿透式、U形回环式和往返式。在穿透式跨链交易模式中，用户利用shapeShift完成一笔跨链交易，将一种链上的虚拟资产转换到另一条链上。这是最常见的跨链交易模式。在U形回环式跨链交易模式中，用户利用shapeshift完成第一笔跨链交易后，在极短时间内发起第二笔跨链交易，将相同数额的加密资产再交换回源链的不同地址。例如，用户将比特币转换为Zcash后，又迅速将相同数量的Zcash转换回比特币，共发现95 576个符合该模式的实例，其中在10 566个实例中，用户使用了相同的Zcash地址。往返式跨链交易模式与U形回环式跨链交易模式不同，往返式跨链交易的目标链地址不同，而源链地址可能相同，因此可以实现对发起跨链交易的源头用户的追踪。例如，在85 057个符合往返式跨链交易模式的实例中，共有10 490个实例使用相同源链地址。

4.3　资金链路追踪技术研究现状总结

目前的资金链路追踪技术如表4所示，根据分析对象的不同，资金链路追踪技术分为针对普通类型交易和特殊类型交易的资金链路追踪。在资金链路追踪方面，发现以下几个研究点。

1) 尽管去中心化跨链交易平台shapeshift声称具有较高安全性，但其接口字段设计不当，导致跨链交易信息泄露。文献[40]发表后，shapeshift修改了接口字段，使原有追踪方法失效。这表明，跨链资金追踪需要根据平台的特性提出针对性方法并适应平台演化。

2) 以太坊的资金追踪方法，如基于personal pagerank的子图搜索方法[66]和有偏随机游走的链路预测方法[75]，可以为比特币资金追踪提供借鉴。

3) 由于追踪策略假设和应用场景不同，一种策略难以适用于所有资金追踪任务。实践中，还需结合侦查人员的进一步分析。

5　比特币监管态势

由于去中心化、匿名等特性，比特币正成为一些非法活动的温床[76]。尽管国际社会已经认识到虚拟货币对金融稳定带来的巨大挑战，但各国在虚拟货币风险评估、政策目标和监管手段等方面存在显著差异。目前，国际社会尚未形成统一的加密货币监管规则。我国现行法律体系下对以比特币为代

表4 资金链路追踪技术总结

分类	文献	方法	效果
针对普通类型交易的资金追踪	文献[57]	提出Poison策略，只要交易包含至少一个受污染地址，交易中所有资金都是受污染资金	—
	文献[57]	提出Haircut策略，受污染比特币与未受污染比特币反复组合后，受污染比例降低而总额不变	—
	文献[61]	提出FIFO策略，考虑交易中多个输入及输出的先后顺序	通过Linode盗窃案和Flexcion攻击案件，FIFO策略受污染的地址明显比Haircut策略低，分别为371 544个和18 208个受污染地址
	文献[63]	实现基于FIFO策略的比特币资金流前后向追踪系统	以一组被报告的涉非法活动比特币地址为起点，将污点追踪过程进行可视化展示
	文献[64]	提出TIHO策略，交易中输出数额较大的地址是接收受污染比特币的主要目标	利用2015年的Bter攻击案例对Poison、Haircut、FIFO、TIHO 4种策略进行评估，TIHO策略追踪到44笔交易和83个地址
	文献[67]	基于图的比特币流量分析，利用BFS算法确定被盗比特币的最可能方向	以Mt.Gox交易所被盗事件中涉及的14个地址作为起始节点，BFS算法遍历到10层相邻节点后，将资金流汇集到单个节点
	文献[68]	提出了基于翻转点的BFS算法检测追踪链路上的可疑节点	针对2015年的某事件中涉及的22个可疑节点构建资金流图，查找到23个可疑地址，并找到可疑地址间的102条资金链路
	文献[70]	提出了一个用于比特币钱包的交互式可视化系统BitAnalysis，可有效地可视化比特币的流动图	—
	文献[71]	提出了一种“双向探索”的新型自动化比特币交易追踪技术，用于识别网络非法活动中的资金关系	在30个恶意软件家族中实验，双向探索可以发现关系中的93%，正向追踪中仅能发现74%。双向探索生成的交易图平均也比正向追踪的大3.6倍，此外，双向探索对关系检测的F1分数下降更慢
	文献[72]	提出微支付策略，向赎金地址发送极小额的比特币追踪勒索软件的资金流向	识别出包含7 093个地址的Locky勒索软件集群和包含8 526个地址的Cerber勒索软件集群
针对特殊类型交易的资金追踪	文献[40]	将跨链交易的用户行为总结为3种模式，并结合跨链交易平台shapeshift的API信息和源链、目标链检测跨链交易	shapeshift在2017年11月至2018年12月共帮助用户完成280万笔跨链交易，该方法成功追踪到130万笔跨链交易
	文献[73]	提出基于交易时间和交易数量的混币交易匹配方法	—
	文献[74]	将现有混币机制划分为交换机制和混淆机制2类，并确定不同混币机制下的混合过程	识别出在4个混币服务提供商的92%的混币交易；对Binance May Hack进行案例分析，识别出157笔混币交易，总价值为4 797 BTC

表的虚拟货币业态实施规制的多是规范性文件，主要有2013年中国人民银行、工业和信息化部等五部委联合印发的文件《关于防范比特币风险的通知》（以下简称《五部委通知》），2017年中国人民银行、中央网信办等七部委联合发布的《关于防范代币发行融资风险的公告》（以下简称《九四公告》），2021年中国人民银行、中央网信办等十部委联合发布的《关于进一步防范和处置虚拟货币交易炒作风险的通知》（以下简称《九二四通知》），以及2021年国家发展改革委、中共中央宣传部等部门发布的《关于整治虚拟货币“挖矿”活动的通知》（以下简称《整治“挖矿”通知》）等[77]。

从上述有关比特币监管的政策性规范文件来看，当前我国官方对比特币持一种“禁止式监管”的基本立场。首先，比特币在我国不具有货币属性。《五部委通知》明确比特币不能作为货币在市场上流通使用，比特币不是由货币当局发行，不是真正意义的货币。其次，我国严格禁止比特币的平台交易。《九四公告》将代币发行融资定义为一种未经批准非法公开融资的行为予以禁止，《九二四通知》则进一步将虚拟货币经营业务、衍生品交易定性为非法金融活动予以严格禁止。另一方面，禁

止为比特币交易提供虚拟货币交易场所。国家不允许金融机构和支付机构从事比特币买卖、登记、交易、清算等相关业务，故机构为比特币兑换法定货币或兑换其他虚拟货币而提供场所的行为是违法的[78]。即便如此，比特币作为一种虚拟商品，其作为一种财产性利益的价值仍然可以得到认可，亦即在上述规范性文件中并未明文禁止公民个人持有以及个人之间的比特币交易，从而允许公众在风险自担的前提下投资买卖比特币。同时，国家允许提供比特币登记、交易等服务的合法网站备案后运转，可以在除金融、支付机构外进行流通和交易。

其他国家或地区对比特币的监管政策也呈现出多元化和差异化特点。在美国，不同机构针对比特币制定了不同的监管规则。例如，美国证券交易委员会将部分加密货币界定为证券进行监管，而美国商品期货交易委员会则将加密货币定性为商品。欧盟对加密货币实施资本监管，要求交易平台实施KYC措施。英国通过立法要求加密货币公司进行反洗钱（AML, anti-money laundering）和KYC监管，对虚拟货币持积极开放态度。其他国家如德国、法国对加密货币也持较为友好的态度。日本很早就通过立法承认加密货币的法律地位，视情况对加密货币交易实施资本监管，是加密货币发展较为友好的国家之一。新加坡则将加密货币纳入清晰的法律框架进行监管。

虚拟货币监管离不开制度和技术的双重保障。在制度层面，通过金融行动特别工作组等国际组织的倡导，各国可以制定和实施统一的监管标准，包括要求交易平台实施AML和KYC规定，记录并审查客户身份信息[79]。在技术层面，融合链上内容、身份标签、网络结构等信息，利用人工智能等技术手段对用户身份和资金信息进行分析。

6　结束语

比特币等虚拟货币系统的监管问题已成为全球关注的焦点。为应对这一挑战，本文从用户身份识别、关联地址识别和资金链路追踪3个关键方面对比特币去匿名化技术进行系统性归纳和梳理。通过深入分析这些技术的原理、方法、实践效果及面临的挑战，本研究为虚拟货币监管提供了理论基础和实践指导。这对于打击虚拟货币开展非法活动、促进区块链技术的健康发展具有重要意义。

参考文献：

[1] SQUAREPANTS S. Bitcoin: a peer-to-peer electronic cash system[R]. 2008.

[2] FOLEY S, KARLSEN J R, PUTNIŅŠ T J. Sex, drugs, and Bitcoin: how much illegal activity is financed through cryptocurrencies?[J]. The Review of Financial Studies, 2019, 32(5): 1798-1853.

[3] 李旭东, 牛玉坤, 魏凌波, 等. 比特币隐私保护综述[J]. 密码学报, 2019, 6(2): 133-149.
LI X D, NIU Y K, WEI L B, et al. Overview on privacy protection in Bitcoin[J]. Journal of Cryptologic Research, 2019, 6(2): 133-149.

[4] 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(5): 1406-1434.
ZHANG A, BAI X Y. Survey of research and practices on blockchain privacy protection[J]. Journal of Software, 2020, 31(5): 1406-1434.

[5] 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.
ZHU L H, GAO F, SHEN M, et al. Survey on privacy preserving techniques for blockchain technology[J]. Journal of Computer Research and Development, 2017, 54(10): 2170-2186.

[6] 沈蒙, 车征, 祝烈煌, 等. 区块链数字货币交易的匿名性: 保护与对抗[J]. 计算机学报, 2023, 46(1): 125-146.
SHEN M, CHE Z, ZHU L H, et al. Anonymity in blockchain digital currency transactions: protection and confrontation[J]. Chinese Journal of Computers, 2023, 46(1): 125-146.

[7] ANTONOPOULOS A M. Mastering Bitcoin: unlocking digital cryptocurrencies[M]. Sebastopol: O'Reilly Media, 2015.

[8] 夏清, 窦文生, 郭凯文, 等. 区块链共识协议综述[J]. 软件学报, 2021, 32(2): 277-299.
XIA Q, DOU W S, GUO K W, et al. Survey on blockchain consensus protocol[J]. Journal of Software, 2021, 32(2): 277-299.

[9] OBER M, KATZENBEISSER S, HAMACHER K. Structure and anonymity of the Bitcoin transaction graph[J]. Future Internet, 2013, 5(2): 237-250.

[10] AL JAWAHERI H, AL SABAH M, BOSHMAF Y, et al. Deanonymizing Tor hidden service users through Bitcoin transactions analysis[J]. Computers & Security, 2020, 89: 101684.

[11] PORTNOFF R S, HUANG D Y, DOERFLER P, et al. Backpage and Bitcoin: uncovering human traffickers[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2017: 1595-1604.

[12] GOLDFEDER S, KALODNER H, REISMAN D, et al. When the cookie meets the blockchain: privacy risks of web payments via cryptocurrencies[J]. Proceedings on Privacy Enhancing Technologies, 2018 (4): 179-199.

[13] KOSHY P, KOSHY D, MCDANIEL P. An analysis of anonymity in Bitcoin using P2P network traffic[C]//Proceedings of the 18th International Conference on Financial Cryptography and Data Security. Berlin: Springer, 2014: 469-485.

[14] BIRYUKOV A, KHOVRATOVICH D, PUSTOGAROV I. Deanonymisation of clients in Bitcoin P2P network[C]//Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2014: 15-29.

[15] BIRYUKOV A, PUSTOGAROV I. Bitcoin over tor isnot a good idea[C]//

Proceedings of the 2015 IEEE Symposium on Security and Privacy. Piscataway: IEEE Press, 2015: 122-134.

[16] 高峰, 毛洪亮, 吴震, 等. 轻量级比特币交易溯源机制[J]. 计算机学报, 2018, 41(5): 989-1004.
GAO F, MAO H L, WU Z, et al. Lightweight transaction tracing technology for Bitcoin[J]. Chinese Journal of Computers, 2018, 41(5): 989-1004.

[17] SHAO W, LI H, CHEN M Q, et al. Identifying Bitcoin users using deep neural network[C]//Algorithms and Architectures for Parallel Processing: 18th International Conference. Berlin: Springer, 2018: 178-192.

[18] JOURDAN M, BLANDIN S, WYNTER L, et al. Characterizing entities in the Bitcoin blockchain[C]//Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW). Piscataway: IEEE Press, 2018: 55-62.

[19] HARLEV M A, SUN Y H, LANGENHELDT K C, et al. Breaking bad: de-anonymising entity types on the bitcoin blockchain using supervised machine learning[C]//Proceedings of the 51st Hawaii International Conference on System Sciences. Piscataway: IEEE Press, 2018: 1-14.

[20] YIN H H S, LANGENHELDT K, HARLEV M, et al. Regulating cryptocurrencies: a supervised machine learning approach to de-anonymizing the Bitcoin blockchain[J]. Journal of Management Information Systems, 2019, 36(1): 37-73.

[21] HAN W L, CHEN D J, PANG J, et al. Temporal networks based industry identification for Bitcoin users[C]//Proceedings of the 16th International Conference on Wireless Algorithms, Systems, and Applications. Berlin: Springer, 2021: 108-120.

[22] WANG K, PANG J, CHEN D J, et al. A large-scale empirical analysis of ransomware activities in Bitcoin[J]. ACM Transactions on the Web, 2022, 16(2): 1-29.

[23] WEBER M, DOMENICONI G, CHEN J, et al. Anti-money laundering in Bitcoin: experimenting with graph convolutional networks for financial forensics[J]. arXiv Preprint, arXiv: 1908.02591, 2019.

[24] LO W W, KULATILLEKE G K, SARHAN M, et al. Inspection-L: self-supervised GNN node embeddings for money laundering detection in Bitcoin[J]. Applied Intelligence, 2023, 53(16): 19406-19417.

[25] ALARAB I, PRAKOONWIT S, NACER M I. Competence of graph convolutional networks for anti-money laundering in Bitcoin blockchain[C]//Proceedings of the 2020 5th International Conference on Machine Learning Technologies. New York: ACM Press, 2020: 23-27.

[26] ELMOUGY Y, LIU L. Demystifying fraudulent transactions and illicit nodes in the Bitcoin network for financial forensics[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2023: 3979-3990.

[27] REID F, HARRIGAN M. An analysis of anonymity in the Bitcoin system[C]//Proceedings of the 3rd International Conference on Privacy, Security, Risk and Trust. Berlin: Springer, 2012: 197-223.

[28] MEIKLEJOHN S, POMAROLE M, JORDAN G, et al. A fistful of Bitcoins: characterizing payments among men with no names[C]//Proceedings of the 2013 conference on Internet measurement conference. New York: ACM Press, 2013: 127-140.

[29] RAPHAEL J. Denial of harm: sex trafficking, backpage, and free speech absolutism[J]. Dignity: A Journal of Analysis of Exploitation and Violence, 2017, 2(1): 1-10.

[30] ENGLEHARDT S, NARAYANAN A. Online tracking: a 1-million-site measurement and analysis[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2016: 1388-1401.

[31] MASTAN I D, PAUL S. A new approach to deanonymization of unreachable Bitcoin nodes[C]//Proceedings of the 16th International Conference on Cryptology and Network Security. Berlin: Springer, 2018: 277-298.

[32] MONACO J V. Identifying Bitcoin users by transaction behavior[C]//Proceedings of the Biometric and Surveillance Technology for Human and Activity Identification XII. Bellingham: SPIE Press, 2015: 25-39.

[33] RON D, SHAMIR A. Quantitative analysis of the full Bitcoin transaction graph[C]//Proceedings of the 17th International Conference on Financial Cryptography and Data Security. Berlin: Springer, 2013: 6-24.

[34] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[J].arXiv Preprint, arXiv: 1706.02216, 2017.

[35] LEE C, MAHARJAN S, KO K, et al. Toward detecting illegal transactions on Bitcoin using machine-learning methods[C]//Proceedings of the 2019 International Conference on Blockchain and Trustworthy Systems. Berlin: Springer, 2019: 520-533.

[36] ZHANG L J, ZHANG J J, TOYODA K, et al. A Bitcoin address multi-classification mechanism based on bipartite graph-based maximization consensus[C]//Proceedings of the 29th InternationalConference on Computational and Experimental Engineering and Sciences. Berlin: Springer, 2023: 473-487.

[37] TOYODA K, OHTSUKI T, MATHIOPOULOS P T. Multi-class Bitcoin-enabled service identification based on transaction history summarization[C]//Proceedings of the 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). Piscataway: IEEE Press, 2018: 1153-1160.

[38] VASSALLO D, VELLA V, ELLUL J. Application of gradient boosting algorithms for anti-money laundering in cryptocurrencies[J]. SN Computer Science, 2021, 2(3): 143.

[39] DUPONT J, SQUICCIARINI A C. Toward de-anonymizing Bitcoin by mapping users location[C]//Proceedings of the 5th ACM Conference on Data and Application Security and Privacy. New York: ACM Press, 2015: 139-141.

[40] YOUSAF H, KAPPOS G, MEIKLEJOHN S. Tracing transactions across cryptocurrency ledgers[J]. arXiv Preprint, arXiv: 1810.12786, 2018.

[41] KALODNER H, MÖSER M, LEE K, et al. BlockSci: Design and applications of a blockchain analysis platform[C]//Proceedings of the 29th USENIX Security Symposium. Berkeley: USENIX Association, 2020: 2721-2738.

[42] NICK J D. Data-driven de-anonymization in bitcoin[D]. Zürich: ETH-Zürich, 2015.

[43] MÖSER M, NARAYANAN A. Resurrecting address clustering in Bitcoin[C]//Proceedings of the 21st International Conference on Financial Cryptography and Data Security. Berlin: Springer, 2022: 386-403.

[44] LIU F, LI Z H, JIA K, et al. Bitcoin address clustering based on change address improvement[J]. IEEE Transactions on Computational Social Systems, 2023, 10(4): 1813-1825.

[45] CHANG T H, SVETINOVIC D. Improving Bitcoin ownership identifi-

cation using transaction patterns analysis[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50(1): 9-20.

[46] KAPPOS G, YOUSAF H, STÜTZ R, et al. How to peel a million: Validating and expanding bitcoin clusters[C]///Proceedings of the 31st USENIX Security Symposium. Berkeley: USENIX Association, 2022: 2207-2223.

[47] ANDROULAKI E, KARAME G O, ROESCHLIN M, et al. Evaluating user privacy in Bitcoin[C]//Proceedings of the 2013 17th International Conference on Financial Cryptography and Data Security. Berlin: Springer, 2013: 34-51.

[48] KIM M, LEE J, KWON H, et al. Get off of chain: unveiling dark web using multilayer Bitcoin address clustering[J]. IEEE Access, 2022, 10: 70078-70091.

[49] QIN F C, WU Y, TAO F, et al. Multi-input address incremental clustering for the Bitcoin blockchain based on Petri net model analysis[J]. Digital Communications and Networks, 2022, 8(5): 680-686.

[50] WANG K, CHENG Y K, TONG M W, et al. Exploring unconfirmed transactions for effective Bitcoin address clustering[C]//Proceedings of the ACM Web Conference 2024. New York: ACM Press, 2024: 1880-1891.

[51] BONNEAU J, NARAYANAN A, MILLER A, et al. Mixcoin: anonymity for Bitcoin with accountable mixes[C]//Proceedings of the 18th International Conference on Financial Cryptography and Data Security. Berlin: Springer, 2014: 486-504.

[52] LEE S, YOON C, KANG H, et al. Cybercriminal Minds: an investigative study of cryptocurrency abuses in the Dark Web[C]//Proceedings of the 2019 Network and Distributed System Security Symposium. Reston: Internet Society, 2019: 1-15.

[53] SCHNOERING H, PORTHAUX P, VAZIRGIANNIS M. Assessing the efficacy of heuristic-based address clustering for Bitcoin[J]. arXiv Preprint, arXiv: 2403.00523, 2024.

[54] REMY C, RYM B, MATTHIEU L. Tracking Bitcoin users activity using community detection on a network of weak signals[C]//Proceedings of the 15th International Workshop on Complex Systems and Networks. Berlin: Springer, 2017: 166-177.

[55] BÉRES F, SERES I A, BENCZÚR A A, et al. Blockchain is watching you: profiling and deanonymizing ethereum users[C]//Proceedings of the 2021 IEEE International Conference on Decentralized Applications and Infrastructures (DAPPS). Piscataway: IEEE Press, 2021: 69-78.

[56] XIA P C, WANG H Y, GAO B Y, et al. Trade or trick?[J]. Proceedings of the ACM on Measurement and Analysis of Computing Systems, 2021, 5(3): 1-26.

[57] MÖSER M, BÖHME R, BREUKER D, Towards risk scoring of Bitcoin transactions[C]//Proceedings of the 18th International Conference on Financial Cryptography and Data Security. Berlin: Springer, 2014: 16-32.

[58] WU J J, LIN D, FU Q S, et al. Toward understanding asset flows in crypto money laundering through the lenses of ethereum heists[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 1994-2009.

[59] TOVANICH N, CAZABET R. Pattern Analysis of Money Flows in the Bitcoin Blockchain[C]//Proceedings of the 2022 International Conference on Complex Networks and Their Applications. Berlin: Springer, 2023: 443-455.

[60] BATTISTA G D, DONATO V D, PATRIGNANI M, et al. Bitconeview: visualization of flows in the Bitcoin transaction graph[C]//Proceedings of the 2015 IEEE Symposium on Visualization for Cyber Security (VizSec). Piscataway: IEEE Press, 2015: 1-8.

[61] ANDERSON R, SHUMAILOV I, AHMED M. Making Bitcoin legal[C]//Proceedings of the 26th International Workshop on Security Protocols XXVI. Berlin: Springer, 2018: 243-253.

[62] 孙鹏. 金钱“占有即所有”原理批判及权利流转规则之重塑[J]. 法学研究, 2019, 41(5): 25-43.
SUN P. A critique on the principle of “money possession that is all” and reconstruction of the rule of “money right transfer”[J]. Chinese Journal of Law, 2019, 41(5): 25-43.

[63] AHMED M, SHUMAILOV I, ANDERSON R. Tendrils of crime: visualizing the diffusion of stolen Bitcoins[C]//Proceedings of the 5th International Workshop on Graphical Models for Security. Berlin: Springer, 2019: 1-12.

[64] TIRONSAKKUL T, MAAREK M, EROSS A, et al. Probing the mystery of cryptocurrency theft: an investigation into methods for cryptocurrency tainting analysis[C]//Cryptocurrency Research Conference. Berlin: Springer, 2019:1-28.

[65] MCCORRY P, MÖSER M, ALI S T. Why preventing a cryptocurrency exchange heist isn’t good enough[C]//Security Protocols XXVI: 26th International Workshop. Berlin: Springer, 2018: 225-233.

[66] WU Z Y, LIU J L, WU J J, et al. TRacer: scalable graph-based transaction tracing for account-based blockchain trading systems[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 2609-2621.

[67] ZHAO C, GUAN Y. A graph-based investigation of Bitcoin transactions[C]//Proceedings of the 2015 IFIP International Conference on Digital Forensics. Berlin: Springer, 2015: 79-95.

[68] PHETSOUVANH S, OGGIER F, DATTA A. EGRET: extortion graph exploration techniques in the Bitcoin network[C]//Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW). Piscataway: IEEE Press, 2018: 244-251.

[69] OGGIER F, PHETSOUVANH S, DATTA A. BiVA: Bitcoin network visualization & analysis[C]//Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW). Piscataway: IEEE Press, 2018: 1469-1474.

[70] SUN Y J, XIONG H, YIU S M, et al. BitAnalysis: a visualization system for Bitcoin wallet investigation[J]. IEEE Transactions on Big Data, 2023, 9(2): 621-636.

[71] GOMEZ G, MORENO-SANCHEZ P, CABALLERO J. Watch your back: identifying cybercrime financial relationships in Bitcoin through back-and-forth exploration[C]//Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2022: 1291-1305.

[72] HUANG D Y, ALIAPOULIOS M M, LI V G, et al. Tracking ransomware end-to-end[C]//Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2018: 618-631.

[73] CHEN L, XU L, SHAH N, et al. Unraveling blockchain based cryptocurrency system supporting oblivious transactions: a formalized approach[C]//Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts. New York: ACM Press, 2017: 23-28.

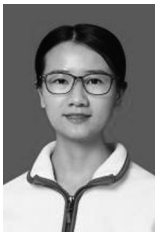
[74] WU L, HU Y F, ZHOU Y J, et al. Towards understanding and demystifying Bitcoin mixing services[C]//Proceedings of the Web Conference 2021. New York: ACM Press, 2021: 33-44.

[75] LIN D, WU J J, XUAN Q, et al. Ethereum transaction tracking: Inferring evolution of transaction networks *via* link prediction[J]. Physica A Statistical Mechanics and Its Applications, 2022, 600: 127504.
[76] KETHINENI S, CAO Y. The rise in popularity of cryptocurrency and associated criminal activity[J]. International Criminal Justice Review, 2020, 30(3): 325-344.
[77] 江溯. 非法获取虚拟货币行为的定性[J]. 人民司法, 2023(11): 15-18.
JIANG S. Nature determination of illegally Gaining other's virtual currency[J]. People's Judicature, 2023(11): 15-18.
[78] 郭志浩, 候柔倩. 论虚拟货币的双重法律属性[J]. 河南社会科学, 2024, 32(3): 78-86.
GUO Z H, HOU R Q. On the dual legal properties of virtual currencies [J]. Henan Social Sciences, 2024, 32(3): 78-86.
[79] FORCE F A T. Updated guidance for a risk-based approach to virtual assets and virtual asset service providers[R]. 2021.

[作者简介]

程杰（1994– ），女，河北秦皇岛人，博士，中国人民公安大学讲师，主要研究方向为经济犯罪侦查、数据分析、可信验证等。

金伟（1994– ），女，北京人，博士，中国信息通信研究院工程师，主要研究方向为网络安全政策标准、访问控制。



夏清（1994– ），女，重庆人，博士，中国科学院软件研究所副研究员，主要研究方向为区块链智能合约、链上攻击检测、大模型应用。

李淼（1993– ），男，湖南郴州人，博士，中国人民公安大学讲师，主要研究方向为刑法中的共同犯罪、网络犯罪。

戴韡（1982– ），男，福建南安人，博士，中央财经大学副教授，主要研究方向为金融科技、区块链、数字货币和不确定理论。

张亚丰（1993– ），男，山西太原人，中国科学院软件研究所工程师，主要研究方向为区块链技术和大模型应用。

戴蓬（1971– ），男，山东济南人，中国人民公安大学教授，主要研究方向为经济犯罪侦查。

李玉成（1961– ），男，浙江温州人，中国科学院软件研究所研究员，主要研究方向为并行软件与计算科学、大数据分析、区块链技术。