Committee-based Blockchains as Games Between Opportunistic players and Adversaries*

Yackolley Amoussou-Guenou
Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France

Bruno Biais
HEC Paris, 1 Rue de la Libération, 78350, Jouy-en-Josas, France

Maria Potop-Butucaru
Sorbonne Université, CNRS, LIP6, F-75005, Paris, France

Sara Tucci-Piergiovanni
Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France

We take a game-theoretic approach to consensus in committee-based blockchains. Proposers send blocks to the other committee members, who can check block validity and vote for blocks. Blocks receiving a qualified majority of votes are accepted. Opportunistic players interact with adversaries. All are rational and play best responses, but opportunistic players maximize expected net rewards, while adversaries seek to disrupt blockchain consensus. Free riding implies the prescribed protocol is not an equilibrium, while there exist equilibria without consensus, even when adversaries are few. There also exists an equilibrium achieving consensus, in which players are endogenously pivotal, which deters free-riding. We propose a simple modification of the prescribed protocol to make this equilibrium a focal point.

Key words: blockchains, committee, consensus, coordination, free-riding, perfect bayesian equilibrium

1. Introduction

Distributed ledgers share information within a network of participants. Blockchains are distributed ledgers recording information in an an evolving list of ordered blocks, each consisting of one or more transactions. There is no central authority deciding which block to add. Instead, participants must reach consensus on the blockchain and its updating. A major challenge is to design a protocol achieving such consensus. Different protocols are employed. In the proof-of-work protocol (see Nakamoto (2008)), participants are miners, and one randomly drawn miner gets to add the new block. The major problem with proof-of-work is that mining consumes a lot of energy (see Stoll et al. (2019), de Vries (2020)). Committee-based protocols, such as, e.g., Algorand, HoneyBadger,

* Thanks for insightful comments to participants at The Economics of Distributed Ledger Technology Conference at Cambridge University, The Central Bank Research Association conference, The Carnegie Mellon University Distinguished Blockchain Seminar, The International Conference on Autonomous Agents and Multi-agents Systems, and also to Denis Gromb, Tristan Tomala, and Luigi Zingales.

HotStuff or Tendermint, are more sustainable. In these protocols, several participants exchange messages to decide (or commit) which block to add (see Gilad et al. (2017), Abraham et al. (2016), Amoussou-Guenou et al. (2018), Decker et al. (2016), Eyal et al. (2016), Kokoris-Kogias et al. (2016), Miller et al. (2016), Yin et al. (2019).)

Computer science offers insights in the interaction between processes in blockchains.[1] An important question in these analyses is whether blockchains can achieve consensus when correct processes, following the prescribed protocol, interact with faulty processes. Put otherwise, the question is whether the blockchain is robust to the risk that some processes are faulty. A particularly damaging type of faulty processes are Byzantine processes, who can take any arbitrary action to disrupt blockchain consensus. Taking a different approach, game theory studies interaction between rational players, following best response strategies which in equilibrium are common knowledge.

At the conjunction between computer science and game theory, in line with the seminal contributions of Abraham et al. (2006), Afek et al. (2014), Aiyer et al. (2005), Groce et al. (2012), Halpern and Vilaça (2016), we study the interaction between opportunistic players and adversaries in committee based blockchains. We assume all players are rational and play best responses, but they have different preferences: *Opportunistic players* maximize expected net rewards, while *adversaries* seek to disrupt blockchain consensus at any cost. We study if consensus arises in the Perfect Bayesian equilibria of that game.

Consensus is achieved when the outcome of the interaction between the processes satisfies the three following properties: *Termination:* every non-adversary participant commits on a value B (i.e., decides which block to add to the chain); *Agreement:* if there is a non-adversary participant that commits a value B, then all the non-adversary participants commit B; *Validity:* a committed value by any non-adversary participant is valid, it satisfies an application-defined predicate locally verifiable by any participant. In our analysis, adversaries endeavour to disrupt consensus.

The workings of the blockchain we study are the following: At each height of the blockchain, a committee is selected. Committee members interact in a sequence of rounds. At the beginning of each round, one of the committee members is selected to be the proposer and sends a block to the others. When receiving a block, committee members decide whether to check its validity or not, and whether to send a vote for the block or not.[2] Checking validity is costly (in terms of electricity and computer space) and so is sending votes. When a block receives a qualified majority of votes in a round it is committed and added to the blockchain. In that case, the network participants move to the next height in the chain. Otherwise, if the number of votes for the block is below

[1] In the present paper, the computer science term "process" can be translated, in game theory terms, as "player;"

[2] Sending a vote message amounts to voting in favour of the block, while not sending any message means that the player is not in favour of the block.

the majority threshold, the block is not added to the chain. In that case, the protocol moves to the next round, in which a new proposer sends a block to the other committee members. When a block is committed, the committee members who voted for the block receive a reward. When an invalid block is committed, opportunistic players incur a cost. While the reward for blocks is specified in the protocol, the cost incurred by opportunistic players reflect the reputational loss of the blockchain in which they participate, and of which they hold the native currency. This sequence of moves and gains and losses defines a dynamic game.

Our first contribution is to write down i) the pseudo-code of the prescribed protocol and the protocols followed by opportunistic players and adversaries, as well as ii) the extensive form of the corresponding dynamic game, which iii) clarifies the link between the two. Our second contribution is to analyze the Perfect Bayesian Equilibria of that game. Key to these equilibria are coordination and free-riding problems among opportunistic players. Coordination problems arise when opportunistic players don't vote, because they anticipate the others won't vote. This problem can easily be solved by reimbursing the cost of sending messages (which are observable). Free-riding problems arise because opportunistic players are tempted not to check a block's validity when they expect the others to check it. Free riding is a severe problem because checking is unobservable and costly. This creates a form of moral hazard. Because of free riding, the prescribed protocol (in which processes are supposed to check validity and vote for valid blocks only) is not an equilibrium. Moreover, there exist equilibria in which consensus is not achieved. We show, however, that there also exists an equilibrium in which consensus obtains, because equilibrium strategies are such that committee members are endogenously pivotal, which deters free riding. We propose a simple modification of the prescribed protocol to make equilibrium consensus a focal point (see Schelling (1960)).

The next section discusses the literature to which our paper is related. Section 3 presents the system model and protocol. Section 4 presents the game and Section 5 its equilibria. Section 6 briefly concludes.

2. Literature

Our analysis is related to the literature on Byzantine consensus-based blockchains. Castro and Liskov (1999) analyzes consensus in a Byzantine Fault Tolerant (BFT) protocol. In order to use a BFT protocol in an open setting such as blockchains, recent research endeavoured to find secure mechanisms to select committees of fixed size over time (e.g. Gilad et al. (2017), David et al. (2018)) and to propose incentives to promote participation Abraham et al. (2016). While, this literature allows messages to be delayed and Byzantine processes to equivocate (see Lamport et al. (1982)), we assume away these difficulties to simplify the problem. On the other hand, while the literature on BFT protocols focuses on correct and Byzantine processes, we in contrast consider rational processes.

Our analysis, therefore, is in line with the papers analyzing blockchains with rational participants. Kiayias et al. (2016), Kroll et al. (2013), and Biais et al. (2019), study Nash equilibria in proof-of-work blockchain protocols. Abraham et al. (2016) presents Solidus, an incentive-compatible BFT protocol for blockchains. Groce et al. (2012) consider an environment with rational and honest participants and provide protocols that tolerate rational adversaries. In contrast with the cooperative game approach in Groce et al. (2012), we take a non-cooperative game approach. Also in a non-cooperative game framework, Halpern and Vilaça (2016), prove that in a fully rational setting, if participants can fail by crashing, there is no *ex post* Nash equilibrium solving the *fair* consensus problem (where fair means that the input of every agent is committed with equal probability), even with only one crash. Afek et al. (2014) proposes protocols solving consensus and renaming under the assumption that participants may be rational and offers a game theoretic analysis.

An important ingredient in models of rational choices is the specification of players' utilities. Lysyanskaya and Triandopoulos (2006) proposes an incentive compatible protocol robust to a coalition of up to f faulty players, when the players' utilities reflect whether a decision is reached or not, as well as the value of the decision. Abraham et al. (2006) proposes an incentive-compatible protocol for secret sharing with rational participants when some utilities can be unknown. Halpern and Vilaça (2020) show there is a Nash equilibrium achieving consensus when an agent's utility depends only on the consensus value achieved and not of the number of messages the agent sends. In contrast with these analyses, in our model utilities are known and reflect not only whether and which blocks are added, but also the costs of the actions (checking, sending) taken by the players. These costs make the analysis more complex, but also give rise to new economic effects, such as free riding.

Manshaei et al. (2018) also offers a non-cooperative game-theoretic analysis of free-riding in committees. In the protocol they consider, multiple committees run in parallel to validate a non-intersecting set of transactions (a shard). They show that rational agents can free-ride when rewards are equally shared. Differences between our analysis and Manshaei et al. (2018) are that we consider i) a dynamic game, ii) the interaction between adversaries and opportunistic players, and that iii) due to the presence of adversaries some blocks can be invalid.

3. System model and protocol

Blockchain: A blockchain is a growing sequence of blocks, to which new blocks can be appended. Once a block is in the blockchain, it cannot be modified nor removed. The block at position $h \geq 0$ in the blockchain is said to be at height h, and the first block at height 0 is the initialization block.

System model: For each given height h, we consider a system composed of a finite and ordered set Π, called *committee*, of synchronous sequential processes or players, namely $\Pi = \{p_1, \ldots, p_n\}$ where process p_i is said to have index i. We assume each player is aware of its index. In the following, we refer to process/player p_i by its index, say process i. Hereafter, the words "player" and "process" are taken to have the same meaning. The protocol defines the sequence of moves followed by the processes when adding blocks to the chain.

Processes behaviour: In this paper, we consider a variant of the BAR model Aiyer et al. (2005) where all processes are rational and are either *opportunistic* or *adversary. Opportunistic processes* are self-interested and seek to maximize their expected utility, which is equal to the expectation of the rewards they obtain net of the costs they incur. Opportunistic processes follow the prescribed protocol if and only if doing so maximizes their expected utility. In line with Aiyer et al. (2005), the objective of *adversary processes* is to prevent the protocol from achieving its goal, and to reduce the opportunistic processes utility, no matter the cost. We denote by f the number of adversary processes in the network.

Consensus properties: Consensus is defined as follows:

Definition 1 *We say that consensus is achieved when the following properties hold:*

- *Termination: every non-adversary process commits on a value (a block);*
- *Agreement: if two non-adversary processes commit respectively on values B and B', then $B = B'$;*
- *Validity: a committed value by any non-adversary process is valid, it satisfies the predefined predicate.*

We use the concept of *external validity* introduced by Cachin et al. (2001). The validity predicate must be known by all processes and is defined by the given application. External validity was later adapted by Crain et al. (2017) as a well-suited validity concept for blockchains.

Rounds, phases and steps: The n processes interact during at most n rounds. Each round is divided in two sequential phases: the PROPOSE phase and the VOTE phase. These two phases encapsulate the main ideas of consensus protocol for committee-based blockchains.[3]

Each phase is divided into three sequential steps: the send step, the delivery step and the compute step. We assume that the send step is atomically executed at the beginning of the phase and the compute step is atomically executed at the end of the phase.

[3] Chan and Shi (2020), extended this two phases approach to multiple communication and failure models. They point out the importance and sufficiency of the PROPOSE and VOTE phases in consensus algorithms for blockchains.

The phase has a fixed duration that allows collecting all the messages sent by the processes at the beginning of the phase during the delivery step. At the end of a phase, a process exits from the current phase and starts the next one.

The processes communicate by sending and receiving messages through a reliable broadcast primitive.[4] Messages are created with a digital signature, and we assume digital signatures cannot be forged. When a process i delivers a message, it knows the process j that created the message. We assume messages cannot be lost.

Algorithm 1 Prescribed Protocol for a given height h at any process i

```
1:  Initialization:
2:    $vote := nil$
3:    $t := 0$                                                    /* Current round number */
4:    $committedValue := nil$

5:  Phase PROPOSE($t$) :
6:    Send step:
7:      if $i ==$ isProposer($t, h$) then
8:        $proposal \leftarrow$ createValidValue($h$)  /* The proposer of the round generates a block, i.e. the value to be proposed */
9:        broadcast $\langle \mathsf{PROPOSE}, h, t, proposal \rangle$
10:   Delivery step:
11:     delivery $\langle \mathsf{PROPOSE}, h, t, v \rangle$ from proposer($t$)        /* The process collects the proposal */
12:   Compute step:
13:     if isValid($v$) then
14:       $vote \leftarrow v$                  /* If the delivered proposal is valid, then the process sets a vote for it */

15: Phase VOTE($t$) :
16:   Send step:
17:     if $vote \neq nil$ then
18:       broadcast $\langle \mathsf{VOTE}_i, h, t, vote \rangle$   /* If the proposal is valid, the process sends the vote for it to all the validators */
19:   Delivery step:
20:     delivery $\langle \mathsf{VOTE}, h, t, v \rangle$          /* The process collects all the votes for the current height and round */
21:   Compute step:
22:     if $|\langle \mathsf{VOTE}, h, t, v \rangle| \geq \nu \wedge committedValue = nil \wedge vote \neq nil \wedge vote = v$ then
23:       $committedValue \leftarrow v$; exit                   /* The valid value is committed if the threshold is reached */
24:     else
25:       $vote \leftarrow nil$
26:       $t \leftarrow t + 1$
```

Prescribed protocol: Algorithm 1 presents the pseudo-code for a correct process, following the prescribed protocol. For each round $t \in \{1, ...n\}$ a committee member is designated to be the proposer for the round in a round robin fashion. The `isProposer`(t, h) function returns the id of the proposer for the current round and height (line 11). The function, by taking as parameter the current height, deterministically selects the proposer on the basis of the information contained in the blockchain up to h (the actual selection mechanism is out of the scope of this paper).

During the PROPOSE phase, the proposer of the round generates a block. In the prescribed protocol, the proposer uses the function `createValidValue`(h), which creates a valid bloc. Because a valid block must include the identifier of the previous block in the blockchain as well as the index

[4] A broadcast is reliable if the following conditions hold: i) safety: every message delivered by a process has been previously sent by a source, and ii) liveliness: every process eventually delivers every message sent by a source.

h where the block should be, the height h is passed as parameter (line 8). Once the block is created, a message broadcasting the proposal is sent (line 9). At line 10 the proposal is received through a delivery function. In the prescribed protocol, each process is correct and checks if the proposal is a valid value (line 13). If so, the process sets its vote to the value (line 14).

During the VOTE phase, any process that set its vote to the current valid proposal sends a message (i.e., a vote) to the other members of the committee (line 18). During the delivery step, sent messages are collected by every process. During the compute step, each process verifies if a quorum of $\nu > 1$ votes for the current proposal has been reached. If the quorum is reached, the process voted for the value and did not already commit for the current height, then it commits for the current proposal (line 23) and the protocol ends. If the quorum is not reached, then a new round starts (line 26).

When there are only correct processes, following the prescribed protocol, consensus is reached.

Algorithm 2 Pseudo-code for a given height h modeling the rational process i's behavior

```
 1: Initialization:
 2:   $vote := nil$
 3:   $t := 0$                                          /* Current round number */
 4:   $committedValue := nil$
 5:   $action^{\text{propose}} := nil$
 6:   $action^{\text{check}} := nil$
 7:   $action^{\text{send}} := nil$
 8:   $validValue[] := \{\bot, \bot, \ldots, \bot\}$     /* $validValue[r] \in \{\bot, 0, 1\}$ */

 9: Phase PROPOSE($t$):
10:   Send step:
11:     if $i == \texttt{isProposer}(h, t)$ then
12:       $action^{\text{propose}} \leftarrow \sigma_i^{\text{propose}}()$   /* $\sigma_i^{\text{propose}} \in \{0, 1\}$ sets the action of proposing a valid block or an invalid one */
13:       if $action^{\text{propose}} == 1$ then
14:         $proposal \leftarrow \texttt{createValidValue}(h)$
15:       else if $action^{\text{propose}} == 0$ then
16:         $proposal \leftarrow \texttt{createInvalidValue}()$
17:       broadcast $\langle \mathsf{PROPOSE}, h, t, proposal \rangle$
18:   Delivery step:
19:     delivery $\langle \mathsf{PROPOSE}, h, t, v \rangle$ from $\mathsf{proposer}(h, t)$
20:   Compute step:
21:     $action^{\text{check}} \leftarrow \sigma_i^{\text{check}}()$   /* $\sigma_i^{\text{check}} \in \{0, 1\}$ sets the action of checking or not the validity of the proposal */
22:     if $action^{\text{check}} == 1$ then
23:       $validValue[r] \leftarrow \texttt{isValid}(v)$   /* The execution of $\texttt{isValid}(\text{v})$ has a cost $c_{\text{check}}$ */
24:     $action^{\text{send}} \leftarrow \sigma_i^{\text{send}}(validValue)$   /* $\sigma_i^{\text{send}} : \{\bot, 0, 1\} \rightarrow \{0, 1\}$ sets the action of sending the vote or not */
25:     if $action^{\text{send}} == 1$ then
26:       $vote \leftarrow v$   /* The process decides to send the vote, the proposal might be invalid */

27: Phase VOTE($t$):
28:   Send step:
29:     if $vote \neq nil$ then
30:       broadcast $\langle \mathsf{VOTE}_i, h, t, vote \rangle$   /* The execution of the broadcast has a cost $c_{\text{send}}$ */
31:   Delivery step:
32:     delivery $\langle \mathsf{VOTE}, h, t, v \rangle$   /* The process collects all the votes for the current height and round */
33:   Compute step:
34:     if $|\langle \mathsf{VOTE}, h, t, v \rangle| \geq \nu \wedge committedValue = nil \wedge vote \neq nil \wedge vote = v$ then
35:       $committedValue = v$; exit
36:     else
37:       $vote \leftarrow nil$
38:       $t \leftarrow t + 1$
```

Pseudo-code for opportunistic processes and adversaries: While correct processes just follow the prescribed protocol, opportunistic and adversary processes choose actions to maximize their objective. Processes choices are represented in the pseudo-code (Algorithm 2) by dedicated variables, namely, $action^{propose}$, $action^{check}$, and $action^{send}$, defined at lines $5-7$. Each action, initialized to nil, can take values from the set $\{0,1\}$. Those values are set by calling the functions $\sigma_i^{\text{propose}}$, σ_i^{check}, and σ_i^{send}, respectively, returning the strategy for the process i.

Strategy $\sigma_i^{\text{propose}}$ determines if the proposer i chooses to produce a valid proposal, in which case $action^{propose}$ takes the value one, or an invalid one, in which case $action^{propose}$ takes the value 0, (lines 12-16). In both cases, the proposal is sent in broadcast (line 17).

Strategy σ_i^{check} determines if the receiving process chooses to check the validity of the proposal, in which case $action^{check}$ takes the value 1, or not, in which case $action^{check}$ takes the value 0. If the process chooses to check the validity (line 22), it also updates its knowledge about the validity of the proposal and it incurs cost c_{check}. Otherwise, the process does not observe if the proposal is valid or not ($validValue[t]$ remains set to $\perp$).

Strategy σ_i^{send} determines if the receiving process chooses to send a vote, in which case $action^{send}$ takes the value 1, or not, in which case $action^{send}$ takes the value 0 (line 24-30). Note that the strategy σ_i^{send} depends on the knowledge the process has about the validity of the proposal. The strategy determines if the process chooses to send its vote for the proposal or not . When a process chooses to send a vote for the proposal, it incurs cost c_{send}.

4. Game

In this section we offer a game theoretic formulation of the protocol presented in the previous section. Recall that the number of adversaries is denoted by f, while the supermajority threshold needed for block acceptance is denoted by ν (number of votes needed to commit a block). Each player i observes its own type, θ_i, which can be adversary ($\theta_i = \theta^a$) or opportunistic ($\theta_i = \theta^s$).

While ν is a parameter of the protocol, commonly known by all participants, f can be random variable, but we assume it is common knowledge that $0 < f < \nu < n - f$. $f < \nu$ means that adversaries don't have the majority, while $\nu < n - f$ means opportunitistic players have the majority. Together these conditions imply that $f < \frac{n}{2}$, i.e, there is a strict majority of opportunistic players. One could expect these assumptions to imply that consensus should be achieved. This is not the case, however. As shown below, termination or validity may fail to hold in equilibrium.

Action space: At each round t the proposer decides which block to send to all the other players. The proposer can thus choose whether to send a valid block or an invalid one. For simplicity we assume that proposing a valid or invalid block is costless.

After receiving the proposed block, each player first decides whether to check the block's validity or not (at cost c_{check}), and then decides whether to send a message (at cost c_{send}) or not.

Information sets: We assume opportunistic players only observe their own type.[5] Adversaries, in contrast are assumed to know the types of all players. This assumption is in line with that, often made in computer science, that adversaries (or Byzantine processes) are very powerful and can collude. The goal is to test whether the blockchain is robust to adversary attacks. If a blockchain resists the attack of very powerful adversaries, it means it is very robust.

The information set of player i, at the beginning each round t, which we denote by η_i^t, includes

- the observation of the round number t,
- the player's own type θ_i, and also when i is an adversary the types of all the other players,
- when $t > 1$, the observation of what happened in previous rounds, namely (i) whether i decided to check validity, and in that case the knowledge of whether the block was valid or not, (ii) how many messages were sent, and (iii) whether a block was produced or not.

Then, in the Compute step of the PROPOSE phase, each player decides whether to check the validity of the current block or not. Denote by b_t the block proposed at round t. When the player does not decide to check validity, $\texttt{isValid}(b_t)$ is the null information set, while if the player decides to check, $\texttt{isValid}(b_t)$ is equal to 1 if the block is valid and 0 otherwise. So, the player information set becomes $H_i^t = \eta_i^t \cup \texttt{isValid}(b_t)$.

Strategies: At each round $t \geq 1$, the strategy of player i is a mapping from its information set into its actions. If the player is selected to propose the block, its choice is given by $\sigma_i^{\text{propose}}(\eta_i^t)$. Then, in the Compute step of the PROPOSE phase, the player's strategy is given by $\sigma_i^{\text{check}}(\eta_i^t)$. Finally, in the Send step of the VOTE phase, the player must decide whether to send a message or not, and that decision is given by $\sigma_i^{\text{send}}(H_i^t)$.

Rewards and Costs for Opportunistic Players: We assume that, when a block is produced, only the players which sent a message receive a reward, denoted by R. This is in line with practice, e.g., in the Tendermint protocol, see Amoussou-Guenou et al. (2018). In addition, we assume that when an invalid block is produced, all opportunistic players incur cost κ.

We assume the reward R is larger than the cost c_{check} of checking validity, which in turn is larger than the cost c_{send} of sending a message. Additionally, we assume that the reward obtained when a block is produced is smaller than the cost κ of producing an invalid block. Thus we overall assume

$$\kappa > R > c_{\text{check}} > c_{\text{send}}.$$

Rewards and Costs for Adversaries. The adversaries have lexicographic preferences over the outcome of the game, in order, they prefer:

1. Outcomes that do satisfy Termination, but not Validity;

[5] If adversaries were detectable, they could be excluded, and therefore could not harm the system.

2. Outcomes that do satisfy Validity, but not Termination;
3. Outcomes that do satisfy Termination and Validity;

Adversaries are assumed to care only about the outcome of the protocol and neglect the costs of their own actions.

Objective of opportunistic players: Let T be the endogenous round at which the game stops. If a block is produced at round $t \leq n$, then $T = t$. Otherwise, if no block is produced, $T = n+1$. In the latter case, the *termination* property is not satisfied.

At the beginning of the first round, the expected gain of opportunistic player i is:

$$U_i = E\left[\begin{array}{c}\left(R * \mathbf{1}_{(\sigma_i^{\text{send}}(H_i^T)=1)} * \mathbf{1}_{(\text{block produced at } T)} - \kappa \mathbf{1}_{(\text{invalid block produced})}\right) \\ -\sum_{t=1}^{T}\left(c_{\text{check}}\mathbf{1}_{(\sigma_i^{\text{check}}(h_i^t)=1)} + c_{\text{send}}\mathbf{1}_{(\sigma_i^{\text{send}}(H_i^t)=1)}\right)\end{array} |\eta_i^1\right],$$

where $\mathbf{1}_{(.)}$ denotes the indicator function, taking the value 1 if its argument is true, and 0 if it is false.

Then, at the beginning of round $t > 1$, if $T \geq t$, the continuation payoff of the opportunistic player with information set η_i^t is

$$W_{i,t}(\eta_i^t) = E\left[\begin{array}{c}\left(R * \mathbf{1}_{(\sigma_i^{\text{send}}(H_i^T)=1)} * \mathbf{1}_{(\text{block produced at } T)} - \kappa \mathbf{1}_{(\text{invalid block produced})}\right) \\ -\sum_{s=t}^{T}\left(c_{\text{check}}\mathbf{1}_{(\sigma_i^{\text{check}}(h_i^s)=1)} + c_{\text{send}}\mathbf{1}_{(\sigma_i^{\text{send}}(H_i^s)=1)}\right)\end{array} |\eta_i^t\right].$$

Equilibrium concept: Since we consider a dynamic game with asymmetric information, the relevant equilibrium concept is Perfect Bayesian Equilibrium Fudenberg and Tirole (1991), intuitively defined as follows:

Definition 2 *A Perfect Bayesian equilibrium an n-tuple of strategies, such that all players 1) choose strategies that are optimal given their preferences and beliefs, 2) rationally anticipate the strategies of the others, and 3) draw rational inferences from what they observe, using their expectations about the strategies of the others and Bayes law, whenever it applies.*

A Perfect Bayesian Equilibrium (PBE) is a Nash equilibrium, Nash (1951), so players best-respond to one another. It imposes additional restrictions, to take into account the fact that the game is dynamic and that players can have private information and must draw rational inferences. Rationality implies that each player's beliefs are consistent with Bayes law, when computing probabilities conditional on events that have strictly positive probability on the equilibrium path. Perfection implies that at each node starting a subgame the players' strategies form a Nash equilibrium of that subgame. In this context, to show that a strategy is optimal it is sufficient to show that it dominates any one-shot deviation, Blackwell (1965).

5. Equilibria

5.1. Equilibrium without termination

Suppose an opportunistic player anticipates the others won't send any vote. In that case, the player anticipates that, even if he/she sent a vote, no block would be committed, since there would be only one vote and $\nu > 1$. So, when an opportunistic player anticipates the others won't vote, he/she prefers to also abstain from voting: his/her message would not change the outcome of the vote, and would only result in the player incurring cost c_{send}. A fortiori, it is suboptimal for the player to check validity, since it would only result in incurring cost c_{check} without otherwise changing the outcome of the game.

Now turn to adversaries, who, in contrast with opportunistic players, can coordinate their moves. Should they choose to all vote for invalid blocks? Such a move would result in $f < \nu$ votes in favour of the invalid block. So the block would still not be committed. Consequently, adversaries (weakly) prefer not to send votes, and, a fortiori, not to check validity.

The above remarks imply that, when players expect the others not to vote, their best response is also not to vote. This implies no block is ever committed. In this context, it is weakly optimal for opportunistic proposers to send valid and for adversary proposers to send invalid blocks.

These remarks are summarized in our first proposition:

Proposition 1 *There exists an equilibrium in which an opportunistic proposer sends a valid block and an adversary proposer sends an invalid block, while all other players do nothing, i.e., they neither check validity nor send messages.*

In the equilibrium of Proposition 1, no block is ever produced, i.e., there is no termination. This failure of consensus reflects a coordination failure among players, who coordinate on a bad equilibrium in which no one votes.

Now, votes are observable. So one could modify the protocol to reimburse c_{send} to the players. As shown below, this would not be enough to eliminate coordination problems.

5.2. Equilibrium without validity

Our next proposition shows there can also be coordination failures with regard to block validity checks.

Proposition 2 *There exists an equilibrium in which: i) An opportunistic proposer sends a valid block while an adversary sends an invalid block. ii) When receiving blocks, opportunistic players don't check validity but send a vote. iii) Adversaries vote in favour of invalid blocks. Thus, when the proposer is opportunistic a valid block is committed, while when the proposer is adversary an invalid block is committed.*

In the equilibrium of Proposition 2 there is termination, at the first round, but validity does not always hold. Opportunistic proposers send valid blocks. It is optimal for them to do so because it leads to the production of a valid block, earning them reward R, while if they had sent an invalid block, which would also have been committed, their payoff would have been $R - \kappa$. Adversaries, however, find it optimal to send invalid blocks, which are committed. Adversaries, of course, also find it optimal to send a vote when an invalid block is proposed, while they find (at least weakly) suboptimal to vote for valid blocks.

The main problem in the equilibrium of Proposition 2 is the behaviour of opportunistic players, who send votes without checking validity. Why do they find it optimal to do so? Why don't they prefer to follow the prescribed behaviour, which is to check validity and send a vote only if the block is valid?

The problem is that, when the other opportunistic players follow the equilibrium strategies, an opportunistic player deviating to the prescribed behaviour would be unable to alter the outcome of the protocol. Valid blocks would still get $n - f$ votes and be committed, while invalid blocks would get $n - 1$ votes and also be committed. Thus following the prescribed behaviour would give opportunistic players the following expected payoff

$$R - c_{\text{check}} - \Pr(\text{valid}) * c_{\text{send}} - \Pr(\text{invalid}) * \kappa,$$

which is lower than their equilibrium payoff

$$R - c_{\text{send}} - \Pr(\text{invalid}) * \kappa$$

because

$$c_{\text{check}} > (1 - \Pr(\text{valid})) * c_{\text{send}}.$$

Another deviation that could be interesting for opportunistic players would be to abstain from voting (and from checking validity).[6] This would however give expected payoff $-\Pr(\text{invalid}) * \kappa$, which is lower than the equilibrium payoff.

Thus, again, consensus fails to obtain due to a coordination failure among opportunistic players. Each opportunistic player finds it suboptimal to check block quality when the others don't. In a sense, checking validity (and voting for valid blocks only) can be interpreted as producing a public good. In the equilibrium of Proposition 2, opportunistic players free ride on the supply of the public good. In a sense the problem is more acute than in Proposition 1: In the case of Proposition 1, one way to solve the coordination problem was to reimburse the cost of sending messages. This does

[6] The other deviations: check but don't vote or check and vote irrespective of validity are obviously dominated by the equilibrium action.

not work for Proposition 2 because validity checks, while costly, are unobservable. In that sense, Proposition 2 reflects moral hazard among the opportunistic players.

Our game theoretic approach, in which all players (or processes) are rational, contrasts with the standard approach in computer science, in which some processes are correct and assumed to follow the prescribed protocol, while other processes are Byzantines. In that standard computer-science approach a typical result is that, if the number of Byzantines is lower than a threshold, then consensus obtains. In contrast, Proposition 2 states that even if f is very low, as long as it is not zero consensus may fail. This is because, in addition to attacks by adversaries, we consider a new source of fragility: coordination problems and free riding among rational, but opportunistic, players.

5.3. The prescribed protocol is not an equilibrium

While the above analysis shows that deviations from the prescribed protocol can form an equilibrium, it does not rule out the possibility that the prescribed protocol would be an equilibrium strategy. We now examine that point. Recall that the prescribed protocol entails proposing valid blocks, ii) checking validity of block received and iii) sending vote for valid blocks only. If the $n-f$ opportunistic players follow that strategy, then consensus obtains: As long as the proposer is an adversary the block is invalid and rejected, and the first time the proposer is opportunistic, the proposed block is valid and a majority $n-f>\nu$ of committee members votes for that block, which is therefore committed. Unfortunately, as stated in the next proposition, it is not an equilibrium that all opportunistic players follow the prescribed strategy at all rounds.

Proposition 3 *If $\nu > f+1$, then it is not an equilibrium that all opportunistic players follow the prescribed protocol at all rounds.*

Proof of Proposition 3: In the candidate equilibrium all opportunistic players follow the prescribed strategy, while adversaries propose invalid blocks and then send votes for invalid blocks. Suppose that during the first $f-1$ rounds the proposer is an adversary, so that the proposed block is rejected and we reach round f. At round f, the block is valid with probability $\frac{n-f}{n-f+1}$ and invalid with the complementary probability. Suppose $n-f-1$ opportunistic players follow the prescribed strategy. If the last opportunistic player also plays the prescribed strategy, he/she gets expected payoff equal to

$$-c_{\text{check}} + \frac{n-f}{n-f+1}(R - c_{\text{send}}) + \frac{1}{n-f+1}(R - c_{\text{send}}),$$

where the first term is the cost of checking the validity of the block proposed at round f, the second term is the probability that that block is valid multiplied by the net reward in that case, and the third term is the expected net reward from round $f+1$. The candidate equilibrium expected payoff at round f simplifies to

$$-c_{\text{check}} + (R - c_{\text{send}}).$$

Now suppose the last opportunistic player deviates from the prescribed strategy by sending a vote at round f without checking validity. If the block is valid, it gets $n-f > \nu$ votes and gets committed, so that the opportunistic player gets payoff $R - c_{\text{send}}$. If the block is invalid, the number of votes it gets is $f+1 < \nu$. So the block is not committed and we move to round $f+1$ at which the block is committed. In that case, the opportunistic player gets payoff $-c_{\text{send}} + (R - c_{\text{send}})$. Thus the overall expected payoff from the deviation is

$$-\frac{c_{\text{send}}}{n-f+1} + (R - c_{\text{send}}),$$

which is larger than the expected payoff from the prescribed strategy.

At rounds prior to f similar arguments apply: Now suppose that during the first $t-1$ rounds the proposer is an adversary, so that the proposed block is rejected and we reach round $t < f$. At round t, the block is valid with probability $\frac{n-f}{(n-f)+(f-(t-1))} = \frac{n-f}{n-t+1}$ and invalid with the complementary probability. Suppose that during the first $t-1$ rounds all opportunistic followed the prescribed strategy. Consider one opportunistic player, who anticipates that the other opportunistic players will follow the prescribed strategy at round t and all opportunistic players will follow that strategy afterwards. If the opportunistic player follows the prescribed strategy, then if the block is valid it is committed, while if it is invalid the block is not committed and we move to the next round. What happens if instead the opportunistic player deviates from the prescribed strategy by sending a vote without checking block validity. If the block is valid, it still gets $n-f > \nu$ votes and gets committed, while if the block is invalid, the block gets is $f+1 < \nu$ votes and is still not committed and we move to the next round. If the block is not committed and we move to the next round, if the opportunistic player reverts to the prescribed behaviour, and the others continue to follow it, all get the payoff obtained when the prescribed behaviour is observed during the whole game. Thus, for the opportunistic player the prescribed behaviour and the deviation give the same payoff, except that at round t the opportunistic player incurs cost c_{check} when following the prescribed behaviour and not when deviating. Hence following the prescribed behaviour is dominated by deviating.

Q.E.D.

Note Proposition 3 result obtains whatever the majority threshold ν, as long as $f+1 < \nu$. In particular the result holds even if $\nu = n-f$ so that the votes of all opportunistic traders are needed for the valid block to be committed. While $\nu = n-f$ makes opportunistic players pivotal in the vote (if one of them fails to vote the valid block is not committed), it does not make them pivotal in validity checks. Thus opportunistic players are tempted to free ride on validity checking, preventing the prescribed strategy from being an equilibrium.

6. Equilibrium consensus

Our results, so far, are negative: While the prescribed protocol would achieve consensus, it is not an equilibrium. Moreover, deviations from the prescribed protocol, leading to consensus failure, constitute equilibria. Is the committee-based protocol we examine doomed to fail? Fortunately, the answer to that question is negative. In this section we show that there exists an equilibrium in which consensus is achieved.

To have termination and validity, it must be that, in equilibrium, while adversaries propose and send messages for invalid blocks, sufficiently many opportunistic players find it in their own interest to check the validity of the block and send messages in support of valid blocks. The problem in the above analysis was that opportunistic players were tempted to free-ride, and let the others bear the cost of checking. To avoid this situation, it must be that (at least some) opportunistic players anticipate they are pivotal, i.e., if they fail to check block validity and send messages in support of valid blocks, this may derail the player at their own expense. In this section, we show present equilibrium in which the strategies of the opportunistic players imply they are endogenously pivotal. To perform this analysis, we need to introduce two additional pieces of notation. First, we denote by $\bar{f}$ the common knowledge upper bound on the number of adversaries. Second, we denote by $f(t)$ the conditional expectation of $\tilde{f}$ given that its realization is larger than or equal to t.

In this equilibrium a key ingredient is that some opportunistic players are expected to check validity, while others are not. It is the presence of opportunistic players who don't check validity that makes the other opportunistic players pivotal with some probability. As a first step in the equilibrium construction, we start with a characterization of the expected continuation payoff of opportunistic players in this context. To present that characterization, we rely on property P defined below:

Definition 3 *A function g satisfies property P, if $g(t) = 1 + \frac{f(t)-t+1}{n-t+1} g(t+1), \forall t < f$.*

Equipped with that definition we can state our next proposition:

Proposition 4 *Consider a candidate equilibrium in which i) some opportunistic players check the validity of the block and send a message if and only if the block is valid, ii) the other opportunistic players send messages without checking validity, iii) when they are proposed, valid blocks are committed while invalid blocks are rejected. In such an equilibrium, if it exists, on the equilibrium path the continuation payoff, at round t, of the opportunistic players who are to check block validity is*

$$\pi_{check}(t) = (R - c_{send}) - \phi(t) c_{check}, \ \forall t \leq \bar{f} + 1, \text{ with } \phi(\bar{f} + 1) = 0, \tag{1}$$

while the expected continuation payoff, at round t, of the opportunistic players who are not to check block validity is

$$\pi_{send}(t) = R - \psi(t)c_{send},\ \forall t \leq \bar{f}+1,\ \textit{with}\ \psi(\bar{f}+1) = 1, \tag{2}$$

where both ϕ and ψ satisfy property P.

The intuition of Proposition 4 the following: In the candidate equilibrium invalid blocks are rejected and, as soon as a valid block is proposed, it is committed. When a valid block is committed, opportunistic participants, who all send a message, get payoff $R - c_{\text{send}}$. This is the first term in $\pi_{check}(t)$, reflecting that eventually a valid block will end up committed. The second part of $\pi_{check}(t)$, $\phi(t)c_{\text{check}}$, is the expected cost of checking the block validity, where $\phi(t)$ is the expected number of times (from round t on) the player expects to check validity before a block is committed. $\phi(\bar{f}) = 1$ reflects that, when round $\bar{f}$ is reached, opportunistic players know this is the last time they have to check block validity. Similarly, in $\pi_{send}(t)$, $\psi(t)c_{\text{check}}$, is the expected cost of sending messages, where $\psi(t)$ is the expected number of times the player expects to send messages before a block is committed. The proof of Proposition 4 offers a more detailed analysis of (1) and (2).

Proof of Proposition 4:

1) In the first part of the proof, we prove that the round t continuation payoff of opportunistic players that are supposed to check validity is as in (1). To do so we proceed by backward induction.

First, we establish that (1) holds at round $t = \bar{f}+1$. At round $t = \bar{f}+1$, players know that all $\bar{f}$ previous proposers were adversary and that now there are only opportunistic proposers. So they don't check block validity and send a message, which gfives them payoff

$$R - c_{\text{send}},$$

so that (1) holds at $t = \bar{f}+1$.

Now turn to round $t < \bar{f}+1$. If round $t < \bar{f}+1$ is reached, the previous $t-1$ proposers were adversaries. There remains $n-(t-1)$ potential proposers. The probability that the next proposer is adversary is

$$\frac{f(t)-(t-1)}{n-t+1},$$

and, with the complementary probability,

$$\frac{n-f(t)}{n-t+1}$$

the next proposer is opportunistic.

Then we prove that if (1) holds at $t+1 \leq \bar{f}+1$, i.e.,

$$\pi_{check}(t+1) = R - c_{\text{send}} - \phi(t+1)c_{\text{check}},$$

then (1) holds at round t. Suppose the opportunistic player follows the equilibrium strategy of checking and sending iff the block is valid. Its expected gain from round t on is

$$-c_{\text{check}} + \frac{n-f(t)}{n-t+1}(R - c_{\text{send}}) + \frac{f(t)-(t-1)}{n-t+1}\pi_{\text{check}}(t+1),$$

where the first term is the cost of checking the block at round t, the second term is the probability that the block is valid and committed multiplied by the payoff in that case, and the third term is the probability that the block is invalid and rejected multiplied by the payoff in that case. Substituting the value of $\pi_{check}(t+1)$ from (1) evaluated at round $t+1$, the expected gain writes as

$$-c_{\text{check}} + \frac{n-f(t)}{n-t+1}(R - c_{\text{send}}) + \frac{f(t)-(t-1)}{n-t+1}(R - c_{\text{send}} - \phi(t+1)c_{\text{check}}).$$

That is

$$R - c_{\text{send}} - \left(1 + \frac{f(t)-(t-1)}{n-t+1}\phi(t+1)\right)c_{\text{check}},$$

which, using property P, is $R - c_{\text{send}} - \phi(t)c_{\text{check}}$. Consequently, (1) holds at round t.

2) In the second part of the proof, we prove that the round t continuation payoff of opportunistic players that are not supposed to check validity is as in (2):

Again, we proceed by backward induction, proving that if the property is satisfied at round $t+1 \leq \bar{f}+1$, i.e., $\pi_{send}(t+1) = R - \psi(t+1)c_{\text{send}}$, then it is satisfied at round t. Suppose the opportunistic player follows the equilibrium strategy of not checking blocks' validity and always sending a message. On the equilibrium path its expected gain from round t on is

$$c_{\text{send}} + \frac{n-f(t)}{n-t+1}R + \frac{f(t)-t+1}{n-t+1}\pi_{send}(t+1),$$

where the first term is the cost of sending a message at round t, the second term is the probability that the block is valid and committed multiplied by the payoff in that case, and the third term is the probability that the block is invalid and rejected multiplied by the payoff in that case. Substituting the value of $\pi_{send}(t+1)$ from (2) evaluated at $t+1$, the expected gain writes as

$$-c_{\text{send}} + \frac{n-f(t)}{n-t+1}R + \frac{f(t)-t+1}{n-t+1}(R - \psi(t+1)c_{\text{send}}).$$

That is

$$R - \left(1 + \frac{f(t)-t+1}{n-t+1}\psi(t+1)\right)c_{\text{send}},$$

which, by Property P, is $R - \psi(t)c_{\text{send}}$.

Q.E.D.

Relying on Proposition 4, we now describe more precisely our candidate equilibrium and prove that it is indeed an equilibrium. Before doing so, we need to introduce some notation: Denote the highest index of all adversary players by i_A; formally, $i_A = \max\{i : \theta_i = \theta^a\}$ and define:

$$\alpha(t) = \frac{(n-t+1)\phi(t) - (f(t)-(t-1))(1-\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t = \theta^a,\ f=\bar{f}))\phi(t+1)}{(f(t)-t+1)\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})}$$

and

$$\beta(t) = \frac{1}{\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})} - 1.$$

$\alpha(t)$ and $\beta(t)$ are complicated functions, for which closed form solutions are not readily available, but they depend only on exogenous parameters and are themselves exogenous objects.[7]

Equipped with these notations, we can now state our next proposition:

Proposition 5 *If the cost κ of committing an invalid block is large enough, in that*

$$\kappa > \alpha(t)c_{check} - \beta(t)c_{send}, \forall t < f, \tag{3}$$

and the reward is large enough in that

$$R \geq \max\left[\frac{n-t+1}{n-f(t)}c_{send}, c_{send} + \frac{n-t+1}{n-f(t)}c_{check}\right], t \leq \bar{f}+1, \tag{4}$$

then there exists a Perfect Bayesian Nash equilibrium achieving consensus, in which the strategy of opportunistic players is the following:

- *As proposer, a opportunistic player proposes a valid block, while an adversary proposes an invalid block.*
- *At any round $t \leq \bar{f}$, when receiving a proposed block, (i) the opportunistic players with index $i \in \{t, \ldots, n-\nu+\bar{f}+1\}$ check the block validity and send a message only if the block is valid, while (ii) the opportunistic players with index $i \in \{n-\nu+\bar{f}+2, \ldots, n\}$ do not check the validity of the block but send a message, and (iii) adversaries check the blocks' validity and send a message if and only if the block is invalid.*
- *If round $t = \bar{f}+1$ is reached, opportunistic players send a message without checking if the block is valid. At this point the block is valid and committed. So in equilibrium termination occurs no later than round $\bar{f}+1$.*

[7] In economics exogenous parameters (concerning, e.g., technologies or prefeences) are given and cannot be modified by agents. In contrast, endogenous variables result from the choices made by the agents.

On the equilibrium path, invalid blocks (proposed by adversaries) are rejected, while valid blocks (proposed by opportunistic players) are committed. This implies that, if round $t = \bar{f} + 1$ is reached, the players know that the number of adversaries was $\bar{f}$ and that during the $\bar{f}$ previous rounds the proposers were adversaries (to draw this inference, the opportunistic players use their anticipation that all participants play equilibrium strategies). Consequently, at round $\bar{f} + 1$, the proposer must be opportunistic, and all players anticipate the proposed block is valid. So, no opportunistic player needs to check the validity of the block but all send a message, which brings them expected gain equal to $R - c_{\text{send}}$. This is larger than their gain from deviating (e.g., by not sending a message or by checking the block.)

Similarly, at rounds $t \leq \bar{f}$, players know that all $t - 1$ previous proposers were adversaries and expect that there remains $f(t) - t + 1$ adversaries with index strictly larger than $t - 1$. In this context, do the equilibrium strategies of the opportunistic players preclude to commit on an invalid block by adversary processes? To examine this point, consider the maximum possible number of messages that can be sent if the proposer is an adversary. In equilibrium the $\nu - \bar{f} - 1$ players with indexes strictly larger than $n - \nu + \bar{f} + 1$ are to send a message without checking it. The worst case scenario (maximizing the number of messages sent when the block is invalid) is that none of these players are adversaries and that $f = \bar{f}$. In that case, in equilibrium, the number of messages sent when the block is invalid is $\bar{f} + (\nu - \bar{f} - 1) = \nu - 1$, so that we narrowly escape validation of the invalid block. If, in that worst case scenario one of the opportunistic players deviated from equilibrium and sent a message without checking the block, this would lead to committing an invalid block. Thus, in that sense, the opportunistic players with index lower than $n - \nu + \bar{f} + 1$ are pivotal. Hence they check block validity, because, under the condition stated in the proposition, the cost of committing an invalid block is so large that opportunistic players do not want to run the risk of tipping the balance.

One could worry that the equilibrium in Proposition 5 is a bit complex, making it hard for players to discover equilibrium and coordinate on it. This difficulty could be circumvented by modifying the prescribed protocol. In the prescribed protocol all players are instructed to check block validity. In the modification of the prescribed protocol we suggest, players would be instructed to check block validity if and only if their index is strictly lower than $n - \nu + \bar{f} + 2$. Proposition 5 implies that, when opportunistic processes expect the others to follow the modified prescribed protocol, their best response is to also follow the modified prescribed protocol: following the modified prescribed protocol is an equilibrium, and it achieves consensus. Thus, the modified prescribed protocol can be interpreted as a way to make equilibrium consensus a focal point, in the sense of Schelling (1960).

Proof of Proposition 5: 1) Our first step is to rule out possible deviations for adversaries. If an adversary deviated by sending a valid block, the block would be immediately committed, which would not make the adversary strictly better off. Moreover, since adversaries neglect the costs of checking validity and sending messages, they weakly prefer to check validity and send messages for invalid blocks, although these two actions have no impact on the outcome of the protocol.

2) The second step is to note that opportunistic players strictly prefer to propose a valid block than an invalid one. This is because, if the opportunistic player follows the equilibrium strategy of proposing a valid block, it is committed and the proposer is rewarded, while if the opportunistic player proposes an invalid block, it is rejected, and we move to the next round, in which case, in equilibrium, the player gets at most $R - c_{\text{check}} - c_{\text{send}}$ and possibly less.

3) The third step is to analyze the actions of the opportunistic players when they receive blocks at round $t = \bar{f} + 1$. At that round, all players know the proposer must be opportunistic and the proposed block valid. In equilibrium, no opportunistic checks validity but all send a message. Any other action would be dominated.

4) The fourth step is to analyze the actions of the opportunistic players with index $i \in \{t, \ldots, n - \nu + \bar{f} + 1\}$ when they receive blocks at round $t < \bar{f} + 1$. On the equilibrium path, these opportunistic players check block validity. To prove equilibrium we must show they prefer to do so rather than deviating once, by voting without checking, and then returning to their equilibrium strategy. When round t is reached, players know that the $t-1$ previous proposers where adversaries. So the average fraction of adversaries, among the $n - t + 1$ players who have not been proposers yet is $\frac{f(t)-(t-1)}{n-t+1}$. In this context, if the next proposer is an adversary, $i_A \leq n - \nu + \bar{f} + 1$ and $f = \bar{f}$, then an opportunistic player who is supposed to check block validity is pivotal. Indeed, in that case the $\bar{f}$ adversaries vote for the block, as well as the $n - (n - \nu + \bar{f} + 2) + 1 = \nu - \bar{f} - 1$ opportunistic players who are not supposed to check validity. So, if an opportunistic player who is supposed to check block validity deviates and votes without checking, the total number of votes is $\bar{f} + (\nu - \bar{f} - 1) + 1 = \nu$ so the block is committed. On the other hand, if $f < \bar{f}$ or $i_A > n - \nu + \bar{f} + 1$, then an opportunistic player who is supposed to check block validity is not pivotal. Thus, the expected gain from the one-shot deviation "vote without checking" at round t is

$$\begin{aligned}
&\left(1 - \frac{f-(t-1)}{n-t+1}\right)(R - c_{\text{send}}) \\
&+ \frac{f(t)-(t-1)}{n-t+1}\Pr(f = \bar{f} | f \geq t-1)\Pr(i_A \leq n - \nu + \bar{f} + 1 | \theta_t = \theta^a,\ f = \bar{f})\,(R - c_{\text{send}} - \kappa) \\
&+ \frac{f(t)-(t-1)}{n-t+1}(1 - \Pr(f = \bar{f} | f \geq t-1)\Pr(i_A \leq n - \nu + \bar{f} + 1 | \theta_t = \theta^a,\ f = \bar{f}))\,(\pi(t+1) - c_{\text{send}})\,.
\end{aligned}$$

- The first term is the expected payoff of the deviating opportunistic player from the case in which the current block is valid and therefore immediately committed.

- The second term is the expected payoff of the deviating player in the "worst case scenario" in which he was pivotal and triggered to commit an invalid block.

- The third term corresponds to the case in which the deviating opportunistic player is not pivotal, and the invalid block is not committed so that we move to the next round.

Substituting the value of $\pi_{check}(t+1) = R - c_{\text{send}} - \phi(t+1)c_{\text{check}}$, the expected continuation value of the deviating player is

$$\begin{aligned}&\left(1 - \frac{f-(t-1)}{n-t+1}\right)(R - c_{\text{send}})\\ &+\frac{f(t)-(t-1)}{n-t+1}\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})\,(R-c_{\text{send}}-\kappa)\\ &+\frac{f(t)-(t-1)}{n-t+1}(1-\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f}))\,(R-c_{\text{send}}-\phi(t+1)c_{\text{check}}-c_{\text{send}})\,.\end{aligned}$$

Or

$$\begin{aligned}&R-c_{\text{send}}-\frac{f(t)-(t-1)}{n-t+1}\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})\kappa\\ &-\frac{f(t)-(t-1)}{n-t+1}(1-\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f}))\,(\phi(t+1)c_{\text{check}}+c_{\text{send}})\,.\end{aligned}$$

The equilibrium condition is that this deviation payoff must be lower than the equilibrium continuation payoff of the player

$$R - c_{\text{send}} - \phi(t)c_{\text{check}}.$$

That is

$$\begin{aligned}&R-c_{\text{send}}-\frac{f(t)-(t-1)}{n-t+1}\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})\kappa\\ &-\frac{f(t)-(t-1)}{n-t+1}(1-\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f}))\,(\phi(t+1)c_{\text{check}}+c_{\text{send}})\\ &\leq R-c_{\text{send}}-\phi(t)c_{\text{check}}.\end{aligned}$$

That is

$$\begin{aligned}&\frac{f(t)-(t-1)}{n-t+1}\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})\kappa \geq \phi(t)c_{\text{check}}\\ &-\frac{f(t)-(t-1)}{n-t+1}(1-\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f}))\,(\phi(t+1)c_{\text{check}}+c_{\text{send}})\,.\end{aligned}$$

Or

$$\begin{aligned}&(f(t)-(t-1))\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})\kappa \geq (n-t+1)\phi(t)c_{\text{check}}\\ &-(f(t)-(t-1))(1-\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f}))\,(\phi(t+1)c_{\text{check}}+c_{\text{send}})\,.\end{aligned}$$

$$\begin{aligned}&\kappa \geq\\ &\frac{(n-t+1)\phi(t)-(f(t)-(t-1))(1-\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f}))\phi(t+1)}{(f(t)-(t-1))\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})}c_{\text{check}}\\ &-\frac{(1-\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f}))}{\Pr(f=\bar{f}|f \geq t-1)\Pr(i_A \leq n-\nu+\bar{f}+1|\theta_t=\theta^a,\ f=\bar{f})}c_{\text{send}}.\end{aligned}$$

which by definition is equivalent to

$$\kappa > \alpha(t)c_{\text{check}} - \beta(t)c_{\text{send}},$$

as stated in the proposition.

The other possible deviations for the opportunistic player supposed to check block's validity are easier to rule out:

First, the player could do nothing (neither check nor send). Relative to the equilibrium payoff, this deviation economizes the cost of checking (c_{check}). If the current proposer is an adversary, from time $t+1$ on the player then obtains the same payoff after this one-shot deviation as on the equilibrium path. If the current proposer is opportunistic, the block gets committed, but the player does not earn any reward. So the deviation is dominated if

$$\frac{n-f(t)}{n-t+1}(R - c_{\text{send}}) \geq c_{\text{check}},$$

which holds under condition (4).

Second, the player could check the block validity, and then send a message irrespective of whether the block is valid or not. This would generate a lower payoff than the main deviation, shown above to be dominated.

Third, the player could check validity but then send no message. When the current proposer is an adversary, this one-shot deviation yields the same payoff as the equilibrium strategy. When the current proposer is opportunistic, this deviation yields a payoff of $-c_{\text{check}}$, which is lower than the equilibrium payoff $R - c_{\text{send}} - c_{\text{check}}$.

Fourth, the player could check the block's validity and send a message only if the block is invalid, which is trivially dominated.

5) The fifth step is to analyze the actions of the opportunistic players with index $i \in \{n - \nu + \bar{f} + 2, ..., n\}$ when they receive blocks at round $t < \bar{f} + 1$. In equilibrium, these opportunistic players send messages without checking blocks' validity. To finalize the proof we need to show they prefer to follow this equilibrium strategy rather than deviating.

First, consider the possibility to abstain from sending a message. This economizes the costs c_{send}, but, in case the block is valid and committed, this implies the agent loses the reward R. So, the deviation is dominated if

$$\frac{n-f(t)}{n-t+1}R \geq c_{\text{send}},$$

which holds under condition (4).

Second, consider the possibility of checking validity and sending a message only for valid blocks. This deviation would imply the agent would have to incur the cost of checking (c_{check}), but it would economize the cost of sending a message when the block is invalid. So the deviation is dominated if

$$c_{\text{check}} \geq \frac{f(t) - t + 1}{n - t + 1} c_{\text{send}},$$

which holds, since by assumption $c_{\text{check}} \geq c_{\text{send}}$.

Other deviations, such as checking validity but never sending messages, or checking validity and always sending messages, or checking validity and sending only if the block is invalid, are trivially dominated.

Q.E.D.

7. Conclusion

In distributed ledgers, such as blockchains, there is no central authority. The advantage is that this eliminates the risk of a malevolent central authority harming network participants. The drawback is that this creates the risk of coordination failures and free riding. We show that coordination failures can arise concerning vote messages. This is because a minimum number of votes is requested for a decision to be made, so that sending a vote is useless when the others don't vote. In that sense, votes are strategic complements. We also show there can be free riding concerning validity checks. In contrast with votes, validity checks are strategic substitutes: a player prefers not to check validity if he/she expects the others to check. Thus, we show that, because of coordination failures and free-riding, there exist equilibria in which Termination or Validity fail. On the other hand, we also show that there exists an equilibrium in which consensus (with Termination and Validity) is achieved. An important issue is how to avoid bad equilibria, without consensus. We suggest a modification of the prescribed protocol to make equilibrium consensus a focal point in the sense of Schelling (1960). Another promising approach to ensuring desirable equilibrium outcomes is to rely on global games and mechanism design, as in Auer et al. (2021).

References

Abraham I, Dolev D, Gonen R, Halpern JY (2006) Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation. *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing, PODC 2006, Denver, CO, USA, July 23-26, 2006*, 53–62.

Abraham I, Malkhi D, Nayak K, Ren L, Spiegelman A (2016) Solidus: An incentive-compatible cryptocurrency based on permissionless byzantine consensus. *CoRR* abs/1612.02916v1.

Afek Y, Ginzberg Y, Feibish SL, Sulamy M (2014) Distributed computing building blocks for rational agents. *ACM Symposium on Principles of Distributed Computing, PODC '14, Paris, France, July 15-18, 2014*, 406–415.

Aiyer AS, Alvisi L, Clement A, Dahlin M, Martin J, Porth C (2005) BAR fault tolerance for cooperative services. *Proceedings of the 20th ACM Symposium on Operating Systems Principles 2005, SOSP 2005, Brighton, UK, October 23-26, 2005*, 45–58.

Amoussou-Guenou Y, Del Pozzo A, Potop-Butucaru M, Tucci-Piergiovanni S (2018) Correctness of tendermint-core blockchains. *22nd International Conference on Principles of Distributed Systems, OPODIS 2018, December 17-19, 2018, Hong Kong, China*, 16:1–16:16.

Auer R, Monnet C, Shin HS (2021) Permissioned distributed ledgers and the governance of money. Technical Report 924, BIS working paper, Basel, Switzerland.

Biais B, Bisière C, Bouvard M, Casamatta C (2019) The blockchain folk theorem. *The Review of Financial Studies* .

Blackwell D (1965) Discounted dynamic programming. *The Annals of Mathematical Statistics* 36(1):226–235.

Cachin C, Kursawe K, Petzold F, Shoup V (2001) Secure and efficient asynchronous broadcast protocols (extended abstract. *in Advances in Cryptology: CRYPTO 2001*, 524–541 (Springer).

Castro M, Liskov B (1999) Practical byzantine fault tolerance. *Proceedings of the Third USENIX Symposium on Operating Systems Design and Implementation (OSDI), New Orleans, Louisiana, USA, February 22-25, 1999*, 173–186.

Chan BY, Shi E (2020) Streamlet: Textbook streamlined blockchains. *IACR Cryptol. ePrint Arch.* 2020:88.

Crain T, Gramoli V, Larrea M, Raynal M (2017) (Leader/Randomization/Signature)-free Byzantine Consensus for Consortium Blockchains. `http://csrg.redbellyblockchain.io/doc/ConsensusRedBellyBlockchain.pdf`.

David B, Gazi P, Kiayias A, Russell A (2018) Ouroboros praos: An adaptively-secure, semi-synchronous proof-of-stake blockchain. *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part II*, 66–98.

de Vries A (2020) Bitcoin's energy consumption is underestimated: A market dynamics approach. *Energy Research & Social Science* 70:101721, ISSN 2214-6296.

Decker C, Seidel J, Wattenhofer R (2016) Bitcoin Meets Strong Consistency. *Proceedings of the 17th International Conference on Distributed Computing and Networking Conference (ICDCN).*

Eyal I, Gencer AE, Sirer EG, van Renesse R (2016) Bitcoin-NG: A Scalable Blockchain Protocol. *13th USENIX Symposium on Networked Systems Design and Implementation, (NSDI).*

Fudenberg D, Tirole J (1991) Perfect bayesian equilibrium and sequential equilibrium. *Journal of Economic Theory* 53(2):236 – 260, ISSN 0022-0531.

Gilad Y, Hemo R, Micali S, Vlachos G, Zeldovich N (2017) Algorand: Scaling byzantine agreements for cryptocurrencies. *Proceedings of the 26th Symposium on Operating Systems Principles, Shanghai, China, October 28-31, 2017*, 51–68.

Groce A, Katz J, Thiruvengadam A, Zikas V (2012) Byzantine agreement with a rational adversary. *Automata, Languages, and Programming - 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part II*, 561–572.

Halpern JY, Vilaça X (2016) Rational consensus: Extended abstract. *Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing, PODC 2016, Chicago, IL, USA, July 25-28, 2016*, 137–146.

Halpern JY, Vilaça X (2020) Rational consensus. *CoRR* abs/2005.10141.

Kiayias A, Koutsoupias E, Kyropoulou M, Tselekounis Y (2016) Blockchain mining games. *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, 365–382.

Kokoris-Kogias E, Jovanovic P, Gailly N, Khoffi I, Gasser L, Ford B (2016) Enhancing Bitcoin Security and Performance with Strong Consistency via Collective Signing. *Proceedings of the 25th USENIX Security Symposium.*

Kroll JA, Davey IC, Felten EW (2013) The economics of bitcoin mining, or bitcoin in the presence of adversaries. *Proceedings of WEIS*, volume 2013, 11.

Lamport L, Shostak R, Pease M (1982) The byzantine generals problem. *ACM Transactions on Programming Languages and Systems* 4(3):382–401, ISSN 0164-0925.

Lysyanskaya A, Triandopoulos N (2006) Rationality and adversarial behavior in multi-party computation. *Advances in Cryptology - CRYPTO 2006, 26th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2006, Proceedings*, 180–197.

Manshaei MH, Jadliwala M, Maiti A, Fooladgar M (2018) A game-theoretic analysis of shard-based permissionless blockchains. *IEEE Access* 6:78100–78112.

Miller A, Xia Y, Croman K, Shi E, Song D (2016) The honey badger of BFT protocols. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, 31–42.

Nakamoto S (2008) Bitcoin: A Peer-to-Peer Electronic Cash System. `https://bitcoin.org/bitcoin.pdf`.

Nash J (1951) Non-cooperative games. *Annals of Mathematics* 54(2):286–295, ISSN 0003486X.

Schelling TC (1960) *The strategy of conflict* (Cambridge, Harvard University Press).

Stoll C, Klaaßen L, Gallersdörfer U (2019) The carbon footprint of bitcoin. *Joule* 3(7):1647–1661, ISSN 2542-4351.

Yin M, Malkhi D, Reiter MK, Golan-Gueta G, Abraham I (2019) Hotstuff: BFT consensus with linearity and responsiveness. *Proceedings of the 2019 ACM Symposium on Principles of Distributed Computing, PODC 2019, Toronto, ON, Canada, July 29 - August 2, 2019.*, 347–356.