Transaction Fee Mining and Mechanism Design

COS521 Final Project

Michael Tang & Alex Zhang

Princeton University

Table of contents

1. Motivation

2. Mining & Mining Attacks

3. Mechanism Design

4. The World After Impossibility

5. Overview of Transaction Fee Mechanisms

6. Conclusion

Motivation

Motivation

Motivation:

- Growing global interest in blockchain: billions of USD worth of transaction volume, tens of millions of users.
- Most of them rely (or plan to rely) primarily on transaction-fee-based incentives

Key Question: How robust is a blockchain under transaction fees?

Overview of Mining & Attacks

The Mining Game

- Assume a set of n miners $m_1, ..., m_n$, each with proportional mining power $\chi(m_i)$ s.t. $\sum_{i=1}^{n} \chi(m_i) = 1$
- Each miner m_i is aware of their private blockchain $G(m_i)$, represented as a tree, and they can choose a block to mine on
- After a time interval that is exponentially distributed with mean $\chi(m_i)^{-1}$, the miner will discover a new block B and add a directed edge to the node they were mining on in $G(m_i)$.
- The miner will select a set of transactions $\mathcal{T}(B)$ to include on their newly mined block if it eventually remains on the public chain.
- Miners may choose when to broadcast their mined block(s).

Our setup will be time-driven, so we observe and analyze the state of the system at some time t.

Selfish Miner (for transaction fees)

The *SelfishMiner* strategy [4] for a miner m_i will

1. Choose to mine on $\mathcal{H}_{m_i}$, which is the highest block on their private chain.
2. Take all remaining available transaction fees $\mathcal{R}(\mathcal{H}_{m_i})$ upon discovering a new block B.
3. If $h(B) = \mathcal{H}$, publish the block immediately. Otherwise, if there exists two distinct blocks of height $\mathcal{H}$ where one is owned by m_i, and $\mathcal{H}_{m_i} = \mathcal{H} + 1$, reveal $b^{\mathcal{H}}_{m_i}$.

tl;dr: Selfish miner will take a risk by hiding a block they found, revealing them when they have a large enough lead over the public chain to force other honest miners to waste computational power mining.

Selfish Mining in Transaction-fee Regime

Theorem

[2] Let $\gamma \in [0, 1]$ be the probability that if the selfish miner m_i is in a race with the public chain, it is not orphaned. Then given $\chi(m_i) = \alpha \in (0, 0.5)$, the expected reward of **`SelfishMiner`** *is*

$$\mathsf{E}(\textit{reward}) = \frac{5\alpha^2 - 12\alpha^3 + 9\alpha^4 - 2\alpha^5 + \gamma(\alpha - 4\alpha^2 + 6\alpha^3 - 5\alpha^4 + 2\alpha^5)}{2\alpha^3 - 4\alpha^2 + 1}$$

Proof sketch:

1. Define different states that the selfish miner can be in.
2. Let f_s be the prob. that a txn is in the block found by m_i if in state s, and p_s be the prob. that m_i is in state s.
3. Use Markov-chain to compute p_s and f_s and get the exp. reward for the miner as $\sum_s f_s \cdot p_s$.

Optimized Selfish Mining in Transaction-fee Regime

Theorem

[2] Let $\gamma \in [0,1]$ be the probability that if the selfish miner m_i is in a race with the public chain, it is not orphaned. For a ***`FeeSelfishMiner`*** *miner m_i using threshold β, given $\chi(m_i) = \alpha \in (0, 0.5)$, the expected reward is*

$$\mathsf{E}(\textit{reward}) = \left(\frac{1 + \beta(1-\alpha)^2(1-\gamma)}{e^{\beta} - 1} + 5\alpha + (1-\alpha)^2\gamma + \frac{2\alpha^2}{1-2\alpha} - 2\alpha^2 \right)$$

$$\times \left(\frac{\alpha(1-2\alpha)(1-e^{-\beta})}{1 - 2e^{-\beta}\alpha - 3(1-e^{-\beta})\alpha^2} \right)$$

Remark

The proof is very similar to the proof for regular selfish mining, but is a bit more convoluted computationally because there is an extra state $0''$ to consider where the selfish miner acts like an honest miner.

Other Attacks

Undercutting Attacks [2, 5]

1. The *PettyCompliant* strategy will choose to mine on the block with the most available transaction fees rather than the oldest $\rightarrow$ strictly better than honest miner.
2. The *FunctionFork* strategy will fork nodes to claim high transaction fees while leaving a large sum behind to incentivize *PettyCompliant* miners to claim them.

Fee-sniping and Whale Transactions

1. Fee-sniping is to fork blocks with high-fees and incentivize other miners to mine on top.
2. Whale transaction attacks [7] fork a history of the chain to undo or double spend a transaction, leaving high-transaction fees behind to incentivize miners to mine on the fork.

Mechanism Design

Revisiting the Mining Game

- Assume a set of n miners $m_1, ..., m_n$, each with proportional mining power $\chi(m_i)$ s.t. $\sum_{i=1}^{n} \chi(m_i) = 1$
- Each miner m_i is aware of their private blockchain $G(m_i)$ and they can choose any node/block on this chain (represented as a tree) to mine on.
- After a time interval that is exponentially distributed with mean $\chi(m_i)^{-1}$, the miner will discover a new block B and add a directed edge to the node they were mining on in $G(m_i)$.
- The miner will select a set of transactions $\mathcal{T}(B)$ to include on their newly mined block if it eventually remains on the public chain.
- Miners may choose when to broadcast their mined block(s).

Transaction Fee Mechanism (TFM)

We consider a single auction instance corresponding to the next mined block. The blockchain is honest, but users and miners are strategic.

1. m users with values $\mathbf{v}$ for having their transaction included in the block, submit bids $\mathbf{b}$
2. the miner picks a subset of k bids $B := \mathbf{I}(\mathbf{b}) \subseteq \mathbf{b}$ to include in the block
3. the blockchain then confirms a further subset $\mathbf{C}(B, \mathbf{b}) \subseteq B$
4. the blockchain enforces payments $\mathbf{p} := \mathbf{P}(B, \mathbf{b}) \geq \mathbf{0}$ for each user whose transaction was confirmed, and revenue $r := \mathbf{R}(B, \mathbf{b}) \geq 0$ for the miner

A TFM is completely described by $(\mathbf{I}, \mathbf{C}, \mathbf{P}, \mathbf{R})$.

Incentive Compatibility

Definition (UIC)

A TFM is *user incentive compatible* (UIC) if truthful bidding is dominant for users.

Definition (MMIC)

A TFM is *myopic miner incentive compatible* (MMIC) if truthful implementation of the mechanism is dominant for miners where only single-round utility is considered

Definition (OCA-proof)

A TFM is *off-chain-agreement-proof* (OCA-proof) if no off-chain agreement between the miner and any number of users can improve joint utility over the outcome from bidding and implementing the mechanism honestly, respectively. We can relax this by denoting TFMs robust against OCAs of the miner and up to c users as c-OCA-proof.

Incentive Compatibility Is Natural

A few desiderata:

1. Simplify user bidding, which reduces overpayment, user-side computational costs, and overall user experience
   $\implies$ **UIC, OCA-proof**
2. Increase network capacity and reduce delays for users
   $\implies$ **MMIC, α-costly**
3. Increase network robustness and decentralization properties
   $\implies$ **mining defenses, OCA-proof**
4. Allow users to pay for priority inclusion in a block
   $\implies$ **UIC**

3-Way Impossibility Theorem

Theorem

Assuming finite block-size, no non-trivial TFM can have UIC and 1-OCA-proof, i.e. three-way incentive compatibility is impossible. [3]

Proof Sketch

1. Any TFM satisfying UIC must satisfy constraints described in Myerson's Lemma.
2. Any TFM satisfying 1-OCA-proof must satisfy a certain inequality.
3. Use above 2 to show that any TFM satisfying both UIC and 1-OCA-proof imply zero miner revenue
4. Use above to show that any TFM with finite block sizes satisfying both UIC and 1-OCA-proof is just the trivial mechanism where no one gets anything.

The World After Impossibility

Weaker Incentive Compatibility

Definition (γ-strict utility)

Let a player, miner, or cartel's utility under a strategy be u, which does not include unconfirmed bids. For each unconfirmed bid i let u_i be the utility resulting from that bid being confirmed. Note that this may include miner revenue coming from the bid. Then, their *γ-strict utility* is $u + \gamma \sum_i \min(u_i, 0)$ for $\gamma \in (0, 1]$. [3]

Definition (γ-weak incentive compatibility)

For $\gamma \in (0, 1]$, let γ-weak UIC be UIC under γ-strict utility. Define γ-weak MIC, γ-weak c-OCA-proof respectively for MIC, c-OCA-proof. [3]

Lemma. *Standard utility $\geq$ γ-strict utility*

Lemma. *u_γ is monotonically decreasing in γ*

Lemma. *Not 1-weak IC $\implies$ not γ-weak IC $\forall \gamma \in (0, 1]$ $\implies$ not IC*

Beyond Incentive Compatibility

Definition (α-costly)

A TFM is α-costly if each confirmed fake transaction decreases its bidder's utility by at least α.

Remark

- Usually compare to an ϵ-costly baseline, (any UIC and MMIC TFM). Assume $\epsilon > 0$, e.g. due to orphan risk
- Looking for large constant $\alpha \gg \epsilon$.

Overview of Transaction Fee Mechanisms

Abridged TFM Summary

	UIC	MMIC	OCA-proof
First-price		✓	✓
Posted-price	✓	✓*	
EIP-1559 (low demand)	✓**	✓*	✓
EIP-1559 (high demand)		✓*	✓

* $(\epsilon + \alpha)$-costly for a constant parameter α

** UIC-like, i.e. the paying the base price is a Nash

Where Classical Designs Fail: First-Price Auction

The *`First-Price Auction`* mechanism is parameterized by the block size B. It behaves as follows: (differences from first-price bolded)

1. **I**: Include the B highest bids $b_1 \geq ... \geq b_B$, breaking ties arbitrarily.
2. **C**: Confirm all bids.
3. **P, R**: The ith confirmed bid pays b_i to the miner, and unconfirmed bids pay nothing.

Lemma. *First-price auctions are not 1-weak UIC*

Proof. Suppose $v_1 = 10, b_2 = 2, b_3 = 1$. The first bidder can save at least 7 by bidding e.g. 3 and still having their bid confirmed. Note that no fake transactions are necessary here so 1-strict utility is the same as regular utility.

Where Classical Designs Fail: Posted-Price Auction

The *`Posted-Price Auction`* mechanism is parameterized by the block size B and a posted price p. It behaves as follows:

1. **I**: Include any B bids (that equal p).
2. **C**: Confirm all bids (that equal p).
3. **P, R**: All confirmed bids pay p to the miner and unconfirmed bids pay nothing.

Lemma. *Posted-price auctions are not 1-weak OCA-proof*

Proof. Suppose $p = 10, \mathbf{b} = (0, 0, 0, 0), \mathbf{v} = (5, 0, 0, 0)$. The miner can form a cartel with user 1 and have them bid the posted 10, which increases their joint utility from 0 to $5 - 10 + 10 = 5$, which does not change under 1-strict utility.

A Realistic TFM: EIP-1559

The *EIP-1559 Mechanism* [1][8][9] is parameterized by the block size B and a posted price p. It behaves as follows: (differences from second-price bolded)

1. **I**: Include highest B bids $b_1 \geq ... \geq b_B \geq$ **p**, breaking ties arbitrarily.
2. **C**: Confirm all bids. Bids must be $\geq p$.
3. **P**: The ith confirmed bid pays b_i, and unconfirmed bids pay nothing.
4. **R**: For the ith confirmed bid, the miner gets $\mathbf{b_i - p}$ **and the remaining** p **is burned**.

Intuitively, p is usually denoted the *base fee* and the remaining payment $b_i - p$ the *tip*.

A Realistic TFM: EIP-1559

Lemma. *EIP-1559 is MMIC*

Proof sketch. FPA on tips

Lemma. *EIP-1559 is in fact $(p + \epsilon)$-costly*

Proof. Any fake transaction incurs at least p

Lemma. *EIP-1559 is OCA-proof*

Proof sketch. Users should bid above or below p honestly, and tips are paid to miner so are arbitrary, so we can set $p_i = v_i$. Then, inclusion is honest.

A Realistic TFM: EIP-1559

Definition (demand)

Bids $\mathbf{b}$ are *low demand* if at most B users have value strictly greater than p, and *high demand* otherwise

Lemma. *If low demand, EIP-1559 is UIC-like in the sense that there is a Nash of paying the base fee. [8][9]*

Proof sketch. $b_i = min(p, v_i)$ is a Nash.

Lemma. *Sadly, EIP-1559 is not 1-weak UIC generally*

Proof sketch. Suppose $p = 5, B = 3, \mathbf{v} = (16, 10, 10, 10), b_2 = b_3 = 5$. First bidder saves 5 by bidding 11 instead of 16. No fake transactions $\implies$ not even 1-weak.

Extended TFM Summary

	UIC	MMIC	OCA-proof
First-price		✓	✓
Second-price	✓		
Monopolistic [6][10]	nearly	✓	
Posted-price	✓	✓*	
EIP-1559 (low demand)	✓**	✓*	✓
EIP-1559 (high demand)		✓*	✓
Tipless EIP-1559 (low demand)	✓**	✓*	✓
Tipless EIP-1559 (high demand)	✓	✓*	
Burning second-price [3]	weak	weak	weak

nearly: honesty is an ϵ-BNE for any fixed F with $\mathbf{v} \sim F$ [6][10]

* $(\epsilon + \alpha)$-costly for a constant parameter α

** UIC-like, i.e. the paying the base price is a Nash

Conclusion

Summary

1. Many ways to deviate under the intuitive transaction-fee setup
    - Mining, bidding, inclusion
2. We can do better!
3. Formal analysis is both useful and needed
    - Analysis of attacks
    - New defenses and better mechanisms
    - Impossibility results

Open Questions (more in paper)

1. Are the currently investigated undercutting and fee sniping attacks resolved completely by the nLockTime protocol, and if not, what new types of equilibrium behavior arise?
2. Where is the provable limit for incentive-compatibility exactly: can we prove tighter impossibility bounds (e.g. with properties beyond zero miner revenue) or show examples of TFMs with stronger guarantees?
3. What is the optimal tradeoff surrounding the failure regime of EIP-1559, and how do we mitigate deviations in those conditions?
4. Given that burning is provably required for some incentive compatibility notions, are there robust ways of paying value forward to miners without encouraging fake transactions and OCAs?

Questions?

References i


V. Buterin, E. Conner, R. Dudley, M. Slipper, I. Norden, and A. Bakhta.
Eip-1559 specification, 2021.


M. Carlsten, H. Kalodner, S. M. Weinberg, and A. Narayanan.
On the instability of bitcoin without the block reward.
In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, page 154–167, New York, NY, USA, 2016. Association for Computing Machinery.


H. Chung and E. Shi.
Foundations of transaction fee mechanism design.
CoRR, abs/2111.03151, 2021.

References ii


I. Eyal and E. G. Sirer.
Majority is not enough: Bitcoin mining is vulnerable.
CoRR, abs/1311.0243, 2013.


T. Gong, M. Minaei, W. Sun, and A. Kate.
Towards overcoming the undercutting problem.
In I. Eyal and J. Garay, editors, *Financial Cryptography and Data Security*, pages 444–463, Cham, 2022. Springer International Publishing.


R. Lavi, O. Sattath, and A. Zohar.
Redesigning bitcoin's fee market.
ACM Trans. Econ. Comput., 10(1), may 2022.

References iii


K. Liao and J. Katz.
Incentivizing blockchain forks via whale transactions.
In M. Brenner, K. Rohloff, J. Bonneau, A. Miller, P. Y. Ryan, V. Teague, A. Bracciali, M. Sala, F. Pintore, and M. Jakobsson, editors, *Financial Cryptography and Data Security*, pages 264–279, Cham, 2017. Springer International Publishing.


T. Roughgarden.
Transaction fee mechanism design for the ethereum blockchain: An economic analysis of EIP-1559.
CoRR, abs/2012.00854, 2020.


T. Roughgarden.
Transaction fee mechanism design.
CoRR, abs/2106.01340, 2021.

References iv


A. C. Yao.
An incentive analysis of some bitcoin fee designs.
CoRR, abs/1811.02351, 2018.