Private Settlement in Blockchain Systems

Alfred Lehar*
Haskayne School of Business
University of Calgary

Motahhareh Moravvej-Hamedani†
Haskayne School of Business
University of Calgary

May 2024

*E-mail: alehar@ucalgary.ca

†Corresponding Author email: motahhareh.moravvejh@ucalgary.ca. We thank Katya Malinova, Stefan Scharnowski, Junli Zhao and Charles M. Kahn and participants of the FutFinInfo 2022 and Economics of Financial Technology Conference 2022, Cardiff Fintech Conference 2022, Sydney 17th Central Bank Conference on the Microstructure of Financial Markets 2022, 2022 Paris December Finance Meeting, 2022 Bonn Workshop on Digital Currencies, University of California Santa Barbara DeFi Seminar 2023, 1st Bank of Canada conference on Networks in Modern Financial and Payments Systems 2023. Lehar is grateful to the Canadian Securities Institute Research Foundation, CFI/JELF, and SSHRC for financial support. We are grateful to Eric Gascoine and Dylan Rae for their research assistance.

Private Settlement in Blockchain Systems

Abstract

We provide evidence that the settlement market in blockchain systems is not transactional and diverges from a simple competitive auction model. Examining the Bitcoin blockchain, 5.88% of transactions, which we label as *private*, bypass the competitive auction and are channelled directly to miners. Despite being more active than the average user, their transactions are confirmed by one miner, which is a statistically unlikely outcome in a competitive auction. Our findings suggest that high-demand users engaging in long-term agreements with miners paying on average, 20% lower fees. We document how settlement contracts are structured in an unregulated market.

Keywords: Blockchain Technology, Bitcoin, Settlement, Distributed Ledger, Cryptocurrency, Private Settlement, Contract

1 Introduction

The Economist magazine recently reported the Ethereum blockchain alone settled transactions worth 2.5 trillion dollars in the second quarter of 2021, roughly the same amount as the Visa network.[1] The premise that decentralized blockchain systems improve efficiency relies on the assumption of a competitive market for settlement of transactions. Nevertheless, very little is known about the settlement agreements that endogenously emerge in an unregulated market with free entry. It is wildly believed that blockchain settlement fees are determined by an auction, in which users attach fees to their transactions, and competitive miners include the transactions with the highest bids in a block and thus settle them. This market is understood to be transactional in nature.

This paper provides evidence consistent with some users and miners bypassing the competitive auction and forging private, exclusive, long-term settlement agreements with miners. We study Bitcoin, which is the oldest and most valuable cryptocurrency. At the time of writing this paper, Bitcoin's market cap is larger than all other cryptocurrencies combined.[2]

Under Bitcoin's original design, users post transactions to a public Mempool from which miners can pick transactions to include in blocks. We document that almost six percent of settled transactions have never appeared in the Mempool and, thus, have not followed the standard settlement process. We argue that miners receive them from private channels outside of the Bitcoin system. One potential reason for the existence of these channels is risk sharing. On the one hand, miners are exposed to variations in average fees and transaction demand while facing fixed operations costs. Small miners rarely find a new block and could face a sparse Mempool or low fees when they win the block, resulting in low revenue. Securing a long term order flow through a private channel allows the miner to collect a fixed minimum amount of fee revenue

[1] https://www.economist.com/leaders/2021/09/18/the-beguiling-promise-of-decentralised-finance
[2] https://www.visualcapitalist.com/Bitcoin-market-cap-compared-to-crypto/

irrespective of the size of the Mempool or the current average fee rate.[3] On the other hand users with a high demand for regular transactions like centralized exchanges or layer 2 protocols face uncertainty on the fees they need to pay to get their transactions mined, creating a risk-sharing opportunity between small miners and regular users using long-term contracts at fixed prices.

Many users need to post transactions frequently. Large centralized exchanges, like Binance or Coinbase, for example, have so many clients who want to deposit or withdraw funds that they have a very predictable transaction demand.[4] Layer two (L2) protocols that are built on top of Bitcoin often run a less secure blockchain but periodically post the state of their chain onto the Bitcoin blockchain to benefit from Bitcoin's high security. These L2s have highly predictable, regular transaction demand. For example, *Stackchain* posts its daily status on the Bitcoin blockchain approximately every 100 Bitcoin blocks. [5]

Using a sample of over a hundred million transactions, we document that private transactions pay lower average fees and exhibit lower fee variation. Using a standard algorithm, we attribute the individual transactions in our sample to 59 million users. Users of private transactions are more likely to insert data into the blockchain as it is done by various L2 protocols with regular transaction demand. We document that users of private transactions post regular and evenly-spaced transactions on the Bitcoin blockchain. On average, users of private transactions in our sample post 566.22 transactions compared to 34.68 transactions for regular users. Despite being more active, private users use only 1.02 distinct miners to process their transactions, while normal users' transactions are, on average, mined by 9.73 distinct miners.

We show that smaller miners are more active in the market for private transactions and

[3]This risk is orthogonal to the uncertainty on how many blocks a miner can find in a given time which is analyzed in Cong, He, and Li (2021).

[4]Clients deposit their Bitcoin typically in a one-time wallet that is created specifically for that deposit. Once the funds are cleared and it is verified that they do not come from a sanctioned address or an address that has been involved with known illicit activity, the exchange then moves the funds to their own wallet in a secondary transaction.

[5]https://gaia.blockstack.org/hub/1AxyPunHHAHiEffXWESKfbvmBpGQv138Fp/stacks.pdf, see an example transaction '7b540ef6983fe904456c7e6a96d051595b2470379a7cc14e3aa0e6de503a7b58'

document that private transactions smooth miners' per-block and monthly incomes. We rule out several alternative explanations. In a series of tests, we rule out longer waiting times for private transactions as the source of lower fees. We also show that private transactions are not of lower priority as miners forgo in some instances the inclusion of higher-paying public transactions to fulfill their commitment to their private users. Finally, we document that private users occasionally switch their designated miner which usually results in lower fee payments for the user, indicating that there is a market for private settlement contracts.

Anecdotal evidence is consistent with the existence of private transaction channels. For example, Bitcoin SV developed an API freely available on GitHub that allows users to interact directly with miners. On their website, they advertise that their product allows for "Direct transaction submission: This allows users to bypass the outer layers of the Bitcoin peer-to-peer network and submit transactions directly to miners."[6] Their product also allows for "User-based fee policies to enable different fee structures for different use cases." Other forum posts also discuss how to pass transactions to specific mining pools by forwarding them directly to a node run by the pool. The pool can then decide whether or not to share these transactions with other nodes.[7]

Our research underscores the complexities of settlement markets with users with heterogeneous transaction demand and settlers with different market shares. In such a setting a simple auction mechanism as it was envisioned by Bitcoin's founder and as it is used in most blockchains today may not span the space of optimal contracts that form endogenously between users and settlers. Both users and miners find gains from trade by bypassing the protocol's default settlement mechanism. Since blockchains are not regulated the set of possible settlement contracts can not be restricted to the simple auction and an 'outside market' has developed that allows users to enter more complex contracts. This observed price discrimination highlights the

[6] https://bitcoinassociation.net/bitcoin-sv-miner-id-and-merchant-api-beta-release/

[7] https://bitcoin.stackexchange.com/questions/5337/how-do-i-send-a-transaction-directly-to-a-miner-or-pool-for-processing

emergence of different tiers of service within the Bitcoin settlement market, where users with higher transaction volumes or strategic relationships with miners pay lower fees and face less uncertainty. In a free market for settlement, the endogenously emerging fee structure is far more complex than the fee structure we see in regulated traditional financial markets.

Dark pools and electronic communication networks (ECNs) offer a mechanism to bypass the public trading venues in traditional financial markets. Menkveld, Yueshen, and Zhu (2017) describes the "pecking order" hypothesis for selecting trading venues. Investors sort dark and lit venues when executing orders by associated costs, like bid-ask spread, price impact, information leakage, and immediacy. They show that dark venues are at the top of the pecking order. Like private transactions in Bitcoin, dark pool transactions reduce trading costs (Hu, Jones, and Zhang 2021). In the United States, this type of transaction executed approximately 13.66% of the equity volume in November 2021.[8] However, the primary reason for utilizing dark pools differs from our setting. Dark pools lessen the negative price impact of large orders. In Bitcoin, there is no price impact of transactions, nor is there price discovery as the blockchain is used to settle previously contracted Bitcoin transfers. Brogaard, Carrion, Moyaert, Riordan, Shkilko, and Sokolov (2018) demonstrate the impact of dark pool transactions on price discovery that leads to inefficient information acquisition and lack of transparency. They also study the mechanism of ECNs and highlight several attractions for both buyers and sellers in ECNs trading. Without a broker-dealer system, both sides can eliminate bid-ask spread limits and communicate privately through the direct link, with lower costs. ECNs are attractive for higher trade speed and anonymity in this infrastructure. While previous research documented deviations from competitive behaviour in transaction settlement on blockchain markets, this paper is, to the best of our knowledge, the first one to document that some users bypass the competitive settlement market. We find evidence consistent with price discrimination where some users who settle many transactions and frequently use the system get a different price than infrequent users. This finding casts doubt on whether all users of decentralized systems can equally benefit

[8] https://www.rblt.com/market-reports/let-there-be-light-us-edition-31

from any cost savings this new technology will bring.

Our paper is part of the nascent literature on settlement in unregulated markets. In the original Bitcoin, paper (Nakamoto 2008) assumes as the first step to run the network that *new transactions are broadcast to all nodes*. Competitive mining is the standard assumption in many works on the Bitcoin network. Easley, O'Hara, and Basu (2019) model the emergence of fees in blockchain systems. Lehar and Parlour (2022b) examines the potential for miners to implicitly collude by strategically managing the effective capacity of the blockchain. They argue that miners' prices discriminate among users to increase fee revenue. Huberman, Leshno, and Moallemi (2021) argue that keen users post higher fees to get their transaction included in the next block when there is uncertainty on the block's arrival time. Both papers assume that all transactions pass through the Mempool and assume a one-off transactional relationship between miners and users. Our paper explores a novel mechanism where transactions bypass the Mempool altogether, and users and miners can sign long-term contracts.

Our paper is related to a literature review concerning regulation in settlement systems. Currently, settlement is highly regulated and often performed by the government near entities that offer their services at a flat price for every participant. Blockchain systems open a private settlement market where users offer miners fees to confirm their transactions (Auer (2019) and Halaburda, Haeringer, Gans, and Gandal (2022)). Schmiedel, Malkamäki, and Tarkka (2006) model the traditional settlement markets considering tight regulation in financial markets. Russo, Hart, Malaguti, and Papathanassiou (2004) examine conflicts of interest typically arising in the securities settlement infrastructure and propose improvements in regulation. We augment this literature by examining a competitive and unregulated settlement market.

The Bitcoin blockchain is an ideal setting to study private settlement markets in the absence of transaction ordering issues that exist in other blockchain systems. In Bitcoin, the ordering of transactions within a block is irrelevant. This is not the case in Ethereum, where users interact with smart contracts based on the order in which the transactions are recorded in the block,

which creates opportunities for front-running (Strehle and Ante 2020). Daian, Goldfeder, Kell, Li, Zhao, Bentov, Breidenbach, and Juels (2019) document several ways in which users can benefit from strategically placing their transactions ahead of the ones from other users. Park (2021) discusses front-running for automated market makers in such systems. Daian, Goldfeder, Kell, Li, Zhao, Bentov, Breidenbach, and Juels (2020) demonstrate how miners in Ethereum can strategically alter the ordering of transactions or execute other users' transactions in their name for financial gain; they pocket the so-called Miner Extractable Value (MEV). Lehar and Parlour (2022a) document that fees do not sort 80% of blocks on Ethereum and that miners collect over a million USD per day from prioritizing some transactions. These opportunities for miners to extract rents are absent in Bitcoin as it offers no smart contract ability, and the ordering of transactions within a block is irrelevant.

We find that private users pay lower than average fees for Bitcoin private settlement, while in Ethereum, users pay higher transaction fees to front-run the existing transaction and claim the profits of the arbitrage transaction. Moreover, Weintraub, Torres, Nita-Rotaru, and State (2022) shows that powerful miners are interested in profiting from such opportunities in Ethereum. In contrast, our findings suggest that miners with smaller market shares have incentives to participate in a private settlement.

2 The Bitcoin System and Private Transactions

Transactions transfer Bitcoin between wallets. A transaction involves at least one sender and one receiver.[9] The former digitally signs a transaction to verify that she is the rightful owner of the Bitcoin and propagates her request through the decentralized peer-to-peer network of

[9]Each Bitcoin transaction can have many inputs and many outputs. The inputs are amounts of Bitcoin individually locked up by a locking script. The sender can unlock all these inputs and transfer funds to output addresses, which do not necessarily belong to the same owner. The process is similar to making a $23 payment with a $20 and a $5 bill (2 inputs) and giving $23 to the merchant and keeping $2 in change (2 outputs).

Bitcoin nodes. These pending transactions constitute the Mempool (or Memory pool). Theoretically, each node could see a different set of pending transactions. Recent innovations such as the fibre network or BIP152 allow rapid transmission of blocks, which ultimately benefits the overall transaction processing and confirmation time in the Bitcoin network. Latency is no longer a practical problem. Empirical results from computer science, for example, Dae-Yong, Meryam, and Hongtaek (2020) show that Mempools are practically identical.[10]

Miners pick pending transactions from the Mempool to be included in a block. Users can attach fees to their pending transactions that the miners can keep when they include the transaction in a block. The Bitcoin protocol has been designed with the idea that all pending transactions go through the Mempool and are ordered based on the aging algorithm in which old and high-fee inputs have higher priority over newer transactions that offer smaller fees. The miners then solve a computationally complex puzzle to create a valid block which can be added to the chain.

In practice, however, miners have complete discretion on which transactions to include in a block. Specifically, they can include transactions that did not go through the Mempool. Users can contract privately with individual miners and forward transactions directly to them (outside the peer-to-peer network). Miners can then include these transactions in a block they will mine. We label such transactions that bypass the Mempool as *private* transactions. Private transactions are not visible to other miners; thus, their fees accrue exclusively to the miner who confirms that transaction.

We argue that miners and users with high transaction demand are incentivized to contract for a fixed fee for risk-sharing. Miners who face medium-term fixed costs for electricity and infrastructure are concerned about variations in future fee revenue. Users with predictable transaction needs are similarly worried about changes in fees. This risk-sharing motive differs from the risk regarding the number of blocks mined in a given time interval.[11]

[10]Moreover, many of our results are independent of latency; for example, latency cannot explain why private users always match with the same miner or why they pay lower fees.

[11]Xue, Xu, Wu, Lu, and Xu (2021), and Cong, He, and Li (2021) discuss and analyze the incentives for rational

To illustrate our point, assume that mining fee revenue varies over time and can either be high or low with equal probability. Variations in fees may be induced by changes in demand for transactions by users, variations in the utility that users place on getting their transactions included in the blockchain, or by the random arrival time of blocks. A miner could solve the puzzle immediately after a previous block was found, in which case there might be fewer transactions waiting for the Mempool, and hence, the miner's fee revenue might be lower. Specifically, assume that a miner can obtain fees worth f or 0 upon finding a block.

Miners are risk-neutral and mine a fixed number of ν blocks in a given period. To introduce a motive for risk sharing, assume that miners are in financial distress and face a loss of l whenever they only mine blocks where the fee revenue is low within a period. One can think of a situation where the miner cannot pay for the electricity costs with the incoming fee revenue and has to obtain costly external financing. The probability of financial distress is then

$$p = \frac{1}{2^{\nu}} \tag{1}$$

The expected profit of a miner of size ν who takes transactions with an expected fee of $f/2$ per block from the Mempool and is exposed to fee risk is, therefore:

$$\pi^{m} = \nu\frac{f}{2} - pl = \nu\frac{f}{2} - 2^{-\nu}l \tag{2}$$

In contrast, a miner who would sign an agreement with a user to mine a certain number of transactions for a fixed fee revenue c can obtain a stable flow of fees and thus avoid financial distress. She obtains

$$\pi^{c} = \nu c \tag{3}$$

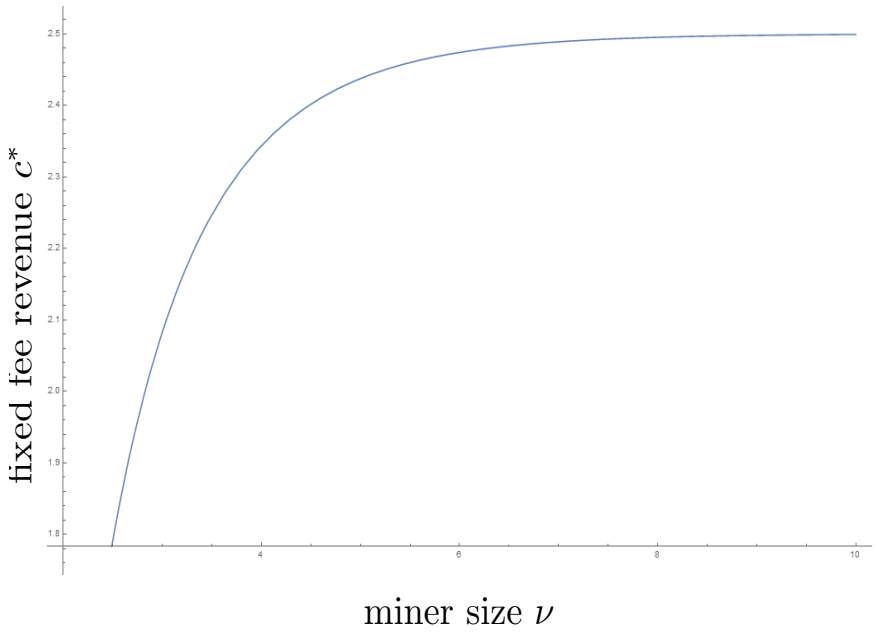
miners to share risk and form pools. A pool will reduce the uncertainty around the number of mined blocks. Participants share the cost of finding a new block and the revenue proportional to their contribution.

Solving for the fixed fee revenue c^* that makes the miner indifferent, we obtain

$$c^* = \frac{2^{-\nu-1}(2^\nu f\nu - 2l)}{\nu} = \frac{f}{2} - \frac{l}{\nu 2^\nu} \tag{4}$$

which is increasing in ν. Figure 1 illustrates the fee revenue as a function of miner size ν for an example.

Figure 1. Fixed fee revenue c^* at which a miner would be indifferent between fixed and variable revenues as a function of miner size ν. The parameters are $f = 5, l = 10$.



From this simple toy model, we can derive several empirical predictions. Miners will offer lower fees in exchange for a fixed or less volatile guaranteed fee flow. We, therefore, expect private transactions to have lower and less variable fees. As illustrated in Figure 1, smaller miners face higher uncertainty over fee revenue and are thus willing to offer a lower fee for private transactions. We, therefore, expect that smaller miners will be more engaged in processing private transactions and offer them at a lower fee.

It only makes sense for a miner to enter a risk-sharing agreement with a user who has a steady demand for transactions. Retail users posting occasional transactions cannot provide a steady flow of fees for the miner to make a long-term contract useful. We, therefore, expect users who engage in private transactions to post more transactions than an average user and post the transactions more regularly than other users. An example of users with a high transaction demand is Layer 2 protocols or side chains synchronizing to the Bitcoin blockchain, such as the Omni-layer, Rootstock, Stacks, or the Liquid Network.[12] These systems use data insertion transactions identified by using the data insertion operation code "OP_RETURN" in the Bitcoin lock script. We, therefore, predict that users who utilize data insertion transactions are more likely to post private transactions. When miners and users enter such an agreement to mine private transactions, we expect users of private transactions to have all their transactions mined by one miner.

3 Data

To trace transactions from their origination to their confirmation, we combine Mempool and blockchain data. For the former, we set up Bitcoin nodes on two separate instances in the Compute Canada cloud and take snapshots from the Mempool every minute. These snapshots include all transactions that wait to be confirmed at that given time. Their unique hash value identifies transactions. We collect blockchain data on transactions mined into blocks directly from a modified Bitcoin node. We know when and in which block it was mined for each transaction. We identify miners based on a unique text most mining pools put in the input script portion of the coinbase transaction, the first transaction of each block that grants the miner

[12]Such systems keep information in a separate ledger, often an independent blockchain, and regularly post a hash of their state to the Bitcoin blockchain. Users of the side-chain or meta-layer can, therefore, verify the system's state by verifying the state with the posted state on the Bitcoin blockchain. These systems take advantage of Bitcoin's security and assure consensus, at least at the points in time when the state is posted to the Bitcoin blockchain. One example of such a meta layer is the Omni-layer, which allows the creation of tokens and was, for a long time, the main home of the USD stablecoin Tether.

newly minted Bitcoin as a reward for finding a block. We describe our dataset in three parts: transactions, miners, and users.

3.1 Transactions

Our sample consists of 97,637,471 confirmed Bitcoin transactions for 2021. We collect senders, receivers, the amounts sent, and the fee paid to the miner for each transaction. The definition of the variables and measures in this study are reported in Table 15 in Appendix A. We classify a transaction as normal if it is observed in the Mempool before being mined into a block. For those transactions, we record the timestamp when the transaction entered the Mempool and when it was mined. Transactions that never enter the Mempool but still get recorded on the blockchain are identified as private transactions. We collect one-year of Mempool data from the 1st of January 2021 to the 1st of January 2022. We tag transactions based on their state as either *pending* or *confirmed*, where pending transactions are still waiting in the Mempool to be added to a block by a miner. We observe 93,374,343 Mempool transactions in our sample period, most of which are eventually confirmed. When a miner confirms a transaction, it will be added to a block and removed from the Mempool. Table 1 reports the summary of statistics for transactions in our sample.

We define the waiting time as the difference between the block number in which the transaction was included and the highest block number when the transaction entered Mempool. We do not measure the delay based on timestamps, so random fluctuations in times between blocks do not drive our delay measure. On average, transactions wait for 6.66 blocks in the Mempool. Some transactions get confirmed immediately; the longest time a transaction waited in the Mempool was for 23 blocks. In our sample, the Mempool is never empty; therefore, we observe positive delays. Our findings are consistent with Pappalardo, Di Matteo, Caldarelli, and Aste (2018), who document an average delay of 6 blocks. The waiting time for private

transactions is not observable as they do not reside in the Mempool; instead, they are transmitted to miners through private channels. To gauge the waiting time for private transactions, we assume that private transactions arrive randomly at the user's designated miner and get included in the next block that she composes. As a proxy for waiting time, we therefore measure half the distance between blocks that are mined by the designated miner. We observe an average waiting time for private transactions of 5.51 blocks, roughly the same order of magnitude as for normal transactions.

Items	*Confirmed Transactions*
Number of Days	365
Start Date	1/1/2021
End Date	31/12/2021
Transactions	97,637,471
Mixers' Transactions	107,639
Miners' Transactions	152,556
Private Transactions	5,738,085
Number of Blocks	52,686
First Block	663,913
Last Block	716,598
Maximum Fee (Sats)	1.83e+08
Average Fee (Sats)	21,940.77
Fee Variation (Sats)	2.18e+10

Table 1. Transactions Summary Statistics This table reports the summary of statistics for the 2021 sample. The *Confirmed Transactions* are recorded in the chain by miners. Fees are in Satoshi. One bitcoin is 100,000,000 Satoshis.
The number of total confirmed transactions before removing mixers and miners is 97,796,453.

3.2 Miners

For each block, we record the mining pool, block height, weight, and creation time. We report a summary of statistics on mined blocks in Table 1. We identify 28 miners for 91.16% of the

blocks in our sample based on the input section of the coinbase transaction.[13] We label miners in the highest quantile of mined blocks as big and the remaining miners as small. Table 2 shows the market share classification of the mining pools in our sample.

Mining Pools	*Block Mined*	*Market Share*	*Category*
F2Pool	6,695	15.65	*Big*
AntPool	6,504	15.20	*Big*
poolin	5,342	12.49	*Big*
ViaBTC	4,799	11.22	*Big*
binance	4,644	10.86	*Big*
BTC.com	3,965	9.27	*Small*
foundry_usa	2,428	5.67	*Small*
Huobi	2,010	4.70	*Small*
SlushPool	1,786	4.17	*Small*
1THash	1,110	2.59	*Small*
WAYI.CN	689	1.61	*Small*
SBICrypto	530	1.23	*Small*
(15 others)	405	5.28	*Small*

Table 2. Mining Pools' Market Share in 2021 The market share per mining pool is calculated based on the number of blocks they mined in 2021 divided by the total blocks mined by all miners in the same period. Miners in the top quantile are classified as *Big*.

Mining is a concentrated industry. The five big miners have 65.42% of the market. F2Pool, Antpool, Poolin, ViaBTC and Binance are our sample's top five mining pools. We rank mining pools based on their mined blocks divided by the total blocks mined in a year.

3.3 Users

To study user behaviour, we use a standard algorithm from the computer science literature, the Union Find Algorithm which is explained in detail in Appendix B. We map 300,147,341 input wallet addresses to 59,086,692 distinct users.

Some transactions on the Bitcoin blockchain are initiated by miners, for example, when they

[13] In some cases, we are unable to attribute the block to certain miners. We report the small or undetectable mining pools in our one-month sample as 'Other.'

pay mining rewards to their pool members. These transactions can be different; for example, miners will likely mine their own transactions as they do not want to pay fees to other miners. We define users as miners if they collect fees or block rewards in one of their wallet addresses. We exclude miner transactions from our empirical analysis. We also exclude transactions from mixers, who on purpose try to obfuscate users' identities by purposefully combining transactions from multiple users. Details can be found in Appendix B.

Users who engage in private submissions send 95.78% of their total transaction demand through the private channel. We classify users engaged in at least one private transaction with a miner as *private* users, while all other users are denoted as *normal*. Our classification algorithm identifies 3,408,352 private users and 55,678,340 normal users in our sample. Consequently, 5.76% of the users in our dataset are categorized as private. Moreover, 88.92% of private users exclusively conduct transactions through private channels.

Private transactions are not equally distributed across pools. Two pools, KanoPool and Ckpool, stand out, with 10.56% and 4.12% of their confirmed transactions being private, respectively. These pools are also the two smallest pools in our sample. Table 3 presents summary statics per user group. Consistent with our predictions, on average private users pay lower fees and have lower fee variation than normal users. Average fees are 10.42 and 12.82 for private and normal users, respectively. Therefore, private users pay lower fees compared to normal ones by 18.72%. The average private user is more active, with 566.22 transactions compared to 34.68 transactions for normal users. Yet, on average, their transactions are mined by only 1.02 miners. If miners and users are competitive, transactions are randomly allocated to multiple miners based on their mining capacity. The probability that a user gets randomly allocated to the same miner for n transactions is p^n, where p is the market share of a miner. Even for the largest pool with a market share of 15.28%, the probability that a user gets allocated to the same miner for five transactions is 0.0083%. We also see that users of private transactions insert more data into the blockchain, which is consistent with the idea that they are frequent users, such as

layer two protocols.

We define the active period for each user as the time difference between the user's first and last transactions in our sample. To measure fees, we use the fee per weight unit. For miners, the opportunity cost of a transaction is determined by its physical size in bytes, as block size is limited. With the introduction of SegWit (segregated witness), parts of the scripts can be outsourced to the witness section, which does not count toward the physical block limit. The SegWit update thus introduced weight units for the measurement of transaction size as weight considers how effectively a transaction utilizes the witness section. Private users post larger transactions with 45.07% more weight units on average. Normal users and private users are waiting for almost the same amount of time, with 101.60 blocks for normal users and 108.71 blocks for private users, respectively.

4 Empirical Findings

4.1 Average Fee

We expect users who contract long-term with one miner to pay lower fees as the miner replaces a risky cash flow with a safe cash flow and is willing to charge a lower fee. We consider the fee paid per weight unit as our measure of the transaction fee as limited block space causes transaction weight to be the opportunity cost for the miner. To make the result more robust toward outliers, for the econometric analysis, we winsorize the fee per weight and the total input value of the transaction at the 99 percentile.

$$\begin{aligned} \textit{FeeperWeight} = \alpha_0 &+ \alpha_1 \textit{PrivateDummy} + \alpha_2 \textit{SumInputs} + \alpha_3 \textit{BlockSize} \\ &+ \alpha_4 \textit{TxCount} + \alpha_5 \textit{TxWeight} + \alpha_6 \textit{WaitTime} + \varepsilon \end{aligned} \tag{5}$$

We regress fee per weight on a dummy for private transactions and control variables as

Items (Mean)	*Normal*	*Private*
Fee per Weight	12.82	10.42
Number of Transactions	34.68	566.22
Number of Wallet Addresses	92.09	61,136.59
Active Period (Months)	3.47	9.22
Transaction Weight	1,867.07	1,905.98
Sum Inputs (Billion)	1.01	1.14
Miner Count	9.73	1.02
Data Insertion (%)	1.38	5.75
Waiting Time (Blocks)	101.60	108.71

Table 3. User Group Types Summary Statistics The summary of statistics for each criterion is calculated on average for each user group. *Fee per Weight* is the measure of comparing fees in each user group type. *Number of Transactions* is the number of transactions submitted by each user group type. *Number of Wallet Addresses* is the distinct number of wallets that each user group uses. The *Active Period* is the number of blocks between the first and the last observation for transaction submission per user and is reported by month. *Transaction Weight* is the variable indicating the size of each submitted transaction by users. The *Sum Inputs* represents the sum of input values measured in BTC, which represents transfers between the sender and receiver and indicates the value of a transaction. *Miner Count* is the number of distinct miners that confirm the transactions for each type of user. *Data Insertion* is the percentage of data insertion transactions to the total transactions submitted by each user type. The *Waiting Time* for private users is calculated based on the waiting time proxy of their designated miners per confirmed transaction, while for normal users, it is the precise duration between entering the Mempool and receiving confirmation in the blockchain. We exclude miners' and mixers' transactions to address sample selection bias.

shown in Equation 5. We present our results in Columns (1), (2) and (3) of Table 4. We find that private users pay on average 2.6 satoshi/weight unit less than normal users. For a typical transaction with 1869.357 weight units, this amounts to 4860.328 sat or 2.27 USD, when converted at the average bitcoin price of 47,438.67 USD/BTC in our sample. On average, private users post 566.22 transactions per year in our sample, resulting in a savings of 1,282.84 USD compared to normal users.

We control for the sum of input values in BTC, which measures the monetary value of the transactions, transaction weight, the size of the block, the number of transactions and waiting time. We cluster the standard errors per block and user to control for heterogeneity in fee

variation. The relationship between the private dummy and the fee per weight is negative and significant. Consistent with our hypothesis, we find that private transactions are cheaper than normal ones.

Table 4. Fee per Weight Analysis The *PrivateDummy* is a dummy set to one if a transaction is sent from a user who utilizes private transactions and is zero otherwise. The regressions in columns (1), (2) and (3) are at the transaction level and the standard errors are clustered at block level. The dependent variable is *Fee per Weight*. The control variables are *Sum Inputs*, *Block Size* and *TX Weight*. In column (3) we remove the unknown miners from the sample and we add the *Wait Time* proxy to control for waiting time. The standard errors for regressions in columns (4), (5), and (6) are clustered at user level, and the dependent variable is *Std Fee per Weight*. The *TX Count* is the number of transactions submitted by a user. *TX Weight* is the weight of a transaction. The *Tx Count * PrivateDummy* is the interaction term of *TX Count* and *PrivateDummy*. This measure captures the effect of increasing the number of transactions on *Fee per Weight* if the submitting user is *Private*. The control variables are *Sum Inputs*, *Block Size*, *TX Weight* and *Wait Time*.

	(1)	(2)	(3)	(4)	(5)	(6)
	Fee per Weight	*Fee per Weight*	*Fee per Weight*	*Std Fee per Weight*	*Std Fee per Weight*	*Std Fee per Weight*
Private Dummy	-2.604***	-2.584***	-2.485***	-0.0360***	-0.0330***	-0.0245***
	(-14.99)	(-38.27)	(-14.25)	(-4.48)	(-4.14)	(-3.65)
Sum Inputs		7.91e-12***	8.21e-12***		-1.20e-13***	-1.10e-13***
		(10.89)	(47.20)		(-4.22)	(-4.42)
Block Size		0.000000607***	5.58e-07***		0.000000177***	0.000000170***
		(11.04)	(3.36)		(24.48)	(24.42)
TX Count					-4.06e-09	7.21e-09***
					(-1.21)	(2.96)
TX Count PrivateD*						-0.0000002 ***
						(-3.33)
TX Weight		-0.0000284***	-0.0000311***		-0.00000179***	-0.00000178***
		(-7.07)	(-71.42)		(-7.67)	(-7.66)
Wait Time			-3.32e-06			4.12e-06
			(-0.78)			(0.82)
Constant	12.92***	11.93***	12.12***	0.00214***	-0.229***	-0.219***
	(206.00)	(88.63)	(59.55)	(4.41)	(-23.51)	(-23.72)
Observations	97,637,471	97,637,471	81,303,198	59,086,692	59,086,692	59,086,692
R^2	0.582	0.677	0.602	0.524	0.613	0.622

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2 Fee Variation

To study the impact of long-term contracts on fee variation, we need to study fees on a per-user basis. We, therefore, compute the standard deviation of fees for each user we identify in our sample. Our regression, presented in Equation (6), follows the specification in the previous section. We present our findings in Table 4 in columns (4), (5) and (6). In line with our hypothesis, we document that the variation of fees for private users is significantly lower than for normal users.

$$\begin{aligned} \textit{Fee Variation} = \alpha_0 + \alpha_1 \textit{PrivateUserDummy} + \alpha_2 \textit{TxCount} + \alpha_3 \textit{TxCount*PrivateUserDummy} \\ + \alpha_4 \textit{SumInputs} + \alpha_5 \textit{BlockSize} + \alpha_6 \textit{TxWeight} + \alpha_7 \textit{WaitTime} + \varepsilon \end{aligned} \tag{6}$$

In Column (6), we add an interaction of transaction count, the number of transactions a user posts, and the private dummy, which we find negative and significant, This finding is consistent with fee variation being lower for private users that post more transactions. For normal users, we observe the opposite, as the coefficient of transaction count is positive and significant.

4.3 Same Miner Likelihood

If private users have an underlying contract with a settlement agent, we expect all of their transactions to be confirmed by the same miner. Since private transactions never enter the Mempool, other miners cannot include these transactions in their blocks. We count the number of distinct miners that process the users' transactions for each user. We regress the private dummy on the number of unique miners and control variables as detailed in Equation 7.

Our findings in Table 5 highlight that private transactions are confirmed by fewer miners. The coefficient of the distinct miner count is negative and significant. Our findings are consistent with the summary statistics in Table 3 that show private users' transactions are mined on average by 1.02 distinct miners while normal users' transactions are confirmed by 9.73 distinct miners.

Our evidence is consistent with our hypothesis that private users forge long-term exclusive contracts with mining pools.

$$\begin{aligned} \textit{PrivateUserDummy} = \alpha_0 + \alpha_1 \textit{DistinctMinerCnt} + \alpha_2 \textit{UserSumInputs} \\ + \alpha_3 \textit{UserTxCount} + \alpha_4 \textit{UserWaitTime} + \varepsilon \end{aligned} \tag{7}$$

Table 5. Likelihood to Stay with the Same Miner The dependent variable is *PrivateUserDummy*, which is a dummy variable set to one if a transaction is sent from a user who has submitted at least one private transaction and zero otherwise. The independent variable is *Number of distinct Miners*, representing the number of unique miners who confirmed transactions for each user. In Column (2) We control the regressions with *User Transaction Count* which is the total number of transactions submitted by each user. We also control for *User Sum Inputs* and *User Wait time* variables which measure the average values of sum inputs and wait time per transaction submitted by each user. Our sample consists of 59,086,692 users including 3,408,352 private and 55,678,340 normal users. Standard errors are clustered at the user level in all columns.

	(1)	(2)
	Private User Dummy	*Private User Dummy*
Number of distinct Miners	-0.00352***	-0.00443***
	(-7.50)	(-10.04)
User Sum Inputs		8.79e-14***
		(7.35)
User Transaction Count		0.000000609***
		(3.34)
User Wait Time		0.32e-07
		(1.32)
Constant	-1.541***	-1.412***
	(-710.74)	(-278.25)
Observations	59,086,692	59,086,692
R^2	0.453	0.611

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.4 Regularity of Transactions

We expect private users to post more transactions, and more evenly spaced transactions than normal users. Users such as exchanges or layer 2 protocols have a high transaction demand which is also evenly spread out over time. To study the regularity of transactions, we define *Time Dispersion* as the standard deviation of the time interval measured in blocks between two consecutive transactions for each user, and regress time dispersion on a dummy for private users and control variables as shown in Equation 8.

$$\begin{aligned} \textit{TimeDispersion} = \alpha_0 &+ \alpha_1 \textit{PrivateUserDummy} + \alpha_2 \textit{SumInputs} \\ &+ \alpha_3 \textit{BlockSize} + \alpha_4 \textit{TxWeight} + \alpha_5 \textit{TxCount} + \varepsilon \end{aligned} \tag{8}$$

We observe that private transactions have significantly lower variation in time intervals between transactions as reported in Table 6. Private transactions are more regular and evenly spaced over time. Strehle and Ante (2020) mentioned crypto exchanges as entities that regularly generate transactions. We manually match some private user addresses to known entities. Anecdotal evidence links several private wallet addresses to crypto exchanges.[14]

4.5 Miner Size

Our toy model suggests that market share determines the miners' reservation price for private transactions. Miners with a greater market share have less desire for the confirmation of private transactions due to fewer incentives for risk-sharing. To test our hypothesis we use a probit model to regress the private dummy on the miner market share and its square as specified in Equation (9). The coefficient of the *market share* variable is negative and significant as reported

[14]We attribute private wallet addresses using the 'https://www.walletexplorer.com' service. For example, the private wallet address '1445TYMH7NYjXyiusGeDrnuoyEvTSC7C2c' is related to BitZLato.com, which is an exchange. This wallet was active from January 22, 2021, to December 12, 2021, and submitted 127,578 private transactions.

Table 6. Regularity of Transactions The dependent variable is the *Std of Block Height Difference*, which is a proxy indicating the regularity of transactions. The independent variable is *Private User Dummy*, a dummy variable set to one if a transaction is sent from a user who utilizes private transactions and zero otherwise. The regression in Column (2) is controlled with variables *Sum Inputs*, *Block Size*, *Transaction Weight*, and *Transaction Count*. Standard errors are clustered at the user level in all columns.

	(1)	(2)
	Time Dispersion	*Time Dispersion*
Private User Dummy	-1,682.2***	-1,276.9***
	(-259.22)	(-191.97)
Sum Inputs		-6.20e-11***
		(-11.65)
Block Size		0.0000808***
		(20.52)
Transaction Weight		-0.000518***
		(-8.37)
Transaction Count		-0.000333***
		(-3.42)
Constant	1,779.0***	1,296.6***
	(696.49)	(199.61)
Observations	36,421,771	28,828,629
R^2	0.486	0.619

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

in Table 7. This finding aligns with our model, indicating that larger miners are significantly less likely to include private transactions in a block. The relationship is convex, indicating that small miners are the most involved in private mining.

$$\begin{aligned} \textit{PrivateDummy} = \alpha_0 + \alpha_1 \textit{MarketShare} + \alpha_2 \textit{(MarketShare)\^2} + \alpha_3 \textit{SumInputs} \\ + \alpha_4 \textit{BlockSize} + \alpha_5 \textit{TxWeight} + \alpha_6 \textit{WaitTime} + \varepsilon \end{aligned} \tag{9}$$

Table 7. Miner Size The dependent variable is *PrivateD*, which is a dummy variable set to one if a transaction is sent from a user who utilizes private transactions and zero otherwise. The independent variables are *Market Share* and *Market Share Square*. *Market Share* variable represents the percentage of mined blocks for each miner out of the total mined blocks. The *Market Share Square* is the square of market share. The regression in Column (2) is controlled with variables *Sum Inputs*, *Block Size*, *Transaction Weight*. In Column (3) we add *Wait Time* as a control variable and unknown miners are removed from the observations. The regression is run at the transaction level. Standard errors are clustered at the miner level in all columns.

	(1)	(2)	(3)
	Private Dummy	*Private Dummy*	*Private Dummy*
Market Share	-0.0983***	-0.0998***	-.01020***
	(-31.25)	(-31.93)	(-4.78)
Market Share Square	0.0077***	0.0078***	0.0011***
	(48.24)	(49.43)	(2.99)
Sum Inputs		-2.06e-14***	-3.41e-14***
		(-2.91)	(-4.32)
Block Size		-2.50e-07***	-4.92e-07***
		(-13.08)	(-35.21)
Transaction Weight		4.52e-07***	4.85e-08***
		(9.48)	(8.84)
Wait Time			1.36e-07
			(0.39)
Constant	-1.768***	-1.439***	-1.312***
	(-117.22)	(-51.03)	(-50.09)
Observations	97,637,471	97,637,471	81,303,198
R^2	0.134	0.137	0.105

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6 Data Insertion

Data insertion is a method to insert arbitrary data into the blockchain network. Sward, Vecna, and Stonedahl (2018) describe and compare various techniques for finding data insertion in the Bitcoin blockchain.[15] Data insertion is often utilized by meta layers or side-chains that process

[15]We find data insertion transactions by examining the content of the output scripts. If the output script field contains "OP_RETURN" string, that transaction is classified as a data insertion transaction. "OP_RETURN" is a script

transactions outside the Bitcoin blockchain but periodically post their chain state on Bitcoin for security. Such users have a steady demand to post transactions on the Bitcoin blockchain and are likely to contract privately with a miner. We study this fact through probit regression analyses based on Equation 11. In Table 8, we report that data insertion transactions are more likely to be sent through a private channel.

$$\begin{aligned} \textit{PrivateDummy} = \alpha_0 + \alpha_1 \textit{DataInsertionDummy} + \alpha_2 \textit{SumInputs} \\ + \alpha_3 \textit{BlockSize} + \alpha_4 \textit{TransactionWeight} + \alpha_5 \textit{WaitTime} + \varepsilon \end{aligned} \tag{10}$$

Table 8. Data Insertion Likelihood The dependent variable is *Private Dummy* and is a dummy variable set to one if a transaction is private and zero if it is normal. The *Data Insertion Dummy* is an independent variable set to one if the transaction type is data insertion. The regression in Column (2) is controlled with *Transaction Weight*, *Sum Inputs* and *Block Size*. We add the *Wait time* control variable to the regression and remove unknown miners from the sample in Column (3). The regression is run at the transaction level. Standard errors are clustered at the block level in all columns.

	(1)	(2)	(3)
	Private Dummy	*Private Dummy*	*Private Dummy*
Data Insertion Dummy	0.0227***	0.0269***	0.0536***
	(22.40)	(3.13)	(4.97)
Sum Inputs		4.45e-14***	3.55e-14***
		(8.55)	(4.50)
Block Size		-9.46e-08***	-4.90e-07***
		(-3.80)	(-34.92)
Transaction Weight		-1.55e-07***	-5.53e-08
		(-2.78)	(-0.96)
Wait Time			1.08e-07
			(0.31)
Constant	-1.573***	-1.447***	-1.301***
	(-203.76)	(-47.76)	(-72.24)
Observations	97,637,471	97,637,471	81,303,198
R^2	0.513	0.523	0.501

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

opcode used to mark a transaction output. Any Bitcoin associated with an output that contains an "OP_RETURN" is unspendable.

5 Robustness Tests

We conduct four tests to evaluate the robustness of our findings. The first test addresses the question of whether private transactions pay lower fees due to extended confirmation times. The second test assesses the lower variation in fees per block when private transactions are included in the block. We also test if the miner variation in fee revenue is lower if a miner has a smaller market share and confirms private transactions. The third test concentrates on evaluating the miner opportunity set at the time of confirming private transactions. The fourth test investigates the switching behaviour of private users among miners.

5.1 Matching Waiting Time

We examine whether private transactions pay lower fees, possibly due to longer waiting times for confirmation compared to regular transactions. In most of our baseline regressions, we control for wait time. In this section, we generate two different control samples. First, we pair each private transaction with all normal transactions with an equivalent waiting time within a six-block window around the private transaction. In the second matched sample, each private transaction is paired with a corresponding attributed synthetic normal transaction, which has the average fee per weight of all normal transactions in a six-block window around the private transaction.

$$\textit{FeeperWeight} = \alpha_0 + \alpha_1 \textit{PrivateDummy} + \varepsilon \tag{11}$$

We run the matching group fixed-effect regression based on Equation 11. We observe that private transactions pay significantly lower fees per weight compared to normal transactions with equivalent waiting times in both implementations, as reported in Table 9. Consequently, the lower fees paid by private transactions cannot be attributed solely to longer waiting times.

Table 9. Matching Waiting Time Regression The dependent variable is *Feeperweight*, while the independent variable *PrivateDummy* equals one for private transactions and zero for normal transactions. Normal transactions falling within the 50% range of the private waiting time proxy are grouped. In Column (1), we present the first implementation, where observations are drawn from 6,000 normal transactions (spanning almost three blocks) within the range of one private transaction. In Column (2), the second implementation is reported, relying on synthetic normal transactions. The total number of private transactions in our sample is 1,741,778 after excluding unknown miners and groups with no normal matching transactions. The number of categories equals the number of private transactions in both regressions. We conduct matching level fixed-effect for both regressions.

	(1)	(2)
	Fee per Weight	*Synthetic Fee per Weight*
Private Dummy	-9.513***	-9.527***
	(-808.17)	(-917.33)
Constant	16.62***	19.40***
	(2302.22)	(2642.44)
Observations	448,855,523	3,483,556
R^2	0.470	0.798

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.2 Miner Revenue Variation

Risk sharing is the motivation for our private transactions in our toy model. The miner has an interest in reducing the variability of her cash flows. We conduct two robustness checks. First, we examine the standard deviation of miner revenue per block in Table 10 as a function of the private transactions in that block, as laid out in Equations 12 and 13. As documented in Table 10 we find that blocks with more private transactions have lower in-block fee variation than normal blocks.

$$\textit{InBlockFeeVariation} = \alpha_0 + \alpha_1 \textit{PrivateBlockDummy} + \alpha_2 \textit{MarketShare} + \varepsilon \tag{12}$$

$$\textit{InBlockFeeVariation} = \alpha_0 + \alpha_1 \textit{PrivatePercentageperBlock} + \alpha_2 \textit{MarketShare} + \varepsilon \tag{13}$$

Second, we analyze the impact of miner market share and involvement in private transaction confirmation on the variation of miners' monthly income based on Equation 14. As shown in Table 11 we find that miners with smaller market shares experience more volatile monthly

Table 10. Miner Revenue in Block Variation The standard deviation of miner revenue per block is the dependent variable and is represented by *Std Total Fee Block*. The independent variable in Column (1) is the *Private Block Dummy*. This variable is a dummy set to one when more than one percent of transactions in a block are confirmed privately. In Column (2), the independent variable is *Private tx Percentage per Block*, representing the percentage of private transactions in each block. We control for the miner's *Market Share* in both regressions. The *Market Share* variable represents the percentage of mined blocks for each miner out of the total mined blocks.

	(1)	(2)
	In Block Fee Variation	*In Block Fee Variation*
Private Block Dummy	-978,558.6**	
	(-6.38)	
Market Share	-1,097,204.1**	-1,037,608.3***
	(-89.53)	(-82.58)
Private tx Percentage per Block		-75,010.1***
		(-20.48)
Constant	4.5***	4.9***
	(197.56)	(246.28)
Observations	68,737	68,737
R^2	0.105	0.110

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

revenue, while miners confirming private transactions exhibit more stable cash flows, aligning with our previous findings. The interaction term 'Small and Private Miner' reveals that miners with both a small market share and involvement in private transactions have a lower standard deviation in total revenue. These findings underscore the incentives for smaller miners to engage in private agreements with users.

$$\begin{aligned} \textit{MonthlyRevenueVariation} = \alpha_0 + \alpha_1 \textit{SmallDummy} + \alpha_2 \textit{PrivateDummy} \\ + \alpha_3 \textit{SmallandPrivateDummy} + \varepsilon \end{aligned} \tag{14}$$

Table 11. Miner's Monthly Revenue Variation The *Monthly Std Miner Revenue* variable is the standard deviation of miner revenue per month as the dependent variable. The *Small Miner* is a dummy variable set to one when a specific miner has less than ten percent of the market share. *Private Miner* is a dummy variable set to one if it is involved in private mining at least once. *Small and Private Miner* is an interaction term of the dummy variables set to one if a miner is both small and involved in private mining. The regression is controlled for month fixed effects.

	(1)
	Monthly STD Miner Revenue
Small Miner	0.206***
	(3.41)
Private Miner	-0.212***
	(-3.53)
Small and Private Miner	-0.151*
	(-1.92)
Constant	0.109**
	(2.34)
Observations	466
R^2	0.150

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.3 Miner Opportunity Set

One potential reason why private transactions pay lower fees is that miners use them as filler transactions, i.e., they include them when there are no better transactions in the Mempool. To investigate this alternative explanation we evaluate the miner's opportunity set at the time of including private transactions in a block.

Figure 2 illustrates the state of the Mempool just before block number 664521 with 799 confirmed transactions mined. The graph shows the number of transactions for quantiles of the fee per weight distribution. In this block, the miner includes 127 private transactions (red) and 672 normal transactions (blue). As expected the normal transactions are at the upper end of the fee/weight distribution consistent with the fact that miners want to maximize fee revenue. The

private transactions, however, are below the 10th percentile due to a lower fee per weight. The miner had a rich opportunity set with many transactions (green) that paid a higher fee per weight but were not included. This observation is consistent with miners having long-term agreements with private users. Without such agreements, confirming private transactions would contradict their short-term incentives, given the evident missed opportunities to maximize fee revenue.

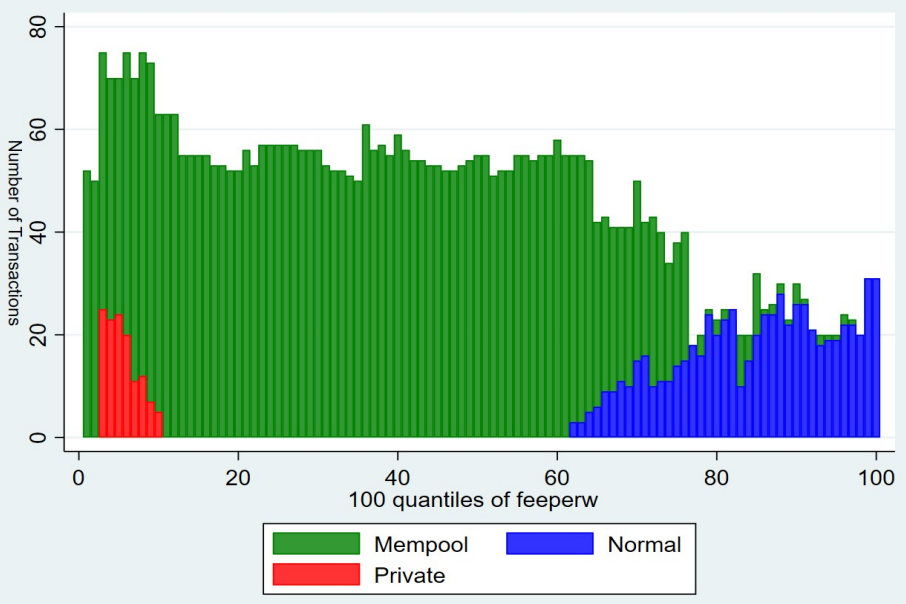


Figure 2. Miner's Opportunity Set This chart displays the fee per weight status for Mempool and confirmed, normal and, private transactions, at block number 664521. The x-axis represents the quantiles of fees per weight, while the y-axis represents the number of transactions. Waiting transactions in the Mempool that were eventually not mined are labelled *Mempoo (green)l*. *Normal (blue)* shows the normal transactions that are confirmed in this block. Private transactions (which are also included in teh block) are labelled as *Private (red)*.

$$PercentileDifference = \alpha_0 + \alpha_1 BlockSize + \alpha_2 MempoolSize + \varepsilon \tag{15}$$

For each block, we compute the average percentile of the fee per weight distribution for waiting Mempool and private transactions. The difference between waiting and private percentiles measures the miner's missed opportunity and is called the *Percentile Difference*. We regress the percentile difference on *Blocksize* and *Mempoolsize* to control for congestion based on Equation 15. The regression results are reported in Table 12. The constant is positive and significant,

highlighting that waiting Mempool transactions in the miner's opportunity set are located in higher percentiles compared to private transactions. Miners confirming private transactions with a lower fee per weight are losing out in the short term.

Table 12. Miners' Opportunity Set For each block, we calculate the average percentile of waiting transactions in Mempool and private transactions in the fee per weight distribution. The left-hand side variable *Percentile Diff* represents the difference between the average fee per weight percentiles of normal and private transactions. The observations represent the number of blocks during the 2021 sample period. The regression is controlled for *Block Size* and *Mempool Size*. In Column (2), the regression includes miner-fixed effects.

	(1)	(2)
	Percentile Diff	*Percentile Diff*
Block Size	0.000104***	0.000088***
	(43.55)	(39.04)
Mempool Size	0.000254***	0.000169***
	(7.91)	(5.57)
Constant	28.99***	51.75***
	(8.82)	(16.55)
Observations	52,686	52,686
R^2	0.039	0.140

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.4 Private User and Miner Switching

To better understand the long-term relationships between private users and miners we investigate cases in which private users switch between miners. Switching miners is very rare among private users. We observe that only 0.79% of private users switch from one miner to another. Private users who switch at least once stay with the same miner on average for 15,534 blocks, or roughly 108 days.

Table 13 represents examples of three private users switching from one miner to another in non-overlapping periods and toward the lower fee per weight payment. Private users generally switch to miners with a smaller market share.

Block Height	*Miner*	*Market Share(%)*	*Avg FeeperWeight*	*Transaction Count*
User Id = 10013				
664137	poolin	10.16	13.10	87
676436	BTC.com	7.54	6.02	72
693154	foundry_usa	4.61	1.21	35
User Id = 675627				
709168	AntPool	12.37	0.33	79
709230	foundry_usa	4.61	0.01	72
User Id = 1001015				
665461	F2Pool	12.73	32.21	70
666506	binance	8.83	31.45	48
668060	lubian.com	0.54	21.10	18

Table 13. Private User-Miner Switch Example The *Profitable Switches* for three private users are reported in this table. For each contract, the miner's *Market Share* and the private user's *Transaction Count* and *Avg Fee per Weight* are reported.

We hypothesize that private users gain profit by switching from bigger miners to smaller ones. A switch is profitable if a private user pays a lower fee per weight on average after the switch. This fee difference is user profit due to the switch, called *UserProfit*. The miners' market share difference is called *DiffMarketShare*. We implement regression analysis based on Equation 16 and findings are reported in Table 14. The coefficient of interest is negative and significant. Switching to a user with a higher market share of one unit results in a -0.04 unit decrease in the user profit. This finding is aligned with our hypothesis that switching to larger miners is not profitable for private users.

$$\textit{UserProfit} = \alpha_0 + \alpha_1 \textit{DiffMarketShare} + \alpha_2 \textit{User TX Count} + \alpha_3 \textit{User Wait Time} + \varepsilon \quad (16)$$

6 Conclusion

We offer novel insights into the fee settlement market within blockchain systems, particularly focusing on Bitcoin. Based on a sample of over a hundred million transactions, we find that private transactions consistently pay lower average fees and display lower fee variation compared to regular transactions. Furthermore, our investigation reveals distinct patterns of activity

Table 14. Profitable Switch and Miner Market Share The dependent variable is *UserProfit* which is the difference between the average fee per weight before and after the switch. *DiffMarketShare* is the difference between the market share of the miners before and after switching, *User TX Count* is the user transaction count, and *User Wait time* which is the proxy for private users' waiting time. The number of observations is the number of switches by private users in our one-year sample. Standard errors are clustered at the user level.

	(1)	(2)
	User profit	*User profit*
Diff Market Share	-0.04283***	-0.04281***
	(-5.00)	(-5.13)
User TX count		0.000023***
		(2.24)
User Wait Time		-0.0000974***
		(-2.58)
Constant	-0.358***	-0.367***
	(-13.89)	(-13.87)
Observations	142,628	142,628
R^2	0.003	0.004

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

among users of private transactions, highlighting their propensity for regular and evenly-spaced interactions with the blockchain. Despite their heightened activity, private users rely on fewer distinct miners for transaction processing compared to their regular counterparts.

Moreover, our research sheds light on the broader perspective of market dynamics within blockchain systems, challenging the idealized notion of Bitcoin as a purely decentralized and egalitarian financial network. The observed price discrimination and the emergence of different tiers of service underscore the complexities inherent in the Bitcoin settlement market, where certain users benefit from privileged access to lower fees and reduced uncertainty. These findings not only contribute to a deeper understanding of blockchain economics but also emphasize the need for further exploration into the evolving nature of financial markets in the era of decentralized technologies. Ultimately, our study highlights the intricacies of market behavior within blockchain systems and underscores the imperative for continued analysis to navigate

the complexities of emerging financial ecosystems effectively.

References

Andola, Nitish, Vijay Kumar Yadav, S Venkatesan, Shekhar Verma, et al., 2021, Anonymity on blockchain based e-cash protocols—A survey, *Computer Science Review* 40, 100394.

Athey, Susan, Ivo Parashkevov, Vishnu Sarukkai, and Jing Xia, 2016, Bitcoin pricing, adoption, and usage: Theory and evidence, .

Auer, Raphael, 2019, Embedded supervision: how to build regulation into blockchain finance, .

Balthasar, Thibault de, and Julio Hernandez-Castro, 2017, An analysis of bitcoin laundry services, in *Nordic Conference on Secure IT Systems* pp. 297–312. Springer.

Bjercke, Bjørn, and Keir Finlow-Bates, 2020, Decoupling Bitcoins from Their Transaction History Using the Coinbase Transaction, .

Brogaard, Jonathan, Allen Carrion, Thibaut Moyaert, Ryan Riordan, Andriy Shkilko, and Konstantin Sokolov, 2018, High frequency trading and extreme price movements, *Journal of Financial Economics* 128, 253–265.

Cong, Lin William, Zhiguo He, and Jiasun Li, 2021, Decentralized mining in centralized pools, *The Review of Financial Studies* 34, 1191–1235.

Cormen, Thomas H, Charles E Leiserson, Ronald L Rivest, and Clifford Stein, 2001, Introduction to algorithms second edition, *The Knuth-Morris-Pratt Algorithm*.

Dae-Yong, Kim, Essaid Meryam, and Ju Hongtaek, 2020, Examining Bitcoin mempools Resemblance Using Jaccard Similarity Index, in *2020 21st Asia-Pacific Network Operations and Management Symposium (APNOMS)* pp. 287–290. IEEE.

Daian, Philip, Steven Goldfeder, Tyler Kell, Yunqi Li, Xueyuan Zhao, Iddo Bentov, Lorenz Breidenbach, and Ari Juels, 2019, Flash boys 2.0: Frontrunning, transaction reordering, and consensus instability in decentralized exchanges, *arXiv preprint arXiv:1904.05234*.

Daian, Philip, Steven Goldfeder, Tyler Kell, Yunqi Li, Xueyuan Zhao, Iddo Bentov, Lorenz Breidenbach, and Ari Juels, 2020, Flash boys 2.0: Frontrunning in decentralized exchanges, miner extractable value, and consensus instability, in *2020 IEEE Symposium on Security and Privacy (SP)* pp. 910–927. IEEE.

Easley, David, Maureen O'Hara, and Soumya Basu, 2019, From mining to markets: The evolution of bitcoin transaction fees, *Journal of Financial Economics*.

Foley, Sean, Jonathan R Karlsen, and Tālis J Putniņš, 2019, Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?, *The Review of Financial Studies* 32, 1798–1853.

Gjermundrød, Harald, Konstantinos Chalkias, and Ioanna Dionysiou, 2016, Going beyond the coinbase transaction fee: Alternative reward schemes for miners in blockchain systems, in *Proceedings of the 20th Pan-Hellenic Conference on Informatics* pp. 1–4.

Greaves, Alex, and Benjamin Au, 2015, Using the bitcoin transaction graph to predict the price of bitcoin, *No data*.

Halaburda, Hanna, Guillaume Haeringer, Joshua Gans, and Neil Gandal, 2022, The microeconomics of cryptocurrencies, *Journal of Economic Literature* 60, 971–1013.

Hu, Danqi, Charles M Jones, and Xiaoyan Zhang, 2021, When Do Informed Short Sellers Trade? Evidence from Intraday Data and Implications for Informed Trading Models, *Evidence from Intraday Data and Implications for Informed Trading Models (February 16, 2021)*.

Huberman, Gur, Jacob D Leshno, and Ciamac Moallemi, 2021, Monopoly without a monopolist: An economic analysis of the bitcoin payment system, *The Review of Economic Studies* 88, 3011–3040.

Khalilov, Merve Can Kus, and Albert Levi, 2018, A survey on anonymity and privacy in bitcoin-like digital cash systems, *IEEE Communications Surveys & Tutorials* 20, 2543–2585.

Lehar, Alfred, and Christine A Parlour, 2022a, Battle of the Bots: Miner Extractable Value and Efficient Settlement, *Working paper*.

Lehar, Alfred, and Christine A Parlour, 2022b, Miner Collusion and the BitCoin Protocol, *Available at SSRN*.

Leskovec, Jure, Kevin J Lang, and Michael Mahoney, 2010, Empirical comparison of algorithms for network community detection, in *Proceedings of the 19th international conference on World wide web* pp. 631–640.

Meiklejohn, Sarah, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage, 2013, A fistful of bitcoins: characterizing payments among men with no names, in *Proceedings of the 2013 conference on Internet measurement conference* pp. 127–140.

Menkveld, Albert J, Bart Zhou Yueshen, and Haoxiang Zhu, 2017, Shades of darkness: A pecking order of trading venues, *Journal of Financial Economics* 124, 503–534.

Nakamoto, Satoshi, 2008, Bitcoin: A peer-to-peer electronic cash system, *Decentralized business review*.

Pakki, Jaswant, Yan Shoshitaishvili, Ruoyu Wang, Tiffany Bao, and Adam Doupé, 2021, Everything you ever wanted to know about bitcoin mixers (but were afraid to ask), in *International Conference on Financial Cryptography and Data Security* pp. 117–146. Springer.

Pappalardo, Giuseppe, Tiziana Di Matteo, Guido Caldarelli, and Tomaso Aste, 2018, Blockchain inefficiency in the Bitcoin peers network, *EPJ Data Science* 7, 1–13.

Park, Andreas, 2021, The Fatal Flaws of Constant Product Automated Market Making, working paper.

Ron, Dorit, and Adi Shamir, 2013, Quantitative analysis of the full bitcoin transaction graph, in *International Conference on Financial Cryptography and Data Security* pp. 6–24. Springer.

Russo, Daniela, Terry L Hart, Maria Chiara Malaguti, and Chryssa Papathanassiou, 2004, Governance of securities clearing and settlement systems, *ECB occasional paper*.

Schmiedel, Heiko, Markku Malkamäki, and Juha Tarkka, 2006, Economies of scale and technological development in securities depository and settlement systems, *Journal of Banking & Finance* 30, 1783–1806.

Strehle, Elias, and Lennart Ante, 2020, Exclusive mining of blockchain transactions, .

Sward, Andrew, Ivy Vecna, and Forrest Stonedahl, 2018, Data insertion in bitcoin's blockchain, *Ledger* 3.

Vlahavas, George, Kostas Karasavvas, and Athena Vakali, 2024, Unsupervised clustering of bitcoin transactions, *Financial Innovation* 10, 25.

Weintraub, Ben, Christof Ferreira Torres, Cristina Nita-Rotaru, and Radu State, 2022, A flash (bot) in the pan: measuring maximal extractable value in private pools, in *Proceedings of the 22nd ACM Internet Measurement Conference* pp. 458–471.

Wu, Lei, Yufeng Hu, Yajin Zhou, Haoyu Wang, Xiapu Luo, Zhi Wang, Fan Zhang, and Kui Ren, 2021, Towards understanding and demystifying bitcoin mixing services, in *Proceedings of the Web Conference 2021* pp. 33–44.

Xi, He, He Ketai, Lin Shenwen, Yang Jinglin, and Mao Hongliang, 2021, Bitcoin Address Clustering Method Based on Multiple Heuristic Conditions, *arXiv preprint arXiv:2104.09979*.

Xue, Gang, Jia Xu, Hanwen Wu, Weifeng Lu, and Lijie Xu, 2021, Incentive mechanism for rational miners in bitcoin mining pool, *Information Systems Frontiers* 23, 317–327.

Yanovich, Yuriy, Pavel Mischenko, and Aleksei Ostrovskiy, 2016, Shared send untangling in bitcoin, *bitfury. com* 2016, 1–25.

A Variable Description

#	Items	Description
1	Block Size	The original Block size limit in the Bitcoin blockchain is one megabyte. Post the segwit update part of the transaction can be outsourced into the witness section and weight units were introduced. Content in the witness section has a lower weight than content in the main block, effectively quadrupling block capacity. A block has a capacity of 4 million weight units, which is also sometimes expressed as 1 million vBytes (virtual bytes).
2	Data Insertion	This method utilizes the underlying blockchain mechanisms to store data in the chain by overwriting the output script. We identify data insertions by the use of the "OP_RETURN" op-code in the Bitcoin script. Other data insertion methods such as posting inserted data use script hash codes.
3	Fee per Weight	For miners, the opportunity cost of including a transaction in a new block is determined by its physical size in weight units, as block size is limited. The fee per weight measure is calculated by dividing the transaction fee by its weight.
4	Input Wallet Address	Input wallet address indicates a specific address from which bitcoins are being sent in a transaction. Each input typically corresponds to an unspent transaction output (UTXO) from a previous transaction. A user can post a transaction with multiple input addresses. In this case, the user should have access to the private key associated with each input wallet address to validate and sign them as a rightful owner of the Bitcoin.
5	Mempool Size	Normal transactions are waiting in the Mempool before including in a block by a miner. The size of the Mempool is the aggregate size in bytes of transactions waiting to be confirmed in Mempool per block.
6	Miner Market Share	Miners compete to win each block creation. Each block in the blockchain is mined by a specific miner. Miners usually signed their mined block. We count the number of blocks mined by each miner during our sample period. The proportion of blocks mined by each miner to the total number of blocks mined in the same period is the miner's market share.
7	Miner Opportunity Set	All transactions waiting in Mempool at the time of selecting a candidate block by the winner miner is the miner opportunity set. The miner has full discretion in selecting transactions from the Mempool and including them in a block.
8	Miner Revenue	Miner revenue consists of two deterministic and non-deterministic components including reward and fee. A miner who can add a block to the blockchain receives 6.25 bitcoin as a reward in our sample period. The block reward is set by the Bitcoin protocol and is roughly cut in half every four years (210,000 blocks). The miner also collects transaction fees for the transactions she confirms and includes in the newly mined block. The fee is offered by users when they post transactions on the Bitcoin blockchain. Miners collect rewards and fees to compensate for the costs of mining a new block and transaction confirmation respectively. The uncertainty of the non-deterministic part of the revenue which is the fee motivates miners to participate in risk-sharing long-term agreements with private users.
9	Private Transaction	A transaction which bypasses the Bitcoin protocol and is sent to the miner directly is private. Private transactions never wait in the Mempool.
10	Profitable Switch	A private user posts its transactions to a designated miner based on a long-term contract. Private users sometimes switch from one miner to another to pay a lower average fee, this switch is profitable for that user.
11	Transaction Fee	The Bitcoin protocol does not set the transaction fee in the blockchain. Users offer a fee to incentivize miners to settle their transactions. The general belief is that transactions with higher fees are more likely to get confirmed by miners.
12	Transaction Weight	Post segwit update, the size of the transaction is measured with weight units. Transaction weight consists of two main parts including the base and witness data. The base weight represents the size of the transaction excluding the witness data. It includes metadata, inputs, outputs and other non-witness data such as fees. Each block in the blockchain has a limited capacity of 4 million weight units. Therefore, a limited number of transactions fills the block.
13	Time Dispersion	The standard deviation of the time interval measured in blocks between two consecutive transactions for each user is the time dispersion. This measure shows how evenly transactions associated with a user are spaced through time. Regular transactions are the ones submitted by the same user which have lower time dispersion.
14	User Group	With the union-find algorithm, we classify input addresses into user groups. Each user group consists of several input wallet addresses. The same input wallet addresses used in various transactions are associated with one user group. Because the same user should have access to the private key of the input addresses.
15	Waiting Time	The difference between the block number in which the transaction was included in the blockchain and the highest block number when the transaction entered Mempool is waiting time. The waiting time for a private transaction is not observable as it never enters Mempool. We use a proxy for measuring private transaction waiting time which is half the distance between blocks that are mined by their designated miner.

Table 15. Description of Variables The variable descriptions are sorted alphabetically.

B User Classification

Based on the address alone, it is impossible to group observed addresses by wallets. Users can have various wallets and modern wallets create new addresses frequently. Therefore, By design, tracing individual users' activities through the Bitcoin blockchain is complex.

Our sample consists of 97,637,471 confirmed Bitcoin transactions for 2021. Each transaction can have multiple input wallet addresses. We observe 300,147,341 transaction hash-input address pairs. The number of unique input wallet addresses is 161,482,396. Table 16 reports the statistics in our sample. We cluster observations to 59,086,692 user groups.

Items	**Values**
Transaction hash - input address pairs	300,147,341
Distinct Input Wallet Addresses	161,482,396

Table 16. Input Wallet Addresses Statistics The summary statistics for input wallet addresses in 2021 are reported. The observations are transaction hash-input address pairs. The miners' and mixers' transactions are removed from the observations (see Appendix C).

We ascribe the hash-input addresses to users based on the Union-Find algorithm. This algorithm is widely used to classify transaction-level data into user-level data and was first introduced by Cormen, Leiserson, Rivest, and Stein (2001) and then widely applied in academic works such as Ron and Shamir (2013), Meiklejohn, Pomarole, Jordan, Levchenko, McCoy, Voelker, and Savage (2013), Khalilov and Levi (2018), Greaves and Au (2015) and Foley, Karlsen, and Putniņš (2019).

The algorithm first attributes all input addresses in a transaction to a user group. Then user groups with overlapping input addresses are merged in an iterative process. We create disjoint sets of input addresses such that each address in a set was (i) used with at least one other address of the same set in one transaction and was (ii) never used together with an input address from another set in one transaction.

To improve the accuracy of our study, we exclude mixers' and miners' transactions (see Appendix C) from the sample before classification. We detect 107,639 mixers' and 152,556 miners' transactions. We identify and exclude miners' wallet addresses through coinbase transactions.[16] Coinbase transactions transfer miner revenue including the total fee of confirmed transactions and block reward bitcoins to the miner's wallet address.

Our approach to detecting Bitcoin users aligns with the previous literature. Xi, Ketai, Shenwen, Jinglin, and Hongliang (2021) used a similar algorithm which they referred to as *common-input* heuristic. Our user classification works well, given the short sample period. Using a sample of 606 million transactions, Foley, Karlsen, and Putniņš (2019) identify 106 million users. Extracting Bitcoin miners' addresses based on the coinbase transactions has been utilized widely before such as (Gjermundrød, Chalkias, and Dionysiou 2016), (Bjercke and Finlow-Bates 2020) and (Vlahavas, Karasavvas, and Vakali 2024). [17]

B.1 Example 1:

This example describes the union-find algorithm that creates two user group disjoint sets in a simple step. Consider we have five transaction-address pairs as follows. First, we consider each pair as a unique user group set.

Pairs:

- *User Group 1*: (T1-add1)

- *User Group 2*: (T1-add2)

- *User Group 3*: (T2-add3)

[16]The first confirmed transaction in each block, called coinbase.

[17]Our user group classification algorithm may not result in perfect clustering if a user never utilizes mutual input addresses. In this case, the algorithm detects disjoint sets leading to different user groups; while they belong to the same user. Our estimate can therefore be seen as an upper bound on the number of users.

- *User Group 4*: (T2-add2)

- *User Group 5*: (T3-add4)

The Algorithm detects that add2 is mutual between two sets and then unites the sets. Since input address two (add2) is mutual between user groups 2 and 4, the sets will unite and create a bigger set containing transaction one (T1) and transaction two (T2) including their related inputs. Therefore, we detect two disjoint user groups in one step.

Step1:

-*User Group 1234*: (T1-add2 ,T2-add2, T1-add1, T2-add3)

-*User Group 5*: (T3-add4)

B.2 Example 2:

This example highlights the recursive nature of the user classification algorithm. In this case, the algorithm detects two disjoint sets in two steps. Consider we have six observations as follows:

Pairs:

- *User Group 1*: (T1-add1, T1-add5)

- *User Group 2*: (T2-add2, T2-add3)

- *User Group 3*: (T3-add3, T3-add1)

- *User Group 4*: (T4-add2)

- *User Group 5*: (T5-add4)

- *User Group 6*: (T6-add1)

Step1: Creating a set to find the relationship between transactions and addresses based on

the mutual address add1 between groups 1, 3 and 6 and mutual address add2 between groups 2 and 4.

- *User Group 136*: (T1-add1, T1-add5, T3-add3, T3-add1, T6-add1)

- *User Group 24*: (T2-add2, T2-add3, T4-add2)

- *User Group 5*: (T5-add4)

Step2: In this step, the address add3 is mutual between user groups 136 and 24. The algorithm detects two disjoint user groups.

- *User Group 12346*: (T1-add1, T1-add5, T3-add3, T3-add1, T6-add1, T2-add2, T2-add3, T4-add2)

- *User Group 5*: (T5-add4)

C Mixed Transaction Detection

Bitcoin mixers provide improved anonymity for Bitcoin users by breaking the connection between wallet addresses and the identity of their true owners. These services leverage inherent characteristics of both Bitcoin and blockchain technology (Pakki, Shoshitaishvili, Wang, Bao, and Doupé 2021). Mixing the transactions of various users in several steps makes it difficult to trace the original source or destination of the funds. Bitcoin mixers provide a layer of privacy by breaking the link between the sender and recipient of the funds. They are often used by individuals who want to enhance the privacy of their transactions, as well as by those who wish to anonymize their Bitcoin holdings for various reasons, including security and confidentiality concerns.

There are two main types of mixers centralized and decentralized in Bitcoin blockchain. The *Centralized* mixers are operated by private service providers such as Yo!Mix and Mixtura. Users send the Bitcoin to the wallet addresses owned by the mixer and pay the service fee. The user should specify the destination address as well. Mixers then combine the received funds from several users and redistribute them in the blockchain. The *Decentralized* mixers utilize protocols like *CoinJoin* to create an automatic mixing algorithm to service several users at the same time and redistribute them to the recipient addresses. CoinJoin operates through open-source protocols allowing users to mix their funds in a peer-to-peer manner.

In general, mixing services have three phases: taking inputs, mixing, and sending outputs. These phases are designed to enhance transactional privacy and make it more challenging to track specific addresses involved in Bitcoin transactions. In *Input* phase, users send their cryptocurrency to the mixing services. In the *Mixing* phase, the mixer service combines various incoming funds from different users to obfuscate the origin of the funds. This phase can take several steps to improve anonymity by making it harder to trace transactions. The last step is

Output phase in which mixed funds are redistributed to the specified destination addresses.

Although Bitcoin mixers offer users anonymity, they are not without risks. The centralized mixers bear a risk of loss of funds if the private service provider shuts down or gets hacked. The decentralized mixers also rely on multiple participants for their mixing processes and aim to remain anonymous. There is also a risk of falling victim to scams and fraud when using Bitcoin mixers, as well as potential vulnerabilities since the implemented mixing service is not disclosed. Mixers are less secure than regular transactions because in order to combine inputs from multiple users the users must hand over their signature to the off-chain mixer because Bitcoin has no smart contract ability or send the funds to the mixers' wallet addresses. In theory, the mixer could divert the money to their own wallet. Mixers therefore rely on reputation.

Detecting mixed transactions improves the user classification algorithm. Although Bitcoin mixing services are evolving, they are not documented to stay untraceable. The mixing services differ with respect to fees, security, and execution time. (Balthasar and Hernandez-Castro 2017) study and rank various mixing services based on their design, tractability and their attack weakness. Typical fees are either proportional to the amount (2% to 3%), or fixed prices, such as 0.01 BTC. There are three general approaches for detecting Bitcoin mixers in literature.

The first strand tries to identify the relations between Bitcoin address and users by clustering (classification) such as (Khalilov and Levi 2018) and (Andola, Yadav, Venkatesan, Verma, et al. 2021) and (Yanovich, Mischenko, and Ostrovskiy 2016). Then the analyzer can detect an out-of-norm (irregular) transaction behaviour by a specific user. Leskovec, Lang, and Mahoney (2010) and Xi, Ketai, Shenwen, Jinglin, and Hongliang (2021) evaluate the modularity of mixed transactions and quantify the quality of the detected transactions as mixers by adding a new transaction to the group.

The second strand studies the representative mixing services. This approach is practical based on the returned information from mixers' services and analyzing that based on user types

and address prefixes. Most public mixers are black-box services, which do not have their code available to the public. To tackle this challenge, Pakki, Shoshitaishvili, Wang, Bao, and Doupé (2021) and Wu, Hu, Zhou, Wang, Luo, Wang, Zhang, and Ren (2021) interact with Five and Four real public mixers, respectively, to identify actual behaviours and indicative of their implementation and their resistance. In this approach, the representative mixing services are detected based on *Bitcoin Talk* and public media.[18] In the *Bitcoin Talk* forum, mixers advertise their services, and there are multiple venues where people talk about mixing services from technical and application points of view. This forum allows users to distinguish trusted mixers as mentioned in (Pakki, Shoshitaishvili, Wang, Bao, and Doupé 2021). Based on interaction with the real mixing services, a sample of input-output transactions will be created based on the return address to analyze. Interacting with mixers has two general types based on cost and technology.

The third strand is based on classifying the mixers' transactions according to transactional specifications such as time, value and input-output addresses. Pakki, Shoshitaishvili, Wang, Bao, and Doupé (2021) finds that delays are maximized in mixed transactions compared to normal ones. Technically input mixing services happen automatically almost one minute out of the chain. Then mixers are a random delay from a few hours to one day to post the batch of mixed transactions to the Bitcoin blockchain. Balthasar and Hernandez-Castro (2017) shows that mixer services may have service limitations that one can utilize to detect and exclude them from the sample. For example, many mixers have an upper limit (1,000 BTC) and a lower limit (0.01 BTC) for the transfer amount. Mixed transactions usually have multiple inputs and multiple outputs. Athey, Parashkevov, Sarukkai, and Xia (2016) shows that transactions with more than 4 inputs and 4 outputs are likely to be created by mixers. Whirlpool mixing services have five inputs and five outputs. In addition, they have equal split amounts (0.001, 0.01, 0.05 and 0.5). Wasabi and Joinmarket are two implementations of the Coinjoin protocol which both

[18] https://bitcointalk.org/

have a specific number of inputs and outputs.[19]

We implement a detecting algorithm based on mixers' transaction specification which is compatible with the third strand. CoinJoin implementations are widely used because of their resistance against join-then-abort attacks, ensuring users cannot disrupt the mix after sending funds to the input address. Our approach is to detect and filter out two transaction types that are typical of the Coinjoin protocol which is a peer-to-peer mixer type. The first type of transaction has at least five equal outputs and five equal inputs. The second type of transaction has more than ten equal outputs and the number of inputs is greater than the number of outputs.[20] In our 2021 Bitcoin sample, our program detects 112,820 mixed transactions, comprising 110,956 for type 1 and 1,864 for type 2. Mixing these two sets, we have 107,639 unique mixed transactions as reported in Table 17. We excluded the transactions of mixers and miners from our sample, leading to 97,637,471 records of data.

Mixers' Transaction Type	**Count**
Type 1	110,956
Type 2	1,864
Total	112,820
Unique	107,639

Table 17. Mixers' Transactions This table reports the summary statistics for mixers transactions in our 2021 sample.

[19]https://wasabiwallet.io/api/v4/btc/ChaumianCoinJoin/unconfirmed-coinjoins
https://github.com/JoinMarket-Org/joinmarket-clientserver/blob/master/scripts/snicker/snicker-finder.py

[20]The Coinjoin protocol implementations are available in GitHub https://github.com/nopara73/Dumplings/blob/master/Dumplings/Sca