Proof of Honesty: Coalition-Proof Blockchain Validation without Proof of Work or Stake

John P. Conley

Vanderbilt University
and
The Geeq Project

NSF-NBER DECENTRALIZATION 2019
University of Michigan, Ann Arbor

March 18, 2019

What is Blockchain?

A Data System:

$$\text{Blockchain} \subset \text{Distributed Ledger Technology} \subset$$
$$\text{State Machine Replication System} \subset \text{Data System}$$

A Consensus System:

- Proof of Work
- Proof of Stake
- Proof of Authority
- Proof of Honesty
- Governance

.....

Consensus Mechanisms

Consensus mechanisms have the two main jobs:

- Establishing a canonical version of the current state of the data
- Making sure the canonical view is correct

In addition, it would be nice if:

- All copies of the database are identical or synchronize quickly
- All copies of the database are available for use
- Altering the data in unauthorized ways is difficult or impossible

The CAP (Brewer) Theorem

CAP Theorem tells us: No distributed data store can simultaneously provide more than two out of the following three:

Consistency: Every read receives the most recent write or an error

Availability: Every request receives a (non-error) response – without the guarantee that it contains the most recent write

Partition Tolerance: The system continues to operate despite an arbitrary number of messages being dropped (or delayed) by the network between nodes.

Honesty and Canonicalness

If we could have all three, we would have a canonical view of the state of the database.

Note, however, that availability and consistency does not imply that the most recent write is correct or honest.

Honesty and Canonicalness are logically different properties

Algorithmic Game Theory

The perspectives that economists and computer scientists bring to the table are different, and each have their value.

Blockchain protocols are have their roots in **algorithmic game theory** which adapts traditional noncooperative game theory for use in computational environments.

Agents using protocols without a complete understanding of how they work may have difficulty determining fully optimal actions. As a consequence, agents are often modeled as following *ad hoc* behavior patterns.

Algorithmic Game Theory

For example, agents might be assumed to be either **honest** or **malicious-type** players since fully rational play may exceed their cognitive limitations.

Algorithmic approaches tend to pay less attention to certain other elements of games and mechanisms:

- Multiple equilibria.
- Refinements of Nash equilibrium.
- Effects of information and belief structures on equilibrium in sequential games.

Mechanisms vs. Protocols

Protocol builders and economic mechanism designers face different problems

Mechanisms: Agents have private information.
Protocols: The truthfulness of validators is externally observable and provable.

Mechanisms: The designer generally sets up a game in which he imposes both a strategy space and a payoff function. Agents can only choose not to participate.
Protocols:The builder also sets up rules that are supposed to be followed and a specific set of messages and actions that are allowed by protocol. Validators, however, can send any messages they wish. Rewards and punishments exist only on/in the blockchain being validated and must be written and agreed upon by the validators themselves.

Where this Bites

A Honesty is endogenous

- Dishonest ≠ Broken

Equilibrium definition

- Nash (example: prisoners' dilemma)
- Dominant Strategy
- Coalition Proof

Multiple equilibrium

- Right side/left side
- All honest/all dishonest
- Information and expectations
- Battle of the sexes
- ETH worth $1000 or $100
- Increasing mining rewards

A Unanimity Game

- Agents are offered a chance to play a game in exchange for a one dollar admission fee.
- Each player who pays the fee is sent to a room where a name is written on the wall. Players are asked to write this name on a piece of paper.
- The papers are then gathered and compared. If they all have the same name, then each player is paid two dollars.
- If there is any disagreement about the name, all players get zero (which gives each a net payoff of negative one dollar).

Note that there are many Nash equilibrium including truth-telling.
This is a feature of most consensus protocols as well.

A Unanimity Game with Auditing

Add the following:

- If all agents write the same name, the named individual gets $1000 (like a transaction on a blockchain ledger).
- All agents sign their papers.
- If there is disagreement about the name, then the door to the room is opened, and the name on the wall is read.
- Any player who wrote down the correct name gets $2 of plus an equal share of a $1000 bonus.
- Players who wrote down an incorrect name receive nothing.

Equilibrium with Auditing

Truth-telling is the unique coalition-proof equilibrium. (implementation)

- Suppose all agents tried to collude and write down one of their own names and then share the $1000 received.
- Any single agent who defected and called for an audit would get the $1000 bonus which is more than an equal share of the $1000 that the coalition tries to steal.
- Knowing that at least one agent will certainly defect, the other agents will abandon the attempt to collude, and so truth-telling is the only equilibrium that remains.

Formalities

For completeness, recall the following definition of a game:

Agents: $n \in \{1, \ldots, N\} \equiv \mathcal{N}$

Strategies: $v_n \in \mathbb{R}_+^1 \quad \forall n \in \mathcal{N}$

Payoff Function: $F_n : \mathbb{R}_+^N \to \mathbb{R}^1 \quad \forall n \in \mathcal{N}$

Thus, when the set of agents $\mathcal{N}$ playing the game choose a *strategy profile* $v \equiv \{v_1, \ldots, v_N\} \in \mathbb{R}_+^N$, any given agent $n \in \mathcal{N}$ receives a payoff of $F_n(v) \in \mathbb{R}^1$.

- We interpret v_n as the amount that agent n plans to steal and move off-chain for his personal gain.
- If $v_n = 0$, then node n is behaving honestly and also reports any thefts by other nodes.
- If $\sum_n v_n > V$, then we interpret the strategy profile as coordination failure which results in all dishonest nodes being unsuccessful in their attempts to steal tokens.

Formalities

To define our equilibrium notion, we will need some additional definitions and notation:

Coalition: $c \subseteq \mathcal{N}$

Deviation Strategy Profile: Consider a strategy profile $v \in \mathbb{R}_+^N$ for the grand coalition $\mathcal{N}$, and a coalition $c \subseteq \mathcal{N}$. Then, $\hat{v} \in \mathbb{R}_+^N$ is a deviation strategy profile from v for the coalition c if for all $n \notin c$, $\hat{v}_n = v_n$.

Credible Deviation: Let $\hat{v} \in \mathbb{R}_+^N$ be any deviation strategy profile from v for the coalition c. Then $\hat{v} \in \mathbb{R}_+^N$ is a credible deviation if for all subcoalitions $c' \subseteq c$ of the deviating coalition, there does not exist a deviation strategy profile $\tilde{v} \in \mathbb{R}^N$ for coalition c' from $\hat{v} \in \mathbb{R}^N$ such that for all $n \in c'$, $F_n(\tilde{v}) > F_n(\hat{v})$.

Coalition-Proof Equilibrium

Coalition-Proof Equilibrium (CPE): A strategy profile $v \in \mathbb{R}^N$ is a coalition-proof equilibrium if for all coalitions $c \subseteq \mathcal{N}$, there does not exist a credible deviation $\hat{v} \in \mathbb{R}^N$ such that $F_n(\hat{v}) \geq F_n(v)$ for every $n \in c$ and there exists at least one $n \in c$ such that $F_n(\hat{v}) > F_n(v)$.

It is well understood that nodes can form coalitions and collude through mining pools or Sybiling.

Thus, for an outcome to be stable in the context of blockchain validation networks, it absolutely must be coalition-proof, not Nash or dominant strategy.

(Note this is actually a refinement of Coalition-Proof Equilibrium)

Catastrophic Dissent Mechanism

We will need some additional notation to formally define the Catastrophic Dissent Mechanism

V – The total value that a coalition of nodes thinks it can gain through dishonesty.
N – The number of validating nodes.
A – The number of independent agents running validating nodes.
H – The number of nodes who decide to honestly report any bad behavior.
D – The number of nodes that follow the conspiracy and behave dishonestly. $(D+H=N)$
G – The amount held in the GBB of each node.
T – The expected present value of transactions processing fees to nodes who behave honestly.

Catastrophic Dissent Mechanism

$$F_n(v_{1,}\ldots,v_n,\ldots v_N)=$$

T	if $v_n=0 \;\; \forall n\in\{1\ldots N\}$	(all nodes honest)
$T+\frac{GD}{H}$	if $v_n=0$ and $\exists m\in\mathcal{N}$ such that $v_m>0$	(honest with dishonest nodes)
$-G$	if $v_n>0$ and $\exists m\in\mathcal{N}$ such that $v_m=0$	(dishonest with honest nodes)
$-G$	if $v_n>0 \;\; \forall n\in\{1\ldots N\}$ and $\sum_n v_n>V$	(dishonest coordination failure)
v_n	if $v_n>0 \;\; \forall n\in\{1\ldots N\}$ and $\sum_n v_n\leq V$	(all nodes dishonest)

Results

Claim 1: *Suppose that all the nodes join a conspiracy to share the largest gain that dishonest strategies allow. That is, the grand coalition chooses a strategy that satisfies the following:*

$$v=(v_1,\ldots,v_N) \text{ where } \forall n\in\{1\ldots N\},\ v_n>0 \text{ and } \sum_n v_n=V.$$

Then if $G>\frac{V}{N(A-1)}$ *and* $A\geq 2$, v *is not a CPE of the CDM.*

Under these same assumptions*:*

Claim 2: *The grand coalition stealing less than V is not a CPE of the CDM.*

Claim 3: *Any strict subcoalition stealing anything is not a CPE of the CDM.*

Claim 5: *Uniform honesty (v=(0, . . ., 0) is a CPE of the CDM.*

The CDM Implements Honesty in CPE

Theorem 1: *If* $G > \frac{V}{N(A-1)}$ *and* $A \geq 2$, *then the CDM implements honesty is CPE*

Where is Proof of Honesty?

- The CDM is part of a more complex protocol in which nodes must prove their honesty to one another and to users.
- There are additional elements that assure termination and restart block writing in the event of a 100% dishonest pool of nodes.
- Formally, Proof of Honesty (PoH) is 99% Byzantine fault tolerant and offers “Strategically Provably Security.”

Conclusion

- Game theory, especially mechanism design is essential for protocol building.
- Computer Science, especially distributed systems and the limitations of faulty networks is essential.
- Cryptography, especially what it does and does not make provable is essential.
- Macroeconomics is often essential (I can't believe I said that!)
- Finally, knowing what real-world problem you are trying to solve is essential.

This is the best example I know of a truly interdisciplinary problem.

Blockchain needs more economic theorists