



追踪比特币新币走向:一种基于交易网络结构特征识别矿工的方案

摘要

比特币是近年关注度最高的虚拟电子货币,“挖矿”是在此生态中最重要的获利方式之一.本文通过分析比特币交易间的引用关系,追踪区块链中挖矿所得的新币走向,构建新币流通网络.通过对比不同矿池所挖得的新币流通网络,设计了启发式算法识别比特币矿工群体,并以此推断矿池常见的收益分配方式.此外,本文还对矿工群体规模的逐年增长趋势进行了刻画.

关键词

比特币;区块链;交易网络;矿池;矿工

中图分类号 O429

文献标志码 A

收稿日期 2018-06-10

资助项目 国家自然科学基金(61304167)

作者简介

刘肖凡(通信作者),男,博士,副教授,博士生导师,主要研究方向为网络科学应用以及区块链技术.xfliu@seu.edu.cn

0 引言

2008年,中本聪发表文章《比特币:一种点对点的电子现金系统》,提出了一种基于P2P(Peer to Peer)的去中心化虚拟电子货币方案^[1].比特币系统于2009年1月上线,截至2017年市值超过1 000亿美元,交易记录超过30亿笔.比特币技术和经济实验的成功吸引了学者们的广泛关注,他们对公开的比特币交易记录展开了分析,以期更好地描述比特币生态.2014年,Kondor等发现在比特币交易网络中枢节点的增长速度比度数较低的节点更快,并概括了比特币交易网络中“富者愈富”的优先依附现象^[2].针对比特币的匿名特性,Reid等^[3]和Fleder等^[4]分别在2013年和2015年提出了几种启发式方法将比特币地址去匿名化,2016年,Sapirshtein等扩展了自私挖掘攻击的底层模型,并提供了一种算法在模型中找到攻击者^[5].2017年,Easley等研究了比特币交易费用在区块链中以采矿为基础的结构向以市场为基础的生态演变过程中所起的作用^[6].

挖矿是在比特币生态中获取利益的最直接方式.2015年,Lewenberg等^[7]提出了比特币矿池团队形成和奖励分配机制的理论模型.2017年,Beccuti等^[8]发表了比特币挖矿机制对矿工诚实性的影响.比特币系统采用POW(Proof of Work)工作量证明机制保证挖矿的安全性和公平性,并通过发行代币奖励付出算力的系统维护者(即矿工)^[1].矿工付出的算力与其获得奖励的概率成正比.然而,当比特币生态中的维护者大量增加后,仅凭单个矿工算力已很难获得奖励.因此,个人矿工纷纷加入矿池,根据自身贡献算力的比重从矿池处获得稳定的报酬.2011年首个矿池Eligius出现后,截至2012年末全球所有矿池的算力总和已达到了全网的50%,如今所有矿池的算力总和甚至已经稳定在全网算力的95%以上(<http://www.8btc.com/bitcoin-mining-pool>).目前,算力排行前五的矿池有4个位于中国,算力已经超过全网的51%(<https://btc.com/stats/pool>).比特币生态担忧,一旦若干大矿池联合对系统安全性发起攻击,将对比特币系统构成严重的威胁.因此,清晰地掌握全网矿工数量和各矿工对算力的贡献分布,对进一步认清比特币经济生态有极其重要的价值.

本文旨在提出一个从比特币交易记录中识别各个矿池旗下矿工

¹ 东南大学 计算机科学与工程学院,南京,211189

群体的方法,并总结比特币矿池的收益分配模型.首先,跟踪每个区块新币产生后的走向,构建新币流通网络;接下来,对比相同矿池和不同矿池在相邻时间所生成的区块中构建出的新币流通网络,通过分析网络的交集与差集比例、位置以及原因,在网络中定位到矿池的矿工群体,并总结出矿池挖矿利益分配模式;最后,分析了近年来矿工群体规模的变化.

1 矿工、矿池和收益分配

“挖矿”实际上是指比特币生态中的系统维护者对比特币账本进行更新的过程.通过给账本的更新数据块加入一个随机数,让数据块哈希值落入一个特定区间,使比特币账本不可篡改.比特币系统维护者寻找正确的随机数的过程就是挖矿的过程.矿池挖矿实际上是一个多用户协议,协议规定由多个客户端基于同一个数据块共同完成哈希计算.矿池最终将挖矿所得的奖励按照贡献算力的比重分发给客户端.加入矿池挖矿的客户端统称为矿工.

比特币矿池一般为矿工提供了 PPLNS、PPS、SOLO 等常见的收益分配方式.其中,PPLNS(Pay Per Last N Shares)模式,“即依据过去的 N 个股份来支付收益”,矿池如能成功挖矿,则在经过几个区块的确认后,依据矿工在过去一段时间提供的算力贡献总和给予矿工分红,若矿池没有挖到新块则矿工无分红;PPS(Pay Per Share)模式,即依据矿工的算力来支付收益,避免了 PPLNS 有时收益非常高,有时没有收益的情况,根据矿工的算力在矿池中的占比,不管矿池有没有挖到新块,都进行与矿工贡献的算力等价的分红,矿池一般在每天固定的时间发放奖励;SOLO 模式下,矿工单独进行挖矿,若挖到新块,矿池将所有奖励分配给此矿工^[8].

由于不同矿池的收益分配模式不同,在比特币交易记录中识别矿工是一个较困难的问题.本文通过复杂网络的视角,追踪一个区块中新币的流通,建立比特币资金流向网络,通过对比不同矿池得到的奖励的分配路径找到矿池的矿工群体.

2 新币流通网络的构建

图 1 为比特币账本中一串典型的交易记录的示意,每个矩形框代表一个交易,实线箭头为交易中资金转移方向,虚线箭头表示交易的引用关系.在每一笔交易中,比特币资金从左侧的输入支付地址转入右侧的输出收账地址.每一笔交易中的输入地址所

花费的比特币必须引用自前一笔交易中输出地址,且一个输出地址只能被引用一次.例如,交易 Tx_2 中的输入地址 a 和 c 分别引用自交易 Tx_1 和 Tx_n 中的输出地址 a 和 c 收到的比特币.Coinbase 交易是每个区块中的首笔交易,用于记录矿工挖矿的奖励.它实际只有输出端的收账地址,没有输入地址.本文将 Coinbase 的输入端地址记为 NoInput,输出端地址 s 为挖矿的收账地址(挖矿地址).

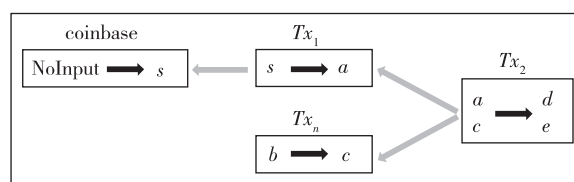


图 1 比特币交易引用关系示意

Fig. 1 Referencing of Bitcoin transaction

本文所构建的新币流通网络指的是从每个区块的 Coinbase 交易在一定时间内被直接或间接引用的所有交易的输出地址为节点所组成的网络,前一笔交易的输出地址和下一笔交易的每个输出地址间存在一条有向边.图 2 为根据图 1 的交易关系构建的新币流通网络.比特币地址 b 和 c 因为并没有直接或间接引用挖矿地址 s ,所以不在新币流通网络中出现.

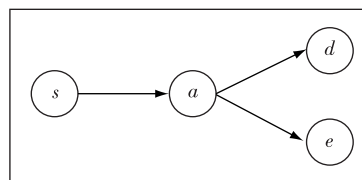


图 2 新币流通网络示意

Fig. 2 Schematic diagram of a fresh coin circulation network

由于矿池一般会在挖到新币的一段时间内向矿工进行收益分配,本文选取比特币新币在 7 d 内产生的可追溯交易构建完整的新币流通网络.图 3 为一个典型的新币流通网络.这个新币在比特币交易记录的第 277 937 个区块被挖出,4 d 内共追踪到 96 714 条交易.构建出的网络共有 168 999 个节点、693 037 条边,网络密度几乎为 0.以无向图的方式计算,节点平均度仅为 4.1,聚类系数为 0.024.同时,如图 4 所示,该网络的度分布呈多模长尾分布,说明了网络中 hub 节点的存在.

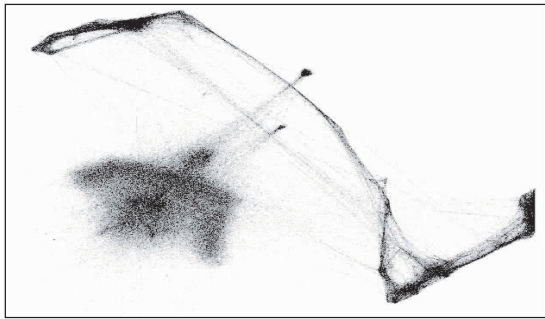


图3 一个比特币新币的流通网络
Fig. 3 A fresh coin circulation network

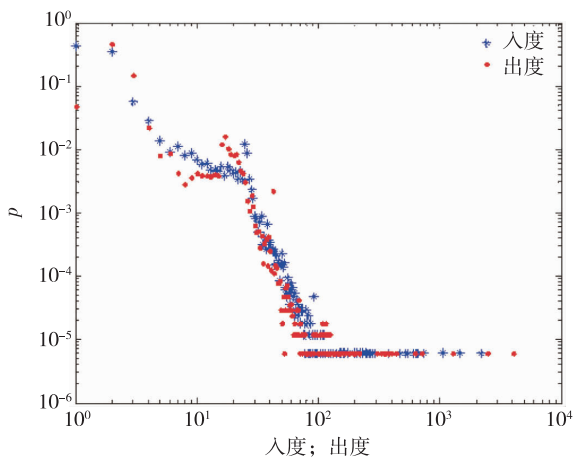


图4 网络出、入度分布
Fig. 4 In-and out-degree distributions of the network

3 矿工群体识别

矿池会在挖到新币之后通过某种收益分配模式将新币按贡献比例转入矿工的个人地址内.一般来说,矿工仅会将自己的全部算力贡献给某一个矿池,

并且只从单个矿池获得利益分配.同时,矿工提交给矿池的收账地址短时间内不会出现变化.因此,我们推测矿工群体可以通过对比不同矿池挖出的新币的流通网络实现.即,选择比特币交易记录中时间距离相邻的两个新币,若两个新币由相同矿池所挖,则新币可能流向同一批矿工群体.若两个区块来自不同矿池,则新币可能流向不同的矿工群体.

本文首先以 BTC Guild 矿池挖到的区块链上的第 277 940 和第 277 941 个区块中的挖矿地址 s_1 和 s_2 为源头构建了新币流通网络 $G_1 = \{V_1, E_1\}$ 和 $G_2 = \{V_2, E_2\}$.如表 1 所示,对比 G_1 和 G_2 可以发现两个网络规模相似、结构相近,网络节点集合 V_1 和 V_2 集合几乎一致.接下来,以 GHash. IO 矿池挖到的第 277 937 个区块和 CloudHashing 矿池挖到的第 277 938 个区块中的挖矿地址 s_3 和 s_4 为源头构建新币流通网络 $G_3 = \{V_3, E_3\}$ 和 $G_4 = \{V_4, E_4\}$.对比可以发现,两个网络大小差异较大,其中 V_4 包含了 V_3 中的大部分节点,同时也包含超过 30% 的 V_3 中没有出现的节点.

进一步分析 G_3 和 G_4 这两对新币流通网络中非重叠节点的网络结构属性,特别关注了它们在两个网络中与网络源节点 s_3 和 s_4 之间的距离.如图 5 所示,网络中所有节点与源节点距离大部分分布在 10 跳左右,非重复节点的距离分布与所有节点的距离分布略有差别.

将网络 G 中距离源节点 s 长度为 i 的节点集合 n_i 的数量记为 N_i ,将非重复节点集合 d_i 的数量记为 D_i ,在不同路径长度下,非重复节点占有所有节点数量的比例为 $r_i = D_i/N_i$.

图 6 所示分别为 G_3 和 G_4 网络中的非重复节点占有所有节点数量的比例.可以看出,若所选两个新币

表 1 相邻新币流通网络特征对比

Table 1 Comparison of two consecutive fresh coin circulation networks

新币流通网络	G_1	G_2	G_3	G_4
起始区块号	277 940	277 941	277 937	277 938
矿池	BTC Guild	BTC Guild	GHash. IO	Cloud Hashing
交易数	343 925	344 240	344 396	344 238
挖矿地址可追溯交易数	180 387	179 386	96 714	147 451
网络节点个数	302 799	301 191	168 998	250 451
重复节点个数	299 359	299 359	168 439	168 439
重复节点占比/%	98.9	99.4	99.7	67.2
网络边数	1 471 647	1 477 792	765 327	1 160 988
网络最长路径	456	456	270	456

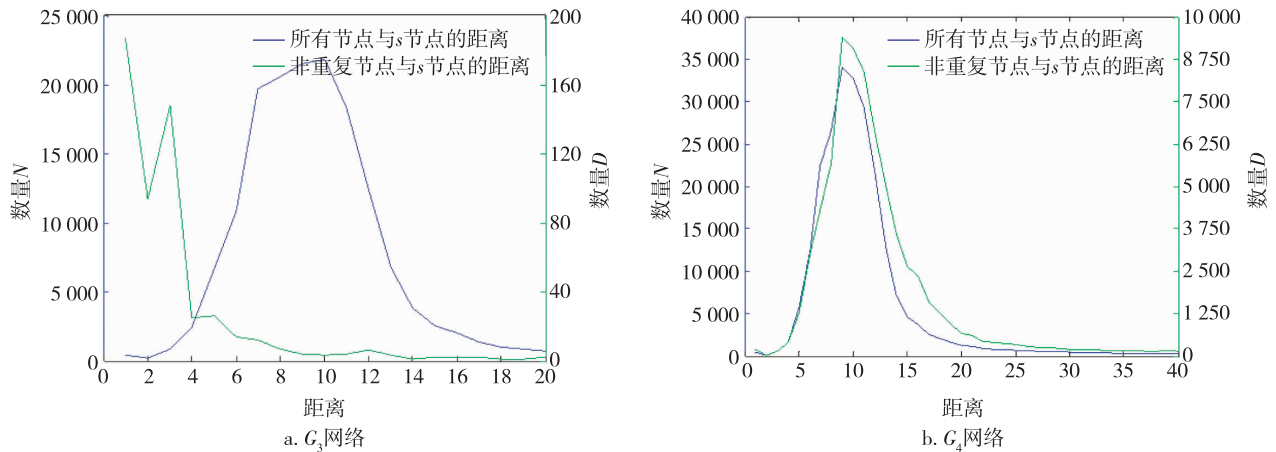


图5 网络节点与源节点s的距离分布

Fig. 5 Distance distributions from all nodes or non-overlapping nodes to source nodes of the network, (a) G_3 network, and (b) G_4 network

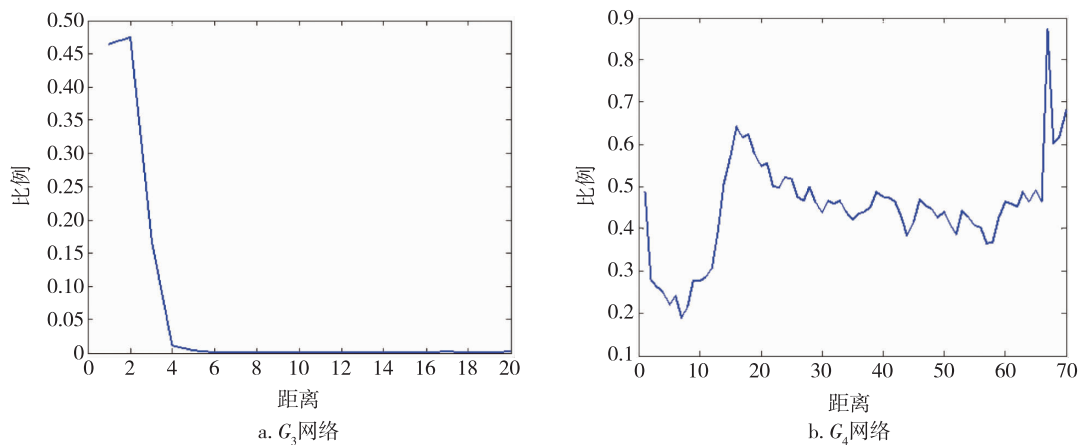


图6 在不同路径长度下,非重复节点占所有节点数量的比例

Fig. 6 Proportions of non-overlapping nodes varied with their distances to the source nodes, (a) G_3 network, and (b) G_4 network

出自相同矿池,则靠近网络源节点 s 的节点完全重复,可以判定为相同的矿工群体;若两个矿池选自不同区块,则非重复节点在靠近源节点出现概率更高,可以认为这些非重复节点为矿池的矿工.通过相邻新币流通网络识别矿工群体的示意如图7所示,其中 s 为挖矿地址, a, b, c, d 为矿工地址.

4 矿池收益分配模式推断

根据第1节所述,不同矿池挖到新块之后有不同的收益分配方法.利用第2、第3节所述的矿工识别方法,本文分析了各大矿池的矿工奖励模式,并将它们总结为以下3种.

1) 模式1:间接分配式.如图8a所示,黄色节点为NoInput地址,红色节点为矿池的挖矿地址 s ,蓝色

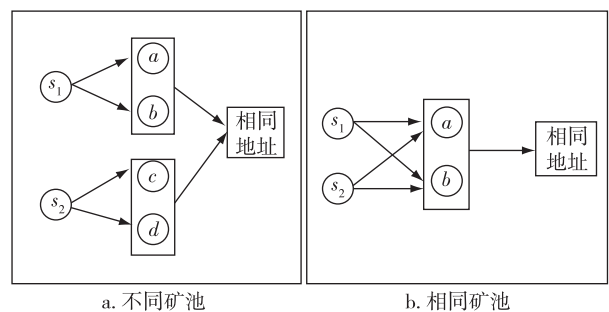


图7 相邻新币流通网络节点重复性示意

Fig. 7 Schematic diagram of overlapping nodes for consecutive fresh coin circulation network, (a) different mining pools, (b) same mining pool

节点为矿池所用的中间地址,绿色节点表示矿工群

体.当矿池挖到一个新块后,首先在新区块中通过构建 Coinbase 交易,将挖矿收益分配到自己的挖矿地址中,随后将挖矿收益转到一个中间地址,一段时间后再由中间地址将挖矿收益统一分配给矿工.使用这一模式的代表矿池为 Ozcoin.

2) 模式 2:直接分配式.如图 9a 所示,黄色节点为 NoInput 地址,红色节点为矿池的挖矿地址 s_1 ,绿色节点表示矿工群体.当矿池挖到一个新块后,首先在

新区块中通过构建 Coinbase 交易,将挖矿收益分配到自己的挖矿地址中,一段时间后将挖矿收益统一分配给矿工.使用这一模式的代表矿池为 GHash.IO.

3) 模式 3:立即分配式.如图 10a 所示,黄色节点为 NoInput 地址,绿色节点表示矿工群体.当矿池挖到一个新块后,直接在新区块中的 Coinbase 交易中将挖矿收益分配给矿工.这种模式较前两种模式更为直接,使用这一模式的代表矿池为 Eligius.

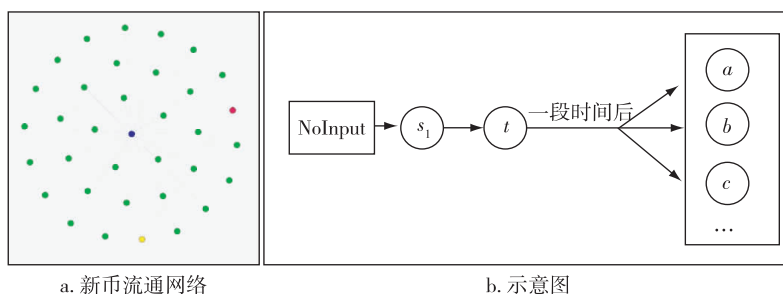


图 8 模式 1:间接分配式

Fig. 8 Mining rewards distribution pattern 1: indirect distribution, (a) coin circulation network, and (b) schematic diagram

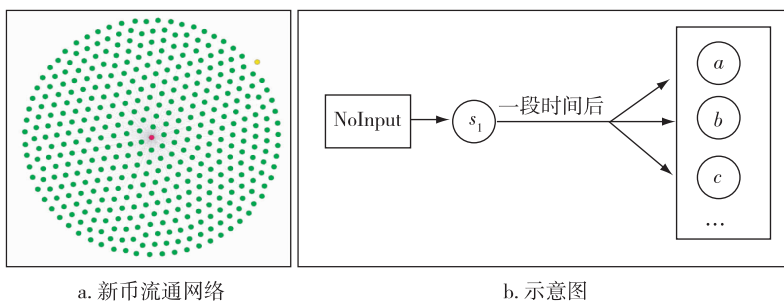


图 9 模式 2:直接分配式

Fig. 9 Mining rewards distribution pattern 2: direct distribution, (a) coin circulation network, and (b) schematic diagram

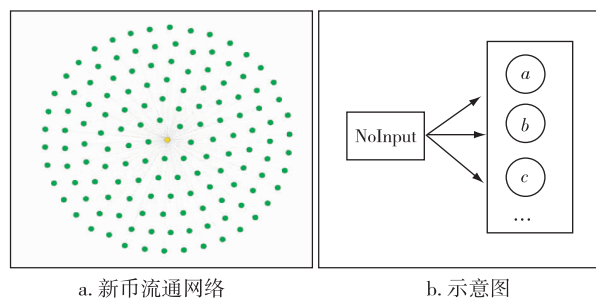


图 10 模式 3:立即分配式

Fig. 10 Mining rewards distribution pattern 3: immediate distribution, (a) coin circulation network, and (b) schematic diagram

5 历年矿工数量分析

2011 年出现了首家比特币矿池 Eligius. 截止 2011 年底,所有矿池的算力之和最高时不足全网算力的 20%.发展至今,全球所有矿池的算力总和已经达到了全网算力综合的 95% 以上.本节通过第 2 节和第 3 节提出的发现矿工的方法对 2012—2016 年挖矿最多的矿池进行了矿工群体的寻找.

图 11 所示为 2012—2016 年全球最大矿池中矿工数量的变化.从 2012 年十几个矿工,到 2016 年接近 5 000 个,矿工数量呈现逐年递增的趋势.由此可见,现在的比特币生态中通过集体挖矿获利已经变成了生态共识.这种变化离中本聪对比特币系统“一个 CPU 一票”的设计初衷已经相去甚远^[1].

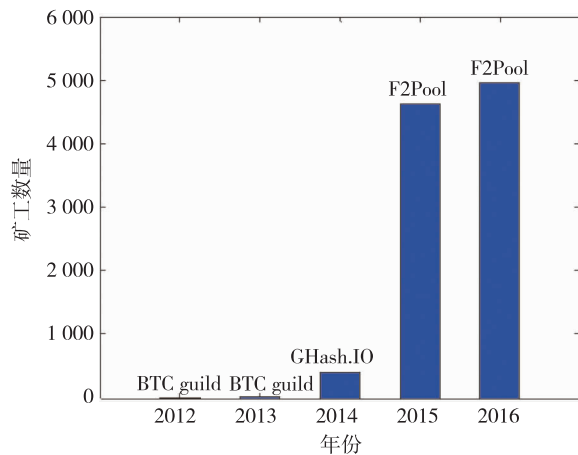


图 11 2012 年到 2016 年最活跃矿池的矿工群体规模

Fig. 11 Scale of miners in the most active mining pools from 2012 to 2016

6 讨论

本文提出了一种识别比特币矿工的可能方法.通过跟踪比特币区块中的新币在一星期内的走向,构建了新币的流通网络.对比不同矿池的时间相邻的新币流通网络之间的差异,发现不同矿池的新币流通网络在靠近网络源头的一端出现了大量非重复节点,并且节点重复率随节点距网络源头的距离慢慢变小并趋于稳定,从而认为靠近网络源头的非重复节点为矿工地址.通过该方法识别出的矿工群体与他们所在的矿池间会定期或不定期进行收益分配.我们也根据比特币交易记录总结出 3 种基本的矿池收益分配模式.最后,本文对矿工群体规模的逐年增长趋势进行了刻画,发现比特币矿工数量在

2015 年有巨大的提升,意味着比特币系统的挖矿生态已经逐渐远离其设计初衷.

识别比特币矿工对进一步推断每个矿工的算力有巨大的帮助.在如今挖矿算力逐渐集中的生态背景下,只需要掌控有限几个矿池,甚至是有限几个拥有巨大算力的矿工,就可以对系统的安全进行有实质性的威胁,比特币系统越来越受到 51% 攻击的风险.

参考文献

References

- [1] Nakamoto S. Bitcoin: a peer-to-peer electronic cash system [J]. Consulted, 2008
- [2] Kondor D, Pósfai M, Csabai I, et al. Do the rich get richer? An empirical analysis of the Bitcoin transaction network [J]. PloS one, 2014, 9(2): e86197
- [3] Reid F, Harrigan M. An analysis of anonymity in the Bitcoin system [J] // Barcelona, Spain: IMC'13, 2013, arXiv: 1107. 4524
- [4] Fleder M, Kester M S, Pillai S. Bitcoin transaction graph analysis [J]. Computer Science, 2015, arXiv: 1502. 01657
- [5] Sapirshstein A, Sompolinsky Y, Zohar A. Optimal selfish mining strategies in Bitcoin [C] // International Conference on Financial Cryptography and Data Security, 2016: 515-532
- [6] Easley D, O'Hara M, Basu S. From mining to markets: the evolution of bitcoin transaction fees [J]. SSRN Electronic Journal, 2017, DOI: 10. 2139/ssrn.3055380
- [7] Lewenberg Y, Bachrach Y, Sompolinsky Y, et al. Bitcoin mining pools: a cooperative game theoretic analysis [C] // International Conference on Autonomous Agents and Multiagent Systems, 2015: 919-927
- [8] Beccuti J, Jaag C. The Bitcoin mining game: on the optimality of honesty in proof-of-work consensus mechanism [R]. Swiss Economics, 2017, Working Papers 0060

Tracking the circulation routes of fresh coins in Bitcoin: a way to identify coinminers based on transaction network structural properties

Zeng-Xian LIN¹ Xiao Fan LIU¹

¹ School of Computer Science and Engineering, Southeast University, Nanjing 211189

Abstract Bitcoin draws the highest degree of attention among cryptocurrencies, while coin mining is one of the most important fashion of profiting in the Bitcoin ecosystem. This paper constructs fresh coin circulation networks by tracking the fresh coin transfer routes through analysis of the transaction referencing in Bitcoin blockchain. A heuristic algorithm is proposed to identify coin miners by comparing coin circulation networks from different mining pools and thereby infer their commonly used mining rewards distribution schemes. Furthermore, this paper characterizes the increasing trend of Bitcoin miner numbers during recent years.

Key words Bitcoin; blockchain; transaction network; pooled mining; coin miners