Order-Fairness for Byzantine Consensus

Mahimna Kelkar[1,2,3], Fan Zhang[1,2,3], Steven Goldfeder[1,2,3], and Ari Juels[1,2,3]

[1] Cornell Tech, New York, USA
{mahimna,fanz}@cs.cornell.edu
{goldfeder,juels}@cornell.edu
[2] Cornell University, Ithaca, USA
[3] The Initiative for CryptoCurrencies & Contracts (IC3), New York, USA

Abstract. Decades of research in both cryptography and distributed systems has extensively studied the problem of state machine replication, also known as Byzantine consensus. A consensus protocol must satisfy two properties: *consistency* and *liveness*. These properties ensure that honest participating nodes agree on the same log and dictate when fresh transactions get added. They fail, however, to ensure against adversarial manipulation of the actual *ordering* of transactions in the log. Indeed, in leader-based protocols (almost all protocols used today), malicious leaders can directly choose the final transaction ordering.

To rectify this problem, we propose a third consensus property: *transaction order-fairness*. We initiate the first formal investigation of order-fairness and explain its fundamental importance. We provide several natural definitions for order-fairness and analyze the assumptions necessary to realize them.

We also propose a new class of consensus protocols called Aequitas[1]. Aequitas protocols are the first to achieve order-fairness in addition to consistency and liveness. They can be realized in a black-box way using existing broadcast and agreement primitives (or indeed using any consensus protocol), and work in both synchronous and asynchronous network models.

1 Introduction

The abstraction of state machine replication has been investigated in cryptography and distributed systems literature for the past three decades. At a high level, the goal of a state machine replication protocol is for a set of nodes to agree on an ever-growing, linearly ordered log of messages (transactions). Two properties need to be satisfied by such a protocol: (1) *Consistency* - all honest nodes must have the same view of the agreed upon log — that is, they must output messages in the same order; and (2) *Liveness* - messages submitted by

The full version of this paper is available at https://eprint.iacr.org/2020/269 [27].

[1] Aequitas (IPA pronunciation: /ˈae̯.kʷi.taːs/) is the Roman personification of fairness.

clients are added to the log within a reasonable amount of time. In this paper, we will use the terms *state machine replication* and *consensus*[2] interchangeably.

Unfortunately, neither consistency nor liveness says anything about the actual ordering of transactions in the final log. A protocol that ensures that all nodes agree on the same ordering is deemed consistent regardless of how the ordering is generated. This leaves room for the definition to be satisfied even if an adversary directly chooses the actual transaction ordering, which is discomforting considering that the ordering is often easy to manipulate [7]. Moreover, in all existing protocols that rely on a designated "leader" node (e.g., [15, 34, 44]), which includes most used in practice, an adversarial leader may choose to propose transactions in any order.

In this paper, we formulate a new property for byzantine consensus which we call *order-fairness*. Intuitively, order-fairness denotes the notion that if a large number of nodes receive a transaction tx_1 before another one tx_2, then this should somehow be reflected in the final ordering.

Importance of fair transaction ordering. The need for a notion of fair transaction ordering is immediately clear when looking at financial systems. Here, the execution order can determine the validity and/or profitability of a given transaction. Suppose Bob has $0, and two transactions are initiated: tx_0, which sends $5 from Alice to Bob, and tx_1, which sends $5 from Bob to Carol. If tx_0 is sequenced before tx_1, then both transactions are valid; the opposite ordering invalidates tx_1. Manipulation of transaction ordering is a well known phenomenon on Wall Street [32], but recent work has shown it to also be commonplace in consensus-based systems such as permissionless blockchains. A recent paper by Daian et al. [20], for example, reports rampant adversarial manipulation of transactions in the Ethereum network [23] by bots extracting upwards of USD 6M in revenue from unsophisticated users.

Comparison to validity in Byzantine agreement. Beyond its critical practical importance, we believe that order-fairness is a key missing theoretical concept in existing consensus literature. To underscore this point, consider Byzantine agreement [30], or *single-shot* agreement, another well-studied problem in consensus literature. For Byzantine agreement, each node starts with a single value within a set $\mathcal{V}$. The goal is for all nodes to agree on the same value. *Validity* requires that if all honest nodes start with the same value v, then the agreed upon value should also be v.

The property of *order-fairness* is a natural analog of validity formulated for the consensus problem, i.e., extension of Byzantine agreement to multiple rounds. If all honest nodes start with the belief that a transaction tx_1 precedes another transaction tx_2, by natural analogy with validity, the final output log should sequence tx_1 before tx_2. Consequently, we maintain that *order-fairness* is a natural property of independent theoretical interest in the consensus literature.

[2] The term "consensus" has been used in systems literature for a number of related primitives, including "single-shot" consensus. However, in this paper, we use "consensus" to refer to the problem of "state machine replication."

1.1 Our Contributions

The main contributions of our paper are three-fold: (1) First, we investigate a natural notion of fair transaction ordering and show why it is impossible to realize. (2) Second, we investigate slightly weaker notions of fair ordering that are intuitive yet achievable. Still, we find that no existing consensus protocol achieves them. (3) Third, we introduce a new class of consensus protocols that we refer to as Aequitas. Aequitas protocols achieve fair transaction ordering while also providing the usual consistency and liveness. We discuss Aequitas protocols in both synchronous and asynchronous settings.

Defining order-fairness and impossibility results. To model our consensus protocols, we use an approach similar to prior work by Pass et al. [39, 40], wherein protocol nodes *receive* transactions from clients and need to *output* or *deliver* them in a way that satisfies consistency and liveness. We detail our model in Section 2. Within this model, we provide the first formalization of the property of order-fairness (Section 4). We start with a natural definition based on when transactions are received by nodes.

Definition 1 (Receive-Order-Fairness, informal; formalized in Definition 9). *If sufficiently many (at least γ-fraction) nodes receive a transaction* tx *before another transaction* tx$'$*, then all honest nodes must output* tx *before* tx$'$*.*

While Definition 1 is intuitive, it turns out that it is impossible to achieve unless we assume very strong synchrony properties and/or a non-corrupting adversary. This result draws from a surprising connection with voter preferences in social choice theory. To highlight this using a simple example, consider three nodes, A, B, and C, that each receive 3 transactions, x, y, and z. A receives them in the order $[x, y, z]$, B in the order $[y, z, x]$ and C in the order $[z, x, y]$. Notice that a majority of nodes have received (x before y), (y before z) and (z before x)! This scenario, often called the Condorcet paradox [18], can cause a non-transitive global ordering *even when* all local orderings are transitive. This is problematic for the notion of receive-order-fairness. Theorem 1 gives an informal description of our impossibility result.

Theorem 1 (Impossibility of receive-order-fairness, informal; formalized in Theorem 2). *Consider a system with n nodes where the external network (between users and protocol nodes) is either asynchronous or the maximum delay δ is at least n rounds. Then, no protocol can achieve all of consistency, liveness, and receive-order-fairness.*

Given this impossibility result, we consider a natural relaxation of receive-order-fairness that we call *block-receive-order-fairness*, or simply *block-order-fairness.* To see the primary difference between the two definitions, we look at two transactions, tx and tx$'$, where sufficiently many nodes have received tx before tx$'$. While receive-order-fairness requires that tx be output "*before*" tx$'$, block-order-fairness relaxes this to "*before or at the same time as.*" We refer to transactions delivered at the same time as being in the same "block."

Definition 2 (Block-Order-Fairness, informal; formalized in Definition 11). *If sufficiently many nodes (at least γ-fraction) receive a transaction* tx *before another transaction* tx′*, then no honest node can deliver* tx *in a block after* tx′*.*

This small relaxation allows us to evade the Condorcet paradox by a simple trick: placing paradoxical orderings into the same "block." We emphasize that block order-fairness does not mean that transactions are partially ordered. Consistency still requires that all nodes output transactions in the same order (within the same block or not). The only difference is that unfair ordering of a set of transactions in our definition without blocks is now, with the use of blocks, considered fair, provided that these transactions appear in the same block.

Further, we note that while receive-order-fairness is impossible to achieve (as pointed out informally in Theorem 1 and formalized later in the paper in Theorem 2), block-order-fairness is not and we provide protocols that guarantee it. We would also like to highlight that our proposed Aequitas protocols actually make minimal use of this relaxation. In particular, they achieve the stronger notion of receive-order-fairness except when non-transitive preferences are observed.

Aequitas: Achieving order-fairness. We present a new class of consensus protocols, Aequitas, that achieve block-order-fairness, in addition to providing consistency and liveness. Aequitas protocols make use of two basic primitives in a black-box way: (1) FIFO Broadcast (FIFO-BC) [26], which is a basic extension of standard reliable broadcast; and (2) Set Byzantine Agreement (Set-BA; defined in Section 3), which can be achieved from Byzantine agreement.

We note that these are weak primitives and any standard consensus protocol (that achieves consistency and liveness) can also be used to build the FIFO-BC and Set-BA primitives. This results in an interesting observation: The Aequitas technique provides a generic compiler that takes any standard consensus protocol and converts it into one that also provides order-fairness. At a high level, Aequitas protocols proceed in three major stages. Each transaction tx goes through these stages before being delivered.

1. **Gossip Stage.** Nodes gossip transactions in the order that they are received. That is, each node gossips its *local* transaction ordering.

   For this purpose, we use the FIFO broadcast primitive (FIFO-BC), which guarantees that broadcasts by an honest node are delivered by other honest nodes in the same order that they were broadcast. Even if the sender is dishonest, FIFO-BC guarantees that all honest nodes deliver messages in the same order. As a result, nodes have a consistent view of the transaction orderings of other nodes.

   We use Log_i^j to denote node i's view of the order in which node j received transactions, according to how j gossiped them. Note that if node j is malicious, Log_i^j may arbitrarily differ from the actual order in which j *received* transactions, but FIFO-BC prevents j from equivocating, i.e., any two honest nodes i and k will have consistent Log_i^j and Log_k^j. When i records enough logs Log_i^k that contain tx, we say that the "gossip phase" for tx is complete.

2. **Agreement Stage.** Nodes agree on the set of nodes whose local orderings should be considered for deciding on the global ordering of a particular transaction.

   To elaborate, at the end of the gossip stage for a transaction tx, a node i ends up with a set U_i^{tx} of other nodes whose local orderings i has obtained. That is, $k \in U_i^{\text{tx}}$ if $\text{tx} \in \mathsf{Log}_i^k$. Note that different nodes may end up with a slightly different set U, but agreement proceeds when enough honest nodes are present in each set. Nodes perform Byzantine agreement to agree on a set L^{tx} of nodes whose ordering will be used to finalize the ordering for tx. For this, we define a new primitive Set-BA whose validity condition guarantees that if $k \in U_i^{\text{tx}}$ for all i, then $k \in L^{\text{tx}}$. It is easy to see how Set-BA can be realized by using standard Byzantine agreement to determine the inclusion of each possible value k individually.

3. **Finalization Stage.** Nodes finalize the global ordering of a transaction tx using the set of local orderings decided on in the agreement stage.

   Suppose that the agreement stage for a transaction tx resulted in the set L^{tx}. Now, if there is any other transaction tx′ such that tx′ is ordered before tx in a large number of these local logs, it signifies that tx should be delivered after tx′. In other words, the finalization of tx depends on waiting until tx′ has been delivered.

   To characterize such ordering dependencies between transactions, a node i maintains a directed graph G_i, where vertices represent transactions and an edge from a to b denotes that b is waiting for a. Since nodes are building this graph on the same "data" (the set of local logs agreed upon in the agreement phase), nodes will have consistent graphs. That is, if an edge (a, b) exists in G_i, then it will also (eventually) exist in G_j, if i and j are both honest.

   We present two finalization techniques, a leader-based one and a leaderless one. For the leader-based technique, resolving any partial ordering within the graph is delegated to a leader node. We emphasize that order-fairness is not lost. The leader is only able to choose the ordering for transactions that are not required to be ordered a certain way. We present another, leaderless technique that requires no further communication between nodes. We find that both realize a slightly weaker notion of liveness than the standard one, even in the synchronous setting. Specifically, future transactions are required to be input to the system in order to "flush out" earlier transactions. We formally define "weak-liveness" in Section 2.

It is worth pointing out that the first two stages (gossip and agreement) are fairly straightforward to understand and easy to achieve. The third stage is somewhat complex, as it needs to avoid the Condorcet paradox while continuing to maintain both consistency and order-fairness.

Aequitas protocols. In summary, we present the first consensus protocols that provide order-fairness. We provide a leader-based and a leaderless protocol each for the synchronous and asynchronous settings, for a total of four protocols that follow the same general outline. These protocols all provide consistency, block order-fairness, and some form of liveness. Fig. 1 shows a comparison.

Protocol	Style	Network	Corruption Bound†	Consistency	Liveness	Order-Fairness
$\Pi^{\mathsf{sync,lead}}_{\mathsf{Aequitas}}$	Leader	Synchronous*	$n > \frac{2f}{2\gamma-1}$	✓	✓ (Weak)	✓
$\Pi^{\mathsf{sync,nolead}}_{\mathsf{Aequitas}}$	Leaderless	Synchronous*	$n > \frac{2f}{2\gamma-1}$	✓	✓ (Weak)	✓
$\Pi^{\mathsf{async,lead}}_{\mathsf{Aequitas}}$	Leader	Any	$n > \frac{4f}{2\gamma-1}$	✓	✓ (Eventual, Weak)	✓
$\Pi^{\mathsf{async,nolead}}_{\mathsf{Aequitas}}$	Leaderless	Any	$n > \frac{4f}{2\gamma-1}$	✓	✓ (Eventual, Weak)	✓

* Completely Synchronous Setting (See Section 2)
† $\frac{1}{2} < \gamma \leq 1$ is the order-fairness parameter (See Section 4)

Fig. 1. The Aequitas protocols

Paper organization. The rest of the paper is organized as follows. We discuss our results in the context of related work in Section 1.2. We describe our formal framework, along with other preliminaries, in Section 2. In Section 3, we provide the building blocks for our protocol constructions. Section 4 formally introduces our notion of order-fairness. Section 5 provides a general overview of our constructions; we detail our leaderless construction for the synchronous setting in Section 6. Due to space constraints, we defer other constructions and results, as well as several proofs to the full version [27] of our paper.

1.2 Related Work

While there is an extensive literature on consensus protocols, to the best of our knowledge, no previous work formally captures a notion of order-fairness like the one we introduce. The term "fairness" has been used widely in blockchain and cryptography literature, but for properties unrelated to ours.

Broadcast primitives. Byzantine broadcast, or the Byzantine Generals Problem [30], is the elementary broadcast primitive where a designated sender broadcasts a single value to a set of receiving nodes. In a Byzantine broadcast protocol with the key property of *consistency*, all honest receivers output the same value. Reliable broadcast is a continuous version of Byzantine broadcast where the sender broadcasts multiple values which must be eventually delivered by nodes if the sender is honest. Three orthogonal properties can be added onto reliable broadcast to give stronger notions. FIFO-ordering provides first-in first-out ordering on the messages broadcast by an honest sender. We refer to such a protocol as FIFO Broadcast or OARcast [26]. Local-ordering (also called causal-ordering) ensures that if a node broadcasts a message m' after receiving some other message m, then m will be ordered before m'. The total-ordering property ensures that all honest nodes deliver messages broadcast potentially by different senders in the same order. This notion is usually called *atomic broadcast* [19], which is well-known to be equivalent to the consensus problem. Adding all three properties to reliable broadcast results in the notion of Causal FIFO Atomic Broadcast which still does not provide the *order-fairness* property that we are looking for. The main problem is none of the requirements consider a global notion of FIFO ordering based on multiple senders.

Our order-fairness property does enforce such a notion according to the following idea: If enough nodes broadcast a message m before another message m', then honest nodes will respect this ordering. Adding this property to atomic broadcast results in a *new broadcast* notion, which we call "Global FIFO Atomic Broadcast." Consequently, requiring order-fairness along with standard consensus properties of consistency and liveness will be equivalent to this new notion of Global FIFO Atomic Broadcast.

We note that our setup is also slightly different than earlier notions. We assume that any message broadcast by an honest node is also eventually broadcast by all honest nodes. This allows us to redefine liveness in terms of being broadcast by enough nodes. This also means that identical messages broadcast by different nodes can now be delivered together as a single message. Global FIFO ordering is defined on the ordering of these messages. Note that it no longer makes sense to talk about (single source) FIFO order or causal order as identical messages, potentially broadcast at different positions by different nodes, are now delivered as a single message.

Consensus protocols. Hundreds of Byzantine fault tolerant consensus protocols have been proposed over the years, with PBFT [15] being perhaps the most well known. Multiple survey papers [7, 10] have aimed to systematize this vast literature. Many papers provide efficiency improvements while maintaining the basic leader-based structure of PBFT. That is, a *leader* or *primary* node is responsible for proposing the transactions in the current round. In such leader-based protocols ([2, 3, 5, 8, 17, 34, 42–44], just to name a few), the leader node can propose transactions in the order of its choosing. The leader is also capable of suppressing transactions, at least temporarily, until an honest node becomes the new leader. We highlight that in previously explored leader-based protocols, nodes do not know the ordering in which transactions were received by everyone. This means that a leader's proposal can only be rejected based on validity of transactions rather than the fairness of their ordering. Order-fairness is thus not achieved in existing leader-based protocols.

Some protocols provide *transaction censorship resistance*, such that malicious nodes cannot censor specific transactions based on their content. For this, in protocols like [4, 11, 36], transactions are encrypted, and the contents are revealed only once their ordering is fixed. Separately, protocols like [4, 29, 31] rely on a reputation based system to detect unfair censorship. Censorship resistance is strictly weaker than the order-fairness we consider for three reasons. First, in practice, even if transaction data is temporarily encrypted, metadata such as a user identifier or a client IP address can be used to censor a particular transaction. Second, a malicious leader can still *blindly* reorder or censor transactions based on just their ciphertext. But perhaps more importantly, a malicious leader colluding with a user will know the ciphertext corresponding to the user's transaction and can thus unfairly order this transaction before others.

Other uses of the word *fairness*. The term *fairness* has been used before in consensus literature for notions unrelated to ours. One popular use case relates to *fairness in block mining* in Proof-of-Work (PoW) blockchains, which intuitively requires that a node's mining rewards be proportional to its relative computational power. That is, no node should be able to mine *selfishly* [24] to obtain more rewards than its fair share. This fairness notion is met by protocols in [1, 31, 33, 35, 37], among others.

Another related definition considers fairness in terms of the opportunities each node gets to append transactions to the ledger. This includes both fair leader election (in leader based protocols) and fair committee election (in hybrid consensus protocols). This definition is considered in [1, 25, 28, 31, 38]. We note that even if the leader election process is fair, the current leader still has the power to manipulate transaction ordering.

Fairness has also been used in the context of "fair exchange," which provides a way for mutually distrusting parties to exchange digital goods in a secure way. This notion is unrelated to ours but we mention it for completeness.

Works that mention fair transaction ordering. Helix [4] alludes to fair transaction ordering, but only considers censorship resistance and fair committee election. It uses threshold encryption to choose a random set of pending transactions for inclusion in the current block. Hashgraph [6] considers our notion of receive-order fairness, but provides no formal definitions. Moreover, it *fails to realize the impossibility* of this notion of fairness resulting from the Condorcet paradox [18]. As a result, we identify an elementary attack on the Hashgraph protocol that allows an adversarial node to control transaction ordering. The main problem in Hashgraph is the use of timestamp based ordering. In Section 5, we provide a brief explanation for why this does not work and defer the description of our attack to the full version [27].

2 Definitions, Framework, and Preliminaries

In this section, we describe the general execution framework that we will use for expressing and analyzing consensus protocols. We adopt an approach like that of Pass and Shi [39, 40] and Chan et al. [16]. We focus on the "permissioned" setting, where the number of consensus nodes n, as well as their identities, are known a priori to all participants. While arbitrary clients can send messages to these nodes, only a fixed set of nodes will take part in the consensus protocol. We are interested in protocols for several network settings (e.g. synchronous, partially synchronous, and asynchronous) and define constrained environments for these settings by imposing restrictions that an adversary must respect. Due to space constraints, we only include the relevant formalism for the constructions in this paper. For the complete details of the model, we refer the reader to the full version [27].

2.1 Protocol Execution Model

Interactive Turing Machines (ITMs). We adopt the widely used Interactive Turing Machine (ITM) approach rooted in the Universal Composability framework [12]. Informally, a protocol details how nodes interact with each other, where each node is represented by an ITM. As standard practice in cryptography literature [12–14], we use an environment $\mathcal{Z}(1^\kappa)$ (where κ is the security parameter) to direct the protocol execution. $\mathcal{Z}$ is responsible for activating nodes as either *honest* or *corrupt*, providing messages as inputs to nodes, and delivering messages between nodes. Honest nodes follow the protocol description while corrupt nodes are assumed to be controlled by an adversary, denoted by $\mathcal{A}$. $\mathcal{A}$ is able to read all inputs/messages sent to corrupt nodes and can set all outputs/messages to be sent. The adversary also decides when messages sent over the network get delivered, subject to any network assumptions.

Rounds. We assume that $\mathcal{Z}$ maintains a global clock. The clock is a global functionality [14] that contains a simple monotonic counter which can be updated adversarially by the environment. In the synchronous setting, we can model protocol execution in discrete time steps or rounds. At the start of each round, each node receives a set txs of transactions from the environment $\mathcal{Z}$. Transactions are assumed to be submitted by clients, but using the environment abstraction avoids having to model clients explicitly. At the end of each round, each node outputs an ordered log LOG to $\mathcal{Z}$ which intuitively represents the list of transactions ordered by the node so far. We assume that $\mathcal{Z}$ always signals the start of a new round to each node. Rounds in the partially synchronous setting work similarly to the synchronous setting. In the asynchronous setting, the clock is not accessible to the protocol nodes. $\mathcal{Z}$ can provide user transactions and communication messages to nodes at any time. Any protocol that works in the asynchronous setting should not rely on the current time. Throughout the paper, we may use the terms "time" and "round" interchangeably.

Notational Conventions. We use κ to denote the security parameter. $\mathcal{N}$ denotes the set of protocol nodes. For a protocol Π, $\mathsf{EXEC}^\Pi(\mathcal{A}, \mathcal{Z}, \kappa)$ represents the random variable for all possible execution traces of Π w.r.t. adversary $\mathcal{A}$ and environment $\mathcal{Z}$. We use $\mathsf{view} \leftarrow\!\!\$\; \mathsf{EXEC}^\Pi(\mathcal{A}, \mathcal{Z}, \kappa)$ to denote randomly sampling an execution. $|\mathsf{view}|$ denotes the number of rounds in view.

Corruption Model. Since we are concerned only with the permissioned setting, we consider environments $\mathcal{Z}$ that spawn a set of nodes, numbered from 1 to n at the start, and never spawn additional nodes. At any point, $\mathcal{A}$ can ask $\mathcal{Z}$ to corrupt a particular node for which $\mathcal{Z}$ sends a `corrupt` signal to that node. When this happens, the internal state of the node gets exposed to $\mathcal{A}$ and $\mathcal{A}$ henceforth fully controls the node. A node is said to be *honest* in a given view if it is never under adversarial control. Otherwise, it is said to be *corrupt* or *Byzantine*. $\mathcal{A}$ can corrupt nodes at any point during the protocol's execution; but once a node is corrupted, it cannot become honest at a later point. The corruption parameter f denotes the maximum number of nodes that $\mathcal{A}$ can corrupt.

Communication and Network Model. As mentioned before, $\mathcal{Z}$ provides transactions sent by users as inputs to nodes and also handles communication between nodes. We assume that a node can broadcast messages to others through authenticated channels. Furthermore, we assume that the adversary $\mathcal{A}$ cannot modify messages sent by honest nodes but can reorder or delay messages, possibly constrained by the specific setting.

We differentiate between two networks in our model — an *internal* network for communication between nodes and an *external* network for how external users send transactions to nodes. We emphasize that $\mathcal{A}$ is only in charge of scheduling message delivery for the internal network. The external network may reside in other parts of the application (not relevant to the consensus protocol) and is managed by $\mathcal{Z}$ (and possibly by some other network adversary). For both networks, we consider the synchronous setting [21] (where the network delay bound is known), the partially synchronous setting [22] (where the network delay bound is finite but unknown), and the asynchronous setting [9] (where the network delay is unbounded).

2.2 Execution Environments

Clients submit transactions by sending them to all nodes. As mentioned before, we do not explicitly model clients, but rather have transactions input by $\mathcal{Z}$.

External Network. The external network models the channel between the system clients and the protocol nodes. By a synchronous external network, we mean that any transaction that is received from $\mathcal{Z}$ by a node reaches all other nodes within a known time. This is formally defined in Definition 3.

Definition 3 (External Synchronous Setting). *We say that* $(\mathcal{A}, \mathcal{Z})$ *respects* $\Delta_{\mathsf{ext}} = (\mathsf{full}, \delta)$ ext-*synchrony w.r.t. protocol* Π *if for every* $\kappa \in \mathbb{N}$ *and* view *in the support of* $\mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$*, the following conditions hold: (1)* $\mathcal{Z}$ *provides* δ *to all nodes upon spawning; (2) If* $\mathcal{Z}$ *provides an input message* m *to a node in the* txs *set at time* t*, then at any time* $t' \geq t + \delta$*, all other nodes will also have received message* m *as input.*

For the partially synchronous setting, we assume that δ exists but is unknown to the nodes, and not provided by $\mathcal{Z}$. For the asynchronous setting, we only assume that transactions are not dropped by the network — they eventually get delivered to all the nodes.

Internal Network. The internal network represents the network between nodes and is usually the standard network considered for consensus problems. We formalize the internal synchrony assumption in Definition 4. The partially synchronous and asynchronous settings are defined similarly to the corresponding notions for the external network.

Definition 4 (Internal Synchronous Setting). *We say that* $(\mathcal{A}, \mathcal{Z})$ *respects* $\Delta_{\mathsf{int}} = (\mathsf{full}, \delta)$ int-*synchrony w.r.t. protocol* Π *if for every* $\kappa \in \mathbb{N}$ *and* view *in the* *support of* $\mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$*, the following conditions hold: (1)* $\mathcal{Z}$ *provides* δ *to all nodes upon spawning; (2) If an honest node sends a message at time* t*, then at any time* $t' \geq t + \delta$*, all recipient(s) will have received the message.*

Network nomenclature. We say that the network is *completely synchronous* (resp. *completely asynchronous*) if both the external and the internal network are synchronous (resp. asynchronous). We use not-async to denote both the synchronous setting and the partially synchronous setting.

Permissioned Setting. For the "permissioned" or "classical" environment, we require that $\mathcal{Z}$ spawn all nodes up front and not spawn any new nodes during the protocol execution. Furthermore, all nodes know the identity of all other nodes in the protocol. We define the permissioned environment in Definition 5.

Definition 5 (Classical Permissioned Environment). *We say that* $(\mathcal{A}, \mathcal{Z})$ *respects* $(n, f, \Delta_{\mathsf{int}}, \Delta_{\mathsf{ext}})$*-classical execution w.r.t. a protocol* Π *if it respects* Δ_{int} int*-synchrony,* Δ_{ext} ext*-synchrony and for every* $\kappa \in \mathbb{N}$ *and* view *in the support of* $\mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$*, the following conditions hold: (1)* $\mathcal{Z}$ *spawns a set of nodes numbered from* 1 *to* n *at the start of the protocol and never spawns any nodes later; (2)* $\mathcal{Z}$ *does not corrupt more than* f *nodes; (3)* $\mathcal{Z}$ *provides all nodes* n, f *as well as any other public parameters upon spawning.*

For all constraints on $(\mathcal{A}, \mathcal{Z})$, when the context is clear, we may choose to exclude the protocol we are referring to.

2.3 The State Machine Replication Abstraction

In the state machine replication or consensus problem, a set of nodes try to agree on a growing, linearly ordered log. At the start of each round, $\mathcal{Z}$ provides a set txs (possibly empty) of transactions to protocol nodes. We assume that the transactions input by $\mathcal{Z}$ are unique. At any time, nodes may also choose to deliver transactions by outputting a LOG to $\mathcal{Z}$. The LOG can be thought of as a totally ordered sequence where each element is an ordered set of transactions. We refer to the set of transactions at an index of the LOG as a "block." The LOG represents the set of transactions committed by a node so far.

Transaction nomenclature. When discussing the trajectory of a transaction, we say that a transaction tx is *received* by a node when it is given as input to the node by $\mathcal{Z}$. A transaction tx is *delivered* or *output* by a node when it is included in a LOG output by the node to $\mathcal{Z}$.

Notation for the ordered log. $\mathcal{T}$ denotes the space of all possible transactions. Let LOG_i represent the most recent log output by node i to the environment, i.e., the ordered list of transactions that node i has delivered so far. For two logs LOG and LOG′, we define a relation $\preceq$ which intuitively signifies a "prefix" notion. $\mathsf{LOG} \preceq \mathsf{LOG}'$ stands for "LOG is a prefix of LOG′." We assume that for any x, we have $x \preceq x$ and $\emptyset \preceq x$. $\mathsf{LOG}[p]$ denotes the p^{th} element in LOG. $\mathsf{LOG}(m)$ denotes the number p such that $\mathsf{LOG}[p]$ contains m.

The security of a state machine replication protocol is now defined as follows:

Definition 6 (Security of state machine replication [40]). *We say that a protocol Π satisfies consistency (resp. $(T_{\text{warmup}}, T_{\text{confirm}})$-liveness) w.r.t. $(\mathcal{A}, \mathcal{Z})$ if there exists a negligible function $\mathsf{negl}(\cdot)$ such that for any $\kappa \in \mathbb{N}$, consistency (resp. $(T_{\text{warmup}}, T_{\text{confirm}})$-liveness) is satisfied except with $\mathsf{negl}(\kappa)$ probability over the choice of $\mathsf{view} \leftarrow_{\$} \mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$ where negl is negligible in κ. For a particular view, we define the properties as below:*

- **(Consistency)** *A view satisfies consistency if the following holds:*
  - *Common Prefix. If an honest node i outputs LOG to $\mathcal{Z}$ at time t and an honest node j outputs LOG' to $\mathcal{Z}$ at time t', then it holds that either $\mathsf{LOG} \preceq \mathsf{LOG}'$ or $\mathsf{LOG}' \preceq \mathsf{LOG}$.*
  - *Future Self Consistency. If a node that is honest between times t and t', outputs LOG at time t and LOG' at time $t' \geq t$ to the environment $\mathcal{Z}$, then it holds that $\mathsf{LOG} \preceq \mathsf{LOG}'$.*
- **(Liveness)** *A view satisfies $(T_{\text{warmup}}, T_{\text{confirm}})$-liveness if the following holds: At a time t such that $T_{\text{warmup}} < t \leq |\mathsf{view}|$, if an honest node either received a transaction m from $\mathcal{Z}$ or output m in its log to $\mathcal{Z}$, then for any honest node i and any time $t' \geq t + T_{\text{confirm}}; t' \leq |\mathsf{view}|$, it holds that m is in the log output by node i at time t'.*

Here, T_{confirm} and T_{warmup} are polynomial functions in κ, n, f, any maximum network delay bounds as defined in Δ_{ext} and Δ_{int}, as well as the actual network delay. T_{warmup} is the protocol's warmup time, until which point liveness need not be satisfied. T_{confirm} is the maximum time it takes for a transaction (input after the warmup time) to be delivered by all honest nodes.

Weak liveness. The standard definition of liveness of a transaction tx (from Definition 6) is independent of what happens in the rest of the protocol's execution. Sometimes however, it may be enough for a protocol to be live only if transactions continue to be received by the system. For example, a transaction tx will only be delivered if there is some transaction that is received by all nodes sufficiently after tx. Intuitively, later transactions will cause earlier ones to be "flushed out" of the system. We note that this subtle distinction between the two liveness definitions is rarely considered in the literature. We found that some leaderless protocols (i.e. those that are not based on a leader node) like [6, 41] implicitly ignore this distinction. Along similar lines, we define a weaker version of conventional liveness, which we call "weak-liveness." Despite the technical difference, we think that it should be acceptable in most real world systems. For a particular view, we define weak-liveness below.

- **(Weak Liveness)** A view satisfies $(T_{\text{warmup}}, T_{\text{confirm}})$-weak-liveness if the following holds: Suppose that at a time t such that $t > T_{\text{warmup}}$, an honest node either received a transaction m from $\mathcal{Z}$ or output m in its log to $\mathcal{Z}$. Let T be a set built recursively as follows: (1) Add m to T; (2) For $m_0 \in \mathsf{T}$, add to T, all transactions m_0' that were received by at least one honest node before m_0. Now if another transaction m' was received at time t' and is such that it was first received by a node after all nodes received all transactions in T, then for any honest node i and any time $t'' \geq t' + T_{\text{confirm}}; t'' \leq |\mathsf{view}|$, it holds that m is in the log output by node i at time t''.

3 Building Blocks

We start by describing some useful primitives that will form the foundation for designing our fair ordering consensus protocols. More specifically, we will utilize two primitives: (1) Set Byzantine Agreement (Set-BA); and (2) FIFO Broadcast (FIFO-BC). We show how to build Set-BA from Byzantine agreement and FIFO-BC from reliable broadcast in the full version [27].

Subroutines and composition. We follow standard conventions to enable secure composition. Each instance of a protocol is spawned with a session identifier sid. We use $\Pi[\mathsf{sid}]$ to denote the instance of protocol Π with session id sid. Each protocol may take inputs from and return outputs to an environment. Note that this "environment" may be different for any subroutines called.

3.1 Set Byzantine Agreement

Definitions. In a (poly) Set Byzantine Agreement protocol (Set-BA), participating nodes will try to agree on a set of values. At the start of the protocol, each node receives any public parameters from $\mathcal{Z}$. Each node i in the set $\mathcal{P}$ of participating nodes also receives a set $U_i \subseteq S$ as input from $\mathcal{Z}$. The set S is also known to all nodes and its description is polynomial in κ. At the end of the protocol, each honest node $j \in \mathcal{P}$ outputs a set of the agreed upon values O_j.

Definition 7 (Security of Set-BA). *A Set-BA protocol Π_{sba} satisfies agreement, inclusion validity, and exclusion validity w.r.t. $(\mathcal{A}, \mathcal{Z})$ if for all $\kappa \in \mathbb{N}$, the following properties hold except with negligible probability over* $\mathsf{view} \leftarrow_\$ \mathsf{EXEC}^{\Pi_{\text{sba}}}(\mathcal{A}, \mathcal{Z}, \kappa)$.

- **(Agreement)** *If honest nodes i and j output the sets O_i and O_j respectively, then $O_i = O_j$.*
- **(Inclusion Validity)** *If an element is in the input sets of all nodes, then it will also be in the output sets of all honest nodes.*
- **(Exclusion Validity)** *If an element is not in any input set, then it is not in any honest output set.*

For a given view, we also say that Π_{sba} satisfies $T^{\text{sba}}_{\text{confirm}}$-liveness, if all honest nodes output in at most $T^{\text{sba}}_{\text{confirm}}$ rounds after all honest nodes have input their starting value. Lemma 1 shows a helpful result that any outputs are "honestly proposed."

Lemma 1. *Consider any set Byzantine agreement protocol* Π_{sba} *that satisfies agreement, inclusion validity, and exclusion validity (w.r.t* $(\mathcal{A}, \mathcal{Z})$*). Except for a negligible number of views,* Π_{sba} *also satisfies the following:*

- **(Honest Proposal)** *If an honest node outputs the set* O*, then for every* $c \in O$*, there exists* $i \in \mathcal{P}$ *such that* i *is honest and* $c \in U_i$*.*

3.2 FIFO Broadcast

Single source FIFO (first in, first out) broadcast (also called Ordered Authenticated Reliable broadcast or OARcast in [26]) is a broadcast primitive in which all honest nodes in the protocol need to deliver messages in the same order as they were broadcast by the sender. In one instantiation of a FIFO broadcast protocol, we consider a single designated sender who broadcasts a sequence of messages to all other nodes. If the sender is honest, each honest node must deliver the messages in the same order as they were broadcast. If the sender is dishonest, all honest nodes must deliver messages in the same order as each other; except now, this order may may be different than the one broadcast by the sender. When composing several FIFO broadcast primitives together with different senders, FIFO order is maintained for each individual sender but different honest nodes may deliver messages from different senders in different orders.

Definitions. At the start of the FIFO Broadcast ($\mathsf{FIFO\text{-}BC}$) protocol, each node receives the appropriate public parameters from the environment. At any time, the designated sender may also receive as input a message m from the environment. At any time, nodes can choose to deliver messages.

Definition 8 (Security of ($\mathsf{FIFO\text{-}BC}$)). *A* $\mathsf{FIFO\text{-}BC}$ *protocol* Π_{fifocast} *satisfies liveness, agreement, and FIFO-order w.r.t.* $(\mathcal{A}, \mathcal{Z})$ *if for all* $\kappa \in \mathbb{N}$*, the following properties hold except with negligible probability over* $\mathsf{view} \leftarrow_{\$} \mathsf{EXEC}^{\Pi_{\text{fifocast}}}(\mathcal{A}, \mathcal{Z}, \kappa)$.

- **(**$(T^{\text{fifocast}}_{\text{warmup}}, T^{\text{fifocast}}_{\text{confirm}})$**-Liveness)** *If the sender is honest and receives a message* m *as input in round* $r > T^{\text{fifocast}}_{\text{warmup}}$*, or if an honest node delivers* m *in round* $r > T^{\text{fifocast}}_{\text{warmup}}$*, then all honest nodes will have delivered* m *by round* $r + T^{\text{fifocast}}_{\text{confirm}}$*.*
- **(Agreement)** *If an honest node delivers a message* m *before* m'*, then no honest node delivers* m' *unless it has already delivered* m*.*
- **(FIFO-Order)** *If the sender is honest and is input a message* m *before* m'*, then no honest node delivers* m' *unless it has already delivered* m*.*

$T^{\text{fifocast}}_{\text{confirm}}$ *is a polynomial in* κ, n, f *and the internal network delay.*

Notation. Let $\Pi_{\text{fifocast}}[(\mathsf{sid}, j)]$ denote the instance of the protocol Π_{fifocast} where node j is the designated sender. In a consensus protocol that invokes $\Pi_{\text{fifocast}}[(\mathsf{sid}, j)]$, we assume that each node i keeps track of the messages delivered (i.e. messages broadcast by node j) in a local log $\mathsf{Log}_i^{(\mathsf{sid}, j)}$. This represents node i's view of broadcasts from node j in the session sid. When the session id is clear from context, we may simply write Log_i^j. Two local logs Log and Log' are called "equal until tx", denoted by $\approx_{\text{tx}}$, if they are equivalent until the occurrence of tx. $\mathsf{Log}[p]$ denotes the p^{th} element in Log. $\mathsf{Log}(m)$ denotes the number p such that $\mathsf{Log}[p]$ contains m. Consequently, $\mathsf{Log}(m) < \mathsf{Log}(m')$ signifies that m appears before m' in Log.

4 Defining Fair Ordering

We formally define fair ordering in this section. As it turns out, providing a definition that is achievable by protocols, yet intuitive, is not trivial. Some natural definitions are not achievable except under strong assumptions. We use this section to also go through these definitions that led to our final definition.

(Attempt 1) – Send-order-fairness. A strawman approach is to require ordering to be in terms of when transactions were *sent* by clients. For instance, if a transaction tx_1 was sent by a client before another transaction tx_2 (possibly by another client), then tx_1 should appear before tx_2 in the agreed upon log. Not surprisingly, this can lead to several problems: most importantly, there needs to be a trusted way to timestamp a transaction at the client side. We discuss the possibility of achieving it in practice using trusted hardware in the full version [27].

(Attempt 2) – Receive-order-fairness. The challenges of send-order-fairness suggest it would be more prudent to define fair ordering in terms of when the consensus nodes actually *receive* transactions. Intuitively, "receive order" means that the fair ordering is defined by looking at when *enough* nodes receive a particular transaction. For instance, if sufficiently many nodes receive a transaction tx_1 before another transaction tx_2, then tx_1 must appear before tx_2 in the final log. "Sufficiently many" is parameterized using γ.

Definition 9 (Receive-order-fairness, restatement of Definition 1). *For a* view *in the support of* $\mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$*, define receive-order-fairness as follows:*

- *A* view *satisfies* $(\gamma, T_{\text{warmup}})$ *receive-order-fairness if the following holds: For any two transactions* m *and* m'*, let* η *be the number of nodes that received both transactions between times* T_{warmup} *and* $|\mathsf{view}|$*. If at least* $\gamma\eta$ *of those nodes received* m *before* m' *from* $\mathcal{Z}$*, then for all honest nodes* i*,* i *does not deliver* m' *unless it has previously delivered* m*.*

A protocol Π *satisfies* $(\gamma, T_{\text{warmup}})$ *receive-order-fairness w.r.t* $(\mathcal{A}, \mathcal{Z})$ *if there is a negligible function* $\mathsf{negl}(\cdot)$ *such that for any* $\kappa \in \mathbb{N}$*, the order-fairness property is satisfied except with probability* $\mathsf{negl}(\kappa)$ *over* $\mathsf{view} \leftarrow\$ \mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$*.*

4.1 Condorcet paradox and the impossibility of fair ordering.

The Condorcet paradox [18], or the "voting paradox", is a result in social choice theory that shows how some situations can lead to non-transitive collective voting preferences even if the preferences of individual voters are transitive. To illustrate how this applies to fair ordering, let us look at a simple example:

Example 1. Suppose that there are 3 nodes: A, B, and C. In the protocol execution, 3 transactions, $\mathrm{tx}_1, \mathrm{tx}_2$, and tx_3 are sent by clients to all the nodes.

- Node A receives transactions in the order $\mathrm{tx}_1, \mathrm{tx}_2, \mathrm{tx}_3$.
- Node B receives transactions in the order $\mathrm{tx}_2, \mathrm{tx}_3, \mathrm{tx}_1$.
- Node C receives transactions in the order $\mathrm{tx}_3, \mathrm{tx}_1, \mathrm{tx}_2$.

Now, 2 nodes (A and C) received tx_1 before tx_2, 2 nodes (A and B) received tx_2 before tx_3, and 2 nodes (B and C) received tx_3 before tx_1. It is easy to see that no protocol can satisfy fair ordering for $\gamma \leq \frac{2}{3}$, since such a protocol would have to include tx_1 before tx_2; tx_2 before tx_3; and tx_3 before tx_1 in its final log.

Theorem 2 generalizes this observation to show an impossibility for $\gamma \leq \frac{n-1}{n}$. Furthermore, it also shows that when $f \geq 1$, even $\gamma = 1$ receive-order-fairness is impossible to achieve.

Theorem 2 (Restatement of Theorem 1). *Consider any $n, f \geq 1, \Delta_{\mathsf{int}}, \Delta_{\mathsf{ext}}$ where Δ_{ext} is either asynchronous or $(\mathsf{not\text{-}async}, \delta_{\mathsf{ext}} \geq n)$. Let $\gamma \leq 1$. If a consensus protocol Π satisfies consistency and $(T_{\mathrm{warmup}}, T_{\mathrm{confirm}})$ liveness w.r.t. all $(\mathcal{A}, \mathcal{Z})$ that respect $(n, f, \Delta_{\mathsf{int}}, \Delta_{\mathsf{ext}})$-classical execution, then it cannot also satisfy $(\gamma, T_{\mathrm{warmup}})$ receive-order-fairness.*

Proof (Sketch). Taking inspiration from the counterexample in Example 1, we first show the result for $\gamma \leq \frac{n-1}{n}$. Denote the nodes in the system by the numbers 1 to n. Suppose that clients submit n transactions tx_1 to tx_n. Further, suppose that node 1 receives the transactions in the order $\mathrm{tx}_1, \mathrm{tx}_2, \cdots, \mathrm{tx}_n$ and any node $i \neq 1$ receives the transactions in the order $\mathrm{tx}_i, \cdots, \mathrm{tx}_n, \mathrm{tx}_1, \cdots, \mathrm{tx}_{i-1}$.

Now, it is straightforward to see that all nodes except node 2 received tx_1 before tx_2, all nodes except node 3 received tx_2 before tx_3 and so on. Finally, all nodes except node 1 received tx_n before tx_1. This means that any consensus protocol that provides order-fairness for $\gamma \leq \frac{n-1}{n}$ must order tx_1 before $\mathrm{tx}_2, \cdots$, tx_{n-1} before tx_n, and tx_n before tx_1 which is a contradiction.

To see the result for $\gamma = 1$, since $f \geq 1$, we observe that the adversary $\mathcal{A}$ can simply crash a single node N. Suppose that all other nodes receive tx_1 before tx_2. Now, since the node N sends no messages, other nodes do not know the order in which it received tx_1 and tx_2. Therefore, any protocol that satisfies receive-order-fairness for $\gamma = 1$ would order tx_1 before tx_2 even when N actually received tx_2 first. In other words it would also need to satisfy receive-order-fairness for $\gamma = \frac{n-1}{n}$, which we showed to be impossible.

4.2 Environments that support receive-order-fairness

We find that the Condorcet paradox can be circumvented in a few ways by assuming specific network properties.

External synchrony assumption. The primary reason for the impossibility of fair-ordering is that different nodes may receive the same client transaction several rounds apart, resulting in non-transitive collective ordering. Suppose that $\Delta_{\mathsf{ext}} = (\mathsf{full}, \delta)$ where $\delta \leq 1$ (e.g., an instant synchronous external network). Then, any client transaction that a node receives will reach all other nodes within 1 round. This implies that if some node receives transactions $\mathrm{tx}_1, \mathrm{tx}_2$ and tx_3 in that order, then no node can receive tx_3 before tx_1. It is now straightforward to see how this circumvents the Condorcet paradox.

Non-corrupting adversary and $\gamma = 1$. If the adversary does not corrupt any nodes, and its power is restricted to influencing network delays, we find that it is possible to achieve receive-order-fairness for $\gamma = 1$. In this setting, a single leader can receive the transaction orderings from individual nodes, and decide on a final ordering that preserves receive order-fairness.

4.3 Towards weaker definitions for order-fairness

We give two natural relaxations of the original definition. The first is *approximate receive order-fairness* (or simply *approximate-order-fairness*) while the second is *block receive order-fairness* (or simply *block-order-fairness*). For approximate-order-fairness, we only look at unfairness in the ordering of two transactions if they were received sufficiently apart in time. We emphasize that approximate-order-fairness only makes sense in synchronous and partially synchronous settings. On the other hand, for block-order-fairness, we choose to ignore the ordering within a block while considering fair ordering. Notably, this allows us to circumvent the Condorcet paradox by aggregating any transactions with non-transitive orderings into the same block. This is reasonable to consider even in asynchronous environments. First, we look at approximate-order-fairness. For a given view in the support of $\mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$, we define the property below.

Definition 10 (Approximate-Order-Fairness). *A* view *satisfies* $(\gamma, T_{\mathrm{warmup}}, \xi)$ *approximate-order-fairness if the following holds: For any two transactions* m *and* m'*, let* η *be the number of nodes that received both transactions between times* T_{warmup} *and* $|\mathsf{view}|$*. If at least* $\gamma\eta$ *of those nodes received* m *more than* ξ *rounds before* m' *from* $\mathcal{Z}$*, then for all honest nodes* i*,* i *does not deliver* m'*, unless it has previously delivered* m*.*

A protocol Π *satisfies* $(\gamma, T_{\mathrm{warmup}}, \xi)$ *approximate-order-fairness w.r.t* $(\mathcal{A}, \mathcal{Z})$ *if there is a negligible function* $\mathsf{negl}(\cdot)$ *such that for any* $\kappa \in \mathbb{N}$*, the above property is satisfied except with probability* $\mathsf{negl}(\kappa)$ *over* $\mathsf{view} \leftarrow\$ \mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$*.*

Quickly, we notice a protocol that satisfies $(T_{\mathrm{warmup}}, T_{\mathrm{confirm}})$-liveness, also satisfies $(1, T_{\mathrm{warmup}}, \xi)$ approximate order-fairness for any $\xi \geq T_{\mathrm{confirm}}$. Clearly, if a transaction tx_2 was received after tx_1 was delivered by all nodes, then tx_2 will be delivered after tx_1. Moreover, we also find that if $\xi < T_{\mathrm{confirm}}$, then any protocol that satisfies $(\gamma, T_{\mathrm{warmup}}, \xi)$ approximate-order-fairness must also satisfy $(\gamma, T_{\mathrm{warmup}})$ receive-order-fairness (for environments with a different network synchrony bound).

Theorem 3. *Consider any* $n, f \geq 1, \Delta_{\mathsf{int}}, \Delta_{\mathsf{ext}}$*. Let* $\Delta_{\mathsf{int}} = (\mathsf{not\text{-}async}, \delta_{\mathsf{int}})$ *and* $\Delta_{\mathsf{ext}} = (\mathsf{not\text{-}async}, \delta_{\mathsf{ext}} \geq 1)$*. Also consider* $\gamma \leq 1$ *and* $\xi < T_{\mathrm{confirm}}$*. If a protocol* Π *achieves consistency,* $(T_{\mathrm{warmup}}, T_{\mathrm{confirm}})$*-liveness, and* $(\gamma, T_{\mathrm{warmup}}, \xi)$ *approximate-order-fairness. w.r.t. all* $(\mathcal{A}, \mathcal{Z})$ *that respect* $(n, f, \Delta_{\mathsf{int}}, \Delta_{\mathsf{ext}})$*-classical execution, then it also satisfies* $(\gamma, T_{\mathrm{warmup}})$ *receive-order-fairness w.r.t all* $(\mathcal{A}', \mathcal{Z}')$ *that respect* $(n, f, \Delta'_{\mathsf{int}}, \Delta'_{\mathsf{ext}})$*-classical execution where* $\Delta'_{\mathsf{int}} = (\mathsf{not\text{-}async}, \delta'_{\mathsf{int}} = \frac{\delta_{\mathsf{int}}}{\xi+1})$ *and* $\Delta'_{\mathsf{ext}} = (\mathsf{not\text{-}async}, \delta'_{\mathsf{ext}} = \frac{\delta_{\mathsf{ext}}}{\xi+1})$*.*

Consequently, approximate-order-fairness doesn't turn out to be very useful since it suffers from the same problems as the previously defined receive-order-fairness. Note that from Section 4.2, we can infer that approximate-order-fairness can be achieved when $\delta_{\mathsf{ext}} \leq \xi$. Still, since it only applies to non-asynchronous networks, we propose a second definition, block-order-fairness, that performs much better since it provides a way to handle any cycles in transaction ordering and also applies to asynchronous networks. We note that our synchronous protocol (Section 6) also satisfies approximate-order-fairness for $\xi \geq \delta_{\mathsf{ext}}$.

For a given view in the support of $\mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$, we state the block-order-fairness property below.

Definition 11 (Block Order-Fairness). *A* view *satisfies* $(\gamma > \frac{1}{2}, T_{\mathrm{warmup}})$*-block-order-fairness if the following holds: For any two transactions* m *and* m'*, let* η *be the number of nodes that received both transactions between times* T_{warmup} *and* $|\mathsf{view}|$*. If at least* $\gamma\eta$ *of those nodes received* m *before* m' *from* $\mathcal{Z}$*, then for all honest nodes* i*,* i *does not deliver* m *at a later index than it delivers* m'*.*

A protocol Π *satisfies* $(\gamma, T_{\mathrm{warmup}})$*-block-order-fairness w.r.t* $(\mathcal{A}, \mathcal{Z})$ *if there is a negligible function* $\mathsf{negl}(\cdot)$ *such that for any* $\kappa \in \mathbb{N}$*, the above property is satisfied except with probability* $\mathsf{negl}(\kappa)$ *over* $\mathsf{view} \leftarrow\$ \mathsf{EXEC}^{\Pi}(\mathcal{A}, \mathcal{Z}, \kappa)$*.*

5 Overview of the Aequitas protocols

We provide a general overview of our Aequitas protocols in this section. Specifically, we give four constructions:

- $\Pi^{\mathsf{sync,nolead}}_{\mathsf{Aequitas}}$ is a leaderless protocol that provides consistency, (weak) liveness, and block-order-fairness in the completely synchronous setting.
- $\Pi^{\mathsf{sync,lead}}_{\mathsf{Aequitas}}$ is a leader-based protocol that provides consistency, (weak) liveness, and block-order-fairness in the completely synchronous setting.
- $\Pi^{\mathsf{async,nolead}}_{\mathsf{Aequitas}}$ is a leaderless protocol that provides consistency, eventual (weak) liveness, and block-order-fairness in any setting.
- $\Pi^{\mathsf{async,lead}}_{\mathsf{Aequitas}}$ is a leader-based protocol that provides consistency, eventual (weak) liveness, and block-order-fairness in any setting.

We present a detailed account only for the synchronous leaderless protocol $\Pi^{\mathsf{sync,nolead}}_{\mathsf{Aequitas}}$ in this paper (Section 6) and defer the other constructions to the full version [27].

Construction overview. Our Aequitas protocols utilize the FIFO-broadcast (FIFO-BC) and the set Byzantine agreement (Set-BA) primitives described in Section 3 in a black-box way to provide order-fairness. We elaborate on the three major stages of our Aequitas protocols below:

- **Stage I: Gossip / Broadcast**. Each node FIFO-broadcasts transactions as they are received as input from the environment. When a node i receives a set of transactions txs from $\mathcal{Z}$, it sends txs as input to the protocol $\Pi_{\text{fifocast}}[(\mathsf{sid}, i)]$ with i as the designated sender. Note that all broadcasts can be sent in the same session sid. Different session ids need to be used only when considering composition of several protocols in the system.

  In parallel to broadcasting transactions, a node also receives and processes broadcasts from other nodes. For a node i, broadcasts sent by node j are appended to a local log Log_i^j when they get delivered to i by $\Pi_{\text{fifocast}}[(\mathsf{sid}, j)]$. Intuitively, Log_i^j denotes node i's view of how transactions were received by node j.

- **Stage II: Agreement on local logs**. To determine the ordering for a particular transaction tx, a node i waits until it has received tx from sufficiently many other nodes. In other words, node i waits until there are sufficiently many k such that its local log Log_i^k contains tx. When both the external and internal networks are synchronous, this can alternatively be achieved by waiting for enough *time*. The properties of FIFO-BC guarantee that if two honest nodes i and j have local logs Log_i^k and Log_j^k respectively that both contain tx, then $\mathsf{Log}_i^k \approx_{\text{tx}} \mathsf{Log}_j^k$. We state this fact as Lemma 2. Recall that $\mathsf{Log}_i^k \approx_{\text{tx}} \mathsf{Log}_j^k$ holds when Log_i^k and Log_j^k are identical until tx occurs.

  Now, the next step is for all nodes to agree on which local logs to use to determine the ordering for tx. For a node i, let U_i^{tx} denote the set of nodes k such that Log_i^k contains tx. Node i starts an instance of the protocol $\Pi_{\text{sba}}[(\mathsf{sid}, \text{tx})]$ and provides it the input U_i^{tx}. Upon the completion of the Set-BA protocol, all honest nodes receive the same set L^{tx}. Intuitively, Set-BA is used to agree which nodes' orderings should be used to determine the final ordering for transaction tx. Recall that Lemma 1 guarantees that if $k \in L^{\text{tx}}$, then there is some honest node j such that $\text{tx} \in \mathsf{Log}_j^k$. This, along with the liveness property for FIFO-BC ensures that all honest nodes will eventually receive tx broadcast by node $k \in L^{\text{tx}}$ (even if k is malicious).

  Finally, we note that at the end of the agreement phase, every honest node has agreed on a set of nodes L^{tx} whose transaction orderings should be used to determine the final ordering for the transaction tx in consideration. We say that a node i has received the agreed logs for tx if for all $k \in L^{\text{tx}}$, it holds that $\text{tx} \in \mathsf{Log}_i^k$.

- **Stage III: Finalization.** To decide on the final ordering for a transaction tx, we provide two options for the finalization step: a leader based one and a leaderless one. For both the leader-based and leaderless finalizations, nodes first build a graph that represents any ordering dependencies between transactions. Specifically, a node i maintains a directed graph G_i, where vertices represent transactions and edges represent ordering dependencies. We refer to G_i as the "dependency graph" or the "waiting graph" maintained by i. After the agreement stage for tx is completed, the protocol now uses the local logs to see if some other transaction might have come before. If there is another transaction tx' that appears before tx in sufficiently many local logs (e.g., $n - f$ times), then i adds an edge from tx' to tx in G_i. Intuitively, an edge $(a, b) \in G_i$ denotes that the finalization stage for b is "waiting" for a to be delivered. Since the same L^{tx} is used by all honest nodes, if an edge (a, b) exists in G_i, then it will at some point exist in G_j, when nodes i and j are both honest. However, we note that G_i is neither guaranteed to be complete nor acyclic. Two vertices in G_i might never have an edge between them. Moreover, the Condorcet paradox can still create cycles in G_i. To break ties between transactions without an edge, we use the following two techniques.

- **Finalization via leader-based proposal.** $\Pi_{\mathsf{Aequitas}}^{\mathsf{sync,lead}}$ and $\Pi_{\mathsf{Aequitas}}^{\mathsf{async,lead}}$ both use a leader-based approach to finalize transactions in the graph. For this, any leader-based consensus protocol can be run along with the gossip and agreement stages above. When a designated leader proposes and broadcasts a new block, instead of just checking the syntactical validity of transactions, each node i checks that the proposal does not conflict with any required order-fairness in the graph G_i. That is, node i checks that for any transaction tx in the proposed block, if $(\mathrm{tx}', \mathrm{tx})$ is in G_i, then either tx' has already been delivered or tx' is also in the current proposed block.

  Abstractly, we allow the leader node to choose the transaction ordering but only as long as order-fairness is still satisfied. For transactions among which there is no clear winner, the leader may choose any ordering.
- **Finalization via local computation.** $\Pi_{\mathsf{Aequitas}}^{\mathsf{sync,nolead}}$ and $\Pi_{\mathsf{Aequitas}}^{\mathsf{async,nolead}}$ both use a leaderless approach to finalize transactions in the graph and require no further communication. At a high level, to order transactions tx_1 and tx_2 between whom there in no edge in G_i, the protocol will wait until tx_1 and tx_2 have a common descendant, with the final ordering being based on which transaction vertex has the most descendants. We prove that any other graph vertex that is a descendant of only one of tx_1 and tx_2 is present in G_i when node i makes the decision for ordering tx_1 and tx_2. This will ensure that all honest nodes will order tx_1 and tx_2 the same way.

We highlight that the above description of the finalization stage is a simplified one. As described, it is not sufficient to avoid the Condorcet paradox. Furthermore, adversarial transactions could result in a node waiting for unbounded periods of time. The actual technique to get around these obstacles is quite nuanced and we dedicate Section 5.1 to its details.

Lemma 2. *If two honest nodes i and j have local logs Log_i^k and Log_j^k respectively where k is any other node such that both logs contain a transaction* tx*, then* $\mathsf{Log}_i^k \approx_{\mathrm{tx}} \mathsf{Log}_j^k$.

Proof. This result follows directly from the agreement property of FIFO-BC.

Before diving into the details of the finalization step, we take a step back to understand why it turns out to be quite non-trivial. We look at a simple strawman protocol based on transaction timestamping that looks intuitive and analyze why it does not work.

The problem with timestamp-based ordering. Consider a simple synchronous protocol $\Pi_{\mathrm{timestamp}}$ that works as follows:

1. When an honest node i receives a transaction tx from $\mathcal{Z}$ in round t, it assigns tx the timestamp t and broadcasts (tx, t) to all other nodes.
2. Upon waiting for $\delta_{\mathsf{ext}} + T_{\mathrm{confirm}}$ rounds where δ_{ext} is the network delay bound for the external network and T_{confirm} is the liveness polynomial for the broadcast primitive, nodes reach agreement on the set of timestamps T to use to calculate the final timestamp for tx.
3. Each node calculates the final timestamp for tx as the median of all the timestamps in T. We represent this final timestamp by $\mathsf{final}(\mathrm{tx})$.

Notice how the first two steps almost perfectly resemble the gossip and agreement stages. The finalization (third) step is also surprisingly simple, but unfortunately can lead to easy manipulation of final timestamps by a single adversary. To see why, consider 5 nodes, A, B, C, D and E, where E is malicious and two transactions, tx_1 and tx_2. tx_1 is received by nodes $A, \ldots, E$ at rounds $1, 1, 4, 4, 2$ while tx_2 is received by the nodes at rounds $2, 2, 5, 5, 3$. Now, all nodes have received tx_1 before tx_2 and consequently, $\mathsf{final}(\mathrm{tx}_1) < \mathsf{final}(\mathrm{tx}_2)$ should hold. However, notice how E can invert the ordering of the final timestamps simply by switching around its own timestamps for tx_1 and tx_2. E can make $\mathsf{final}(\mathrm{tx}_1) = 3$ and $\mathsf{final}(\mathrm{tx}_2) = 2$ which results in a timestamp of 3 for tx_1 (median of (1, 1, 3, 4, 4)) and 2 for tx_2 (median of (2, 2, 2, 5, 5)), and thus an unfair ordering.

5.1 The Finalization Stage

We describe the general theme of the finalization stage here.

Ordering two transactions. For a pair of transactions tx and tx′, how does a node i choose which one to deliver first? Suppose that the agreement phases for tx and tx′ result in the outputs L^{tx} and $L^{\mathrm{tx}'}$. Define $l_{(\mathrm{tx},\mathrm{tx}')}$ as below.

$$l_{(\mathrm{tx},\mathrm{tx}')} = \left| \left\{ k \in L^{\mathrm{tx}} \cup L^{\mathrm{tx}'} \mid \mathsf{Log}_i^k(\mathrm{tx}) \leq \mathsf{Log}_i^k(\mathrm{tx}') \right\} \right|$$

$l_{(\mathrm{tx},\mathrm{tx}')}$ denotes the number of logs Log_i^k where tx was ordered at or before tx′. Now, if $l_{(\mathrm{tx},\mathrm{tx}')}$ is "small," it means that a large number of nodes have received tx′ before tx. This means that the finalization stage for tx should wait until tx′ has been delivered. This provides a partial ordering between any two transactions.

Additional notation. Let $\text{tx} \lhd_i \text{tx}'$ represent that i is waiting to deliver tx' before proceeding with the finalization phase for tx. Lemma 3 shows that $l_{(\text{tx},\text{tx}')}$ and $l_{(\text{tx}',\text{tx})}$ cannot both be "small". Consequently, both tx and tx' will not wait for each other or equivalently, at most one of $\text{tx} \lhd_i \text{tx}'$ and $\text{tx}' \lhd_i \text{tx}$ will be true.

Lemma 3. $l_{(\text{tx},\text{tx}')} + l_{(\text{tx}',\text{tx})} \geq \left| L^{\text{tx}} \cup L^{\text{tx}'} \right|$

Proof. Let $X = L^{\text{tx}} \cup L^{\text{tx}'}$. For any $k \in X$, at least one of $\mathsf{Log}_i^k(\text{tx}) \leq \mathsf{Log}_i^k(\text{tx}')$ and $\mathsf{Log}_i^k(\text{tx}') \leq \mathsf{Log}_i^k(\text{tx})$ is true. k is therefore counted in either $l_{(\text{tx},\text{tx}')}$ or $l_{(\text{tx}',\text{tx})}$ which proves the required result.

Adversarial transactions. The calculation of $l_{(\text{tx},\text{tx}')}$ needs to wait for the agreement phases of both tx and tx' to finish. Now, if an adversarial node FIFO-broadcasts a transaction tx_{fake} claiming it to be a real user transaction, then the ordering between tx_{fake} and a real transaction tx cannot be calculated since the agreement phase for tx_{fake} will never finish. So that this does not happen, the protocol needs to ensure that at least one honest node has received tx_{fake} before tx (from $\mathcal{Z}$). For the synchronous protocol, this is done by checking that a transaction tx' is added to the graph only when there is another transaction tx that has finished its agreement stage and tx' is present in at least $|L^{\text{tx}}|-(n-f)+1$ among the local logs in L^{tx}. Note that the agreement stage will only finish for honest transactions.

Non-transitive waiting. The Condorcet paradox can still cause non-transitive waiting. It is still possible to have transactions tx_1, tx_2, and tx_3 such that $\text{tx}_1 \lhd \text{tx}_2$; $\text{tx}_2 \lhd \text{tx}_3$; and $\text{tx}_3 \lhd \text{tx}_1$. The way we get around this is by delivering such transactions at the same time—by placing them in the same block.

Graph based approach. Instead of a separate thread waiting for the resolution of each transaction, representing the "waiting" between transactions as a graph provides a nice way to modularize the protocol. Suppose that each node i maintains a directed graph $G_i = (G_i.V, G_i.E)$ where $G_i.V$ denotes the set of vertices and $G_i.E$ denotes the set of edges in G_i. Each vertex represents a transaction and an edge from y to x (equiv. $(y, x) \in G_i.E$) represents that x is waiting on y i.e. $x \lhd_i y$. When the agreement phase for a transaction tx completes, i does the following:

- Add tx to the graph G_i if it does not already exist.
- For all transactions tx' such $\text{tx} \lhd_i \text{tx}'$, first, if tx' does not exist in the graph, add a new vertex. Then, add the edge (tx', tx) to G_i.

As mentioned before, G_i may not be acyclic. In order to deal with the Condorcet paradox, we consider the *strongly connected components* of G_i. Recall that a subgraph G' of a directed graph G is called strongly connected if every vertex in G' can reach every other vertex in G'. A strongly connected component is a maximal strongly connected subgraph.

Intuitively, all transactions in a strongly connected component will be delivered in the same block. A cycle that exists in G_i (due to non-transitivity of transactions) will be entirely contained in the same strongly connected component. On the other hand, if a transaction does not need to wait on any other one, then it will be in a strongly connected component by itself. We can collapse G_i into a new graph G_i^* where each strongly connected component is represented as a single vertex. G_i^* is also called the *condensation* of G_i. Each vertex in G_i^* will now denote a set of transactions. We note that G_i^* will now be acyclic.

Graph Notation. Since a vertex in G_i contains a single transaction, we may use a transaction and its corresponding vertex interchangeably when referring to the vertex in G_i. Let $\mathsf{TXS}_i(v)$ be the set of transactions for a vertex $v \in G_i^*.V$. Let $\mathsf{SCC}_i(v)$ denote the strongly connected component of G_i that contains the vertex v. $\mathsf{SCC}_i(v)$ also denotes the corresponding vertex in the condensation graph G_i^*.

Ordering incomparable vertices in G_i^* and breaking ties. As mentioned before, not all pairs of vertices in G_i^* are connected by an edge. This only gives a partial ordering for delivering transactions. We still need a way to totally order vertices in G_i^*. In the leader-based version of the finalization step, we delegate this responsibility to the leader node. We elaborate on the technique used in the synchronous leaderless protocol in Section 6.

Delivering a transaction. Recall that a transaction enters the *finalization* stage when it has completed the agreement stage, while it is *delivered* when it gets output to $\mathcal{Z}$ as part of the LOG. For the leaderless protocols, the set of transactions $\mathsf{TXS}_i(v)$ corresponding to the vertex $v \in G_i^*.V$ can be delivered in the LOG output to $\mathcal{Z}$ when it is not waiting for any other transaction and is preferred over any other transaction that it is incomparable with in the graph. For this, care must be taken to ensure that the set of transactions that tx is incomparable with is the same when all honest nodes are deciding to deliver tx, which we defer to the actual protocol description in Section 6.

6 The Synchronous Aequitas protocol

We describe $\Pi_{\mathsf{Aequitas}}^{\mathsf{sync,nolead}}$, the leaderless Aequitas protocol for the completely synchronous setting. By "complete synchrony," we mean that both the external and internal networks are synchronous. For this section, we assume that $(\mathcal{A}, \mathcal{Z})$ respects $\Delta_{\mathsf{ext}} = (\mathsf{full}, \delta_{\mathsf{ext}})$ ext-synchrony and $\Delta_{\mathsf{int}} = (\mathsf{full}, \delta_{\mathsf{int}})$ int-synchrony.

To build the $\Pi_{\mathsf{Aequitas}}^{\mathsf{sync,nolead}}$ protocol, we assume a secure FIFO-BC protocol Π_{fifocast} (from Definition 8) and a secure Set-BA protocol Π_{sba} (from Definition 7) that both work for any $(\mathcal{A}, \mathcal{Z})$ that respects $(n, f, \Delta_{\mathsf{int}}, \Delta_{\mathsf{ext}})$-classical execution. Let $(T_{\text{warmup}}^{\text{fifocast}}, T_{\text{confirm}}^{\text{fifocast}})$ and $T_{\text{confirm}}^{\mathsf{Set\text{-}BA}}$ denote the liveness parameters for Π_{fifocast} and Π_{sba} respectively. We note that any bound for the number of corruptions f will be at least as restrictive as bounds required by Π_{fifocast} and Π_{sba}.

6.1 Protocol Description

The $\Pi^{\mathsf{sync,nolead}}_{\mathsf{Aequitas}}$ protocol follows much of the same general techniques from Section 5. The gossip and agreement stage take place exactly as described there. In the gossip stage, a node i forks an instance of $\Pi_{\text{fifocast}}[(\mathsf{sid}, i)]$ and uses it to broadcast transactions as they are received from $\mathcal{Z}$. After broadcasting a transaction tx, it waits until the broadcasts from all honest nodes would have arrived. Let U_i^{tx} denote the set of nodes k such that $\text{tx} \in \mathsf{Log}_i^k$. Note that all honest nodes are present in U_i^{tx}. In the agreement stage, i forks an instance of $\Pi_{\text{sba}}[(\mathsf{sid}, \text{tx})]$ to agree on a set L^{tx} indicating the nodes whose logs to use to order tx.

For the finalization stage, we now present the remaining details that were deferred from Section 5.1. Please refer to Section 5 for any notation.

Building the "waiting" graph G_i. Recall that each node i builds a graph G_i where vertices are transactions and edges denote ordering dependencies between transactions. For two transactions tx and tx′, an edge (tx', tx) is added to G_i if $l_{(\text{tx},\text{tx}')} \leq \left| L^{\text{tx}} \cup L^{\text{tx}'} \right| - \gamma n + f$. Each node i also maintains the condensation graph G_i^* where each strongly connected component in G_i is condensed to a single vertex.

Ordering incomparable vertices in G_i^*. Suppose that v and v' are two vertices in G_i^* that are are currently not comparable i.e. they do not have an edge between them. To determine which vertex to deliver first, we wait until they have a common descendant, after which we order based on number of descendants. We note that once a common descendant arrives, any other transaction that arrives will also be a descendant of both v and v'. In other words, the vertex with the higher number of descendants will become fixed allowing for a consistent ordering across protocol nodes. Lemma 4 shows a helpful result on when vertices can be "incomparable."

A subtle point to note here is that the common descendant itself can cause v and v' to be combined into the same strongly connected component if it creates a cycle containing them. This is precisely why our protocol achieves weak-liveness, where we achieve liveness, if a transaction arrives late enough that it cannot create a cycle with transactions in v and v'. Effectively, we need to wait for a transaction to arrive sufficiently late in order to "flush out" earlier transactions.

Lemma 4. *Let v_1 and v_2 be two vertices in G_i^* that do not have an edge between them. Let r_{first} denote the time when any transaction in $\mathsf{TXS}_i(v_1)$ was first received by a node. Let r_{last} denote the time when any transaction in $\mathsf{TXS}_i(v_2)$ was last received by a node. Then $r_{\mathit{last}} - r_{\mathit{first}} \leq 2\delta_{\mathsf{ext}}$.*

Breaking ties. We use an a priori known ordering relation to break any ties that arise (e.g., two vertices with equal number of descendants). In particular, suppose that Ord is a binary relation on $2^{\mathcal{T}} \times 2^{\mathcal{T}}$ that is known a priori to all nodes. $2^{\mathcal{T}}$ represents the power set of $\mathcal{T}$. The relation is defined on sets of transactions (rather than individual transactions only) since we may deliver several transactions at once. We assume that Ord is supplied to all nodes on initialization by $\mathcal{Z}$. We will use this function to deterministically break ties between two sets of transactions when neither should clearly come before the other. For two sets S_1 and S_2, $(S_1, S_2) \in \mathsf{Ord}$ implies that all nodes agree S_1 should come before S_2 if there is no clear winner. Ord can also be used to order transactions in the same block. We note that Ord can be defined using a simple alphabetical or ascending order. In general, Ord needs to satisfy two properties:

- $\forall (a, b) \in 2^{\mathcal{T}} \times 2^{\mathcal{T}}; a \neq b$, exactly one of (a, b) and (b, a) is in Ord.
- $\forall a, b, c \in 2^{\mathcal{T}}$, if $(a, b) \in \mathsf{Ord}$ and $(b, c) \in \mathsf{Ord}$ then $(a, c) \in \mathsf{Ord}$.

Delivering transactions. The transactions $\mathsf{TXS}_i(v)$ of a vertex v in G_i^* can be delivered when:

- v is a source vertex i.e., it has no incoming edge. This ensures that v is not waiting on any other transaction to be delivered first.
- $2\delta_{\mathsf{ext}}$ rounds have passed since v was added to the graph. This ensures that any other vertex v' that v is incomparable to, is also present in the graph.
- For any other source vertex v', v has a common descendant with v' and either has more descendants or has an equal number of descendants and $(\mathsf{TXS}_i(v), \mathsf{TXS}_i(v')) \in \mathsf{Ord}$ holds. This ensures that every node will order v before v'.

Bound on f. Suppose that $(\gamma, \cdot)$ order-fairness needs to be realized. This implies that if γn nodes receive transactions in a particular order, it must be reflected in the final ordering. Since f nodes can be adversarial, the output must be the same even if $\gamma n - f$ of those orderings are seen. Now, as we don't want a bi-directed edge to be added to G_i, $\gamma n - f > \frac{n}{2}$ must hold. Equivalently, $n > \frac{2f}{2\gamma - 1}$. For block-order-fairness with $\gamma = 1$, we require an honest majority.

6.2 Protocol Pseudocode

Initialization. At the start of the protocol, we assume that i receives the identities of other protocol nodes, n, f, the maximum network delays δ_{int}, δ_{ext}, and the binary relation Ord. A $\mathsf{FIFO\text{-}BC}$ protocol Π_{fifocast} and a $\mathsf{Set\text{-}BA}$ protocol Π_{sba} have also been agreed upon a priori. Let $T^{\text{fifocast}}_{\text{confirm}}$ and $T^{\text{sba}}_{\text{confirm}}$ represent the liveness bounds for Π_{fifocast} and Π_{sba} respectively. Now, for each $j \in \mathcal{N}$, i initializes $\mathsf{Log}_i^j \leftarrow []$. It also initializes an empty graph G_i and a final output log LOG_i.

• At the start of round r, when a node i receives a set of transactions txs from $\mathcal{Z}$, it does the following:

1. **(Gossip)**
   (a) Fork an instance of $\Pi_{\text{fifocast}}[(\mathsf{sid}, i)]$, if it does not already exist.
   (b) Send txs as input to $\Pi_{\text{fifocast}}[(\mathsf{sid}, i)]$.
   (c) Record $(\mathsf{sid}, \mathsf{gossip\text{-}end}, \mathsf{txs}, r + \delta_{\mathsf{ext}} + T^{\text{fifocast}}_{\text{confirm}})$

2. **(Agreement)**
   (a) Check if there is any recorded tuple $(\mathsf{sid}, \mathsf{gossip\text{-}end}, \mathsf{txs}', r')$ such that $r = r'$.
   (b) For such a tuple for txs', for each $\mathrm{tx} \in \mathsf{txs}'$, fork an instance of $\Pi_{\mathrm{sba}}[(\mathsf{sid}, \mathrm{tx})]$ and provide it the input U_i^{tx}.
   (c) Record $(\mathsf{sid}, \mathsf{agreement\text{-}end}, \mathrm{tx}, r + T^{\mathrm{sba}}_{\mathrm{confirm}})$ for each $\mathrm{tx} \in \mathsf{txs}'$.

3. **(Build Graph)**
   (a) Check if there is any recorded tuple $(\mathsf{sid}, \mathsf{agreement\text{-}end}, \mathrm{tx}, r')$ such that $r = r'$.
   (b) For such a tuple for tx, first add a vertex denoted by tx to G_i if it does not already exist. Now, for any other transaction tx' seen so far that has not yet been delivered,
      i. Let $u = \left|\left\{k \in L^{\mathrm{tx}} \mid \mathrm{tx}' \in \mathsf{Log}_i^k\right\}\right|$.
      ii. If $u \geq |L^{\mathrm{tx}}| - (n - f) + 1$, compute $l_{(\mathrm{tx},\mathrm{tx}')}$ as per Section 5.1.
      iii. If $l_{(\mathrm{tx},\mathrm{tx}')} \leq \left|L^{\mathrm{tx}} \cup L^{\mathrm{tx}'}\right| - \gamma n + f$, then record $\mathrm{tx} \lhd \mathrm{tx}'$. Add an edge $(\mathrm{tx}', \mathrm{tx})$ to G_i if it does not already exist.
   (c) Record $(\mathsf{sid}, \mathsf{graph\text{-}end}, \mathrm{tx}, r + 2\delta_{\mathsf{ext}})$ for tx.

4. **(Finalization)**
   (a) Compute the *condensation* graph G_i^* of G_i by collapsing each strongly connected component into a single vertex.
   (b) Let V_{source} be the set of vertices in G_i^* where $v \in V_{\mathsf{source}}$ if it satisfies:
      - All transactions in $\mathsf{TXS}(v)$ have been received.
      - v is a source vertex in G_i^*. That is, v has no incoming edges.
   (c) Let $V_{\mathsf{finalize}} \subseteq V_{\mathsf{source}}$ be the set of vertices v that also satisfy:
      - For all $\mathrm{tx}^* \in \mathsf{TXS}(v)$, there is any previously recorded tuple $(\mathsf{sid}, \mathsf{graph\text{-}end}, \mathrm{tx}^*, r')$ with $r \geq r'$
   (d) For $v \in V_{\mathsf{source}}$, let $\mathsf{Desc}(v)$ denote the descendants of v in G_i^*. Let $\mathsf{nDesc}(v) = |\mathsf{Desc}(v)|$ i.e. the number of descendants.
   (e) For $v \in V_{\mathsf{finalize}}$ and $v' \in V_{\mathsf{source}}$, let $\mathsf{common\text{-}desc}_{(v,v')}$ be a boolean that denotes whether v and v' have a common descendant. That is, we define $\mathsf{common\text{-}desc}_{(v,v')} := (\mathsf{Desc}(v) \cap \mathsf{Desc}(v') \neq \emptyset)$
   (f) If there is a $v \in V_{\mathsf{finalize}}$ such that for all other $v' \in V_{\mathsf{source}}$,
      - $\mathsf{common\text{-}desc}_{(v,v')} = \mathsf{true}$
      - Either $\mathsf{nDesc}(v) > \mathsf{nDesc}(v')$ holds or $(\mathsf{nDesc}(v) = \mathsf{nDesc}(v')) \wedge (\mathsf{TXS}(v), \mathsf{TXS}(v')) \in \mathsf{Ord}$.

      then, deliver transactions in v by appending $\mathsf{TXS}(v)$ to LOG_i. Remove v from G_i^* and the corresponding vertices form G_i.
   (g) Repeat steps 4b to 4f until there is no such v in step 4f.
   (h) Output the current LOG_i to $\mathcal{Z}$.

- When i receives txs from $\Pi_{\mathrm{fifocast}}[(\mathsf{sid}, j)]$, it appends txs to Log_i^j and adds j to the set U_i^{tx}.
- When i receives the output from $\Pi_{\mathrm{sba}}[(\mathsf{sid}, \mathrm{tx})]$, it stores it as L^{tx}.

Transaction Lifecycle. Suppose that a transaction tx is input to node i in round r_0. Since the external network is synchronous, by round $r_0+\delta_{\mathsf{ext}}$, all nodes will have been input tx by $\mathcal{Z}$. Consequently, by round $r_1 = r_0 + \delta_{\mathsf{ext}} + T^{\text{fifocast}}_{\text{confirm}}$, node i will have received the gossip broadcasts from all other honest nodes. By round $r_2 = r_1 + T^{\text{sba}}_{\text{confirm}}$, node i will receive the output of the agreement stage for tx, and tx can be added to the graph G_i. Now by round $r_3 = r_2 + 2\delta_{\mathsf{ext}}$, any other transaction that tx could be incomparable with will also get added to G_i. Waiting for this time ensures that tx does not get delivered before ensuring that all relevant transactions have been placed in the graph.

6.3 Consistency, Liveness, and Order-Fairness Results

We present the consistency, liveness, and order-fairness results for $\Pi^{\mathsf{sync,nolead}}_{\mathsf{Aequitas}}$ in Theorem 4. We provide brief proof sketches, and defer the formal proofs to the full version [27]. As a corollary, we also note that $\Pi^{\mathsf{sync,nolead}}_{\mathsf{Aequitas}}$ also satisfies receive-order-fairness, and (conventional) liveness when the external network has $\delta_{\mathsf{ext}} = 1$, since non-transitive Condorcet cycles can no longer arise.

Theorem 4 (Consistency, Liveness, and Order-Fairness of $\Pi^{\mathsf{sync,nolead}}_{\mathsf{Aequitas}}$). *Consider any $n, f, \gamma > \frac{1}{2}, \Delta_{\mathsf{ext}} = (\mathsf{full}, \delta_{\mathsf{ext}}), \Delta_{\mathsf{int}} = (\mathsf{full}, \delta_{\mathsf{int}})$ with $n > \frac{2f}{2\gamma-1}$. Let Π_{fifocast} be a secure* FIFO-BC *protocol and Π_{sba} be a secure* Set-BA *protocol. Further, suppose that Π_{fifocast} satisfies $(T^{\text{fifocast}}_{\text{warmup}}, T^{\text{fifocast}}_{\text{confirm}})$ liveness, and Π_{sba} satisfies $T^{\text{sba}}_{\text{confirm}}$ liveness. Then $\Pi^{\mathsf{sync,nolead}}_{\mathsf{Aequitas}}$ satisfies consistency, $(T^{\text{fifocast}}_{\text{warmup}}, T^{*}_{\text{confirm}})$ weak-liveness where $T^{*}_{\text{confirm}} = 2\delta_{\mathsf{ext}} + T^{\text{fifocast}}_{\text{confirm}} + T^{\text{sba}}_{\text{confirm}}$, and $(\gamma, T^{\text{fifocast}}_{\text{warmup}})$ block-order-fairness w.r.t. any $(\mathcal{A}, \mathcal{Z})$ that respects $(n, f, \Delta_{\mathsf{int}}, \Delta_{\mathsf{ext}})$-classical execution.*

Consistency proof sketch. To show consistency, we need to prove that two honest nodes i and j remove transactions from their graphs G^*_i and G^*_j in the same order. For this, we first present a helpful lemma (Lemma 5).

Lemma 5. *Suppose that when an honest node i delivers* tx, *$v = \mathsf{SCC}_i(\text{tx})$ is the vertex that contains* tx *in G^*_i. Now, if another honest node j delivers* tx *and $v' = \mathsf{SCC}_j(\text{tx})$ at that point, then $\mathsf{TXS}_i(v) = \mathsf{TXS}_j(v')$, or equivalently $\mathsf{SCC}_i(\text{tx}) = \mathsf{SCC}_j(\text{tx})$ when* tx *is output by each of the nodes. This means that we can drop the node subscripts.*

Now, suppose that node i delivers a transaction tx_1 before another one tx_2. Let $v_1 = \mathsf{SCC}_i(\text{tx}_1)$ and $v_2 = \mathsf{SCC}_i(\text{tx}_2)$ be vertices in G^*_i when tx_1 and tx_2 were delivered. Note that by Lemma 5, we can also use v_1 and v_2 to denote the vertices when j delivers tx_1 and tx_2. Now, either tx_1 was delivered even before tx_2 was added to G_i, or there is an edge from v_1 to v_2 in G^*_i (which caused tx_1 to be output before) or v_1 and v_2 are incomparable.

- If tx_1 was delivered before tx_2 was added to G_i, then at least $\gamma n - f$ nodes received tx_1 before tx_2. Therefore, even if tx_2 gets added to G_j before tx_1, there will be an edge from tx_1 to tx_2 in G_j. By Lemma 5, tx_1 cannot be in the same SCC as tx_2 either, which implies that j cannot deliver tx_2 first.

- If (v_1, v_2) is an edge in G_i^*, then it will also be in G_j^* when j delivers $\mathsf{TXS}(v_2)$. This means that j cannot deliver $\mathsf{TXS}(v_2)$ before it delivers $\mathsf{TXS}(v_1)$.
- If there is no edge between v_1 and v_2 in G_i^*, then node i delivers $\mathsf{TXS}(v_1)$ before because v_1 had more descendants (or because of the deterministic tie-breaker). Since j waits for $2\delta_{\mathsf{ext}}$ time, both v_1 and v_2 are present in its graph G_j^* when j outputs $\mathsf{TXS}(v_2)$, causing j to wait for a common descendant of v_1 and v_2 to be added. By this time, any other vertex that is not a common descendant will also be in G_j^*, and the difference in the number of descendants of v_1 and v_2 will remain constant henceforth. This means that j will take the same decision as i to deliver $\mathsf{TXS}(v_1)$ before $\mathsf{TXS}(v_2)$.

Weak-Liveness proof sketch. To show weak-liveness for a transaction tx, first, in Lemma 6, we prove that if a transaction is input sufficiently after tx, it cannot be coalesced into the same strongly connected component as tx.

Lemma 6. *Consider a transaction* tx *and build the set* T *as per the weak-liveness definition. Now, let* tx′ *be a transaction that is input to all nodes after all transactions in* T. *Then* $\mathsf{SCC}_i(\text{tx}) \neq \mathsf{SCC}_i(\text{tx}')$ *for any honest* i.

Now, suppose that tx was first input by $\mathcal{Z}$ in round $r > T_{\text{warmup}}^{\text{fifocast}}$. Consider the set T built form tx as in the weak-liveness definition. Suppose now that a transaction $\text{tx}_{\mathsf{flush}}$ is input to all nodes after all transactions in T. Let r_{flush} be the round that $\text{tx}_{\mathsf{flush}}$ is first input to some node. Then, $\text{tx}_{\mathsf{flush}}$ is received by all nodes by round $r_{\mathsf{flush}} + \delta_{\mathsf{ext}}$ and therefore added to all honest graphs G_i by round $r_{\mathsf{flush}} + 2\delta_{\mathsf{ext}} + T_{\text{confirm}}^{\text{fifocast}} + T_{\text{confirm}}^{\text{sba}}$. From Lemma 6, $v = \mathsf{SCC}_i(\text{tx}) \neq \mathsf{SCC}_i(\text{tx}_{\mathsf{flush}})$ for any honest i. Now, any transaction tx′ that tx is incomparable was input to at least one honest node no later than tx, i.e. $\text{tx}_{\mathsf{flush}}$ was received after tx′ by all honest nodes. Consequently, $\text{tx}_{\mathsf{flush}}$ will be a descendant of both tx and tx′. This means that node i can deliver $\mathsf{TXS}_i(\text{tx})$ when $\text{tx}_{\mathsf{flush}}$ gets added to its graph, which happens by round $r_{\mathsf{flush}} + T_{\text{confirm}}^*$.

Order-Fairness proof sketch. First, we note that if γn nodes receive tx_1 before tx_2, then at least $\gamma n - f$ honest ones do. This means that there will be an edge from tx_1 to tx_2 in all honest G_i. Consequently, either tx_1 will be delivered before tx_2 by all nodes, or it will end up in the same strongly connected component as tx_2 and be delivered at the same time.

Acknowledgements. This work was funded by NSF grants CNS-1564102, CNS-1704615, and CNS-1933655 as well as support from IC3 industry partners. We would also like to thank Mic Bowman at Intel for drawing attention to potential applications.

References

[1] Ittai Abraham et al. "Solida: A Blockchain Protocol Based on Reconfigurable Byzantine Consensus". In: *OPODIS*. 2017, 25:1–25:19.

[2] Ittai Abraham et al. *Sync HotStuff: Simple and Practical Synchronous State Machine Replication*. Cryptology ePrint Archive, Report 2019/270. 2019.
[3] Yair Amir et al. "Prime: Byzantine Replication Under Attack". In: *IEEE TDSC* 8.4 (2011), pp. 564–577.
[4] Avi Asayag et al. "A Fair Consensus Protocol for Transaction Ordering". In: *ICNP*. 2018, pp. 55–65.
[5] Pierre-Louis Aublin, Sonia Ben Mokhtar, and Vivien Quéma. "RBFT: Redundant Byzantine Fault Tolerance". In: *ICDCS*. 2013, pp. 297–306.
[6] Leemon Baird. *The Swirlds Hashgraph Consensus Algorithm: Fair, Fast, Byzantine Fault Tolerance*. `https://www.swirlds.com/downloads/SWIRLDS-TR-2016-01.pdf`. 2016.
[7] Shehar Bano et al. *Consensus in the Age of Blockchains*. arXiv:/1711.03936. 2017.
[8] Alysson Bessani, João Sousa, and Eduardo E.P. Alchieri. "State Machine Replication for the Masses with BFT-SMART". In: *DSN*. 2014, pp. 355–362.
[9] Gabriel Bracha and Sam Toueg. "Asynchronous Consensus and Broadcast Protocols". In: *J. ACM* 32.4 (1985), pp. 824–840.
[10] Christian Cachin and Marko Vukolić. *Blockchain Consensus Protocols in the Wild*. arXiv:/1707.01873. 2017.
[11] Christian Cachin et al. "Secure and Efficient Asynchronous Broadcast Protocols". In: *CRYPTO*. 2001, pp. 524–541.
[12] Ran Canetti. "Universally Composable Security: A New Paradigm for Cryptographic Protocols". In: *FOCS*. 2001, pp. 136–147.
[13] Ran Canetti and Tal Rabin. "Universal composition with joint state". In: *CRYPTO*. 2003, pp. 265–281.
[14] Ran Canetti et al. "Universally composable security with global setup." In: *TCC*. 2007, pp. 61–85.
[15] Miguel Castro and Barbara Liskov. "Practical Byzantine Fault Tolerance". In: *OSDI*. 1999, pp. 173–186.
[16] T.-H. Hubert Chan, Rafael Pass, and Elaine Shi. "Consensus through Herding". In: *EUROCRYPT*. 2019, pp. 720–749.
[17] Allen Clement et al. "Making byzantine fault tolerant systems tolerate byzantine faults". In: *NDSI*. 2009, pp. 153–168.
[18] *Condorcet Paradox*. `https://wikipedia.org/wiki/Condorcet_paradox`.
[19] Flaviu Cristian et al. "Atomic Broadcast: From Simple Message Diffusion to Byzantine Agreement". In: *Information and Computation* 118.1 (1995), pp. 158–179.
[20] Philip Daian et al. "Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability". In: *IEEE S&P*. 2020, pp. 585–602.
[21] Danny Dolev and H. Raymond Strong. "Authenticated Algorithms for Byzantine Agreement". In: *SIAM J. Comput* 12 (1983), pp. 656–666.

[22] Cynthia Dwork, Nancy Lynch, and Larry Stockmeyer. “Consensus in the presence of partial synchrony”. In: *J. ACM* 35.2 (1988), pp. 288–323.

[23] *Ethereum*. `https://ethereum.org/`.

[24] Ittay Eyal and Emin Gün Sirer. “Majority is not Enough: Bitcoin Mining is Vulnerable”. In: *FC*. 2014, pp. 436–454.

[25] Yossi Gilad et al. “Algorand: Scaling Byzantine Agreements for Cryptocurrencies”. In: *SOSP*. 2017, pp. 51–68.

[26] Chi Ho, Danny Dolev, and Robbert van Renesse. “Making distributed systems robust”. In: *OPODIS*. 2007, pp. 232–246.

[27] Mahimna Kelkar et al. *Order-Fairness for Byzantine Consensus*. Cryptology ePrint Archive, Report 2020/269. 2020.

[28] Aggelos Kiayias et al. “Ouroboros: A Provably Secure Proof of Stake Blockchain Protocol”. In: *CRYPTO*. 2017, pp. 357–388.

[29] Eleftherios Kokoris-Kogias et al. “OmniLedger: A Secure, Scale-Out, Decentralized Ledger via Sharding”. In: *IEEE S&P*. 2018, pp. 583–598.

[30] Leslie Lamport, Robert Shostak, and Marshall Pease. “The Byzantine Generals Problem”. In: *TOPLAS* 4.3 (1982), pp. 382–401.

[31] Kfir Lev-Ari et al. “FairLedger: A Fair Blockchain Protocol for Financial Institutions”. In: *OPODIS*. 2019, 1:1–1:16.

[32] Michael Lewis. *Flash Boys: A Wall Street Revolt*. WW Norton & Company, 2014.

[33] Loi Luu et al. “SmartPool: Practical Decentralized Pooled Mining”. In: *USENIX Security*. 2017, pp. 1409–1426.

[34] Jean-Philippe Martin and Lorenzo Alvisi. “Fast Byzantine Consensus”. In: *IEEE TDSC* 3.3 (2006), pp. 202–215.

[35] Andrew Miller et al. “Non-outsourceable scratch-off puzzles to discourage bitcoin mining coalitions”. In: *ACM CCS*. 2015, pp. 680–691.

[36] Andrew Miller et al. “The Honey Badger of BFT Protocols”. In: *ACM CCS*. 2016, pp. 31–42.

[37] Rafael Pass and Elaine Shi. “FruitChains: A Fair Blockchain”. In: *PODC*. 2017, pp. 315–324.

[38] Rafael Pass and Elaine Shi. “Hybrid Consensus: Efficient Consensus in the Permissionless Model”. In: *DISC*. 2017, pp. 1–16.

[39] Rafael Pass and Elaine Shi. “Rethinking Large-Scale Consensus”. In: *CSF*. 2017, pp. 115–129.

[40] Rafael Pass and Elaine Shi. “Thunderella: Blockchains with Optimistic Instant Confirmation”. In: *EUROCRYPT*. 2018, pp. 3–33.

[41] Team Rocket et al. *Scalable and Probabilistic Leaderless BFT Consensus through Metastability*. arXiv:/1906.08936. 2019.

[42] Giuliana Santos Veronese et al. “Efficient Byzantine Fault-Tolerance”. In: *IEEE Transactions on Computers* 62.1 (2013), pp. 16–30.

[43] Giuliana Santos Veronese et al. “Spin One’s Wheels? Byzantine Fault Tolerance with a Spinning Primary”. In: *SRDS*. 2009, pp. 135–144.

[44] Maofan Yin et al. “HotStuff: BFT Consensus with Linearity and Responsiveness”. In: *PODC*. 2019, pp. 347–356.