Revisiting the Primitives of Transaction Fee Mechanism Design[*]

Aadityan Ganesh[†]   Clayton Thomas[‡]   S. Matthew Weinberg[§]

July 8, 2024

Abstract

Transaction Fee Mechanism Design—a rapidly-evolving research agenda initiated by Roughgarden (2021)—studies auctions run by untrusted miners for transaction inclusion in a blockchain. Under previously-considered desiderata, an auction is considered 'good' if, informally-speaking, each party (i.e., the miner, the users, and coalitions of both miners and users) has no incentive to deviate from the fixed and pre-determined protocol. In other words, previous works posit that a 'good' auction should be 'simple for users', 'simple for miners', and 'resistant to collusion'.

In this paper, we propose an alternative desiderata for Transaction Fee Mechanisms. We say that a TFM is *off-chain influence proof* when the miner cannot achieve additional revenue by running a separate auction off-chain. While the previously-highlighted EIP-1559 satisfies previously-considered desiderata (simplicity for miners and users, as well as collusion resistance), we show that it *does not* satisfy off-chain influence proofness. Intuitively, this holds because a Bayesian revenue-maximizing miner can strictly increase profits by threatening to censor any bids that do not transfer a tip directly to the miner off-chain. We prove a strong impossibility result: no mechanism satisfies all previously-considered properties along with off-chain influence proofness. On the other hand, we reconsider the Cryptographic (multi-party computation assisted) Second Price Auction mechanism (Shi et al., 2023). We argue that the space of TFMs can be naturally expanded to solicit an input from the miner, for example, by asking them to set the reserve price of the auction. We show that in this model, the cryptographic second price auction satisfies simplicity for users and miners and off-chain influence proofness, since it allows a Bayesian miner to maximize their revenue by posting an optimal reserve price.

[†]Princeton University — *E-mail*: aadityanganesh@princeton.edu.

[‡]Microsoft Research — *E-mail*: thomas.clay95@gmail.com.

[§]Princeton University — *E-mail*: smweinberg@princeton.edu.

1   Introduction

Transaction fee mechanisms (henceforth, TFMs) are auctions used by blockchains to decide which transactions are included in each block, along with a price ("transaction fee") charged to the user submitting the transaction. TFMs are an integral part of blockchain infrastructure, and are hotly debated. In one high-profile example, the Ethereum blockchain switched from the status-quo first-price auction (which is still in use in Bitcoin) to a mechanism known as EIP-1559. In essence, EIP-1559 is a posted-price mechanism with a variable price exogenously set by market activity, in which all user-paid transaction fees are discarded ("burned") and the miner[1] who constructs the block receives only a positive constant revenue ("block reward").

The redesign of Ethereum's TFM was meant to simplify bidding behavior for users (who may be small-scale retail users valuing ease-of-participation), while avoiding placing trust in the miner running the TFM algorithm (who may be a strategically sophisticated, untrusted, profit-maximizing agent). Balancing these two objectives is a challenging task. For example, in the pre-existing First-Price Auction, users were frustrated that they had to bid strategically,[2] while under a Second-Price Auction, we would expect the pseudonymous miner to submit a shill bid equal to (just under) the highest submitted bid. In contrast, a series of works (discussed in detail below) has shown that EIP-1559 provides formal guarantees on agents' incentives: participants (users, miners, or even coalitions containing both the miner and some set of users) cannot increase their utility by manipulating their inputs to the protocol.

Consider, however, the following miner strategy in EIP-1559. The miner publicly announces *off-chain* that they will censor all bids that do not directly transfer the miner an "entry fee" γ via some off-chain payment platform. If *any* user transfers the miner this γ in order to get their transaction considered, then the miner will strictly profit from this attack. Thus, despite the fact that the miner cannot profit in EIP-1559 by deviating from the protocol while remaining entirely on-chain, the miner *can* profit by making convincing threats and collecting payments off-chain.

In this paper, we take "off-chain" attacks of the above form seriously, and propose a model in which the miner has the ability to conduct *any off-chain mechanism whatsoever* in order to determine the inputs of the TFM algorithm. Using this model, we formalize and highlight the above novel attack on EIP-1559. Alongside this, we also make the natural suggestion that the miner be *allowed* to strategize on-chain, so long as this does not harm the incentives of the users. We consider a cryptographic implementation of a second-price auction in which the miner strategizes on-chain by choosing the reserve price. We show that (while adopting our new desiderata involves a provably-unavoidable trade-off with prior definitions of resistance to user-miner collusion) the Cryptographic Second-Price Auction achieves good incentive properties both on-chain and off-chain under our revised notions.

Prior frameworks on TFM design. Before giving more exposition into our framework, we review past approaches to the study of transaction fee mechanism design. Drawing on classical auction theory, Roughgarden (2021) initiated the study of TFM design and proposed three elegant desiderata that a transaction fee mechanism should satisfy, with a rapidly developing agenda around them (Chung and Shi, 2023; Wu et al., 2024; Chung et al., 2024; Gafni and Yaish, 2024). At a high level, these desiderata are:

---

[1]Following Roughgarden (2021) and the common terminology used in proof-of-work blockchains, we use the language of "miner", even though block creators are now known as "builders" in Ethereum since the switch to proof-of-stake.

[2]Anecdotally, "autobidders" like MetaMask were not particularly good at optimal bidding either, due to volatility in demand.

- **Simple for Users.** Users who know their value v for their transaction being included in a block should optimize their expected utility by bidding v. This is called User Incentive Compatibility (UIC) in Roughgarden (2021), and is equivalent to familiar notions of (Dominant Strategy) Incentive Compatibility.

- **Simple for Miners.** Miners should maximize their revenue by executing the intended auction, rather than deviating. This is called Myopic Miner Incentive Compatibility (MMIC) in Roughgarden (2021) (which also has conceptual connections to the notion of credibility (Akbarpour and Li, 2020) in classical auction settings).

- **Resistant to Collusion.** No miner together with a collection of users should profit by first writing a binding side-contract and then altering their behavior in the auction. Roughgarden (2021) terms an auction Off-Chain Agreement-Proof (OCA-proof) if the miner together with all users cannot jointly profitably deviate, and Chung and Shi (2023) term an auction c-Side Contract Proof (c-SCP) if the miner together with a set of any c users cannot jointly profitably deviate.

These desiderata have proven very useful for describing the properties of different TFMs. However, in our paper, we argue that as previously studied, these desiderata fail to capture important strategic aspects of two notable mechanisms—EIP-1559, and the Cryptographic Second Price Auction.

EIP-1559, and Off-Chain Influence Proofness. We now discuss the first of our two main motivating mechanisms, EIP-1559. The ideal version of EIP-1559 is an unlimited supply posted-price mechanism with an exogenous price, where all revenue is burned. That is, any bid exceeding an exogenously set price p is included and pays p, but those payments are burned and the miner always receives zero revenue.[3] Prior desiderata consider this ideal (i.e., unlimited supply) EIP-1559 as "the dream TFM"—Roughgarden (2021) establishes that it is UIC, MMIC, and OCA-proof. [4]

However, observe that a miner receives zero revenue under the ideal EIP-1559,[5] and MMIC essentially just observes that censoring bids or including fake bids cannot change that (in fact, including fake bids would even harm the miner, as now the miner has to pay the burned bids). But consider the miner behavior discussed above: the miner publicly announces *off-chain* that they intend to censor all bids that don't pay a fee of γ. This transfer could take place off-chain, or in the actual implementation of EIP-1559, the miner could demand this transfer *as part of the included transaction tip* (a part of the EIP-1559 protocol designed to handle periods of high demand). This strategy dominates honestly implementing EIP-1559—the worst that can happen is that all bidders ignore the miner (in which case the miner still gets zero), but if *any* user is willing to pay γ on top of the exogenous posted price, then the miner gets strictly positive utility. This deviation is not in violation of MMIC—it is an *off-chain* deviation, and cannot be implemented simply by censoring transactions or including fake bids (the main types of deviations considered by prior desiderata).[6]

---

[3] Of course, there is not actually unlimited supply of blockspace in Ethereum. EIP-1559 behaves like this in the event that the exogenous price p is sufficiently high so that supply exceeds demand. In the event that demand exceeds supply, EIP-1559 devolves into a first-price auction with reserve p, except that p of each bidder's payment is burned and only the excess 'tip' is transferred to the miner.

[4] And moreover, with limited supply, Chung and Shi (2023) show no mechanism satisfies all three of UIC, MMIC, and 1-SCP.

[5] This has been observed several times, and motivates follow-up work such as Wu et al. (2024) which we discuss in Section 1.1.

[6] Nor is it a violation of OCA-proofness/c-SCP – this miner strategy does not cause the sum of miner and users' on-chain utilities to increase. It just adds a direct transfer from users to miners.

We therefore argue that MMIC is insufficient to claim that a transaction fee mechanism is simple for miners.

Inspired by the above, we propose to add a new desiderata, *off-chain influence-proofness*, which informally states that off-chain behavior of the form above is not profitable, and the miner can always optimize her revenue by engaging directly with the on-chain mechanism (see Section 4 for a formal definition).

Cryptographic Second-Price Auction, and On-Chain Miner-Set Reserves.   We now discuss our second main motivating mechanism, the Cryptographic Second-Price Auction (C2PA). Motivated by the recent push towards heavyweight cryptography in blockchains, we consider a model which abstracts from the actual cryptographic technology used, in which users can submit encrypted bids, and the only role of the miner is again to decide which bids (including fake bids of their own) are given as input to a cryptographic protocol.[7] From the viewpoint of simplicity for users and miners, the C2PA intuitively seems perfect, with the only drawback being the use of heavyweight cryptography. However, it follows from Chung and Shi (2023) that C2PA is, in fact, not MMIC. Specifically, C2PA does not have a built-in reserve, so technically the miner will always want to add a fake bid to act as a reserve.

While correct, we argue that to some extent, this classification seems to be missing the point. The blockchain community initially considered MMIC because we don't trust the miner, and we intuitively conclude that when the TFM is MMIC, we can assume the miner will "actually" implement the auction, and when it is additionally UIC, users can bid their true values without worry. But in C2PA, it is still the case the each user should really go ahead and bid their actual value! In fact, due to our discussion above on off-chain influence proofness, this is perhaps even more true in C2PA than in EIP-1559! We argue therefore that MMIC in prior models may be too restrictive, because (in mechanism design terminology) it requires the miner to implement exactly a prior-free mechanism, and does not give the miner any flexibility to augment the mechanism based on their belief about the distribution of user values, even when this flexibility could never impact the users' incentives.

We instead propose that transaction fee mechanisms should allow for miner augmentation. Specifically, we propose that a transaction fee mechanism should allow for input from the miner (this could take the form of a reserve, but is not restricted to this). But importantly, it should hold that these inputs from the miner do not impact the users' incentives. To accommodate this shift in the model, and to be coherent with our framework more broadly, we rephrase UIC to *on-chain user simplicity*, and MMIC to *on-chain miner simplicity* (see Section 4 for the formal definitions).

A Bayesian Formalization.   EIP-1559 and C2PA are our main motivating examples for revisiting the primitives of transaction fee mechanism design. While doing so, we also propose a formalization that imagines a revenue-maximizing miner in a Bayesian setting,[8] which is more in line with classical auction design (Myerson, 1981) and also some recent work that considers untrusted auctioneers (Akbarpour and Li, 2020; Ferreira and Weinberg, 2020; Essaidi et al., 2022).

---

[7]That is, in our model, the miner must decide which users' bids to forward to the TFM algorithm without *any* information on the actual contents of the bid. Our model and results are thus agnostic to whether Multi-Party Computation (MPC), or Verifiable Delay Functions (VDFs), or some other cryptographic primitives are used in the implementation. Shi et al. (2023) consider a slightly different model termed MPC-assisted Second-Price Auction, in which the miner cannot even censor bids.

[8]Throughout the paper, we restrict attention to the i.i.d. setting, where every user has the same distribution over values. This seems an especially apt assumption for blockchain settings, since users can freely create new identifier if they wish to avoid the types of price discrimination-like behavior that Myersonian optimal auctions require for non-i.i.d. bidders.

Specifically, we propose to consider a miner working within the independent private values model (Myerson, 1981), and we examine the properties of different Bayes-Nash equilibria involving users.

Using our model, we formalize the simplicity desiderata and properties satisfied by EIP-1559 and C2PA discussed above. We also prove a strong impossibility theorem: *no* non-trivial auction can be simultaneously on-chain simple (for both users and miners), off-chain influence proof, and strongly collusion proof (a definition which essentially rephrases 1-SCP in our framework). This impossibility holds *even in the unlimited supply setting*. Remember that EIP-1559 is UIC, MMIC, and c-SCP for all $c \geq 1$ in the unlimited supply setting. Since C2PA abandons strong collusion proofness, but maintains all our other desirable properties, this proves a formal tradeoff: among on-chain simple TFMs, one must choose to adopt an off-chain influence proof auction (such as C2PA) or a strongly collusion proof auction (such as EIP-1559), but cannot satisfy both.

As an aside, we also take the opportunity to make some remarks on different notions of collusion resistance for TFMs. We observe that prior definitions (OCA proofness, 1-SCP, and our strong collusion proofness) mostly seem to capture collusion *between trusted entities who truly profit-share*. These definitions may make sense in some instances (if, for example, Coinbase is a miner and also paying for its own transactions rather than passing on fees to customers). But in other instances, this strong level of collusion may be unrealistic (if, for example, a prominent miner and a prominent dApp try to jointly collude). In such settings, we argue that the miner would likely establish a concrete contract of transfers to redistribute any joint surplus, and moreover, *colluding users would likely act strategically* to maximize their returns while taking this contract into consideration. Thus, following our framework more broadly, the colluding users would play in a Bayes-Nash equilibrium. Since users will still be playing in a Bayes-Nash equilibrium of the off-chain game in the end, this means that off-chain influence proof TFMs *already* guarantee some form of such "Bayesian" collusion-resistance (with some caveats; see Section 4). We thus argue that such more permissive collusion-resistance definitions are best suited to capture an untrusted miner colluding with a separate user. Still, as discussed above, strong collusion proofness better captures situations where miners genuinely have their own transactions, so we believe both criteria are worth considering.

Results and Paper Structure.   After giving preliminaries on (Bayesian) mechanism design in Section 2, our paper proceeds as follows:

- Section 3 introduces our model of TFMs, allowing for both on-chain manipulations as in previous papers, and allowing for the miner to conduct arbitrary off-chain auctions to determine behavior on-chain.

- Section 4 gives our proposed simplicity and incentive-compatibility definitions for TFMs, and provides discussions and comparisons to pre-existing definitions.

- Section 5 analyzes several commonly-studied auctions in the transaction fee mechanism space, (namely, EIP-1559, C2PA, and first-price auction) and classifies these auctions by our desiderata.

- Section 6 proves our impossibility result: no nontrivial mechanism simultaneously satisfies on-chain (user and miner) simplicity, off-chain influence proofness, and strong collusion proofness.

1.1   Related work

Transaction Fee Mechanism Design.   Lavi et al. (2019) are the first to formally view Bitcoin's

transaction fees as an auction, and propose alternatives (such as the kth price auction, or the monopolistic-price auction), and similar questions are also considered in Basu et al. (2019) and Yao (2018). Buterin et al. (2019) propose EIP-1559 for Ethereum, the first significant change to the transaction fee mechanism of a major blockchain, which is analyzed rigorously in Roughgarden (2020). Roughgarden (2020, 2021) kicks off the more recent study of transaction fee mechanisms by proposing the desiderata mentioned above and justifying Ethereum's EIP-1559 through this lens. In particular, Roughgarden (2021) establishes that without a supply constraint (or when EIP-1559's exogenous price is high enough so that supply exceeds demand), EIP-1559 is UIC, MMIC, and OCA-proof. Chung and Shi (2023) prove a strong impossibility result, that no auction is simultaneously UIC and c-SCP with limited supply (for any c).

Further follow-up work of Gafni and Yaish (2022, 2024); Chung et al. (2024) investigate the distinction between OCA-proof and c-SCP, while Gafni and Yaish (2022) and Chen et al. (2024) also consider a relaxation of UIC to Bayesian-UIC (without miner-set advice as in our paper). Shi et al. (2023) are the first to introduce cryptography into transaction fee mechanism design—they design non-trivial TFMs assuming Multi-Party Computation, but also observe that the C2PA is not MMIC. Wu et al. (2024) design mechanisms (also leveraging cryptography) that achieve non-zero revenue in a Bayesian setting, under the assumption that some users are honest.

While some of these works indeed tweak the core TFM model of Roughgarden (2021), they do so in directions that are orthogonal to our revised axioms. For example, Gafni and Yaish (2022) propose to consider Bayesian-IC instead of DSIC for users. Chung and Shi (2023) propose a model where it is costly for the miner to include fake bids—this approach could be combined with our revised definitions (although we do not investigate this particular avenue ourselves). To our knowledge, our work is the first proposal to significantly revise the core initial TFM desiderata, which we believe will provide additional insight to future works alongside this agenda (and in a way that is backwards-compatible with existing tweaks to the vanilla model of Roughgarden (2021)).

Variants of Transaction Fee Mechanism Design. In line with all previously-cited works, we consider a one-shot Transaction Fee Mechanism played by users who want their transactions to be included in the present block and miners who are mypoic (i.e. because they are a small player in a decentralized ecosystem). While TFMs "in the wild" are indeed a repeated game (both because users may be content to have their transaction in a later block, and becuase miners of this block may also be miners of the next block), this one-shot setting is the core upon which later research lies, and it is therefore crucial to nail the core setting. Some recent works have explored the repeated game aspect. For example, Ferreira et al. (2021); Leonardos et al. (2021, 2023) study the dynamics of how EIP-1559's exogenous price is updated over time, but do not study further strategic aspects that appear in the one-shot game. Gafni and Yaish (2022) consider a particular Bayesian model of user values arriving over time and propose a simple BIC and MMIC mechanism (essentially the revelation principle applied to a first-price auction). Our work has little overlap with these, as we prioritize a deeper understanding of the core single-block setting.

Since EIP-1559, an orthogonal paradigm shift has taken over transaction fee mechanism design in practice: Proposer-Builder Separation. Specifically, while some transactions are still simple peer-to-peer payments that reward the miner with a direct payment of (say) ETH, many transactions are significantly more complicated and allow a miner to extract profit through so-called Maximal Extractable Value (MEV) (Daian et al., 2020). For example, one user's transaction might be a trade of two cryptoassets on a DeFi Exchange, and the miner might profit by inserting 'fake' transactions to sandwich them – their profit comes *not at all by increased revenue in the TFM itself*, but *through external opportunities not at all captured by the TFM* (indeed, this is profitable

even in EIP-1559 where the miner always receives zero direct revenue). Bahrani et al. (2023) are the first to investigate the impact of MEV on Transaction Fee Mechanism design, and largely provide impossibility results. This work is also orthogonal to ours, as we aim for a deeper understanding of the core TFM regime.

Broader Literatures Of course, blockchains relate to a wide variety of economic phenomena. The rapidly-evolving economics literature on the economics of blockchain is far too vast to summarize here, but includes Huberman et al. (2017); Arnosti and Weinberg (2022); Budish (2023); Gans and Halaburda (2024). Our work also relies on the large literature on Bayesian mechanism design following Myerson (1981). Our framework is also conceptually related to the long line of work in the cryptography literature studying mechanism design—including Halpern and Teague (2004); Abraham et al. (2006); Kol and Naor (2008); Izmalkov et al. (2011); Garay et al. (2013); Canetti et al. (2023), among many others—though we do not work directly with precise cryptography notions relying on complexity theory, and instead consider a simple idealized model of cryptographic auctions.

2 Preliminaries

Before we introduce our model of a transaction fee mechanism (Section 3), we review the basics of (Bayesian) mechanism design for the reader not familiar with these concepts.

In mechanism design terminology, we consider single-parameter quasi-linear Bayesian environments. In these environments, users' values for being allocated to a binary outcome (e.g., receiving an item or not, being included in a block or not) are drawn from some publicly-known distribution, and some feasible set of users are allocated and each charged some price.

Formally, we define an *environment* $(n, \mathcal{X}, \mathcal{T}_1 \times \cdots \times \mathcal{T}_n)$ to consist of a number n of users $\{1, 2, \ldots, n\}$ (variously called bidders, players, or agents in the literature), a set $\mathcal{X} \subseteq 2^{\{1, \ldots, n\}}$ of feasible allocations (where $X \in \mathcal{X}$ indicates that each user $i \in X$ is allocated, and each $i \notin X$ is not), and independent type distributions $\mathcal{T}_1 \times \cdots \times \mathcal{T}_n$, which are arbitrary continuous distributions over $\mathbb{R}_{\geq 0}$. User i's value from being allocated is denoted by v_i, and is drawn from the distribution $\mathcal{T}_i$, denoted $v_i \sim \mathcal{T}_i$. We typically focus on the i.i.d. case, in which $\mathcal{T}_i = \mathcal{T}$ is identical for each $i = 1, \ldots, n$.

The set of *outcomes* $\mathcal{Y} = \mathcal{X} \times \mathbb{R}^n$ in an environment consists of tuples $(X, P_1, \ldots, P_n)$, where $X \in \mathcal{X}$ is the set of allocated users, and P_i is the payment charged to each user i. Given an outcome $(X, P_1, \ldots, P_n)$, each user i such that $i \in X$ receives utility $v_i - P_i$, and every non-allocated bidder receives a utility $-P_i$. Users receive a quasi-linear utility, i.e, given a distribution $\mathcal{D}$ over outcomes $(X, P_1, \ldots, P_n) \in \mathcal{Y}$, user i with value v_i receives an expected utility $\mathbb{E}_{(X, P_1, \ldots, P_n) \sim \mathcal{D}} [v_i \cdot \mathbb{1} [i \in X] - P_i]$, where $\mathbb{1} [\cdot]$ denotes the indicator function. An *auction* (or a *mechanism*) over some environment is any mapping $\mathcal{M} : \mathcal{T}^n \to \Delta(\mathcal{Y})$ from bids of each user to a distribution over outcomes.

2.1 Dominant Strategy Incentive Compatible Auctions

Consider a mechanism $\mathcal{M}$. For a profile $v = (v_i)_{1 \leq i \leq n}$, let $X_i(v)$ denote the probability that user i gets allocated and $P_i(v)$ be the expected payment charged from i (where expectation is taken over the randomness in the mechanism). Since users are assumed to be expected utility maximizers, conditioned on the bids b_{-i} placed by the other users, user i bids b_i such that his utility $v_i \cdot X_i(b_i, b_{-i}) - P_i(b_i, b_{-i})$ is maximized. The mechanism $\mathcal{M}$ is *dominant strategy incentive*

compatible (DSIC) if bidding truthfully, i.e, bidding v_i, is the optimal strategy for each user i irrespective of the bids b_{-i} placed by the other bidders. Formally, $\mathcal{M}$ is DSIC if for all users i and for all profiles $v = (v_1, \ldots, v_n)$ of values, and all alternative bids $\widetilde{v}_i$, we have

$$v_i \cdot X_i(v_i, v_{-i}) - P_i(v_i, v_{-i}) \geq v_i \cdot X_i(\widetilde{v}_i, v_{-i}) - P_i(\widetilde{v}_i, v_{-i}).$$

A standard example of a DSIC mechanism is the $(k+1)^{\text{th}}$ price auction with a reserve r. To sell k items, the auctioneer collects bids from users and allocates the user with the highest k bids (as long as they are larger than the reserve r) and charges the maximum among the kth highest bid and r. We skip the proof of incentive compatibility here, since we will be revisiting the $(k+1)^{\text{th}}$ price auction later in the paper (Section 5.2).

2.2   Bayesian Incentive Compatible Auctions

Dominant strategy incentive compatibility enforces a strong condition— irrespective of the bids b_{-i} placed by the other users, user i best responds by bidding his value v_i. However, if user i knows the distribution $\mathcal{T}_{-i} = \Pi_{j \neq i} \mathcal{T}_j$ of values of other users and believes other users are going to bid truthfully, user i might want to bid v_i if it optimizes his utility in expectation over the values v_{-i} of the other users.

In accordance to the discussion above, we define the interim allocation and payment rules of the mechanism $\mathcal{M}$.

Definition 2.1 (Interim Allocation and Payment Rules)**.** For a mechanism $\mathcal{M}$ with an allocation and payment rule X and P respectively, and distributions $\Pi_{i=1}^n \mathcal{T}$ of user values, the interim allocation rule $x_i : \mathcal{T}_i \to [0,1]$ and the interim payment rule $p_i : \mathcal{T}_i \to \mathbb{R}$ are given as follows:

$$x_i(v_i) = \mathop{\mathbb{E}}_{v_{-i} \sim \mathcal{T}_{-i}} [X_i(v_i, v_{-i})] \qquad \text{and} \qquad p_i(v_i) = \mathop{\mathbb{E}}_{v_{-i} \sim \mathcal{T}_{-i}} [P_i(v_i, v_{-i})].$$

Intuitively, the interim allocation and payment rules describe the user's expected allocation and payment assuming all other users bid truthfully when their values are drawn from $\mathcal{T}_{-i}$. Writing user i's expected utility in terms of the interim rules is extremely convenient— when he has a value v_i and submits a bid b_i, his expected utility equals $v_i \cdot x_i(b_i) - p_i(b_i)$.[9]

An auction $\mathcal{M}$ is *Bayesian incentive compatible* (BIC) if bidding truthfully optimizes the expected utility for every user, taking expectation over the values of all other users (motivating the assumption that x_i and p_i are defined in expectation over the other user's values and not other arbitrary bids). In other words, $\mathcal{M}$ is BIC if for all $v_i, \widetilde{v}_i \in \mathbb{R}_{\geq 0}$,

$$v_i \cdot x_i(v_i) - p_i(v_i) \geq v_i \cdot x_i(\widetilde{v}_i) - p_i(\widetilde{v}_i).$$

Observe that DSIC is a stronger condition than BIC. BIC requires incentive compatibility in expectation while DSIC requires incentive compatibility to hold pointwise over the other users' values.

The Myerson's lemma (a.k.a the payment identity; Myerson, 1981) characterizes all mechanism $\mathcal{M}$ with allocation and payment rules X and P such that $\mathcal{M}$ is BIC. It requires no user i should expect to get allocated with a lower probability when placing a larger bid, i.e, the interim allocation rule x_i is monotonically non-decreasing in the user's value v_i. Further, given a set of interim allocation rules $(x_i)_{1 \leq i \leq n}$, there exist a unique (up to additive constants) set of interim payment

---

[9]Note that the term "interim" refers to the fact that from user i's perspective, his value v_i is already known, but other users' valuations are not. For clarity, throughout this paper we use lower-case letters such as $x_i(v_i)$ to denote such interim rules, and use upper-case letters such as $X_i(v) = X_i(v_1, \ldots, v_n)$ to denote the "ex-post" outcomes conditioned on the bids of all users.

rules $(p_i)_{1 \le i \le n}$ such that the resulting mechanism is BIC.[10]

Theorem 2.2 (Myerson, 1981; Payment Identity)**.** *A mechanism $\mathcal{C}$ with interim allocation rules $(x_i)_{1 \le i \le n}$ and interim payment rules $(p_i)_{1 \le i \le n}$ is BIC if and only if for all users i,*

1. *the interim allocation rule $x_i(v_i)$ is monotonically non-decreasing in v_i, and*

2. *the interim payment rule $p_i(v_i)$ satisfies $p_i(v_i) = \int_o^{v_i} z x_i'(z) \, dz + c$ for some constant c.*

2.3 Bayes-Nash Equilibria and the Direct Revelation Principle

We also consider auctions that are neither DSIC nor BIC. Even though utility maximizing users will not bid truthfully in such mechanisms, users will play strategies that are (simultaneously) best responses to each other's strategies (in expectation). Such a profile of simultaneous best responses is called a *Bayes-Nash equilibrium* (BNE).

More formally, consider a mechanism $\mathcal{M}$ that takes as inputs messages from an arbitrary message space $\mathsf{Bid}_1, \ldots, \mathsf{Bid}_n$ with an allocation rule $X : \mathsf{Bid}_1 \times \cdots \times \mathsf{Bid}_n \to [0,1]^n$ and a payment rule $P : \mathsf{Bid}_1 \times \cdots \times \mathsf{Bid}_n \to \mathbb{R}^n$. A user i's strategy $s_i : \mathbb{R} \to \mathsf{Bid}_i$ maps the value v_i of user i to i's bid (input in Bid_i) to the mechanism $\mathcal{M}$. A profile of strategies $(s_1, \ldots, s_n)$ is a Bayes-Nash equilibrium if for all users i, conditioned on all other users playing the strategies s_{-i}, user i with value v_i optimizes his expected utility by bidding $s_i(v_i)$. In other words, for all users i and for all values v_i

$$\mathbb{E}_{v_{-i} \sim \mathcal{T}_{-i}} \left[v_i \cdot X_i(s_i(v_i), s_{-i}(v_{-i})) - P_i(s_i(v_i), s_{-i}(v_{-i})) \right]$$

$$\ge \mathbb{E}_{v_{-i} \sim \mathcal{T}_{-i}} \left[v_i \cdot X_i(b_i, s_{-i}(v_{-i})) - P_i(b_i, s_{-i}(v_{-i})) \right]$$

for all $b_i \in \mathsf{Bid}_i$.

For any BNE of any mechanism $\mathcal{M}$, a corresponding interim allocation and interim payment rule can be defined as follows:

Definition 2.3 (Interim Allocation and Interim Payment Rules For a BNE)**.** For a BNE of user strategies $\sigma = (s_1, \ldots, s_n)$ induced by a mechanism $\mathcal{M}$ with allocation and payment rules X and P respectively, and distributions $\Pi_{i=1}^n \mathcal{T}$ of user values, the *interim allocation rule* $x_i : \mathcal{T}_i \to [0,1]$ and the *interim payment rule* $p_i : \mathcal{T}_i \to \mathbb{R}$ corresponding to σ are given as follows:

$$x_i(v_i) = \mathbb{E}_{v_{-i} \sim \mathcal{T}_{-i}} \left[X_i(s_i(v_i), s_{-i}(v_{-i})) \right] \qquad \text{and} \qquad p_i(v_i) = \mathbb{E}_{v_{-i} \sim \mathcal{T}_{-i}} \left[P_i(s_i(v_i), s_{-i}(v_{-i})) \right].$$

We now describe the classical revelation principle, which says that any BNE of any mechanism induces a BIC mechanism which implements the same interim allocation and payment rules. Intuitively, consider an alternative mechanism $\widetilde{\mathcal{M}}$ that implements $\mathcal{M}$ supplemented with a "strategizing service". Instead of having to devise the optimal strategy s_i, user i can instead submit a bid equal to his value v_i. The strategizing service bids $s_i(v_i)$ on user i's behalf. Assuming other users bid truthfully in $\widetilde{\mathcal{M}}$, user i bidding v_i in $\widetilde{\mathcal{M}}$ corresponds to the strategy profile $(s_i(v_i), s_{-i}(v_{-i}))$ in $\mathcal{M}$. Since $(s_1, \ldots, s_n)$ is a BNE of $\mathcal{M}$, we see that v_i is user i's expected-utility-maximizing strategy in $\widetilde{\mathcal{M}}$; since this is true for all users, this means $\widetilde{\mathcal{M}}$ is BIC. The *direct revelation principle* follows directly from this argument:

---

[10]Note however that not every set of monotone interim allocation rules will correspond to a feasible auction which always produces feasible outcomes in $\mathcal{X}$.

Lemma 2.4 (Direct Revelation Principle). *Fix a mechanism $\mathcal{M}$ and a Bayes-Nash equilibria $\sigma = (s_1, \ldots, s_n)$ of user strategies. Then, there exists a BIC mechanism $\widetilde{\mathcal{M}}$ such that the interim allocation and payment rules corresponding to σ in $\mathcal{M}$ both equal the interim allocation and payment rules of the BIC mechanism $\widetilde{\mathcal{M}}$.*

In particular, the payment identity (Theorem 2.2) can be applied to any BNE of $\mathcal{M}$ via the interim payment rules. The mechanism $\widetilde{\mathcal{M}}$ is BIC, and therefore, the interim payment rule p_i is uniquely determined by the interim allocation rule x_i. We get the following direct consequence of Theorem 2.2:

Lemma 2.5 (Revenue Equivalence). *For any two BNEs $s^{(1)}, s^{(2)}$ of any two mechanisms, if the interim allocation rules $(x_i)_{1 \le i \le n}$ induced by the BNEs $s^{(1)}$ and $s^{(2)}$ are the same, then the interim payment rules $(p_i)_{1 \le i \le n}$ induced by $s^{(1)}$ and $s^{(2)}$ are also the same. Further, the expected total payment made by users from all such BNEs equal $\sum_{i=1}^{n} \mathbb{E}_{v_i \sim \mathcal{T}_i}[p_i(v_i)]$.*

2.4 Revenue Optimization and Virtual Values

In classical auctions, the auctioneer collects the payments made by the users and hence, an auctioneer optimizing her revenue would optimize the total payments received from users. Working with payment rules directly while trying to optimize revenue can be mathematically challenging, for instance, since one must ensure that the payment rule of the revenue optimal mechanism simultaneously satisfies the payment identity (Theorem 2.2). Instead, Myerson (1981) provides a recipe to compute the expected payments via the expected sum of *virtual values* (a.k.a *virtual welfare*), described below. Thus, optimizing revenue is equivalent to optimizing the total virtual welfare, which is mathematically similar to optimizing the total welfare (i.e, optimizing the total value).

Definition 2.6 (Virtual Values). For a (continuous) type distribution $\mathcal{T}$ with cumulative distribution function $F_{\mathcal{T}}(t) = Pr_{v \sim \mathcal{T}}(v \le t)$ and probability density function $f_{\mathcal{T}}(t) = \frac{dF_{\mathcal{T}}(t)}{dt}$, define the *virtual value* function $\phi(\cdot)$ as follows:

$$\phi_i(v_i) = v_i - \frac{1 - F_{\mathcal{T}}(v_i)}{f_{\mathcal{T}}(v_i)}.$$

Theorem 2.7 (Myerson, 1981; Revenue Equals Virtual Welfare). *Given distributions $\Pi_{i=1}^{n}\mathcal{T}_i$ of user values with virtual values $(\phi_i)_{1 \le i \le n}$, for any BIC mechanism with interim allocation rules $(x_i)_{1 \le i \le n}$ and interim payments $(p_i)_{1 \le i \le n}$ (or equivalently, any BNE with value interim allocation and payment rules $(x_i)_{1 \le i \le n}$ and $(p_i)_{1 \le i \le n}$), we have:*

$$\mathbb{E}_{v \sim \mathcal{T}^n}\left[\sum_{i=1}^{n} p_i(v_i)\right] = \mathbb{E}_{v \sim \mathcal{T}^n}\left[\sum_{i=1}^{n} \phi_i(v_i) \cdot x_i(v_i)\right]. \tag{*}$$

For a simple example, consider the posted-price mechanism with a single user and a single good. In this auction, the user is posted a price p, and is sold the good at the price p when the user has a value $v_i \ge p$. Observing that the posted price mechanism is DSIC is straightforward. Suppose the user's value v is drawn from the uniform distribution $U[0, 1]$ (with a cumulative distribution function $F(v) = v$ and a density function $f(v) = 1$ for all $0 \le v \le 1$). By posting a price p, the good is sold with probability $1 - F(p) = 1 - p$, fetching an expected revenue equal to $p(1 - p)$. The user's virtual value $\phi(v) = v - \frac{1 - F(v)}{f(v)} = v - \frac{(1-v)}{1} = 2v - 1$. The user is allocated ($x(v) = 1$) whenever $v \ge p$ and not allocated ($x(v) = 0$). Plugging the above into the right hand side of (*),

we get the expected virtual welfare equals $\int_p^1 (2v - 1)\, dv = p\,(1 - p)$.

We now discuss revenue-optimal auctions. Theorem 2.7 says that optimizing revenue corresponds to optimizing the virtual welfare. However, for the mechanism (or equivalently, the value interim allocation rules) to be BIC, the interim allocation rules must be monotonically non-decreasing. For general distributions $\mathcal{T}_i$ of user i's value, the virtual value function $\phi_i(v_i)$ need not be monotone non-decreasing, and therefore optimizing for virtual welfare does not automatically guarantee that interim allocation rules are monotone. The revenue optimal mechanism involving such distributions can be fairly complex. However, a standard simplifying assumption is the notion of regularity, which states that the virtual value function is monotone non-decreasing.

For regular distributions, it follows that optimizing for virtual welfare also ensures that the interim allocation rules all satisfy monotonicity. Moreover, for a given distribution $\mathcal{T}$, if there exists a value r such that $\phi(r) = 0$, then for all values $v \leq r$, the virtual value is non-positive (by monotonicity). Thus, when the feasibility constraints are downwards-closed (i.e., when a user can always be removed from a feasible allocation and this results in another feasible allocation), we can thus simply exclude any user with a negative virtual value. This must increase the expected virtual welfare, and thus the expected revenue. We call r to be the *monopoly reserve* for the regular distribution $\mathcal{T}$.

Definition 2.8 (Regular Distributions and Monopoly Reserve). A distribution $\mathcal{T}$ is *regular* if the virtual value function $\phi(v)$ is monotonically non-decreasing in v. A *monopoly reserve* of the regular distribution $\mathcal{T}$ is given by any value r such that $\phi(r) = 0$.[11]

Theorem 2.9 (Myerson, 1981). *Let the set of feasible allocations $\mathcal{X}$ consist of any subset of users of cardinality at most k. Let users' values be drawn i.i.d. from from a regular distribution $\mathcal{T}$ for any number of users n. Then, the revenue optimal BIC mechanism is the $(k+1)$st price auction with monopoly reserve r.*

3 Model of Transaction Fee Mechanisms

3.1 Formal Model

We now define our model of transaction fee mechanisms. We model TFMs in three layers, starting from the primitives of the protocol, and adding increasing levels of strategic flexibility. The innermost layer is the "block-building process" $\mathcal{B}$, which is an algorithm that dictates how all new blocks must be created, and is fixed by the blockchain designer (e.g., the Ethereum community). The second layer, the "on-chain game" $\mathcal{C}$, represents all strategic manipulations available on-chain to the miner and users based only on manipulating the inputs passed to $\mathcal{B}$. Previously-considered notions typically work in terms of the on-chain game.

The third stage of our model, the "off-chain game" $\mathcal{D}$, allows the miner to conduct a separate off-chain auction to determine the inputs to $\mathcal{C}$ (and hence to $\mathcal{B}$). This substantial additional ability granted to the miner is our primary departure point from previous works.

To begin, we define the set of possible outcomes of a TFM. Given a set of bids indexed by $[N]$, a (deterministic) *outcome* of a TFM (given bids indexed by $[N]$) is a triple $\big(X, (P_i)_{i \in [N]}, \mathsf{Rev}\big)$, where $X \subseteq [N]$ specifies a binary outcome for each bid (i.e., a $X \subseteq [N]$ where bids indexed by $i \in X$ are

---

[11]For simplicity, we assume the existence of an r such that $\phi(r) = 0$; any such r can serve as the monopoly reserve. However, if there does not exist such a value r, the monopoly reserve can be defined as $r = \sup\{v | \phi(v) \leq 0\}$. With slight modification, all definitions and results discussed in the paper continue to hold under this definition.

included in the block), a payment P_i for each bid, and a revenue Rev transferred to the miner.[12] Denote the set of all outcomes by $\mathcal{X}$.

We now define our main model. A running example of the definitions below is provided in Section 3.2.

Definition 3.1 (Transaction Fee Mechanisms). A transaction fee mechanism (TFM) is specified by a block-building process $\mathcal{B}$. Strategic play can proceed according to on-chain game $\mathcal{C}$ and off-chain game $\mathcal{D}$. These are defined as follows:

1. The *Block-Building Process* is a function $\mathcal{B}(a_{\mathsf{mi}}, b_1, \ldots, b_N)$ from a miner advice a_{mi} and bids $(b_1, \ldots, b_N)$ to distributions over outcomes $(X, (P_i)_{i \in [N]}, r) \in \mathcal{X}$. In more detail, this is any algorithm that takes any number of bids b_i from a pre-defined bid space $\mathsf{Bid}_{\mathsf{usr}}$,[13] and some advice a_{mi} from an advice space $\mathsf{Adv}_{\mathsf{mi}}$ provided by the miner. We restrict attention to simultaneous, single-round mechanisms. The block building process returns a distribution over outcomes $\mathcal{X}$.

   The block-building process also specifies a cryptographic model, which determines the level of cryptography in the communication between users and the miner in $\mathcal{B}$. We primarily consider two abstract cryptographic models: the "plaintext" model, where the miner will be able to see the bids before submitting to B, and the "miner-gatekeeper" model, where the bids are encrypted by users before sending them to the miner.[14] The effects of the cryptographic model is captured in the miner's strategy space defined in the on-chain game, as we describe below.

2. The block-building process $\mathcal{B}$ induces an *On-Chain Game* $\mathcal{C}$ involving the miner and the users. In $\mathcal{C}$, users' strategy space $S_{\mathsf{usr}}^{\mathcal{C}}$ is submitting zero or more bids. Formally, each user i's strategy $s_{\mathsf{usr},i}^{\mathcal{C}}$ is a mapping from the user's type (i.e., willingness-to-pay) to a set of bids to be submitted to $\mathcal{B}$.[15]

   The miner's strategy space $S_{\mathsf{mi}}^{\mathcal{C}}$ is different under the two levels of cryptography. In the plaintext model, the miner can arbitrarily condition their behavior on the submitted bids. Formally, in the plaintext model, the miner's strategy $s_{\mathsf{mi}}^{\mathcal{C}} \in S_{\mathsf{mi}}^{\mathcal{C}}$ is a mapping from the set H of all submitted bids to an advice $a_{\mathsf{mi}} \in \mathsf{Adv}_{\mathsf{mi}}$, a subset $I \subseteq H$ of submitted bids, and another set of "fabricated" bids J. The outcome of $\mathcal{C}$ is then the block built according to $\mathcal{B}(a_{\mathsf{mi}}; I, J)$.

   In the miner-gatekeeper model, the miner can only condition their behavior on *who* submits a bid, not the contents of the bid. Formally, in the miner-gatekeeper model, the miner's strategy $s_{\mathsf{mi}}^{\mathcal{C}} \in S_{\mathsf{mi}}^{\mathcal{C}}$ is a mapping from the set H of all *bidders i* who submit a bid, to an advice $a_{\mathsf{mi}} \in \mathsf{Adv}_{\mathsf{mi}}$, a subset of submitted bidders $I \subseteq H$, and another set of fabricated bids J. The outcome of $\mathcal{C}$, analogous to the plaintext model, is then the block built according to

---

[12]Note that we allow miner revenue to be less or greater than the sum of the transaction fees (money creation, e.g. from a constant block reward, or money burning).

[13]We typically consider the bid space $\mathsf{Bid}_{\mathsf{usr}}$ to be $\mathbb{R}_{\geq 0}$ (like in all sealed bid auctions). Other natural bid spaces include {yes, no} in a posted-price mechanism (signifying the answer to the question "are you willing to buy the good at the given price?"), or {yes, no} $\times \mathbb{R}_{\geq 0}$ in EIP-1559 (with the real number signifying the tip).

[14]Technically, the input to the miner-gatekeeper model are encrypted commitments to bids. However, for simplicity, we assume that the input to $\mathcal{B}$ is just the set of bids, and model the fact that the miner is oblivious to the exact contents of the bids by restricting the miner's strategy space.

[15]Note that, as in Section 2, we use the term "strategies" for mappings from users' types to their actions / messages in games.

11

$\mathcal{B}(a_{\mathsf{mi}}; (b_i)_{i \in I}, J)$.[16,17]

Given a miner strategy $s_{\mathsf{mi}}^{\mathcal{C}}$ and a profile of user strategies $s_{\mathsf{usr}}^{\mathcal{C}} = \left(s_{\mathsf{usr},i}^{\mathcal{C}}\right)_i$ and values $v = (v_i)_i$, we let $\mathcal{C}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v))$ denote the result of the on-chain game when the miner plays according to $s_{\mathsf{mi}}^{\mathcal{C}}$ and users play according to $s_{\mathsf{usr}}^{\mathcal{C}}(v) = \left(s_{\mathsf{usr},i}^{\mathcal{C}}(v_i)\right)_i$. Let $X_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v))$ denote the probability that user i is allocated and let $P_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}})$ denote i's expected transaction fee;[18] bidder i's total expected utility is then $v_i \cdot X_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v)) - P_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v))$. Let $\mathsf{Rev}^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v))$ denote the miner's expected reward returned by $\mathcal{B}$, minus the expected payments made by the fabricated bids injected into $\mathcal{B}$ by the miner (where the expectation is taken over the randomness in the mechanism); this is the miner's total utility from $\mathcal{C}$.

3. The *Off-Chain Game* $\mathcal{D}$ induced by $\mathcal{B}$ is a game defined as follows. To begin, the miner announces and commits to an "off-chain mechanism" $\mathcal{M}_{\mathsf{off}}$ with message space $\mathcal{U}_{\mathsf{off}}$. This off-chain mechanism allows the miner to run any mechanism of her choosing to determine how she and the users will behave in $\mathcal{C}$. Formally, $\mathcal{M}_{\mathsf{off}}$ is a function from any number N' of reports in $\mathcal{U}_{\mathsf{off}}$, to the miner's and users' behavior $s_{\mathsf{mi}}^{\mathcal{C}}$ and $s_{\mathsf{usr}}^{\mathcal{C}} = (s_{\mathsf{usr},i}^{\mathcal{C}})_{i \in \{1,\ldots,N'\}}$ in $\mathcal{C}$, as well as off-chain payments $p_{\mathsf{off}} = (p_{\mathsf{off},i})_{i \in \{1,\ldots,N'\}}$.

The only assumption we make about $\mathcal{M}_{\mathsf{off}}$ is that each user always has the option of not participating in the off-chain mechanism and instead submits bids directly to $\mathcal{C}$. Formally, we assume that in $\mathcal{M}_{\mathsf{off}}$, for all users i and every strategy $s_{\mathsf{usr},i}^{\mathcal{C}}$ in $\mathcal{C}$, there exists at least one message $\perp_{\mathsf{off}}(s_{\mathsf{usr},i}^{\mathcal{C}}) \in \mathcal{U}_{\mathsf{off}}$ (called an abstaining bid) such that

(a) whenever user i submits $\perp_{\mathsf{off}}(s_{\mathsf{usr},i}^{\mathcal{C}})$, the result of $\mathcal{M}_{\mathsf{off}}$ specifies that i will play $s_{\mathsf{usr},i}^{\mathcal{C}}$ in $\mathcal{C}$ and pay $p_{\mathsf{off},i} = 0$, and,

(b) the result of $\mathcal{M}_{\mathsf{off}}$ for all users j that submit non-abstaining bids to the off-chain mechanism is invariant to the set of abstaining bids.

Property (b) signifies that whenever a user i refuses to participate in the off-chain mechanism, the miner and other users lack any information about i, including its existence.

After the miner commits to an off-chain mechanism $\mathcal{M}_{\mathsf{off}}$, users' strategies consist of submitting some message to $\mathcal{M}_{\mathsf{off}}$. The on-chain game proceeds according to the result of $\mathcal{M}_{\mathsf{off}}$, and utilities are given by the sum of utilities from the on-chain game and the off-chain payments from $\mathcal{M}_{\mathsf{off}}$. Formally, a strategy $s_{\mathsf{usr},i}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}} \in S_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}$ of user i is any mapping from user i's type to some set of messages to $\mathcal{M}_{\mathsf{off}}$.[19] We let $\mathcal{D}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))$ denote the result of running the on-chain game $\mathcal{C}$ according to $\mathcal{M}_{\mathsf{off}}(s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))$, along with off-chain payments. In particular, denote by $X_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)) = X_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}})$, where $s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr}}^{\mathcal{C}}$ are set by $\mathcal{M}_{\mathsf{off}}(s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))$.

---

[16]Note in particular that in both models, the bidder's index i serves as an "identifier" for the particular bid. However, we do not restrict users to a single identifier, so a bidder i can submit a fake bid with identifier i' for some i' different than any other bidders' (though we typically leave this implicit in the notation for simplicity).

[17]We are agnostic to the cryptographic machinery used by the TFM. While Shi and Chung (2023) recommend the use of multi-party computation, fully homomorphic encryption, digital signature schemes, or verifiable random functions are also potential cryptography tools that can be used by the TFM.

[18]If user i submits more than one bid, we let $X_i(\cdot)$ denote the probability that at least one of these bids is allocated in $\mathcal{B}$, and we let $P_i(\cdot)$ denote the expected sum of the payments charged to these bids. This reflects our assumption that each user has just one bid they want to submit, or equivalently, that the users' total value is additive over the different transactions they want to submit.

[19]As in the case of $\mathcal{C}$, we allow users to submit multiple messages $\mathcal{M}_{\mathsf{off}}$; in this case, the result of $\mathcal{M}_{\mathsf{off}}$ will specify separate strategies and charge separate payments in $\mathcal{C}$ for the different messages. As for $\mathcal{C}$, we typically leave this implicit in the notation for simplicity.
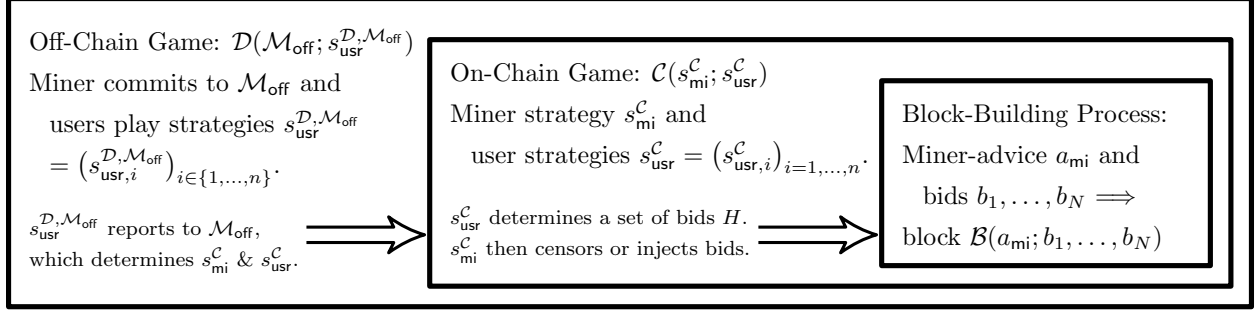
Figure 1: Model of TFMs.

Denote by $P_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)) = P_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}) + p_{\mathsf{off},i}$, where $p_{\mathsf{off},i}$ is additionally set by $\mathcal{M}_{\mathsf{off}}(s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))$. User i's total expected utility is then

$$v_i \cdot X_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)) - P_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)).$$

Finally, let $\mathsf{Rev}^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)) = \mathsf{Rev}^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v)) + \sum_i p_{\mathsf{off},i}$ denote the miner's expected reward from $\mathcal{C}$ plus the sum of the off-chain payments.

We say that the miner runs a *trivial off-chain mechanism* if $\mathcal{M}_{\mathsf{off}}$ has a message space $\mathcal{U}_{\mathsf{off}}$ which consists of only abstaining bids (and hence the miner's on-chain strategy specified by the result of $\mathcal{M}_{\mathsf{off}}$ must be a constant). This corresponds to the miner not running an off-chain mechanism and instead playing a fixed strictly on-chain strategy. We will sometimes abuse notation to describe an on-chain profile of strategies $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr}}^{\mathcal{C}})$ as a strategy profile in the off-chain game, which corresponds to the trivial mechanism $\mathcal{M}_{\mathsf{off}}$ which always specifies that the miner plays the fixed strategy $s_{\mathsf{mi}}^{\mathcal{C}}$, and users play strategies $\perp_{\mathsf{off}}(s_{\mathsf{usr}}^{\mathcal{C}})$.

See Figure 1 for a high-level illustration of our model. As we will see in Section 4, previously-considered notions such as UIC and MMIC are naturally expressed in in the language of the game $\mathcal{C}$. In contrast, modeling the miner attack on EIP-1559 motivating our paper (as discussed in Section 1) requires the miner manipulating the off-chain game $\mathcal{D}$.

3.2 Example: The Squared Revenue Second Price Auction (SR2PA)

We run over the above notations with an example of a novel auction format under which the miner can benefit by using a completely separate off-chain auction to determine her behavior on-chain. We define the Squared Revenue Second Price Auction (SR2PA) to auction off inclusion in a block of size 1 (a single item).

Informally, the SR2PA implements the second-price auction in the plaintext model, while allowing the miner to set a reserve, and burning some function of the on-chain payments. In particular, we will consider cases where users have values less than 1 and will always bid $b_i \le 1$, and the miner will receive only the *square* of the payment p_i made by the highest bidder (so that $p_i - p_i^2 \ge 0$ is burned). Formally, the block-building process $\mathcal{B}_{\mathsf{SR2PA}}$ of the SR2PA is the following:

1. Users submit a bid from $\mathsf{Bid}_{\mathsf{usr}} = \mathbb{R}_{\ge 0}$, as in any sealed-bid auction. Bids are submitted in the plaintext model.

2. $\mathsf{Adv}_{\mathsf{mi}} = \mathbb{R}_{\ge 0}$, allowing the miner to set a reserve price.

3. The allocation rule X_i to user i is given by

$$X_i(a_{\mathsf{mi}}; b_i, b_{-i}) = \begin{cases} 1 & \text{if } b_i \geq b_j \text{ for all } j \neq i \text{ and } b_i \geq a_{\mathsf{mi}} \\ 0 & \text{otherwise} \end{cases}$$

with ties broken lexicographically.

4. The payment rule P_i is as follows.

$$P_i(a_{\mathsf{mi}}; b_i, b_{-i}) = \begin{cases} 0 & \text{if } X_i(b_i, b_{-i}) = 0 \\ \max_{j \neq i}\{b_j, a_{\mathsf{mi}}\} & \text{otherwise} \end{cases}$$

5. The miner receives the square of the collected payments if the total payments is less than 1, and receives nothing otherwise.

$$\mathsf{Rev}(a_{\mathsf{mi}}; b) = \Big(\sum_{i=1}^{N} P_i(a_{\mathsf{mi}}; b_i, b_{-i}) \cdot \mathbb{1}\left[P_i(a_{\mathsf{mi}}; b_i, b_{-i}) \leq 1 \right] \Big)^2$$

We now illustrate the on-chain game $\mathcal{C}$ arising from $\mathcal{B}_{SR2PA}$ through a specific profile of strategies. Consider the miner strategy which, given a set of bids H, sets reserve $a_{\mathsf{mi}} = \max_i\{b_i\} - \epsilon$ (for some small $\epsilon > 0$) to be just below the largest submitted bid. This way, the miner gets a revenue of $(\max_i\{b_i\} - \epsilon)^2$, instead of the square of the second largest bid. As a consequence, the user i with the highest bid b_i always gets allocated and pays $b_i - \epsilon$. The auction is reduced to a first-price auction, since the winner (essentially) pays its bid. The miner gets a revenue of $(\max_i\{b_i\} - \epsilon)^2$ whenever the largest bid is less than 1. In the first-price auction, the equilibrium bidding strategy $s^{\mathcal{C}}_{\mathsf{usr},i}$ for user i involve shading their bids according to the distribution of values $\mathcal{T}$, i.e, placing a bid $b_i(v_i)$ less than their value v_i. These strategies constitute one equilibrium in the on-chain game $\mathcal{C}$. We discuss an equilibrium with a larger revenue from the off-chain game below.

In the off-chain game $\mathcal{D}$ arising from $\mathcal{B}_{SR2PA}$, the miner can use the off-chain mechanism to avoid having a portion of the revenue burnt, as we illustrate next. In $\mathcal{M}_{\mathsf{off}}$, the miner simulates a second-price auction with an optimal reserve chosen according to the distribution $\mathcal{T}$, and collects off-chain payments as dictated by the second-price auction. The miner also commits to setting reserve $a_{\mathsf{mi}} = 0$ in the on-chain game $\mathcal{C}$. Further, the miner censors all bids except the highest bid in $\mathcal{M}_{\mathsf{off}}$, and the highest bidder is instructed to bid zero on-chain. Since the user with the highest bid is the only uncensored bid on-chain, the user gets allocated and does not have to make any payments on-chain. All told, this off-chain mechanism implements a revenue-optimal second-price auction (when the distribution $\mathcal{T}$ is regular), while avoiding money burning on-chain and thus getting a strictly higher miner revenue than the equilibrium in $\mathcal{C}$ discussed above.

Formally, the off-chain mechanism and user strategies are:

1. $\mathcal{M}_{\mathsf{off}}$: The message space $\mathcal{U}_{\mathsf{off}}$ equals $\mathbb{R}_{\geq 0}$. $\mathcal{M}_{\mathsf{off}}(u_1, \ldots, u_{N'})$ returns:

   - $s^{\mathcal{C}}_{\mathsf{mi}} = (0, \{b_i : u_i \geq u_j \text{ for all } j \neq i \text{ and } u_i \geq r_0 \text{ for some reserve } r_0\}, \emptyset)$, i.e, the miner sets a zero reserve on-chain, censors all but the largest off-chain bid (if it is larger than the off-chain reserve r_0) and injects no fake bids.

   - $s^{\mathcal{C}}_{\mathsf{usr},i} = 0$. The mechanism recommends bidding zero to each user.

   - The payments in the off-chain mechanism are determined as follows:

$$p_{\mathsf{off},i}(u_1, \ldots, u_{N'}) = \begin{cases} 0 & \text{if } u_i < r_0 \text{ or } u_i \neq \max_j\{u_j\} \\ \max_{j \neq i}\{u_j\} & \text{otherwise} \end{cases}$$

14

2. $s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)$ and $\mathcal{D}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))$:

   - User i plays the strategy $s_{\mathsf{usr},i}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}$ that bids truthfully in $\mathcal{M}_{\mathsf{off}}$, i.e, bidding $u_i = v_i$. This is the user's dominant strategy since the miner has committed to running a second-price auction off-chain.

   - $X_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))$: the highest bidder is included in the block if the highest bid is larger than the reserve. The block is empty otherwise.

   - The users make zero payment on-chain. Thus, $P_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)) = p_{\mathsf{off},i}(u_1, \ldots, u_{N'})$.

In summary, $\mathcal{M}_{\mathsf{off}}$ defines a completely separate off-chain auction which allows the miner to collect payments and maximize their revenue, while avoiding any of the users' transactions being burnt on chain. Besides using SR2PA above to illustrate the off-chain game, we revisit the various equilibria of this mechanism in more detail in Section B.4 in order to discuss the relationship between our main definitions.

4   Simplicity Definitions

In this section, we give the main definitions of our paper. Throughout, we assume that users values are distributed i.i.d from a distribution $\mathcal{T}$ (though these definitions can be extended to environments where values are drawn from independent, but not necessarily identical distributions).

To begin, we define two simple pieces of terminology for convenience. First, we consider a strategy profile $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C}})$ in the on-chain game $\mathcal{C}$—or $(\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ in the off-chain game $\mathcal{D}$—such that conditioned on the miner's strategy, the users' strategy profile is a Bayes-Nash equilibrium. We call such a strategy profile to be a *user Bayes-Nash equilibrium* (user BNE), and all of our definitions restrict attention to user Bayes-Nash equilibria.

Second, we call a miner strategy *compliant* if there exists a miner advice $a_{\mathsf{mi}} \in \mathsf{Adv}_{\mathsf{mi}}$ such that, for any submitted set of user bids H, the action played by the miner is $(a_{\mathsf{mi}}, H, \emptyset)$. In other words, a compliant miner strategy chooses a constant miner advice, does not censor any bids, and does not fabricate any bids; informally, these are exactly the miner strategies we think of as "following the protocol". For any a_{mi}, we denote the corresponding compliant miner strategy by $s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}})$.

4.1   On-Chain Simplicity

We now define on-chain simplicity. On-chain simplicity is a property of a strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C}})$ in the on-chain game $\mathcal{C}$, and specifies that (i) all agents "follow the protocol" (for users, this means bidding their value, and for the miner, this means playing a compliant strategy) in $\sigma^{\mathcal{C}}$, and (ii) $\sigma^{\mathcal{C}}$ forms an equilibrium. Thus, on-chain simplicity is very close in spirit to the combination of UIC and MMIC (Roughgarden, 2020, 2021; see also Remark 4.3 below).

Definition 4.1 (On-Chain Simplicity). Consider the on-chain game $\mathcal{C}$ induced by some block-building process, and a distribution $\mathcal{T}$ of user values. The user BNE $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C}})$ is *on-chain simple* for $\mathcal{C}$ and $\mathcal{T}$ if it satisfies the following two properties:

   - *(On-Chain User Simple)* Conditioned on the miner's strategy $s_{\mathsf{mi}}^{\mathcal{C}}$, the resulting auction is DSIC and individually rational for users. In other words:

       1. $s_{\mathsf{usr},i}^{\mathcal{C}}(v_i) = v_i$, i.e, users report their values truthfully in $\mathcal{C}$.[20,21]

---

[20] We abuse notation to describe the action suggested by a strategy to be a real number instead of a set of bids (as is necessitated by our model). Technically, the above mentioned strategy should be $\{v_i\}$, signifying that the user submits only one bid, equal to his value.

[21] We are assuming that bids are non-negative real numbers in the above definition. However, the assumption is

2. For any set of bids b_{-i} submitted by other users, user i's best response is to bid truthfully. Formally, for any value v_i of user i and strategy $\widetilde{s}_{\mathsf{usr},i}^{\mathcal{C}}(\cdot) \in S_{\mathsf{usr}}^{\mathcal{C}}$ for user i,

$$v_i \cdot X_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; v_i, b_{-i}) - P_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; v_i, b_{-i}) \geq v_i \cdot X_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; \widetilde{s}_{\mathsf{usr},i}^{\mathcal{C}}(v_i), b_{-i}) - P_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; \widetilde{s}_{\mathsf{usr},i}^{\mathcal{C}}(v_i), b_{-i}).$$

3. $v_i \cdot X_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; v_i, b_{-i}) - P_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; v_i, b_{-i}) \geq 0$. The user is guaranteed a non-negative utility by bidding his value. Otherwise, (as a consequence of DSIC), the user always nets a negative utility and is better off from not participating in the auction.

- *(On-Chain Miner Simple)* The miner's strategy $s_{\mathsf{mi}}^{\mathcal{C}}$ is compliant, and is a best response to the users' strategies. Formally, the latter condition says that conditioned on users bidding according to the strategy profile $(s_{\mathsf{usr},1}^{\mathcal{C}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C}})$, the miner optimizes her expected revenue by playing $s_{\mathsf{mi}}^{\mathcal{C}}$, i.e., for any strategy $\widetilde{s}_{\mathsf{mi}}^{\mathcal{C}} \in S_{\mathsf{mi}}^{\mathcal{C}}$ of the miner,

$$\mathop{\mathbb{E}}_{v \sim \mathcal{T}} \left[\mathsf{Rev}^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v)) \right] \geq \mathop{\mathbb{E}}_{v \sim \mathcal{T}} \left[\mathsf{Rev}^{\mathcal{C}}(\widetilde{s}_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v)) \right].$$

The following corollary follows directly from the definition.

Corollary 4.2. *If the strategy profile $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C}})$ is on-chain miner simple for the on-chain game $\mathcal{C}$ induced by some block-building process $\mathcal{B}$ and distribution $\mathcal{T}$ of user values, then $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C}})$ is a Bayes-Nash equilibrium in $\mathcal{C}$ (considering both users and the miner as strategic agents).*

Remark 4.3 (On-Chain Simplicity vs. Previous Desiderata)**.** We remark that the ideas behind on-chain user and on-chain miner simplicity are essentially the same as the ideas behind UIC and MMIC Roughgarden (2020), respectively. However, there are two important differences between our definition of on-chain simplicity and these previously-considered definitions in TFM design.

First, we use a more-expansive model of TFM protocols: our model of TFM protocols allows for the miner to submit an advice. Thus, our notion of a compliant miner strategy must be more permissive than previous ones, which only consider a single unique miner strategy as "following the protocol" (namely, divulging all users' bids without fabricating any of her own).

Second, following Akbarpour and Li (2020), we adopt different semantics in that our definition posits that a *particular* strategy profile $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr}}^{\mathcal{C}})$ is on-chain simple, where prior definitions posit that the mechanism is 'simple' if some such equilibrium exists. We posit that this approach is more clear when certain properties (such as being DSIC) might hold in some equilibria of the mechanism and might not hold in others, or when the equilibrium strategies are not obvious from the mechanism. Relatedly, our definitions specify that the strategy profile is simple for a particular distribution of user values.

4.2 Off-Chain Influence Proofness

We now proceed to define the main novel simplicity condition of our paper, off-chain influence proofness. Off-chain influence proofness is a property of a strategy profile $(\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ in the off-chain game $\mathcal{D}$, and specifies that (i) the miner and users stay entirely on-chain, and (ii) the miner cannot profit even by conducting any alternative off-chain mechanism. Formally:

---

without loss of generality and can be easily extended to other bid spaces. Instead of requiring "bidding their value" being the dominant strategy, we require that there exists a bid that optimizes the user's utility irrespective of the bids placed by the other users. For instance, if the bid space is {yes, no} in a posted-price mechanism (see footnote 13 for details and more examples of alternate bid spaces), a user with value larger (lower) than the price would always want to bid yes (no) irrespective of the bids of other users.

Definition 4.4 (Off-Chain Influence Proofness). For some off-chain game $\mathcal{D}$ and distribution $\mathcal{T}$, a strategy profile $(\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ is *off-chain influence proof* if:

1. $\mathcal{M}_{\mathsf{off}}$ is a trivial off-chain mechanism (i.e., $\mathcal{M}_{\mathsf{off}}$ has no off-chain component; see Definition 3.1, Item 3). Further, $(\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ is a user BNE in the off-chain game $\mathcal{D}$.

2. $\mathcal{M}_{\mathsf{off}}$ is revenue optimal over all off-chain mechanisms that the miner can commit to, and any corresponding user Bayes-Nash equilibrium, conditioned on using $\mathcal{B}$ to finalize the block. Formally, for any off-chain mechanism $\widetilde{\mathcal{M}}_{\mathsf{off}}$ and any user Bayes-Nash equilibrium strategy profile $(\widetilde{\mathcal{M}}_{\mathsf{off}}, \widetilde{s}_{\mathsf{usr},1}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}, \ldots, \widetilde{s}_{\mathsf{usr},n}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}})$ for users in $\widetilde{\mathcal{M}}_{\mathsf{off}}$, we have

$$\mathbb{E}_{v\sim\mathcal{T}}\left[\mathsf{Rev}^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))\right] \geq \mathbb{E}_{v\sim\mathcal{T}}\left[\mathsf{Rev}^{\mathcal{D}}(\widetilde{\mathcal{M}}_{\mathsf{off}}; \widetilde{s}_{\mathsf{usr}}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v))\right].$$

Since Item 1 of this definition states that the miner and users stay entirely on-chain, we can also view on-chain influence proofness as a property of a user BNE $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{D}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D}})$ in the on-chain game $\mathcal{C}$. Throughout this paper, we refer to such a strategy profile in $\mathcal{C}$ as off-chain influence proof if the corresponding off-chain profile $(\mathcal{M}_{\mathsf{off}}^{s_{\mathsf{mi}}^{\mathcal{C}}}, \perp(s_{\mathsf{usr},1}^{\mathcal{D}}), \ldots, \perp(s_{\mathsf{usr},n}^{\mathcal{D}}))$ satisfies off-chain influence proofness.

From the perspective of a revenue-maximizing miner, Definition 4.4 gives a very strong property of the user BNE $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{D}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D}})$. Namely, even though the miner stays completely on-chain in $\sigma^{\mathcal{C}}$, in *any* other off-chain mechanism, and *any* user Bayes-Nash equilibrium induced by the other off-chain mechanisms, the revenue cannot increase beyond that of $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{D}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D}})$. Thus, the miner cannot profit by deviating from $\sigma^{\mathcal{C}}$, and this is true in a much stronger sense than typically considered under most formal notions of an equilibrium. In Section A.1, we examine several different notions of influence-proofness, and observe that off-chain influence-proofness gives the strongest guarantee among all that we consider.[22]

4.3 Strong Collusion Proofness

Lastly, we introduce the notion of strong collusion proofness, which stipulates that the miner and users cannot jointly increase their utility by colluding. This definition is close in spirit to side-contract proofness (SCP) as defined in (Chung and Shi, 2023). However, similar to Remark 4.3, we make two modifications of this definition to fit our framework. First, we consider arbitrary user Bayes-Nash equilibria (not just equilibria where users truthfully bid their values). Second, we allow the miner to increase the cartel's utility by manipulating her advice to the on-chain game $\mathcal{C}$.

Definition 4.5 (1-1-Strong Collusion Proofness). Let $\mathcal{C}$ be the on-chain game induced by some block building process, and consider some distribution $\mathcal{T}$ of user values. The strategy profile $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C}})$ in user Bayes-Nash equilibrium is *1-1-strong collusion proof*[23] (or *strong collusion proof* for short) if for each user i and value v_i for user i, the miner and user i should not be

---

[22]For example, we consider definitions which allow the miner to influence the selected user BNE without using an off-chain mechanism, and we consider whether the miner needs commitment power or not.

Of all previous definitions of which we are aware, off-chain influence proofness is most connected to the notion of a subgame perfect equilibrium. Specifically, an off-chain influence proof strategy profile is a subgame perfect equilibrium of the off-chain game, in which the miner moves first in order to commit to the mechanism $\mathcal{M}_{\mathsf{off}}$, and all users play simultaneous after learning $\mathcal{M}_{\mathsf{off}}$.

[23]For $a \in \mathbb{N}$ and $b \in \{0, 1\}$, we call a mechanism to be a-b-strong collusion proof if it is resistant to collusion when the colluding cartel has a users and b miners. Given this notation, it is not hard to extend the above definition of 1-1-strong collusion proofness to cartels of larger sizes.

able to increase their joint utility by deviating from playing $s_{\mathsf{mi}}^{\mathcal{C}}$ and $s_{\mathsf{usr},i}^{\mathcal{C}}(v_i)$ respectively. Formally, over all values v_i and bids b_i for user i and all strategies $\widetilde{s}_{\mathsf{mi}}^{\mathcal{C}}$ for the miner,

$$
\underset{v_{-i} \sim \mathcal{T}^{n-1}}{\mathbb{E}} \left[\mathsf{Rev}^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v)) + v_i \cdot X_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v)) - P_i^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; s_{\mathsf{usr}}^{\mathcal{C}}(v)) \right]
$$

$$
\geq \underset{v_{-i} \sim \mathcal{T}^{n-1}}{\mathbb{E}} \left[\mathsf{Rev}^{\mathcal{C}}(\widetilde{s}_{\mathsf{mi}}^{\mathcal{C}}; b_i, s_{\mathsf{usr},-i}^{\mathcal{C}}(v_{-i})) + v_i \cdot X_i^{\mathcal{C}}(\widetilde{s}_{\mathsf{mi}}^{\mathcal{C}}; b_i, s_{\mathsf{usr},-i}^{\mathcal{C}}(v_{-i})) - P_i^{\mathcal{C}}(\widetilde{s}_{\mathsf{mi}}^{\mathcal{C}}; b_i, s_{\mathsf{usr},-i}^{\mathcal{C}}(v_{-i})) \right]
$$

At a high level, being robust to collusion is a very desirable notion, especially in decentralized settings (where agents may be able to collude via smart contracts, and where centralization is a strong concern). However, it turns out that achieving strong collusion proofness is impossible when combined with on-chain simplicity and off-chain influence proofness, as we demonstrate in Section 6. While a full investigation of different notions of collusion is beyond the scope of our paper, we include here a brief discussion of why notions of collusion proofness similar to Definition 4.5 may be too strong, and why blockchains might be better served by adopting off-chain influence proof TFMs instead.

Most significantly, we observe that Definition 4.5 assumes that all agents in the cartel are not strategic with each other. This assumption is certainly appropriate if the miner herself is submitting a bid to the TFM. However, such an assumption is necessarily not appropriate when agents do not fully incorporate into a single entity. We posit that even while colluding, when the miner and users are distinct parties, they would be colluding according to a formally specified contract or profit-sharing function, and would be playing strategies which are in equilibrium with each other.[24] Thus, even though Definition 4.5 captures relevant concerns of collusion for some settings, it may be much stronger than needed in many cases.

In Section A.2, we investigate weaker definitions of collusion resistance which capture the above idea of colluding agents (a.k.a the "cartel"), being strategic with each other. In Section A.2.1, we observe that in fact, off-chain influence proofness directly implies *trustless collusion proofness*, a weak form of collusion resistance where all users play in equilibrium according to a proposed contract of collusion. This models collusion where all agents (including the ones not in the cartel) learn about the collusion. In Section A.2.2, we define weak collusion proofness, a stronger (but still weaker than Definition 4.5) model of collusion in which the cartel best-responds to a proposed contract, but agents not in the cartel are oblivious to the contract, and thus cannot react to the cartel deviating strategically. The cryptographic $(k+1)^{\text{th}}$ price auction (Definition 5.5), for instance, satisfies weak collusion proofness (Proposition 5.8), but not strong collusion proofness (Proposition 5.7). We believe that these observations strengthen our proposal to consider off-chain influence proof TFMs such as the cryptographic $(k+1)^{\text{th}}$ price auction, even though it is not strongly collusion proof. we leave a more thorough investigation of these and other notions of collusion as a promising direction for future work.

5 Evaluating Canonical Mechanisms

In this section, we consider various mechanisms, and ask whether they satisfy on-chain simplicity, off-chain influence proofness and strong collusion proofness. As in Section 4, we assume user values are drawn i.i.d from a regular distribution $\mathcal{T}$.

---

[24]See Section A.2 for an example of this idea in a posted price mechanism.

5.1 EIP-1559

To begin, we review the definition of EIP-1559 (Buterin et al., 2019; Roughgarden, 2021)—the transaction fee mechanism employed by the Ethereum blockchain—for blocks with an unlimited capacity. In this idealized setting, EIP-1559 essentially becomes a posted price mechanism with an exogenously-determined price, where all payments are burnt. We adopt this simplified version of EIP-1559 for our analysis.

Definition 5.1 (EIP-1559 with Unlimited Supply)**.** For an exogenously given price p, the block-building algorithm $\mathcal{B}$ of EIP-1559 is as follows:

- The mechanism takes no advice from the miner. Formally, this means that the miner advice is always some fixed $a_{\mathsf{mi}} = \emptyset$, i.e. $\mathsf{Adv}_{\mathsf{mi}} = \{\emptyset\}$.

- $\mathsf{Bid}_{\mathsf{usr}} = \mathbb{R}_{\geq 0}$. Bids consist of a non-negative real number each. Bids are submitted in the plaintext model.

- All bids greater than the price p are included in the block. For a set of bids b,

$$X_i(\emptyset; b_i, b_{-i}) = \begin{cases} 1 & \text{if } b_i \geq p \\ 0 & \text{otherwise} \end{cases}$$

- All included bids pay the price p posted by the mechanism. In other words, for a set of bid b submitted to the mechanism,

$$P_i(\emptyset; b_i, b_{-i}) = \begin{cases} p & \text{if } b_i \geq p \\ 0 & \text{otherwise} \end{cases}$$

- Finally, all payments transferred by the users are burnt. The miner receives zero revenue (or equivalently, any constant block reward).

$$\mathsf{Rev}(\emptyset; b) = 0$$

Remark 5.2. The above definition is a simplified version of EIP-1559, specific to the unlimited supply setting, which captures all necessary components required for its analysis. The original proposal and analysis of EIP-1559 (Buterin et al., 2019; Roughgarden, 2020) differs from Definition 5.1 as follows. The bid space $\mathsf{Bid}_{\mathsf{usr}} = \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ consists of two real numbers (δ_i, c_i). The first coordinate δ_i of the bid b_i is the *tip* to be transferred to the miner, while the second coordinate c_i is the *fee cap*. Bid (δ_i, c_i) indicates that user i will pay $p + \delta_i$, where p is burnt and δ_i is the tip to the miner, up to a maximum of a total payment of c_i. For a given price p, any bid i with a fee cap $c_i \geq p$ can be included (as long as the block capacity is not exceeded, in case of a finite capacity). User i pays $\min\{p + \delta_i, c_i\}$ if included, and 0 otherwise. The miner receives a revenue of $\sum_i \min\{\delta_i, c_i - p\}$, where the sum is over all allocated users i (plus a constant block reward).

While the fee cap and tips are important in case of blocks with a finite supply, the case of an unlimited supply (or a supply much larger than the demand) essentially reduces to Definition 5.1. In practice, we indeed often observe users including only very small tips and essentially just choosing if they are willing to pay the current posted price, particularly in periods of low demand. We refer the reader to Roughgarden (2020, 2021) for a thorough description of tips and fee caps. Note that the fact that we work in the simplified infinite supply case makes our negative result for EIP-1559 only stronger.[25]

---

[25]This is in contrast to the results of Chung and Shi (2023), who argue that EIP-1559 is the gold standard when supplemented with a pricing oracle which ensures that the demand is always smaller than the supply.

For the remainder of this subsection, we analyze the simplicity notions satisfied by the strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset), v_1, \ldots, v_n)$ in the on-chain game, where $s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset)$ denotes the compliant miner strategy that specifies playing $(\emptyset, H, \emptyset)$ for a set H of submitted bids (i.e, the miner does not submit an advice, does not censor any bids and does not fabricate bids), and v_i denotes user i bidding truthfully (footnote 20). We argue that $\sigma^{\mathcal{C}}$ satisfies on-chain simplicity and strong collusion proofness, but not off-chain influence proofness.

Theorem 5.3 (Adaptation of Roughgarden (2020))**.** *The strategy profile $\sigma^{\mathcal{C}}$ in the on-chain game $\mathcal{C}$ induced by EIP-1559 for any distribution $\mathcal{T}$ of user values is on-chain simple and strong collusion proof.*

Our arguments are similar to the proof of the analogous claims in Roughgarden (2020). For completeness, we prove Theorem 5.3 formally in Section C.1. We sketch the proof below.

To begin, note that on-chain user simplicity of $\sigma^{\mathcal{C}}$ (and the fact that $\sigma^{\mathcal{C}}$ is a user BNE) follows immediately since a posted-price mechanism is DSIC for users. On-chain miner simplicity follows from the fact that the miner can only decrease her revenue by getting some of her fabricated bids included and burning her own payments in the process. Thus, she maximizes her revenue by playing the compliant strategy. Finally, we verify strong collusion proofness. Since the miner gets constant revenue (unless she submits fabricated bids, which as argued above only leads to burning her own money), optimizing the joint utility of a coalition of the miner and a user reduces to optimizing the user's utility. As argued in on-chain user simplicity, bidding truthfully optimizes the user's utility, so strong collusion proofness follows.

In contrast, EIP-1559 does not satisfy off-chain influence proofness.

Theorem 5.4. *There exists regular distributions $\mathcal{T}$ of values such that the strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset), v_1, \ldots, v_n)$ is not off-chain influence proof in the off-chain game $\mathcal{D}$.*

Proof. Recall that the miners' on-chain revenue is zero (which is strategically equivalent to receiving a constant block reward); hence, it would suffice to construct any off-chain mechanism in which the miner must collect off-chain payments and receives positive expected revenue. In fact, we sketch the miner's optimal off-chain mechanism, and observe that it indeed (typically) gives the miner a positive expected revenue.

Consider a general off-chain mechanism $\mathcal{M}_{\mathsf{off}}$ and fix a user BNE under $\mathcal{M}_{\mathsf{off}}$. For any profile of user values v and any user i, let $p_i(v)$ be the expected (on-chain and off-chain) payment made by user i under this user BNE of $\mathcal{M}_{\mathsf{off}}$ in $\mathcal{D}$. By including a user i with value v_i, the miner earns a virtual welfare $\phi(v_i)$ but expends p towards burns. Thus, the miner's expected revenue equals

$$\underset{v \sim \mathcal{T}^n}{\mathbb{E}} \left[\sum_{i=1}^{n} p_i(v) - p \cdot \mathbb{1}\left[\text{user } i \text{ is included when values are } v\right] \right]$$

$$= \underset{v \sim \mathcal{T}^n}{\mathbb{E}} \left[\sum_{i=1}^{n} \left(\phi(v_i) - p \right) \cdot \mathbb{1}\left[\text{user } i \text{ is included when values are } v\right] \right],$$

where the equality follows from Theorem 2.7. Now, due to the unlimited supply model, user i can be included independently of the other users. Thus, the miner optimizes expected revenue by allocating to users with a virtual value larger than the price p. By regularity, this corresponds to allocating users with a value $v_i \geq \phi^{-1}(p)$.

Thus, the miner's optimal off-chain mechanism can be described as follows. The miner posts a price $\phi^{-1}(p)$ off-chain. Users willing to get included pay $\phi^{-1}(p) - p$ to the miner and burn p on-chain. All other users are censored. This off-chain mechanism gets the miner revenue strictly

higher than any on-chain strategy whenever the value distribution $\mathcal{T}$ has positive mass above p; hence, EIP-1559 is not off-chain influence proof. □

5.2 The Cryptographic $(k+1)^{\text{th}}$ Price Auction

Following EIP-1559, we discuss the *Cryptographic $(k+1)^{\text{th}}$ Price Auction* (abbreviated C$(k+2)$PA). The C$(k+1)$PA implements the $(k+1)^{\text{th}}$ price auction in the miner-gatekeeper model, while allowing the miner to set a reserve price.

Definition 5.5 (Cryptographic $(k+1)^{\text{th}}$ Price Auction with Reserve)**.** The block building process $\mathcal{B}$ of the cryptographic $(k+1)^{\text{th}}$ auction is as follows:

- $\mathsf{Adv}_{\mathsf{mi}} = \mathbb{R}_{\geq 0}$. The mechanism is advised by the miner on the reserve price a_{mi}.

- $\mathsf{Bid}_{\mathsf{usr}} = \mathbb{R}_{\geq 0}$, similar to typical sealed-bid auctions. Bids are submitted in the miner-gatekeeper model.

- All bids b_i larger than the reserve price, up to k bids are included in the block. In other words, for a set of bids b and a reserve a_{mi},

$$X_i(a_{\mathsf{mi}}; b_i, b_{-i}) = \begin{cases} 1 & \text{if } b_i \geq a_{\mathsf{mi}} \text{ and } |\{j \neq i | b_j > b_i\}| \leq k-1 \\ 0 & \text{otherwise} \end{cases}$$

- All included users pay the minimum bid needed to get included, i.e, the maximum over the reserve price and the $(k+1)^{\text{th}}$ largest bid. For a set of bids b such that $b^{(k+1)}$ is the $(k+1)^{\text{th}}$ largest bid (set $b^{(k+1)} = 0$ if there are less than $k+1$ bids),

$$P_i(a_{\mathsf{mi}}; b_i, b_{-i}) = \begin{cases} \max\{b^{(k+1)}, a_{\mathsf{mi}}\} & \text{if } b_i \geq a_{\mathsf{mi}} \text{ and } |\{j \neq i | b_j > b_i\}| \leq k-1 \\ 0 & \text{otherwise} \end{cases}$$

- All collected payments are transferred to the miner.

$$\mathsf{Rev}(a_{\mathsf{mi}}; b_i, b_{-i}) = \sum_i P_i(a_{\mathsf{mi}}; b_i, b_{-i})$$

Consider the strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), v_1, \ldots, v_n)$ in the on-chain game $\mathcal{C}$, where $s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}})$ specifies playing $(a_{\mathsf{mi}}, H, \emptyset)$ for each set of bids H, and where a_{mi} is the monopoly reserve for a regular distribution $\mathcal{T}$ (as defined in Definition 2.8). We will argue that $\sigma^{\mathcal{C}}$ is on-chain simple and off-chain influence proof, but typically does not satisfy strong collusion proofness.

Theorem 5.6. *For all regular distributions $\mathcal{T}$ with a monopoly reserve a_{mi}, the strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), v_1, \ldots, v_n)$ in the on-chain game $\mathcal{C}$ induced by C$(k+1)$PA is on-chain simple and off-chain influence proof.*

We sketch the proof of Theorem 5.6. The $(k+1)^{\text{th}}$ price auction is DSIC for users (Vickrey, 1961; Myerson, 1981; we include a proof in Section C.2 for completeness). Thus, $\sigma^{\mathcal{C}}$ is a user BNE and on-chain user simple for all distributions $\mathcal{T}$.

Next, we show that C$(k+1)$PA is off-chain influence proof. From Theorem 2.9, we know $(s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), v_1, \ldots, v_n)$ is the revenue optimal mechanism over all possible BNE for a regular distribution $\mathcal{T}$. Any off-chain mechanism announced by the miner is going to induce a user BNE in

the off-chain game and cannot have a larger revenue than $\sigma^{\mathcal{C}}$.

Intuitively, to show on-chain miner simplicity, all we need to do is ensure that the miner doesn't want to change the reserve or drop and fabricate bids based on the number of submitted bids (since the miner-gatekeeper cryptographic model says that the miner cannot see the value of the bids). To formally prove this, one can use a similar idea to off-chain influence proofness, by first showing that truthful bidding remains a user BNE irrespective of the miner's on-chain strategy (assuming users are required to submit exactly one bid)[26], and second, by using Theorem 2.9 to conclude that by setting the monopoly reserve, the miner earns revenue at least as high as in any BNE of any mechanism. Fix some $s_{\mathsf{mi}}^{\mathcal{C}}$, some user i, and on-chain strategies $s_{\mathsf{usr},-i}^{\mathcal{C}}$ of users other than i. Since the C$(k+1)$PA is executed in the miner-gatekeeper model, the miner's strategy is independent of the values of any bids. The resulting mechanism is identical to executing C$(k+1)$PA only on the uncensored bids, in which truthful bidding is a dominant strategy for users. Now suppose the miner also fabricates bids, which once again are independent of the values of the users' bids. From user i's vantage, he gets the same utility while participating in the C$(k+1)$PA where all fake bids are replaced by a real user with an identical bid. Once again, truthful bidding is the optimal strategy. Since any strategy played by the miner induces a user BNE, the miner is better off inducing the revenue optimal BNE by playing $s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}})$.

However, $\sigma^{\mathcal{C}}$ in C$(k+1)$PA is not strongly collusion proof.

Proposition 5.7. *There exist regular distributions $\mathcal{T}$ of user values such that the strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), v_1, \ldots, v_n)$, for the monopoly reserve a_{mi}, is not strongly collusion proof in the on-chain game $\mathcal{C}$ induced by $C(k+1)PA$.*

Proof. We describe the optimal collusion strategy between the miner and a user i with value v_i, and observe that this optimal collusion strategy typically requires the miner and user i to deviate away from $\sigma^{\mathcal{C}}$. In fact, our argument parallels a result for revenue optimality provided in Bulow and Roberts (1989) for settings where the auctioneer has a non-zero value for the good.

To state this argument, we treat the cartel containing the miner and a user i to be a single entity, who we refer to as the *grand auctioneer*. Fix the value v_i of user i. Imagine that the grand auctioneer can always allocate user i, provided that at lease one of the k slots is not filled by other bidders. Then the grand auctioneer incurs a reward v_i equal to the value of user i if all k slots in the block are not filled.

As argued above for on-chain miner simplicity, any strategy played by the grand auctioneer will result in a DSIC on-chain game for all users apart from i (formally, when constraining all user to submitting exactly one bid, similar to footnote 26). Thus, if the grand auctioneer induces an allocation rule $(\widetilde{X}_j)_{j \neq i}$ for each bidder other than i, the expected revenue is given by (applying Theorem 2.2)

$$\mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[v_i \cdot \mathbb{1} \left[\sum_{j \neq i} \widetilde{X}_j(v_{-i}) \leq k \right] + \sum_{j \neq i} \phi(v_j) \cdot \mathbb{1} \left[\widetilde{X}_j(v_{-i}) \right] \right]$$

The grand auctioneer can optimize the above pointwise (i.e., for all $(v_j)_{j \neq i}$) by ensuring that user i is included whenever the kth-largest bid (among users other than i) has a virtual value less

---

[26]Assuming users can submit only one bid is important. For instance, if the miner plays the following on-chain strategy— if an odd number of bids are submitted, run the C$(k+1)$PA with the monopoly reserve a_{mi}, else censor all bids. Some user is incentivized to fabricate a fake bid equal to zero if an even number of bids are submitted. However, since we are proving on-chain miner simplicity, we would want to argue the miner's optimal strategy conditioned on users bidding truthfully is setting the monopoly reserve on-chain. Thus, we can assume that users do not fabricate bids or censor themselves and bid truthfully. We argue that users are anyways best responding by truthful bidding as a stepping-stone to proving on-chain miner simplicity.

than v_i (or, if there are less than k bids). This can be achieved by the colluding user i placing a bid equal to $\phi^{-1}(v_i)$ (and the miner continuing to play $s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}})$). Under the above mentioned strategy profile, whenever the kth largest bid has a value less than $\phi^{-1}(v_i)$, it will have a virtual value smaller than v_i (by monotonicity) and will be outbid by user i; thereby optimizing the grand auctioneer's (and the cartel's) objective. Indeed, for typical regular distributions, there will be many values v_i such that $v_i \neq \phi^{-1}(v_i) > a_{\mathsf{mi}}$, and hence the utility of the grand auctioneer strictly increases via collusion. $\qquad\square$

Revisiting the notions of collusion resistance. The Cryptographic $(k + 1)^{\text{th}}$-Price Auction satisfies all the desiderata highlighted in Section 4, except for strong collusion proofness. We revisit our arguments on the appropriateness of strong collusion proofness from Section 4.3, and argue that $\mathrm{C}(k + 1)\mathrm{PA}$ may be intuitively considered fairly resistant to collusion. In particular, we argue that truthful bidding in $\mathrm{C}(k + 1)\mathrm{PA}$ satisfies *weak* collusion proofness, which says that when the miner and a colluding user strategically participate in some specific contract of collusion, either the miner or the colluding user cannot profit. We defer the full definition of this property to Definition A.4, and sketch the proof next.

Consider a coalition between the miner and a user i. An analogous argument to on-chain miner simplicity would prove that truthful bidding for users $j \neq i$ is still their optimal response (conditioned on having to submit exactly one bid), regardless of the actions played by the miner and user i. Since user i knows about the collusion contract, user i would play a best response to the miner's strategy and truthful bidding by users apart from i. Thus, the strategy profile played by users is a user Bayes-Nash equilibrium in some mechanism (which treats the colluding user i differently from others). The miner achieves the optimal revenue over all user Bayes-Nash equilibria by playing $s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}})$, where a_{mi} is the monopoly reserve of the regular value distribution $\mathcal{T}$. Thus, the miner cannot increase its expected revenue through collusion.

Proposition 5.8. $(s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), v_1, \ldots, v_n)$ *is weak collusion proof for a regular distribution $\mathcal{T}$ of values, where a_{mi} is the monopoly reserve.*

Now, we argue that even when the miner submits her own bids (informally speaking, this means that she engages in collusion is the sense of strong collusion proofness), other users' best-responses are still truth-telling. Suppose the miner had a few transactions of her own, with transaction i valued $v_{\mathsf{mi}}^{(i)}$ by the miner. The above scenario is similar to the miner (strongly) colluding with users with values $v_{\mathsf{mi}}^{(i)}$, where the miner and the users are not strategic with each other. Thus, the miner has an incentive to not bid truthfully and place bids equal to $\phi^{-1}(v_{\mathsf{mi}}^{(i)})$ (as argued in the proof of Proposition 5.7). However, we noted that irrespective of the miner's strategy, users not in the coalition best respond by bidding truthfully. Thus, even when the miner engages in non-truthful behaviour that increases her revenue, users do not have an ex-post regret on bidding truthfully in the TFM.

Remark 5.9. Ethereum burns money in EIP-1559 not just to ensure simple participation for users, but also to maintain a deflationary pressure on the cryptocurrency. The $\mathrm{C}(k + 1)\mathrm{PA}$ can be supplemented with an (exogenous) burn too, while still allowing the user to set a reserve on top of the burn. All of our analysis of the $\mathrm{C}(k + 1)\mathrm{PA}$ continues to hold with very minor modifications. In particular, there exists an optimal reserve that the miner can set such that truthful bidding continues to remain on-chain simple and off-chain collusion proof. Getting the exogenously set reserve price so that the demand does not exceed the capacity of the block is necessary for EIP-1559 to satisfy on-chain user simplicity. On the other hand, $\mathrm{C}(k+1)\mathrm{PA}$ satisfies on-chain simplicity

and off-chain influence proofness irrespective of the burn fee. The burn fee has to now satisfy one less constraint— the capacity meeting the demand.

5.3 The Winner-Pays-Bid Mechanism

We now consider the winner-pays-bid mechanism, the generalization of the first-price auction to a block capacity k, and was the status quo in TFMs before EIP-1559 and still the TFM deployed in Bitcoin.

The block-building algorithm $\mathcal{B}$ of the winner-pays-bid mechanism is identical to the $\text{C}(k+1)\text{PA}$, except the payment rule—instead of paying the minimum amongst the reserve or the $(k+1)^{\text{th}}$ largest bid, users included in the block pay their bid.

Definition 5.10 (Winner-Pays-Bid Mechanism with Reserve)**.** The block building process $\mathcal{B}$ of the winner-pays-bid mechanism is as follows:

- The miner sets a reserve price a_{mi} as part of its advice, i.e, $\text{Adv}_{\text{mi}} = \mathbb{R}_{\geq 0}$.

- $\text{Bid}_{\text{usr}} = \mathbb{R}_{\geq 0}$. Bids are submitted in the miner-gatekeeper model.

- All bids b_i larger than the reserve price, up to k bids are included in the block. In other words, for a set of bids b and a reserve a_{mi},

$$X_i(a_{\text{mi}}; b_i, b_{-i}) = \begin{cases} 1 & \text{if } b_i \geq a_{\text{mi}} \text{ and } |\{j \neq i | b_j > b_i\}| \leq k - 1 \\ 0 & \text{otherwise} \end{cases}$$

- All included users pay their bid. For a set of bids b,

$$P_i(a_{\text{mi}}; b_i, b_{-i}) = \begin{cases} b_i & \text{if } b_i \geq a_{\text{mi}} \text{ and } |\{j \neq i | b_j > b_i\}| \leq k - 1 \\ 0 & \text{otherwise} \end{cases}$$

- All collected payments are transferred to the miner.

$$\text{Rev}(a_{\text{mi}}; b_i, b_{-i}) = \sum_i P_i(a_{\text{mi}}; b_i, b_{-i})$$

We observe two similar looking, yet different user BNEs satisfying different set of desiderata. First, note that bidding truthfully is a dominated strategy which always results in zero utility for users, and thus users would want to shade their bid (i.e. bid below their value). Indeed, it turns out that for any given reserve a_{mi} and regular distribution $\mathcal{T}$ of user values, (Myerson, 1981; Chawla and Hartline, 2013) show the existence of a unique (for all values $v_i \geq a_{\text{mi}}$) strategy $s_{\text{usr}}^{\mathcal{C}, \mathcal{T}, a_{\text{mi}}}$ for users to shade their bids such that the strategy profile $(s_{\text{mi}}^{\text{comp}}(a_{\text{mi}}), s_{\text{usr}}^{\mathcal{C}, \mathcal{T}, a_{\text{mi}}}, \ldots, s_{\text{usr}}^{\mathcal{C}, \mathcal{T}, a_{\text{mi}}})$ is a user BNE. We consider two variants of this equilibrium below.[27]

At a high level, the two user Bayes-Nash equilibria differ in how users with a probability zero of getting included (have a values less than the reserve a_{mi}) behave. Whenever a user sees a value $v_i \geq a_{\text{mi}}$, he bids according to $s_{\text{usr}}^{\mathcal{C}, \mathcal{T}, a_{\text{mi}}}$ in both the strategy profiles. However, when their values are

---

[27]The existence of the two user Bayes-Nash equilibria $\sigma_{\text{val}}^{\mathcal{C}}$ and $\sigma_0^{\mathcal{C}}$ in the winner-pays-bid mechanism that satisfy two different set of simplicity assumptions motivate describing simplicity as a property of the equilibrium as opposed to being a property of the mechanism (see Remark 4.3). While the two equilibria here are similar, see Section B.3 for an example of an auction which induces two very different looking equilibria satisfying two very different set of simplicities.

24

smaller than the reserve, in the first profile, they bid their values and in the second, they bid zero. In both cases, they end up bidding below the reserve and do not get allocated. Thus, both the profiles have the same interim allocation and payment rules, which in turn is the same as the user BNE $(s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), s_{\mathsf{usr}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}, \ldots, s_{\mathsf{usr}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}})$. Thus, both the strategy profiles are user BNEs. However, we will verify that these user BNEs differ with respect to on-chain miner simplicity.

Formally, let $\sigma_{\mathsf{val}}^{\mathcal{C}}$ be the user BNE $(s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), s_{\mathsf{usr,val}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}, \ldots, s_{\mathsf{usr,val}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}})$, where a user playing according to $s_{\mathsf{usr,0}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}$ bids the same as $s_{\mathsf{usr}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}$ for all values $v_i \geq a_{\mathsf{mi}}$ and bids his value v_i otherwise. Similarly, let $\sigma_0^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), s_{\mathsf{usr,0}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}, \ldots, s_{\mathsf{usr,0}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}})$, where the output of $s_{\mathsf{usr,0}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}$ equals that of $s_{\mathsf{usr}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}$ for all values $v_i \geq a_{\mathsf{mi}}$ and 0 otherwise.

Proposition 5.11. *Fix a regular distribution $\mathcal{T}$ of user values, and let a_{mi} be the monopoly reserve of $\mathcal{T}$. Then, the user BNE $\sigma_{\mathsf{val}}^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), s_{\mathsf{usr,val}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}, \ldots, s_{\mathsf{usr,val}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}})$ is not on-chain user simple, on-chain miner simple and not strong collusion proof. However, $\sigma_{\mathsf{val}}^{\mathcal{C}}$ satisfies off-chain influence proofness.*

Proof. Users are not bidding truthfully, and thus, $\sigma_{\mathsf{val}}^{\mathcal{C}}$ is not on-chain user simple.

We argue off-chain influence proofness next. Similar to off-chain influence in C2PA, we will argue that the revenue optimal equilbrium is already induced on-chain. For the monopoly reserve a_{mi}, observe that the highest k bids, conditional on being larger than a_{mi} get allocated. Since all users play the same strategy $s_{\mathsf{usr,val}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}$, the users with the k largest values (conditional on being larger than the reserve a_{mi}) are allocated. Thus, $\sigma_{\mathsf{val}}^{\mathcal{C}}$ has the same interim allocation rule as C$(k+1)$PA. By revenue equivalence (Lemma 2.5), the miner's expected revenue from $\sigma_{\mathsf{val}}^{\mathcal{C}}$ equals the miner's revenue from the truth-telling user BNE discussed in Section 5.2 for the reserve a_{mi}. But, from Theorem 2.9, the truth-telling user BNE with reserve a_{mi} is revenue optimal over all possible user Bayes-Nash equilibria. Thus, $\sigma_{\mathsf{val}}^{\mathcal{C}}$ is a revenue optimal user BNE, and the miner will not want to induce a different equilibria via an off-chain mechanism.

Next, we show the lack of strong collusion proofness and the lack of on-chain miner simplicity. In fact, these two statements will follow from the same arguments. Consider the miner deviating and setting the reserve to zero (but continues to not fabricate or censor bids). At a high-level, this can potentially include a user with value $v_i < a_{\mathsf{mi}}$. In detail, with a non-zero probability, all users are going to have a value less than a_{mi}. In such scenarios, the miner gets a non-zero revenue when setting the reserve to zero, but a zero reserve when setting the reserve to a_{mi}. However, since winners pay their bids, all users with values larger than the reserve continue to pay their bids. Thus, the miner gets a strictly larger revenue by setting the reserve to zero. Indeed, a colluding user i with a value $v_i < a_{\mathsf{mi}}$ always gets zero utility (either by paying v_i and getting included, or by being excluded). Thus, by increasing her revenue, the miner also increases the joint utility of the cartel. Thus, $\sigma_{\mathsf{val}}^{\mathcal{C}}$ is neither on-chain miner simple, nor strong collusion proof. $\square$

Proposition 5.12. *Fix a regular distribution $\mathcal{T}$ of user values, and let a_{mi} be the monopoly reserve of $\mathcal{T}$. Then, the user BNE $\sigma_0^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), s_{\mathsf{usr,0}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}, \ldots, s_{\mathsf{usr,0}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}})$ is not on-chain user simple and not strong collusion proof. However, $\sigma_0^{\mathcal{C}}$ satisfies on-chain miner simplicity and off-chain influence proofness.*

Proof. The arguments regarding lack of on-chain simplicity and off-chain influence proofness are identical to $\sigma_{\mathsf{val}}^{\mathcal{C}}$ (Proposition 5.11).

Proof of lack of strong collusion non-proofness is also similar, except the miner sets a reserve zero and the colluding user i with a value $v_i < a_{\mathsf{mi}}$ bids his value. Following the above deviation allocates user i, for instance, when all other users have a value less than the reserve (which happens

with a non-zero probability). Instead of getting a joint utility zero, the cartel gets a joint utility v_i. In other scenarios, the above mentioned deviation does not decrease the joint utility since included users pay their bids. User i may get included, increasing the joint utility, or get excluded, which does not decrease the cartel's utility.

$\sigma_0^{\mathcal{C}}$ is on-chain miner simple. By decreasing the reserve, the miner can potentially include users who would not have been included at a reserve a_{mi}. However, all such users would be bidding 0, and thus, does not increase the miner's revenue. $\square$

Remark 5.13. Interestingly, the notions of collusion resistance proposed in Roughgarden (2020) (off-chain agreement proof, OCA-proof) and Chung and Shi (2023) (side contract proofness, SCP) both classify the winner-pays-bid mechanism as collusion resistant. However, we show above that the winner-pays-bid mechanism is not strongly collusion proof (i.e., our adaptation of OCA proofness and SCP). This difference comes from the fact that OCA-proofness and SCP both assume that users play truth-telling strategies. We view this as a further reason to consider simplicity properties of *particular equilibria*, rather than the mechanism itself (as discussed in Remark 4.3). (As an aside, we note that by our Corollary A.3, the winner-pays-bid mechanism satisfies a weaker form of collusion resistance which we call trustless collusion proofness (Definition A.2), albeit for a different reason than suggested by prior frameworks).

Remark 5.14. Note that the optimal shading strategy for users $s_{\mathsf{usr}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}(v_i)$ is dependent on the number of users participating in the mechanism. For instance, if $k = 1$, $\mathcal{T}$ equals $U[0,1]$ and the reserve a_{mi} is set to zero, $s_{\mathsf{usr}}^{\mathcal{C},\mathcal{T},a_{\mathsf{mi}}}(v_i) = \frac{(n-1)}{n} \cdot v_i$ (Myerson, 1981; Chawla and Hartline, 2013) where n is the number of participants. However, in a decentralized environment, it might be tricky for a user to gauge the number of other users participating in the auction. Thus, even when informed of the distribtuion $\mathcal{T}$ and the optimal strategy, it could be difficult for users to bid optimally in practise.

In Appendix B, we discuss various other standard mechanisms like the plaintext $(k+1)^{\mathrm{th}}$ price auction and the posted-price mechanism. We also study a variety of auctions like the bonus-on-matching bids auctions and the squared revenue second price auction so that the equilibria in our examples satisfy all possible combinations of our core definitions—on-chain user simplicity, on-chain miner simplicity and off-chain collusion proofness.

6 Impossibility Results

In this section, we present our main impossibility result: we prove that no nontrivial mechanism is on-chain simple, off-chain influence proof, and collusion proof. At a high level, we first show that no mechanism that satisfies on-chain simplicity and off-chain influence proof can award the miner a non-trivial revenue. However, a miner awarded a trivial revenue can achieve a larger revenue by running a separate auction off-chain.

6.1 On-Chain Simplicity and Strong Collusion Proofness Implies Constant Expected Revenue

We begin by showing an impossibility result similar to Chung and Shi (2023), who show that any mechanism satisfying user incentive compatibility and side contract proofness will have to award the miner a zero revenue. Analogously, we show that the expected revenue of a miner (expectation taken over the randomness of the mechanism and values of users) must be independent of the number of users participating in the mechanism if the mechanism is on-chain simple and collusion

proof.

Lemma 6.1. *For a distribution $\mathcal{T}$ of user values, consider a block-building process $\mathcal{B}$ and a sequence of user Bayes-Nash equilibria $\left((s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C},n}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C},n}) \right)_{n \in \mathbb{N} \cup \{0\}}$ in the on-chain game $\mathcal{C}$, where the strategy profile $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C},n}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C},n})$ corresponds to the equilibrium induced on-chain when n users participate. Suppose for each $n \in \mathbb{N} \cup \{0\}$, $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C},n}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C},n})$ is on-chain (user and miner) simple, and strongly collusion-proof. Then:*

1. *In expectation over users' values and the randomness of the TFM, the miner's expected revenue is independent of the number of users participating in the TFM.*

2. *The expected miner revenue is invariant to the number of users even after conditioning on the bid of some single user.*

Proof. Fix the distribution $\mathcal{T}$ of user values. Consider the strategy profile $\sigma^{\mathcal{C},n} = (s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C},n}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C},n})$. Let $x_i^{(n)}$ and $p_i^{(n)}$ be the interim allocation and payment rules (Definition 2.1) induced by $\sigma^{\mathcal{C},n}$ for user i.

First, for each user i, we prove that the miner revenue, in expectation over the values of all other users, is independent of user i's value. By on-chain user simplicity, we have that the induced mechanism is DSIC (and therefore BIC) and thus,

$$p_i^{(n)}(v_i) = \int_0^{v_i} z x_i^{(n)\prime}(z) \, dz + c \tag{1}$$

for some constant c (by the payment identity, Theorem 2.2).

Let $\mathsf{rev}^{(n)}(v_i) = \mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}^{\mathcal{C}}(s_{\mathsf{mi}}^{\mathcal{C}}; v_i, v_{-i}) \right]$ be the miner's "interim" revenue when user i bids v_i. We now apply collusion-proofness on the cartel containing the miner and the user i. Fixing the miner's strategy $s_{\mathsf{mi}}^{\mathcal{C}}$ but varying user i's bid b_i, strong collusion-proofness means that the cartel optimizes its joint utility when user i bids his value v_i:

$$v_i \cdot x_i^{(n)}(v_i) - p_i^{(n)}(v_i) + \mathsf{rev}^{(n)}(v_i) \geq v_i \cdot x_i^{(n)}(\widetilde{v}_i) - p_i^{(n)}(\widetilde{v}_i) + \mathsf{rev}^{(n)}(\widetilde{v}_i) \tag{2}$$

for all possible bids $\widetilde{v}_i$ that user i can submit. Define the expected net *expenditure* of the cartel, $\mathsf{exp}^{(n)}(v_i) = p_i^{(n)}(v_i) - \mathsf{rev}^{(n)}(v_i)$, to be the difference in money held by the cartel before and after the mechanism. Then, Equation 2 becomes

$$v_i \cdot x_i^{(n)}(v_i) - \mathsf{exp}^{(n)}(v_i) \geq v_i \cdot x_i^{(n)}(\widetilde{v}_i) - \mathsf{exp}^{(n)}(\widetilde{v}_i). \tag{3}$$

Now, the main observation towards our proof is the following. If we pretend that user i pays all of the expenditure $\mathsf{exp}^{(n)}$, then *Equation 3 is precisely the definition of BIC*. Thus, the payment identity (Theorem 2.2) can be applied to the expenditure $\mathsf{exp}^{(n)}$ to get

$$\mathsf{exp}^{(n)}(v_i) = \int_0^{v_i} z x_i^{(n)\prime}(z) \, dz + \widetilde{c}.$$

for a constant $\widetilde{c}$, maybe different from c. Thus, applying Equation 1 and the definition of $\mathsf{exp}^{(n)}$, we get

$$p_i^{(n)}(v_i) - C = \mathsf{exp}^{(n)}(v_i) = p_i^{(n)}(v_i) - \mathsf{rev}^{(n)}(v_i),$$

and thus, $\mathsf{rev}^{(n)}(v_i) = C$ for some constant C and for all $v_i \in \mathbb{R}_{\geq 0}$.

Next, we prove that the $\mathsf{rev}^{(n)}(0) = \mathsf{rev}^{(n-1)}(0)$. Since $\mathsf{rev}^{(n)}(v_i) = \mathsf{rev}^{(n)}(0)$ and $\mathsf{rev}^{(n-1)}(v_i) = \mathsf{rev}^{(n-1)}(0)$, this will prove the second claim in Lemma 6.1. Intuitively, this follows due to on-chain miner simplicity (and strong-collusion proofness), since the miner (and the cartel, respectively) has no incentive to add a zero bid (or drop a zero bid, respectively). We now go through this argument fully.

To begin, we argue $\mathsf{rev}^{(n)}(0) \geq \mathsf{rev}^{(n-1)}(0)$. Suppose that the miner is colluding with a user i with value 0. The cartel's expected joint utility when user i bids truthfully equals $\mathsf{rev}^{(n)}(0) + 0 \times x_i^{(n)}(0) - p_i^{(n)}(0) = \mathsf{rev}^{(n)}(0) - p_i^{(n)}(0)$. By individual rationality that follows due to on-chain user simplicity, user i will make no payments by bidding zero. Thus, the joint utility of the cartel equals $\mathsf{rev}^{(n)}(0)$. Consider the deviation from $\sigma^{\mathcal{C},n}$ where the miner censors user i. User i continues to contribute nothing to the joint utility, while the miner's revenue becomes $\mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_{-i}) \right]$. Thus, by collusion proofness, we have

$$\mathsf{rev}^{(n)}(0) \geq \mathop{\mathbb{E}}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_{-i}) \right].$$

On the other hand, we have proved $\mathsf{rev}^{(n-1)}(v_i)$ is a constant, and

$$\mathop{\mathbb{E}}_{v_i \sim \mathcal{T}} \left[\mathsf{rev}^{(n-1)}(v_i) \right] = \mathop{\mathbb{E}}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_{-i}) \right],$$

by definition. Thus, $\mathsf{rev}^{(n-1)}(0) = \mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_{-i}) \right]$. Combining the above arguments, we have $\mathsf{rev}^{(n)}(0) \geq \mathsf{rev}^{(n-1)}(0)$.

It remains to prove $\mathsf{rev}^{(n)}(0) \leq \mathsf{rev}^{(n-1)}(0)$. As argued above $\mathsf{rev}^{(n-1)}(0) = \mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_{-i}) \right]$. Consider a miner considering fabricating a bid $\tilde{b} = 0$ with $n - 1$ participating users. Prior to fabricating the bid, the miner received an expected revenue equal to $\mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_{-i}) \right]$. By fabricating $\tilde{b}$, the miner makes no payments (by on-chain user simplicity) but receives an expected revenue equal to $\mathsf{rev}^{(n)}(0)$. By on-chain miner simplicity,

$$\mathsf{rev}^{(n)}(0) \leq \mathop{\mathbb{E}}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_{-i}) \right] = \mathsf{rev}^{(n-1)}(0).$$

The above concludes the proof that $\mathsf{rev}^{(n)}(0) = \mathsf{rev}^{(n-1)}(0)$, and hence the proof of the second claim in Lemma 6.1. From here, the first claim is straightforward.

$$\mathsf{rev}^{(n)}(v_i) = \mathsf{rev}^{(n-1)}(v_i)$$

for all v_i, and taking expectation over v_i on both sides, the proof follows. $\quad\square$

As a quick sanity check, remember that the strategy profile $(s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset), v_1, \ldots, v_n)$ in EIP-1559 with unlimited supply satisfies on-chain simplicity and strong collusion proofness (Theorem 5.3). As Lemma 6.1 claims, the miner gets a constant block reward, independent of the number of users.

Remark 6.2. To be precise, Lemma 6.1 claims that, in expectation over users' bids, the miner's revenue is independent over the number of users. We believe that the miner revenue must be independent of the bids submitted to the on-chain game too. More formally, we conjecture that if a sequence of user Bayes-Nash equilibria $\left(\sigma^{\mathcal{C},n} \right)_{n \in \mathbb{N} \cup \{0\}}$ satisfy on-chain simplicity and collusion proofness, then the miner revenue must be a constant with probability 1 over the bids placed by users. We leave this conjecture as an open problem for future work.

6.2 On-Chain Simplicity, Strong Collusion Proofness, and Off-Chain Influence Proofness Implies No Non-Trivial Mechanism

Next, building on Lemma 6.1, we show that if a TFM satisfies previously-considered desiderata as well as off-chain influence proofness, then the mechanism must be trivial (namely, the mechanism cannot award allocate any user with positive probability). Intuitively, this follows because off-chain influence proofness claims that the miner is already receiving the optimal revenue over all user Bash-Nash equilibria that can be induced in the on-chain game $\mathcal{C}$ over the block-building process $\mathcal{B}$, and this clashes with the claim of Lemma 6.1 that claims the miner revenue is independent of the number of users participating in the TFM.

Our proof proceeds in two steps. First, we show that in a user BNE satisfying on-chain simplicity, off-chain influence proofness, and strong collusion proofness, all users must get zero utility. Second, we prove that any mechanism where users are always awarded a zero utility to users must be trivial in in a formal sense.

Lemma 6.3. *For a given distribution $\mathcal{T}$ of user values, and a sequence of on-chain simple and collusion proof user Bayes-Nash equilibria $\left((s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C},n}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C},n}) \right)_{n \in \mathbb{N} \cup \{0\}}$ in the on-chain game $\mathcal{C}$ Then, for all $n \in \mathbb{N}$ and all users i, the expected utility of user i equals zero. Formally,*

$$\mathop{\mathbb{E}}_{v \sim \mathcal{T}^n} \left[v_i \cdot x_i^{(n)}(v_i) - p_i^{(n)}(v_i) \right] = 0.$$

Proof. Let $\mathsf{util}_i^{(n)}(v_i) = v_i \cdot x_i^{(n)}(v_i) - p_i^{(n)}(v_i)$ be user i's interim utility function. Suppose there exists a user i with an expected utility $\mathbb{E}_{v_i \sim \mathcal{T}} \left[\mathsf{util}_i^{(n)}(v_i) \right] > 0$. The miner runs the following off-chain mechanism $\mathcal{M}_{\mathsf{off}}$. The miner asks user i (and only user i) to pay $(1/2) \cdot \mathbb{E}_{v_i \sim \mathcal{T}} \left[\mathsf{util}_i^{(n)}(v_i) \right]$ off-chain, barring which user i would be censored by the miner in the on-chain game. Irrespective of user i's strategy, the miner plays $s_{\mathsf{mi}}^{\mathcal{C}}$ on-chain.

From the perspective of all user apart from i, conditioned on the miner playing the strategy $s_{\mathsf{mi}}^{\mathcal{C}}$ on-chain, the resulting game is on-chain user simple. Thus, bidding v_{-i} continues to be the best response in the off-chain game induced by the miner's off-chain mechanism $\mathcal{M}_{\mathsf{off}}$.

From user i's perspective, i receives an expected utility $\mathbb{E}_{v_i \sim \mathcal{T}} \left[\mathsf{util}_i^{(n)}(v_i) \right]$, and thus, with positive probability, must be willing to pay $(1/2) \cdot \mathbb{E}_{v_i \sim \mathcal{T}} \left[\mathsf{util}_i^{(n)}(v_i) \right]$ to be included in the block. Thus, $\mathbb{E}_{v \sim \mathcal{T}^n} \left[p_{\mathsf{off},i}(\mathcal{M}_{\mathsf{off}}; v_i, v_{-i}) \right] > 0$.

From the miner's perspective, the miner receives the same on-chain revenue after announcing $\mathcal{M}_{\mathsf{off}}$ as playing the trivial off-chain mechanism corresponding to $s_{\mathsf{mi}}^{\mathcal{C}}$. This is because, (since the user Bayes-Nash equilibria satisfy on-chain simplicity and collusion proofness) from the second part of Lemma 6.1,

$$\mathop{\mathbb{E}}_{v_{-i} \sim \mathcal{T}^{n-1}} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_i, v_{-i}) \right] = \mathop{\mathbb{E}}_{v \sim \mathcal{T}^n} \left[\mathsf{Rev}(s_{\mathsf{mi}}^{\mathcal{C}}; v_i, v_{-i}) \right]$$

and thus, in expectation over the values v_{-i} of other users, the miner's revenue is independent of user i's on-chain strategy. However, the miner also receives strictly positive revenue off-chain, and thus, prefers the equilibrium induced by $\mathcal{M}_{\mathsf{off}}$.

Thus, $(s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C},n}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C},n})$ is not off-chain influence proof, as required. $\qquad\square$

Finally, we argue that in general auction settings, if every user gets an expects a zero utility in a mechanism, then the mechanism allocates with a positive probability only if the user draws a value equal to the supremum of the distribution $\mathcal{T}$.

Lemma 6.4. *Consider some DSIC and individually rational mechanism given by an allocation and payment rule (X, P). Let users' values be distributed according to $\mathcal{T}$ with a supremum $v_{\sup} \in \mathbb{R} \cup \{\infty\}$.*[28]

1. *If user i receives a zero expected utility in (X, P) (in expectation over the randomness in the mechanism and values of all users, including user i), i.e,*

$$\mathop{\mathbb{E}}_{v \sim \mathcal{T}^n} [v_i \cdot x_i(v_i, v_{-i}) - p_i(v_i, v_{-i})] = 0,$$

then $\mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} [x_i(v_i, v_{-i})] = 0$ for all $v_i \neq v_{\sup}$.

2. *Suppose $\mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} [x_i(v_i, v_{-i})] = 0$ for all $v_i \neq v_{\sup}$. Then, there exists a user that can receive a non-zero expected utility (in expectation over the outcomes of the mechanism and the values v of all users including i) only if the following conditions are met:*

   - $\mathbb{P}_{v_i \sim \mathcal{T}} [v_i = v_{\sup}] > 0$ *and* $\mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} [x_i(v_{\sup}, v_{-i})] > 0$

   - *The supremum of the distribution $\widetilde{\mathcal{T}}$ given by $\mathcal{T}$ conditioned on the draw being less than $v_{\sup}$. Then, the supremum $\tilde{v}_{\sup}$ of $\widetilde{\mathcal{T}}$ must be strictly less than $v_{\sup}$.*

We will unpack the second claim before we prove the lemma. The second claim is a partial converse to the first, and helps characterize all mechanisms where each user gets an expected utility equal to zero. Consider the distribution $\mathcal{T}$ given by

$$\mathop{\mathbb{P}}_{v_i \sim \mathcal{T}} [v_i = 0] = 1/2 \qquad\qquad \mathop{\mathbb{P}}_{v_i \sim \mathcal{T}} [v_i = 1] = 1/2$$

Conditioned on the draw from $\mathcal{T}$ not being equal to 1, its supremum, the resulting distribution is just the constant distribution at 0, whose supremum is indeed less than 1. If a user with value distributed according to $\mathcal{T}$ is posted a price $\frac{1}{2}$, he receives an expected positive utility. The second part in the lemma claims a user can potentially receive a non-zero utility only if the supremum of the distribution $\mathcal{T}$ has a non-zero probability of getting realized, and is not a limit point of the support of the distribution $\widetilde{\mathcal{T}}$.

Proof.   1. Suppose a user i with value v_i is included in the block with a non-zero probability $x_i(v_i)$. By monotonicity of the allocation rule, whenever the user has a value $\tilde{v}_i > v_i$, he has to be included with at least the probability $x_i(v_i)$. Further, if the user pretends to have a value v_i, he receives a utility

$$\tilde{v}_i \cdot x_i(v_i) - p_i(v_i) > v_i \cdot x_i(v_i) - p_i(v_i) \geq 0$$

where p_i is the interim payment rule for user i. Thus, user i gets strictly positive utility whenever he has a value larger than v_i. If he receives an expected utility of zero, then, the probability of realizing a value larger than v_i must be zero.

2. Suppose that $\mathbb{E}_{v_{-i} \sim \mathcal{T}^{n-1}} [x_i(v_{\sup}, v_{-i})] = 0$ for all $v_i \neq v_{\sup}$ and user i receives a strictly positive expected utility (in expectation over the values v of all users). Whenever i has a value strictly smaller than $v_{\sup}$, he gets included with probability zero. By individual rationality, i nets a zero utility. Thus, user i receives a positive utility when his value $v_i = v_{\sup}$. For his expected utility to be positive, he needs to make an expected payment p per unit probability of getting included, strictly smaller than $v_{\sup}$.

---

[28] As is standard, the supremum of a distribution $\mathcal{T}$ equals $\inf\{s : Pr_{x \sim \mathcal{T}}(x \leq s) = 1\}$

Assuming $Pr_{v_i \sim \mathcal{T}}(p < v_i < v_{\sup}) > 0$ for contradiction, then user i can bid $v_{\sup}$ to make an expected payment p per unit probability of getting included and receive a positive utility. This is a contradiction, since the mechanism is DSIC.

$\square$

From here, proving our main impossibility result becomes straightforward.

Theorem 6.5. *For a given distribution $\mathcal{T}$ of user values, and a sequence of on-chain simple, off-chain influence proof, and strongly collusion proof user Bayes-Nash equilibria $\left((s_{\mathsf{mi}}^{\mathcal{C}}, s_{\mathsf{usr},1}^{\mathcal{C},n}, \ldots, s_{\mathsf{usr},n}^{\mathcal{C},n}) \right)_{n \in \mathbb{N} \cup \{0\}}$. Then, for all $n \in \mathbb{N}$ and all users i, user i is allocated with probability zero whenever his value v_i is smaller than the supremum $v_{\sup}$ of $\mathcal{T}$.*

The partial converse in Lemma 6.4 also rules out any user being allocated with a positive probability when the supremum of the distribution does not have a point mass. Of course, most commonly occurring continuous distributions like the uniform distribution, the normal distribution, the exponential distribution, etc., satisfy the above property.

Theorem 6.6. *There can exist no prior-independent block-building process $\mathcal{B}$ such that for all distribution $\mathcal{T}$, there exist strategy profiles that satisfy on-chain simplicity, off-chain influence proofness and collusion proofness and also includes users with a positive probability.*

Remark 6.7. We believe an impossibility result similar to Theorem 6.6 can be achieved by assuming only on-chain miner simplicity, collusion-proofness, and off-chain influence-proofness. While users need not bid truthfully, the corresponding interim allocation and payment rules (Definition 2.3) are BIC. We believe that the BIC guarantee over the value allocation rule, coupled with on-chain miner simplicity, collusion proofness and off-chain influence proofness should lead to an impossibility result similar to Theorem 6.6.

7 Conclusion

We propose a set of desiderata for TFMs, and under these desiderata, users can be advised to simply bid their values, without worrying about either the bids placed by other users, or the strategy played by the miner, including any off-chain chatter. If a user BNE in a TFM is off-chain influence proof, even a miner with unchecked commitment power would not want to induce a different equilibrium since she is receiving the optimal revenue over all user BNEs. Moreover, analogous to UIC and MMIC, on-chain user simplicity guarantees a DSIC mechanism to users conditioned on the miner's *expected* behaviour, and on-chain miner simplicity says the miner will not deviate on-chain from the expected on-chain behaviour.

We suggest that TFMs consider the cryptographic $(k+1)^{\text{th}}$ auction, which satisfies on-chain simplicity and off-chain influence proofness, but not strong collusion proofness. This is in contrast to the truth-telling user BNE in EIP-1559, which satisfies on-chain simplicity and strong collusion proofness, but fails off-chain influence proofness. We prove that this tradeoff between strong collusion proofness and off-chain influence proofness is unavoidable, and argue that off-chain influence proofness has innate collusion resistance guarantees which may suffice by themselves. Still, even after our paper, the optimal TFM design should remain in consideration; for one, because the best explicitly cryptographic implementations of a second-price auction is not clear. Moreover, in the era of "MEV", novel new design considerations for TFMs may emerge beyond the ones we consider here (see e.g. Bahrani et al. (2023)), and we hope that our framework provides relevant insights into future investigations.

References

Ittai Abraham, Danny Dolev, Rica Gonen, and Joe Halpern. 2006. Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation. In *Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing.* 53–62. 6

Mohammad Akbarpour and Shengwu Li. 2020. Credible Auctions: A Trilemma. *Econometrica* 88, 2 (2020), 425–467. 2, 3, 16, 38

Nick Arnosti and S Matthew Weinberg. 2022. Bitcoin: A natural oligopoly. *Management Science* 68, 7 (2022), 4755–4771. 6

Maryam Bahrani, Pranav Garimidi, and Tim Roughgarden. 2023. Transaction fee mechanism design with active block producers. *arXiv preprint arXiv:2307.01686* (2023). 6, 31

Soumya Basu, David Easley, Maureen O'Hara, and Emin Gün Sirer. 2019. Towards a Functional Fee Market for Cryptocurrencies. *CoRR* abs/1901.06830 (2019). arXiv:1901.06830 http://arxiv.org/abs/1901.06830 5

Eric Budish. 2023. *Trust at scale: The economic limits of cryptocurrencies and blockchains.* Technical Report. Working Paper, U. of Chicago. 6

Jeremy Bulow and John Roberts. 1989. The simple economics of optimal auctions. *Journal of political economy* 97, 5 (1989), 1060–1090. 22

Vitalik Buterin, Eric Conner, Rick Dudley, Matthew Slipper, Ian Norden, and Abdelhamid Bakhta. 2019. EIP-1559: Fee market change for ETH 1.0 chain. Ethereum Improvement Proposals, no. 1559. https://eips.ethereum.org/EIPS/eip-1559 5, 19

Ran Canetti, Amos Fiat, and Yannai A Gonczarowski. 2023. Zero-Knowledge Mechanisms. *arXiv preprint arXiv:2302.05590* (2023). 6

Shuchi Chawla and Jason D Hartline. 2013. Auctions with unique equilibria. In *Proceedings of the fourteenth ACM conference on Electronic commerce.* 181–196. 24, 26, 39, 41, 42

Xi Chen, David Simchi-Levi, Zishuo Zhao, and Yuan Zhou. 2024. Bayesian mechanism design for blockchain transaction fee allocation. *arXiv preprint arXiv:2209.13099* (2024). 5

Hao Chung, Tim Roughgarden, and Elaine Shi. 2024. Collusion-resilience in transaction fee mechanism design. *EC 2024* (2024). arXiv preprint arXiv:2402.09321. 1, 5

Hao Chung and Elaine Shi. 2023. Foundations of Transaction Fee Mechanism Design. (2023), 3856–3899. https://doi.org/10.1137/1.9781611977554.ch150 arXiv:https://epubs.siam.org/doi/pdf/10.1137/1.9781611977554.ch150 1, 2, 3, 5, 17, 19, 26

Philip Daian, 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).* IEEE, 910–927. 5

Meryem Essaidi, Matheus VX Ferreira, and S Matthew Weinberg. 2022. Credible, strategyproof, optimal, and bounded expected-round single-item auctions for all distributions. *arXiv preprint arXiv:2205.14758* (2022). 3

Matheus VX Ferreira and S Matthew Weinberg. 2020. Credible, truthful, and two-round (optimal) auctions via cryptographic commitments. In *Proceedings of the 21st ACM Conference on Economics and Computation*. 683–712. 3

Matheus V. X. Ferreira, Daniel J. Moroz, David C. Parkes, and Mitchell Stern. 2021. Dynamic posted-price mechanisms for the blockchain transaction-fee market. In *AFT '21: 3rd ACM Conference on Advances in Financial Technologies, Arlington, Virginia, USA, September 26 - 28, 2021*, Foteini Baldimtsi and Tim Roughgarden (Eds.). ACM, 86–99. https://doi.org/10.1145/3479722.3480991 5

Yotam Gafni and Aviv Yaish. 2022. Greedy Transaction Fee Mechanisms for (Non-) myopic Miners. *arXiv preprint arXiv:2210.07793* (2022). 5

Yotam Gafni and Aviv Yaish. 2024. Barriers to collusion-resistant transaction fee mechanisms. *EC 2024* (2024). arXiv preprint arXiv:2402.08564. 1, 5

Joshua S Gans and Hanna Halaburda. 2024. "Zero Cost" Majority Attacks on Permissionless Proof of Work Blockchains. *Management Science* (2024). 6

Juan Garay, Jonathan Katz, Ueli Maurer, Björn Tackmann, and Vassilis Zikas. 2013. Rational protocol design: Cryptography against incentive-driven adversaries. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 648–657. 6

Joseph Halpern and Vanessa Teague. 2004. Rational secret sharing and multiparty computation. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*. 623–632. 6

Gur Huberman, Jacob Leshno, and Ciamac C. Moallemi. 2017. An Economic Analysis of the Bitcoin Payment System. *SSRN* (2017). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3025604 6

Sergei Izmalkov, Matt Lepinski, and Silvio Micali. 2011. Perfect implementation. *Games and Economic Behavior* 71, 1 (2011), 121–140. 6

Gillat Kol and Moni Naor. 2008. Cryptography and game theory: Designing protocols for exchanging information. In *Theory of cryptography conference*. Springer, 320–339. 6

Ron Lavi, Or Sattath, and Aviv Zohar. 2019. Redesigning Bitcoin's fee market. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*. 2950–2956. https://doi.org/10.1145/3308558.3313454 4

Stefanos Leonardos, Barnabé Monnot, Daniël Reijsbergen, Stratis Skoulakis, and Georgios Piliouras. 2021. Dynamical Analysis of the EIP-1559 Ethereum Fee Market. *CoRR* abs/2102.10567 (2021). arXiv:2102.10567 https://arxiv.org/abs/2102.10567 5

Stefanos Leonardos, Daniël Reijsbergen, Barnabé Monnot, and Georgios Piliouras. 2023. Optimality Despite Chaos in Fee Markets. In *Financial Cryptography and Data Security - 27th International Conference, FC 2023, Bol, Brač, Croatia, May 1-5, 2023, Revised Selected Papers, Part II (Lecture Notes in Computer Science, Vol. 13951)*, Foteini Baldimtsi and Christian Cachin (Eds.). Springer, 346–362. https://doi.org/10.1007/978-3-031-47751-5_20 5

Roger B. Myerson. 1981. Optimal Auction Design. *Mathematics of Operations Research* 6, 1 (1981), 58–73. 3, 4, 6, 7, 8, 9, 10, 21, 24, 26

Michal Penn, Maria Polukarov, and Moshe Tennenholtz. 2005. Congestion games with failures. In *Proceedings of the 6th ACM Conference on Electronic Commerce*. 259–268. 37

Tim Roughgarden. 2020. Transaction fee mechanism design for the Ethereum blockchain: An economic analysis of EIP-1559. *arXiv preprint arXiv:2012.00854* (2020). 5, 15, 16, 19, 20, 26, 44

Tim Roughgarden. 2021. Transaction fee mechanism design. *ACM SIGecom Exchanges* 19, 1 (2021), 52–55. Full version at https://arxiv.org/abs/2106.01340. 1, 2, 5, 15, 19

Elaine Shi and Hao Chung. 2023. What can cryptography do for decentralized mechanism design. *Innovations in Theoretical Computer Science* (2023). 12

Elaine Shi, Hao Chung, and Ke Wu. 2023. What can cryptography do for decentralized mechanism design. *Innovations in Theoretical Computer Science (ITCS)* (2023). 1, 3, 5

William Vickrey. 1961. Counterspeculations, Auctions, and Competitive Sealed Tenders. *Journal of Finance* 16, 1 (1961), 8–37. 21

Ke Wu, Elaine Shi, and Hao Chung. 2024. Maximizing Miner Revenue in Transaction Fee Mechanism Design. *ITCS* (2024). 1, 2, 5

Andrew Chi-Chih Yao. 2018. An Incentive Analysis of some Bitcoin Fee Designs. *CoRR* abs/1811.02351 (2018). arXiv:1811.02351 http://arxiv.org/abs/1811.02351 5

A A Zoo of Simplicity Definitions

A.1 Weaker Notions of Influence Proofness

In this appendix, we explore weaker versions of off-chain influence proofness. Our primary purpose is not to provide thorough investigations into any of these notions, but merely to emphasize that off-chain influence proofness gives a very strong garantee on the incentives of the miner: she cannot raise additional revenue, even if she is able to influence the way users behave in a wide variety of ways, so long as users play in a user BNE (consistent with the block-building process).

Recall that off-chain influence proofness stipulates that (i) the users and miner stay on-chain, while (ii) the miner gets her optimal revenue even when considering alternative miner-strategies that use nontrivial off-chain mechanisms. To capture the different alternative miner strategies (i.e., ways the miner may try to "game the system"), we focus on part (ii). Thus, we phrase this subsection as giving various *optimality* properties. See Figure 2 for a list of different sets of strategy profiles, $\Sigma^{\mathcal{C}}_{\text{Simple-and-Equilibrium}}, \Sigma^{\mathcal{C}}_{\text{Equilibrium}}, \Sigma^{\mathcal{C}}_{\text{On-Chain}}, \Sigma^{\mathcal{D}}_{\text{Off-Chain}}$. The set of equilibria respectively correspond to: truthful user BNEs in which the miner's strategy is on-chain, compliant, and a best-response to the users' strategies; arbitrary user BNEs in which the miner's strategy is on-chain and a best-response to the users' strategies; arbitrary user BNEs in which the miner's strategy is on-chain, but *need not* be a best response; and finally arbitrary user BNEs of arbitrary miner strategies in the off-chain game. (Note that $\Sigma^{\mathcal{C}}_{\text{Simple-and-Equilibrium}}$ may be empty if all equilibria of $\mathcal{C}$, but all other sets can be shown to be nonempty by general results.)

We consider optimality conditions for each such set, as follows:

$$\Sigma^{\mathcal{C}}_{\text{Simple-and-Equilibrium}} = \big\{ \sigma = (s^{\mathcal{C}}_{\text{mi}}, s^{\mathcal{C}}_{\text{usr},1}, \dots, s^{\mathcal{C}}_{\text{usr},n})$$

$$\big| \; \sigma \text{ is a user-BNE, } s^{\mathcal{C}}_{\text{mi}} \text{ is compliant and is the miner's best-response}$$

$$\text{to } (s_{\text{usr},1}, \dots, s_{\text{usr},n}), \text{ and } s_{\text{usr},i} \text{ is truth-telling and DSIC for each } i = 1, \dots, n \big\}$$

$$\text{I} \cap$$

$$\Sigma^{\mathcal{C}}_{\text{Equilibrium}} = \big\{ \sigma = (s^{\mathcal{C}}_{\text{mi}}, s^{\mathcal{C}}_{\text{usr},1}, \dots, s^{\mathcal{C}}_{\text{usr},n})$$

$$\big| \; \sigma \text{ is a user-BNE and } s^{\mathcal{C}}_{\text{mi}} \text{ is the miner's best-response to } (s_{\text{usr},1}, \dots, s_{\text{usr},n}) \big\}$$

$$\text{I} \cap$$

$$\Sigma^{\mathcal{C}}_{\text{On-Chain}} = \big\{ \sigma = (s^{\mathcal{C}}_{\text{mi}}, s^{\mathcal{C}}_{\text{usr},1}, \dots, s^{\mathcal{C}}_{\text{usr},n})$$

$$\big| \; \sigma \text{ is a user-BNE} \big\}$$

$$\text{I} \cap$$

$$\Sigma^{\mathcal{D}}_{\text{Off-Chain}} = \big\{ \sigma = (\mathcal{M}_{\text{off}}, s^{\mathcal{C}}_{\text{usr},1}, \dots, s^{\mathcal{C}}_{\text{usr},n})$$

$$\big| \; \sigma \text{ is a user-BNE and off-chain influence proof} \big\}$$

Figure 2: Sets of strategies a miner might be able to influence the equilibrium to.

Definition A.1. For each $E \in \{\text{Simple-and-Equilibrium, Equilibrium, On-Chain, Off-Chain}\}$, We say that a user-BNE σ (in either on on-chain or off-chain game) is *E-miner-optimal* if $\sigma \in \Sigma_E$, and moreover, the miner's revenue under σ is at least as high as the miner's revenue in any other strategy profile in Σ_E.

Observe that a strategy profile σ is off-chain influence proof if and only if it is Off-Chain-miner-optimal, and also, the miner's strategy in σ is completely on-chain. As mentioned in Section 4, we thus observe that off-chain influence proofness gives a stronger grantee than all of the other optimality notions we consider. Moreover, we observe that when σ is on-chain (miner and user) simple, and also off-chain influence proof, then the optimal miner revenue from $\Sigma^{\mathcal{C}}_{\text{Simple-and-Equilibrium}} \ni \sigma$ equals the optimal miner revenue from $\Sigma^{\mathcal{D}}_{\text{Off-Chain}}$. We believe this gives compelling reasons that everyone will "just follow σ"; perhaps more compelling than any previous set of desiderata.

A.2 Alternate Notions of Collusion Resistance

In this appendix, we explore various (*weaker*) notions of collusion resistance than the one given by Definition 4.5. All the notions discussed below capture collusion where agents in the cartel continue to remain strategic with each other even while colluding.

To begin with, we sketch an example of a mechanism that is not collusion proof (via Definition 4.5) but how agents in the cartel being strategic with each other could make colluding difficult. We consider a miner and a user colluding in a simple posted-price mechanism for a good with an unlimited supply, in which all payments are transferred to the miner (which we define formally in Section B.2). We argue that truthful bidding in the posted-price mechanism is not collusion-proof. For a price p posted by the mechanism, the miner can collude with a user i whose value v_i is less

than p, via the following "contract" of collusion. The miner recommends the user to bid p, thereby procuring the good. The miner keeps (say) $\frac{v_i}{2}$ from the on-chain payment made by the user and rebates the rest (i.e., transfers $\frac{v_i}{2}$). The user gets included and gets a positive utility even though his value is less than the price, and the miner is able to achieve a non-zero revenue by transacting from user i. Both the agents in the cartel are better off from colluding, which shows that this mechanism fails strong collusion proofness.

However, when colluding under the above (informally specified) contract of collusion, truthfully reporting their value is no longer an equilibrium for the users in the cartel. Colluding users would be incentivized to tell the miner that their value is very close to zero, even if their values are well above the price. This way, the miner would only dock a payment close to zero, which is far lesser than the price p the user would have to pay if he were truthful.

A.2.1 Trustless Collusion Proofness

For the following definition and the other alternate notion of collusion resistance in Section A.2.2, we find defining collusion resistance through the off-chain game much more natural than the on-chain game.[29]

Definition A.2 (1-1-Trustless Collusion Proofness)**.** For some off-chain game $\mathcal{D}$ and distribution $\mathcal{T}$, a user BNE $(\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ where $\mathcal{M}_{\mathsf{off}}$ is a trivial off-chain mechanism is 1-1-trustless collusion proof[30] if, for any off-chain mechanism $\widetilde{\mathcal{M}}_{\mathsf{off}}$ (called the *collusion contract*) and any user Bayes-Nash equilibrium $(\widetilde{\mathcal{M}}_{\mathsf{off}}, \widetilde{s}_{\mathsf{usr},1}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}, \ldots, \widetilde{s}_{\mathsf{usr},n}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}})$ in $\widetilde{\mathcal{M}}_{\mathsf{off}}$, at least one of the following conditions holds:

1. The miner prefers the the trivial mechanism $\mathcal{M}_{\mathsf{off}}$ over the collusion contract $\widetilde{\mathcal{M}}$.

$$\mathbb{E}_{v\sim\mathcal{T}}\left[\mathsf{Rev}^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))\right] \geq \mathbb{E}_{v\sim\mathcal{T}}\left[\mathsf{Rev}^{\mathcal{D}}(\widetilde{\mathcal{M}}_{\mathsf{off}}; \widetilde{s}_{\mathsf{usr}}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v))\right].$$

2. No user i, irrespective of his value v_i, prefers the collusion contract $\widetilde{\mathcal{M}}$ over the trivial mechanism $\mathcal{M}_{\mathsf{off}}$. Thus, for all users i,

$$\mathbb{E}_{v_{-i}\sim\mathcal{T}^{n-1}}\left[v_i \cdot X_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)) - P_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))\right]$$
$$\geq \mathbb{E}_{v_{-i}\sim\mathcal{T}^{n-1}}\left[v_i \cdot X_i^{\mathcal{D}}(\widetilde{\mathcal{M}}_{\mathsf{off}}; \widetilde{s}_{\mathsf{usr}}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v)) - P_i^{\mathcal{D}}(\widetilde{\mathcal{M}}; \widetilde{s}_{\mathsf{usr}}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v))\right].$$

Summarizing the above definition, a user Bayes-Nash equilibrium is trustless collusion proof if signing a collusion contract leads to an equilibrium that is adverse either to the miner or to the user colluding with the miner.

Suppose that a strategy profile $(\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ in the off-chain game $\mathcal{D}$ for some distribution $\mathcal{T}$ satisfies off-chain influence proof. Then, the miner prefers $(\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ over any other user BNE induced by any alternate off-chain mechanism. The first condition in the definition of trustless collusion proofness (Definition A.2) is automatically satisfied, leading to the following conclusion.

---

[29]Similar to off-chain influence proof, both weaker notions of collusion resistance can be viewed partly as a property of an on-chain user BNE since we claim no cartel containing the miner and a user would want to communicate off-chain. We will, therefore, abuse notation to say a user BNE in the on-chain game $\mathcal{C}$ satisfies trustless collusion proofness and weak collusion proofness (defined in Section A.2.2).

[30]Similar to collusion proofness (footnote 23), the above definition can be extended to a-b-trustless collusion proofness.

Corollary A.3. *If* $\sigma = (\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ *is off-chain influence proof for some off-chain game* $\mathcal{D}$ *and distribution* $\mathcal{T}$ *of values, then,* σ *is also trustless collusion proof.*

A.2.2 Weak Collusion Proofness

In this section, we address the issue of secrecy with trustless collusion proofness and the natural collusion resistance provided by off-chain influence proofness. Trustless collusion proofness models situations where users outside a cartel learn about the collusion and adjust their strategies to optimize their utilities in the subsequent off-chain game following the miner's announcement of the collusion contract. However, if the cartel's formation is kept secret, assuming all users adjust their strategies based on the collusion contract might not be accurate. To address this, we assume that agents outside the cartel cannot modify their actions based on the miner's announcement. Note that we define weak collusion proofness only for user BNE induced by trivial off-chain mechanisms. In this scenario, all users, particularly those outside the cartel, would play an abstaining bid. An abstaining bid is a valid strategy for users outside the cartel in any off-chain mechanism declared by the miner. Therefore, assuming users cannot modify their strategies based on the miner's announcement does not contradict the syntax of the framework.

Definition A.4 (1-1-Weak Collusion Proofness)**.** Let $\mathcal{D}$ be the off-chain game induced by some block-building process for some distribution $\mathcal{T}$ of user values. A strategy profile $(\mathcal{M}_{\mathsf{off}}, s_{\mathsf{usr},1}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}, \ldots, s_{\mathsf{usr},n}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}})$ in user BNE for a trivial off-chain mechanism $\mathcal{M}_{\mathsf{off}}$ is 1-1-weak collusion proof (or weak collusion proof) if for all off-chain mechanisms $\widetilde{\mathcal{M}}_{\mathsf{off}}$ and all users i, the following property holds. Consider any strategy $\widetilde{s}_{\mathsf{usr},1}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}$ for i in $\widetilde{\mathcal{M}}_{\mathsf{off}}$ which is a best response for user i to $\widetilde{\mathcal{M}}_{\mathsf{off}}$ conditioned on other users being oblivious to the non-trivial off-chain component and playing the strategy $s_{\mathsf{usr},-i}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}$. In other words, suppose that for all strategies $\hat{s}_{\mathsf{usr},i}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}$ for user i, we have

$$\mathbb{E}_{v_{-i}\sim\mathcal{T}^{n-1}}\left[v_i \cdot X_i^{\mathcal{D}}(\widetilde{\mathcal{M}_{\mathsf{off}}}; \widetilde{s}_{\mathsf{usr},i}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v_i), s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v_{-i})) - P_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; \widetilde{s}_{\mathsf{usr},i}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v_i), s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v_{-i}))\right]$$

$$\geq \mathbb{E}_{v_{-i}\sim\mathcal{T}^{n-1}}\left[v_i \cdot X_i^{\mathcal{D}}(\widetilde{\mathcal{M}_{\mathsf{off}}}; \hat{s}_{\mathsf{usr},i}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v_i), s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v_{-i})) - P_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; \hat{s}_{\mathsf{usr},i}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}, s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v_{-i}))\right].$$

Then, user i and miner cannot both receive higher utility under the colluding strategies; formally, one of the following two properties must hold:

1. the miner has a higher expected revenue in $\mathcal{M}_{\mathsf{off}}$ than in $\widetilde{\mathcal{M}}_{\mathsf{off}}$, i.e,

$$\mathbb{E}_{v\sim\mathcal{T}}\left[\mathsf{Rev}^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))\right] \geq \mathbb{E}_{v\sim\mathcal{T}}\left[\mathsf{Rev}^{\mathcal{D}}(\widetilde{\mathcal{M}}_{\mathsf{off}}; \widetilde{s}_{\mathsf{usr},i}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v_i), s_{\mathsf{usr},-i}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v_{-i}))\right],$$

2. or, user i has a higher expected utility in $\mathcal{M}_{\mathsf{off}}$ than in $\widetilde{\mathcal{M}}_{\mathsf{off}}$, i.e,

$$\mathbb{E}_{v_{-i}\sim\mathcal{T}^{n-1}}\left[v_i \cdot X_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v)) - P_i^{\mathcal{D}}(\mathcal{M}_{\mathsf{off}}; s_{\mathsf{usr}}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v))\right]$$

$$\geq \mathbb{E}_{v_{-i}\sim\mathcal{T}^{n-1}}\left[v_i \cdot X_i^{\mathcal{D}}(\widetilde{\mathcal{M}}_{\mathsf{off}}; \widetilde{s}_{\mathsf{usr},i}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v_i), s_{\mathsf{usr},-i}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v_{-i})) - P_i^{\mathcal{D}}(\widetilde{\mathcal{M}}; \widetilde{s}_{\mathsf{usr},i}^{\mathcal{D},\widetilde{\mathcal{M}}_{\mathsf{off}}}(v_i), s_{\mathsf{usr},-i}^{\mathcal{D},\mathcal{M}_{\mathsf{off}}}(v_{-i}))\right].$$

Weak collusion proofness is similar to the notion of a semi-strong Nash equilibrium (Penn et al., 2005). A Nash equilibrium is semi-strong if fixing the strategies of all agents not in the coalition and for any strategy played by agents in the collusion, at least one of the following two conditions are satisfied:

1. At least one of the agents in the coalition achieve a larger utility prior to colluding that after.

2. If all agents in the coalition bag a larger utility, at least one colluding agent has an incentive to betray the cartel and deviate further, increasing their individual utility.

To our knowledge, weak collusion proofness is not a straightforward consequence of off-chain influence proofness. While off-chain influence proofness compares the miner's revenue in two user BNEs, weak collusion proofness compares the miner's (and the colluding user's) revenue between a user BNE, and a deviation from the user BNE just for the miner and the colluding user. Other users' strategy profiles need not be an equilibrium given the new strategies played by the miner and the colluding user. On the other hand, collusion proofness implies weak collusion proofness.

Truthful bidding by users and the miner setting the monopoly reserve (a_{mi} such that the virtual value $\phi(a_{\mathsf{mi}}) = 0$) in the cryptographic $(k + 1)^{\mathrm{th}}$ price auction for a regular distribution $\mathcal{T}$ is an example of a user Bayes-Nash equilibrium that is not collusion proof, but satisfies weak collusion proofness (Proposition 5.8).

B Evaluating Other Mechanisms

In this section, we run over various standard mechanisms like the posted-price mechanism and the (plaintext) $(k + 1)^{\mathrm{th}}$ price auction. We also provide examples like the bonus-on-matching-bids auction and revisit SR2PA (Section 3.2), which along with the examples discussed in Section 5 help separate the three core definitions in Section 4—on-chain user simplicity, on-chain miner simplicity, and off-chain influence proofness—i.e., show each property is logically independent. See Table 1.

B.1 The Plaintext $(k + 1)^{\mathrm{th}}$ Price Auction

The *plaintext $(k + 1)^{\mathrm{th}}$ price auction* (P$(k + 1)$PA) is identical to C$(k + 1)$PA except that the bids are submitted in the plaintext model.

Proposition B.1. *For all regular distributions $\mathcal{T}$ with a monopoly reserve a_{mi}, the strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), v_1, \ldots, v_n)$ in the on-chain game $\mathcal{C}$ induced by P$(k+1)$PA is on-chain user simple and off-chain influence proof. However, $\sigma^{\mathcal{C}}$ is not necessarily on-chain miner simple and collusion proof for all regular distributions $\mathcal{T}$.*

The proofs for on-chain user simplicity, off-chain influence proofness and lack of collusion proofness are identical to the analogous claims for C$(k + 1)$PA. Note that off-chain influence proofness does not depend on the cryptographic model. Since the miner commits to her behaviour in the off-chain game and users best respond to the miner's commitment, a user BNE is always induced in the off-chain game. Since $\sigma^{\mathcal{C}}$ is the revenue optimal equilibrium, the miner would not want to induce a different equilibrium in the off-chain game.

On-chain miner simplicity on the other hand depends on the cryptographic model. By being able to look at the bids before playing her action, the miner can fabricate bids to increase her revenue (Akbarpour and Li, 2020). Given a set of bids b, the miner sets the reserve $\tilde{a}_{\mathsf{mi}} = \arg\max_{1 \le i \le k} i \cdot b^{(i)}$, where $b^{(i)}$ is the ith-highest bid. The miner gets a revenue equal to $\max_{1 \le i \le k} i \cdot b^{(i)} \ge k \cdot b^{(k)} \ge k \cdot b^{(k+1)}$, the miner's revenue if all k slots get filled. Even if only j bids are larger than the monopoly reserve a_{mi} (i.e, $b^{(j)} \ge a_{\mathsf{mi}} \ge b^{(j+1)}$), $\max_{1 \le i \le k} i \cdot b^{(i)} \ge k \cdot b^{(k)} \ge j \cdot b^{(j)} \ge j \cdot a_{\mathsf{mi}}$.

Before we conclude the section, we showcase a different user Bayes-Nash equilibrium in the on-chain game when $k = 1$. Suppose that the miner sets the reserve just below the largest bid and users best respond to the above miner's strategy. The highest bid wins and since the reserve is (just

Mechanism	Off-Chain Influence Proof	On-Chain	
		User Simple	Miner Simple
C2PA	✓	✓	✓
EIP-1559	✗	✓	✓
P2PA	✓	✓	✗
BoMB, posted-price equilibrium	✓	✗	✓
Winner-pays-bid mechanism, $\sigma_0^{\mathcal{C}}$	✓	✗	✗
SR2PA, second-price equilibrium	✗	✓	✗
BoMB, winner-pays-bid equilibrium	✗	✗	✓
SR2PA, first-price equilibrium	✗	✗	✗

Table 1: Different auctions and equilibria satisfying all combinations of our core desiderata.
Notes: The block-building processes of C2PA (Section 5.2), P2PA (Section B.1), the winner-pays-bid mechanism (Section 5.3) and SR2PA (Section B.4) allow the miner to set a reserve on-chain, while EIP-1559 (Section 5.1) and BoMB (Section B.3) admit no advice from the miner. In C2PA, P2PA, the winner-pays-bid mechanism and the second-price equilibrium on SR2PA (Section B.4.1) we consider the miner's compliant strategy of setting the optimal reserve. In both equilibria of BoMB and EIP-1559, we consider the miner playing the compliant strategy by giving no advice, while in the first-price equilibrium of SR2PA (Section B.4.2), we consider the miner fabricating a fake bid just below the highest bid. In C2PA, EIP-1559, P2PA and the second-price equilibrium of SR2PA we consider the truth-telling strategy of users. In the winner-pays-bid mechanism, the winner-pays-bid equilibrium in BoMB (Section B.3.2) and SR2PA, we consider user shading their bid optimally according to the equilibrium in the winner-pays-bid mechanism from classical settings (Chawla and Hartline, 2013). In the posted-price equilibrium in BoMB (Section B.3.1), users willing to get included bid the reserve price while the rest bid zero.

below) the highest bid, the winner pays his bid. Thus, the mechanism reduces to the first-price auction. Thus, the user BNE equilibrium discussed in the winner-pays-bid mechanism (Section 5.3) is induced on-chain.

B.2 The Posted-Price Mechanism

We discuss the posted-price mechanism for a block with unlimited supply. The posted-price mechanism differs from EIP-1559 in two ways—the miner sets the price a_{mi} for inclusion in the block as part of her advice, and all payments collected from users are transferred to the miner, as opposed to getting burnt. We discuss the posted-price mechanism in both the plaintext and the miner-gatekeeper model.

Definition B.2 (Posted-Price Mechanism with Unlimited Supply). The block-building algorithm $\mathcal{B}$ of the posted-price mechanism is as follows:

- The miner informs the mechanism of the price a_{mi} she wants to post. $\mathsf{Adv}_{\mathsf{mi}} = \mathbb{R}_{\geq 0}$.

- $\mathsf{Bid}_{\mathsf{usr}} = \mathbb{R}_{\geq 0}$. Bids consist of a non-negative real number each. The *plaintext posted-price mechanism* elicits bids in the plaintext model, while the *cryptographic posted-price mechanism* collects bids in the miner-gatekeeper model.

- All bids greater than the price a_{mi} are included in the block. For a set of bids b,

$$X_i(a_{\mathsf{mi}}; b_i, b_{-i}) = \begin{cases} 1 & \text{if } b_i \geq a_{\mathsf{mi}} \\ 0 & \text{otherwise} \end{cases}$$

- All included bids pay the price a_{mi} posted by the mechanism. In other words, for a set of bid b submitted to the mechanism,

$$P_i(a_{\mathsf{mi}}; b_i, b_{-i}) = \begin{cases} p & \text{if } b_i \geq a_{\mathsf{mi}} \\ 0 & \text{otherwise} \end{cases}$$

- All payments made by the users are transferred to the miner.

$$\mathsf{Rev}(a_{\mathsf{mi}}; b) = \sum_i P_i(a_{\mathsf{mi}}; b)$$

We will analyze the simplicity notions of the strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), v_1, \ldots, v_n)$ for the monopoly reserve a_{mi} for regular distributions $\mathcal{T}$ of user values in both the cryptographic models.

B.2.1 The Cryptographic Posted-Price Mechanism

Proposition B.3. *The strategy profiles $\sigma^{\mathcal{C}}$ is on-chain simple and off-chain influence proof for all regular distributions $\mathcal{T}$. However, there exist regular distributions $\mathcal{T}$ such that $\sigma^{\mathcal{C}}$ is not strongly collusion proof.*

We observe that the cryptographic posted-price mechanism is identical to the cryptographic $(k+1)^{\text{th}}$ price auction when $k = \infty$. Thus, arguments similar to the proofs of Theorem 5.6 will prove on-chain simplicity and off-chain influence proofness of the cryptographic posted-price mechanism.

The argument that $\sigma^{\mathcal{C}}$ is not strongly collusion proof is much simpler. For a given price $a_{\text{mi}} > 0$, consider a user i with value $v_i < a_{\text{mi}}$. The miner and the user can increase their joint utility by the user bidding a_{mi} on-chain. By not colluding, the joint utility of the cartel from i's transaction equals zero, while the joint utility increases to v_i if the user bids a_{mi}. The revenue from other users remains constant irrespective user i's behaviour. Thus, $\sigma^{\mathcal{C}}$ is not strongly collusion proof.

B.2.2 The plaintext posted-price mechanism

Proposition B.4. *The strategy profiles $\sigma^{\mathcal{C}}$ is on-chain user simple and off-chain influence proof for all regular distributions $\mathcal{T}$. However, there exist regular distributions $\mathcal{T}$ such that $\sigma^{\mathcal{C}}$ is not on-chain miner simple and not collusion proof.*

Similar to the connection between the cryptographic posted-price mechanism and C$(k+1)$PA, the plaintext posted-price mechanism is the plaintext $(k+1)^{\text{th}}$ price auction (Section B.1) for $k = \infty$. Thus, $\sigma^{\mathcal{C}}$ is on-chain user simple and off-chain influence proof but not on-chain miner simple. The proof that $\sigma^{\mathcal{C}}$ is not collusion proof is identical to the analogous claim for the cryptographic posted-price mechanism.

B.3 Bonus-on-Matching-Bids (BoMB) Auction

The *Bonus-on-Matching-Bids* (BoMB) auction (a variant of an auction originally proposed in Chawla and Hartline (2013)) allows two different looking equilibria— one similar to the winner-pays-bid mechanism and the other similar to the posted-price mechanism.

At a high level, when the largest bid is unique, the BoMB auction is identical to the winner-pays-bid mechanism for $k = 1$ (a.k.a the first-price auction). However, if the largest bid is not unique, all users that placed the largest bid get included. Even though the mechanism is similar to the winner-pays-bid mechanism and all included users pay their bid, we argue that every user with a value greater than the reserve price would bid the reserve price in equilibrium, inducing an equilibrium similar to the posted-price mechanism with an unlimited supply.

Definition B.5 (Bonus-on-Matching-Bids Auction)**.** The block building process $\mathcal{B}$ of the bonus-on-matching-bids auction is as follows:

- $\mathcal{B}$ takes no advice from the miner and thus, $\text{Adv}_{\text{mi}} = \{\emptyset\}$. We will assume that $\mathcal{B}$ knows the monopoly reserve r of the regular distribution $\mathcal{T}$ of user values.

- $\text{Bid}_{\text{usr}} = \mathbb{R}_{\geq 0}$. Bids are submitted in the miner-gatekeeper model.

- Allocate all users with the largest bid $b_{\max}$, if $b_{\max} \geq \text{r}$. Formally, for a set of bids b such that $b_{\max} = \max_i b$ is the highest bid,

$$X_i(\emptyset; b_i, b_{-i}) = \begin{cases} 1 & \text{if } b_i \geq \text{r} \text{ and } b_i = b_{\max} \\ 0 & \text{otherwise} \end{cases}$$

- All included users pay their bid. For a set of bids b,

$$P_i(\emptyset; b_i, b_{-i}) = \begin{cases} b_i & \text{if } b_i \geq \text{r} \text{ and } b_i = b_{\max} \\ 0 & \text{otherwise} \end{cases}$$

- All collected payments are transferred to the miner.

$$\text{Rev}(\emptyset; b_i, b_{-i}) = \sum_i P_i(\emptyset; b_i, b_{-i})$$

B.3.1   The Posted-Price Equilibrium

First, we consider user strategies that will induce an equilibrium similar to the posted-price mechanism with an unlimited supply. Consider user i playing the strategy $s_{\mathsf{usr},\mathsf{pp}}^{\mathcal{C},\mathsf{r}}$ given by

$$s_{\mathsf{usr},\mathsf{pp}}^{\mathcal{C},\mathsf{r}}(v_i) = \begin{cases} \mathsf{r} & \text{if } v_i \geq \mathsf{r} \\ 0 & \text{otherwise} \end{cases}$$

We analyze the strategy profile $\sigma_{\mathsf{pp}}^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset), s_{\mathsf{usr},\mathsf{pp}}^{\mathcal{C},\mathsf{r}}, \ldots, s_{\mathsf{usr},\mathsf{pp}}^{\mathcal{C},\mathsf{r}})$. Getting included would yield a negative utility to users with a value smaller than the reserve r. All users with value at least r get included in $\sigma_{\mathsf{pp}}^{\mathcal{C}}$. Bidding higher would only result in a larger payment while bidding any lower than r will get them excluded. Thus, $\sigma_{\mathsf{pp}}^{\mathcal{C}}$ is a user BNE.

Since users are not bidding truthfully, $\sigma_{\mathsf{pp}}^{\mathcal{C}}$ is not on-chain user simple. From the miner's viewpoint, fabricating a bid larger than r drive the miner revenue to zero, while censoring bids or fabricating bids below the reserve does not help the miner's revenue either. Thus, $\sigma_{\mathsf{pp}}^{\mathcal{C}}$ is on-chain miner simple.

The proof of off-chain influence proofness is identical to the cryptographic posted-price mechanism, since the miner already receives the optimal revenue over all possible BNEs and mechanisms for an unlimited supply. A manipulation by a colluding cartel similar to the one described in the cryptographic posted-price mechanism (Section B.2.1) shows $\sigma_{\mathsf{pp}}^{\mathcal{C}}$ is not strongly collusion proof. The colluding user with value less than r can bid r instead of zero to increase the joint utility of the cartel.

Proposition B.6. *The on-chain strategy profile $\sigma_{\mathsf{pp}}^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset), s_{\mathsf{usr},\mathsf{pp}}^{\mathcal{C},\mathsf{r}}, \ldots, s_{\mathsf{usr},\mathsf{pp}}^{\mathcal{C},\mathsf{r}})$ in $\mathcal{C}$ induced by the BoMB auction satisfies on-chain miner simplicity, off-chain influence proofness but not on-chain user simplicity or strong collusion proofness.*

B.3.2   The Winner-Pays-Bid Equilibrium

In the section on first-price auctions (winner-pays-bid mechanisms for $k = 1$; Section 5.3), we discussed the existence of an equilibrium strategy for users $s_{\mathsf{usr},\mathsf{val}}^{\mathcal{C},\mathcal{T},\mathsf{r}}$, for a given number of users n and the distribution $\mathcal{T}$ with monopoly reserve r. We discuss properties satisfied by the strategy profile $\sigma_{\mathsf{wpb}}^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset), s_{\mathsf{usr}}^{\mathcal{C},\mathcal{T},\mathsf{r}}, \ldots, s_{\mathsf{usr}}^{\mathcal{C},\mathcal{T},\mathsf{r}})$ in the on-chain game of the BoMB auction.

We will argue that $\sigma_{\mathsf{wpb}}^{\mathcal{C}}$ is a user BNE for most common regular distributions like $U[0,1]$ which do not have point masses. The strategy $s_{\mathsf{usr},\mathsf{val}}^{\mathcal{C},\mathcal{T},\mathsf{r}}$ is injective when the distribution $\mathcal{T}$ has no point mass (Chawla and Hartline, 2013). Thus, the probability of any arbitrary bid b_i of user i matching the bid b_j equals zero (particularly in the miner-gatekeeper model, when users cannot see each other's bids). Any deviation by users will fetch the user same utility as the deviation in the first-price auction. Since $\sigma_{\mathsf{wpb}}^{\mathcal{C}}$ is a user BNE in the first-price auction with reserve r, it is also a user BNE in the BoMB auction.

Arguing other properties are fairly straightforward. $\sigma_{\mathsf{wpb}}^{\mathcal{C}}$ is not on-chain user simple, but satisfies on-chain user simplicity. Censoring bids will only decrease the miner's revenue, while fabricating bids will not extract a larger payment from included users either. $\sigma_{\mathsf{wpb}}^{\mathcal{C}}$ is not strongly collusion proof for the same reason as the first-price auction (Proposition 5.11).

Finally, $\sigma_{\mathsf{wpb}}^{\mathcal{C}}$ is not off-chain influence proof. The miner only needs to communicate to the users to steer them towards $\sigma_{\mathsf{pp}}^{\mathcal{C}}$, which yields a larger expected revenue. We summarize the properties in the proposition below.

Proposition B.7. *The strategy profile $\sigma_{\mathsf{wpb}}^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset), s_{\mathsf{usr,val}}^{\mathcal{C},\mathcal{T},\mathsf{r}}, \ldots, s_{\mathsf{usr,val}}^{\mathcal{C},\mathcal{T},\mathsf{r}})$ induced by the BoMB auction is on-chain miner simple but does not satisfy on-chain user simplicity, strong collusion proofness and off-chain influence proofness.*

Remark B.8. Even though $\sigma_{\mathsf{wpb}}^{\mathcal{C}}$ is not off-chain influence proof, the miner's optimal strategy in the off-chain game is to steer users away from $\sigma_{\mathsf{wpb}}^{\mathcal{C}}$ towards $\sigma_{\mathsf{pp}}^{\mathcal{C}}$, which is also an entirely on-chain equilibrium. $\sigma_{\mathsf{wpb}}^{\mathcal{C}}$ is therefore, not on-chain-miner-optimal. The miner's optimal off-chain strategy does not involve payments and only assists in selecting the optimal amongst all equilibria in the on-chain game. Thus, the "consequence" of having a off-chain deviation from the miner is not as bad as the "consequence" in EIP-1559, where the miner runs a separate auction altogether. On-chain optimality helps tease apart this distinction of coordinating equilibrium through off-chain messages and running an entire auction off-chain.

B.4   The Squared Revenue Second-Price Auction, Revisited

We revisit the squared revenue second-price auction (SR2PA) described in Section 3.2. As a quick recap, (assuming the distribution $\mathcal{T}$ is supported on $[0, 1]$), the SR2PA implements the plaintext second-price auction, allowing the miner to set the reserve. However, instead of transferring the entire payment made by the included user to the miner, the SR2PA only transfers the square of the payment (which is smaller than the payment since users have values and therefore bids supported in $[0, 1]$) and burns the rest.

We consider two strategy profiles. The first where the miner sets the monopoly reserve and users bid truthfully and the second where the miner fabricates a bid just below the highest bid, reducing the on-chain allocation and payments rule into a first-price auction (winner-pays-bid mechanism with $k = 1$).

B.4.1   The Second-Price equilibrium

We analyze the properties satisfied by the strategy profile $\sigma_{\mathsf{2PA}}^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(a_{\mathsf{mi}}), v_1, \ldots, v_n)$, where a_{mi} is the monopoly reserve of the distribution $\mathcal{T}$.

Conditioned on the miner playing $\mathsf{comp}(a_{\mathsf{mi}})$, users respond to a second-price auction with a reserve on-chain and thus, truthful bidding is a user BNE and the on-chain mechanism induced is also DSIC, and thus, on-chain user simple. Indeed, it is not on-chain miner simple since the miner can increase her revenue by injecting a fake bid just below the largest bid. It is also not collusion proof due to the simple deviation described next. For the monopoly reserve a_{mi} and a user i with value $v_i \in [a_{\mathsf{mi}} - a_{\mathsf{mi}}^2, a_{\mathsf{mi}}]$, the miner advices the block-building process to set a reserve zero. and i bids a_{mi}. When a different user has a value larger than a_{mi}, the allocation and payments are not different between the miner setting a reserve a_{mi} or faking a reserve a_{mi} by having user i bid a_{mi}. However when all users have a value less than a_{mi}, instead of not including any user and obtaining a zero joint utility, the user i gets included to bag a utility $v_i - a_{\mathsf{mi}} \geq a_{\mathsf{mi}} - a_{\mathsf{mi}}^2 - a_{\mathsf{mi}} = -a_{\mathsf{mi}}^2$. The miner gets a revenue a_{mi}^2, and thus, the joint utility of the cartel is at least zero.

We argued that $\sigma_{\mathsf{2PA}}^{\mathcal{C}}$ is not off-chain influence proof in Section 3.2. At a high level, the miner runs a second-price auction off-chain and includes the winner in the block (censoring all other bids). The winner only has to bid and pay zero on-chain. Thus, even though the total expected payments made by the users remains the same, the miner increases her revenue by saving on the money that gets burnt on-chain.

Proposition B.9. *Given a distribution supported over $[0, 1]$ with monopoly reserve a_{mi}, the user*

BNE $\sigma_{2PA}^{\mathcal{C}} = (s_{mi}^{comp}(a_{mi}), v_1, \ldots, v_n)$ *in the on-chain game* $\mathcal{C}$ *is on-chain user simple, but not on-chain miner simple or collusion proof. The corresponding user BNE* $\sigma_{2PA}^{\mathcal{D}} = (\mathcal{M}_{off}^{triv,a_{mi}}, \perp_{off}(v_1), \ldots, \perp_{off}(v_n))$ *in the off-chain game is not off-chain influence proof.*

B.4.2 The First-Price Equilibrium

Consider the miner setting the reserve to be just below the maximum bid b_{max} (for convenience, we pretend the miner sets the reserve to be equal to the maximum bid). Thus, the largest bidders always gets included and has to pay his bid. Thus, the on-chain game is reduced to the winner-pays-bid mechanism for $k = 1$ with reserve zero. For the optimal strategy $s_{usr}^{\mathcal{C},\mathcal{T},0}$ for users in the winner-pays-bid mechanism (Section 5.3), we consider the user BNE $\sigma_{wpb}^{\mathcal{C}} = ((b_{max}, H, \emptyset), s_{usr}^{\mathcal{C},\mathcal{T},0}, \ldots, s_{usr}^{\mathcal{C},\mathcal{T},0})$.

$\sigma_{wpb}^{\mathcal{C}}$ is neither on-chain user simple, nor is it on-chain miner simple (since the miner is not playing a compliant strategy). It is not collusion proof and off-chain influence proof for reasons identical to the second-price equilibrium $\sigma_{2PA}^{\mathcal{C}}$.

Proposition B.10. *Given a distribution supported over* $[0, 1]$, *the user BNE* $\sigma_{wpb}^{\mathcal{C}} = ((b_{max}, H, \emptyset), s_{usr}^{\mathcal{C},\mathcal{T},0}, \ldots, s_{usr}^{\mathcal{C},\mathcal{T},0})$ *in the on-chain game* $\mathcal{C}$ *and the corresponding strategy profile* $\sigma_{wpb}^{\mathcal{D}} = (\mathcal{M}_{off}^{triv,(b_{max},H,\emptyset)}, \perp_{off}(s_{usr}^{\mathcal{C},\mathcal{T},0}), \ldots, \perp_{off}(s_{usr}^{\mathcal{C},\mathcal{T},0}))$ *satisfies none of the four properties: on-chain user simplicity, on-chain miner simplicity, collusion proofness and off-chain influence proofness.*

C Omitted Proofs

C.1 On-Chain Simplicity and Collusion Proofness in EIP-1559

We now prove Theorem 5.3, directly adapting the arguments from Roughgarden (2020) to our setting.

As mentioned in Section 5.1, the first step is to verify that $\sigma^{\mathcal{C}}$ is indeed a user BNE for all distributions $\mathcal{T}$ in both the on-chain game $\mathcal{C}$ and the off-chain game $\mathcal{D}$.

Lemma C.1. $\sigma^{\mathcal{C}} = (s_{mi}^{comp}(\emptyset), v_1, \ldots, v_n)$ *is a user BNE in both, the on-chain game* $\mathcal{C}$ *and off-chain game* $\mathcal{D}$ *induced by EIP-1559 for any distribution* $\mathcal{T}$.

Proof. We verify that $\sigma^{\mathcal{C}}$ is a user Bayes-Nash equilibrium in $\mathcal{C}$. The proof for $\mathcal{D}$ is analogous.

Observe that the allocation and payments made by user i is independent of the other bids submitted to the mechanism. Consider a value v_i larger than the price p. Placing any bid larger than the price p will lead to getting included and making a payment equal to p. The user bags a utility equal to $v_i - p \geq 0$, which is larger than the zero utility from bidding below the price and getting excluded. Similarly, for a value v_i smaller than the price p, bidding below the price leads to a zero utility, while bidding above the price leads to a utility $v_i - p < 0$. Bidding v_i optimizes user i's utility in both cases, as required. □

Next, we show that $\sigma^{\mathcal{C}}$ is on-chain simple.

Lemma C.2. *Consider the on-chain game* $\mathcal{C}$ *induced by EIP-1559 for some distribution* $\mathcal{T}$. *The strategy profile* $\sigma^{\mathcal{C}}$ *is on-chain simple.*

Proof. We have verified that $\sigma^{\mathcal{C}}$ is a user Bayes-Nash equilibrium in Lemma C.1. On-chain user simplicity follows, since all users bid truthfully in $\sigma^{\mathcal{C}}$. The miner will not be able to receive any

revenue on-chain irrespective of the set of fake bids or censored bids. The miner could potentially end up burning her own money if she fabricates bids larger than the price posted by the mechanism. Thus, the miner optimizes her revenue by playing $s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset)$. □

Finally, we verify that $\sigma^{\mathcal{C}}$ is collusion proof.

Lemma C.3. *The strategy profile $\sigma^{\mathcal{C}} = (s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset), v_1, \ldots, v_n)$ is collusion proof for the on-chain game $\mathcal{C}$ induced by EIP-1559 for all distributions $\mathcal{T}$.*

Proof. Suppose that the miner colludes with a user i with value v_i. Observe that user i's inclusion in the block and utility is invariant to any fake bids injected by the miner or any bids censored by the miner (as long as the miner does not censor i's bid). As argued in the proof of on-chain miner simplicity (Lemma C.2), the miner will not be able to increase her revenue by injecting fake bids, nor would it increase user i's utility. Thus, without loss of generality, assume that the miner does not fabricate bids or censor bids of all users apart from i.

Irrespective of whether the miner includes or censors i's bid, the miner always gets a zero revenue. Thus, the joint utility of the coalition equals user i's utility. If the user has a value v_i larger than the price p, i's utility is optimized by bidding truthfully and getting included. Similarly, if the user has a value smaller than p, i's utility is optimized by bidding truthfully and getting excluded.

Playing $s_{\mathsf{mi}}^{\mathsf{comp}}(\emptyset)$ and v_i optimizes the joint utility of the coalition, implying $\sigma^{\mathcal{C}}$ is collusion proof for $\mathcal{C}$ for all distributions $\mathcal{T}$. □

Lemma C.1, Lemma C.2 and Lemma C.3 conclude the proof of Theorem 5.3.

C.2   The $\mathbf{C}(k+1)\mathbf{PA}$ is DSIC

Lemma C.4. *For a given reserve a_{mi}, bidding truthfully is utility optimizing for users irrespective of the strategies employed by the other users.*

Proof. Fix the bids b_{-i} of users other than i. For a value v_i for user i larger than the reserve a_{mi}, and amongst the k largest bids, we will prove that i optimizes his utility by bidding v_i. The proof of the other cases are analogous.

By being amongst the k largest bids and the reserve, user i will get included by bidding v_i. He would need to make payment $P_i(v_i, b_{-i}) = \max\{a_{\mathsf{mi}}, b^{(k+1)}\}$ and therefore, bag a utility $v_i - \max\{a_{\mathsf{mi}}, b^{(k+1)}\} \geq 0$.

By placing a bid b_i larger than v_i, user i continues to get allocated and still have to pay $P_i(b_i, b_{-i}) = \max\{a_{\mathsf{mi}}, b^{(k+1)}\}$, resulting in the same utility as bidding v_i. Under-bidding by placing a bid b_i less than v_i can result in two outcomes — where i continues winning the good and where i does not win the good. If i wins the good, he still has to make a payment equal to $P_i(b_i, b_{-i}) = \max\{a_{\mathsf{mi}}, b^{(k+1)}\}$. If i does not win the good anymore, he gets a utility equal to zero, which is worse than the utility from truthful bidding.

Thus, each user optimizes his utility by bidding truthfully. □