BATscope: 比特币恶意地址及混币交易识别

王大宇[1], 殷婷婷[2], 李 赟[2], 秦嗣量[3], 任 歆[4], 罗夏朴[5],
王浩宇[6], 尹 霞[1], 张 超[2]

[1]清华大学 计算机科学与技术系 北京 中国 100084
[2]清华大学 网络科学与网络安全研究院 北京 中国 100084
[3]中国科学院大学 网络空间安全学院 北京 中国 100049
[4]厦门大学 软件工程系 厦门 中国 361005
[5]香港理工大学 计算系 香港 中国
[6]华中科技大学 网络空间安全学院 武汉 中国 430074

摘要 *比特币作为第一个也是最主流的基于区块链技术的数字货币, 吸引了越来越多用户的关注和投资。因为匿名性和去中心化的特点, 比特币也是不法分子常用的洗钱工具。据报道, 最近几年比特币已被用于许多案件, 包括黑客、暗网市场、资金走私、诈骗和勒索。为了打击此类恶意行为, 准确识别比特币地址的类型和比特币交易目的尤为重要。然而, 现有的解决方案仅能部分地解决这个问题, 并且在识别准确率上表现不佳。在本文中, 我们提出了一种基于机器学习的解决方案 **BATscope**, 可以准确地识别比特币地址的类型及一些交易的目的(例如, 混币交易)。其核心是通过一些可靠的启发式方法和一种新颖的先导预测方法, 可以自动化的迭代增加训练集中的比特币地址, 从而不断反馈给模型再次训练, 稳定提升机器学习模型的性能。评估结果表明, **BATscope** 可以在公开数据集中以 0.99 的精度识别基于混淆的混币交易, 并在识别比特币地址的类型(例如, 恶意地址)中达到 0.9621/0.9567 的 Micro/MacroF1 分数, 远高于现有的解决方案。此外, 结果还表明我们的启发式方法可以有效地增强可靠的地址标签数据, 先导预测也可以准确的进行纠错并进一步提升模型性能。我们利用 **BATscope** 进一步分析了混币交易, 揭示了混币行为和恶意地址之间的关系。为了证明其鲁棒性和实用性, 我们还使用 **BATscope** 来验证已知恶意地址, 并帮助执法部门分析未知地址并提供线索。进一步证明在实际应用中, **BATscope** 的结果是可靠的。

关键词 *比特币; 地址分类; 机器学习
中图法分类号 TP309.2 DOI 号 10.19363/J.cnki.cn10-1380/tn.2023.07.01

BATscope: Demystifying Malicious Addresses and Mixing Transactions in Bitcoin

WONG Taiyu[1], YIN Tingting[2], LI Yun[2], QIN Siliang[3], REN Xin[4], LUO Xiapu[5],
WANG Haoyu[6], YIN Xia[1], ZHANG Chao[2]

[1] Department of Computer Science, Tsinghua University, Beijing 100084, China
[2] Institute for Network Science and Cyberspace, Tsinghua University, Beijing 100084, China
[3]School of Cyber Science and Technology. University of Chinese Academy of Sciences, Beijing 100049, China,
[4]Department of Software Engineering, Xiamen University, Xiamen 361005, China,
[5]Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China,
[6]School of Computer Science, Beijing University of Posts and Telecommunications Beijing 100876, China,

Abstract *Bitcoin, the first and the most popular Blockchain-based cryptocurrency, has attracted more and more users and investment. Because of the anonymity and decentralization of the Bitcoin, it has become one of the most common ways for malicious entities to launder money. In recent years, it is reported that Bitcoin has been used as a medium in many illegal actions, including cyberspace hacking, darknet marketplaces, money smuggling, scams, and blackmails. To combat such malicious behaviors, it is crucial to identify the roles of Bitcoin addresses and purposes of Bitcoin transactions of interest. However, existing solutions only partially addressed this problem and had poor performance in recognition. In this paper, we propose a novel machine learning (ML) based solution **BATscope** to address this problem. **BATScope** can accurately identify the Bitcoin address type and the purpose of some transaction behaviors (e.g., mixing transactions). At the core, it iteratively and automatically augments the training set of Bitcoin address labels with some *reliable heuristics* and a novel *pilot prediction* method, and thereby continuously promotes the ML model's performance. Evalua-

通讯作者: 张超, 博士, 副教授, Email: chaoz@tsinghua.edu.cn。
本课题得到国家重点研发计划资助(No. 2021YFB2701000); 国家自然科学基金资助(No. 61972224, No. U1736209)。
收稿日期: 2021-12-29; 修改日期: 2022-03-15; 定稿日期: 2023-04-18

tion results showed that **BATscope** can recognize obfuscating-based mixing transactions with a precision of 0.99 in the public dataset and recognize the type of Bitcoin addresses (e.g., attackers) with a micro/macro-F1 score of 0.9621/0.9567, much higher than existing solutions. Besides, the result also proved that our *reliable heuristics* can augment valid address labels with high confidence and *pilot prediction* corrected mislabeled addresses to further promote model's performance. We use **BATscope** to further analyze the mixing transactions in Bitcoin, which revealed the relationship between malicious addresses and mixing transactions. To demonstrate its robustness and usefulness, we also used **BATscope** to verify known malicious addresses and help law enforcement authorities analyze unknown addresses and close cases. The case studies showed that the result of **BATscope** is reliable in practical application.

Key words * bitcoin, address classification, machine learning

1 引言

比特币[1] 作为第一个也是最主流的基于区块链技术的数字货币, 吸引了越来越多用户的关注和投资。在比特币生态系统中, 每个用户或实体都可以拥有一个或多个比特币地址(类似于银行账号)。然而, 与银行账号不同, 比特币地址是通过密码学算法生成的, 与用户的现实真实身份没有直接关系, 因此提供了一定程度的匿名性。

除了匿名性外, 比特币还以去中心化的方式运作, 并拥有很高的价值, 因此比特币也成为了犯罪分子洗钱和逃避政府部门监管的主要手段之一。据报道, 比特币已被用作许多非法行为的工具, 包括黑客、暗网市场、资金走私、诈骗和勒索。例如, 2020 年 7 月, 许多知名推特账户被黑客攻击, 并利用比特币进行诈骗[2], 造成超过 110000 美元的损失。2021 年 5 月, Colonial Pipeline 被勒索软件勒索, 支付了 75 个比特币(或 440 万美元)用于数据恢复[3]。2019 年, 被中国警方破获的 PlusToken 庞氏骗局[4]造成的损失超过 42 亿美元。

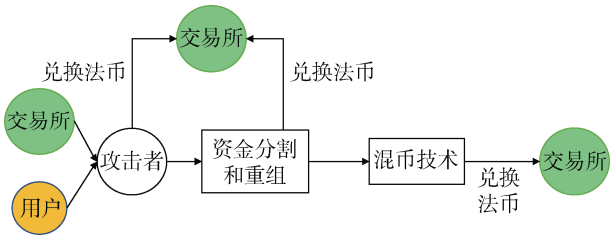


图 1 比特币洗钱简化过程

Figure 1 Illustration of a classic money laundering process in Bitcoin

圆形和矩形节点分别代表用户/地址和交易。

Circle and rectangle nodes represent different types of users/addresses and transactions, respectively.

为了打击此类违法行为, 我们需要了解比特币地址的类型和比特币交易的目的来追踪非法资金(即比特币)的流动。这是个非常具有挑战性的任务。图 1 展示了比特币洗钱的简化流程。除了比特币的匿名性和去中心化外, 恶意实体还利用分割、聚合、混币、剥离链(Peeling Chain)交易等技术来进一步增加追踪资金流的难度。因此, 识别地址(例如攻击者、交易所、个人钱包等)和交易意图(例如混币交易等)在比特币监管中的作用至关重要, 在流程中可以看到, 攻击者, 交易所, 混币交易等都可以对分析起到关键的作用。

然而, 现有的解决方案仅能部分地解决这个问题, 并且在很多情况下表现不佳。一般来说, 目前有两种类型的解决方案, 即基于启发式和基于机器学习(Machine Learning , ML)的。基于启发式的解决方案[5-9]利用某些启发式方法, 例如比特币白皮书[1] 中提出的多输入启发式方法, 可以对同一用户或者实体控制的地址进行聚类。在已知集群中某一个地址标签的情况下, 可以将地址将标签从已知地址扩展到集群剩余的未知地址。这种启发式方法通常具有误报, 可能会被攻击者绕过, 从而导致不准确的结果[10]。更不用说这样的解决方案只能将标签扩展到有限数量的地址, 而无法识别其余地址(不属于同一集群)。基于 ML 的解决方案利用手动提取的特征(例如文献[11-12])或图神经网络(例如文献[13])来表征区块链地址和交易。此类解决方案受限于训练数据规模小, 泛化性能较差。

在本文中, 我们提出了一种新的基于 ML 的解决方案 BATscope 来解决这个问题。BATscope 将启发式方法和机器学习相结合, 它可以自动化地迭代增加 ML 模型的训练集, 以不断提升模型的性能。具体来说, 我们应用以下两条规则来扩充训练数据。

首先, 它改进了现有的启发式方法, 使用可靠的启发式方法将标签从已知地址扩展到未知地址。请注意, 在对混币交易进行操作时, 某些启发式方法可能会被破坏, 例如, 多输入启发式方法不适用于混币交易。因此, 为了使启发式方案更可靠, BATscope 会利用交易的输出金额分布特征来识别混币交易并在应用启发式方法时跳过他们以避免误报。

其次, BATscope 采用了一种新颖的先导预测方法来准确地标记未知地址的类型, 从而进一步增加训练数据。它 (1) 首先使用当前的 ML 模型预测未知地址的类型, (2) 然后使用一种纠错机制来纠正预测结果, 并且(3)使用新标记的未知地址来扩充训练数据。第二步可以阻止不准确的预测数据及其派生数据污染模型。

上述数据增强过程可以重复执行以获取足够的训练数据并不断提高模型的性能。我们已经实现了 BATscope 的原型并将其应用于比特币区块链。评估结果表明, BATscope 能够以 0.99 的精度识别混币交易, 远高于最先进的解决方案[14]。通过查询最先进的商业软件, 我们确认 BATscope 使用的可靠启发式可以有效地扩展未知地址的标签, 证明启发式算法确实可靠。此外, 通过应用先导预测方法, BATscope 可以识别比特币地址(例如, 攻击者等)的类型, 其 Micro/Macro F1 可以达到 0.9621/0.9567, 远高于现有的基于 ML 的解决方案[11-12, 15]。最后, 我们将 BATscope 应用于实际场景以验证其鲁棒性和实用性, 并表明它可以(1)正确识别在最近的安全事件中使用的已知(但不在训练集中)恶意地址[16] 和(2)成功帮助执法部门识别未知地址。

在本文中, 我们做出以下贡献。

- 我们提出了一种新颖的基于机器学习的解决方案 BATscope, 能够识别混币交易并利用它来高精度地识别比特币地址的类型。
- 我们从各种开源信息收集了一组包括 43k 比特币地址标签数据, 据我们所知这是目前最大的数据集, 并会在未来将其开源。
- 我们提出利用输出金额分布来可靠地识别混币交易, 并提出了基于它的可靠的启发式方案来扩展地址标签。
- 我们提出了一种新的先导预测方法, 通过利用当前模型和特殊的纠错机制将标签扩展到未知地址。我们可以将标签地址从 43k 扩展到 1.6M 以上, 并且因此大大提高了模型的性能。
- 我们将 BATscope 应用于最近攻击事件中报告的一些已知恶意地址和执法部门查询的未知地址, 并证明了其鲁棒性和实用性。

2 背景

2.1 比特币

比特币是第一个也是目前最主流的加密数字货币之一, 由中本聪于 2008 年提出[1]。比特币通过 P2P 网络以及工作量证明(POW)的分布式共识协议解决了双花问题。每个比特币用户或实体控制的账户都由一个地址标识, 并且一个用户可以有多个地址。每个地址都是通过对用户公钥进行一系列密码学算法和不可逆的哈希计算得到的, 这种算法的碰撞概率极低。此外, 比特币社区鼓励用户为接收一笔交易生成新的地址, 这使得交易几乎无法被追踪。因此, 很难将用户的某一个地址关联到她/他的其他地址, 也很难关联到她/他的真实身份。通过这种方式, 比特币提供了一定程度的匿名性。

比特币的支付是通过将比特币从(多个)输入地址转移到(多个)输出地址的交易进行的。一笔交易需要所有输入地址的私钥对交易进行数字签名才能成功执行。由于私钥只有输入地址的所有者知道, 因此通常假设多输入交易中的输入地址都属于一个用户, 但这在实践中很可能不正确的。

矿工是比特币网络中的特殊节点, 负责验证交易并将它们打包在一个区块中, 这些区块将链接在一起形成分布式账本。挖矿的过程是一种计算哈希值的过程, 当矿工通过应用不同随机数计算出一个满足条件的区块哈希值的时候(如哈希值的前几位都为 0), 可以认为产出了一个可用的新区块, 这个新区块将会被连接到当前区块链的末尾。矿工在生成区块时将获得奖励的比特币, 除了挖矿的奖励还有用户交易时向矿工支付手续费用。通过这种方式, 许多节点愿意充当矿工的角色, 这使得比特币能够以公平和去中心化的方式运作。我们可以从奖励交易(Coinbase 交易)中精确推断出矿工的地址。

近些年来, 为了让挖矿的概率变得更高, 许多矿工将算力集合起来, 形成一个具有强大算力的新实体-矿池。矿池将每次挖矿的收益按照参与矿工的算力按比例进行分配。

2.2 混币交易

比特币使用 UTXO 模型, 该模型要求交易的输入必须是先前交易的输出, 每个 UTXO 上都包含一个地址信息(除 OP_RETURN 外), 从而允许用户跟踪比特币的流动。在 UTXO 的模型下, 由于任何一笔 UTXO 的金额不能分割和合并, 因此对于需要找零的情况, 用户一般会生成一笔两个输出的交易, 其中一个输出的目标地址是真实接受支付的地址, 另一个地址是属于输入支付方控制的地址, 用作接受多余的零钱, 并形成一个新的 UTXO。通过 UTXO, 交易和交易之间互相连接形成一个复杂的交易网络。此外, UTXO 还允许交易的输入端和输出端有多个地址相同的输入或输出, 形成聚合交易或者分片交易, 使得比特币的交易方式更加灵活。

混币交易使用户可以将自己的资金与其他用户的资金混合, 以隐藏其资金的流动并保护交易参与者的隐私。目前有两种流行的混合技术[14]: 基于交换和基于混淆的混币技术。基于交换的混币依赖于受信任的第 3 方来交换来自不同用户的输入和输出, 以保持输入输出关系的匿名性并打破 UTXO 的可追溯性。基于混淆的混币通过将多个用户的交易合并为一个交易来混淆输入和输出之间的匹配关系, 并可以用分布式的协议实现。

一种常见的混淆解决方案是 CoinJoin[17], 它被现在主流的混币服务提供商广泛使用, 如 Wasabi 钱包[18]、JoinMarket [19]和 Samourai 钱包[20]。为了进一步增强匿名性, 这些混币服务应用匿名集(一组具有相等值的输出)使得多个输出不可区分, 导致确定输入和输出之间的支付关系变得更加困难。图 2 展示了一笔标准的混币交易, 它将两笔独立的交易合并为一笔, 使得输入和输出地址的关联关系变得模糊复杂。此外, CoinJoin 还可以通过匿名集, 使得两笔实际的支付输出金额变得相等, 如图中右侧的 C 与 D, 在其输出金额相等的情况下我们是无法区分 C 和 D 的, 因此 CoinJoin 可以打破 UTXO 的可追溯性, 增强参与用户的隐私性。

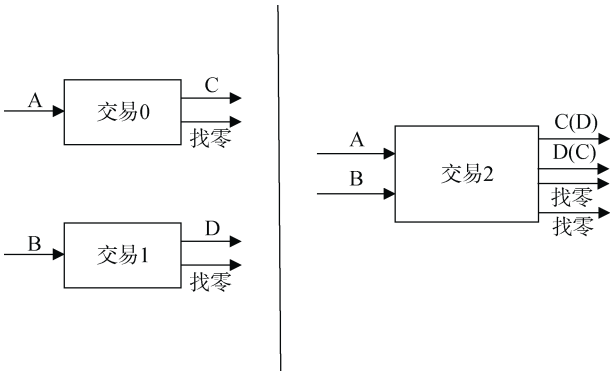


图 2 CoinJoin 交易示例
Figure 2 Case of CoinJoin transaction

除了 CoinJoin 交易, 还有一种类似的交易手段称为 Chip Generation, 被 Chipmixer 用于混币服务。类似于 CoinJoin, 它依旧是将多笔交易混合到一起, 并且具有多组输出相等的金额, 但是, 其输出金额有独特的特点, 金额一般为 2 的幂, 如 0.02BTC, 0.04BTC….8.192BTC。Samourai 服务采用的则是 CoinJoin 模式的变体, 通常来说, 其混币交易往往为 5 个输入和 5 个输出, 并且输入端金额和输出端金额分布极为相似。图 3 展示了一笔 Samourai 采用的 CoinJoin 交易模式, 其输入输出的金额基本相等, 在输入端有几个输入会稍多一点以作为交易的手续费, 如图中两个 0.101BTC 的输入。相比于其它 CoinJoin 变体, Samourai 采用的 Whirlpool 模式, 具有着极强的隐私性。

0.101BTC	交易	0.1BTC
0.1BTC		0.1BTC
0.101BTC		0.1BTC
0.1BTC		0.1BTC
0.1BTC		0.1BTC

图 3 Samourai 混币交易示例
Figure 3 Case of Samourai mixing transaction

在本文中, 我们专注于基于混淆方案的混币交易, 因为它们是现实世界中部署最为广泛的混币服务类型。

3 识别混币交易

准确地识别混币交易, 可以有助于执法部门和监管机构识别流向混币服务的资金并调查洗钱等非法活动。此外, 它可以帮助我们过滤掉不可靠的启发式方法并提升地址识别模型的性能, 这将在下一节中讨论。

许多以往的工作试图识别基于 CoinJoin 的混币交易。文献[21]假设具有超过五个输入和输出的交易可能与 CoinJoin 大致相关。文献[22]假设 CoinJoin 交易中的输入数量必须至少是输出数量的一半。BlockSci[23-24]还提供了一种严格遵循 Greg Maxwell[17]对 CoinJoin 原始定义的启发式方法来识别 CoinJoin。然而, 所有这些启发式或模式并不适用于现在流行的混币服务, 如 Wasabi 钱包[18]、JoinMarket [19]和 Samourai 钱包[20]等, 因为它们使用 CoinJoin 的变体来实现混币交易。

本节介绍了我们的识别方法, 可以更通用的识别基于混淆机制的混币交易, 包括一个基础算法和一个进阶算法。

3.1 基础识别算法

CoinJoin 将多笔交易组合为一笔交易, 因此很难追踪交易的输入和输出之间的关系。然而, 输入和输出的金额仍然会泄漏足够的信息, 并且可以使区块链研究人员根据输入端金额的组合与输出端进行比较, 有很大可能将组合的交易进行拆分[25], CoinJoin sudoku 曾经宣称其成功破解了 Shared Coin 服务的初始混币交易。因此, Maxwell[17] 建议有效的 CoinJoin 应具有相同值的交易输出, 为所有潜在的真实交易接收者形成一个匿名集合。

基于这一观察, 我们提出了一种评估交易输出

的多样性并量化其混淆程度的方法, 如下所示:

$$\alpha=Nu/Nt$$

其中, Nu 是唯一输出值(不同输出值)的数量, Nt 是交易中的输出总数。α 越低, 表明交易的匿名性越强。我们认为如果一笔交易的 α 低于一个阈值, 那么这个交易就具有混淆资金流的作用, 是一笔混币交易。这个阈值可以从最简单的混币交易里推断得出。在最简单的情况中, 混币交易的最小输入数量是两个, 输出的最小数量是四个, 即两个实际支付的输出和两个找零的输出。为了进一步增强匿名性, 一个有效的混币交易会有至少两个相同金额的输出(一般是两个实际支付的输出), 形成一个匿名集合(Anonymous set)。因此, 最简单的混币交易最多应该有三个唯一的输出值, 其 α 应该小于 0.75(即 3/4)。因此, 如果某笔交易至少有两个输入和四个输出, 并且 α 的值低于 0.75, 我们就会认为其为混币交易。

为了排除与 Omni[26]和灰尘攻击(Dust attack)等第 2 层协议相关的交易带来的误报, 我们在应用混币交易识别算法前需要对交易进行预处理。由于比特币的防尘机制和攻击成本, 灰尘攻击和第二层协议的输出一般低于 10000 satoshi(比特币的最小单位), 高于 247 satoshi。因此, 我们从每笔交易中删除值低于 10000 satoshi 的输出, 排除二层协议, 灰尘攻击以及 OP_RETURN(无价值输出), 然后评估调整后交易的 α 以及输入和输出的数量。为了进一步排除同一个地址将多个比特币 UTXO 聚合到一个地址的聚合交易或拆分资金并向同一地址发送不同值的拆分交易(交易的多个输入或输出的的地址是相同的), 我们认为混币交易至少有 2 个不同的的输入地址和 4 个不同的输出地址。

3.2 进阶识别算法

通过分析最新的公开的混币交易数据集[14], 我们发现匿名集可以是具有相似(而不是相同)值, 但具有细微差异的一组输出, 这一点在以往的工作中鲜有论述。而这种模式进一步混淆了交易输出, 且不能通过传统的 Coninjoin 检测算法轻易识别。事实上, 我们确实发现了一些遵循这种模式的混币交易[27]。

为了处理输出的细微差异, 我们设计了一种进阶算法来评估交易输出的方差而不是唯一性。具体来说, 我们使用非参数估计方法, 即核密度估计(KDE)[28-29], 来估计随机变量的概率密度函数(PDF)。KDE 可以估计交易的输出值分布。在概率密度函数中分布更接近的一组输出金额会形成一个有极大值的波峰, 可以被看作一个潜在的匿名集合。因此, 我们使用估计的 PDF 中局部极大值的数量作为 Nu, 并相应地计算 α 以识别混币交易。

我们使用的 KDE 算法设置如下。首先, 我们选择具有两个以上输入和四个以上输出的交易作为候选, 并从交易中删除低于 10000 satoshi 的输出值以避免误报。然后, 我们为每笔交易找到最大的输出值 V_{0max} 和第二小的输出值 V_{1min}。之后, 我们这样选择 KDE 的带宽: (1)在等比数列 10^n 中找到项 P, 其中 n 的范围从 $-\infty$ 到 ∞, s.t. $P <= V_{1min} < 10 * P$; (2)设置带宽为 $bandwidth = P/10$。例如, 如果 V_{1min} 为 0.5, 则 P 为 0.1, 带宽为 0.01。最后, KDE 中使用的采样点数设置为 $V_{0max}/bandwith$, 核函数设置为高斯函数。这里取第二小的输出值是因为避免交易里的多个输出偏差过大, 最小值相对其他输入太小以至于得到的 $bandwith$ 过小, 让本应该归为一组匿名集的输出没有归在一起导致算法漏报。

通过应用这种 KDE 算法, 我们可以计算输出值的 PDF 中极大值的数量, 并相应地计算 α 以识别具有不同输出值而不是相同输出值的混币交易。

4 识别比特币地址类型

图 4 展示了 BATscope 识别比特币地址类型的整体流程。如图 4 所示, 我们首先从开源数据中收集地址标签数据集, 训练机器学习模型(如 LightGBM), 然后按照如下方式扩充训练数据以不断改进模型。具体来说, 我们分析地址的历史交易行为, 将某些可靠的启发式方法应用于与已知地址相关的非混币交易, 以将标签扩展到未知地址。为了增强模型的泛化能力, 我们采用了一种新颖的先导预测方法来进一步标记未知地址。它首先使用当前模型来预测未知地址的类型, 并使用特殊的纠错机制纠正潜在的预测错误, 然后用纠正后的地址标签数据来扩充训练集。BATscope 通过重复这个增强过程以获得足够的训练数据来不断完善模型。

4.1 特征提取和机器学习模型

我们首先使用 BlockSci[24]从比特币地址的交易中提取本地和交易特征。本地特征包括地址本身的本地信息, 例如其最终余额、发送和接收的比特币总数、涉及的输入/输出交易数量和交易频率等。交易特征描述了一个地址涉及的所有交易的行为抽象, 仅与交易本身有关, 如地址所有交易输出/输入端平均数, 所有交易总金额平均数等, 特征的具体细节如[30]。

我们应用 LightGBM 来解决多分类问题并标记每个地址。LightGBM 具有出色的性能和较低开销, 在以前的工作中被广泛采用[12, 15]。文献[12]表明

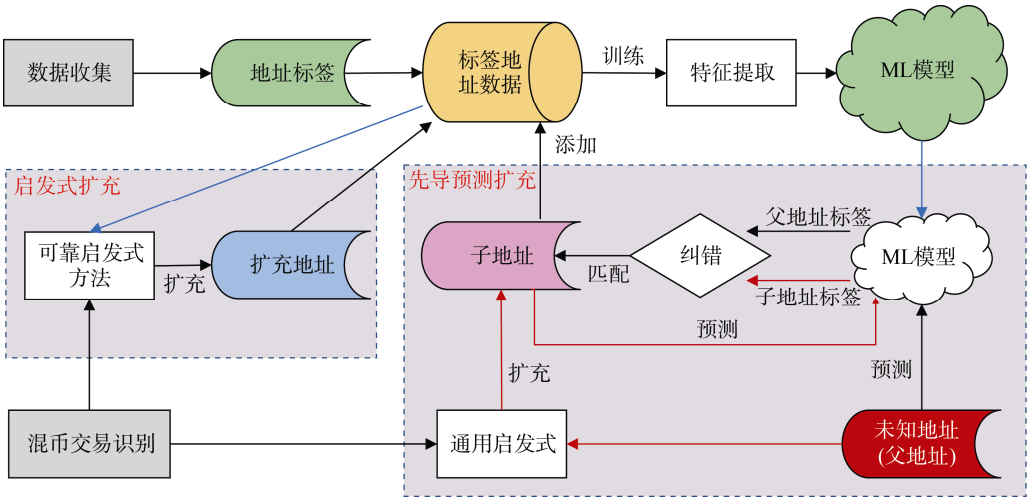


图 4　BATscope 设计

Figure 4　Overall design of BATscope

LightGBM 在地址分类方面的性能优于其他 ML 模型, 包括 Logistic Regression、SVM、Random Forest、XGBoost 和神经网络。此外, LightGBM 在一定程度上还提供了可解释性, 可以帮助我们分析哪些特征对地址分类有效。

4.2　启发式数据增强

我们改进了现有的启发式方法, 总结并提出了三种更可靠的启发式方法, 即适用于任何类型地址的通用启发式方法(General Heuristic)、适用于交易所和矿池地址的剥离链启发式方法(Peeling Chain Heuristic)以及适用于个人用户的普通用户启发式方法(Normal User Heuristic), 来从已知地址扩展标签到未知地址。

4.2.1　通用启发式方法

该启发式源自经典的多输入启发式, 这种启发式方法广泛用于聚类比特币地址。在多输入启发式中, 同一交易的多个输入地址会分配相同的标签。尽管这种启发式方法在大多数情况下有效, 但它仍然可能会失败, 即多个输入地址的标签可能不同。例如, 很多交易所与矿池有密切的合作, 甚至很多交易所有自己的矿池。矿池可能会在多输入交易中将奖励的比特币转移到交易所, 在这种情况下, 矿池和交易所会被错误地归入同一类(标签)。

我们通过考虑每个输入的比特币金额来改进多输入启发式并提出以下通用启发式方法:

如果两个或多个输入在同一笔交易中, 并且它们的比特币价值非常接近(即差异低于 10%), 那么我们认为这些输入地址属于同一类型。

上述启发式是可以用于任何类型地址的通用方法。同一个交易的输入首先确保它们被分配到同一个集群, 即被同一个实体控制。它们的接近的金额表明它们在这个交易中具有相似的状态或功能, 这意味着地址可能属于同一类型。此外, 交易中的输入必须来自之前交易的 UTXO, 而这些 UTXO 接近的金额表明这些不同的地址由于相同的交易目的作为交易的输出地址收到了相似数量的比特币, 证明了通用启发式方案的合理性。金额相似但不完全相同可能是由于比特币价格波动导致的(美元计价)。在恶意地址中的一个例子是攻击者通常以美元为攻击目标设置赎金, 但受害者在不同的时间以比特币的形式发送赎金, 导致每笔赎金交易中比特币价值的微小差异。在通用启发式中这个差异应小于 10%, 这是根据自 2015 年比特币开始流行以来的平均每周波动计算得出的。

在误报方面, 我们认为其中大部分误报应来自输入混合技术(例如 CoinJoin)的影响, 这也打破了原始的多输入启发式方法。在 CoinJoin 的影响下, 启发式方案会将不同类型的地址归为同一类。使用第三节中提出的混币识别算法, 我们可以跳过对混币交易应用通用启发式, 从而有效地减少误报。

其他误报来自交易所地址的多重身份。当用户需要将自己的比特币存入中心化交易所时, 交易所会将自己的地址分配给用户进行存币操作, 一些不法分子可以直接使用他们的交易所的存款地址接收赎金。当交易所进行多对一交易将比特币聚合到交易所主钱包时, 如果当攻击者的地址在多个输入地址中, 通用启发式会导致我们将交易所的其他地址错误标记为恶意的地址, 导致误报。因此, 对于非交易所地址, 当使用通用启发式算法时, 我们排除了超过 50 个输入的多对一交易。阈值(50)是我们在对

主流交易所的聚合交易的分析中找到的聚合交易的最小输入数量，是 Bitzlato[31]交易所采用的聚合交易的模式。由于阈值越小，造成的误报越少，我们认为 50 是一个合适的阈值。

4.2.2 剥离链启发式方法

第二个启发式基于一种称为剥离链(Peeling Chain)交易的特殊交易模式。这种交易模式将输入金额进行分割，绝大部分的金额发回输入实体控制的另一个地址，其余的输出地址分配一小部分剩余的金额。获得输出金额最大的地址不断做同样的事情，从而形成一个交易链，逐渐将金额剥离到其他地址。剥离交易通常是一对多的交易，输入地址和输出地址中价值最大的可以认为是同一种类型，甚至是同一个实体。这种模式广泛用于一些中心化的实体，如交易所和矿池。交易所使用剥离链进行用户的提现操作，矿池通过剥离链向每个矿工发送奖励，因此在这种情况下，输出值小的地址可能是矿工地址。图 5 展示了矿池使用剥离链向矿工发送奖励的示例。

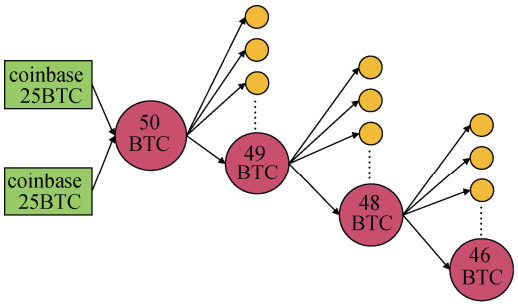


图 5 剥离链交易模式分配矿工奖励

Figure 5 Sending rewards to miners with peeling chain pattern

我们可以利用剥离链模式来扩展更多的地址标签，可以概括为以下启发式方法:

在剥链交易中，如果输入是一个交易所，那么最大的输出是一个交易所地址；如果输入是矿池，那么最大的输出是一个矿池地址，其他输出值小的地址是矿工地址。

4.2.3 普通用户价启发式方法

对于使用不属于任何实体的比特币的个人，没有有效的方法来检测他们，但我们可以通过恶意地址来识别与其交易的普通用户的钱包。由于比特币用户将比特币发送到这些恶意地址，我们可以搜索这些地址为输出的交易，并分析输入地址是否为普通用户。据我们了解，个人用户的钱包默认只向单个地址发送比特币，这表明用户钱包产生的交易总是有一个或两个输出(商家和找零地址)。文献[32] 中也提到了类似的启发式方法，称为消费者启发式方法(Consumer Heuristic)，可用于主流的钱包，例如 Bitcoin Core、Electrum、MultiBit、Armory 和 Android 比特币钱包等。我们进一步改进消费者启发式，只分析输入数量为 1 的交易(多个输入地址可能属于交易所行为)。因此，我们可以使用以下启发式方法来确定普通用户的比特币地址。

对于输出包含恶意地址的交易，如果它只有一个输入和两个或更少的输出，我们可以将输入地址标记为普通用户。

4.3 先导预测数据增强

使用可靠的启发式方案扩充固定的标签数据集来提升模型的边际效应会随着地址的迭代次数而递减。因为从原始标签地址扩展而来的新地址是有限的并且容易同质化。在最极端的情况下，我们的启发式方法可能会退回到多输入启发式，并且无法推导出更多新地址的标签。因此，有必要对未知地址应用启发式算法，解决收益递减的问题，进一步增强模型的泛化能力。

为了进一步将标签扩展到未知地址，我们设计了一种新颖的先导预测方法。对于标签未知的地址，我们首先使用当前模型预测一个类型作为它的标签。但是，标签可能是错误的，因为模型不能完全准确，用该地址标签以及从其派生出的新地址数据会对模型的迭代训练会产生负面反馈。因此有必要进行纠错。在图像识别等其他机器学习任务中，如果不进行人工检查，很难判断输出的正确性。但是，在比特币场景中，可以通过地址之间的交易关系来评估模型的输出结果。

为了纠正模型的潜在错误，我们使用了通用启发式方法，该方法可以从多输入交易中推断出与已知标签相同标签的地址。为方便起见，我们将要检查其标签的已知地址称为父地址，从父地址派生的地址称为子地址。如果模型输出对于父地址是正确的，那么大部分子地址也应该被模型归类为同一类型。因此，我们用模型对通用启发式派生出的子地址进行分类，并统计不同类型各自的数量。如果其中最主要的类型(数量最多)与当前模型标记的父地址的标签相同，可以认为父亲的标签是正确的。如果没有一个主要的类型，或者主要类型与父地址标签不符，我们不会使用子地址来不断地重新训练当前模型以进行模型优化。

5 实验评估

在本节中，我们进行了几个实验来评估我们的

解决方案并回答以下问题:

- 问题 1: BATscope 在识别混币交易方面是否有效?
- 问题 2: 我们的启发式方案从已知地址扩展到未知地址的标签是否可靠?
- 问题 3: 先导预测方法在提升模型性能方面是否有效? 纠错机制是否表现良好?
- 问题 4: 我们的解决方案在实践中是否鲁棒且实用?

5.1 混币交易识别评估

5.1.1 公开数据集上的算法性能

我们首先评估混币交易识别的方法。我们将基本算法和进阶算法与 BlockSci 中 CoinJoin 启发式算法和 Mo □ser 的文献[22]提出的算法在公共混币交易数据集[14]上进行比较。该数据集包含来自 Wasabi 钱包的 13, 581 笔混币交易和 9, 372 笔 ChipMixer 交易, 其中对于 ChipMixer 我们使用 7, 683 笔有效混淆交易(即具有多个输入)进行评估。我们使用覆盖率, 即算法识别出的混币交易数量占混币交易总数的比例来衡量算法的有效性, 结果如表 1 所示。它显示了我们的基本和进阶算法在 ChipMixer 数据集上实现了 100%的覆盖率, 在 Wasabi 数据集上实现了接近 100%的覆盖率, 以极大的优势超越了其他两种方法。此外, 在 Wasabi 数据集上, 我们的进阶算法略好于基本算法。

表 1 混币识别算法在公开数据中的效果

Table 1 Identification of mixing transaction on public data

	Blocksci	Mo □ser	基础算法	进阶算法
ChipMixer	0%	10.30%	100%	100%
Wasabi	7.26%	94.79%	99.65%	99.94%
总和	4.63%	64.21%	99.78%	99.96%

5.1.2 混币交易测量

我们现在从比特币客户端提取的区块链数据来量化混币技术在历史中的使用情况。数据包含 2021 年 6 月 8 日之前的所有交易(区块高度为 686795), 我们分别使用 BlockSci 启发式、基本算法和进阶算法来识别其中的混币交易。我们分别统计每周三种算法识别出的混币交易数量, 得到图 6。

结果表明我们的进阶算法可以识别大多数潜在的混币交易, 而 BlockSci 启发式算法由于过于严格出现很多漏报, 而我们的识别算法可以识别出更多的混币交易。

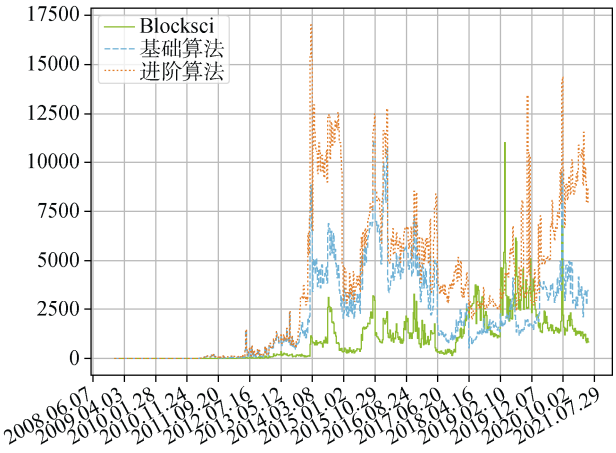


图 6 每周混币交易数量

Figure 6 Number of mixing transactions per week

评估结果也反映了混币交易的几个重要时间点。2013 年之前还没有混币交易, 而此时类似于混币交易的交易模式基本上是由链上博彩地址(satoshi dice)发起的, 这可以通过交易中涉及的虚荣地址(Vanity address, 指比特币地址中含有可识别的单词, 如地址 1dice7fUkz5h4z2wPc1wLMPWgB5mDwKDx) 中的特殊字符(dice 和 lucky)来证明。2014 年潜在混币交易的数量突然从零增长到相对较高的水平, 这可以通过 2013 年 8 月提出的 CoinJoin 技术[17] 和 2013 年 11 月 Shared Coin[33]提供的免费混币服务来解释。之后混币交易数量逐渐上升, 在 2014 年 3 月达到峰值, 这一现象也如文献[21]所述, 此外, 经过计算块平均混币交易数量, 我们也得到了和文献[21]相同的结论。

2015 年 1 月, 混币交易的数量突然下降, 评估结果与[22]中所述的结果相符, 作者推测这是由于 Shared Coin 的实现进行了更改, 增加了对额外 API 端点的支持[34]。2015 年 5 月, 因 JoinMarket[35]发布, 混币交易的数量又一次快速增长; 随后出现下跌, 可能与 Shared Coin 在 2016 年上半年停止服务有关。2017 年 6 月 20 日到 2018 年 4 月 16 日左右, 混币交易数量整体处于低谷, 这一趋势符合比特币的总体交易量[36]。Wasabi 钱包和 Samourai 钱包随后于 2018 年 7 月和 2019 年夏季开始提供混币服务, 导致混币交易逐渐增加。

综合来看, 我们的进阶识别算法可以比基础算法识别出更多的混币交易, 并且其随历史发展的分布也符合比特币历史中有关混币交易的发展趋势, 证明我们提出的识别算法十分可靠。

5.2 启发式数据增强评估

5.2.1 数据收集

我们首先描述数据收集的方法。基本标签数据

来自之前的工作[11-12], 包含 26000 多个地址, 最初从 Blockchain.info/tags、WalletExplorer 和 BitcoinTalk 等公开网络渠道获取, 涵盖 7 种不同类型, 即交易所、HYIP(高收益投资项目)、水龙头、博彩、混币器、矿池和市场。

我们从 2018 年 1 月 4 日至 2021 年 6 月 8 日的区块中的 Coinbase 交易中重新收集矿池地址。从 18 年起比特币的挖矿市场趋于稳定, 由于全网算力的不断提高, 基本不会再出现个体的矿工, 因此我们可以认为这些 Coinbase 交易的几乎所有输出地址都属于矿池。经过去重和剥离链启发式算法, 我们得到了 434 个矿池和 3, 319 个矿工地址。我们通过实际交易和 RichList[37]等公开网络资源收集了额外的 250 个交易所地址。我们直接从 2021 年 5 月 15 日之前 Bitcoinabuse[16]的报告中获得恶意地址, 该网站每天报告和跟踪勒索软件、钓鱼邮件、欺诈等攻击事件中使用的比特币地址。Bitcoinabuse 中报告的地址还包含一些在被盗币和赃款的交易流程中涉及的交易所地址。我们通过商业软件检查报告中超过 1000 笔交易的地址来过滤掉其中潜在的交易所, 最终得到 13, 379 个恶意地址。此外, 我们将数据集[11-12] 中的 HYIP(高风险投资项目)地址添加到恶意类别中。对于普通比特币用户, 我们对恶意地址应用普通用户启发式算法获取用户地址。为了平衡不同类型的数据, 我们抽取了 10%的恶意类型地址, 最终得到了 2060 个用户地址。

在本文中, 为了着眼于分析恶意地址和其恶意行为, 我们只关注交易所(Exchange)、恶意地址(Malicious)、矿池(Pool)、矿工(Miner)和普通用户(User), 并将其他地址归为“其他类型”(Others), 并对这些地址进行分类。

5.2.2 启发式方法评估

现在我们评估地址扩展启发式方案的可靠性。我们在交易所地址上测试了通用启发式方法, 并将我们的解决方案与一款商业软件的工具进行了比。因为商业工具与交易所有合作, 可以收集相对更多的交易所地址, 在一些知名地址中表现良好, 可以作为参考的 baseline。我们选择了两组交易所地址, 即直接交易的 10 个交易所地址和收集的另外 240 个交易所地址。使用通用启发式方法, 我们最终从两组地址中分别推导出 75k 和 386k 候选交易所地址, 并用商业软件查询这些候选地址的类型。

考虑到商业软件查询大规模地址的巨大时间开销, 对于 75k 候选者, 我们抽取其中的 1%, 并通过商业软件测试所抽取的地址, 并重复该过程 10 次。对于 386k 候选者, 我们分别抽取其中的 1%和 10%进行测试。我们使用命中率来量化我们的通用启发式算法的性能, 即被商业软件识别为交易所的地址占采样的候选地址中的比例。此外, 我们还针对矿池测试了通用启发式方法。由于矿池可能控制多个地址并承担不同的任务(如在剥离链中承担中转节点), 所以矿池地址并不一定直接参与 Coinbase 交易并作为输出地址。如果一个地址交易中的输入可以在几跳内溯源到一个 Coinbase 交易, 我们认为这个地址属于一个矿池。通过通用启发式, 我们得到 1, 076 个矿池候选地址, 并将目标跳数分别设置为 3、5 和 8。

结果如表 2 所示。从表中可以看出, 我们的通用启发式方法是可靠的。在对采样数据的所有十次测试中, 商业软件报告每次交易所候选者的命中率约为 90%, 并且每次查询的命中率都保持相对稳定。在更大的候选交易所(交易所 2)中, 我们的启发式仍然可以保持较高的命中率。鉴于商业服务为了避免误报通常是比较保守, 仅对完全有信心的地址进行标注, 而大多数地址通常被认为是无标签的。因此, 评估结果可以证明我们的启发式方法的可靠性。此外, 我们的通用启发式方法也适用于矿池。大部分的候选地址可以在 8 跳内可以溯源到 Coinbase 交易, 说明通用启发式方法确实可以从矿池地址派生出新的矿池地址。

表 2 通用启发式方法评估

Table 2 Evaluation of the general heuristic

交易所 1	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
	89.2%	89.6%	89.7%	92.1%	90.3%	90.3%	88.4%	90.5%	90.3%	88.9%
交易所 2	1%采样 86.1%							10%采样 88.7%		
矿池	3 跳 67.8%				5 跳 77.9%				8 跳 86.1%	

由于剥离链启发式是通过我们的观察和真实交易经验, 普通用户启发式在文献[32]和在钱包的日常使用中都有得到证明, 我们这里不对这些启发式进行实验评估。

5.2.3 模型提升评估

通过可靠的地址扩展启发式方法, 我们可以量化启发式方法给机器学习模型带来的提升。我们将收集到的数据中 20%作为测试集, 另外 20%作为用于测试先导预测的未知标签地址, 其余作为基本训练数据集。我们用基本数据集训练的 LightGBM 模型作为 baseline, 然后对基础数据的不同比例(即 20%、50% 和 80%)进行采样, 并使用我们的启发式方法来扩充这些标记地址, 用扩充出来的标签地址数据重新训练我们的基本模型以获得更好的性能。采样的比例可以说明扩充地址的迭代次数对模型性能的影响, 并用 Micro F1 分数和 Macro F1 分数进行评估。

启发式方法在 100%的基本训练集上迭代后, 我们使用先导预测的方法, 继续不断提升模型性能。表 3 描述了实验结果。从仅用基础数据训练到使用启发式迭代不同规模的基础数据, 模型在同一测试集中的性能不断提升, 在基础训练集 100%迭代的情况下, Micro-F1 达到 95.049, Macro-F1 达到 94.412。

表 3 模型表现

Table 3 Model performance

	基础	20%	50%	80%	100%	先导预测
Micro F1	88.219	93.035	94.155	94.841	95.049	96.206
Marco F1	87.686	92.238	93.329	94.190	94.412	95.667

最明显的提升出现在从基础训练集到 20%的迭代, 模型 F1 提升了 5 左右, 而随着迭代次数的增加, 模型改进的收益逐渐放缓, 从 80%到 100%的迭代过程中, Micro F1 分数和 Macro F1 分数仅提升了 0.2 左右, 如第四节所述, 呈现出边际效应递减的趋势。然后我们对另外 20%的未知标签的地址数据应用先导预测, 将模型的 Micro/Macro-F1 大幅提升到 96.206 和 95.667, 证明先导预测的方法确实可以通过对未知地址的标签扩充让模型性能进一步提升, 并且可以有效缓解仅对于固定数据扩充带来的边际效应递减现象。

图 7、8 和 9 分别展示了具有基本数据、100% 地址迭代和先导预测的模型的混淆矩阵。很明显, 经过 100%的地址迭代, 模型的性能在每个类中都有了明显的提升, 达到了 90%以上的准确率。通过先导预测, 模型性能进一步提升; 几乎所有的类都达到了 95%

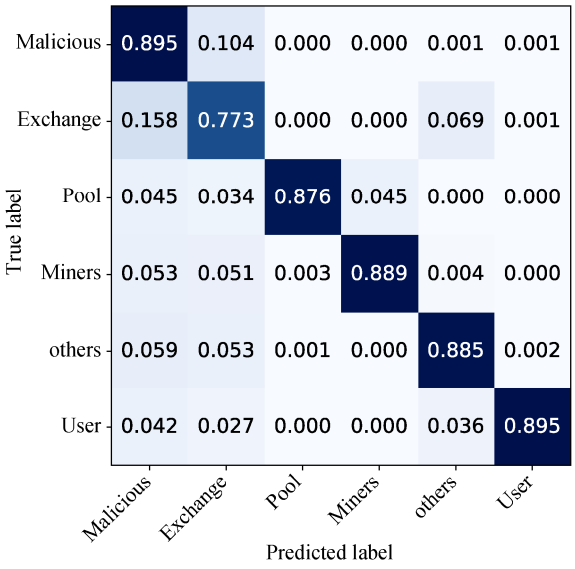


图 7 基础数据混淆矩阵

Figure 7 The confusion matrix of model with basic data

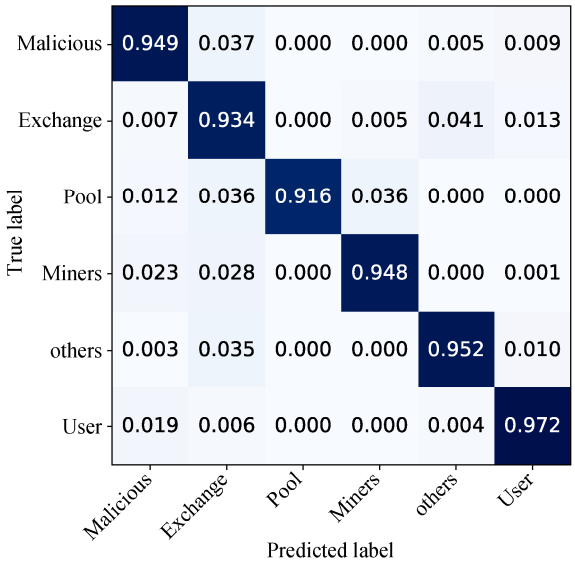


图 8 100%迭代混淆矩阵

Figure 8 The confusion matrix of model with 100% address iteration

的准确率。提升后的模型也优于之前的工作, 优于文献[11]准确率 0.70、文献[15] F1 分数 0.91 和文献[12] Micro-F1/Macro-F1 分数 87%/86%。

5.3 先导预测数据增强评估

如上一小节所示, 先导预测方法确实有效提升了模型的性能, 实现了比 100%地址迭代更高的 Micro/Macro-F1。在这一小节, 我们将深入研究这种方法, 评估我们的纠错机制。

如上一小节中所述, 我们使用 20%的收集数据作为标签未知数据进行测试, 它实际上具有真实标签作为 Groudtruth; 该数据集包含 8, 521 个地址, 其

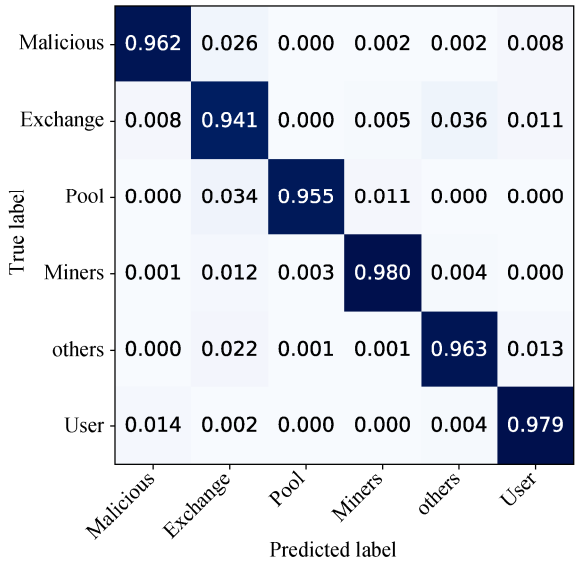


图 9 先导预测混淆矩阵

Figure 9 The confusion matrix of model with pilot prediction

中 3, 406 个是可扩展的, 而对于其余无法扩展的地址, 为了避免潜在的污染, 我们只会给出当前模型对他的类型推断, 而不会直接用预测的标签去扩充地址并继续训练模型。我们首先使用模型(100% 基础数据迭代)得到每个地址的预测标签, 然后使用纠错机制纠正潜在错误的预测标签, 对于每一个父地址, 我们可以得到子地址的主要类型。我们在表 4 中给出了结果, 展示了预测标签、主要标签和它们的真实标签之间的关系。如图所示, 先导预测适用于 94.6%((3109+113)/3405) 个地址, 其中主要标签与地址的真实标签一致。在我们当前模型没有准确预测的数据中, 纠错机制仍然可以避 77.6% ((113+12)/161) 个未知地址在再训练过程中给模型带来负反馈。

表 4 纠错效果

Table 4 Performance of the error correction

预测标签	主要标签	数量
T	T	3109
T	F	136
F	T	113
F	F*	12
F	F**	36

T/F 分别代表与地址真实标签一致/不一致。F**/F*分别代表在与真实标签不一致的情况下, 主要标签和预测标签一致/不一致

T/F represents consistency/inconsistency with true label; F□□/F□ represents inconsistency with true label, as well as consistency/inconsistency with prediction label

我们进一步分析了纠正失败的 36 个地址。其中 26 个地址是交易所, 但每个交易所扩展的新地址数量很少, 只有不到 7 个。扩展地址的数量过少容易导致纠错的误报, 因为在这种情况下主要标签并不稳定。其他 10 个地址也仅扩展了少数几个新地址。但是, 由于扩展地址很少, 即使纠错失败, 这些地址对模型演化的负面影响也非常有限(占总体训练数据集的比例极低)。因此综合而言, 我们的纠错模型可以有效的避免在对模型提升过程中的负反馈。

5.4 恶意地址检测的实际应用

为了进一步展示 BATscope 在实践中检测恶意地址的能力, 我们额外收集了 2021 年 5 月 15 日之后来自 Bitcoinabuse[16]的恶意地址, 这些地址未在我们用于训练和测试 BATscope 的数据集中。在 787 个有实际交易记录的恶意地址中, BATscope 成功识别了其中的 748 个, 准确率达到 95.04%。由于 Bitcoinabuse 中记录的地址在现实世界中被不法分子实际应用, 其中一些来自已知事件的攻击, 因此我们认为 BATScope 在实践中可以有效识别恶意地址。

6 案例分析

现在我们使用我们的混币交易识别算法和 BATScope 来分析现实世界中的恶意比特币地址并进行一些详细的案例研究。

6.1 混币交易在恶意地址中的使用

我们分析 Bitcoinabuse 中恶意地址的所有发送交易, 跟踪他们 UTXO 的后续 10 跳交易来检测是否存在混币交易。在全部 13, 379 个恶意地址中, 我们发现其中 5, 027 个使用混币技术来混淆资金流, 共产生 20, 659 个混币交易。为了进行更深入的分析, 我们设计了一种粗略的方法来识别混币服务。我们的结果显示 2, 027 笔交易属于 Wasabi 钱包, 633 笔交易符合 ChipMixer 模式。有趣的是, 来自不同恶意地址的比特币可能会流向同一个混币交易, 这表明这些不同的恶意地址可能由同一个攻击者实体控制。

表 5 显示了恶意地址资金流入最多的前 10 个混币交易。我们进一步分析这些交易的每个地址。在交易 875e2f 中, 所有地址都是用邮件进行勒索诈骗, 不同的地址有不同的邮件内容。攻击者要求的比特币赎金的数量(美元计价)在不同地址之间很接近。此外, 这些不同地址的邮件是在同一时期(2018 年末至 2019 年初)发送的, 证明这些地址可能由单个实体控制。d098319 相关的地址不同。他们进行虚假捐赠, 承诺将双倍数量的比特币退回。

表 5 10 个最多恶意地址资金流入的混币交易

Table 5 Top 10 mixing Tx with most malicious addresses

交易 hash	地址数量	混币服务商
875e2f...	39	ChipMixer
619791...	28	ChipMixer
e5d856...	27	ChipMixer
eb5366...	25	unknown
d09831...	24	Wasabi
27ff1a...	24	Wasabi
43b0ed...	24	Wasabi
13056b..	22	ChipMixer
320c86...	22	ChipMixer
d6c334...	22	unknown

通过对混币交易的分析, 我们可以得出结论, 与同一个混币交易相关的不同恶意地址可能由同一个实体控制, 这对于识别和关联攻击者控制的恶意地址集群带来了新的思路, 在未来可以被应用作为发现更多比特币恶意使用者的新启发式方法, 可能会对比特币的反洗钱工作提供帮助。

6.2 Luno 钱包诈骗

恶意地址有时不使用混币交易, 而是直接将币存入交易所用来直接兑换法币。在某些情况下, 他们甚至使用由交易所直接控制的存币地址来接收赎金。我们发现 coinboom24[38]中的投资诈骗符合这种模式。通过检查我们的地址标签库, 攻击者收到的比特币流向地址 17ac9t (17ac9txhxu1nxdlglu9wyk7vr8ggfn5gkh) 被 BATscope 标记为交易所, 具体而言是 Luno 钱包。攻击者会请求一笔投资, 并会返还 100% 的保证金。当攻击者地址收到比特币时, 他们会自动通过一笔有数百个输入的多对一交易将比特币发送到 17ac9。此模式表明这些地址是实际上是 Luno 钱包控制的用户存款地址。类似的攻击也会发生在其他社交媒体上, 如 telegram 和 instagram[39, 40], 例如地址 35vEY(35vEYaj43QdciY9tR9QiRXLdmfxCuC-rmC3)。它利用诈骗手段一共做了 22 笔交易, 总共收到 0.08352287 比特币, 并在 18 笔交易中将所有硬币发送到 17ac9, 大约有 750 个输入。 在 Bitcoinabuse 的报告中向 17ac9 转币的恶意地址有 988 个, 可见 Luno 钱包滥用比特币的规模之大。

6.3 Twitter 黑客诈骗

Twitter 黑客攻击发生在 2020 年 7 月 15 日, 造成超过 110, 000 美元的损失[2]。我们使用 BATscope 来查看和分析本次事件中使用的比特币地址, 并在图 10 中绘制了交易图(2020 年 7 月 31 日之前)。该图表明攻击者通过 ChipMixer 混币服务使用的大量混币交易(黑色矩形)。我们还使用 BATscope 来识别将比特币转移到交易所的交易(绿色矩形)。我们检查由 BATscope 识别的潜在交易所实体并用我们收集的实体标签进行匹配。我们找到了在报道中[41]攻击者试图转移的交易所 Coinbase; 其他涉案交易所如币安、火币和 Kraken 也匹配了 BATscope 识别的地址, 进一步证明了 BATscope 在实践中的可用性。

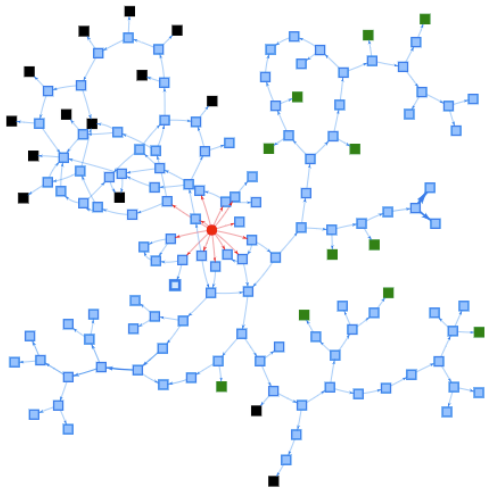


图 10 Twitter 勒索事件中的部分交易图

Figure 10 Part of transaction graph in twitter hack case

6.4 勒索邮件

在这种攻击场景, 我们分析了与 MySQL 服务器攻击事件中 PLEASE READ ME 勒索软件相关的地址 1BLYhUD3[42]。根据勒索邮件的内容, 攻击者删除受害者的数据库并勒索 0.06、0.04 或 0.03 比特币以恢复被勒索丢失的数据库。其中五笔交易用于转移赎金比特币。同样, 我们跟踪所有五笔交易并找到交易所或混币交易。交易图如图 11 所示。我们发现部分比特币流向了 BATscope 识别的交易所地址。他们的实体包括 Fcoin、Bitzlato、Bitpay 和 Kraken。这个地址还使用了 3 个混币交易, 包括 ChipMixer 和 Wasabi 钱包。同时, 该恶意地址的三个发送交易合并为交易 18cbe, 最终将硬币转移到交换 Fcoin。而 Fcoin 交易所也于 2020 年 2 月 17 日正式宣布关停。

虽然该攻击者采取邮件勒索的方式获取比特币, 但与一般勒索诈骗不同, 攻击者确实攻击了受害者的数据库, 因此相比于其他诈骗地址, 这个案例的交易记录更值得关注。我们额外分析了这个案例的

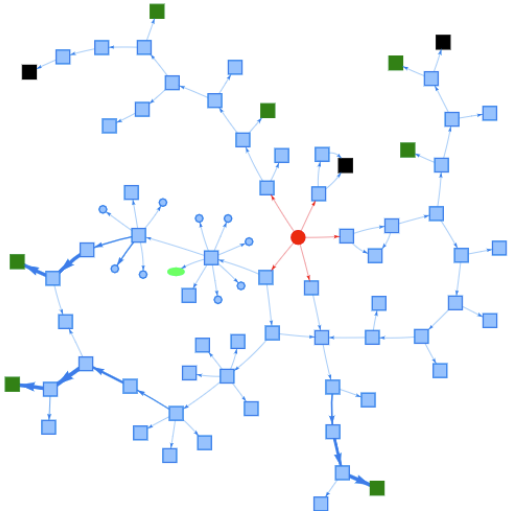


图 11 勒索邮件案例中的部分交易图

Figure 11 Part of transaction graph in blackmail case

受害者地址身份信息。我们遍历 1BLYhUD3 地址的所有接收交易，交易的输入地址即为受害者地址，我们逐个分析并用 BATscope 分析他们的地址类型，其结果如图 12 所示。我们注意到受害者地址中有76%属于交易所地址，21%属于普通用户地址。这是由于如今用户购买和交易比特币都是通过中心化的交易所，资金全部由交易所托管，在这个案件中，反应为赎金的支付也为中心化交易所代为支付，仅有21%的用户可能直接使用自己控制的钱包地址支付赎金。这反映了为了交易，投资的方便和效率，用户更倾向于把资金托管给交易所而不是自己实际持有。我们还发现了极少量矿工地址，分析其交易，我们确实发现其处于一个矿池的剥离链交易链中。

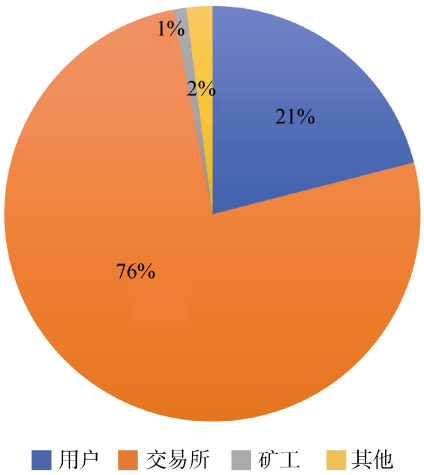


图 12 受害者地址类型

Figure 12 Proportion of victim address type

6.5 执法机关案件

我们还使用 BATscope 来分析执法部门提供的未知地址。具体来说，我们帮助执法部门分析了 3 个地址，并进一步跟踪了相关比特币的流动。这些地址都通过一系列交易发送比特币，并最终转移到由 BATscope 识别的潜在交易所地址。这些潜在的交易所地址进一步将比特币转移到我们实体标签集中包含的交易所，这可以进一步确认他们的实体信息。根据交易所的反馈，执法部门能够得到充足的线索帮助破案。

7 相关工作

7.1 混币交易检测

一些工作专门研究比特币的混币技术。在文献[43]中，作者评估了 Blockchain.info 的 Shared Coin，Bitcoin Fog 和 BitLaundry 三种混币服务，发现 BitLaundry 的输入和输出之间存在直接关系。然而，许多服务已经关闭，这些工作今天已经缺乏实际意义。

除了中心化的混币服务，很多工作都专注在像 CoinJoin 这样的分布式协议上。对于 CoinJoin 交易，文献[21] 认为它应该有五个以上的输入和输出，而文献[22]认为其中的输入数量必须至少是输出数量的一半。Blocksci[23-24]还提供了遵循 Greg Maxwell [17] 原始规范的启发式方法。但是，这些启发式方法或模式不适用于 Wasabi 或 Samourai 等如今流行的混币服务，因为它们部署的是 CoinJoin 的变体。

文献[14]使用种子扩展算法从已知交易中找到更多混币交易，并获得 92% 的覆盖率。但是这个算法只对 Wasabi 和 chipmixer 有效，无法检测到其他服务的一般混币交易。

7.2 地址去匿名

一些论文[5-9]在 2015 年之前就进行了大规模的地址去匿名化研究，试图将地址与现实信息联系起来。他们使用启发式方法对比特币地址进行聚类，如原始比特币论文[1]中首次提到的多输入启发式方法和找零地址启发式方法。文献[7, 32, 44]改进现有的启发式方法以避免误报并防止超大集群。他们还提出了新的启发式方法，例如消费者钱包和优化的找零启发式方法。然而许多钱包或者服务的找零策略都不相同，因此一直很难设计出一个通用的找零启发式方案。

文献[10]采用机器学习的方法，通过训练一个模型去识别一个交易的找零地址，但在数据预处理中仅使用 Blocksci 的 CoinJoin 交易识别，仍会导致数据清洗不够充分。文献[45]和文献[46]利用多输入启发式生成比特币网络的用户图并进行图分析以推断异常行为。然而，如今比特币的整个生态系统比过去几

年复杂得多, 很难做类似的分析。同时, 这些工作没有考虑到那些启发式算法存在许多误报(例如, CoinJoin 会打破多输入启发式)。

7.3 使用 ML 进行地址分类

许多工作开始使用机器学习来识别地址或实体。在文献[11]中, 作者收集了超过 26000 个标记地址, 包括七种类型, 通过随机森林方案实现了 0.70 的地址类型识别的准确率。文献[12]沿用了文献[11]中的数据集, 并提出了交易时刻的特征, 在地址分类中使用 LightGBM 实现了 87%/86%的 Micro-F1/Macro-F1。文献[47]专注于实体分类, 作者使用 Gradient Boosting 算法, 在 10 个类别的 434 个实体中实现了 77%的准确率和 0.75 的 F1 分数。Bitscope[48]、文献[15]和文献[49]做了类似的工作。文献[15]结合交易序列特征并达到 0.91 的 F1 分数。基于 ML 的地址分类虽然在特定的数据集可以达到比较高的准确率, 但其数据集所含的地址规模都太小, 训练出的模型泛化能力可能仍有提升的空间。

文献[13]分析了对属于非法实体的比特币交易。作者利用图卷积网络(GCN)及其扩展模型 Evolve GCN 来识别非法的比特币交易, 在非法交易识别中实现了 0.72 的 F1, 在 MicroAVG 中实现了 0.968 的 F1。但其数据分布十分不均匀, 且经过脱敏, 无原本的交易原始的信息, 很难进行更深入的分析。

7.4 商业解决方案

各种公司, 例如 Elliptic[50]、Chainalysis[51]、Coinholmes[52], BEI[53]和 Blocksec[54]等, 都有专门研究“比特币区块链分析”的模型和工具。这些公司提供了一套软件和分析引擎来分析区块链并识别地址和实体的类型, 并提供可视化的追踪工具方便识别并追溯检测潜在的非法地址。

7.5 其他数字货币相关研究

除了比特币, 很多工作还对其他数字货币做了相关的研究。文献[55]对 Zcash 的匿名性做了试验性分析, 通过与比特币类似的启发式方法, 可以通过特定的交易模式识别出其中一些地址的身份信息如矿工, 矿池, 创建者等。文献[56]通过三种以太坊交易图(转账, 合约调用, 合约创建)的图分析可以识别出一些潜在攻击者的恶意交易行为。

8 结论

比特币系统是目前应用最广泛的基于区块链技术的数字货币, 但同时比特币的使用也伴随着犯罪和非法行为的大量滥用。在数字货币市场日益繁荣的今天, 如何对数字货币进行有效的监管一直是学界和业界重点关注的问题。在本文中, 我们提出了 BATscope, 针对比特币而言, 它通过可靠的启发式算法和先导预测可以自动迭代地增强机器学习模型的训练集, 从而不断提升模型的性能。为了避免启发式的误报并更好地了解恶意地址行为, 我们设计了混币交易检测算法。评估结果表明, BATscope 可以准确识别地址类型和混币交易, 部分实验代码在[57]。我们还使用 BATScope 来分析恶意地址的案例, 在实践中证明了它的实用性。

9 致 谢

本课题得到国家重点研发计划资助(No. 2021YFB2701000)和国家自然科学基金资助(No. 61972224, No. U1736209)。

参考文献

[1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system[J]. *Decentralized Business Review*, 2008: 21260.

[2] Twitter hack: 130 accounts targeted in attack: https://www.bbc.com/news/technology-53445090. July 2020.

[3] Colonial pipeline paid hackers nearly $5 million in ransom: https://www.bloomberg.com/news/articles/2021-05-13/colonial-pipeline-paid-hackers-nearly-5-million-in-ransom. May 2021.

[4] Chinese authorities have seized a massive $4b in crypto from plustoken scam: https://www.coindesk.com/chinese-authorities-have-seized-a-massive-4-billion-in-crypto-from-plustoken-scam. Nov 2019.

[5] F. Reid F, Harrigan M. An analysis of anonymity in the bitcoin system[M]. *Security and privacy in social networks,* New York, NY, 2013: 197-223.

[6] D. Ron and A. Shamir,. Quantitative analysis of the full bitcoin trans- action graph[C]. *International Conference on Financial Cryptography and Data Security*, 2013: 6-24.

[7] Meiklejohn S, Pomarole M, Jordan G, et al. A Fistful of Bitcoins: Characterizing Payments among Men with no Names[C]. *The 2013 conference on Internet measurement conference*, 2013: 127-140.

[8] P. Koshy, D. Koshy, and P. McDaniel. An analysis of anonymity in bitcoin using p2p network traffic[C]. *International Conference on Financial Cryptography and Data Security*, 2014: 469- 485.

[9] E. Androulaki, G. O. Karame, M. Roeschlin, et al. Evaluating user privacy in bitcoin[C]. *International Conference on Financial Cryptography and Data Security*, 2013: 34-51.

[10] Möser M, Narayanan A. Resurrecting Address Clustering in Bitcoin[EB/OL]. 2021: arXiv: 2107.05749. https://arxiv.org/abs/2107.05749

[11] Toyoda K, Ohtsuki T, Mathiopoulos P T. Multi-Class Bitcoin-Enabled Service Identification Based on Transaction History Summarization[C]. *2018 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*, 2019: 1153-1160.

[12] Lin Y J, Wu P W, Hsu C H, et al. An Evaluation of Bitcoin Address Classification Based on Transaction History Summarization[C]. *2019 IEEE International Conference on Blockchain and Cryptocurrency*, 2019: 302-310.

[13] Weber M, Domeniconi G, Chen J, et al. Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics[EB/OL]. 2019: arXiv: 1908.02591. https:// arxiv.org/abs/1908.02591

[14] Wu L, Hu Y F, Zhou Y J, et al. Towards Understanding and Demystifying Bitcoin Mixing Services[C]. *The Web Conference 2021*, 2021: 33-44.

[15] Jourdan M, Blandin S, Wynter L, et al. Characterizing Entities in the Bitcoin Blockchain[C]. *2018 IEEE International Conference on Data Mining Workshops*, 2019: 55-62.

[16] Bitcoin abuse database: https://www.bitcoinabuse.com/. March 2022

[17] G. Maxwell. CoinJoin: Bitcoin privacy for the real world: https://bitcointalk.org/index.php?topic=279249.0. Aug 2013

[18] Wasabi wallet: https://wasabiwallet.io/. 2021

[19] Joinmarket.: https://github.com/JoinMarketOrg/joinmarket-clientserver. 2021

[20] Samourai wallet: https://samouraiwallet.com/. 2021

[21] Meiklejohn S, Orlandi C. Privacy-Enhancing Overlays in Bitcoin[C]//International Conference on Financial Cryptography and Data Security. Berlin, Heidelberg: Springer, 2015: 127-141.

[22] Möser M, Böhme R. Anonymous Alone? Measuring Bitcoin's Second-Generation Anonymization Techniques[C]. *2017 IEEE European Symposium on Security and Privacy Workshops*, 2017: 32-41.

[23] Kalodner H, Möser M, Lee K, et al. BlockSci: Design and Applications of a Blockchain Analysis Platform[C]. *The 29th USENIX Conference on Security Symposium*, 2020: 2721-2738.

[24] Blocksci: A high-performance tool for blockchain science and exploration.: https://github.com/citp/BlockSci. 2020

[25] Boltzmann, a python script computing the entropy of bitcoin transactions and the linkability of their inputs and outputs.: https://github.com/Samourai-Wallet/boltzmann

[26] Omni layer.: https://www.omnilayer.org/

[27] Mixing transaction: https://www.blockchain.com/btc/tx/8b7cc7f9ca4231d498045cc6920a8fe1cd995151428586151293f452cbda6d35

[28] Davis R A, Lii K S, Politis D N. Remarks on Some Nonparametric Estimates of a Density Function[M]. Selected Works of Murray Rosenblatt. New York: Springer, 2011: 95-100.

[29] Parzen E. On Estimation of a Probability Density Function and Mode[J]. *The Annals of Mathematical Statistics*, 1962, 33(3): 1065-1076.

[30] Address feature.: https://docs.google.com/spreadsheets/d/1o39Zd0YETuHCPvO1cMC vdLXBgc84zZb7ifMxBetzVz4/edit?usp=sharing

[31] An example transaction of bitzlato: https://www.blockchain.com/btc/tx/6608aa7e9ac84bde9d75d0e 399da20d797fba6be0ae3784e41a671f6d3a0993e

[32] Nick J D. Data-driven de-anonymization in bitcoin[D]. ETH-Zürich, 2015.

[33] Blockchain. shared coin announcement tweet: https://twitter.com/blockchain/status/40222401049. 2013

[34] AB. Reeves. Use apache http client + multiple api endpoints.: https://github.com/blockchain/Sharedcoin/commit/550b39. 2014

[35] JoinMarket. joinmarket now running on the main bitcoin network: https://twitter.com/joinmarket/status/ 596463678066708480. 2020

[36] Confirmed transactions per day: https: //www.blockchain.com/charts/n- transactions 2021

[37] Bitcoin rich list: https://bitinfocharts.com/ top- 100- richest- bitcoin- addresses.html. 2020

[38] Investment scam coinboom24: https://reportscam.com/coinboom 24com. 2020

[39] Bitcoin abuse database report history for 3jvqf-gyf5 dahsnqrtcrhqud7 ysniardj3z: https://www.bitcoinabuse.com/reports/ 3JvqFgYF5DA HsNQRTCRhqUd7ysniarDj3z. 2021

[40] Bitcoin abuse database report history for 1geathyvba- gaij9e3 uzhwxarubk3emwgxq: https://www.bitcoinabuse.com/reports/1GeathyvBagaiJ9e3uZhwXarUBk3Emwgxq. 2021

[41] Twitter hackers were caught after sending bitcoin to verified coin-base accounts: https://siliconangle.com/2020/08/02/twitter- hackers- caught-sending- bitcoin- verified- coinbase- accounts/. 2020

[42] Please read me ransomware attacks 85k mysql servers: https://threatpost.com/please read me- ransomware-mysql-servers/162136/. 2020

[43] Möser M, Böhme R, Breuker D. An inquiry into money laundering tools in the Bitcoin ecosystem[C]. *2013 APWG eCrime research-ers summit,* 2013: 1-14.

[44] Neudecker T, Hartenstein H. Could Network Information Facilitate Address Clustering in Bitcoin? [C]. *International Conference on Financial Cryptography and Data Security*, 2017: 155-169.

[45] Di Francesco Maesa D, Marino A, Ricci L. Uncovering the Bitcoin Blockchain: An Analysis of the Full Users Graph[C]. *2016 IEEE International Conference on Data Science and Advanced Analytics*, 2016: 537-546.

[46] D. D. F. Maesa, A. Marino, and L. Ricci. An analysis of the bitcoin users graph: inferring unusual be- haviours[C]. *International Workshop on Complex Networks and their Applications*, 2016: 749-760.

[47] Harlev M A, Yin H S, Langenheldt K C, et al. Breaking Bad: De-Anonymising Entity Types on the Bitcoin Blockchain Using Supervised Machine Learning[J]. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2018, 2018-January: 3497-3506.

[48] Z. Zhang, T. Zhou, and Z. Xie. Bitscope: Scaling bitcoin address deanonymization using multi-resolution clustering[C]. *The 51st Hawaii International Conference on System Sciences*, 2018.

[49] X. Lv, Y. Zhong, and Q. Tan. A study of bitcoin de-anonymization: Graph and multidimensional data analysis[C]. *2020 IEEE Fifth International Conference on Data Science in Cyberspace*. IEEE, 2020:339-345.

[50] Elliptic: Blockchain analytics for crypto compliance.: https://www.elliptic.co/. 2020

[51] Chainalysis: The blockchain analysis company: https://www.chainalysis.com/ 2020

[52] Aml solution for digital assets: https://coinholmes.com/ 2020

[53] Blockchain ecosystem intelligence: https://www.anchain.ai/bei. 2020

[54] Blocksce Team: https://www.blocksecteam.com/. 2021

[55] Kappos G, Yousaf H, Maller M, et al. An Empirical Analysis of

Anonymity in Zcash[EB/OL]. 2018: arXiv: 1805.03180. https://arxiv.org/abs/1805.03180

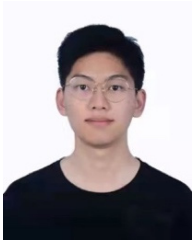
[56] Chen T, Zhu Y X, Li Z H, et al. Understanding Ethereum via Graph Analysis[C]. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018: 1484-1492.

[57] Experiment code: https://github.com/aaasdsada/BATscope. 2022.

王大宇 于 2019 年在北京邮电大学电子信息工程专业获得学士学位。现在清华大学计算机科学与技术专业攻读硕士学位。研究领域为系统安全方向。研究兴趣包括: 区块链安全, DeFi 数据分析等。Email: wdy19@mails.tsinghua.edu.cn

殷婷婷 于 2018 年在北京交通大学信息安全专业获得学士学位。现在清华大学网络空间安全专业攻读硕士学位。研究领域为系统安全。研究兴趣包括: 二进制安全、自动化漏洞挖掘、区块链安全等。Email: ytt18@mails.tsinghua.edu.cn

李赟 于 2019 年在北京邮电大学信息安全专业获得学士学位。现在清华大学系统安全专业攻读博士学位。研究领域为应用密码学, 研究兴趣包括: 多方安全计算、零知识证明。Email: liyun19@mails.tsinghua.edu.cn



秦嗣量 现在中国科学院大学网络空间安全专业攻读学士学位。研究领域为软件与系统安全。研究兴趣包括软件分析与测试、人工智能安全。Email: qinsiliang18@mails.ucas.ac.cn

任歆 现在厦门大学软件工程专业攻读学士学位。研究领域为区块链。研究兴趣包括: DeFi, rewritable blockchain。Email: mbt1809505@xmu.edu.my。

罗夏朴 于 2007 年在香港理工大学获得博士学位, 并在佐治亚理工学院担任 2 年博士后研究员. 现香港理工大学计算机系副教授. 主要研究方向为网络和系统安全、区块链和智能合约、移动和物联网安全。Email: csxluo@comp.polyu.edu.hk

王浩宇 于 2016 年于北京大学获得博士学位, 现为华中科技大学教授, 主要研究领域为移动安全和区块链系统安全等。Email: haoyuwang@hust.edu.cn

尹霞 于 2000 年在清华大学计算机科学与技术专业获得博士学位。现任清华大学计算机科学与技术系教授。研究领域为下一代互联网和协议测试。Email: yxia@tsinghua.edu.cn

张超 于 2013 年在北京大学大学计算机应用专业获得博士学位。现任清华大学网络科学与网络空间研究院副教授。研究领域为软件与系统安全。研究兴趣包括: 漏洞挖掘、漏洞利用、漏洞防利用、人工智能安全等。Email: chaoz@tsinghua.edu.cn