基于动机分析的区块链数字货币异常交易行为识别方法

沈蒙[1),2)] 桑安琪[1)] 祝烈煌[1)] 孙润庚[1)] 张璨[1)]

[1)](北京理工大学 计算机学院, 北京 100081)

[2)](密码科学技术国家重点实验室, 北京 100878)

摘 要 当前区块链数字货币被众多恶意交易者利用，导致了“粉尘”注入、“空投”操作、勒索、骗局等一系列异常交易行为。因此，研究区块链数字货币异常交易行为的识别方法对于规范交易行为、保障网络空间安全具有重要意义。在众多区块链数字货币中，比特币市值超过所有区块链数字货币市值和的一半，具有高代表性。比特币系统的用户数量多、交易规模大、地址匿名化等特性，为异常交易行为的准确识别带来巨大挑战。鉴于任何比特币异常交易行为背后都存在着明确的动机，本文以分析交易动机为切入点，设计了一种新颖的比特币异常交易行为识别方法。具体地，我们以空投糖果和贪婪注资两类异常交易行为作为典型代表，分别设计了两类异常交易行为的判定规则，进而抽象出异常交易模式图。在此基础上，利用子图匹配技术设计实现了比特币异常交易行为的识别算法。为了评估本方法的效果，我们收集了近 30 个月的比特币历史交易数据，通过人工分析确定了异常交易行为的真值集。实验结果显示，空投糖果行为的识别召回率为 85.71%、准确率为 43.62%，贪婪注资行为的识别召回率为 81.25%、准确率为 54.32%。此外，我们重点分析展示了三个比特币异常交易行为的典型实例，通过真实案例进一步验证了本文所提方法的有效性。

关键词 区块链；比特币；异常交易行为；动机分析；交易图

中图法分类号 TP309

Abnormal Transaction Behavior Recognition Based on Motivation Analysis in Blockchain Digital Currency

SHEN Meng[1),2)] SANG An-Qi[1)] ZHU Lie-Huang[1)] SUN Run-Geng[1)] ZHANG Can[1)]

[1)](School of Computer Science, Beijing Institute of Technology, Beijing 100081)

[2)](State Key Laboratory of Cryptology, P.O. Box 5159, Beijing 100878)

Abstract Due to the chaos in the current cryptocurrency market, blockchain digital currency is used by many malicious traders, leading to a series of abnormal trading behaviors such as "dust" injection, "airdrop" operations, extortion, and scams. Therefore, research on the identification method of abnormal transaction behavior of blockchain digital currency is of great significance for regulating transaction behavior and ensuring cyberspace security. Among the many blockchain digital currencies, the market value of Bitcoin exceeds half of the total market value of all blockchain digital currencies, and is highly representative. Bitcoin is the most successful blockchain application scenario at present and one of the most popular topics in the field of digital currency investment and research in the recent decade. The Bitcoin system has a large number of users, a large transaction

本课题得到广东省重点领域研发计划(No.2019B010137003)、国家自然科学基金(No. 61972039、No. 61872041)、北京市自然科学基金(No.4192050)资助.沈蒙，男，1988年生，博士，副教授，计算机学会(CCF)会员,主要研究领域为网络安全、云计算隐私保护.E-mail: shenmeng@bit.edu.cn.桑安琪，女，1996年生，硕士研究生,主要研究领域为网络与信息安全.E-mail: anqi_960123@163.com.祝烈煌（**通信作者**），男，1976年生，博士，教授，计算机学会(CCF)会员,主要研究领域为密码学、网络与信息安全.E-mail: liehuangz@bit.edu.cn.孙润庚，男，1997年生，硕士研究生,主要研究领域为网络与信息安全.E-mail: 17864290295@163.com.张璨，男，1996年生，博士研究生,主要研究领域为网络与信息安全.E-mail: canzhang@bit.edu.cn.

scale, and anonymization of addresses, which bring great challenges to the accurate identification of abnormal transaction behavior. So far, many researchers have focused on a particular type of illegal and abnormal trading behavior. But different from their method, given that there is a clear motivation behind any Bitcoin abnormal transaction behavior, this article designs a novel method for identifying Bitcoin's abnormal transaction behavior based on the analysis of transaction motivation. Specifically, we take the two types of abnormal transaction behaviors of airdrop candy and greedy capital injection as typical representatives, and design the two types of abnormal transaction behavior determination rules (i.e. judgment rules for airdrop candy behavior and greed injection behavior), and then abstract the abnormal transaction pattern diagram (i.e. airdrop candy behavior trading pattern and greedy capital injection behavior trading pattern). Based on this, the algorithm for identifying abnormal transaction behaviors of Bitcoin was designed and implemented using subgraph matching technology. In order to evaluate the effectiveness of this method, we collected the historical transaction data of Bitcoin for nearly 30 months, and determined the ground-truth set of abnormal transaction behavior through manual analysis. The experimental results show that the recognition recall rate of airdrop candy behavior is 85.71%, the accuracy is 43.62%, the recognition recall rate of greedy fund injection behavior is 81.25%, and the accuracy is 54.32%. In addition, we focus on the analysis and display of three typical examples of Bitcoin's abnormal transaction behavior (i.e. "dust" injection behavior, WannaCry ransomware, SOXex exchange scam), and further verify the effectiveness of the method proposed in this paper through real cases. At the same time, it also shows that there are many abnormal trading activities in the cryptocurrency market, and the cryptocurrency investment market is constantly being disrupted. Therefore, research to identify Bitcoin's abnormal trading behavior has the potential to provide insights into the wider cryptocurrency ecosystem and the trading behavior of thousands of digital currencies now included. It can also help Bitcoin investors understand the dangers of investing in the market and reduce investment risk in the market. In addition, it is more conducive for national authorities to use the abnormal transaction behavior of cryptocurrencies to regulate investors' investment behavior.

Key words blockchain; Bitcoin; abnormal trading behavior; motivation analysis; transaction graph

1 引言

比特币是迄今为止最为成功的区块链应用场景，也是近十年在数字货币投资领域和研究领域中最流行的话题之一。由 2008 年化名为“中本聪”的学者发表的比特币奠基性的论文可知，比特币是由密码学支持保护的、可以在参与者之间实现价值转移的数字货币，也是最著名的首个具有变革潜力的、不依赖第三方权威的分布式加密货币[1]。此外，由于比特币允许用户在比特币网络中使用一个与真实身份无关的假名，并且无需地址重用，所以比特币具有一定的匿名性。同时，因为具有分布式和匿名性这两个特点，近些年比特币吸引大量用户资本，积累大量交易数据。据统计，比特币在全球至少有 1000 万的用户使用量，每天的交易量高达 2 亿美元[2]。比特币的实际价值和用户量的潜在价值是投资者参与投资比特币的主要因素。一方面，如图 1 所示，由历史均价可知，比特币的实际价值非常高。另一方面，比特币庞大的基础用户量和匿名性，使得其用户的身份更容易被隐藏。所以许多不怀好意之人正是利用了比特币的这种高价值属性与匿名性来实施恶意行为。

本文主要关注“粉尘”注入、“空投”操作、勒索、骗局等比特币异常交易行为。这些异常交易行为普遍存在，并造成了重大的经济损失。如，SOXex 交易平台利用假空投福利和高回报活动诱导投资者注资，最终套现约 4000 万人民币[3]；2019

1 S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf

2 C. Burniske, A. White. Bitcoin: Ringing the bell for a new asset class. 2019. [Online]. Available: https://research.ark-invest.com/hubfs/1Download Files ARK-Invest/ White Papers/Bitcoin-Ringing-The-Bell-For-A-New-Asset-Class.pdf

3 https://www.qubi8.com/archives/309081.html

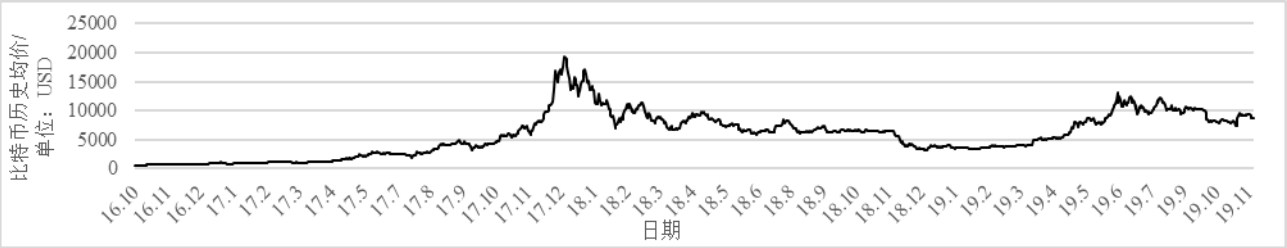

图 1　比特币历史均价趋势图

年 1 月至 3 月，仅 3 个月的时间就出现了 6 起比特币勒索事件 [1]。本文关注的异常交易行为中的“粉尘”是指微量的币，一般这种金额量级的币不会被用于进行交易。但很多比特币地址会被注入“粉尘”，因为这样有助于攻击者破坏比特币的匿名性，从而实现对这些地址的追踪。同时，近几年出现了一种称作“空投”的商业行为，利用比特币的大量用户基础，以零成本进行传播，创造营销高潮。交易平台通过“空投”操作来吸引大量用户进行交易，从而赚取佣金；新币发布利用“空投”操作来吸引场外资本；犯罪分子试图通过“空投”操作非法获取更多用户的数字资产，利用自己的资金进行资本化。勒索行为则通常会采用禁止访问文件的方式，来向用户索要加密数字货币。骗局会采用欺骗的方式让用户认为其是有利可图的，但其本质是骗局的发起者将用户的数字资产转移进自己所有权的账户。由此可知，每一个异常交易行为背后都有其对应的动机。因此，从交易动机出发，有助于更好地实现对比特币异常交易行为的识别。

比特币独特的自身属性（如用户数量多、交易规模大、地址匿名化）使得界定与识别上述异常交易行为面临诸多挑战。首先，由于一个用户可以生成多个比特币地址，使得整个账本数据中涵盖了大量不重复的比特币地址。同时这些地址也都涉及了海量的比特币交易。因此识别比特币异常交易行为这一过程需要基于海量复杂的比特币交易数据，分析效率低、计算量大。第二，比特币的整体交易模式为多对多（即输入和输出地址均可以是多个）。这种交易模式使得交易中的输入和输出地址之间缺乏显式的链接。同时这种弱链接也更有效地在大量交易数据中模糊了潜在比特币异常交易行为的特征，使得难以通过分析地址链接的方式识别基于比特币的异常交易行为。

本文从比特币异常交易行为的动机分析入手，以比特币地址集群为单位进行探索研究。根据异常行为的交易动机提取其交易特征，并设计了异常交易行为的判定规则。然后基于判定规则构建交易模式图，从而提出利用子图匹配技术来进行比特币异常交易行为的识别。最后采用真值匹配和实例验证的方式对识别方法进行了验证分析。实验结果表明，本文的识别方法可有效识别出比特币异常交易行为。此识别方法有利于规范加密货币市场及其内部的交易行为。同时，能为用户提供更加安全的服务，以及更加健康的投资环境。

本文的主要贡献包括三方面：

- 提出了比特币系统中空投糖果行为和贪婪注资行为的判定规则，进而抽象出两类异常交易行为的交易模式，可以将地址集群与“粉尘”注入、“空投”操作、勒索、骗局等异常交易行为进行关联。
- 提出了一个基于动机分析，利用子图匹配技术有效识别比特币异常交易行为的方法。并构建了一个比特币异常交易行为真值集。基于此数据集，识别方法的空投糖果行为召回率为 85.71%、准确率为 43.62%，贪婪注资行为召回率为 81.25%、准确率为 54.32%。
- 通过研究分析“粉尘”注入、WannaCry 勒索事件、SOXex 交易所骗局的真实案例，证明本文所提出的比特币异常交易行为识别方法的有效性。本方法可以帮助降低比特币投资者的市场投资风险，并对加密货币市场及其内部的交易行为进行规范。

2 问题定义

2.1　比特币交易数据

“未花费的交易输出”，即 UTXO（unspent transaction outputs），是比特币交易的基本构建单元和价值单元[1]。本质上，UTXO 的出现是为了防止发生双重花费[2]。且众所周知，一个比特币用户可以拥有任意多个地址，一笔交易也可能同时涉及多个地址，即一笔合法的比特币交易可以有多个输入

1 https://forex.cngold.org/c/2019-03-14/c6267409.html

和多个输出。除此之外，比特币交易是可能存在所谓的“找零”的。出于隐私保护，找零地址通常与输入地址不同，采用当前所有者的一个新地址。同时，比特币交易输入总额和输出的总额不需要相等，不管是否存在找零地址。当输入总额略高于输出总额，两者的差值就是手续费，被称为“矿工费”。

如图 2 描述了一笔典型的记账记录比特币交易 TX。此交易示例有 2 个输入地址（*A1* 和 *A2*）和三个输出地址（*B*，*C* 和 *Ac*），其中地址 *A1*、*A2* 和 *Ac* 属于同一个用户 *A*，*Ac* 是用户 *A* 在 TX 交易中新生成的找零地址。由图可知，地址 *A1* 转移了 2.1 比特币（BTC），地址 *A2* 花费了 4.5BTC 的 UTXO；地址 *B* 接收到了 1BTC，地址 *C* 新生成了一个 5BTC 的 UTXO；找零用户 *A* 0.5BTC。图中输入总额与输出总额的差为 0.1BTC，是此交易的矿工费。

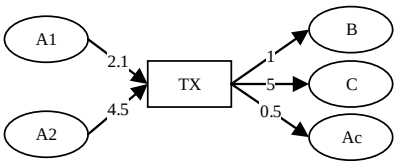

图 2 典型比特币交易示意图

在比特币交易中存在着一个例外——“Coinbase 交易”，如图 3 所示。它是每个区块中的第一笔交易，没有输入，不消耗 UTXO，可以产生新的比特币。其目的是生产新的可花费的比特币，用作对“赢家”矿工挖矿的奖励。

```
{“inputs”: [{“address”: “coinbase”}],
 “blocktime”: 1573624517,
 “blockhash”: 000000000000000000055785d6eaec5c349a30c63cae5
              66d76235b061852cb7e,
 “outputs”: [{“value”: 12.84336697,
             “address”: “bc1qjl8uwezzlech723lpnyuza0h2cdkvxvh5
                        4v3dn”}]
}
```

图 3 Coinbase 交易实例图

2.2 典型的比特币异常交易行为

2.2.1 空投糖果行为

空投糖果行为的本质是按一定规则免费发放加密货币。这里泛指在一段时间内，大量数字货币持有者的账户中无故（或因为前期简单操作，如注册等）多出一部分数字资产的现象。这些数字资产的金额可能非常小，也可能价值非常相近。

一方面，空投糖果行为的发起者可能会为了在用户追加投入后卷款跑路，而前期增加自身用户量，提高自身的用户使用率。另一方面，空投糖果行为的发起者可能会利用注入的“粉尘”获取用户未知信息，从而非法获利。所谓的“粉尘”注入行为，就是向大量用户比特币地址发送“粉尘”。恶意者通过关联该比特币地址，追踪此用户的交易活动，从而找出与此用户相关的其他比特币地址。然后分析得到一个地址集群背后的所有者，从而破坏比特币本身的匿名性。

这种投资行为破坏了数字货币投资的公平性，在某种程度上存在产生跟踪地址行为或者欺诈用户行为的可能，比如“粉尘”注入行为 [1]、OMG 空投事件 [2]、EOS 空投事件 [3]等。空投糖果行为的发起者可以是基于区块链的服务、交易平台或者一种新代币的所有者，乃至普通用户等等。

2.2.2 贪婪注资行为

贪婪注资行为通常是在一定时间段内存在某个或者某几个数字货币账户收到大量转账交易的行为。这里的“贪婪”具体指代两种不同角度的“贪婪”。一种是恶意用户通过非法的方式，贪婪地向普通用户索取加密数字货币资产。另一种是贪婪用户希望通过注资这种行为（如，投资加密数字货币）来获得财富增长。

对于第一种角度，以勒索行为为例。勒索行为是指威胁用户，强行索要以比特币为主的赎金的行为，即通过非法占有的手段来获取非法收入。比如，全球最大勒索病毒 GandCrab 就是一款比特币勒索病毒，截至目前，其获利资金已高达 20 亿美金，平均每周获利 250 万美元 [4]。

对于第二种角度，许多骗局都是符合的。即从表面上看数字货币持有者可以轻松获利，但实际上这些“政策”和机会是为了让更多的用户尽力去投资，也就是一种非法集资诈骗行为。攻击者利用项目的高回报率，或者更优惠的汇率来吸引用户的注意，诱惑用户投资。但实际上，这些被吸引的用户最终大部分都没有真实得到他们应得的东西。而他们的数字资产早已被转移到骗局发起者的账户。像 BTC Promo、btcQuick 和 CoinOpend 交易所都已被确认为比特币交易所骗局。

1 https://www.jinse.com/news/blockchain/448255.html

2 https://baijiahao.baidu.com/s?id=1638477545580119343

3 http://www.zyfree.net/QuKuaiLian/2018-03/889.htm

4 https://www.360kuai.com/pc/9c919ca602771a128?cota=3&kuai_so=1&sign=360_57c3bbd1&refer_scene=so_1

3 相关工作

区块链是应用于比特币等加密货币的底层技术，具有开放性和匿名性，所以截至目前，有大量围绕加密货币去匿名，试图挖掘用户隐私的研究。早期的加密货币去匿名研究主要集中在比特币上[3-6]。通过采用目前所熟知的“多输入”聚类规则和找零地址的方式进行实体识别[3,4]。我们在第 4 节中也采用了这两种方法来获得共享所有权集群。

随着区块链技术的应用越来越成熟，也陆续出现了针对其他加密货币的去匿名研究工作。包括 Ripple（瑞波币）[7]、Dash（达世币）[8,9]、Monero（门罗币）[10,11]和 Zcash（零币）[12]的单币种传统去匿名工作，以及文献[13]跨多个币种和文献[14,15]结合暗网的去匿名工作。

但是现有的针对加密货币的去匿名研究工作，并不能阻止因人们滥用区块链匿名性而形成的正在不断泛滥的非法异常加密货币交易行为。如，洗钱[3,16]、勒索[17-19]、各种骗局[20,21]、市场操控[22]、暗网非法交易[15]等，具体可参看文献[23]。所以，从区块链交易中识别出特殊的交易模式，从而发现相关的非法异常交易行为是一个非常值得探索的问题。有不少研究人员专注于某一特定类型的非法异常交易行为进行研究。下面重点介绍基于区块链技术的洗钱、勒索和骗局的相关研究。

洗钱是将非法所得合法化的一种加密货币非法服务，主要通过各种手段隐瞒非法所得的来源和性质，使其在形式上合法化。文献[16]利用 3 种增强交易匿名性的洗钱服务，来系统性说明反洗钱政策在比特币上的效果和局限性。文献[3]概括了 3 种典型疑似洗钱的交易模式：汇聚、折叠和分割。

勒索软件通常利用封锁文件的方式向用户勒索非法赎金。文献[17]使用数据驱动方法识别 35 个勒索软件相关交易，并经验分析这些勒索软件造成的最直接的经济损失。文献[19]以勒索软件公开的比特币地址为起点，通过分析比特币交易数据，研究关联交易，最终找到了 968 个属于该组织的地址，识别出价值 1128.40BTC 的赎金交易。

文献[20]通过整合可获得的各种骗局报告，最终获取 192 个基于比特币的骗局案例，使我们可以了解到当前比特币骗局的大规模和严重性。并且这些骗局被划分为 4 种不同类型：庞氏骗局、挖矿骗局、诈骗钱包和虚假交易所等。文献[21]是利用数据挖掘方法分析比特币庞氏骗局的研究。

除此之外，还有不少研究将区块链网络可视化，以检测比特币交易中潜在的异常或特殊交易模式，从而识别出非法异常交易行为[24,25]。现有的研究还没有基于动机分析的解决方案，但是由第 2 章问题描述可知，其实每个异常交易行为都存在一个明确的动机。所以区别于以上分析异常交易行为的方法，本文提出基于动机分析来进行比特币异常交易行为的检测。

4 比特币异常交易行为识别模型

4.1 方法概述

本文按图 4 所示的技术路线，对比特币交易背后的异常行为进行了识别与实证探索分析。首先根据对异常交易行为的动机分析提出了空投糖果行为判定规则和贪婪注资行为判定规则，并对它们进行了分析和讨论。然后将两个判定规则分别转化为可以用于识别异常交易行为的交易模式，即空投糖果行为交易模式和贪婪注资行为交易模式。从而提出了基于动机分析的比特币异常交易行为识别方法。最后本文通过真值匹配和识别真实案例的实验来验证识别方法的有效性，详细内容会在第 5 章和第 6 章进行进一步的阐述。

4.2 异常交易行为识别方法设计

4.2.1 异常交易行为动机分析

（1）空投糖果行为动机分析

如 2.2.1 节所述，空投糖果行为可能是一种异常交易行为。这种行为的主要特征是在一定时间段内，大量用户接收到具有一定规则性的转账交易，并且这些转账交易发起源于一个或一组数字货币账户的交易行为。这种转账行为表现出具有一定的发散性，可能包含以下四种动机，涉及项目方（或犯罪分子）Pro、用户 User 以及资金 Ac 三种实体。

第一种，Pro（新币或新交易所）初期的商业行为，为了提高人气。Pro 试图尽可能多的将数字货币持有者吸引为他们真正的 User。

第二种，项目上市后 Pro 的商业行为，为了提高影响力。增加 User 参与度和提高影响力，有助于发掘更多的潜在 Ac。这种行为不一定是具有恶意的（如市场操控），可能只是正常的商业行为。

第三种，Pro 为了筹措 Ac。从合法的角度来考虑，可以是 Pro 为了资助项目的未来发展和建设。从非法的角度来考虑，可能是 Pro 为了吸引大量的

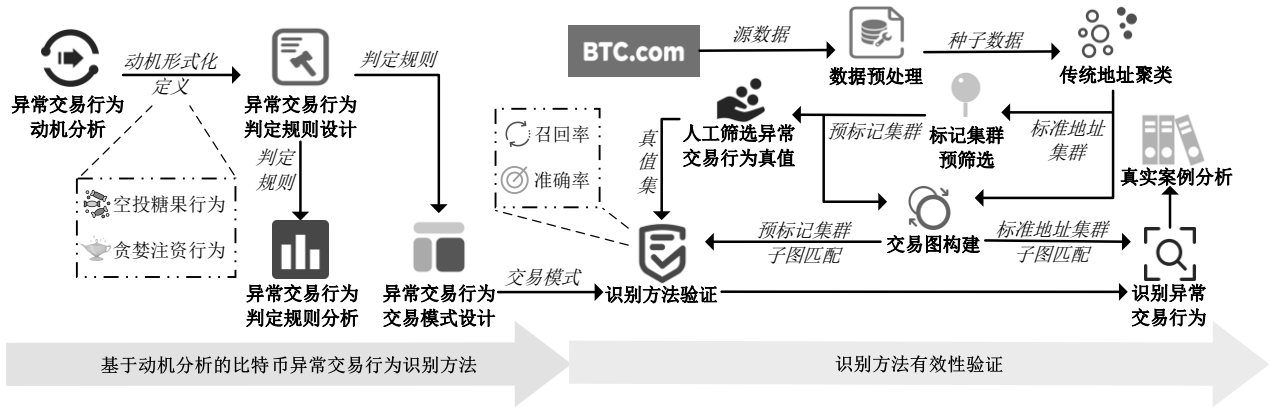

图 4 比特币异常交易行为识别方法技术路线

User 关注后，采取利诱等非法恶意行为来骗取 User 的 Ac。

第四种，不同于前三种，可能是 Pro 为了恶意追踪 User 地址，即“粉尘”注入行为。Pro 有意地识别地址的真实所有权，帮助其进一步实施诈骗等非法活动。

（2）贪婪注资行为动机分析

如 2.2.2 节所述，贪婪注资行为的主要特征是在一定时间段内，转账行为具有汇聚性，即存在大量用户向某个或者某几个数字货币账户进行转账操作的行为。其包含两种不同角度的“贪婪”，对应也有如下两大类不同的动机，涉及勒索操纵者 Black、骗局发起者 Scam、用户 User 以及资金 Ac 四种实体。

第一类，站在非法 Black 的角度。Black 试图通过锁定 User 的文件，来威胁大量 User 向 Black 的地址集群进行存款交易，从而贪婪地获得大量非法 Ac，实现快速获利。

第二类，先站在 User 的角度。User 通常会被所谓的优惠汇率、高回报等推销手段所吸引，并想要尽快注资，企图获得更多的 Ac。再站在 Scam 的角度。Scam 正是利用了 User 的这种贪婪心理，来欺骗 User 投资，从而骗取 User 的 Ac。

4.2.2 异常交易行为判定规则设计

（1）空投糖果行为判定规则设计

我们根据上述动机分析，结合空投糖果行为特征，提出了第一个异常交易行为判定规则：

空投糖果行为判定规则. 在 T_1 时间内，同一所有权的地址集群中，若存在将 A 个近似金额（浮动范围为 Gap）发送到不包含在此集群内的其他数字货币持有者的地址，则认为此地址集群存在空投糖果行为。

就误报而言，最明显的一个误报就是 Coinbase 交易，其输出地址可以归属于同一个实体[26]。我们采用删除 Coinbase 交易的方法来防止误报。因为如图 3 所示，Coinbase 交易在其输入的地址处是有特殊标识的，所以通过识别这个特殊的标识，可以在应用此判定规则之前排除 Coinbase 交易，从而大大降低误报的概率。

（2）贪婪注资行为判定规则设计

我们根据上述动机分析，结合贪婪注资行为特征，提出了第二个异常交易行为判定规则：

贪婪注资行为判定规则. 在 T_2 时间内，同一所有权的地址集群接收到 B 笔高于正常值（该集群历史平均接收交易金额 c）$Gmul$ 倍的交易，且这些交易由不包含在此集群内的其他数字货币持有者地址发送，则认为此地址集群存在贪婪注资行为。

就误报而言，可能存在同一个用户将自己所有的比特币资产在一段时间内汇入一个或几个自己新拥有的账户。这种交易行为特征与贪婪注资行为非常类似。但经过统计，在贪婪注资行为真值中产生这种误报的概率几乎为零，足以被忽略。

此外，上述两个判定规则中的时间参数（即，T_1 和 T_2）或金额参数（即，Gap 和 $Gmul$）如果选取的不恰当，也可能会产生误报或者漏报。即，一种是把正常的集群判定为异常集群。另一种是异常的集群没有被判定出来。空投糖果行为判定规则可以通过增加 T_1 或增大 Gap 来减少漏报，但代价是增加了更多计算，并且可能增加误报。也可以通过减少 T_1 或减小 Gap 来减少误报，但很可能增加漏报。贪婪注资行为判定规则可以通过增加 T_2 或减少 $Gmul$ 来减小漏报率，但代价是增加了更多计算，并且可能增加误报。也可以通过减少 T_2 或增加 $Gmul$ 来减小误报率，但很可能会增加漏报率。

4.2.3 异常交易行为交易模式设计

我们将一个地址集群定义为有向图中的一个

节点 b，同时，有向图中的每条边 e 都代表了两个节点间的至少一笔直接交易。当由地址集群 m_1 向地址集群 v_1 产生了一笔发送交易时，则存在一个有向边(m_1,v_1)。当由地址集群 m_2 从地址集群 v_2 获得了一笔接收交易时，则存在一个有向边(v_2,m_2)。这些有向边都具有权重，此值等于沿此边进行的一笔或多笔交易所转移的比特币总量加权。

具体地，本文提出的交易及交易图构建方式如下所示。

定义 1(交易)**.** 交易 $D_T=\{b_s, b_r, w, d\}$是一个四元组. 其中，b_s 和 b_r 分别表示交易的发送方集群和接收方集群，w 表示此交易的交易金额（单位为 BTC），d 是交易中地址集群的角色标识。也就是当此笔交易中的发送方（优先考虑）地址集群 s 存在空投糖果行为时，d 被赋 0 值；当此笔交易中的接收方地址集群 r 存在贪婪注资行为时，d 被赋 1 值；其余情况 d 被赋值为-1。

定义 2(交易图)**.** 交易图 $G=\{N, E, W\}$是一个三元组. 其中 N 代表一组节点，即是一组 m_1 或一组 m_2。E 是对应连接 N 的一组边，即(m_1,v_1)或(v_2,m_2)。W 是一组边的权重函数，即 E 中每条边的权重都是沿这条边转移的比特币总量之和。

为了能更加细致地区分、识别和分析不同的异常交易行为，我们根据 4.2.2 节所述的异常交易行为判定规则，设计了空投糖果行为交易模式和贪婪注资行为交易模式。

（1）空投糖果行为交易模式

空投糖果行为交易图由多个空投糖果行为节点（N_k）与他们作为发送方的交易（E_k）及其权重（W_k）所组成。每一个空投糖果行为节点都是符合空投糖果行为交易模式的。我们根据空投糖果行为判定规则可知，空投糖果行为交易模式要求交易图中的节点 b 及其边同时满足：e 的方向为(m_1,v_1)；$w(1\pm Gap)$的出度数$\geq A$；这 A 笔交易的时间戳最大差值$\leq T_1$。

若此 b 及其边满足以上要求，则标记此 b 为空投糖果行为节点（b_k）。同时更新其作为 b_s 的交易元组中的 d，赋值为 0。并且 b_k 其实是具有实际意义的空投糖果行为地址集群，所以其中所包含的比特币地址也被标记为空投糖果行为地址。

图 5 所示为空投糖果行为交易模式示意图，其中用不同填充的点和实线箭头共同表示空投糖果行为交易模式。值得注意的是，图中用大圆圈圈起来的一个或多个地址，表示拥有共同所有权的地址集群。并且图中用箭头的宽度表示两个地址集群之间的交易金额（交易金额大，则箭头宽；交易金额小，则箭头窄）。此交易模式代表存在共享所有权的空投糖果行为地址集群向集群外的其他地址发起存款交易的资金流量。

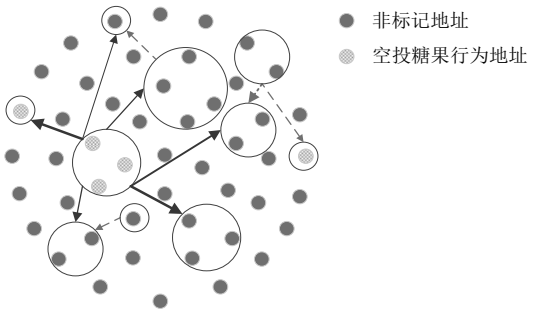

图 5 空投糖果行为交易模式示意图

（2）贪婪注资行为交易模式

贪婪注资行为交易图由多个贪婪注资行为节点（N_t）与他们作为接收方的交易（E_t）及其权重（W_t）所组成。每一个贪婪注资行为节点都是符合贪婪注资行为交易模式的。我们根据贪婪注资行为判定规则可知，贪婪注资行为交易模式要求交易图中的节点 b 及其边同时满足：e 的方向为(v_2,m_2)；$w/c\geq Gmul$ 的入度数$\geq B$；这 B 笔交易的时间戳最大差值$\leq T_2$。

若此 b 及其边满足以上要求，则标记此 b 为贪婪注资行为节点（b_t）。同时更新其作为 b_r 的交易元组中的 d，赋值为 1。并且 b_t 其实是具有实际意义的贪婪注资行为地址集群，所以其中所包含的比特币地址也被标记为贪婪注资行为地址。

图 6 所示为贪婪注资行为交易模式示意图，其中用不同填充的点和实线箭头共同表示贪婪注资行为交易模式；用大圆圈圈起来的一个或多个地址，表示拥有共同所有权的地址集群；用箭头的宽度表示两个地址集群之间的交易金额。此交易模式代表存在共享所有权的贪婪注资行为地址集群从集群外的其他地址接收到存款交易的资金流量。

4.3 检测算法

基于前文所述方法，我们设计了算法 1——识别空投糖果行为的检测算法。其中设 G 为带权有向图，Gap 表示权值浮动范围，满足 $Gap\in[0,1]$；A 表示满足条件的输出数量下限；W_b 表示 G 中以 b 为起点的边的权值集合；T_b 表示筛选出的以 b 为起点的边的时间集合；T_{diff} 表示 T_b 中最早时间和最晚时间的差；T_{begin} 表示节点筛选范围的起始天数；T_{end} 表示节点筛选范围的结束天数；Day 为常量

24*60*60。

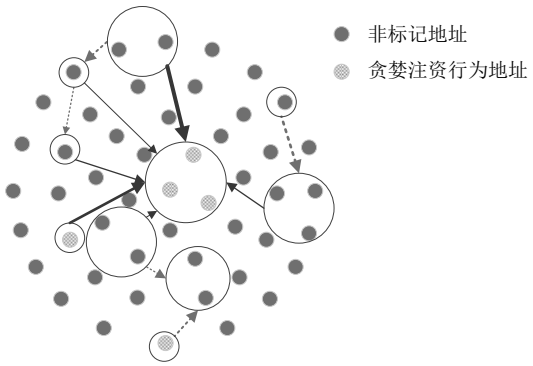

图 6 贪婪注资行为交易模式示意图

算法 1. 空投糖果行为检测算法.

```
输入：G=(N, E, W)
输出：空投糖果行为节点集合 Q
Q = {}
FOR b∈N DO
    W_b.sort() #对权值进行升序排序
    FOR i = 1 TO W_b.length() DO
        count = 0 #计数器
        FOR j = i - A TO i + A DO
            IF (abs(W_b[i] – W_b[j]) <= W_b[i] * Gap)
                count = count + 1
            ENDIF
        ENDFOR
        IF ( count >= A )
            IF ( T_diff >= Day * T_begin and T_diff <= Day * T_end )
                Q = Q ∪ {b}
            ENDIF
        ENDIF
    ENDFOR
ENDFOR
输出空投糖果行为节点集合 Q
```

算法 1 时间复杂度分析：G 中的节点数量用 n 表示，每个节点的发送交易数用 m 表示，则空投糖果行为检测算法的时间复杂度为 O(nmA)。由于执行算法时 A 为常量，所以时间复杂度近似为 O(nm)。在第 6 章的实验中，经过我们多次测试，可知算法 1 平均耗时为 14 毫秒，其效率可以接受。

算法 2 描述了识别贪婪注资行为的检测算法，其中 $Gmul$ 表示权值需满足的倍数；B 表示满足条件的输入数量下限；W_b 表示 G 中以 b 为终点的边的权值集合；其余变量含义与算法 1 相同。

算法 2. 贪婪注资行为检测算法.

```
输入：G=(N, E, W)
输出：贪婪注资行为节点集合 Q
Q = {}
FOR b∈N DO
    W_b.sort() #对权值按交易时间进行升序排序
    i = 1
    WHILE i < W_b.length() – 1 DO
        j = 1
        IF ( W_b[i+1] / W_b[i] ) >= Gmul )
            count = 0 #计数器
            WHILE j < W_b.length() – 1 DO
                IF ( W_b[i+j] / W_b[i] ) >= Gmul )
                    count = count + 1
                    IF ( count >= B )
                        IF ( T_diff >= Day * T_begin and T_diff <= Day * T_end )
                            Q = Q ∪ {b}
                            GOTO(5)
                        ENDIF
                    ENDIF
                ENDIF
            ENDWHILE
        ENDIF
        i = i + j
    ENDWHILE
ENDFOR
输出贪婪注资行为节点集合 Q
```

算法 2 时间复杂度分析：G 中的节点数量用 n 表示，每个节点的接收交易数用 m 表示，则贪婪注资行为检测算法的时间复杂度为 O(nmlog(m))。在第 6 章的实验中，经过我们多次测试，可知算法 2 平均耗时为 5 毫秒，其效率可以接受。

5 数据采集与处理

5.1 数据收集与解析

本文使用 BTC.com[1]区块链浏览器的多个公共接口下载了近 30 个月（从 2017 年 5 月 1 日到 2019 年 11 月 9 日）的历史比特币区块链数据。接下来利用自定义的 Python 脚本对获得的交易详细信息进行解析。并将解析后的详细比特币区块链数据存储在数据库文件中。图 7 显示了根据我们获得的历史比特币区块链数据描绘的比特币交易量和交易数随时间变化的曲线。将两条曲线分开来看，比特币交易量的整体趋势随着时间的推移而增长，但是在这个过程中也存在着一些峰值。这些峰值与经济增长（或下降）和政府的支持有关。再来看比特币交易数曲线，其整体趋势随时间推移而趋于平稳。将比特币交易数曲线与图 1 的比特币价格曲线相融合，可以看出融合后的曲线基本与比特币交易量曲线的趋势相一致。

最后，我们借鉴文献[27,28]中的方法，将解析得到的比特币区块链数据进行人工预处理。即删除重复交易，并删除验证无效的地址以及其对应的交易，从而得到种子数据。然后使用自定义的脚本对这些种子数据信息进行了相应的分析。

1 https://btc.com/

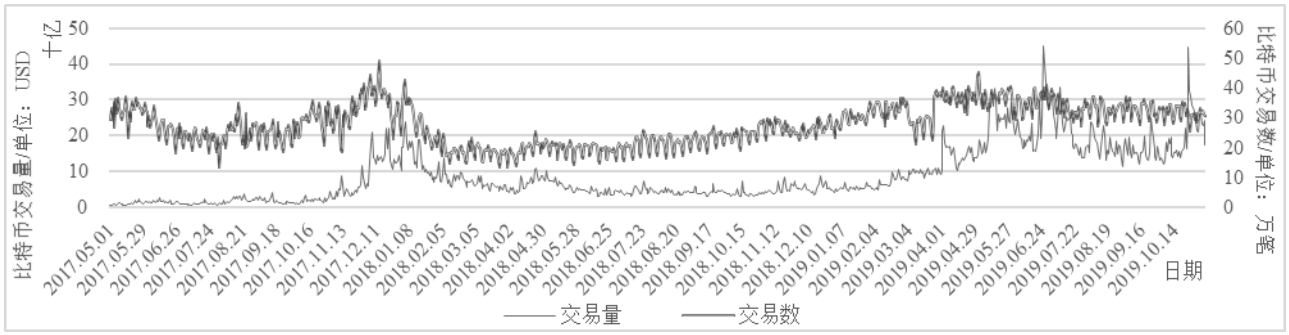

图 7 比特币历史交易量和交易数趋势图

5.2 传统地址聚类

为了更好地利用历史比特币区块链数据分析出更多有意义的信息，我们首先考虑使用传统的“多输入”比特币地址和找零地址聚类规则，将属于同一个实体（个人、机构或者服务）的地址进行聚类[19]。虽然聚类结果的准确性会受 CoinJoin 等增强隐私的技术的影响，但是现在已经存在相应的解决方法和技术，来恢复这种“多输入”聚类规则的聚类效果[9]。不过由文献[3]可知，这种找零地址聚类规则方法仍存在误报。

所以我们对已经利用上述两种规则聚类产生的集群内的比特币地址数量设定了一个阈值，即每个集群的地址数量不超过 90 万个。如果聚类结果规模超出这个阈值，则选择保留其对应的只运行“多输入”聚类规则后的集群规模。

最终我们将得到的所有集群按规模从大到小的顺序进行编号。其中我们使用的聚类结果如图 8 所示，共有 5642 个标准地址集群（包含 6,938,481 个比特币地址）。其中 55.48%的集群只包含 1 个比特币地址，还有大部分（42.54%）的集群是包含大于 1 个且不超过 10 个的比特币地址。由此可以看出，在庞大数量的比特币用户中，还是小规模实体居于主要地位（比如个人用户）。但同时，我们在后面人工筛选的过程中发现，在进行大量转账交易的情况下，确实也存在着很多只作为一次性转账中介的比特币地址。

5.3 真值获取

目前针对本文所关注的异常交易行为，未找到公开确定的数据集。所以本文通过先由程序粗略筛选，再由人工细致筛选的方式确定两组不同的异常交易行为真值。在此过程中，由于每种异常交易行为都要满足其交易数量的特征，所以我们先利用程序在标准地址集群中进行一轮预筛选，从而得到每种异常交易行为的预标记集群。

这里我们设置预筛选的规则为：每个空投糖果行为预标记集群应包含至少 40 笔发送交易；每个贪婪注资行为预标记集群应包含至少 40 笔接收交易。最终我们获得了 94 个空投糖果行为预标记集群和 162 个贪婪注资行为预标记集群。

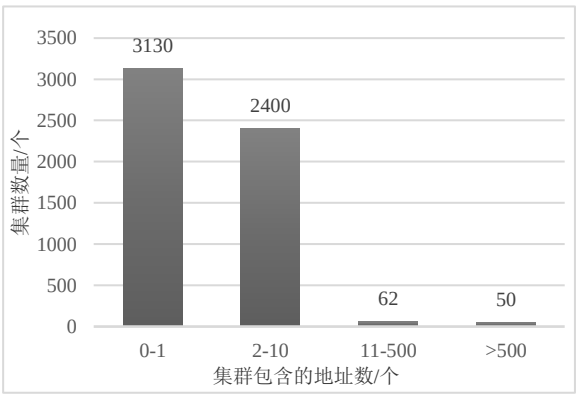

图 8 地址聚类结果图

然后我们要求 10 名安全研究人员根据下述给定的详细筛选规则，分别对两组异常交易行为预标记集群中的每个地址集群通过查询 BTC.com 和 WalletExplorer[1]进行细致的筛选。对于某一个地址集群，若有 8 名安全研究人员判断其是异常交易行为，则此地址集群为该异常交易行为的真值。

两组异常交易行为人工经验筛选规则中的时间参数和金额浮动/倍数参数，分别根据公开资料[2][3]进行确定。交易数参数是经过统计分析 5642 个标准地址集群的交易情况后，在上述时间和金额参数的背景下，以百分位数值（空投糖果行为是第 99.9 百分位，贪婪注资行为是第 99.8 百分位）进行确定的。当存在真实公开的可参考样本数据集之后，上述所有参数均可根据新样本集进行调整。具体每种异常交易行为的人工经验筛选规则和真值

1 https://www.walletexplorer.com/

2 https://www.tdapk.com/

3 http://www.120btc.com/zt/20/

结果如下所述。

5.3.1 空投糖果行为真值

基于 2.2.1 节描述的空投糖果行为的特征，我们规定空投糖果行为人工经验筛选规则如下：

（1）以集群为单位，存在大于等于 40 个金额较为固定（相近）的输出。

（2）满足（1）的所有交易的时间戳差值在 1 天到 30 天内。

同时考虑到执行空投糖果行为的通常是一些需要吸引大量用户注意的新项目方，或者试图窥探别人隐私（比如追踪用户资金流量）的恶意分子。所以筛选出的集群，需要尽量满足如下（3）所述的规则。

（3）集群内的地址第一次发生交易的时间戳与满足（1）交易的时间戳之差小于 5 个月，且相差的这段时间内，地址所涉及的交易数量小于 100 笔。

上述人工筛选规则可以组成两种组合，即（1）（2）和（1）（2）（3）。如果某一地址集群满足两种中的一种，则可以确定为空投糖果行为真值。且满足第二种组合的真值结果更优。

此外，“粉尘”注入行为作为空投糖果行为的一种特殊形式，在比特币区块链中，一般情况下满足：

（4）一笔交易的手续费大于此交易金额的 1/3。

如果在上述两种组合的基础上满足（4）则可以进一步确定此集群的异常交易行为是“粉尘”注入行为。

我们经过人工细致筛选，最终得到了 7 个符合要求的集群，其中 1 个集群被确定为“粉尘”注入行为。我们将这些集群作为后续研究所需的空投糖果行为真值。

5.3.2 贪婪注资行为真值

基于 2.2.2 节描述的贪婪注资行为的特征，我们规定贪婪注资行为人工经验筛选规则如下：

（1）以集群为单位，存在大于等于 40 笔相对金额较大的接收交易。这里相对金额较大是指，集群在这段时间内的平均接收交易金额大于等于在这段时间之前的历史平均接收交易金额的 10 倍。

（2）满足（1）的所有交易的时间戳差值在 1 小时到 9 个月内。

同时考虑到这些潜在非法的贪婪注资异常交易行为普遍生存周期较短，具有一定的突发性。所以筛选出的集群，需要尽量满足如下（3）所述的规则。

（3）集群地址活跃时间小于等于 6 个月（即，生存周期较短），甚至此集群没有历史交易。

上述人工筛选规则可以组成两种组合，即（1）（2）和（1）（2）（3）。如果某一地址集群满足两种中的一种，则可以确定为贪婪注资行为真值。且满足第二种组合的真值结果更优。

我们经过人工筛选，我们得到了 16 个符合要求的集群，并将这些集群作为后续研究所需的贪婪注资行为真值。

6 实验验证与分析

6.1 实验设置

6.1.1 实验环境

本文实验通过 Python 实现。处理器为 Intel(R) Xeon(R)，24 核；内存 256GB；硬盘 3.6T；操作系统为 Windows 10，64 位；运行环境为 Python 3.7。

6.1.2 数据集

本节采用的数据集全部来源于第 5 节中描述的种子数据，我们使用其经过聚类处理后的第 5.2 节标准地址聚类结果的部分结果。即，使用以下三个集合来验证本文提出的识别方法。（1）真值集合：该集合包括第 5.3.1 节中描述的 7 个空投糖果行为真值，以及第 5.3.2 节中描述的 16 个贪婪注资行为真值；（2）预标记集群集合：该集合包括第 5.3 节中描述的 94 个空投糖果行为预标记集群和 162 个贪婪注资行为预标记集群；（3）标准地址集群集合：此集合包含的是图 8 描绘的 5642 个标准地址集群。

6.1.3 评价指标

本文使用常见的召回率和准确率作为评价指标，并结合识别出比特币异常交易行为实例，来对本文提出的识别方法的执行结果进行评价。在识别过程中，召回率是针对真值集合来说的，准确率是针对预标记集群集合来说的。结合比特币异常交易行为实例来进行识别方法评价是指，如果此识别方法可以在标准地址集群集合中针对不同异常交易行为识别出对应真实存在的某一实例，则证明识别方法有效可行。

定义 3(召回率)**.** 召回率=预标记集群中识别出的真值结果数/此异常交易行为真值数.

定义 4(准确率)**.** 准确率=（2×预标记集群中识别出的真值结果数－此异常交易行为真值数＋预标记集群中未被识别出的结果数）/此异常交易行

为预标记集群数.

6.2 实验结果分析

针对两种比特币异常交易行为，我们采取检验识别方法的召回率和准确率，以及利用识别方法识别出实例并分析，来验证识别方法的有效性。

6.2.1 异常交易行为召回率及准确率

（1）空投糖果行为

根据 4.2.3 节描述的方法，我们利用预标记集群集合中的空投糖果行为预标记集群构建交易图。然后在不同参数设置的情况下，使用识别方法在此交易图中识别出的部分典型结果在表 1 列出。从表中可以看出，时间参数越大、交易数参数越小、金额浮动参数越大，越容易达到空投糖果行为的高召回率。其高召回率意味着可以更全面地识别出空投糖果行为。也就是在将此识别方法应用于判断实际异常交易行为时，识别结果非常具有参考意义，其有效性是值得信赖的。

同时，可以看到表 1 中第 5 行，不仅其召回率最高（85.71%），在相同召回率下，其准确率也是最高（43.62%）的。由此可知，5.3 节获取空投糖果行为真值所用的经验参数相比于此行参数，其时间参数相对减小、交易数参数相对增大、金额浮动参数相对减小。这样可以在保证更全面地识别异常交易行为（高召回率）的情况下，更准确严格地筛选出空投糖果行为。因此由实验结果证明，5.3.1 节经验参数的选取是客观合理的。

此外，由于本识别方法的侧重点在尽可能全面地识别出异常交易行为，因此表 1 中所示的准确率普遍在 50%左右。不过考虑到识别出的异常交易行为本身所具有的特殊属性（涉事范围广、影响深），当前的准确率是可以接受的。

表 1 识别空投糖果行为召回率和准确率统计表

金额浮动参数（%）	交易数参数（个）	时间参数(天)	识别结果中的真值数（个）	召回率（%）	准确率（%）
5	30	1-20	4	57.14	51.06
5	30	1-100	5	71.43	42.55
5	35	1-20	4	57.14	52.13
10	35	1-20	5	71.43	44.68
10	35	1-50	6	85.71	43.62
10	40	1-20	5	71.43	48.94
15	35	1-50	6	85.71	38.3
15	45	1-50	3	42.86	50

最优识别结果如表 3 所示，可以看到，识别出的空投糖果行为集群中包含的真值数为 6 个，占 5.3.1 节中的空投糖果行为真值数的 6/7（即召回率为 85.71%）。

我们将识别出的这 6 个真值构建为一个交易图，如图 9 所示。共有 1887 个节点（其中有 6 个是识别出的空投糖果行为节点），1921 条边（涉及 2791 个不同的节点间发送关系），579.43126536BTC 被转移，1881 个未被标记的地址集群。

对于每个空投糖果行为节点，以该标记地址集为中心且与其相邻的所有节点被定义为一个输出集合，即此标记地址集直接指向的所有比特币地址的集合。由于此交易图是按出度设计的，所以对输出集合进行由大到小的排序，可以对应找到发起了最多比特币交易的实体（即空投糖果行为节点）。其中最大输出集合由 241,874,351 个地址组成，对应的空投糖果行为节点包含 16 个比特币地址。通过对上述结果的分析，可以得出这种识别方法较为有效的结论。

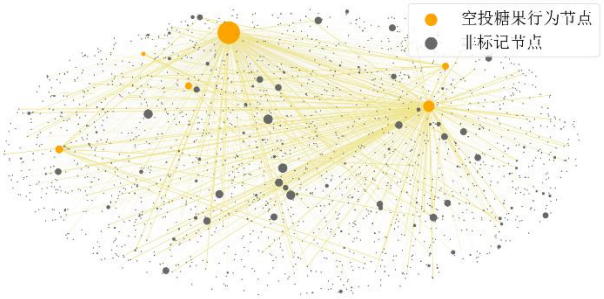

图 9 空投糖果行为交易图

（2）贪婪注资行为

同理，利用预标记集群集合中的贪婪注资行为预标记集群构建交易图。然后在不同参数设置的情况下，使用识别方法在此交易图中识别出的部分典型结果在表 2 列出。从表中可以看出，时间参数越大、交易数参数越小、金额倍数参数越小，越容易达到贪婪注资行为的高召回率，意味着可以更全面地识别出贪婪注资行为。

同时，也可以看到表 2 中第 4 行的召回率最高（81.25%），其对应的准确率为 54.32%。由此可知，5.3 节获取贪婪注资行为真值所用的经验参数相比于此行参数，其时间参数相对减小、交易数参数相对增大、金额倍数参数相对增大。这样可以在保证更全面地识别异常交易行为（高召回率）的情况下，更准确严格地筛选出贪婪注资行为。因此由实验结果证明，5.3.2 节经验参数的选取是客观合理的。并

且，由于本识别方法的侧重点在尽可能全面的识别出异常交易行为，因此表 2 中所示的准确率普遍在55%左右，是可以接受的。

表 2 识别贪婪注资行为召回率和准确率统计表

金额倍数参数（倍）	交易数参数（个）	时间参数（天）	识别结果中的真值数（个）	召回率（%）	准确率（%）
7	30	1-60	1	6.25	62.96
7	30	1-180	12	75	51.85
7	35	1-180	12	75	52.47
8	35	1-360	13	81.25	54.32
8	40	1-180	12	75	58.64
8	45	1-180	11	68.75	59.88
11	35	1-360	11	68.75	54.94
11	45	1-360	10	62.5	56.17

最优识别结果如表 3 所示，可以看到，识别出的贪婪注资行为集群中包含的真值数为 13 个，占 5.3.2 节中的贪婪注资行为真值数的 13/16（即召回率为 81.25%）。

表 3 识别结果统计表

比特币异常交易行为	真值数（个集群）	识别结果中的真值数（个）	召回率（%）	标准地址聚类识别结果（个集群）
空投糖果行为	7	6	85.71	77
贪婪注资行为	16	13	81.25	72

我们将识别出的这 13 个真值构建为一个交易图，如图 10 所示。共有 965 个节点（其中有 13 个是识别出的贪婪注资行为节点），1064 条边（涉及 1080 个不同的节点间接收关系），189.9286056BTC 被转移，952 个未被标记的地址集群。

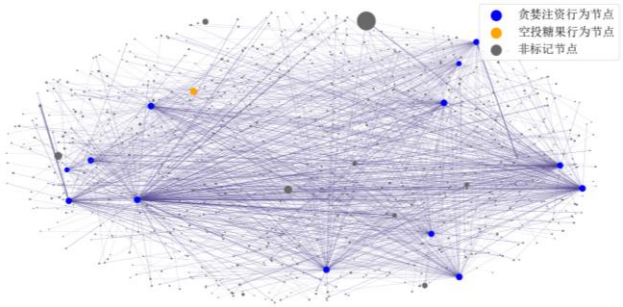

图 10 贪婪注资行为交易图

对于每个贪婪注资行为节点，以该标记地址集为中心且与其相邻的所有节点被定义为一个输入集合，即所有直接指向此标记地址集的比特币地址的集合。由于此交易图是按入度设计的，所以对输入集合进行由大到小的排序，可以对应找到收到最多比特币交易的实体（即贪婪注资行为节点）。其中最大输入集合由 106 个地址组成，对应的贪婪注资行为节点仅包含 1 个比特币地址。通过对上述结果的分析，可以得出此识别方法较为有效的结论。

6.2.2 异常交易行为识别

我们使用标准地址集群集合及其涉及的比特币交易，按 4.2.3 节描述的方法构建交易图，如图 11 所示。共有 25,853 个节点（其中有 5642 个是标准地址集群的节点），49,474 条边（涉及 72,550 个不同的节点间关系），包含标准地址集群中 5619 个未被标记的地址集群。

（1）空投糖果行为识别

我们利用识别方法，在标准地址集群集合中识别空投糖果行为。在表 3 中可以看到识别的结果，共识别出 77 个空投糖果行为集群（节点）。这些集群包含了 1,515,715 个比特币地址，涉及 14,367 笔发送交易（价值 14167.8769268BTC）。

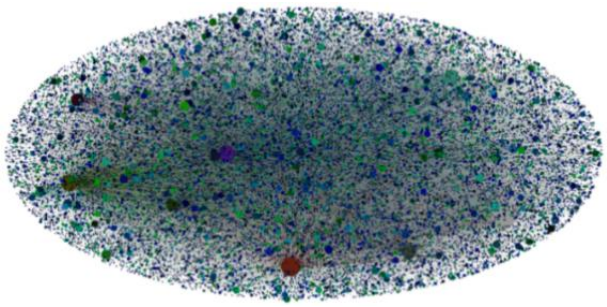

图 11 标准地址集群交易图

案例 1 分析：我们利用 WalletExplorer 在这些识别出的集群中发现了一个具有分析价值的地址集群。这个集群的部分发送交易行为与“粉尘”注入行为基本一致。同时也满足 5.3.1 节所提到的，“粉尘”交易的一笔交易的交易金额小于此交易手续费的 3 倍。

以 P2PKH 交易为例，其最小体积时为一个输入，一个输出，总共 182 字节。且比特币中默认手续费为 0.00001BTC/KB，所以此交易的手续费等于 0.00000182BTC，其 3 倍就是 0.00000546BTC。我们识别出的这个空投糖果行为集群，其虽然存在大量交易金额为 0.00000546BTC 的交易，但此金额的每笔交易所对应的手续费并不是 0.00000182BTC，而是 0.0001818BTC，远超交易金额的 1/3。所以考虑此集群存在“粉尘”注入行为，不排除有试图跟踪分析用户的动机。图 12 显示的是此集群在今年 11 月 8 日发送的“粉尘”交易数量随时间变化的累

积增长趋势。由图可知，仅单日，此集群就有 3602 个 0.00000546BTC 的输出。其在短时间内将大量相同金额发送到不包含在此集群内的其他地址的特征，满足空投糖果异常交易行为特征，可由空投糖果异常交易行为识别算法识别出来。并且可以看出“粉尘”输出数量的增长速率与时间具有相关性。不排除此集群试图在某个时间段隐藏其“粉尘”注入行为。

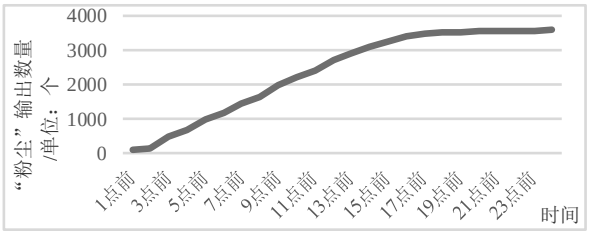

图 12 “粉尘”交易数量累积图

（2）贪婪注资行为识别

我们利用识别方法，在标准地址集群集合中识别贪婪注资行为。在表 3 中可以看到识别的结果，共识别出 72 个贪婪注资行为集群（节点）。这些集群包含了 291,476 个比特币地址，涉及 6129 笔接收交易（价值 526.90302134BTC）。

案例 2 分析：我们在这些识别出的集群地址中发现了 WannaCry 勒索事件的三个涉事地址（“ 115p7UMMngoj1pMvkpHijcRdfJNXj6LrLn ”，“12t9YDPgwueZ9NyMgw519p7AA8isjr6SMw”和“13AM4VW2dhxYgXeQepoHkHSQuy6NgaEb94”）。我们针对这三个涉事地址进行了一些统计和分析。截至 2019 年 11 月 9 日，这三个地址共进行了 413 笔交易，其中 407 笔存款交易，6 笔取款交易（每个地址 2 笔）。图 13 综合统计了这三个地址的交易数量。

由图可知，仅 2017 年 5 月一个月，此集群就有 333 笔存款交易。也就是在无历史交易的情况下，此集群在短时间内从不包含在此集群的地址接收到了大量资金的特征，满足贪婪注资异常交易行为特征，可由贪婪注资异常交易行为识别算法识别出来。2017 年 8 月的 7 笔交易中，有 6 笔是取款交易，这意味着犯罪分子在此时间段内将赃款进行了转移。经过详细的查询发现，此 6 笔取款交易发生在美国时间 2017 年 8 月 3 日的凌晨四点至五点，共转移了 51.9269094BTC。之所以选择在这个时间节点转移赃款，可能是因为 2017 年比特币价值上涨趋势明显，犯罪分子急于卖币，从而获取高资产；同时，他们可能也发现存款交易的数量随着时间的推移越来越少，从而决定收手。观察图 13，还可以发现，在事件过去半年多以后，还有人在往三个地址里存款。我们猜想，这可能是在世界的某个角落还存在此次事件的受害人，或者犯罪分子还在利用这三个地址进行着一些非法交易活动。

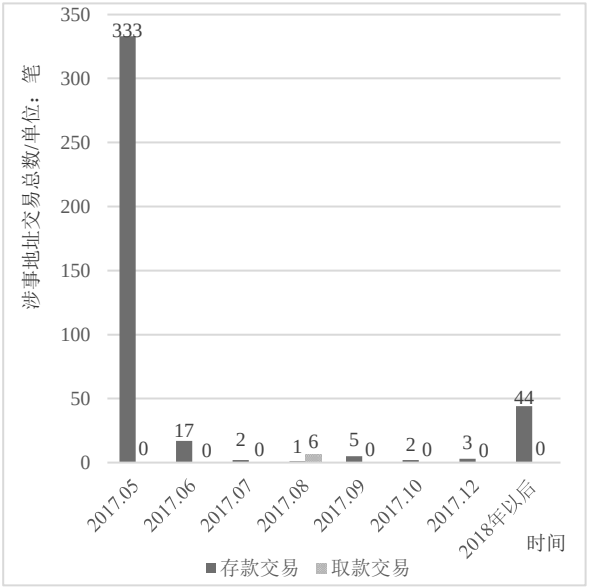

图 13 涉事地址交易数量统计图

由于此事件发生在 2017 年 5 月，所以我们又着重统计分析了 2017 年 5 月这三个地址所涉及的交易情况。三个涉事地址在 2017 年 5 月共收到了 333 笔存款交易，交易金额共 50.65432BTC，涉及 565 个输入地址和其余 3823 个输出地址。如图 14 所示，显示的是按天综合统计三个地址的交易金额和交易数量的趋势。由图可知，交易金额和交易数量特征也是满足贪婪注资异常交易行为特征的。从图中的交易数量和交易金额可以看出，在 5 月 12 日 WannaCry 出现后，勒索行为并没有立刻爆发。而是在第二天，5 月 13 日出现了第一个交易峰值（71 笔，10.202544BTC）。值得注意的是 5 月 15 日，周一，第二个交易峰值出现（89 笔，13.814941BTC）。可能是迫于工作日刚刚开始，业务的需要，所以许多人选择给犯罪分子注资，以获得相应的工作文件。随着时间的推移，很快，犯罪分子基本就接收不到太多的存款交易了。

案例 3 分析：我们在这些识别出的地址集群中又探究了一个实例，其整个诈骗过程综合了空投糖果行为与贪婪注资行为，是一个非常具有代表性的骗局案例。

2019 年 6 月 27 日，SOXex 交易平台精心地策划了一场骗局。在短短几天的时间里，先利用空投

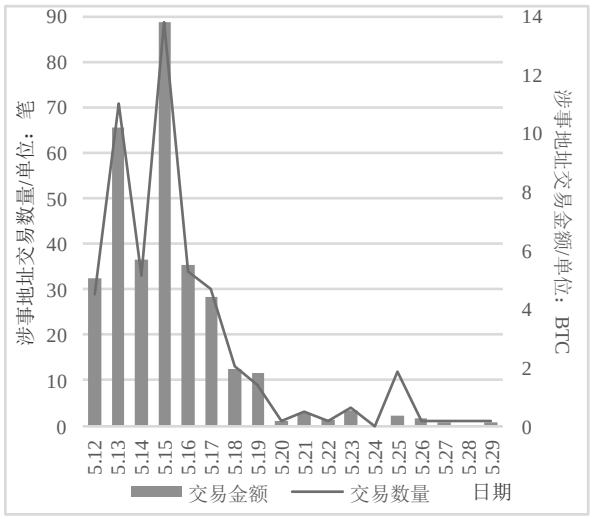

图 14 2017 年 5 月涉事地址交易数量和金额趋势图

糖果行为吸引投资者（注册交易平台即送 0.001BTC，进群即送 0.0005BTC），再利用高额的收益让投资者的贪婪心理驱使其疯狂注资认购（五折购买比特币活动，发行平台币 sox）。当交易所操作者认为骗取到了目标数额后，便即刻闭网消失。截至 7 月 5 日，交易所钱包地址中的金额已经大部分通过交易所套现，套现金额约为 4000 万人民币，诈骗金额则高达数亿之多。

我们基于公开信息找到了涉事地址“1HckjUpRGcrrRAtFaaCAUaGjsPx9oYmLaZ”。为了在标准地址集群中验证匹配到这一实例，我们进一步筛选那些包含这个实例发生时间段的标记地址集群，并跟踪记录标记地址集群的后续全部资金流量。最终，我们通过查询访问一些交易所可公共访问的 API，追踪到了上述涉事地址（“1HckjUpRGcrrRAtFaaCAUaGjsPx9oYmLaZ”），以及其他 720 个属于同一集群的涉事地址（以“19H6VMc2TvEqrddUcAdcYKq4HXeaowsKza”和“1KkF25QJRNDVYeN1DYFtoWhoirzcfjmvcr”为例）。表 4 展示的是上述三个地址在骗局发生期间的作用情况，且不排除该资金后续存在洗钱的可能。

表 4 涉事地址在骗局发生期间的作用情况说明表

涉事地址	骗局交易数（笔）	骗局接收交易金额（BTC）	特征
1HckjUpRGcrrRAtFaaCAUaGjsPx9oYmLaZ	约 9657	约 70.4614636	地址生存周期长，没有因涉及骗局而被冻结，仍在持续小额输出
19H6VMc2TvEqrddUcAdcYKq4HXeaowsKza	7	0.00185146	在 7 月 3 日和 4 日两天内多次接收并转出全部比特币
1KkF25QJRNDVYeN1DYFtoWhoirzcfjmvcr	4	0.00001638	仅在 7 月 3 日一天接收并转出全部比特币

7 结束语

本文旨在基于动机分析，探索识别比特币交易背后的异常行为。为了解决上述问题，我们收集了比特币历史交易数据，并确定了异常交易行为真值集。通过分析可能存在的交易动机，提出了两个异常交易行为的判定规则，并且通过交易图转换得到了两种异常交易行为的交易模式，从而提出比特币异常交易行为识别模型。最后，通过对异常交易行为进行识别来验证识别方法的有效性。实验结果显示，空投糖果行为召回率为 85.71%、准确率为 43.62%，贪婪注资行为召回率为 81.25%、准确率为 54.32%。并且本文重点关注分析了三个真实的异常交易行为案例。此识别方法有助于及时对基于比特币的异常交易活动进行遏制。

本文提出的基于动机分析的比特币异常交易行为识别方法是一种新的见解。但本文在基础数据范围、真值集覆盖和参照信息质量三方面依然存在局限。即扩大数据采集范围可以增加数据库现有集群所涉及的交易关系，获得更多提示；若存在公开确定且可以参考使用的真值集，则有助于正向影响本文的识别方法；大量的公开高质量参照信息，有助于识别比特币中的异常交易行为。

在未来的研究中我们将顺应市场和技术发展趋势，对区块链数据进行更深入的研究，主要包含以下工作：（1）将强模拟扩展算法[29]与子图匹配检测算法相结合，对强模拟扩展算法的边权值等属性进行增强约束。从而对检测算法进行改进处理，使其可以从海量数据中有效且高效地识别出异常交易行为，并进一步提升识别方法的召回率和准确率；（2）探索研究更多的新型异常交易行为及其特点，如非法资金跨境转移行为和市场操控行为等等；（3）基于更多种类的加密数字货币区块链进行异常交易行为的研究[30]；（4）尝试与跨领域知识进行结合，更加全面精准地揭示市场内潜在的异常交易行为。

致　谢　感谢广东省重点领域研发计划资助（2019B010137003）、国家自然科学基金资助（61972039、61872041）、北京市自然科学基金（4192050）资助，感谢向本文提出宝贵建议的审稿专家。

参 考 文 献

[1] Antonopoulos A. Mastering Bitcoin: unlocking digital cryptocurrencies. USA: O'Reilly Media, 2017.

[2] Shao Qi-Feng, Jin Che-Qing, Zhang Shao, et al. Blockchain: architecture and research progress. Journal of Computer Research and Development, 2018, 41(5): 969-988 (in Chinese)
(邵奇峰, 金澈清, 张召等. 区块链技术:架构及进展. 计算机学报, 2018, 41(5): 969-988)

[3] Meiklejohn S, Pomarole M, Jordan G, et al. A fistful of bitcoins: characterizing payments among men with no names//Proceedings of the 2013 conference on Internet measurement conference. Barcelona, Spain, 2013: 127-140.

[4] Zheng Bao-Kun, Zhu Lie-Huang, Shen Meng, et al. Identifying the vulnerabilities of bitcoin anonymous mechanism based on address clustering. SCIENCE CHINA Information Sciences, 2020, 63(3): 132101.

[5] Gao Feng, Mao Hong-Liang, Wu Zhen, et al. Lightweight transaction tracing technology for bitcoin. Chinese Journal of Computers, 2018, 41(5): 989-1004 (in Chinese)
(高峰, 毛洪亮, 吴震等. 轻量级比特币交易溯源机制. 计算机学报, 2018, 41(5): 989-1004)

[6] Fu Shuo, Xu Hai-Xia, Li Pei-Li, et al. A survey on anonymity of digital currency. Chinese Journal of Computers, 2019, 42(5): 1045-1062 (in Chinese)
(付烁, 徐海霞, 李佩丽等. 数字货币的匿名性研究. 计算机学报, 2019, 42(5): 1045-1062)

[7] Di Luzio A, Mei A, Stefa J. Consensus robustness and transaction de-anonymization in the ripple currency exchange system//Proceedings of the 37th IEEE Int Conf on Distributed Computing Systems. Piscataway, USA, 2017: 140-150.

[8] Meiklejohn S, Orlandi C. Privacy-enhancing overlays in bitcoin//Proceedings of the International Conference on Financial Cryptography and Data Security. Berlin, Germany，2015: 127–141.

[9] Möser M, Böhme R. Anonymous alone? measuring bitcoin's second-generation anonymization techniques//Proceedings of the 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). Paris, France, 2017: 32-41.

[10] Hinteregger A, Haslhofer B. An empirical analysis of monero cross-chain traceability//Proceedings of the 23rd International Conference on Financial Cryptography and Data Security. Basseterre, St. Kitts, 2019: 150-157.

[11] Yu Zuo-Xia, Au Man-Ho, Yu Jiang-Shan, et al. New empirical traceability analysis of CryptoNote-style blockchains//Proceedings of the 23rd International Conference on Financial Cryptography and Data Security. Basseterre, St. Kitts, 2019: 133-149.

[12] Kappos G, Yousaf H, Maller M, et al. An empirical analysis of anonymity in zcash//Proceedings of the 27th USENIX Security Symposium. Baltimore, USA, 2018: 463-477.

[13] Yousaf H, Kappos G, Meiklejohn S. Tracing transactions across cryptocurrency ledgers//Proceedings of the 28th USENIX Security Symposium. Santa Clara, USA, 2019: 837-850.

[14] Foley S, Karlsen J R, Putniņš, et al. Sex, drugs, and bitcoin: how much illegal activity is financed through cryptocurrencies. The Review of Financial Studies, 2019, 32(5): 1798-1853.

[15] Lee Seunghyeon, Yoon Changhoon, Kang Heedo, et al. Cybercriminal minds: an investigative study of cryptocurrency abuses in the dark web//Proceedings of the 26th Annual Network and Distributed System Security Symposium. San Diego, USA, 2019: 1-15.

[16] Moser M, Bohme R, Breuker D. An inquiry into money laundering tools in the bitcoin ecosystem//Proceedings of the 2013 APWG eCrime Researchers Summit. San Francisco, USA, 2013: 1-14.

[17] Paquet-Clouston M, Haslhofer B, Dupont B. Ransomware payments in the bitcoin ecosystem. Journal of Cybersecurity, 2019, 5(1): tyz003.

[18] Huang D.Y, McCoy D, Aliapoulios M.M, et al. Tracking ransomware end-to-end//Proceedings of the IEEE Symposium on Security and Privacy. San Francisco, USA, 2018: 618-631.

[19] Liao K, Zhao Zi-Ming, Doupé A, et al. Behind closed doors: measurement and analysis of cryptolocker ransoms in bitcoin//Proceedings of the 2016 APWG Symposium on Electronic Crime Research. Toronto, Canada, 2016: 1-13.

[20] Vasek M, Moore T. There's no free lunch, even using bitcoin: tracking the popularity and profits of virtual currency scams//Proceedings of the Financial Cryptography and Data Security. San Juan, Puerto Rico, 2015: 44-61.

[21] Bartoletti M, Pes B, Serusi S. Data mining for detecting bitcoin ponzi schemes//Proceedings of the Crypto Valley Conference on Blockchain Technology. Zug, Switzerland, 2018: 75-84.

[22] Chen Wei-Li, Wu Jun, Zheng Zi-Bin, et al. Market manipulation of bitcoin: evidence from mining the Mt. Gox transaction network//Proceedings of the 2019 IEEE Conference on Computer Communications. Paris, France, 2019: 964-972.

[23] Chen Wei-Li, Zheng Zi-Bin. Blockchain data analysis: a review of status, trends and challenges. Computer Research and Development, 2018, 55(9): 1853-1870.

[24] Di Battista G, Di Donato V, Patrignani M, et al. Bitconeview: visualization of flowa in the bitcoin transaction graph//Proceedings of the 12th IEEE Symp on Visualization for Cyber Scurity. Piscataway, USA, 2015: 1-8.

[25] McGinn D, Birch D, Akroyd D, et al. Visualizing dynamic bitcoin transaction patterns. Big Data, 2016, 4(2): 109-119.

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

[27] Feder A, Gandal N, Hamrick J, et al. The impact of ddos and other security shocks on bitcoin currency exchanges: evidence from mt. gox. Journal of Cybersecurity, 2018, 3(2): 137–144.

[28] Gandal N, Hamrick J, Moore T, et al. Price manipulation in the bitcoin ecosystem. Journal of Monetary Economics, 2018, 95(5): 86-96.

[29] Ma Shuai, Cao Yang, Fan Wen-Fei, et al. Strong simulation: Capturing topology in graph pattern matching. ACM Transactions on Database Systems, 2014, 39(1): 1-46.

[30] Gao Feng, Zhu Lie-Huang, Ding Kai, et al. Research progress on stable coins. Journal of Nanjing University of Information Science & Technology, 2019, 11(5): 499-512 (in Chinese)
(高峰, 祝烈煌, 丁凯等. 区块链稳定代币研究进展. 南京信息工程大学学报, 2019, 11(5): 499-512)

SHEN Meng, born in 1988, Ph. D. , associate professor. His research interests include network security and privacy-preserving algorithms in cloud computing.

SANG An-Qi, born in 1996, M. S. candidate. Her research interests include network and information security.

ZHU Lie-Huang, born in 1976, Ph. D. , professor. His research interests include cryptography, network and information security.

SUN Run-Geng, born in 1997, M. S. candidate. His research interests include network and information security.

ZHANG Can, born in 1996, Ph. D. candidate. His research interests include network and information security.

Background

The current blockchain digital currency is used by many malicious traders, leading to a series of abnormal trading behaviors such as "dust" injection, "airdrop" operations, extortion, and scams. Therefore, research on the identification method of abnormal transaction behavior of blockchain digital currency is of great significance for regulating transaction behavior and ensuring cyberspace security. Among the many blockchain digital currencies, the market value of Bitcoin exceeds half of the total market value of all blockchain digital currencies, and is highly representative. The Bitcoin system has a large number of users, a large transaction scale, and anonymization of addresses, which bring great challenges to the accurate identification of abnormal transaction behavior. So far, many researchers have focused on a particular type of illegal and abnormal trading behavior. But different from their method, given that there is a clear motivation behind any Bitcoin abnormal transaction behavior, this article designs a novel method for identifying Bitcoin's abnormal transaction behavior based on the analysis of transaction motivation.

Specifically, we take the two types of abnormal transaction behaviors of airdrop candy and greedy capital injection as typical representatives, and design the two types of abnormal transaction behavior determination rules, and then abstract the abnormal transaction pattern diagram. Based on this, the algorithm for identifying abnormal transaction behaviors of Bitcoin was designed and implemented using subgraph matching technology. In order to evaluate the effectiveness of this method, we collected the historical transaction data of Bitcoin for nearly 30 months, and determined the truth set of abnormal transaction behavior through manual analysis. The experimental results show that the recognition recall rate of airdrop candy behavior is 85.71%, the accuracy is 43.62%, the recognition recall rate of greedy fund injection behavior is 81.25%, the accuracy is 54.32%. In addition, we focus on the analysis and display of three typical examples of Bitcoin's abnormal transaction behavior (i.e. "dust" injection behavior, WannaCry ransomware, SOXex exchange scam), and further verify the effectiveness of the method proposed through real cases. It also shows that there are many abnormal trading activities in the cryptocurrency market.

This work is partially supported by the Key research and Development Program for Guangdong Province (No. 2019B010137003), the National Natural Science Foundation of China (No. 61972039, No. 61872041), the Beijing Natural Science Foundation (No. 4192050). The purpose of this project is to explore the abnormal transaction behavior in Bitcoin, thereby ensuring the healthy development of the cryptocurrency market.