Characterizing Ethereum's Mining Power Decentralization at a Deeper Level

Liyi Zeng*§, Yang Chen†§, Shuo Chen†, Xian Zhang†, Zhongxin Guo†, Wei Xu*, Thomas Moscibroda‡
*Institute for Interdisciplinary Information Sciences, Tsinghua University
†Microsoft Research ‡Microsoft Azure
§Contacts: zengly17@mails.tsinghua.edu.cn, yachen@microsoft.com

***Abstract*—For proof-of-work blockchains such as Ethereum, the mining power decentralization is an important discussion point in the community. Previous studies mostly focus on the aggregated power of the mining pools, neglecting the pool participants who are the source of the pools' power. In this paper, we present the first large-scale study of the pool participants in Ethereum's mining pools. Pool participants are not directly observable because they communicate with their pools via private channels. However, they leave "footprints" on chain as they use Ethereum accounts to anonymously receive rewards from mining pools. For this study, we combine several data sources to identify 62,358,646 pool reward transactions sent by 47 pools to their participants over Ethereum's entire near 5-year history. Our analyses about these transactions reveal interesting insights about three aspects of pool participants: the power decentralization at the participant level, their pool-switching behavior, and why they participate in pools. Our results provide a complementary and more balanced view about Ethereum's mining power decentralization at a deeper level.**

***Index Terms*—Blockchain, Ethereum, mining power decentralization, mining pool participant.**

I. Introduction

Major blockchains like Bitcoin [1] and Ethereum [2] are based on Proof-of-Work mining [3]. The decentralization of the mining power is a major concern of the community. Previous studies show that mining pools aggregate most of the mining power [4]–[8]. For example, the combined power of the top 2∼4 pools for Ethereum and Bitcoin, which are two of the biggest public blockchains, has already exceeded the 50% threshold! However, the pool participants, who are the source of the pools' power, are largely missing in these studies.

In this paper, we conduct the first large scale characterization of pool participants. Our study focuses on Ethereum, which is one of the most popular public blockchains. The challenge is that pool participants are not directly observable because they communicate with their pools via private channels. Fortunately, they leave "footprints" on chain as they use Ethereum accounts to anonymously receive rewards from mining pools. For this study, we combine several data sources to identify pool reward transactions sent by pools to their participants. Although these transactions only carry indirect and partial information about pool participants, our analyses reveal the following important insights:

- Insight 1: Despite the power concentration at the pool level, the number of participants required to control more than 50% of the total power has grown from several hundred to several thousand. Overall, the power is more decentralized at the participant level than 4 years ago. However, we also find that this number varied significantly over time, which means it requires continuous tracking. Additionally, as our current data and methodology cannot de-anonymize the participants, it's possible that some participants split themselves into many smaller ones for various reasons, which could make our estimation inaccurate if not completely off the target. Further study to improve the estimation accuracy is important.
- Insight 2: Our study about *multi-pool* miners, who have switched pools or participated in multiple pools, shows that they control a significant amount of the total mining power. This suggests that a large share of the mining power is not owned by their pools or not loyal to them. Even if the top pools could collude to launch a 51% attack (which is a public noticeable event, as we explain in Section II), they might have to factor in the heavy price to pay if their participants switch to other honest pools.
- Insight 3: Concerned about the pools' power, multiple research groups have developed and published solutions [9]–[11], such as FruitChain [11], to disincentivize miners from joining mining pools. They attempt to reduce the pools' power concentration by removing a financial benefit, i.e., by making it possible for small-power miners to solo mine and enjoy a frequent and stable stream of rewards. Our empirical evidence about *solo-able* miners indicates that miners participate in pools even when they are powerful enough to get stable rewards by solo mining, so the benefits of pool participation are more than the reward stability. Therefore, any solution that disincentivizes pool participation by offering reward stability alone may not be effective. Researchers need to identify and address other reasons for pool participation.

We publish the data set collected for the study[1]. It contains 62,358,646 identified pool reward transactions, which are sent by 47 pools to their participants over Ethereum's entire history since 2015. In the most recent 12 months[2], the reward transactions identified on average cover more than 76% of

[1] https://github.com/yangsrc/pool-dataset
[2] From 2019-04-01 to 2020-04-01.

the total mining power. Since the data are from the public Ethereum, our results can be independently verified.

In addition to releasing a new data set, our contribution includes 1) the first longitudinal and cross-pool analysis of Ethereum's pool participants; 2) the first to reveal and document that mining power decentralization at the participant level might be significantly different from the situation at the pool level; 3) the first attempt to quantify the *multi-pool* and *solo-able* mining power, which adds to a more complete understanding of the fear and solutions about mining pools.

We review the important concepts and background in Section II. A summary of our methodology and data set is presented in Section III. We present our characterization results and findings in Section IV. Section V presents related work. Section VI concludes the paper.

II. Background

Miners and mining pools. Starting with Bitcoin [1], many popular cryptocurrencies (Ethereum [2], Litecoin [12] and others [13]–[15]) employ proof-of-work (PoW) consensus algorithms to maintain a decentralized ledger among peers in a P2P network who do not need to trust each other. These PoW algorithms incentivize ***miners***, who are peers in the network, to continuously grow a global chain of ***blocks***. Miners are required to compete for each new block by solving a unique and computation-heavy puzzle, which is called block ***mining***. In return, miners are allowed to record their account addresses in the *coinbase* field of a block to receive newly generated crypto coins as reward[3]. The block puzzles are designed in such a way that a miner, whose mining power accounts for $X\%$ of the combined total in the network, has $X\%$ chance to win the competition for each new block. For example, a miner with 1% of the total power is expected to win and receive a block reward once every 100 blocks. If 10 such miners combine their power to mine collaboratively, they together, with 10% of the total power, can expect to win once every 10 blocks. By splitting the reward for each block, each of them can receive rewards 10 times more frequently (each time one tenth of the reward). This practice can also reduce the variance of the interval between two consecutive rewards received by the miners. In the real world, many ***mining pools*** are created to coordinate such collaborative mining. The miners who join pools are called ***mining pool participants***. People often believe that small-power miners can especially benefit from pool participation [16]. Otherwise, they might have to operate without an income for a long and uncertain amount of time.

Figure 1 depicts Ethereum's mining ecosystem, in which mining pools play an important role. The operators of mining pools (denoted as O_1, O_2) create blocks and set their accounts as coinbases to receive rewards (denoted as $Miner : O_1$ and $Miner : O_2$ in the boxes representing blocks). Instead of using their own computation power to solve the block puzzles as what solo miners (denoted as S_1) do, the pool operators use the combined mining power of the mining pool participants (denoted as P_1, P_2 and P_3). As shown in the graph, one participant can join several pools. Pool operators send *pool reward transactions* (denoted as $TX : O_1 \rightarrow P_1$, $TX : O_2 \rightarrow P_3$ and etc.) to split and distribute the block rewards to participants based on their estimated mining power contribution. Pool operators often charge participants a fixed fee and have craftily designed reward allocation and payment policies [17].

[3]Miners also collect *transaction fees* if they pack transactions into their mined blocks.

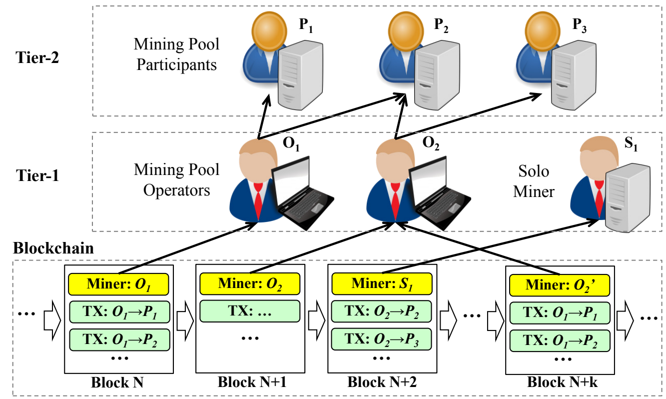

Fig. 1. Ecosystem of Ethereum mining. Miners (pools or solo miners) benefit from block rewards ("Miner:..."). Pool operators further allocate the block rewards to pool participants via on-chain transactions (TXs, "TX:...").

A 51% attack is very noticeable when it is launched. A 51% attack by top pools can be detected as they have to divert more than 51% of the total mining power off the honest chain to build a forked chain that grows faster than the honest one. After they release the malicious fork to replace the honest one, both forks of the chain are automatically recorded by the nodes in the network, which could be independently examined after the fact. Even if the victims of the attack do not cry out loudly for their loss, a significant drop of mining power on the replaced honest chain can be easily observed.

III. Methodology and Data Set

A. Methodology

We study mining pools and pool participants by observing the amount, the source and the time they receive block rewards[4] and pool rewards respectively. The amount of rewards received is used to estimate the mining power of pools and participants [19]. We are interested in the relative mining power a pool or participant has, which is calculated as a percentage of the combined total power in the network. More precisely, given a time interval (e.g., a week, month, year), we estimate the average mining power of a pool in that interval as the ratio of the mining rewards received by the pool over the

[4]Ethereum's variant of PoW algorithm allows a block to include no more than two recent stale blocks as uncles (ommers) to reward their miners and factor in their mining power contribution [18]. Our block reward calculation includes these uncle (ommer) blocks. The rest of the paper will not explicitly mention this again.

total rewards of all the mining pools and solo miners received in the interval.

$$P_{pool} = \frac{mining\ rewards\ received\ by\ this\ pool}{total\ mining\ rewards}$$

The average mining power of a pool participant is estimated similarly:

$$P_{participant} = \frac{pool\ rewards\ received\ by\ the\ participant}{total\ mining\ rewards}$$

In the rest of the paper, for brevity, we use "mining power" to refer to the "average mining power in a given time interval".

Mining pools receive rewards by setting the coinbase account of the blocks they mine as the Ethereum accounts they own. For transparency, mining pools usually disclose the accounts they use. With this information, it is straightforward to collect all the rewards received by each of the public mining pools. However, collecting the pool rewards sent out by pools to their participants is more challenging, because not all the transactions that transfer Ethers out of pools' accounts are for rewarding participants. Some of the transactions might be used for moving pools' own cut of the rewards (pool fees) or even for other unrelated purposes. To identify whether an Ether transfer transaction from a pool's account is for sending pool rewards to a participant, we need to know whether the receiver account of the transaction belongs to a pool participant. A difficulty for us is that pool participants usually stay anonymous and do not disclose their accounts as there is no obvious reason for them to do so. Fortunately, we discover a data source that can help us identify a large number of accounts used by pool participants. Among all the public mining pools, we find that 5 large pools, namely, Ethermine [20], SparkPool [21], NanoPool [22], F2Pool [23], EthPool [24], provide public APIs which could be used to verify whether a given Ethereum account is owned by their participants. Note that a participant account may be used by the participant for other unrelated purposes. We only assume that if a participant account receives an Ether transfer from an identified pool account (not necessarily the pool whose API is used to identify the account), this particular transfer is for the pool to send pool rewards to the participant. More explicitly, if an account *A* is verified as a participant account by pool *P*'s API, we assume that:

- Assumption 1: all the Ether transfers from the accounts of pool *P* to account *A* are for distributing pool rewards.
- Assumption 2: all the Ether transfers from the accounts of other pools to account *A* are also for these pools to distribute their rewards.

In Section III-C, we will provide evidence to show why we believe that the above two assumptions are mostly true, although we do not depend on them being 100% true. For example, even without Assumption 2, the core findings about the multi-pool participants in Section IV-B are still valid. The APIs we discovered do not tell us whether two or more accounts belong to the same participant. De-anonymization techniques [25], [26] might help, which we leave to be explored in future work. Our results do not depend on this knowledge, unless otherwise noted and addressed.

B. Data Set Collection

Based on the above methodology and assumptions, we use the following steps to collect the accounts used by pools and participants and the rewards received by these accounts.

1) We collect the name of the pools and the accounts they use by crawling the mining pool database of Etherscan [27], which is the most widely used blockchain explorer for Ethereum.
2) We collect the accounts of pool participants by using the following steps.
    a) From step 1, we get the accounts of the 5 large pools (see Section III-A) whose APIs can be used to verify participant accounts[5].
    b) We scan all the Ethereum transactions and collect those transactions that are sent out by the accounts of the 5 pools to transfer Ethers.
    c) We collect all the receiver accounts of the transactions we get from the previous step as candidate accounts of participants.
    d) We use the candidate accounts to query the APIs and recognize an account as a participant's account as long as it can be verified by any of the pools' APIs.
3) We collect the mining rewards received by all the identified pools and other miners (i.e., unidentified pools and solo miners).
4) We collect the pool rewards sent to all the participant accounts by scanning all the Ethereum transactions to find out those which transfer Ethers from all the identified pools to all the identified participant accounts.

We develop and assemble a cloud-based pipeline to collect, store, and process the data set. It consists of the following components: a blockchain data collector, which queries an Ethereum full node to get all the blocks and transactions and uploads them to a cloud-based data lake in JSON format; a pool collector, which crawls the Etherscan website to get all the publicly known Etheruem mining pools and the accounts they use; a participant collector, which leverages Apache Spark to scan all the Ethereum transactions stored in the data lake, aggregates the receiver accounts of all pool initiated transactions, and verifies the receiver accounts by calling the APIs of the five large pools in a rate-throttled way[6]; a data analysis and visualization platform, which includes Apache Spark, Python, Jupyter Notebook and Matplotlib. Our data set is accessible at https://github.com/yangsrc/pool-dataset, allowing reproducing the results and further study.

Our data set (first 9,847,646 blocks) covers Ethereum's entire history, starting from inception (July 30, 2015) through

[5]These APIs cannot be used to enumerate all the participant accounts of the pools. They can only be used to verify whether a given account is a participant account or not.

[6]The pools' web APIs use rate limiting to prevent abuse. It takes a long time to verify a large number of candidate participant accounts.

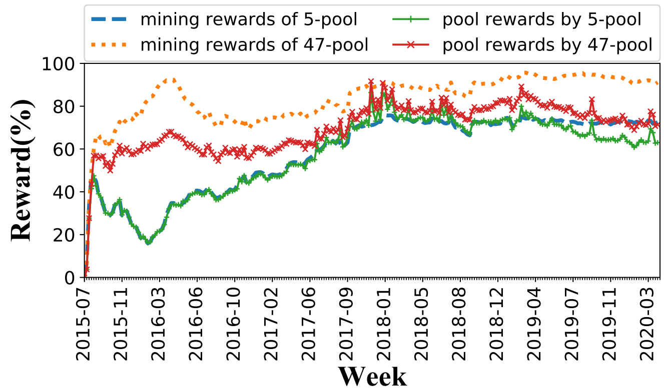

Fig. 2. Rewards received by recognized pools and participants as a percentage of the total mining rewards for each week.

April 10, 2020, which is near 5 years. It contains all the 10,852,005 block (and uncle) rewards in the target period, most of which are received by the identified pools. We find 47 public mining pools. They, in total, use 67 accounts. We identify 980,168 participant accounts with the APIs we discovered (see Section III-A). Out of the 69,285,505 Ether transfer transactions sent out by all the 47 pools, 62,358,646 are identified as pool reward transactions by our approach.

C. Data Set Characterization

In the following, we show that the data set we collect covers a significant amount of mining activities of Ethereum and provides supporting evidence for the two assumptions we make in section III-A.

From our data set, we can calculate the following statistics for all the history over 4+ years covered:

- Total mining rewards: the total mining rewards received by all the miners (mining pools and solo miners) who have ever directly produced Ethereum blocks: 38,456,433 Ethers.
- Mining rewards of 47-pool: the mining rewards received by the 47 mining pools we recognize: 31,237,271 Ethers.
- Mining rewards of 5-pool: the mining rewards received by the 5 large mining pools whose APIs we use to recognize pool participants' accounts: 20,621,001 Ethers.
- Pool rewards by 47-pool: the pool rewards received by all the recognized accounts of pool participants from all the 47 recognized mining pools: 25,854,355 Ethers.
- Pool rewards by 5-pool: the pool rewards received by all the recognized accounts of participants from the 5 large pools whose APIs we use to recognize these accounts: 20,420,738 Ethers.

Figure 2 shows a weekly breakdown of the above metrics. In this figure, we normalize the latter four as a percentage of the first one (total mining rewards), which can be interpreted as their weekly average relative mining power (see Section III-A). The line "mining rewards of 47-pool" shows that the 47 mining pools we recognize control a great majority of the total mining power in most weeks of the whole 4+ years. In the recent two years, they control more than 93% of the total mining power on average. The line "pool rewards by 47-pool" reflects the amount of mining power contributed by the pool participants whose accounts we recognize. We can see that these pool participants control more than 60% of the total power in 80% of the weeks of the whole 4+ years. In the recent two years, the percentage number is even greater, 77% on average. It clearly shows that the pools and participants, whose activities are captured by our data set, control a significant amount of total mining power. Figure 2 also provides evidence to support Assumption 1 introduced in Section III-A, which says that if account A is verified as a participant account by pool P, all the Ethers transferred from P to A are for pool reward distribution. To recall, we use the APIs of 5 large pools (see Section III-A) to verify participant accounts. Line "mining rewards of 5-pool" reflects the amount of Ether received by the 5 large pools for mining blocks. Line "pool rewards by 5-pool" reflects the amount of Ether sent out by the 5 large pools to the participants recognized by their APIs. We can observe that the two lines almost overlap with each other during most of the time. It means the total pool rewards we identified, which are sent out by the 5 large pools, closely match the total mining income of the 5 pools for most of the weeks in the past 4+ years. As a result, the following two points are highly likely true: (a) the 5 pools' APIs can recognize most, if not all, of the accounts used by their participants to receive pool rewards; (b) all the Ethers sent by these pools to their participants are for pool reward distribution. Point (b) suggests that Assumption 1 is likely true. The reasoning is that: if Assumption 1 is true, then point (b) above should be true. If, in addition, point (a) above is also true, we should observe that the total amount of pool rewards, which are recognized by our approach, is equal to the total mining rewards received by the 5 pools (minus a small fraction of pool fees). In reality, our observation is indeed very close to this, which gives us good confidence about Assumption 1 and the two points above [7].

There is also evidence that leads us to make Assumption 2 in Section III-A. This assumption is needed for us to collect the pool rewards sent out by the rest 42 pools, which do not have APIs to directly verify the accounts of their participants. We use the following indirect method to verify accounts for these pools. If an account (a) is verified by any of the 5 large pools, and (b) receives any Ether transfer from any of the 42 pools, we treat that account as a verified participant account of the sending pool. Then we collect the transfer as a pool reward sent out by the pool to the participant. Obviously this method might introduce misidentified participant accounts, so it is necessary to estimate the precision of the method, i.e., the percentage of misidentified accounts. Since we do not have the ground truth for the rest of pools, what we do is a leave-one-out cross-validation on the 5 large pools with ground truth for the estimation. Specifically, we pick one of the 5 large pools as the validation pool and pretend that the validation pool

[7]Note that we, by no means, claim that this mathematically proves Assumption 1.

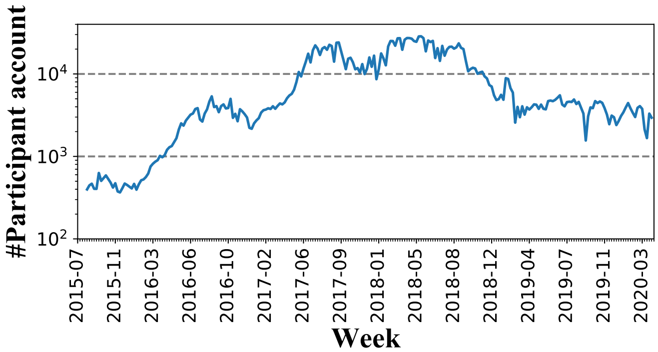

Fig. 3. Weekly minimum number of participant accounts having over 50% of the **total mining power**. The weekly mining power of each account is the sum of its mining power across all the identified pools.

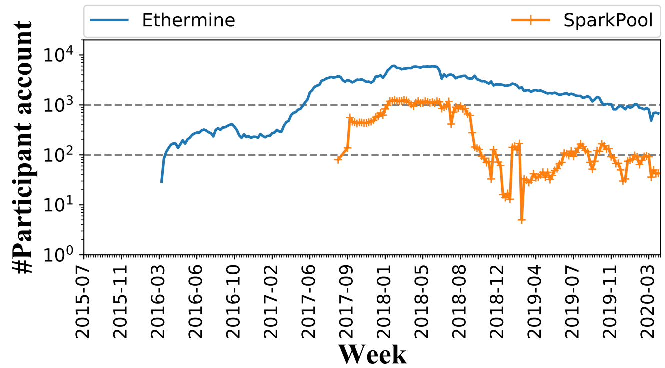

Fig. 4. Weekly minimum number of participant accounts having over 50% of the **pool mining power**.

does not have the verification API. We use the above indirect verification method, in which we only use the APIs of the other 4 pools, to find out all the indirectly verified accounts for the validation pool. After that, we count the number of these accounts that are not misidentified by using the API of the validation pool. The latter number divided by the former number is the precision. By repeating the above process five times, each time picking a different pool as validation, we have five precision numbers. The average precision of this method is as high as 97.7%, which strongly supports our confidence on Assumption 2.

IV. Detailed Studies and Insights

The data analysis results in this section provide insights about three aspects of pool participants: the power decentralization at the participant level, their pool-switching behavior, and why they participate in pools.

A. *Mining Power at the Participant Level: More Decentralized*

The mining powers of pools come from their participants. Previous work [4], [5], [7] was primarily concerned about the power concentration trend at the pool level. In this subsection, we study the trend at the participant level. This can help us evaluate the likelihood of another possible strategy of collusion: what if the top pools collude not only among themselves but also with their participants, so that they can attack without the consequence of losing these participants to other (more honest) pools? Specifically, the question is whether this strategy requires the pools to bribe only a few very large participants, or a large number of them?

Observation: We measure the mining power of pool participants based on the amount of pool rewards they receive. More precisely, we estimate the weekly power associated with each account used by participants to receive the rewards. Figure 3 shows, for each week in the 4+ years, the minimum number of participant accounts, whose combined mining power exceeds 50% of the whole network. We can see that the number of participant accounts that control the majority of the mining power went through a long and gradual growth from several hundred to more than twenty thousand, and then fluctuated around the level of several thousand. It is possible that one participant uses multiple accounts to receive rewards. As acknowledged earlier, the changes in the number of accounts may not match exactly with the changes in the number of participants. In the extreme case, the number of participants could have even decreased, while each of them uses many more accounts. Due to the limitation of our data set and methodology, we leave the de-anonymization [25], [26] of the participant accounts to future work. In the following discussions, we do not consider the extreme cases and assume that the large increase of accounts observed is positively correlated with the participant growth. The increase of participants contrasts with the history at the pool level. This increase in decentralization could be caused by a large number of newly joined participants as well as a flatter power distribution among existing participants.

Despite the large increase of participant accounts over the near 5-year period, the increase is not monotonic. For example, Figure 4 shows the power distribution in the two largest pools, SparkPool and Ethermine. Each curve shows the minimum number of participant accounts who can control the majority power of each pool. It is worth the attention that, since the 2018-08, the curve of SparkPool trends down considerably and the number falls to a few and dozens, while the Ethermine curve is relatively flat[8]. This suggests that the degree of power concentration at the participant level could change drastically over time. A data analytic work that monitors the trend can help the community to more precisely foresee the risk.

Discussion: Given this level of decentralization, we believe that there is no imminent concern about pools being able to collude with enough participants so that they could attack without risk losing their power. To a certain extent, our results can be used as evidence to confirm that the memory-intensive design of Ethereum's mining algorithm, which favors general purpose computing device (GPU) over customized single-purpose ASICs, has succeeded in making the Ethereum network more decentralized than Bitcoin, whose mining algorithm favors ASICs, in terms of the power concentration at

[8] There are still around thousands of participant accounts that receive pool rewards from Ethermine in the same period of time.

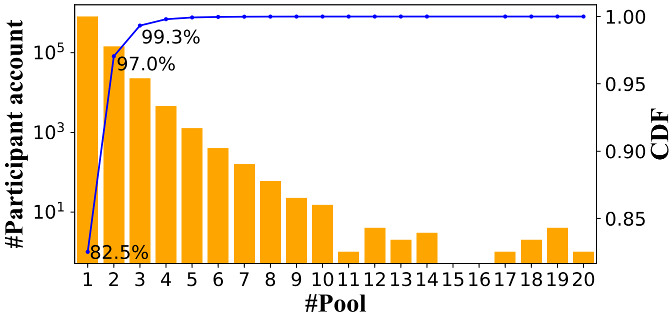

Fig. 5. The number of participant accounts who have ever received rewards from a given number of pools and the corresponding cumulative distribution.

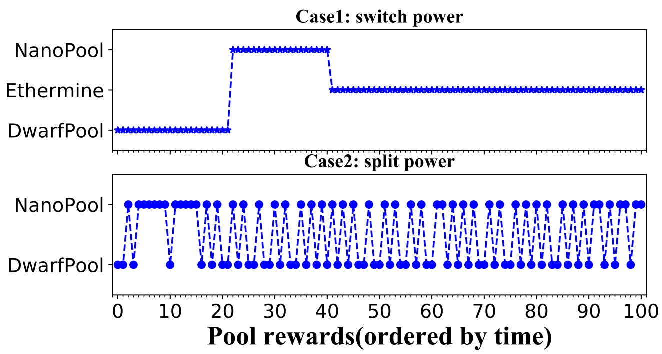

Fig. 6. Two cases to show different kinds of behaviors of the multi-pool participants. Each sub-graph represents one participant. Each dot represents one pool reward received from a specific pool. All reward payments are sorted in time order along the X-axis.

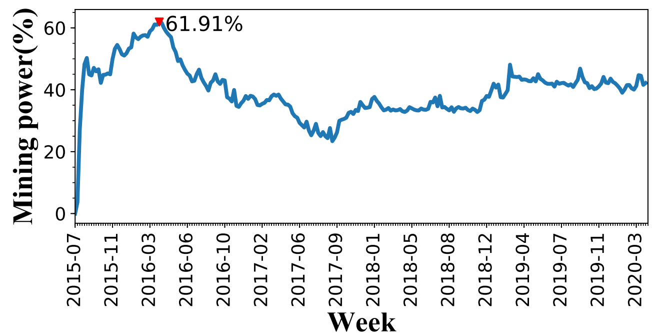

Fig. 7. Weekly mining power of the multi-pool participants

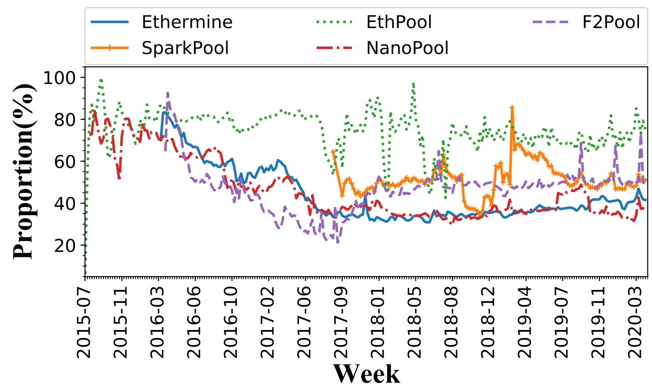

Fig. 8. Weekly mining power proportion of the 5 large pools controlled by the multi-pool participants.

the participant level [28].

B. The Power of the Multi-Pool: Significant Deterrence

The total mining power of *multi-pool* participants who do change pools is another metric to indicate how stable the relation between pools and their participants is. If such *multi-pool* participants control significant power, it suggests that dishonest top pools might pay a heavy price if they launched a 51% attack (which is a public noticeable event, as we explained in Section II), because many participants can easily leave these dishonest pools as the public's trust on them has collapsed.

Observation: We identify *multi-pool* participants by observing whether their accounts are used to receive pool rewards from more than one pools. For each of the recognized 980,168 participant accounts, we count the number of pools from which they have received pool rewards. The results are aggregated and shown in Figure 5. We can see that over 17% of the accounts receive rewards from more than one pools, ranging from 2 pools to 20 pools. Our analyses reveal a wide spectrum of participation behaviors including migration of mining power from one pool to another in one go, simultaneous participation in multiple pools and many variations in between. In Figure 6, we present two examples[9] about how participants receive pool rewards from multiple pools over time. Each dot in the figure represents a pool reward from a particular pool, whose name is shown on the Y-axis. The rewards are ordered by the time of rewarding along the X-axis. The top subgraph of Figure 6 shows a participant who migrated twice: it first received rewards only from DwarfPool, then migrated to NanoPool, and then moved again to Ethermine. The bottom one shows a participant who simultaneously participated in two pools as it received pool rewards alternatively from NanoPool and DwarfPool over a long time.

In Figure 7, for each week, we show all the mining power controlled by the *multi-pool* participants. We find 171,354 such participant accounts and they always control a significant part of the mining power, once as much as 61.91% of the total in 2016. In 2020, they control at least 39% of the network total mining power. Figure 8 shows, for each of the 5 large pools whose APIs are used to verify participants, the percentage of their mining power contributed by the multi-pool participants. We can see that they account for a significant share (at least 31% since 2019) of the total power controlled by each pool. They could significantly weaken these pools if they choose to move their power out of these pools.

In Figure 9, we further quantify what would happen if all the multi-pool participants in the top 3 pools move their mining power elsewhere. We can see that, should the multi-pool participants choose to move their power out of the top

[9]The accounts are 0xa06b0fa1384e28b87354b459a5798cdf5b6fa094 and 0x013fb52a8d412739aae37745db813478ee6f9996.

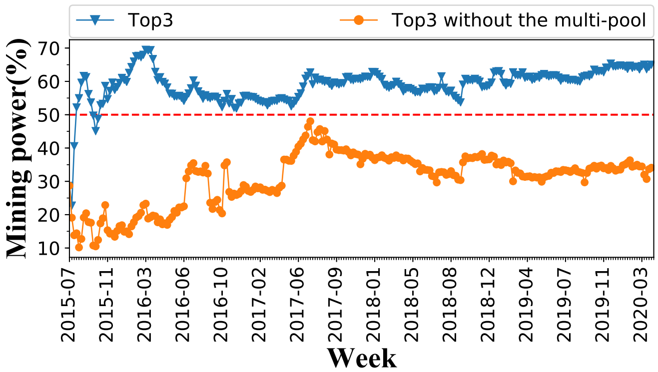

Fig. 9. Weekly mining power of the top 3 pools if multi-pool participants move their power out of the 3 pools.

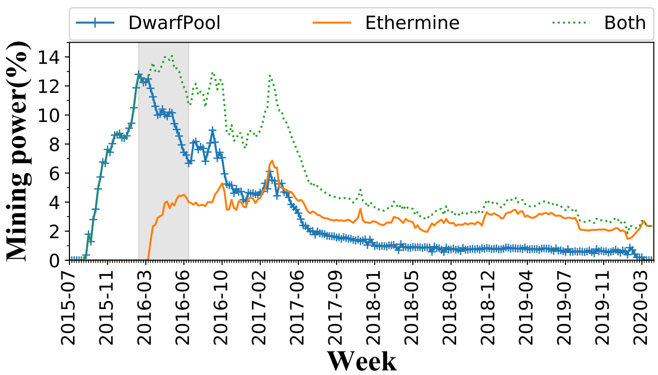

Fig. 10. Weekly mining power of these two-pool participants which migrate their power from one pool (DwarfPool) to another one (Ethermine).

3 pools, the pools' power share could drop far below 50%, making them unable to control the majority power anymore in all the past weeks. This is not just speculation. Figure 10 provides a real-world example in which the migration of pool participants causes the rise of a pool (Ethermine) and the fall of another (DwarfPool). This figure shows, for each week, the total power these participants allocate to each of the two pools (see the "DwarfPool" line and the "Ethermine" line) and the sum of the two (see the "Both" line). We can clearly observe that, from the 30th week (in February, 2016) to the 51st week (in July, 2016), the power of DwarfPool drops sharply from 12.8% to 6.7%, while the power of Ethermine rises quickly from nearly zero to 4%. Since the total power that these participants contribute to the two pools stays around 13% during this time, the simultaneous power drop and rise of the two pools are partly caused by a significant amount of participant migration from one pool to another.

Discussion: The results indicate that a significant portion of the power used by the pools is not owned by them. Thus, it is unlikely that pools can simply lure participants to join them and do whatever they want after that. Multi-pool participation is a rational behavior with several potential benefits. For example, a multi-pool participant may pay less pool fees, have better software or support, gain more rewards by pool hopping, and reduce risk by not putting all eggs into one basket. For the sake of decentralization, not having loyalty to a specific pool is a rational and healthy behavior.

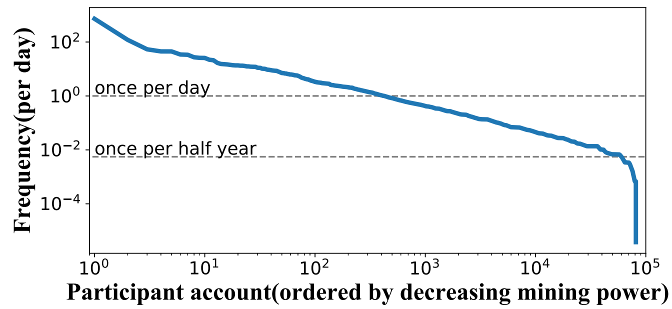

Fig. 11. Distribution of the expected reward frequency for the participants to successfully mine a block if they solo mine. The participant accounts are ordered by decreasing mining power along the X-axis. It is the snapshot of the power distribution in the latest week covered by our data set (from April 4 to April 10 in 2020).

C. Disincentivizing Pool Participation: Harder Than It Seems

Researchers often believe that a stable financial reward is the primary reason why miners join pools [10], [11], [29]. They proposed solutions to remove this incentive, hoping to disincentivize pool participation as a result. Do the real-world data support this hypothesis? We focus on **solo-able** participants. These are participants who, if solo mine, would get rewards as frequently as they do from pool participation. Hence, a stable reward from the pool seems not to be the primary incentive for these solo-able participants.

Observation: For the study of solo-able participants, we first estimate the frequency each participant can be expected to successfully mine blocks [10] if she solo mines. As explained in Section II, there is a linear relationship between the amount of mining power controlled by a participant [11] and the expected time she has to wait before she can successfully solo mine a block. More precisely, if a participant has $X\%$ of the total mining power, she is expected to successfully solo mine one block every $\frac{100}{X}$ blocks. Multiplying it by $block\_time$, which is the average time it takes to mine a block, we get the expected time for her to solo mine a block: $\frac{100}{X} \times block\_time$. For example, if a participant has 0.1% of the total mining power and the average block time is 15 seconds [12], her expected reward frequency is once per 15,000 seconds (or once per 250 minutes).

In Figure 11, we show the distribution of the reward frequency of the participant accounts in the most recent complete

[10]Solo miners receive mining rewards automatically as soon as they successfully mine a block.

[11]More accurately, it is the percentage of the total mining power a participant has.

[12]In theory, a PoW consensus algorithm makes sure that the average block time converges to a predefined parameter, e.g. 13 seconds, no matter how the total mining power of the blockchain changes over time. However, Ethereum's block time parameter has been adjusted many times and its targeting algorithm is an approximate one. In the following calculations, we use the actual measured weekly average block time instead of the parameter value directly.

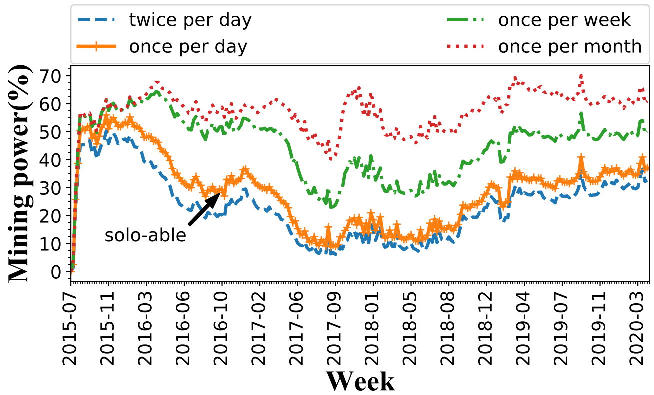

Fig. 12. Weekly mining power controlled by the solo-able participants. We assume the solo-able threshold to be **once per day**, and also show the results for the threshold set as other values.

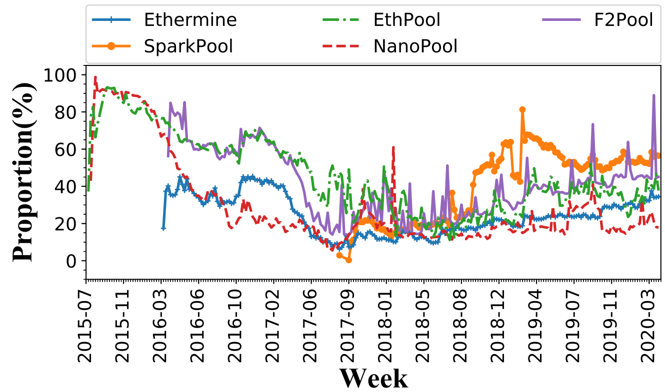

Fig. 13. Weekly percentage of the large pools' mining power controlled by the solo-able participants.

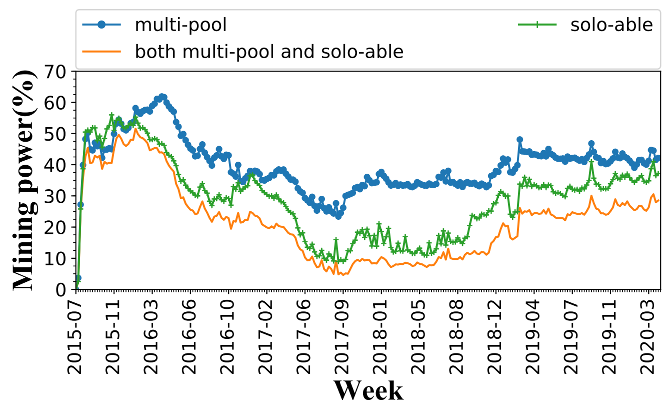

Fig. 14. Weekly mining power of the participants who are multi-pool, solo-able, both multi-pool and solo-able.

week[13] covered by our data set. Among the 81,290 participant accounts, we can see that 21,219 accounts have so little power that they are expected to get solo mining rewards no more frequently than once every half year. However, there are 423 accounts[14] whose solo mining reward frequency is at least once per day. In other words, the participants who own these accounts have enough power to solo mine at least one block per day[15]. As a comparison, the current largest pools, like SparkPool [21] and Ethermine [20], cap their pool payments to the same frequency of once per day. This means, these large pools assume that most of the participants could survive at least this frequency. Thus, we use **once per day** as one reasonable frequency threshold to determine which participants are **solo-able**[16].Compared to the current mining difficulty, the rest of the accounts seem to have too little power to solo mine. They have to join pools for stable rewards. However, the very existence of the solo-able participants suggests that the currently non-solo-able might still stay in pools even after they are capable of solo mining.

Figure 12 shows, for each week, the percentage of the total power controlled by all the solo-able participants (see the line for "once per day")[17]. For several months in the first year, they once have more than half of the total power. In 2020, they on average controlled 36.2% of the total mining power. Figure 13 further shows that they also account for a significant share of the power in the top pools. For example, in the most recent week, the solo-able account for 56.5% of SparkPool's total mining power, 34.5% of Ethermine, 45.1% of F2Pool, which are the three largest pools of that week. Additionally, some of these solo-able participants also mine in multiple pools (see Section IV-B for more findings on multi-pool participants). Figure 14 shows the total mining power controlled by such both *multi-pool and solo-able* participants in each week. On average, they account for 33.8% of the total mining power weekly in 2020. It provides evidence that a considerable number of solo-able participants would like to mine in multiple pools. The reasons that these solo-able participants would rather split their power across multiple pools than combine all their power to solo mine are especially interesting, which could be future studied.

In summary, we present strong evidence indicating that there exist participants who join and stay in pools not because they have too little mining power to solo mine. And these solo-able participants, who control a large amount of mining power, are definitely not a rare minority. Their significant existence challenges a key assumption made by previous proposals [9]–[11]: a stable financial reward is the primary reason why miners join pools. For example, FruitChain [11] introduces fruit blocks, which are much easier to mine than normal blocks. They assume that most of the participants would not join pools as long as their mining power is enough for them to solo mine these fruit blocks. This assumption is not supported by our evidence.

[13]The week is from April 4 to April 10 in 2020.

[14]As will be shown in the next paragraph, these accounts control a significant amount of the total mining power in the network, although they only account for a small fraction of all the participant accounts.

[15]If some of these participants use several accounts to receive rewards, they can solo mine blocks more frequently than what we estimate.

[16]At the time of this writing, to be solo-able within one day, a participant needs to control at least 0.0154% of the total power.

[17]This figure also shows similar results for lower (once per week, once per month) and higher (twice per day) solo-able thresholds.

Discussion: Our results suggest that we should identify other incentives when designing protocols to effectively disincentivize pool participation. Admittedly, some possible incentives might be very difficult to remove because pools do offer good value to participants. For example, pools take care of participants' operational overhead (e.g., running a well-connected full node, handling software upgrade), so that participants' job can be simplified into providing computing power. This is a strong incentive for participants, which may be hard to remove in practice.

V. Related Work

In this section, we summarize the related work from the following aspects:

Ethereum. The Ethereum blockchain has been widely studied. Many studies [29]–[32] focus on the security of Etherum's smart contracts. Some [33], [34] reveal various attacks at the p2p level. Other works [35], [36] propose layer-2 protocols to improve Ethereum's scalability and confidentiality. In addition, researchers also study how to build decentralized applications (Dapps) on Ethereum [37], [38]. This paper studies the mining power decentralization problem from the view of Ethereum's pool participants and the effectiveness of the existing mitigating proposals.

Mining pools/participants. Kroll *et. al.* theoretically analyzed the economy of Bitcoin mining and whether the Bitcoin protocol can survive attacks from Goldfinger-type adversaries [39]. Eyal *et. al.* proposed selfish mining attack [40] and block withholding attack [41] that can be employed by mining pools to gain an unfair advantage. Kwon *et. al.* proposed fork after withholding (FAW) attack, which is more efficient than selfish mining [42]. Gervais *et. al.* proposed a framework which can analyze the security and performance of a POW-based blockchain regarding different parameters [3]. For pool participants, Wang *et. al.* studied the computation power distribution among participants of F2pool in Bitcoin [43]. Lewenberg *et. al.* have studied the advantage a participant may achieve with pool switching strategies [44]. Belotti *et. al.* studied the pool-hopping phenomenon in Bitcoin [45]. Our work conducts a large-scale empirical study on the mining power evolution and mining behaviors of pool participants in Ethereum.

Centralization. Several works have studied mining power centralization in PoW blockchains [4]–[7]. The work done by Gencer *et.al.* [4] shows that the top three Ethereum mining pools have a combined 61% mining power of the network, indicating that Ethereum is quite "centralized" at the pool level. In addition, various attacks related to centralization such as collusion attacks [46], [47], are discussed. However, none of them provides the large-scale and cross-pool study of Ethereum's pool participants, which turns out to be another perspective for evaluating decentralization. They also have not gone beyond the simplistic standpoint of technical feasibility to study social and economic factors inherent in blockchain systems. **To mitigate the centralization due to pool mining,** Miller *et. al.* proposed nonoutsourceable mining puzzles to disincentivize pool formation and a multitier reward scheme to pay small-power miners more frequently [10]. Luu *et. al.* proposed SmartPool, a decentralized pool that improves security and reduces income variance for participants [9]. Similarly, Fruitchain is introduced to make the mining reward steady for single miners, to remove the incentive of participating in mining pools [11]. Our work suggests that participants join and stay in pools for more reasons than getting a more stable income. Proposals only focusing on the latter issue might not work as well as expected. **On the mining hardware side,** researchers find ways, such as memory-hard puzzles, to reduce the performance advantage of ASICs over CPUs and GPUs, which lowers the entry barrier and investment risk to attract more miners. For example, cryptocurrencies such as Ethereum, Litecoin, Monero, Zcash, and Grin [48] have adopted Ethash [18], Scrypt [49], CryptoNight [50], Equihash [51] and Cuckoo cycle [52], respectively. Our findings show that this does increase decentralization for Ethereum at the participant level.

Empirical study of blockchain. Since blockchain data are public, many studies have looked into the data and made many observations. Meiklejohn *et. al.* analyze the bitcoin ledger to deanonymize addresses [53]. Similarly, Zcash [25], Monero [54], and cross-ledger transactions [55] are analyzed. Chen *et. al.* generally study Ethereum ledger data and propose efficient algorithms to detect DDoS-like attacks [56]. Compared to [56], our study focuses on the evolving mining powers of Ethereum with much more details. Different from the previous work which analyzes the ledger data, Kim *et. al.* collect data from the p2p layer of Ethereum and conduct comprehensive analytics [57]. Gencer *et. al.* also study the p2p network of both bitcoin and ethereum [4].

VI. Conclusions

In this paper, we present a large-scale and longitudinal analysis of Ethereum's pool participants. To our best knowledge, it is the first empirical characterization of the mining power decentralization of Ethereum beneath the surface of pools. We show that a new perspective of the mining power could deepen and challenge the current understanding of the decentralization problem. We admit that this paper raises more questions than it answers. Our results invite similar comparative studies of the pool participants of other major blockchains (e.g., Bitcoin), as well as future work that uncovers new data sets or uses new angles to analyze pool participants.

Acknowledgment

We would like to show great appreciation to anonymous reviewers for their insightful comments. This work is supported in part by National Natural Science Foundation of China (NSFC) Grant 71872094, the Zhongguancun Haihua Institute for Frontier Information Technology and Nanjing Turing AI Institute.

References

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[2] "Ethereum homepage," https://www.ethereum.org/, 2020.
[3] "Proof of work," https://www.investopedia.com/terms/p/proof-work.asp, 2020.
[4] A. E. Gencer, S. Basu, I. Eyal, R. van Renesse, and E. G. Sirer, "Decentralization in bitcoin and ethereum networks," *arXiv preprint arXiv:1801.03998*, 2018.
[5] A. Gervais, G. O. Karame, V. Capkun, and S. Capkun, "Is bitcoin a decentralized currency?" *IEEE security & privacy*, vol. 12, no. 3, pp. 54–60, 2014.
[6] A. Beikverdi and J. Song, "Trend of centralization in bitcoin's distributed network," in *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE, 2015, pp. 1–6.
[7] L. J. Valdivia, C. Del-Valle-Soto, J. Rodriguez, and M. Alcaraz, "Decentralization: The failed promise of cryptocurrencies," *IT Professional*, vol. 21, no. 2, pp. 33–40, 2019.
[8] M. Romiti, A. Judmayer, A. Zamyatin, and B. Haslhofer, "A deep dive into bitcoin mining pools: An empirical analysis of mining shares," *arXiv preprint arXiv:1905.05999*, 2019.
[9] L. Luu, Y. Velner, J. Teutsch, and P. Saxena, "Smartpool: Practical decentralized pooled mining," in *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Association, pp. 1409–1426.
[10] A. Miller, A. Kosba, J. Katz, and E. Shi, "Nonoutsourceable scratch-off puzzles to discourage bitcoin mining coalitions," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2015, pp. 680–691.
[11] R. Pass and E. Shi, "Fruitchains: A fair blockchain," in *Proceedings of the ACM Symposium on Principles of Distributed Computing*. ACM, 2017, pp. 315–324.
[12] "Litecoin homepage," https://litecoin.org/, 2020.
[13] "Zcash homepage," https://z.cash/, 2020.
[14] "Dash homepage," https://www.dash.org/, 2020.
[15] "Monero homepage," https://monero.org/, 2020.
[16] "Benefits of a mining pool," https://www.investopedia.com/terms/m/mining-pool.asp, 2019.
[17] "Ethereum mining pool comparation," https://www.poolwatch.io/coin/ethereum, 2020.
[18] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, pp. 1–32, 2014.
[19] "What is mining in ethereum?" https://github.com/ethereum/wiki/wiki/Mining#so-what-is-mining-anyway, 2018.
[20] "Ethermine homepage," https://ethermine.org/, 2020.
[21] "Sparkpool homepage," https://www.sparkpool.com/, 2020.
[22] "Nanopool homepage," https://nanopool.org/, 2020.
[23] "F2pool homepage," https://www.f2pool.com/, 2020.
[24] "Ethpool homepage," https://ethpool.org/, 2020.
[25] G. Kappos, H. Yousaf, M. Maller, and S. Meiklejohn, "An empirical analysis of anonymity in zcash," in *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, 2018.
[26] P. Evans-Greenwood, "Distributed ledgers & linked data," in *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017, pp. 1451–1451.
[27] "Etherscan," https://etherscan.io/, 2020.
[28] G. Hileman and M. Rauchs, "Global cryptocurrency benchmarking study," *Cambridge Centre for Alternative Finance*, vol. 33, 2017.
[29] L. Luu, D.-H. Chu, H. Olickel, P. Saxena, and A. Hobor, "Making smart contracts smarter," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 254–269.
[30] Y. Zhou, D. Kumar, S. Bakshi, J. Mason, A. Miller, and M. Bailey, "Erays: Reverse engineering ethereum's opaque smart contracts," in *USENIX Security*, 2018.
[31] S. Kalra, S. Goel, M. Dhawan, and S. Sharma, "Zeus: Analyzing safety of smart contracts," in *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*, 2018.
[32] J. Krupp and C. Rossow, "teether: Gnawing at ethereum to automatically exploit smart contracts," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 1317–1333.
[33] K. Wüst and A. Gervais, "Ethereum eclipse attacks," ETH Zurich, Tech. Rep., 2016.
[34] Y. Marcus, E. Heilman, and S. Goldberg, "Low-resource eclipse attacks on ethereum's peer-to-peer network." *IACR Cryptology ePrint Archive*, vol. 2018, no. 236, 2018.
[35] J. Poon and V. Buterin, "Plasma: Scalable autonomous smart contracts," *White paper*, pp. 1–47, 2017.
[36] R. Khalil and A. Gervais, "Revive: Rebalancing off-blockchain payment networks," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 439–453.
[37] H. Kalodner, S. Goldfeder, X. Chen, S. M. Weinberg, and E. W. Felten, "Arbitrum: Scalable, private smart contracts," in *Proceedings of the 27th USENIX Conference on Security Symposium*. USENIX Association, 2018, pp. 1353–1370.
[38] L. Breidenbach, I. Cornell Tech, P. Daian, F. Tramer, and A. Juels, "Enter the hydra: Towards principled bug bounties and exploit-resistant smart contracts," in *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, 2018.
[39] J. A. Kroll, I. C. Davey, and E. W. Felten, "The economics of bitcoin mining, or bitcoin in the presence of adversaries," in *Proceedings of WEIS*, vol. 2013, 2013, p. 11.
[40] I. Eyal and E. G. Sirer, "Majority is not enough: Bitcoin mining is vulnerable," *Communications of the ACM*, vol. 61, no. 7, pp. 95–102, 2018.
[41] I. Eyal, "The miner's dilemma," in *Security and Privacy (SP), 2015 IEEE Symposium on*. IEEE, 2015, pp. 89–103.
[42] Y. Kwon, D. Kim, Y. Son, E. Vasserman, and Y. Kim, "Be selfish and avoid dilemmas: Fork after withholding (faw) attacks on bitcoin," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 195–209.
[43] L. Wang and Y. Liu, "Exploring miner evolution in bitcoin network," in *International Conference on Passive and Active Network Measurement*. Springer, 2015, pp. 290–302.
[44] Y. Lewenberg, Y. Bachrach, Y. Sompolinsky, A. Zohar, and J. S. Rosenschein, "Bitcoin mining pools: A cooperative game theoretic analysis," in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2015, pp. 919–927.
[45] M. Belotti, S. Kirati, and S. Secci, "Bitcoin pool-hopping detection," in *2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI)*. IEEE, 2018, pp. 1–6.
[46] K. Liao and J. Katz, "Incentivizing double-spend collusion in bitcoin," in *Financial Cryptography Bitcoin Workshop*, 2017.
[47] C. Decker, J. Seidel, and R. Wattenhofer, "Bitcoin meets strong consistency," in *Proceedings of the 17th International Conference on Distributed Computing and Networking*. ACM, 2016, p. 13.
[48] "Grin homepage," https://hq.grin.ninja/, 2020.
[49] J. Alwen, B. Chen, K. Pietrzak, L. Reyzin, and S. Tessaro, "Scrypt is maximally memory-hard," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2017, pp. 33–62.
[50] M. J. Seigen and T. Nieminen, "Neocortex, and am juarez,"cryptonight hash function,"," 2013.
[51] A. Biryukov and D. Khovratovich, "Equihash: asymmetric proof-of-work based on the generalized birthday problem," *Proceedings of NDSS 2016*, p. 13, 2016.
[52] J. Tromp, "Cuckoo cycle: a memory bound graph-theoretic proof-of-work," in *International Conference on Financial Cryptography and Data Security*. Springer, 2015, pp. 49–62.
[53] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: characterizing payments among men with no names," in *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 2013, pp. 127–140.
[54] A. Miller, M. Möser, K. Lee, and A. Narayanan, "An empirical analysis of linkability in the monero blockchain," *arXiv preprint*, vol. 1704, 2017.
[55] H. Yousaf, G. Kappos, and S. Meiklejohn, "Tracing transactions across cryptocurrency ledgers," *arXiv preprint arXiv:1810.12786*, 2018.
[56] T. Chen, Y. Zhu, Z. Li, J. Chen, X. Li, X. Luo, X. Lin, and X. Zhange, "Understanding ethereum via graph analysis," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1484–1492.
[57] S. K. Kim, Z. Ma, S. Murali, J. Mason, A. Miller, and M. Bailey, "Measuring ethereum network peers," in *Proceedings of the Internet Measurement Conference 2018*. ACM, 2018, pp. 91–104.