The Economics of Transaction Fee Mechanism Design

David Lancashire
david.lancashire@gmail.com

December 7, 2024

Abstract

A socially optimal transaction fee mechanism must be *incentive compatible* for users and block producers and *pareto optimal* for all participants. This paper offers a succinct proof it is possible to overcome the objection Paul Samuelson raised to achieving both properties in the presence of a blockchain that offers both excludable and non-excludable forms of utilitty.

Introduction

Start with the standard *utility possibilities frontier* for a single individual choosing between two goods.[1]

$$\frac{u_a^i}{u_b^i} = \frac{F_a}{F_b}$$

We can sum across all individuals to find the *utility possibilities frontier* for society as a whole.

$$\sum_{i=1}^{s} \frac{u_a^i}{u_b^i} = \frac{F_a}{F_b}$$

All points on this curve are *pareto optimal* – we cannot increase utility for any individual without decreasing it for at least one other person. Voluntary-trade pulls this equation into equilibrium for any two private goods under classical assumptions. That implies equilibrium across any N-dimensional basket of private goods. We therefore set *good a* as our blockchain and *good b* as our basket of all other private goods competing for consumption of the transaction fee.

It follows that unless a fee mechanism incentivizes payment of the amount of fees defined by this equation said mechanism cannot be *incentive compatible* or *collusion resistant* as a subset of users must exist who can costlessly increase utility through the byzantine strategy of paying a different fee. If also follows that any fee mechanism which is defined by this equation is collusion-proof if the blockchain is a private good, as – in the absence of strategic misdirection during the negotiating stage – negotiations between block producers and users must be utility-optimizing for both.

Modelling Blockchain Provision as a Public Good

It follows that in the absence of available strategies for strategic misdirection, the blockchain must be a public good for collusion to result in suboptimal provision. Modelling it as a private good also fails to capture the non-excludable public benefits created when block producers compete to include *publicly-circulating, fee-bearing transactions* in blocks:

- **higher security** - competition induces nodes to spend a greater percentage of their income on the security function than would otherwise be rational.
- **faster inclusion** - competition makes confirmation times more sensitive to gradiations in transaction fees as multiple nodes prioritize the highest-paying transactions for inclusion in blocks.
- **censorship resistance** - it is more profitable for non-censoring nodes to join the network if the fee-paying transactions that existing producers refuse to include are circulating publicly.

We thus follow Samuelson (1954) and model it as a public good that provides both *private* and *public* benefits:

$$\sum_{i=1}^{s} \frac{u_{pub+priv}^i}{u_b^i} = \frac{F_{pub+priv}}{F_b}$$

Observing this equation, Samuelson hypothesized that free-rider pressures make it impossible to induce any decentralized mechanism to induce optimal provision. Hurwicz generalized Samuelson's critique to all informationally decentralized games where "byzantine" strategies permit misleading signalling about utility during the course of price-signalling, but for the sake of examining Samuelson's objection we assume temporarily that strategic misdirection is not present.

[1] The LHS shows the marginal rate of substitution between two goods. This is the amount of utility that *individual i* gets from each good expressed as a ratio. The RHS shows a similar ratio that reflects the cost of producing each. The individual purchases the optimal amount of utility with the resources they exspend when these ratios are equal. It is easy to see that if these equations are not in alignment the individual can costlessly increase their utility by adjusting the way in which they allocate their resources. If the LHS is smaller than the RHS, for instance, the individual can increase their utility by consuming more of good b.

Escaping the Samuelson Trap

Public blockchains induce provision of non-excludable benefits in proportion to the percentage of fee-bearing transactions that circulate publicly. Optimal provision must occur when all transactions and fees are available for competitive inclusion. We therefore simplify our cost function and add a variable x that reflects the probability that transactions and fees are circulating publicly:

$$\sum_{i=1}^{s} \frac{u^i_{(pub*x)+priv}}{u^i_b} = \frac{F_{priv}}{F_b}$$

The value of x depends on whether users and block producers share unconfirmed fee-bearing transactions. Observe:

- **rational users** - prefer to share transactions publicly, as sharing transactions publicly increases the utility to the user, while rational block producers will not discriminate against transactions that are publicly available.
- **rational nodes** - prefer to free-ride on publicly-circulating transactions but not share their own transactions.

The obstacle to achieving optimality is the preference of block producers to restrict transaction propagation to reduce competition for collection of the fee.

Without incentivizing block producers to share transactions publicly we cannot achieve *pareto optimality.* Without *pareto optimality* we cannot achieve *incentive compatibility.* And without *incentive compatibility* we cannot design a fee mechanism that is robust against collusion. This formally proves the *transaction fee mechanism design* requires a solution to the *sybil problem.*

A theoretical proof it is impossible to induce rational actors to share transactions in proof-of-work and proof-of-stake networks may be found in the paper *On Bitcoin and Red Balloons.* We dissent from this proof by observing the existence of routing work mechanisms which incentivize block producers to share unconfirmed transactions with peers. If these mechanisms are confirmed to solve this specific technical problem, they necessarily also consitute socially optimal transaction fee mechanisms as our *utility possibilities frontier* simplifies to the following once x becomes 1:

$$\sum_{i=1}^{s} \frac{u^i_{pub+priv}}{u^i_b} = \frac{F_{priv}}{F_b}$$

Social optimality is achieved at all points on this curve by definition.

Incentive compatibility is achieved since the fee paid by users reflects their private utility.

References

[1] Moshe Babaioff, Shahar Dobzinski, Sigal Oren, and Aviv Zohar. On bitcoin and red balloons. In *SeCO Workshops*, 2011.

[2] Oguzhan Ersoy, Zhijie Ren, Zekeriya Erkin, and Reginald L. Lagendijk. Information propagation on permissionless blockchains. *CoRR*, abs/1712.07564, 2017.

[3] Leonid Hurwicz. On the concept and possibility of informational decentralization. *The American Economic Review*, 59(2):513–524, 1969.

[4] Leonid Hurwicz. But who will guard the guardians. *The American Economic Review*, 98(3):577–585, 2008.

[5] Paul Samuelson. *Economics: An Introductory Analysis.* McGraw-Hill Book Company, 1948.

[6] Paul Samuelson. The pure theory of public expenditure. *The Review of Economics and Statistics*, 36(4):387–389, 1954.

[7] Paul Samuelson. Diagrammatic exposition of a theory of public expenditure. *The Review of Economics and Statistics*, 37(4):350–356, 1955.

[8] Youjia Zhang and Pingzhong Tang. Collusion-proof and sybil-proof reward mechanisms for query incentive networks, 2023.