Transaction Pricing for Maximizing Throughput in a Sharded Blockchain Ledger

James R. Riehl * **Jonathan Ward** *

* *Fetch.ai, St. John's Innovation Centre, Cowley Road, Cambridge, CB4 0WS, UK (e-mail: {james.riehl, jonathan.ward*

@fetch.ai).}

Abstract: In this paper, we present a pricing mechanism that aligns incentives of agents who exchange resources on a decentralized ledger with the goal of maximizing transaction throughput. Subdividing a blockchain ledger into shards promises to greatly increase transaction throughput with minimal loss of security. However, the organization and type of the transactions also affects the ledger's efficiency, which is increased by wallet agents transacting in a single shard whenever possible while collectively distributing their transactions uniformly across the available shards. Since there is no central authority to enforce these properties, the only means of achieving them is to design the system such that it is in agents' interest to act in a way that benefits overall throughput. We show that our proposed pricing policy does exactly this by inducing a potential game for the agents, where the potential function relates directly to ledger throughput. Simulations demonstrate that this policy leads to near-optimal throughput under a variety of conditions.

Keywords: blockchain, decentralized ledger, network throughput, potential game

1. INTRODUCTION

Decentralized ledgers, commonly implemented as encrypted linked lists of transaction records, or *blockchains*, allow individuals to trade resources and maintain a common state machine securely and without a central authority (Narayanan et al., 2016; Crosby et al., 2016). As the demand for such systems grows, the slow throughput of established systems, e.g. 7-15 transactions per second on Bitcoin and Ethereum (Croman et al., 2016), is becoming a major obstacle to more widespread adoption and success, especially in applications requiring high frequency or time-critical transactions. One innovation that promises to significantly increase transaction throughput is subdividing each the transaction records into distinct *shards*, allowing parallel communication, execution and storage of transactions that use different shards (Saraph and Herlihy, 2019). These subdivision methods have proven effective in conventional databases (Corbett et al., 2013). By evenly distributing transactions across multiple shards, the system can process transactions much faster than a serial blockchain ledger. However, in a decentralized system, different external users submit transactions, and there is no guarantee that they will choose to distribute their transactions in a way that enables the system to operate efficiently. In this paper, we investigate the use of small transaction surcharges as incentives to align the goals of the individual users with the system-wide goal of maximizing throughput.

Decentralized ledgers pose additional challenges due to various stakeholders having different and often competing objectives. For example, it is in the agents' interest that transactions are fast and cheap, which is more likely to occur when congestion is low, but *miners* or *validators* (agents responsible for reaching a consensus on which transactions are valid) benefit when transactions are expensive, which may be the case when congestion is high. These conflicting incentives must be accounted for in the design of an effective sharding system for blockchain ledgers.

Several proposals for sharding block chains have previously been put forth (Kokoris-Kogias et al., 2018; Luu et al., 2016; Zamani et al., 2018; Buterin, 2016), each of which aims to make the ledgers more scalable while maintaining appropriate levels of security. However, these approaches all rely on randomization for distributing transactions among shards and do little to explicitly mitigate the problems of load imbalance and frequent cross-shard transactions, which consume resources in communication between shards and potentially, depending on the sharding implementation, force the pausing of execution threads on one or both shards.

We address these problems here with a transaction pricing policy that incentivizes agents to choose shards in a way that maximizes ledger throughput. In particular, the proposed pricing function is based on a novel transaction efficiency measure that induces a potential game for the agents, where the potential function relates directly to overall transaction throughput.

The problem considered here resembles that of congestion games, a classic example of potential games, in which the

978-1-7281-9390-8/20/$31.00 ©2020 IEEE
DOI 10.1109/CVCBT50464.2020.00008

goal is to minimize congestion in transportation or communication networks, for example, by setting prices to align agent incentives with this goal (Monderer and Shapley, 1996). Indeed if the only goal were to minimize congestion on the transaction network, this would be a straightforward application of congestion game theory. However, the reduced efficiency caused by cross-shard transactions introduces additional complexity that must be accounted for in the pricing mechanism. By combining transaction size and the distribution of transactions across shards into a single quantity called *transaction efficiency*, we are able to express the throughput objective as a function of agents' shard choices and previous transactions, which we then use to set the price. Decentralized ledgers sometimes use transaction fees as incentives for other agents to maintain the ledger by validating transactions, so our proposal would simply weight such fees to promote overall efficiency.

2. LEDGER AND TRANSACTION MODEL

The model consists of a set of agents who transact with their neighbors in the network via a blockchain ledger.

2.1 Agents and network

The network consists of a set $\mathcal{A}$ of n agents who are interconnected by the edges $\mathcal{E} \subseteq \mathcal{A}^2$, where an edge $(i,j) \in \mathcal{E}$ means that agent i can request a transaction from agent j. We denote the set of neighboring agents from which agent i requests transactions by $\mathcal{N}_i := \{j \in \mathcal{A} : (i,j) \in \mathcal{E}\}$. Each agent i maintains a balance of resources $\mathbf{b}_i := [b_i^1, \ldots, b_i^m]^\top$, where b_i^s denotes the amount of resources agent i owns in shard $s \in \mathcal{S} := \{1, \ldots, m\}$, and its objective is to choose shards and execute transactions in a way that minimizes transaction fees.

2.2 Transactions

Assume that transactions are fully asynchronous and arrive in a sequence where one agent (i, the *receiver*) requests a transaction of an amount (a) from another agent (j, the *sender*), such that the sender is in the neighbor set of the receiver ($j \in \mathcal{N}_i$). Let $T := (i, j, \boldsymbol{\tau})$ denote a transaction requested by agent i from agent j, where $\boldsymbol{\tau} := [\tau_1, \ldots, \tau_m]^\top$ lists the amounts to be transferred in each shard, which we assume are non-negative (each $\tau_s >= 0$), and sum to the total amount ($\sum_{s=1}^{m} \tau_s = a$). The *cardinality* $|\mathcal{S}_T|$ is the number of shards used in the transaction. Let $\mathcal{P}$ denote the set (or *pool*) of transactions that are waiting to be added to a block. We define the price of a transaction as a function of the transaction itself and the current transaction pool: $\phi := f(T, \mathcal{P})$, to be given in precise terms in section 4. The transaction process proceeds as follows:

(1) **Request:** Agent i requests a transaction from agent $j \in \mathcal{N}_i$ and specifies an amount a and a shard $s_i \in \mathcal{S}$ in which to receive the transaction, with the goal of minimizing current and expected future transaction fees. Only the sender pays the transaction fees, but the receiver has an incentive to minimize these in order to maximize the probability that the sender accepts and fulfills the transactions. We can express the receiver's shard choice as

$$s_i \in \arg\min_{s \in \mathcal{S}} \left[\gamma_r f(T, \mathcal{P}) + (1 - \gamma_r) E[\phi_{ij}; s_i]\right], \quad (1)$$

where $E[\phi_{ij}; s_i]$ denotes the expected price of future transactions from the sender to the receiver if the receiver requests the current transaction in shard s_i, and $\gamma_r \in [0, 1]$ denotes the priority receiving agents place on the current transaction fee relative to future transactions. Estimating $E[\phi_{ij}; s_i]$ is a key element in the pricing mechanism and is discussed further in section 3.1.

(2) **Fulfillment:** Agent j accepts and fulfills the transaction if there are sufficient funds (including fees) and chooses a set of shards $\mathcal{S}_j \subseteq \mathcal{S}$ from which to send resources, with the goal of minimizing current and expected future transaction fees. Since the final transaction always includes the receiver's requested shard s_i, the transaction set is given by $\mathcal{S}_T := \mathcal{S}_j \cup \{s_i\}$. The set of all feasible transaction shard sets is then

$$\mathbb{S}_j(a) := \Big\{\mathcal{S}_T \subseteq \mathcal{S} : \sum_{s \in \mathcal{S}_j} b_j^s \geq a + f(T, \mathcal{P})\Big\}.$$

For a requested transaction, we assume that the sender chooses a shard set that minimizes the following expression:

$$\mathcal{S}_j \in \arg\min_{\mathcal{S}_T \in \mathbb{S}_j(a)} \left[\gamma_s f(\bar{\mathcal{S}}_T, \mathcal{P}) + (1 - \gamma_s) E[\phi_{ij}; s]\right], \quad (2)$$

where $\gamma_s \in [0, 1]$ denotes the priority sending agents place on the current transaction fee relative to future transactions. The sender withdraws the resources to be transferred plus transaction fees from the shards in $\mathcal{S}_T^*$, and the receiver adds the transferred balance to shard s_i. The result is a transaction T that goes into the transaction pool $\mathcal{P} \leftarrow \mathcal{P} \cup \{T\}$.

(3) **Block assembly:**

We assume that the blockchain is divided into m shards, each of which contains σ slots, and that the blocks in each shard are produced synchronously.

Let $\mathcal{P}_s$ denote the set of transactions in pool $\mathcal{P}$ that use shard s:

$$\mathcal{P}_s := \{(i, j, \boldsymbol{\tau}) \in \mathcal{P} : \tau_s > 0\}.$$

In this paper, we assume that when any shard becomes full (i.e., there exists a shard s such that $|\mathcal{P}_s| > \sigma$) the block is assembled from all transactions in the pool and $\mathcal{P}$ is reset to empty. The maximum theoretical capacity of the blockchain ($m \times \sigma$) is reached when the cardinality of all transactions is one (no cross-shard transactions) and each shard contains exactly the maximum number of transactions. Note that it may not be possible to execute both sides of multi-shard transactions leading to failure of transactions of this type. Therefore, assuming that all transactions in the pool are executed, as we do in this study, leads to optimistic estimates of the throughput in the case of frequent cross-shard transactions. This leads to conservative estimates of the performance gains that would arise from implementing our proposed pricing policy. Although this serves as a reasonable approximation for our purposes, coordinating states between shards is a nontrivial problem and provides further motivation to incentivize single-shard transactions. We note that this analysis generalizes to blockchains with complex state execution rules such as smart

contracts where cross-shard transactions require at least twice the computation of an otherwise identical single-sharded transaction.

3. TRANSACTION THROUGHPUT

Transaction throughput measures the number of transactions processed in a given unit of time, but since we do not explicitly include time in our model, we seek a time-independent alternative quantity. Specifically, we define the *transaction efficiency* of a pool $\mathcal{P}$ as the fraction of the theoretical maximum number of transactions that could be included in a block in which all transactions have cardinality one and are evenly distributed across shards.

There are two primary factors that determine transaction efficiency: *cardinality* and *shard balance*. We say that shard balance is high when the transaction pool uses the shards in roughly equal proportions and low when the transaction pool uses some shards much more than others. Figure 1 illustrates three partially completed blocks from transaction pools exhibiting varying degrees of cardinality and shard balance.

We measure shard balance in terms of the deviation from uniform shard usage in the transaction pool. The usage of each shard relative to the pool is given by $\mathbf{u}(\mathcal{P}) := [u_1(\mathcal{P}), \ldots, u_m(\mathcal{P})]^\top$, where

$$u_s(\mathcal{P}) := \frac{|\mathcal{P}_s|}{\sum_{T\in\mathcal{P}} |\mathcal{S}_T|}.$$

Since we consider only positive transaction fees, we focus on those shards with greater usage than average and define the loading of shard $s \in \mathcal{S}$ by

$$\lambda_s(\mathcal{P}) := \max\left(0, u_s(\mathcal{P}) - \frac{1}{m}\right).$$

These values are collected in the loading vector $\boldsymbol{\lambda}(\mathcal{P}) := [\lambda_1(\mathcal{P}), \ldots, \lambda_m(\mathcal{P})]^\top$, and we can now quantify the shard balance in the transaction pool as

$$B_{\mathcal{P}}(\mathcal{S}) := 1 - \sum_{s\in\mathcal{S}} \lambda_s(\mathcal{P}). \tag{3}$$

Note that by definition, $B_{\mathcal{P}}(\mathcal{S}) \in [0, 1)$. We can now express the total transaction efficiency as the shard balance divided by the mean cardinality of all transactions in the pool:

$$F_{\mathcal{P}} := \frac{B_{\mathcal{P}}(\mathcal{S})}{\frac{1}{|\mathcal{P}|}\sum_{T\in\mathcal{P}} |\mathcal{S}_T|}. \tag{4}$$

As an example, for the middle block in Figure 1, the shard usage is $\mathbf{u}(\mathcal{P}) = (\frac{0}{8}, \frac{4}{8}, \frac{4}{8}, \frac{0}{8})^\top$, resulting in the loading vector $\boldsymbol{\lambda}(\mathcal{P}) = (0, \frac{1}{4}, \frac{1}{4}, 0)^\top$ and total shard balance $B_{\mathcal{P}}(\mathcal{S}) = \frac{1}{2}$. Since the mean cardinality is one, the transaction efficiency is $\frac{1}{2}$. For the left block in Figure 1, since the shard balance is one and the mean cardinality is two, the transaction efficiency is also $\frac{1}{2}$. The block on the right achieves the maximum transaction efficiency of 1.

3.1 Expected transaction efficiency

We introduce for each edge $(i, j) \in \mathcal{E}$ along which a transaction can take place, a *shard request distribution* $\mathbf{w}_{ij} := [w_{ij1}, \ldots, w_{ijm}]^\top$, where w_{ijs} denotes the probability that agent i will choose shard s when requesting a transaction from agent j. Similarly, we define a *shard sending distribution* $\mathbf{v}_{ij} := [v_{ij1}, \ldots, v_{ijm}]^\top$, where v_{ijs} denotes the probability that agent j will choose shard s when sending a transaction to agent i. These randomized distributions, which lie on a probability simplex (each $w_{ijs} \geq 0$ and $\sum_{s=1}^m w_{ijs} = 1$, and similarly for $\mathbf{v}_{ij}$), model the initial uncertainty about the shards used by neighboring agents and how such uncertainty evolves toward deterministic choices.

Expected cardinality: We can now write the expected cardinality of transactions between agents i and j as:

$$\begin{aligned} E\left[|\mathcal{S}_T|\right] &= \mathbf{w}_{ij}^\top P_{\text{card}} \mathbf{v}_{ij} \\ &= \begin{bmatrix} w_{ij1} & w_{ij2} & \cdots & w_{ijm} \end{bmatrix} \begin{bmatrix} 1 & 2 & \cdots & 2 \\ 2 & 1 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 2 & 2 & \cdots & 1 \end{bmatrix} \begin{bmatrix} v_{ij1} \\ v_{ij2} \\ \vdots \\ v_{ijm} \end{bmatrix}, \end{aligned} \tag{5}$$

where P_{card} is a matrix that encodes the expected cardinality when the agents request transactions from each other in the shards corresponding to the row and column of each entry.

Expected shard balance: Similarly, we can express the expected shard balance as follows:

$$E\left[1 - \sum_{s\in\mathcal{S}_T} \lambda_s(\mathcal{P})\right] = \mathbf{w}_{ij}^\top P_{\text{bal}} \mathbf{v}_{ij} =$$
$$\begin{bmatrix} w_{ij1} & \cdots & w_{ijm} \end{bmatrix} \begin{bmatrix} 1-\lambda_1 & 1-\lambda_1-\lambda_2 & \cdots & 1-\lambda_1-\lambda_m \\ 1-\lambda_1-\lambda_2 & 1-\lambda_2 & \cdots & 1-\lambda_2-\lambda_m \\ \vdots & \vdots & \ddots & \vdots \\ 1-\lambda_1-\lambda_m & 1-\lambda_2-\lambda_m & \cdots & 1-\lambda_m \end{bmatrix} \begin{bmatrix} v_{ij1} \\ \vdots \\ v_{ijm} \end{bmatrix}, \tag{6}$$

where P_{bal} encodes the shard balance values corresponding to the shards used in the transaction (we omit the argument $\mathcal{P}$ for a more compact expression).

Expected transaction efficiency: Based on the definition of transaction efficiency for the entire transaction pool (4), we define the efficiency of a single transaction in a given pool as the shard balance of the transaction shards divided by the cardinality:

$$F_{\mathcal{P}}(T) := \frac{B_{\mathcal{P}}(\mathcal{S}_T)}{|\mathcal{S}_T|}. \tag{7}$$

Using (7), we can write the expected efficiency of transactions from agent j to i:

$$E\left[F_{\mathcal{P}}(T)\right] = \mathbf{w}_{ij}^\top P_{\text{eff}} \mathbf{v}_{ij} =$$
$$\begin{bmatrix} w_{ij1} & w_{ij2} & \cdots & w_{ijm} \end{bmatrix} \begin{bmatrix} 1-\lambda_1 & \frac{1-\lambda_1-\lambda_2}{2} & \cdots & \frac{1-\lambda_1-\lambda_m}{2} \\ \frac{1-\lambda_1-\lambda_2}{2} & 1-\lambda_2 & \cdots & \frac{1-\lambda_2-\lambda_m}{2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-\lambda_1-\lambda_m}{2} & \frac{1-\lambda_2-\lambda_m}{2} & \cdots & 1-\lambda_m \end{bmatrix} \begin{bmatrix} v_{ij1} \\ v_{ij2} \\ \vdots \\ v_{ijm} \end{bmatrix}. \tag{8}$$

4. TRANSACTION PRICING

We propose the use of transaction fees to align the individual goals of minimizing fees with the system-wide goal of maximizing throughput. A natural choice is to make the fee proportional to the desired objective, which we have quantified as the transaction efficiency. Hence, we propose the following pricing function:

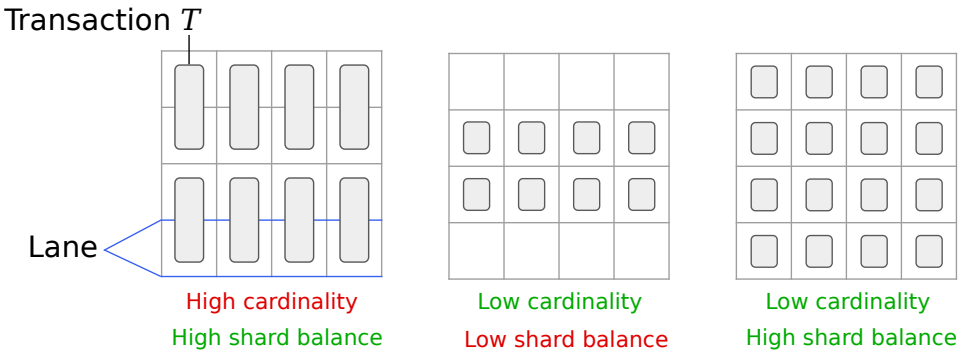

Fig. 1. Cardinality and shard balance in three different transaction pools assembled into blocks. Rows corresponds to shards and columns represent block slices. The gray boxes symbolize transactions on the underlying shards.

$$f(T,\mathcal{P}) := p_0(T) + \left(1 - \frac{B_{\mathcal{P}}(\mathcal{S}_T)}{|\mathcal{S}_T|^\alpha}\right)\phi_{\max}, \tag{9}$$

where $p_0(T)$ is the nominal transaction price, which can vary with computational requirements and market demand, $\phi_{\max}$ is the maximum transaction fee, and α is a free parameter that can be used to calibrate the cardinality estimate or to further discourage multi-shard transactions. Note that the price includes one minus the transaction efficiency since the price should be low when the efficiency is high. To simplify the remaining analysis, we assume that $p_0(T) = 0$ and $\phi_{\max} = 1$ unless otherwise stated. However, it is straightforward to extend the analysis to include these parameters.

While (9) defines the price for a particular transaction, the expected price of a future transaction requested by agent i from agent j is given by:

$$E\left[\phi_{ij}\right] = 1 - \mathbf{w}_{ij}^\top P \mathbf{v}_{ij} = 1 - \begin{bmatrix} w_{ij1} & w_{ij2} & \cdots & w_{ijm} \end{bmatrix} \begin{bmatrix} 1-\lambda_1 & \frac{1-\lambda_1-\lambda_2}{2^\alpha} & \cdots & \frac{1-\lambda_1-\lambda_m}{2^\alpha} \\ \frac{1-\lambda_1-\lambda_2}{2^\alpha} & 1-\lambda_2 & \cdots & \frac{1-\lambda_2-\lambda_m}{2^\alpha} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-\lambda_1-\lambda_m}{2^\alpha} & \frac{1-\lambda_2-\lambda_m}{2^\alpha} & \cdots & 1-\lambda_m \end{bmatrix} \begin{bmatrix} v_{ij1} \\ v_{ij2} \\ \vdots \\ v_{ijm} \end{bmatrix}. \tag{10}$$

The expression (10) provides a direct link from the transaction price (9) to the agents' optimal choice of shards in which to request transactions for the case of agents that wish to minimize long-term expected transaction fees ($\gamma_r = \gamma_s = 0$), allowing us to rewrite (1) as follows:

$$\mathbf{w}_{ij}^* := \arg \max_{\mathbf{w}_{ij} \in \Delta_m} \mathbf{w}_{ij}^\top P \mathbf{v}_{ij}, \tag{11}$$

where $\mathbf{w}_{ij}^*$ denotes an update to the shard request distribution $\mathbf{w}_{ij}$ and Δ_m denotes the m-dimensional probability simplex. The update (11) is indeed a *best response* of agent i (in mixed-strategy space) to the mixed strategy of agent j. Similarly, the sender's optimal shard choice distribution is given by

$$\mathbf{v}_{ij}^* := \arg \max_{\mathbf{v}_{ij} \in \Delta_m} \mathbf{w}_{ij}^\top P \mathbf{v}_{ij}, \tag{12}$$

Since the sender also seeks to minimize long-term expected transaction fees, this distribution update is indeed independent from the choice of the receiver.

There is an important class of multi-player games called *potential games*, in which players choose actions to maximize their individual utility functions, which in turn increases some global utility function (Monderer and Shapley, 1996). A key property of potential games is that when agents act to improve their utility functions, the system is guaranteed to converge to a Nash equilibrium, which is a state in which no action by any single agent will increase their utility. In potential games, Nash equilibria also correspond to maxima of the global utility function. We define potential games in precise terms below.

Let $\mathcal{S}_i$ denote the space of actions for a single agent and let $\boldsymbol{\mathcal{S}} := \mathcal{S}_i^n$ denote the set of all actions in the system. Denote by u_i : the utility function of agent i. Together these define a game for the n agents, and such a game is said to be a *potential game* if there exists a function $H : \boldsymbol{\mathcal{S}} \to \mathbb{R}$ such that for any agent $i \in \mathcal{A}$ and any pair of actions $a, a' \in \mathcal{S}_i$ (where a_{-i} denotes the actions of all agents except i):

$$H(a_i', a_{-i}) - H(a_i, a_{-i}) = u_i(a_i', a_{-i}) - u_i(a_i, a_{-i}).$$

Indeed we can show that the proposed pricing mechanism induces a potential game over the transaction edges with the following potential function:

$$H := \sum_{(k,l)\in\mathcal{E}} \mathbf{w}_{kl}^\top P \mathbf{v}_{kl}. \tag{13}$$

Theorem 1. The game where n agents connected by the edges $\mathcal{E}$ update their shard request and sending distributions according to the edge-utility functions $u_{ij} := \mathbf{w}_{ij}^\top P \mathbf{v}_{ij}$ is a potential game with the potential function (13).

Proof. Given an edge $(i,j) \in \mathcal{E}$, suppose agent i updates its shard request distribution for this edge from $\mathbf{w}_{ij}$ to $\mathbf{w}_{ij}'$. Then, the change in the edge utility function is equal to $u_{ij}' - u_{ij} = (\mathbf{w}_{ij}' - \mathbf{w}_{ij})^\top P \mathbf{v}_{ij}$. The resulting change in the potential function is

$$H' - H := (\mathbf{w}_{ij}' - \mathbf{w}_{ij})^\top P \mathbf{v}_{ij},$$

since the only change was to agent i's shard request distribution from agent j, which is exactly equal to the change in the edge utility function u_{ij}. Similarly, if agent j updates its shard sending distribution for this edge from $\mathbf{v}_{ij}$ to $\mathbf{v}_{ij}'$, then the change in the edge utility function is equal to $H_{ij}' - H_{ij} = \mathbf{w}^\top(\mathbf{v}_{ij}' - \mathbf{v}_{ij})$, which is also equal to the change in the potential function, completing the proof.

Although the case we have analyzed is somewhat simplified, the fact that it constitutes a potential game is important because it ensures not only that the agents' incentives are aligned with the global objective, but that rational choices by the agents will result in convergence of the system to a maximum of the global potential function, which in our case corresponds to transaction throughput. And their choices need not be optimal – the only requirement is that agents take actions that increase their local utility function (lower their expected future transaction price). This means that even deterministic shard choices (pure strategy best or better responses) will lead to convergence. For example, a pure strategy best response update is given by replacing the probability simplex Δ_m in (11) with the set of all pure strategies $\Pi_m := \{\mathbf{w} \in \Delta_m : w_s \in \{0, 1\} \text{ for each } s \in \mathcal{S}\}$.

A potential problem with the approach described so far is that the update (11) assumes that agents know the shard request or sending distribution of their transacting neighbors along each edge, which would require additional communication between agents. To resolve this, let's assume that agents keeps estimates ($\hat{\mathbf{v}}_{ij}$ for the receiver, $\hat{\mathbf{w}}_{ij}$ for the sender) of the shard request and sending distributions of each neighbor, constructed simply from the normalized histogram of past transaction requests. The new update rules that rely only on information available to the respective agents are then:

$$\mathbf{w}_{ij}^{+} := \arg \max_{\mathbf{w}_{ij} \in \Delta_m} \mathbf{w}_{ij}^{\top} P \hat{\mathbf{v}}_{ij}, \tag{14}$$

$$\mathbf{v}_{ij}^{+} := \arg \max_{\mathbf{v}_{ij} \in \Delta_m} \hat{\mathbf{w}}_{ij}^{\top} P \mathbf{v}_{ij}. \tag{15}$$

This turns out to be an example of *fictitious play* in game theory, where agents estimate strategies of other players based on empirical distributions. Indeed, multiplayer potential games in which players act to improve their utility using fictitious play are known to converge to a Nash equilibrium (Marden et al., 2009).

The optimizations (14)-(15) are readily solved via linear programming, and the deterministic (pure-strategy) case is an integer program that reduces to finding the maximum entry in a m-dimensional vector:

$$\begin{aligned} \mathbf{w}_{ij}^{+} &:= \arg \max_{\mathbf{w}_{ij} \in \Pi_m} \mathbf{w}_{ij}^{\top} P \hat{\mathbf{w}}_{ji} \\ &= \mathbf{e}_{s^*}, \text{ where } s^* = \arg \max_{s \in \mathcal{S}} P_s \hat{\mathbf{w}}_{ji}, \end{aligned} \tag{16}$$

where $\mathbf{e}_s$ refers to column s of the $m \times m$ identity matrix, and P_s denotes row s of the matrix P. A similar modification can be made for the sending shard update. The computational complexity of the pure-strategy and mixed-strategy optimizations are linear and polynomial (due to the complexity of linear programming, e.g. (Cohen et al., 2019)), respectively.

5. SIMULATIONS

In this section, we investigate the performance of a simulated blockchain ledger with transaction price (9) in which agents update their shard request distributions with pure strategy best response updates (16).

5.1 Ideal case

We begin with a simple scenario to test the pricing mechanism under ideal conditions. Suppose that 20 agents transact with two neighbors each via a ring network on a ledger with 4 shards. Each block contains 2500 slices meaning that the maximum capacity is 10000 transactions per block. In this scenario, agents start with an arbitrarily large initial balance ($1e6$) in one shard (such that the resources in these shards will not be depleted), staggered among the agents, and transactions of a small fixed amount (10) are generated randomly by rounds. That is, each of the 38 edges in the network executes a transaction in random order, and then the process repeats in a new random sequence until 5 blocks are eventually completed.

In the figures that follow, the x-axis corresponds to the transaction index, where anytime one shard reaches maximum capacity, a block is assembled (indicated by the vertical grid lines) and the transaction pool resets to zero. The top panel shows the proportion of transactions contained in each shard, resulting in the balance value shown on the second panel. The third panel shows the mean transaction cardinality and the fourth shows the transaction efficiency calculated for each block. As a baseline, Figure 2 shows the result of assigning transactions to random shards, modeling a standard hash-based sharding protocol.

We observe that the randomized policy achieves a reasonably even distribution among the shards, but since the mean cardinality is quite high, the efficiency is only about 50%, yielding 25899 transactions in 5 blocks.

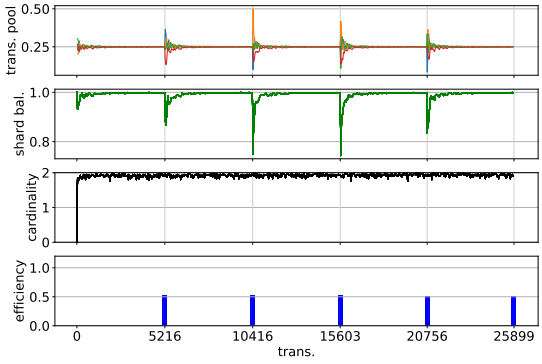

Fig. 2. Fixed price and random shard requests on 20-agent ring network with 4 shards.

Next, we apply load-minimizing pricing by setting $\alpha = 0$ in (9) and allowing the agents to update their shard requests with best-responses. Figure 3 shows that the loading stays very low, but the transaction efficiency remains below 60%, because the agents still have no incentive to transact in a single shard and therefore engage in frequent multi-shard transactions.

Finally, we see in Figure 4 that combining the proposed pricing mechanism ($\alpha = 0.001$) with best-response updates results in high shard balance and low cardinality and therefore almost perfect transaction efficiency.

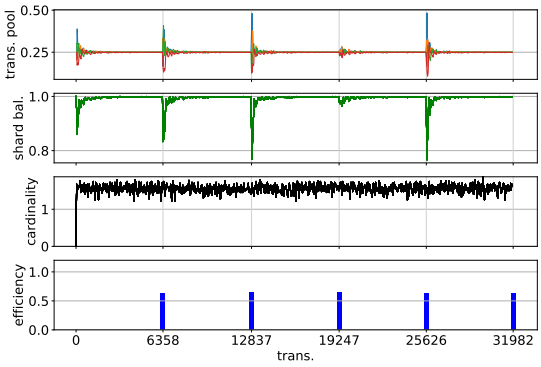

Fig. 3. Congestion pricing with best-response update on 20-agent ring network with 4 shards.

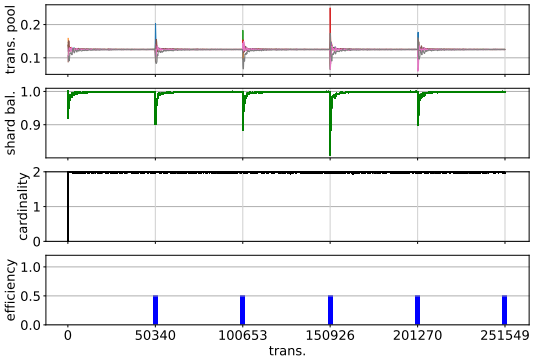

Fig. 5. Fixed pricing with random shard requests on 100-agent scale-free network with 8 shards.

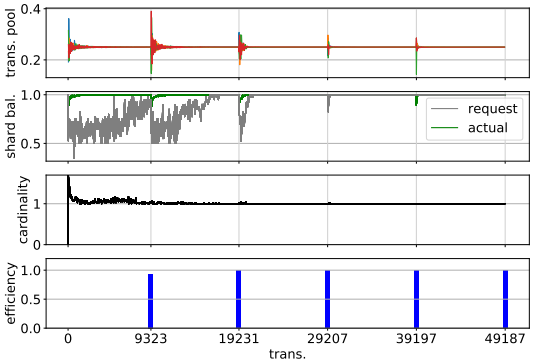

Fig. 4. Efficiency pricing with best-response update on 20-agent ring network with 4 shards ($\alpha = 0.001$).

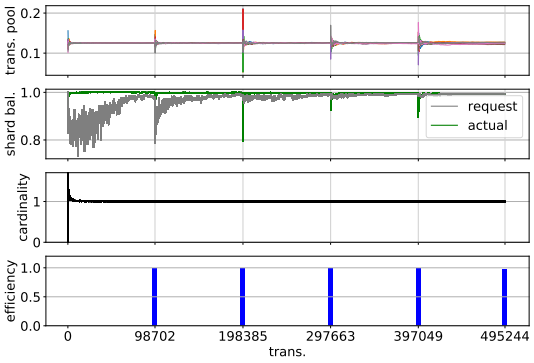

Fig. 6. Efficiency pricing with best-response update on 100-agent directed random network with 8 shards ($\alpha = 0.00015$).

5.2 Larger random networks

In the next simulation study, we construct a directed random network with a long-tailed degree distribution using a preferential attachment model. Also, instead of generating transactions on repeated random sequences of edges, we generate them uniformly at random on the edges. Blocks in this scenario contain 12500 slices and thus a maximum capacity of 100000 transactions. In this case, we initialized the agents with arbitrarily large balances in each shard. Figure 6 shows that network still converges to a high transaction efficiency compared to the baseline shown in Figure 5. The gray data in the loading plot shows how evenly distributed the shard requests are throughout the network. As the agents acquire more information about the shard preferences of their neighbors, the shard requests converge to align with the corresponding send requests while maintaining a balance across shards. The values of α in both simulations were chosen by manual tuning, and the results are somewhat sensitive to small changes in this parameter. An automated tuning method for α would be a useful direction for future research.

6. CONCLUSIONS AND FUTURE WORK

The pricing function (9) induces a potential game for the users of a sharded blockchain ledger, such that is in their interests to choose shards in a way that promotes ideal throughput conditions on the ledger. Both the estimation of shard request distributions and the deterministic best-response policy are implementable with simple and fast computations, and guarantee convergence to a Nash equilibrium under some practical simplifying assumptions. Simulations demonstrate the effectiveness of the policy in various conditions.

We plan to extend the approach in several ways. For example, there is currently no mechanism that would encourage agents to request from different neighbors in the same lane, although this would seem reasonable in practice, especially if resources are scarce. This could be achieved by adding a term that encourages agents to align shard requests with their resource balance. We also intend to more explicitly account for varying costs of transaction and smart contract execution and storage.

ACKNOWLEDGEMENTS

We are grateful to Marcin Abram and Jin-Mann Wong for insightful technical discussions and for help in editing the paper.

REFERENCES

Buterin, V. (2016). Ethereum: Platform review. *Opportunities and Challenges for Private and Consortium Blockchains*.

Cohen, M.B., Lee, Y.T., and Song, Z. (2019). Solving linear programs in the current matrix multiplication time. In *Proceedings of the 51st annual ACM SIGACT symposium on theory of computing*, 938–942.

Corbett, J.C., Dean, J., Epstein, M., Fikes, A., Frost, C., Furman, J.J., Ghemawat, S., Gubarev, A., Heiser, C., Hochschild, P., et al. (2013). Spanner: Google's globally distributed database. *ACM Transactions on Computer Systems (TOCS)*, 31(3), 8.

Croman, K., Decker, C., Eyal, I., Gencer, A.E., Juels, A., Kosba, A., Miller, A., Saxena, P., Shi, E., Sirer, E.G., et al. (2016). On scaling decentralized blockchains. In *International Conference on Financial Cryptography and Data Security*, 106–125. Springer.

Crosby, M., Pattanayak, P., Verma, S., Kalyanaraman, V., et al. (2016). Blockchain technology: Beyond bitcoin. *Applied Innovation*, 2(6-10), 71.

Kokoris-Kogias, E., Jovanovic, P., Gasser, L., Gailly, N., Syta, E., and Ford, B. (2018). Omniledger: A secure, scale-out, decentralized ledger via sharding. In *2018 IEEE Symposium on Security and Privacy (SP)*, 583–598. IEEE.

Luu, L., Narayanan, V., Zheng, C., Baweja, K., Gilbert, S., and Saxena, P. (2016). A secure sharding protocol for open blockchains. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 17–30. ACM.

Marden, J.R., Arslan, G., and Shamma, J.S. (2009). Joint strategy fictitious play with inertia for potential games. *IEEE Transactions on Automatic Control*, 54(2), 208–220.

Monderer, D. and Shapley, L.S. (1996). Potential games. *Games and economic behavior*, 14(1), 124–143.

Narayanan, A., Bonneau, J., Felten, E., Miller, A., and Goldfeder, S. (2016). *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton University Press.

Saraph, V. and Herlihy, M. (2019). An empirical study of speculative concurrency in ethereum smart contracts. *arXiv preprint arXiv:1901.01376*.

Zamani, M., Movahedi, M., and Raykova, M. (2018). Rapidchain: Scaling blockchain via full sharding. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 931–948. ACM.