

---

# Combining GHOST and Casper

---

**Vitalik Buterin**  
Ethereum Foundation  
v@buterin.com

**Diego Hernandez**  
SJSU  
diego.hernandez@sjsu.edu

**Thor Kamphofner**  
SJSU  
thor.kamphofner@sjsu.edu

**Khiem Pham**  
VinAI Research  
v.khiempd1@vinai.io

**Zhi Qiao**  
SJSU  
qiaozhi9092@gmail.com

**Danny Ryan**  
Ethereum Foundation  
danny@ethereum.org

**Juhyeok Sin**  
SJSU  
homin203@gmail.com

**Ying Wang**  
SJSU  
yingwang1001@gmail.com

**Yan X Zhang**  
SJSU  
yan.x.zhang@sjsu.edu

## Abstract

We present “Gasper,” a proof-of-stake-based consensus protocol, which is an idealized version of the proposed Ethereum 2.0 beacon chain. The protocol combines Casper FFG, a finality tool, with LMD GHOST, a fork-choice rule. We prove safety, plausible liveness, and probabilistic liveness under different sets of assumptions.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Setup and Goals</b>	<b>3</b>
2.1	Consensus Protocols, Validators, Blockchain . . . . .	3
2.2	Messages and Views . . . . .	4
2.3	Proof-of-stake . . . . .	5
2.4	Byzantine Validators, PBFT . . . . .	7
2.5	Safety and Liveness . . . . .	7
2.6	Time, Epochs, and Synchrony . . . . .	8
<b>3</b>	<b>Main Ingredients</b>	<b>8</b>
3.1	Casper FFG . . . . .	8
3.2	LMD GHOST Fork-Choice rule . . . . .	10

<b>4</b>	<b>Main Protocol: Gasper</b>	<b>11</b>
4.1	Epoch Boundary Blocks and Pairs . . . . .	11
4.2	Committees . . . . .	12
4.3	Blocks and Attestations . . . . .	12
4.4	Justification . . . . .	14
4.5	Finalization . . . . .	15
4.6	Hybrid LMD GHOST . . . . .	16
4.7	Slashing Conditions . . . . .	17
4.8	Rewards and Penalties . . . . .	18
<b>5</b>	<b>Safety</b>	<b>18</b>
<b>6</b>	<b>Plausible Liveness</b>	<b>19</b>
<b>7</b>	<b>Probabilistic Liveness</b>	<b>20</b>
7.1	High Weight After the First Slot - The Equivocation Game . . . . .	21
7.2	High Weight after First Slot Leads to Justification . . . . .	21
7.3	Probabilistic Justification Leads to Probabilistic Finalization . . . . .	24
<b>8</b>	<b>Practice versus Theory</b>	<b>26</b>
8.1	Sharding . . . . .	26
8.2	Implementing the View . . . . .	26
8.3	Attestation Inclusion Delay . . . . .	27
8.4	Attestation Consideration Delay . . . . .	27
8.5	A Four-Case Finalization Rule . . . . .	28
8.6	Safety - Dynamic Validator Sets . . . . .	28
8.7	Extreme Cases; Hard Forks . . . . .	31
<b>9</b>	<b>Conclusion</b>	<b>31</b>
<b>A</b>	<b>Technicalities of Views</b>	<b>35</b>
<b>B</b>	<b>The Equivocation Game</b>	<b>35</b>
B.1	The Pessimistic Regime - High Latency . . . . .	36
B.2	The Optimistic Regime - Low Latency . . . . .	37
B.3	An Example Inbetween . . . . .	38

# 1 Introduction

Our goal is to create a consensus protocol for a proof-of-stake blockchain. This paper is motivated by the need to construct Ethereum’s “beacon chain” in its sharding design. However, the construction can also be used for a solo blockchain with no sharding.

We present Gasper, a combination of Casper FFG (the Friendly Finality Gadget) and LMD GHOST (Latest Message Driven Greediest Heaviest Observed SubTree). Casper FFG [7] is a “finality gadget,” which is an algorithm that marks certain blocks in a blockchain as *finalized* so that participants with partial information can still be fully confident that the blocks are part of the canonical chain of blocks. Casper is not a fully-specified protocol and is designed to be a “gadget” that works on top of a provided blockchain protocol, agnostic to whether the provided chain were proof-of-work or proof-of-stake. LMD GHOST is a *fork-choice rule* where *validators* (participants) *attest* to blocks to signal support for those blocks, like voting. Gasper is a full proof-of-stake protocol that is an idealized abstraction of the proposed Ethereum implementation.

In Sections 2 and 3, we define our primitives, provide background knowledge, and state our goals. In Section 4, we give our main protocol Gasper. In Sections 5, 6, and 7, we formally prove Gasper’s desired qualities. In Section 8, we summarize some differences between Gasper and the actual planned Ethereum implementation, such as delays in accepting attestations, delaying finalization, and dynamic validator sets. We conclude with some thoughts and ideas for future research in Section 9.

## 2 Setup and Goals

The main goal of this paper is to describe and prove properties of a proof-of-stake blockchain with certain safety and liveness claims. In this section, we give some general background on consensus protocols and blockchain. As the literature uses a diverse lexicon, the primary purpose of this section is to pick out a specific set of vocabulary for the rest of the paper.

### 2.1 Consensus Protocols, Validators, Blockchain

Our goal is to create a *consensus protocol* (or *protocol* for short from this point on), which is an agreed suite of algorithms for a set of entities (nodes, people, etc.) to follow in order to obtain a consensus history of their state, even if the network is unreliable and/or if many validators are malicious.

We call the entities (people, programs, etc.) who participate in protocol *validators*, denoted by the set  $\mathcal{V}$ . They are called “validators” because of their data-validation role in the Ethereum 2.0 beacon chain, though the specifics of this role are not part of our abstract protocol. Other works use many different words for the same concept: e.g. “replicas” in classic PBFT literature, “block producers” in DPoS and EOS, “peers” on Peercoin and NXT, or “bakers” on Tezos. The validators are connected to each other on a (peer-to-peer) *network*, which means they can broadcast *messages* (basically, packets of data) to each other. In the types of protocols we are interested in, the primary types of messages are *blocks*, which collect bundles of messages. The first block is called the *genesis block* and acts as the initial “blank slate” state. Other blocks are descriptions of state transitions with a pointer to a *parent* block<sup>1</sup>. A *blockchain* is just an instantiation (i.e. with specific network conditions, validators, etc.) of such a consensus protocol where we use blocks to build the overall state.

We do not want all blocks seen on the network to be accepted as common history because blocks can conflict with each other. These conflicts can happen for honest reasons such as network latency, or for malicious reasons such as byzantine validators wanting to double-spend. *Byzantine* refers to a classic problem in distributed systems, the Byzantine Generals Problem, and is used in this instance to describe a validator who does not follow the protocol, either because they are malicious or experiencing network latency. Graph-theoretically, one can think of a choice of history as a choice of a “chain” going from the genesis block to a particular block. Thus, a consensus history is

---

<sup>1</sup>Our choice that each block has a unique parent is almost taken for granted in blockchains; one can imagine blockchains where blocks do not need to have parents, or can have more than one parent. See, for instance, Hashgraph [2]. Even the very general structure of “block” itself may be limiting!

a consensus of which chain of blocks to accept as correct. This is why we call such an instantiation of our protocol a “blockchain.”

**Example 2.1.** Assume the goal of a protocol is to keep a ledger of every validator’s account balances. Then their collective state is the ledger, and each block captures a state transition, such as a monetary transaction from one validator to another (e.g., “Yunice sends coin with ID 538 to Bureau-gard,”). One type of conflict we may want to avoid is having both the aforementioned transaction and another transaction with content “Yunice sends coin with ID 538 to Carol” be accepted into consensus history, an act commonly called *double-spending*. The current consensus state can be determined by starting from a “genesis state” and sequentially processing each message in the consensus history (hopefully one that all honest validators will agree with), starting from the genesis block.

## 2.2 Messages and Views

In our model, a validator  $V$  interacts with the network by broadcasting *messages*, which are strings in some language. When an honest validator  $V$  broadcasts a message  $M$ , we assume  $M$  is sent to all validators on the network<sup>2</sup>.

The main type of message *proposes a block*, which is a piece of data, to the network. Other messages can be bookkeeping notices such as voting for blocks (“attestation”), putting new validators on the blockchain (“activation”), proving bad actions of other validators (“slashing”), etc. depending on the specific protocol. We assume that each message is digitally signed by a single validator, which means we can accurately trace the author of each message and attacks such as impersonation of honest validators are not possible.

Due to network latency and dishonest validators (delaying messages or relaying incorrect information), validators may have different states of knowledge of the full set of messages given to the network; we formalize this by saying that a validator either *sees* or *does not see* each message given to the network at any given time.

Each message may have one or more *dependencies*, where each dependency is another message. At any time, we *accept* a message if and only if all of its dependencies (possibly none) are accepted, defined recursively. We now define the *view* of a validator  $V$  at a given time  $T$ , denoted as  $\text{view}(V, T)$ , as the set of all the **accepted** messages the validator has seen so far. We also have a “God’s-eye-view” that we call the *network view*, defined to be the set of accepted messages for a hypothetical validator that has seen (with no latency) all messages any validator has broadcast at any time (this includes messages sent by a malicious validator to only a subset of the network). We will treat the network as a “virtual validator”, so we use  $\text{view}(\text{NW}, T)$  to denote the network view at time  $T$ . For any validator  $V$  and any given time  $T$ ,  $\text{view}(\text{NW}, T)$  includes all the messages in  $\text{view}(V, T)$ , though the timestamps may be mismatched. Finally, since usually the context is that we are talking about a specific point in time, we will usually just suppress the time and use notation such as  $\text{view}(V)$  to talk about  $\text{view}(V, T)$ , unless talking about the specific time is necessary.

**Example 2.2.** At time  $T$ , suppose validator  $V$  sees the message  $M$  with content “Yunice sends coin number 5340 to Bob,” but has not yet seen the message  $M'$  containing “Yunice obtains coin number 5340” (which is on the network but has not yet made its way to  $V$ ). Also suppose that in our protocol  $M$  depends on  $M'$  (and no other message), which captures the semantic meaning that we need to obtain a coin before spending it, and that  $V$  cannot and should not act on  $M$  without seeing  $M'$ . In this situation we say that  $V$  has seen but has not accepted  $M$ , so  $M \notin \text{view}(V, T)$ . However, the network has seen both; so if the network accepts  $M'$ , we have  $M \in \text{view}(\text{NW}, T)$ .

We now make some assumptions on the structure of views:

- Everyone (the validators and the network) starts with a single message (with no signature) corresponding to a single agreed-upon *genesis block*, denoted  $B_{\text{genesis}}$ , with no dependencies (so it starts out automatically accepted into all views).

<sup>2</sup>We model the validators as sending a single message to an abstract “network,” which handles the message. In practice, it may be part of the protocol that e.g., every honest validator rebroadcasts all messages that they see, for robustness. We consider these implementation details not important to our paper.

- Each block besides  $B_{\text{genesis}}$  refers to (and depends on) a *parent block* as part of its data. Thus, we can visualize  $\text{view}(V)$  as a directed acyclic graph (in fact, a directed tree<sup>3</sup>) rooted at  $B_{\text{genesis}}$ , with an edge  $B \leftarrow B'$  if  $B$  is the parent of  $B'$ , in which case we say  $B'$  is a *child* of  $B$ . We call these *parent-child* edges to differentiate from other types of edges.
- We say that a block is a *leaf* block if it has no children.
- A *chain* in such a protocol would then be a sequence of pairwise parent-child edges  $B_1 \leftarrow B_2 \leftarrow \dots$ . If there is a chain from  $B$  to  $B'$ , we say  $B'$  is a *descendant* of  $B$ . We say  $B$  is an *ancestor* of  $B'$  if and only if  $B'$  is a descendant of  $B$ . We say two blocks  $B$  and  $B'$  *conflict* if they do not equal and neither is a descendant of the other.
- The above setup implies each block  $B$  uniquely defines a (backwards) chain starting from  $B_{\text{genesis}}$  to  $B$ , and this must be the same chain in any view that includes  $B$  (thus we do not need to include a view as part of its definition). We call this *the chain of  $B$* , or  $\text{chain}(B)$ .

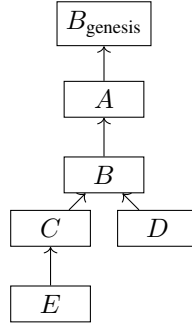


Figure 1: A graph visualization of a view (only visualizing blocks and not other messages) that occurs in the types of protocols we care about.

**Example 2.3.** See Figure 1 for an example of a view, limited to blocks. Arrows portray parent-child edges. The blocks in a view form a tree rooted at a single genesis block  $B_{\text{genesis}}$ . Here blocks  $C$  and  $D$  conflict (as well as  $E$  and  $D$ ).  $E$  has the longest chain, which is  $\text{chain}(E) = (B_{\text{genesis}}, A, B, C, E)$ .

There are some technicalities to our wording in this section that motivates our definition of *views* but are not very important to the core of this paper. To not distract the reader, we defer these details to Appendix A.

### 2.3 Proof-of-stake

The first and most influential approach to blockchain protocols is Bitcoin [15], using a proof-of-work model. This approach makes it computationally difficult to propose blocks, using the simple and intuitive fork-choice rule, “the block with the heaviest chain is the head of the chain.” Miners (the analogy of validators) simply propose blocks to build on the heaviest chain (the chain with the most amount of computational work), which does not mathematically guarantee any non-genesis block as “correct” but offers a probabilistic guarantee since it becomes less and less likely to conflict with a block once other miners build on top of it. Then the consensus history is simply the chain with the heaviest amount of work, which ideally keeps growing. The elegance of proof-of-work is paid for by the expensive cost of computational work / electricity / etc. to propose a block.

In this paper, we want to create a *proof-of-stake* blockchain, where a validator’s voting power is proportional to their bonded stake (or money) in the system. Instead of using computational power to propose blocks, proposing blocks is essentially free. In exchange, we need an additional layer of mathematical theory to prevent perverse incentives that arise when we make proposing blocks

<sup>3</sup>This statement already captures the primary reason we define both seeing and accepting; the blocks a validator *accepts* must be a single rooted tree, though the blocks a validator *sees* can be some arbitrary subset of the tree, so limiting our attention to the tree allows cleaner analysis and avoids pathological cases. Beyond this reason, the distinction between the two words is unimportant for the rest of this paper.

“easy.” We now shift our paper’s focus to a family of blockchain designs with proof-of-stake in mind.

To start, we assume we have a set of  $N$  validators  $\mathcal{V} = \{V_1, \dots, V_N\}$  and that each validator  $V \in \mathcal{V}$  has an amount of *stake*  $w(V)$ , a positive real number describing an amount of collateral. We make the further assumption that the average amount of stake for each validator is 1 unit, so the sum of the total stake is  $N$ . All of our operations involving stake will be linear, so this scaling does not change the situation and makes the bookkeeping easier. We assume validator sets and stake sizes are fixed for the sake of focusing on the theoretical essentials of the protocol. We address issues that occur when we relax these conditions in practical use in later sections, especially Section 8.

The main ingredients we need are:

- A *fork-choice rule*: a function  $\text{fork}()$  that, when given a view  $G$ , identifies a single leaf block  $B$ . This choice produces a unique chain  $\text{fork}(G) = \text{chain}(B)$  from  $B_{\text{genesis}}$  to  $B$  called the *canonical chain*. The block  $B$  is called the *head of the chain* in view  $G$ . Intuitively, a fork-choice rule gives a validator a “law” to follow to decide what the “right” block should be. For example, the *longest chain rule* is a fork-choice rule that returns the leaf block which is farthest from the genesis block. This rule is similar with the “heaviest chain rule” used by Bitcoin.
- A concept of *finality*: formally, a deterministic function  $F$  that, when given a view  $G$ , returns a set  $F(G)$  of *finalized* blocks. Intuitively, finalized blocks are “blocks that everyone will eventually think of as part of the consensus history” or “what the blockchain is sure of.”
- *Slashing conditions*: these are conditions that honest validators would never violate and violating validators can be provably caught, with the idea that violators’ stake would be *slashed*, or destroyed. Slashing conditions incentivize validators to follow the protocol (the protocol can reward honest validators for catching dishonest validators violating the conditions, for example).

The key concept we use to define these ingredients are *attestations*, which are votes (embedded in messages) for which blocks “should” be the head of the chain. To prevent validators from double-voting or voting in protocol-breaking ways, we enforce *slashing conditions* which can be used to destroy a validator’s stake, incentivizing validators to follow the protocol.

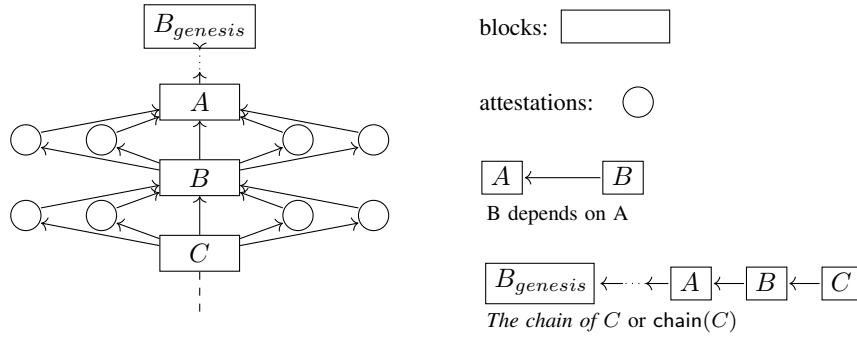


Figure 2: An ideal view with both blocks and attestations.

One can see an idealized version of the protocol we will present in Figure 2; in each time period, a new block is created with the parent block being the last head of chain (which is ideally just the last block created), and validators in the corresponding committee all perfectly attest to the new block. In less idealized situations, the blocks may fork, we may have missing attestations, etc. and the attestations will come to play by helping validators decide the correct head of the chain. In Section 4 we will define all the details.

## 2.4 Byzantine Validators, PBFT

In a protocol, we call a validator *honest* if he/she follows the protocol, and *byzantine* otherwise. We typically assume strictly less than  $p = \frac{1}{3}$  of validators are byzantine. This constant can be traced to Practical Byzantine Fault Tolerance (PBFT) from [8], a classic consensus protocol in byzantine tolerance literature; PBFT ensures that the system runs correctly as long as less than  $\frac{1}{3}$  of the *replicas* (synonymous with our *validators*) are byzantine, as proven in [3]. This constant  $\frac{1}{3}$  tolerance appears frequently in many works based on PBFT (the main idea is that a pigeonhole principle argument needs the  $1/3$  to work, such as in our proof in Section 5), with Casper FFG [7] being the most directly relevant to our work, which we review in Section 3.1.

Concretely, we define a blockchain running the protocol to be *p-slashable* if validators with a total of  $pN$  stake can be provably slashed by a validator with the network view, and most of our results will take the form “at any time, our blockchain either has (good) property X or is  $(1/3)$ -slashable,” because we really cannot guarantee having any good properties when we have many byzantine actors, so this is the strongest type of statement we can have. Also note that having  $(1/3)$ -slashable is a stronger and much more desirable result than just having  $(1/3)$  of the stake belonging to byzantine validators, because if we only had the latter guarantee, we may still incentivize otherwise honest validators to cheat in a way they do not get caught.

One of the most important aspects of PBFT is that it works under *asynchronous* conditions, which means we have no bounds on how long it may take for messages to be received. Thus, it is robust to the demands of blockchains and cryptocurrencies, where we often worry about adversarial contexts. We will discuss synchrony assumptions further in Section 2.6.

## 2.5 Safety and Liveness

The ultimate goal for the honest validators is to grow a finalized chain where all blocks form “logically consistent” state transitions with each other (despite having validators potentially go offline, suffer latency problems, or maliciously propose conflicting state changes). Formally, this translates to two desired properties:

**Definition 2.4.** We say a consensus protocol has:

- *safety*, if the set of finalized blocks  $F(G)$  for any view  $G$  can never contain two conflicting blocks. A consequence of having safety is that any validator view  $G$ ’s finalized blocks  $F(G)$  can be “completed” into a unique subchain of  $F(\text{view}(\text{NW}))$  that starts at the genesis block and ends at the last finalized block, which we call the *finalized chain*.
- *liveness*, if the set of finalized blocks can actually grow. There are different ways to define liveness. For our paper, we say the protocol has:
  - *plausible liveness*, if regardless of any previous events (attacks, latency, etc.) it is always possible for new blocks to be finalized (alternatively, it is impossible to become “deadlocked”). This is to prevent situations where honest validators cannot continue unless someone forfeits their own stake.
  - *probabilistic liveness*, if regardless of any previous events, it is probable for new blocks to be finalized (once we make some probabilistic assumptions about the network latency, capabilities of attackers, etc.).

On first glance, the latter implies the former, but the situation is more subtle; the former is a purely non-probabilistic property about the logic of the protocol; the latter requires (potentially very strong) assumptions about the context of the implementation to guarantee that the protocol “usually works as intended.” We discuss this contrast further in Section 7.3.

Our main goal in this paper is to prove these vital properties for our main protocol in Section 4.

**Remark** (Syntax versus Semantics). It should not be immediately obvious that chains have anything to do with logical consistence of state transitions, and this requires more work on the part of the protocol designer to ensure. For example, one can dictate a protocol where a block that allows a user to spend a coin  $S$  forbids an ancestor block from using  $S$  in another transaction. In this situation, if Xander obtained a coin  $S$  from block  $B_x$  and writes two blocks  $B_y$  and  $B_z$ , such that  $B_y$ ’s data includes Xander paying  $S$  to Yeezus and  $B_z$ ’s data includes Xander paying  $S$  to Zachariah, then

the logical idea that these actions are inconsistent corresponds to the graph-theoretical property that  $B_y$  and  $B_z$  conflict as blocks. The language of chains and conflicting blocks is enough to serve any such logic, so we assume in our paper that the syntax of what blocks can be children of what other blocks has already been designed to embed the semantics of logical consistency, which allows us to ignore logic and only think in terms of chains and conflicting blocks.

## 2.6 Time, Epochs, and Synchrony

In our model, we measure time in *slots*, defined as some constant number of seconds (tentatively 12 seconds in [11]). We then define an *epoch* to be some constant number  $C$  (tentatively 64 slots in [11]) of slots. The *epoch  $j$  of slot  $i$*  is  $\text{ep}(i) = j = \lfloor \frac{i}{C} \rfloor$ . In other words, the blocks belonging to epoch  $j$  have slot numbers  $jC + k$  as  $k$  runs through  $\{0, 1, \dots, C - 1\}$ . The genesis block  $B_{\text{genesis}}$  has slot number 0 and is the first block of epoch 0.

The main purpose of epochs is to divide time into pieces, the boundaries between which can be thought of as “checkpoints.” This allows concepts from Casper FFG to be used, as we will see in Section 3.1.

We should not assume the network is guaranteed to have the validators see the same messages at any given time, or even that different validators have the same view of time. The study of consensus protocols address this issue by having different *synchrony conditions*, such as:

- a *synchronous* system has explicit upper bounds for time needed to send messages between nodes;
- an *asynchronous* system has no guarantees; recall that PBFT [8] works under asynchrony.
- a *partially synchronous* system can mean one of two things depending on context: (i) explicit upper bounds for delays exist but are not known a priori; (ii) explicit upper bounds are known to exist after a certain unknown time  $T$ . The work [10] establishes bounds on fault tolerance in different fault and synchrony models, focusing on partial synchrony.

For us, we make no synchrony assumptions when studying safety and plausible liveness. When studying probabilistic liveness (i.e. trying to quantify liveness under “realistic” conditions), we will use the notion (ii) of partial synchrony above. We say that the network is  *$t$ -synchronous* at time  $T$  (where  $T$  and  $t$  are both in units of slots) if all messages with timestamps at or before time  $(T - t)$  are in the views of all validators at time  $T$  and afterwards; e.g., if each slot is 12 seconds, then  $(1/2)$ -synchrony means that all messages are received up to 6 seconds later.

**Remark.** It is possible to e.g., receive messages “from the future.” Suppose Alexis sends a message timestamped at 00:01:30 PT and Bob receives it 1 second later with a clock that is 3 seconds behind Alexis’s. Bob sees this 00:01:30 PT message on his own clock timestamped at 00:01:28 PT because he is a net total of 2 seconds behind. To make analysis easier, we can assume that all (honest) validators always delay the receiving of a message (i.e., do not add it into their view) until their own timestamps hit the message’s timestamp. This allows us to assume no messages are read (again, by honest validators) in earlier slots than they should.

## 3 Main Ingredients

### 3.1 Casper FFG

Buterin and Griffith introduce Casper the Friendly Finality Gadget (FFG) in [7]. This tool defines the concepts of *justification* and *finalization* inspired by practical Byzantine Fault Tolerance (PBFT) literature. Casper is designed to work with a wide class of blockchain protocols with tree-like structures.

- Every block has a *height* defined by their distance from the genesis block (which has height 0). Equivalently, the height of  $B$  is the length of  $\text{chain}(B) - 1$ .
- We define *checkpoint* blocks to be blocks whose height is a multiple of a constant  $H$  (in [7],  $H = 100$ ). We define the *checkpoint height*  $h(B)$  for a checkpoint block  $B$  as the height of  $B$  divided by  $H$ , which is always an integer. Thus, we can think of the subset of



checkpoints in the view as a subtree, containing only blocks whose heights are multiples of  $H$ .

- *Attestations* are signed messages containing “checkpoint edges”  $A \rightarrow B$ , where  $A$  and  $B$  are checkpoint blocks. We can think of each such attestation as a “vote” to move from block  $A$  to  $B$ . The choices of  $A$  and  $B$  depend on the underlying blockchain and is not a part of Casper. Each attestation has a *weight*, which is the stake of the validator writing the attestation. Ideally,  $h(B) = h(A) + 1$ , but this is not a requirement. For example, if  $H = 100$ , it is possible for an honest validator to somehow miss block 200, in which case their underlying blockchain may want them to send an attestation from checkpoint block 100 to checkpoint block 300.

**Remark.** Observe we setup time in units of *slots*, which are grouped into *epochs*. This is analogous to (but not exactly identical to) block heights and checkpoint blocks. We discuss this further in Section 4.1.

Casper FFG also introduces the concepts of *justification* and *finalization*, which are analogous to phase-based concepts in the PBFT literature such as *prepare* and *commit*, e.g., see [8]:

- In each view  $G$ , there is a set of *justified* checkpoint blocks  $J(G)$  and a subset  $F(G) \subset J(G)$  of *finalized* checkpoint blocks. The genesis block is always both justified and finalized.
- In a view  $G$ , a checkpoint block  $B$  is *justified* (by a justified checkpoint block  $A$ ) if  $A$  is justified and there are attestations voting for  $A \rightarrow B$  with total weight at least  $2/3$  of total validator stake. Equivalently, we say there is a *supermajority link*  $A \xrightarrow{J} B$ . This is a view-dependent condition, because the view in question may have or have not seen all the relevant attestations to break the  $2/3$  threshold (or even having seen the block  $B$  itself), which is why the set of justified blocks is parametrized by  $G$ .
- In a view  $G$ , if  $A \in J(G)$  (equivalently,  $A$  is justified), and  $A \xrightarrow{J} B$  is a supermajority link with  $h(B) = h(A) + 1$ , then we say  $A \in F(G)$  (equivalently,  $A$  is *finalized*).

Finally, Casper introduces *slashing conditions*, which are assumptions about honest validators. When they are broken by a validator  $V$ , a different validator  $W$  can *slash*  $V$  (destroy  $V$ ’s stake and possibly getting some sort of “slashing reward”) by offering proof  $V$  violated the conditions.

**Definition 3.1.** The following *slashing conditions*, when broken, cause a validator violating them to have their stake slashed:

- (S1) No validator makes two distinct attestations  $\alpha_1$  and  $\alpha_2$  corresponding to checkpoint edges  $s_1 \rightarrow t_1$  and  $s_2 \rightarrow t_2$  respectively with  $h(t_1) = h(t_2)$ .
- (S2) No validator makes two distinct attestations  $\alpha_1$  and  $\alpha_2$  corresponding respectively to checkpoint edges  $s_1 \rightarrow t_1$  and  $s_2 \rightarrow t_2$ , such that

$$h(s_1) < h(s_2) < h(t_2) < h(t_1).$$

As Casper is a finality gadget and not a complete protocol, it assumes the underlying protocol has its own fork-choice rule, and at every epoch all the validators run the fork-choice rule at some point to make one (and only one) attestation. It is assumed that honest validators following the underlying protocol will never be slashed. For example, if the protocol asks to make exactly one attestation per epoch, then honest validators will never violate (S1).

The main theorems in Casper are (slightly paraphrased):

**Theorem 3.2** (Accountable Safety). Two checkpoints on different branches cannot both be finalized, unless a set of validators owning stake above some total provably violated the protocol (and thus can be held accountable).

**Theorem 3.3** (Plausible Liveness). It is always possible for new checkpoints to become finalized, provided that new blocks can be created by the underlying blockchain.

**Remark.** Palmskog et al. [17] provided a mechanical proof assistant for Casper FFG’s accountable safety and plausible liveness in the Coq Proof Assistant. The authors modified the blockchain model in Coq, Toychain, to use for the Casper FFG with Ethereum’s beacon chain specs in order to see how Casper FFG’s proofs will work in practice.

### 3.2 LMD GHOST Fork-Choice rule

The Greediest Heaviest Observed SubTree rule (GHOST) is a fork-choice rule introduced by Sompolinsky and Zohar [20]. Intuitively, GHOST is a greedy algorithm that grows the blockchain on sub-branches with the “most activity.” Zamfir [22], in looking for a “correct-by-construction” consensus protocol, introduces a natural adaptation of GHOST that happens to be well-suited for our setup<sup>4</sup>. We call this variant *Latest Message Driven Greediest Heaviest Observed SubTree (LMD GHOST)*, which we can define after having a notion of *weight*:

**Definition 3.4.** Given a view  $G$ , Let  $M$  be the set of latest attestations, one per validator. The *weight*  $w(G, B, M)$  is defined to be the sum of the stake of the validators  $i$  whose last attestation in  $M$  is to  $B$  or descendants of  $B$ .

---

**Algorithm 3.1** LMD GHOST Fork Choice Rule.

---

```

1: procedure LMD-GHOST( $G$ )
2:    $B \leftarrow B_{\text{genesis}}$ 
3:    $M \leftarrow$  the most recent attestations of the validators (one per validator)
4:   while  $B$  is not a leaf block in  $G$  do
5:      $B \leftarrow \arg \max_{B' \text{ child of } B} w(G, B', M)$ 
6:     (ties are broken by hash of the block header)
7:   return  $B$ 

```

---

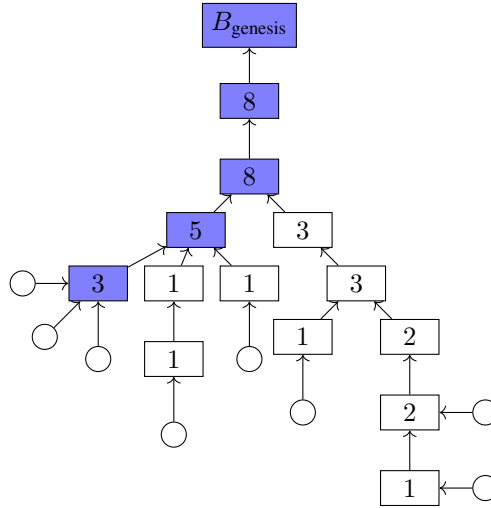


Figure 3: An example of the LMD-GHOST fork-choice rule. The number in each block  $B$  is the weight (by stake), with all attestations (circles) having weight 1 in our example. A validator using this view will conclude the blue chain to be the canonical chain, and output the latest blue block on the left, with weight 3, to be the head of the chain.

The idea of LMD GHOST is that at any fork, we use the weights of the subtrees created by the fork as a heuristic and assume the subtree with the heaviest weight is the “right” one, as evident from the name of the algorithm. We will always end up at a leaf block, which defines a canonical chain. See Figure 3 for a graphical visualization. In [22], LMD GHOST was tied to a particular protocol. Our presentation of it in this paper treats the specifics of attestations as details to be filled in by the protocol, like how Casper does not specify details on how blocks are constructed. We merely need that the concepts of *blocks* and *attestations* exist, and then LMD GHOST defines an algorithm to find a block given a view. In Section 4 we will modify the rule to satisfy our other purposes.

<sup>4</sup>This part of the field can create phantasmal confusion: the name of [22] is “Casper the Friendly GHOST,” which introduces Casper CBC (also see [5]), a completely separate protocol from Casper FFG. Our paper is in some sense recombining these two ideas, but the “Casper” in [22] is very different from Casper FFG.

## 4 Main Protocol: Gasper

We now define our main protocol Gasper, which is a combination of the GHOST and Casper FFG ideas. The main concepts are:

- (*Epoch boundary*) *pairs* of a chain: given a chain, certain blocks are picked out, ideally one per epoch, to play the role of Casper’s *checkpoints*. However, a block may appear more than once as a checkpoint on the same chain (this is a nuance not found in Casper; we expound on this in Section 4.1), so we use ordered pairs  $(B, j)$ , where  $B$  is a block and  $j$  is an epoch, to disambiguate. These will be called *epoch boundary pairs*, or *pairs* for short.
- *Committees*: the validators are partitioned into *committees* in each epoch, with one committee per slot. In each slot, one validator from the designated committee proposes a block. Then, all the members of that committee will attest to what they see as the head of the chain (which is hopefully the block just proposed) with the fork-choice rule HLMD GHOST (a slight variation of LMD GHOST).
- *Justification and Finalization*: these concepts are virtually identical to that of Casper FFG, except we justify and finalize pairs instead of justifying and finalizing checkpoint blocks, i.e., given a view  $G$ ,  $J(G)$  and  $F(G)$  are sets of pairs.

### 4.1 Epoch Boundary Blocks and Pairs

Recall any particular block  $B$  uniquely determines a chain,  $\text{chain}(B)$ . For a block  $B$  and an epoch  $j$ , define  $\text{EBB}(B, j)$ , the  $j$ -th *epoch boundary block* of  $B$ , to be the block with the highest slot less than or equal to  $jC$  in  $\text{chain}(B)$ . Let the latest such block be  $\text{LEBB}(B)$ , or the *last epoch boundary block* (of  $B$ ). We make a few observations:

- For every block  $B$ ,  $\text{EBB}(B, 0) = B_{\text{genesis}}$ .
- More generally, if  $\text{slot}(B) = jC$  for some epoch  $j$ ,  $B$  will be an epoch boundary block in every chain that includes it.
- However, without such assumptions, a block could be an epoch boundary block in some chains but not others.

To disambiguate situations like these, we add precision by introducing *epoch boundary pairs* (or *pairs* for short)  $(B, j)$ , where  $B$  is a block and  $j$  is an epoch. Our main concepts of justification and finalization will be done on these pairs. Given a pair  $P = (B, j)$ , we say  $P$  has *attestation epoch*  $j$ , using the notation  $\text{aep}(P) = j$ , which is **not** necessarily the same as  $\text{ep}(B)$ .

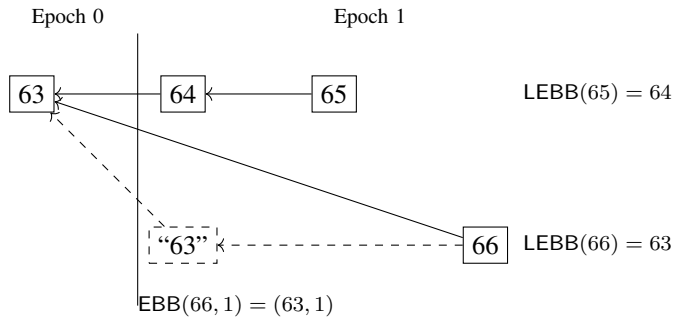


Figure 4: Example of epoch boundary blocks and pairs. Blocks are labeled with their slot numbers. “63” is not an actual block but illustrates the perspective of 66 needing an epoch boundary block at slot 64 and failing to find one.

**Example 4.1.** Suppose blocks (labeled by slots) form a chain  $63 \leftarrow 64 \leftarrow 65$  and there is a fork  $63 \leftarrow 66$ . Then  $\text{EBB}(65, 1) = 64$  since 64 is in  $\text{chain}(65)$ . When we try to find  $\text{EBB}(66, 1)$ , we look for block 64 because we are in epoch 1 and realize it is not in  $\text{chain}(66)$ , so we have to look backwards and “pull up” block 63 from the previous epoch to serve as the epoch boundary block in

this chain. This is visualized by the dashed components in Figure 4. Observe even though  $\text{aep}(63, 1)$  and  $\text{ep}(63) = 0$ ,  $\text{ep}()$  is a local property that only depends on the block’s slot, and epoch boundary concepts like  $\text{aep}()$  depend on the context of the chain.

**Remark.** We further address why we use pairs instead of checkpoint blocks. Casper FFG is a “finality gadget,” meaning it is designed to place a layer of finality on top of a blockchain which has probabilistic liveness, which gives a steady new supply of checkpoint blocks. Probabilistic liveness assumes good synchrony conditions. For the safety part of the design, we would want to make fewer assumptions and take into account worst-case scenarios where we may be put into a state where we have not seen a new block for a while.

As an example, block  $B$  may be the head of the chain from epoch 1, but be currently at epoch 3 with no new block built on top of  $B$ . In the original Casper FFG, we expect probabilistic liveness so we would have a different checkpoint at every epoch. However, when we have no synchrony assumptions, in the analysis we would need to differentiate the idea of “checkpoint in epoch 2” and “checkpoint in epoch 3” even though the best candidate for both is  $B$ , which is itself still only in epoch 1! This naturally induces the concepts  $(B, 2)$  and  $(B, 3)$ , which represents the best approximation to a checkpoint in a later epoch if the checkpoint block that “should have been there” is missing.

Also, Casper FFG makes no assumptions about time in the underlying blockchain – only block heights are important, with no notion of *slots* or *epochs*. The use of block height is a natural choice in proof-of-work blockchains due to the Poisson process of mining new blocks serving somewhat as a system clock, with blocks coming at fairly irregular times. Gasper, as a proof-of-stake protocol, can have blocks coming in at controlled regular intervals as part of the protocol (instead of depending on a random process), so instead the notion of time is explicitly desired in the protocol. To capture this notion of time, each object in our blockchain then naturally requires the knowledge of both the data (captured in a block) and the time (captured in the epoch count), which also leads naturally to the idea of pairs.

## 4.2 Committees

The point of committees is to split up the responsibilities among the validators. To start, assume the validators have access to a sequence of random length- $N$  permutations  $\rho_0, \rho_1, \dots$ , as functions  $\{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, N\}$ . In the scope of this article, we assume we get these random permutations from a random oracle<sup>5</sup>.

Recall time is split into epochs, of  $C$  slots each. Each permutation  $\rho_j$  will be used only during epoch  $j$ . Its role is to pseudorandomly select validators into  $C$  *committees*, each of whom has responsibilities for one slot of the epoch. To be precise:

- During epoch  $j$ , we would like to split the set of validators  $V$  into  $C$  equal-size committees  $S_0, S_1, \dots, S_{C-1}$  (we assume  $C|N$  for easier notation; dealing with “roughly equal” size committees does not change the essence of our approach).
- Therefore, for each  $k \in \{0, 1, \dots, C-1\}$ , we define  $S_k$  to be the set of the  $N/C$  validators of the form  $V_{\rho_j(s)}$ , where  $s \equiv k \pmod{C}$ . Note, for an epoch  $j$ , the sets  $S_0, S_1, \dots, S_{C-1}$  partition the entire set of validators  $\{V_1, \dots, V_N\}$  across all the slots of epoch  $j$ , as desired.

To summarize, in each epoch  $j$ ,  $\rho_j$  allows us to shuffle the validators into  $C$  committees. Our work does not assume more than this intuition.

## 4.3 Blocks and Attestations

Now, in each slot, the protocol dictates 2 types of “committee work” for the committee assigned to their respective slot: one person in the committee needs to *propose* a new block, and everyone in the committee needs to *attest* to their head of the chain. Both proposing blocks and attesting mean immediately adding a corresponding message (a block and an attestation, respectively) to the validator’s own view and then broadcasting the message to the network. Recall the messages proposing (non-genesis) blocks and publishing attestations have digital signatures.

<sup>5</sup>Achieving non-exploitable randomness on the blockchain is itself an interesting problem, stimulating research such as verifiable delay functions; however, in this article we will take this randomness for granted.

Both proposing and attesting requires the committee member to run the same fork-choice rule  $\text{HLMD}()$  on the validator's own view, which is a variation of the LMD GHOST fork-choice rule.  $\text{HLMD}()$ 's definition requires concepts that we have not introduced yet, so we delay its definition to Section 4.6. For now, the only thing we need to know is, like LMD GHOST,  $\text{HLMD}()$  takes a view  $G$  and returns a leaf block  $B$  as the head of the chain, making  $\text{chain}(B)$  the canonical chain of  $G$ .

We now present the responsibilities of the protocol during slot  $i = jC + k$ , where  $k \in \{0, 1, \dots, C - 1\}$ . All mentions of time are computed from the point-of-view of the validator's local clock, which we assume to be synced within some delta.

**Definition 4.2.** At the beginning of slot  $i = jC + k$ , designate validator  $V = V_{\rho_j(k)}$ , i.e., the first member of the committee  $S_k$  of epoch  $j$ , as the *proposer* for that slot. The proposer computes the canonical head of the chain in his/her view, i.e.,  $\text{HLMD}(\text{view}(V, i)) = B'$ , then *proposes a block*  $B$ , which is a message containing:

1.  $\text{slot}(B) = i$ , the slot number.
2.  $P(B) = B'$ , a pointer to the parent block; in other words, we always build a block on top of the head of the chain.
3.  $\text{newattests}(B)$ , a set of pointers to all the *attestations* (to be defined next)  $V$  has accepted, but have not been included in any  $\text{newattests}(B')$  for a block  $B'$  who is an ancestor of  $B$ .
4. Some implementation-specific data (for example, “Yunice paid 4.2 ETH to Brad” if we are tracking coins), the semantics of which is irrelevant for us.

For dependencies,  $B$  depends on  $P(B)$  and all attestations in  $\text{newattests}(B)$ . (so for example, if we see a block on the network but do not see its parent, we ignore the block until we see its parent).

In typical (honest) behavior, we can assume each slot number that has a block associated with it also has at most one such block in  $\text{view}(\text{NW})$ . By default,  $B_{\text{genesis}}$  is the unique block with slot 0. A dishonest validator can theoretically create a block with a duplicate slot number as an existing block, but we can suppose the digital signatures and the pseudorandom generator for block proposal selection are set up so such behavior would be verifiably caught. For example, every time a digitally signed block enters the view of a validator, the validator can immediately check if the block proposer is the unique person allowed to propose a block for that respective slot. A validator can also prove the same person proposed two blocks in the same slot by just pointing to both blocks.

**Definition 4.3.** At time  $(i + 1/2)$ , the middle of slot  $i = jC + k$ , each validator  $V$  in committee  $S_k$  computes  $B' = \text{HLMD}(\text{view}(V, i + 1/2))$ , and publishes an *attestation*  $\alpha$ , which is a message containing:

1.  $\text{slot}(\alpha) = jC + k$ , the slot in which the validator is making the attestation. We will also use  $\text{ep}(\alpha)$  as shorthand for  $\text{ep}(\text{slot}(\alpha))$ .
2.  $\text{block}(\alpha) = B'$ . We say  $\alpha$  *attests to*  $\text{block}(\alpha)$ . We will have  $\text{slot}(\text{block}(\alpha)) \leq \text{slot}(\alpha)$ , and “usually” get equality by quickly attesting to the block who was just proposed in the slot.
3. A *checkpoint edge*  $\text{LJ}(\alpha) \xrightarrow{V} \text{LE}(\alpha)$ . Here,  $\text{LJ}(\alpha)$  and  $\text{LE}(\alpha)$  are epoch boundary pairs in  $\text{view}(V, i + 1/2)$ . We define them properly in Section 4.4.

For dependencies,  $\alpha$  depends on  $\text{block}(\alpha)$ . So we ignore an attestation until the block it is attesting to is accepted into our view (this is one of the bigger differences between theory and implementation; we discuss this more in in Section 8).

Intuitively,  $\alpha$  is doing two things at once: it is simultaneously a “GHOST vote” for its block and also a “Casper FFG vote” for the transition between the two epoch boundary pairs (akin to Casper’s checkpoint blocks).

#### 4.4 Justification

**Definition 4.4.** Given a block  $B$ , we define<sup>6</sup>  $\text{view}(B)$ , the *view of  $B$* , to be the view consisting of  $B$  and all its ancestors in the dependency graph. We define  $\text{ffgview}(B)$ , the *FFG view of  $B$* , to be  $\text{view}(\text{LEBB}(B))$ .

The definition  $\text{view}(B)$  is “agnostic of the viewer” in which any view that accepted  $B$  can compute an identical  $\text{view}(B)$ , so we do not need to supply a validator (or NW) into the argument. Intuitively,  $\text{view}(B)$  “focuses” the view to  $\text{chain}(B)$  and  $\text{ffgview}(B)$  looks at a “frozen” snapshot of  $\text{view}(B)$  at the last checkpoint. Casper FFG operates only on epoch boundary pairs, so the FFG view of a block  $B$  extracts exactly the information in  $\text{chain}(B)$  that is relevant to Casper FFG.

We now define the main concepts of justification. To start, recall an attestation  $\alpha$  for Gasper contains a *checkpoint edge*  $\text{LJ}(\alpha) \xrightarrow{\text{V}} \text{LE}(\alpha)$ , acting as a “FFG vote” between two epoch boundary pairs. After a couple of new definitions, we can finally explicitly define these notions.

**Definition 4.5.** We say there is a *supermajority link* from pair  $(A, j')$  to pair  $(B, j)$  if the attestations with checkpoint edge  $(A, j') \xrightarrow{\text{V}} (B, j)$  have total weight more than  $\frac{2}{3}$  of the total validator stake. In this case, we write  $(A, j') \xrightarrow{\text{J}} (B, j)$ .

**Definition 4.6.** Given a view  $G$ , we define the set  $J(G)$  of *justified pairs* as follows:

- $(B_{\text{genesis}}, 0) \in J(G)$ ;
- if  $(A, j') \in J(G)$  and  $(A, j') \xrightarrow{\text{J}} (B, j)$ , then  $(B, j) \in J(G)$  as well.

If  $(B, j) \in J(G)$ , we say  $B$  is *justified in  $G$  during epoch  $j$* .

**Definition 4.7.** Given an attestation  $\alpha$ , let  $B = \text{LEBB}(\text{block}(\alpha))$ . We define:

1.  $\text{LJ}(\alpha)$ , the *last justified pair of  $\alpha$* : the highest attestation epoch (or last) justified pair in  $\text{ffgview}(\text{block}(\alpha)) = \text{view}(B)$ .
2.  $\text{LE}(\alpha)$ , the *last epoch boundary pair of  $\alpha$* :  $(B, \text{ep}(\text{slot}(\alpha)))$ .

While we define  $J(G)$  for any view  $G$ , we are typically only interested in  $J(G)$  for  $G = \text{ffgview}(B)$ . We can think of each chain as having its own “state” of justified blocks/pairs that is updated only at the epoch boundaries.

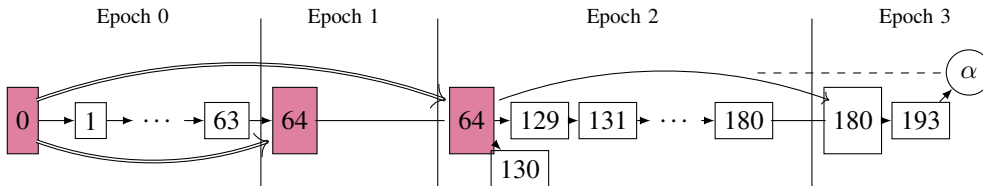


Figure 5: A validator’s view  $G$  as she writes an attestation in epoch 3. During epoch 1, latency issues make her not see any blocks, so block 64 is both  $\text{EBB}(193, 1)$  and  $\text{EBB}(193, 2)$ . She ends up writing an  $\alpha$  with a “GHOST vote” for  $\text{block}(\alpha) = 193$  and a “FFG vote” checkpoint edge (single arc edge)  $(64, 2) \xrightarrow{\text{V}} (180, 3)$  for  $\alpha$ . We use slot numbers for blocks, so block 0 is just  $B_{\text{genesis}}$ . Blocks in red are justified (in  $G$ ). Double edges corresponding to supermajority links.

**Example 4.8.** A validator runs  $\text{HLMD}(G)$  on her view  $G$  to obtain her head of the chain, which is block 193. She is then supposed to attest to 193 with an attestation  $\alpha$ .

On  $\text{chain}(193)$  portrayed in Figure 5,  $\text{LE}(\alpha) = (180, 3)$ , even though  $\text{ep}(180) = 2$ , because our attestation has epoch 3, and we were looking for  $3 * 64 = 192$  but did not see it (one can imagine,

<sup>6</sup>We have tyrannically overworked the notation  $\text{view}()$  by this point, but there should be no ambiguity when we know the type of its parameter.

like in our figure, that we “pull up” block 180 to show that it is  $\text{EBB}(193, 3)$ ). In  $\text{ffgview}(193) = \text{view}(180)$ , the last justified (by epoch number, not slot) pair is  $(64, 2)$ , so  $\text{LJ}(\alpha) = (64, 2)$ .

Note that this relies on  $\text{view}(180)$  to include attestations worth at least  $(2N/3)$ -stake pointing to  $(64, 2)$ . It would be possible to have a case where this did not happen by 180, but the chain does see  $(2N/3)$  worth of attestations by 193. In this case,  $(64, 2)$  is in  $J(G)$ , but  $(64, 2)$  is not in  $J(\text{ffgview}(193))$ , so the resulting checkpoint edge would be, e.g.,  $(64, 1) \xrightarrow{V} (180, 3)$  instead, assuming the attestations for  $(64, 1)$  are in  $\text{view}(180)$ . The checkpoint edge could even end up being  $(0, 0) \xrightarrow{V} (180, 3)$ , if 180 did not see those attestations; note that this is very possible if the attestations for  $(64, 1)$  were only included in the blocks that forked off of  $\text{chain}(193)$  (such as block 130), as the forking may be an indication of network problems causing those attestations to be temporarily unavailable to validators on  $\text{chain}(193)$ .

Everything here is analogous to Casper FFG. Inside the chain of  $\text{block}(\alpha)$  is a sub-chain created by the epoch boundary blocks of that chain, starting from  $B_{\text{genesis}}$  and ending at  $B = \text{LEBB}(\alpha)$ . We want to focus on this subchain of blocks (represented by pairs to allow for boundary cases) and justify epoch boundary pairs with many attestations; this sub-chain is exactly what’s captured by  $\text{ffgview}(\text{block}(\alpha))$ .  $\alpha$  is a vote to transition from some last justified pair  $(B', j')$  to the new pair  $(B, j)$ , visualized as the checkpoint edge  $(A, j') \xrightarrow{V} (B, j)$ . If  $(2N/3)$  stake worth of such votes happen, we create a supermajority link  $(A, j') \xrightarrow{J} (B, j)$  and justify  $(B, j)$ .

#### 4.5 Finalization

Given our notion of *justification* and a new fork-choice rule, we are now ready to define the notion of *finalization*. Finalization is a stronger notion of justification in the sense that the moment **any** view considers a block  $B$  as finalized for some  $j$ , **no** view will finalize a block conflicting with  $B$  unless the blockchain is  $(1/3)$ -slashable.

**Definition 4.9.** For a view  $G$ , we say  $(B_0, j)$  is *finalized* (specifically, *k-finalized*) in  $G$  if  $(B_0, j) = (B_{\text{genesis}}, 0)$  or if there is an integer  $k \geq 1$  and blocks  $B_1, \dots, B_k \in \text{view}(G)$  such that the following holds:

- $(B_0, j), (B_1, j+1), \dots, (B_k, j+k)$  are adjacent epoch boundary pairs in  $\text{chain}(B_k)$ ;
- $(B_0, j), (B_1, j+1), \dots, (B_{k-1}, j+k-1)$  are all in  $J(G)$ ;
- $(B_0, j) \xrightarrow{J} (B_k, j+k)$ .

We define  $F(G)$  to be the set of finalized pairs in the view  $G$ ; we also say that a block  $B$  is *finalized* if  $(B, j) \in F(G)$  for some epoch  $j$ .

We expect 1-finalized blocks for the vast majority of time. Specifically, this just means we have blocks  $B_0$  and  $B_1$  such that  $(B_0, j)$  is justified in  $G$  and  $(B_0, j) \xrightarrow{J} (B_1, j+1)$ , or “a justified epoch boundary block justifies the next adjacent epoch boundary block.” This is essentially the original definition of *finalized* in [7]. With  $(1/2)$ -synchrony and without implementation details that interfere with accepting messages, we should only get 1-finalization. We include the  $k > 1$  cases to account for situations where network latency and attestation inclusion delays (see Section 8.3) are relevant.

**Remark.** In Figure 6, we see some examples of finalization. On the left, we have 1-finalization, which is what we expect to happen in the vast majority of the time. At the center, we have an example of 2-finalization to account for attestation delays. The  $k = 3$  example is mainly for the sake of illustration; it takes some contrived orchestration to create. In practice, the planned implementation for finalization in [11] does not even include finalization for  $k \geq 3$ . See Section 8.5.

If we define a set of blocks  $F$  as finalized and prove safety for them, then safety remains true automatically if we change the definition of *finalized* to any subset of  $F$ , because safety is defined by the lack of incomparable pairs of finalized blocks. Thus, it does not hurt to define a very general class of blocks as finalized if we can prove safety. This is why we include all  $k$  in our definition; the more general definition is easier to analyze.

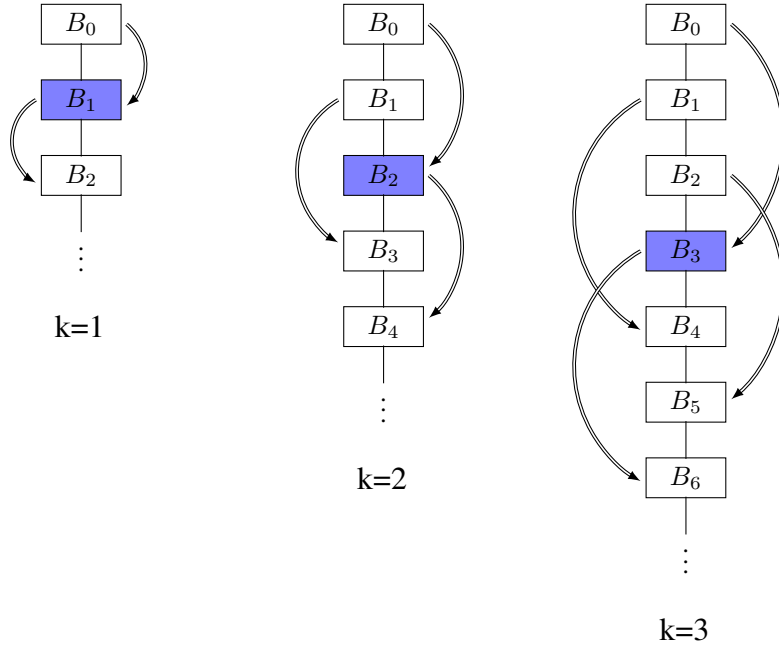


Figure 6: Illustrative examples of Definition 4.9 for  $k = 1, 2, 3$ , respectively. Double arrows are supermajority links. The blue block is the one being finalized in each case.

#### 4.6 Hybrid LMD GHOST

We now have enough notation to define HLMD. Since the final design is fairly complicated at first glance, we first present Algorithm 4.1, a “prototype” version; simply put, it starts with the last justified block in the view and then runs LMD GHOST. It is not obvious  $(B_J, j)$  should be unique, but it can be proven to be well-defined as a consequence of Lemma 4.11 when the blockchain is not  $(1/3)$ -slashable. For implementation it is costless to add a tiebreaker (say via block header hash).

---

##### Algorithm 4.1 Prototype Hybrid LMD GHOST Fork Choice Rule

---

```

1: procedure HLMD( $G$ )
2:    $(B_J, j) \leftarrow$  the justified pair with highest attestation epoch  $j$  in  $G$ 
3:    $B \leftarrow B_J$ 
4:    $M \leftarrow$  most recent attestations (one per validator)
5:   while  $B$  is not a leaf block in  $G$  do
6:      $B \leftarrow \arg \max_{B' \text{ child of } B} w(G, B', M)$ 
7:     (ties are broken by hash of the block header)
8:   return  $B$ 

```

---

However, using this prototype will run into some subtle problems. Because the “FFG part” of an attestation is a vote between epoch boundary pairs, and  $\text{LJ}(B)$  depends only on  $\text{ffgview}(B)$ , we can think of the “last justified pair in a chain” as “frozen” at the beginning of every epoch, which does not mix well with the “non-frozen”  $(B_J, j)$  being instinctively defined as the last justified block in our entire view, as the latter definition can shift during an epoch (the reader is encouraged to think of problems with this, as an exercise). Thus, we need a version of  $(B_J, j)$  that does not change in the middle of an epoch.



Another problem occurs when the algorithm forks: forked blocks can have drastically different last justified pairs, even if they are next to each other. As a result of this mismatch, we can create pathological situations where an honest validator may slash themselves when following Algorithm 4.1, if a validator previously attested to a higher last justification epoch but then forks into a chain whose last justification epoch is older (c.f. Definition 4.10, S2).

To resolve these issues, first we “freeze” the state of the latest justified pair  $(B_J, j)$  to the beginning of the epochs; formally, this means when defining  $(B_J, j)$  we consider the views of  $\text{ffgview}(B_l)$  over the leaf blocks  $B_l$  as opposed to the entire view. Then, we filter the branches so we do not go down branches with leaf nodes  $B_l$  where  $\text{LJ}(B_l)$  has not “caught up” to  $(B_J, j)$ ; formally, we create an auxiliary view  $G'$  from  $G$  that eliminates the problematic branches. These fixes give Algorithm 4.2, the final design.

---

**Algorithm 4.2** Hybrid LMD GHOST Fork Choice Rule

---

```

1: procedure HLMD( $G$ )
2:    $L \leftarrow$  set of leaf blocks  $B_l$  in  $G$ 
3:    $(B_J, j) \leftarrow$  the justified pair with highest attestation epoch  $j$  in
        $J(\text{ffgview}(B_l))$  over  $B_l \in L$ 
4:    $L' \leftarrow$  set of leaf blocks  $B_l$  in  $G$  such that  $(B_J, j) \in J(\text{ffgview}(B_l))$ 
5:    $G' \leftarrow$  the union of all chains  $\text{chain}(B_l)$  over  $B_l \in L'$ 
6:    $B \leftarrow B_J$ 
7:    $M \leftarrow$  most recent attestations (one per validator)
8:   while  $B$  is not a leaf block in  $G'$  do
9:      $B \leftarrow \arg \max_{B' \text{ child of } B} w(G', B', M)$ 
10:    (ties are broken by hash of the block header)
11:  return  $B$ 

```

---

A dual (and possibly simpler) way of understanding Algorithm 4.2 is from the perspective of a state-based implementation of the protocol. We can think of each chain of a leaf block  $B_l$  as storing the state of its own last justified pair. During an epoch, new attestations to blocks in the chain updates the GHOST-relevant list of latest attestations  $M$  but **not** the FFG-relevant justification and finalization information of the chain until the next epoch boundary block. This way, the “FFG part” of the protocol always works with the “frozen until next epoch” information, while the “GHOST part” of the protocol is being updated continuously with the attestations. This careful separation allows us to avoid pathological problems from mixing the two protocols.

## 4.7 Slashing Conditions

In this subsection, we add *slashing* conditions, analogous to those from Casper in Definition 3.1. We prove a couple of desired properties, after which we are ready to use these conditions to prove safety of Gasper in Section 5.

**Definition 4.10.** We define the following *slashing conditions*:

- (S1) No validator makes two distinct attestations  $\alpha_1, \alpha_2$  with  $\text{ep}(\alpha_1) = \text{ep}(\alpha_2)$ . Note this condition is equivalent to  $\text{aep}(\text{LE}(\alpha_1)) = \text{aep}(\text{LE}(\alpha_2))$ .
- (S2) No validator makes two distinct attestations  $\alpha_1, \alpha_2$  with
 
$$\text{aep}(\text{LJ}(\alpha_1)) < \text{aep}(\text{LJ}(\alpha_2)) < \text{aep}(\text{LE}(\alpha_2)) < \text{aep}(\text{LE}(\alpha_1)).$$

We now prove a very useful property of the protocol: unless we are in the unlikely situation that we have enough evidence to slash validators with at least  $1/3$  of the total stake, in a view  $G$  we may assume all elements in  $J(G)$  have unique attestation epochs (in other words, the view sees at most one pair justified per epoch).

**Lemma 4.11.** In a view  $G$ , for every epoch  $j$ , there is at most 1 pair  $(B, j)$  in  $J(G)$ , or the blockchain is  $(1/3)$ -slashable. In particular, the latter case means there must exist 2 subsets  $\mathcal{V}_1, \mathcal{V}_2$  of  $\mathcal{V}$ , each with total weight at least  $2N/3$ , such that their intersection violates slashing condition (S1).

*Proof.* Suppose we have 2 distinct pairs  $(B, j)$  and  $(B', j)$  in  $J(G)$ . This means in epoch  $j$ , more than a total stake of  $2N/3$  attested with a checkpoint edge to  $(B, j)$  and more than  $2N/3$  stake attested with a checkpoint edge to  $(B', j)$ . These are our desired  $\mathcal{V}_1$  and  $\mathcal{V}_2$ .  $\square$

**Remark.** The importance of Lemma 4.11 is that assuming reasonable conditions (we are not  $(1/3)$ -slashable), the choice of  $(B_j, j)$  in Algorithm 4.2 is unique and thus well-defined. More generally, we can assume any view only sees at most one justified pair  $(B, j)$  per epoch  $j$  over **all** chains.

**Proposition 4.12.** An honest validator following the protocol will never accidentally violate the slashing conditions.

*Proof.* First, notice an honest validator cannot violate (S1) because each validator is assigned to exactly one committee in each epoch, and is thus asked to attest exactly once per epoch. Thus, it suffices to consider (S2).

Suppose an honest validator  $V$  is about to write an attestation that violates (S2) in epoch  $u$ . This means there are epochs  $r < s < t < u$  where in epoch  $t$ ,  $V$  wrote an attestation  $\alpha_t$  with checkpoint edge  $(B_2, s) \xrightarrow{V} (B_3, t)$  and the protocol is telling  $V$  now to write an attestation  $\alpha_u$  with checkpoint edge  $(B_1, r) \xrightarrow{V} (B_4, u)$ . This means running HLMD GHOST on  $V$ 's current view ends up at some leaf block  $B$ , which must be a descendent of  $B_4$  such that  $\text{LE}(B) = (B_4, u)$ .

Because  $V$  wrote  $\alpha_t$  in epoch  $t < u$ , we know  $(B_2, s)$  was a justified pair in  $\text{ffgview}()$  of some leaf block at that point in epoch  $t$ . As justified blocks remain justified when their chains grow, we know that now in epoch  $u$ , the starting justified pair  $(B_J, j)$  as defined in HLMD GHOST (Algorithm 4.2) must satisfy  $j \geq s$ . Since  $s > r$ , we know that  $j > r$ .

However, because of the filtering  $L'$  in Algorithm 4.2, we know that the resulting leaf block  $B$  of running HLMD GHOST (in epoch  $u$ ) starting at  $(B_J, j)$  must satisfy  $(B_J, j) \in J(\text{ffgview}(B))$ . Since  $\text{LJ}(B) \in J(\text{ffgview}(B))$  by definition and  $\text{LJ}(B) = (B_1, r)$ , we know  $j \leq r$ , a contradiction. Therefore, the protocol cannot force an honest validator to violate (S2).  $\square$

## 4.8 Rewards and Penalties

A validator should be rewarded (i.e. have stake increased) for either including valid attestations in his/her proposed block (in the beacon chain specs [11], *proposer reward*) or attesting to the correct block which is justified and finalized as the chain grows (in [11], *attester reward*). Meanwhile, a validator should be penalized (have stake decreased) for violating slashing conditions.

The reward and penalty amounts can be adjusted based on the security level that needs to be achieved, and the game theory of the situation should be such that validators are incentivized to perform their tasks and to not violate slashing conditions. Additionally, it is worth considering more complex incentives, such as incentivizing validators to catch other misbehaving validators. While this makes for potentially interesting game theory work, adding a layer of analysis of these parameters is distracting for the scope of this paper, where we want to focus on the consensus aspect of Gasper. To be pragmatic, we abstract this analysis away by assuming these reward and penalty mechanisms provide enough game-theoretical incentives such that the following holds:

- Honest validators follow the protocol.
- Any honest validator seeing a slashing condition violated will slash the dishonest validator.

## 5 Safety

Recall our main goals are proving safety and liveness. In this section, we prove safety, which is done similarly as in Casper FFG [7]; the main differences are that we are using epoch boundary pairs (as opposed to Casper's checkpoint blocks) and our finalization definition is more complex. Thus, we need the following Lemma, which is necessarily more complicated than its counterpart idea in Casper.

**Lemma 5.1.** In a view  $G$ , if  $(B_F, f) \in F(G)$  and  $(B_J, j) \in J(G)$  with  $j > f$ , then  $B_F$  must be an ancestor of  $B_J$ , or the blockchain is  $(1/3)$ -slashable – specifically, there must exist 2 subsets  $\mathcal{V}_1, \mathcal{V}_2$  of  $\mathcal{V}$ , each with total stake at least  $2N/3$ , such that their intersection all violate slashing condition (S1) or all violate slashing condition (S2).

*Proof.* Anticipating contradiction, suppose there is a pair  $(B_J, j)$  with  $j > f$  and  $B_J$  is not a descendant of  $B_F$ . By definition of finalization, in  $G$ , we must have  $(B_F, f) \xrightarrow{j} (B_k, f + k)$ , where we have a sequence of adjacent epoch boundary pairs  $(B_F, f), (B_1, f + 1), \dots, (B_k, f + k)$ .

Since  $(B_J, j)$  is justified and  $B_J$  is not a descendant of  $B_F$ , without loss of generality (by going backwards with supermajority links), we can assume  $(B_J, j)$  is the earliest such violation, meaning we can assume  $(B_l, l) \xrightarrow{j} (B, j)$  where  $l < f$  but  $j > f$  (here we are using Lemma 4.11, which tells us no two justified blocks have the same  $\text{aep}()$ , else we are done already with 2 validator subsets of weight  $2N/3$  each violating (S1); this is why we do not worry about the equality case). Since  $B_1, \dots, B_k$  are all justified but are descendants of  $B_F$ , we know  $B_J$  cannot be any of these blocks, so we must have  $j > f + k$ . This means the view  $G$  sees some subset  $\mathcal{V}_1$  of  $\mathcal{V}$  with total stake more than  $2N/3$  have made attestations justifying a checkpoint edge  $(B_l, l) \rightarrow (B_J, j)$ , so for any such attestation  $\alpha_1$ ,  $\text{aep}(\text{LJ}(\alpha_1)) = l$  and  $\text{ep}(\alpha_1) = j$ . Similarly,  $G$  also sees more than  $2N/3$  weight worth of validators  $\mathcal{V}_2$  have made attestations justifying  $(B_F, f) \rightarrow (B_k, f + k)$  so for any such attestation  $\alpha_2$ ,  $\text{aep}(\text{LJ}(\alpha_2)) = f$  and  $\text{ep}(\alpha_2) = f + k$ . Thus, for anyone in the intersection  $\mathcal{V}_1 \cap \mathcal{V}_2$ , they have made two distinct attestations  $\alpha_1$  of the former type and  $\alpha_2$  of the latter type. Because

$$l < f < f + k < j,$$

we know

$$\text{aep}(\text{LJ}(\alpha_1)) < \text{aep}(\text{LJ}(\alpha_2)) < \text{ep}(\alpha_2) < \text{ep}(\alpha_1),$$

which allows them to be provably slashed by (S2).  $\square$

We are now ready to prove safety.

**Theorem 5.2** (Safety). In a view  $G$ , if we do not have the following properties, then  $G$  is  $(1/3)$ -slashable.

1. any pair in  $F(G)$  stays in  $F(G)$  as the view is updated.
2. if  $(B, j) \in F(G)$ , then  $B$  is in the canonical chain of  $G$ .

*Proof.* The first property is straightforward from the definitions of  $F(G)$  and  $J(G)$ , so we omit the proof.

The definition of Algorithm 4.2 shows that the canonical chain always goes through the justified pair with highest attestation epoch  $j$ , so by Lemma 5.1 must go through the highest finalized pair in  $F(G)$ . Thus, it suffices to show that no two finalized blocks conflict, because then all the finalized blocks must lie on the same chain, which we just showed must be a subset of the canonical chain.

We now show that if  $(B_1, f_1), (B_2, f_2) \in F(G)$  but  $B_1$  and  $B_2$  conflict, then  $G$  is  $(1/3)$ -slashable. Without loss of generality,  $f_2 > f_1$ . Since  $(B_2, f_2)$  is finalized, it is justified, and we can apply Lemma 5.1. This shows that either  $B_2$  must be a descendant of  $B_1$  (assumed to be impossible since they conflict) or  $G$  is  $(1/3)$ -slashable, as desired.  $\square$

Finally, to get our original notion of safety from the Theorem, note that the definition of finalized blocks automatically persist in future views,

## 6 Plausible Liveness

This section is similar to Casper FFG [7, Theorem 2]. Since our setup is more complex (in particular, Algorithm 4.2 is much more involved), the results do not immediately follow from the previous result, despite having the same main ideas.

To do this, we give an auxiliary definition: given a chain  $c = \text{chain}(B)$  and an epoch  $j$ , we say that  $c$  is *stable at  $j$*  if  $\text{LJ}(\text{EBB}(B, j)) = (B', j - 1)$  for some block  $B'$ .

**Theorem 6.1.** If at least  $2N/3$  stake worth of the validators are honest, then it is always possible for a new block to be finalized with the honest validators continuing to follow the protocol, no matter what happened previously to the blockchain.

*Proof.* Suppose we are starting epoch  $j$ . Specifically, suppose our current slot is  $i = Cj$ ; then it is plausible (by having good synchrony) for everyone to have the same view. In particular, the proposer, who is plausibly honest, would then propose a new block  $B$  with slot  $i$ , which is a child of the output block of  $\text{HLMD}()$  run on the old view. Call this current view (including  $B$ )  $G$  and define  $c = \text{chain}(B)$ . Keep in mind that  $\text{LEBB}(B) = B$ ; we are introducing a new epoch boundary block.

Now, we claim that it is plausible that  $c$  extends to a stable chain at the beginning of the next epoch  $(j + 1)$ . To see this, note that since at least  $2N/3$  stake worth of the validators are honest, it is plausible that they all attest for  $B$  or a descendant (for example, if they are all synced with the network view and vote immediately after  $B$  is created). This creates a supermajority link  $\text{LJ}(B) \xrightarrow{J} (B, j)$ , justifying  $B$ . Now, in the next epoch  $(j + 1)$ , it is plausible for the new block  $B'$  with slot  $i + C = (j + 1)C$  to include all of these attestations (which would happen with good synchrony), so  $\text{LJ}(B') = (B, j)$  and  $\text{chain}(B')$  is indeed stable at epoch  $(j + 1)$ . Thus, by assuming good synchrony and honest validators, it is plausible (possibly having to wait 1 extra epoch) that the chain of the first epoch boundary block of the new epoch is stable, no matter what the state of the blockchain was initially.

Thus, we can reduce our analysis to the case that  $c$  was stable to begin with, meaning that we have a supermajority link  $(B', j - 1) \xrightarrow{J} (B, j)$  in  $\text{ffgview}(B)$ . Then by the same logic above, it is plausible for the next epoch boundary block  $B''$  with slot  $(j + 1)C$  to also be stable, with another supermajority link  $(B, j) \rightarrow (B'', j + 1)$ . This finalizes the pair  $(B, j)$ ; in particular, it is the special case of 1-finalization.  $\square$

The original Casper FFG is a “finality gadget” on top of a (presumed probabilistically live) blockchain, so its focus was not on the probabilistic liveness of the underlying chain, rather that the participants do not get into a situation where we are forced to slash honest validators to continue. Since our paper also provides the underlying protocol, we also want a probabilistic guarantee of liveness, which we will cover in Section 7.

## 7 Probabilistic Liveness

In this section, we prove probabilistic liveness of our main protocol Gasper from Section 4, given some assumptions. Our proof consists of the following steps, which are basically the main ideas of the proof of Theorem 6.1:

1. (assumptions lead to high weight after first slot) Under “good” conditions (such as having enough honest validators, synchrony conditions, etc. to be formalized) honest validators have a high probability of finding a block with a high weight after the first slot.
2. (high weight after first slot leads to high probability of justification) if we have a block with high weight after one slot, its weight “differential” over competitors is likely to increase throughout the epoch, meaning that this block (or a descendent) is likely to be justified.
3. (high probability of justification leads to high probability of finalization) if every epoch is likely to justify a block, then it is very likely for at least one block to be finalized in a stretch of  $n$  epochs as  $n$  increases.

This division is not just organizational; it also encapsulates the assumptions into different parts of the argument so that each part would remain true even if assumptions change. For example, the main theorem corresponding to the 3rd item, Theorem 7.5, does not make any synchrony assumptions and only relies on a probability  $p$  that a block gets justified (though synchrony assumptions may be required to bound  $p$  given the previous parts). Also, we assume in the second part (as we do for simulations in Appendix B) that all validators have equal stake, whereas the third part does not.

This section is the most intricate and dependent on parameters: for one, the current planned implementation for Ethereum 2.0 actually skirts a nontrivial portion of the analysis due to the attestation consideration delay (see Section 8.4); also, the bounds assume equal stake per validator, which may or may not be a useful assumption depending on external factors. However, it is still worthwhile to include an analysis for the “pure” protocol because the proof strategies here, such as using concentration inequalities, are very generalizable to a whole class of these potential slot-based approaches to proof-of-stake, and we believe our analysis captures the heart of what makes these approaches

work probabilistically. In fact, “patches” such as the aforementioned attestation consideration delay exist primarily because of the types of attacks against the pure protocol, so it is useful for intuition-building to address these attacks even if the actual implementation sidesteps them. Thus, the primary value of this section is as a proof-of-concept of the analysis involved in designing probabilistically-live protocols such as Gasper, as opposed to a mathematical guarantee of the actual implementation.

## 7.1 High Weight After the First Slot - The Equivocation Game

We define the *equivocation game* to be the following:

1. The game is parametrized by  $(\mathcal{V}, a, \epsilon_1, \epsilon_2)$ , where  $\mathcal{V}$  is a set of validators and  $(a, \epsilon_1, \epsilon_2)$  are real numbers that parameterize network / synchrony conditions (for ease of reading we push the details to the Appendix).
2. As in Gasper,  $|\mathcal{V}| = N$  and each  $V \in \mathcal{V}$  has some fixed stake  $w(V)$ . We assume the total stake is  $N$  (so the average stake is 1) and the total amount of stake of honest validators is at least  $2N/3$ .
3. There are 2 options to vote for, which we call  $O_1$  and  $O_2$  for short. These are abstractions for voting on 2 conflicting blocks in Gasper, and the concept of “blocks” are not included in this equivocation game. We (meaning the honest validators) *win* if either  $O_1$  or  $O_2$  obtain at least  $2N/3$  stake worth of votes, and we lose (i.e., we are in a state of equivocation) otherwise. Note that this is equivalent to the idea that there is a stake differential of  $N/3$  between the two choices after the game.

The equivocation game is defined to capture the idea of Gasper with just 2 choices in one slot. Most of the necessary assumptions for liveness are encoded into the game so that the other liveness results in this section can be as independent as possible.

In Appendix B, we flesh out the details of the equivocation game and conduct the simulations of equivocation game with three regimes (a pessimistic regime, an optimistic regime, and a regime in between). The main idea is that under “reasonable” conditions  $(\mathcal{V}, a, \epsilon_1, \epsilon_2)$ , it is very likely that after the first slot, we win the game, meaning that we should expect one of the pairs in Gasper to have  $2N/3$  worth of stake supporting it after the first slot.

For the big picture, the only thing we need for the rest of Probabilistic Liveness is the idea that we “win” the equivocation game (specifically, have a block with a high number of attestations over the second-best option after the first slot) with some probability  $r$ , and having higher  $r$  increases the guarantees that a block will be justified in Theorem 7.4. For practical purposes, the work in the Appendix seems like an  $r$  of around 80% seems to be a reasonable heuristic under not too strong or weak synchrony conditions, and that is already assuming a very bad case where close to  $1/3$  of validators are dishonest.

**Remark.** Clearly, the concept of equivocation games can be generalized, and similar games could be used to study other proof-of-stake models. There are many directions (the protocols for honest validators, latency modeled by something more than the uniform distribution, etc.) which may be interesting to study for future work, but may be too distracting for the purpose of our paper.

## 7.2 High Weight after First Slot Leads to Justification

In this section, we focus on the  $j$ -th epoch  $[T, T + C]$ , where  $T = jC$ . Let  $S = N/C$ . We define the following set of assumptions, which we call  $A(\epsilon)$ :

- the network is  $(1/2)$ -synchronous starting at time  $T = jC$ ;
- each validator has 1 stake;
- the total number of byzantine validators is equal to  $N/3 - C\epsilon$ , meaning the average number of byzantine validators in each slot’s committee is  $S/3 - \epsilon$ .

For each  $i \in \{0, 1, \dots, C - 1\}$ , define  $h_i$  to be the number of guaranteed honest attestors in slot  $jC + i$  and  $b_i$  to be the number of dishonest attestors in the slot. Let  $S = N/C$ , so for all  $i$ ,  $h_i + b_i = S$ . We show that the distribution of honest / dishonest attestors should not stray too much from expectation in Proposition 7.2, using ubiquitous concentration inequalities:

**Proposition 7.1.** Let  $\mathcal{X} = (x_1, \dots, x_N)$  be a finite list of  $N$  values with  $a \leq x_i \leq b$  and let  $X_1, \dots, X_n$  be sampled without replacement from  $\mathcal{X}$ . The following hold:

$$\mathbb{P}\left(\sum_{i=1}^n X_i - n\mathbb{E}[X] \geq n\delta\right) \leq \exp\left(\frac{-2n\delta^2}{(1 - (n-1)/N)(b-a)^2}\right)$$

$$\mathbb{P}\left(\sum_{i=1}^n X_i - n\mathbb{E}[X] \leq -n\delta\right) \leq \exp\left(\frac{-2n\delta^2}{(1 - (n-1)/N)(b-a)^2}\right)$$

*Proof.* See e.g. [19]. □

**Proposition 7.2.** Suppose the assumptions  $A(\epsilon)$  are met. Assume that  $C = 2^L$ . Then the conjunction of the following events (we purposefully skip  $h_0$ ):

$$\begin{aligned} E_1 &: h_1 \geq 2 \cdot \frac{S}{3} \\ E_2 &: h_2 + h_3 \geq 2^2 \cdot \frac{S}{3} \\ E_3 &: h_4 + h_5 + h_6 + h_7 \geq 2^3 \cdot \frac{S}{3} \\ &\dots \\ E_L &: h_{C/2} + \dots + h_{C-1} \geq 2^L \cdot \frac{S}{3} \end{aligned}$$

has probability

$$\mathbb{P}\left(\bigcap_{i=1}^L E_i\right) \geq 1 - \sum_{i=1}^L \exp\left(-\frac{2^i \epsilon^2}{\left(1 - \frac{2^{i-1}S-1}{2^L S}\right)S}\right) \geq 1 - \sum_{i=1}^L \exp\left(-\frac{2^i \epsilon^2}{S}\right).$$

*Proof.* We use the hypergeometric distribution model; that is, we consider the set  $\mathcal{X} = (x_1, \dots, x_N)$  representing the validators  $\mathcal{V}$ , where  $x_j = 1$  if the  $j$ -th validator is honest and 0 otherwise. Let  $X_1, \dots, X_{2^{i-1}S}$  be  $2^{i-1}S$  samples from  $\mathcal{X}$  without replacement. Then for each  $j$ ,  $1 \leq j \leq 2^{i-1}S$ ,  $\mathbb{E}[X_j] = 2/3 + \epsilon/S$ , so:

$$\mathbb{E}\left[\sum_{j=1}^{2^{i-1}S} X_j\right] = 2^{i-1}S(2/3 + \epsilon/S) = 2^i S/3 + 2^{i-1}\epsilon.$$

We can then bound each  $E_i$  (as a sum of  $X_j$ 's) by

$$\begin{aligned} \mathbb{P}\left(\sum_{j=1}^{2^{i-1}S} X_j \geq 2^i \frac{S}{3}\right) &= 1 - \mathbb{P}\left(\sum_{j=1}^{2^{i-1}S} X_j < 2^i \frac{S}{3}\right) \\ &= 1 - \mathbb{P}\left(\sum_{j=1}^{2^{i-1}S} X_j - \left(\frac{2^i S}{3} + 2^{i-1}\epsilon\right) < -(2^{i-1}S) \frac{\epsilon}{S}\right) \\ &\geq 1 - \exp\left(-\frac{2^i \epsilon^2}{\left(1 - \frac{2^{i-1}S-1}{2^L S}\right)S}\right), \end{aligned}$$

where the last inequality results from Proposition 7.1, recalling that  $N = 2^L S$ . The probability of each event  $E_i$  when considered independently of each other is then bounded below by this value. Using the intersection bound on all  $E_i$  we get the first inequality in the desired statement; the second inequality holds by comparing denominators. □

**Remark.** It suffices to also use e.g. Hoeffding's inequality with a binomial model where the validators are assigned as honest or byzantine with replacement. We leave it as an exercise to the reader that this method immediately gets the weaker bound

$$\mathbb{P}\left(\bigcap_{i=1}^L E_i\right) \geq 1 - \sum_{i=1}^L \exp\left(-\frac{2^i \epsilon^2}{S}\right).$$

While the binomial model is only an approximation, the model with replacement is strictly more mean-reverting than the model without, so the approximation is in the correct direction for us. This could be made rigorous with e.g. coupling methods. We use the weaker bound because it is algebraically cleaner and loses very little compared to the stronger bound.

It can be seen that the exponential terms in the result of proposition 7.2 is fairly small when  $\epsilon$  is on the order of  $\sqrt{S}$ . For example, when  $C = 64$  (a power of 2 actually makes Proposition 7.2 work cleanly, though it is certainly not a crucial part of the bound) and thus  $S = 900$ , picking  $\epsilon = 30$  (meaning we have as many as 17280 byzantine validators out of 57600) gives a probability bound of around 85%; changing  $\epsilon$  to 40 jumps the probability to around 97%.

**Lemma 7.3.** Suppose the assumptions in  $A(\epsilon)$  are met. Then, if  $\text{view}(\text{NW}, T + C)$  justifies a new block  $B$  not in  $J(\text{view}(\text{NW}, T))$ ,  $B$  must be a descendant of the last justified block in  $J(\text{view}(\text{NW}, T))$ .

*Proof.* Recall that honest attestors wait for  $1/2$  time before attesting to their assigned slot. Thus, in the upcoming epoch  $j$ , all of the honest attestors in this epoch attest in the time period  $[T + 1/2, T + C]$ . Because we have  $(1/2)$ -synchrony, it means their views during this period all include  $\text{view}(\text{NW}, T)$ . By the nature of Algorithm 4.2, this means all of their block proposals and attestations must be to either:

- a descendant of  $B_J$ , the last justified block in  $\text{view}(\text{NW}, T)$ , or
- a descendant of some new block justified in epoch  $j$ .

Our Lemma only fails in the second case, and if some new block in this epoch obtains  $2/3$  of the attestations of this epoch.

Luckily, the chicken-and-egg favors us. It is possible for byzantine validators to propose a new block  $B'_J$  that's not a descendant of  $B_J$ . However, it would be impossible for  $B'_J$  to receive enough votes in this epoch, as all of the attestations before  $B'_J$  is justified must go to a descendant of  $B_J$  by Algorithm 4.2. Thus, we know that if we justify a new block, it must be a descendant of  $B_J$ .  $\square$

**Theorem 7.4** (Justification Probabilistic Liveness). Suppose the assumptions in  $A(\epsilon)$  are met, and suppose we win the equivocation game corresponding to the first slot with probability  $r$ . Let  $B_J$  be the last justified block in  $J(\text{view}(\text{NW}, T))$ . Then,  $\text{view}(\text{NW}, T + C)$  will justify a new descendant of  $B_J$  with probability at least

$$r - \sum_{i=1}^L \exp\left(-\frac{2^i \epsilon^2}{S}\right) - \frac{1}{3^{C-1}}.$$

*Proof.* We can think of the first slot of this epoch  $[T, T + 1]$  as an equivocation game with  $S$  validators ( $h_0$  honest and  $b_0$  byzantine) where all of the options are  $B_J$  or a descendant<sup>7</sup> of  $B_J$ . Thus, with probability  $r$ , one of these options receives at least  $2S/3$  attestations after slot  $jC$ . We call this winning block  $B_w$  (ideally,  $B_w$  is simply just the block  $jC$ , though this is not necessary). For sake of weighting in Algorithm 4.2, note that against any other option in the equivocation game,  $B_w$ 's weight is winning by at least  $(2S/3 - S/3) = S/3$ .

<sup>7</sup>As stated in the proof of Lemma 7.3, it is possible that there are options that are not, but they will never receive attestations from honest validators, so it would be strictly worse for dishonest validators to create such options.

By the intersection bound, with probability at least

$$r - \sum_{i=1}^L \exp\left(-\frac{2^i \epsilon^2}{S}\right)$$

we also satisfy the events in Proposition 7.2. We now show that when these events are satisfied, **all remaining honest validators in slots  $(jC+1), \dots, (jC+C-1)$  will vote for  $B_w$  or a descendant.**

Consider slot  $jC + 1$ . Because of  $E_1$ , we know  $b_1 < S/3$ , so even if all  $b_1$  potentially byzantine actors conspire to vote for some other option, the weight advantage of  $B_w$  cannot be diminished to 0 by the time the honest validators vote at time  $(jC + 1.5)$ . This means that all honest validators will keep attesting to  $B_w$  or a descendant block (an honest proposer in slot  $jC + 1$  would propose block  $jC + 1$  as a descendant of  $B_w$ , for example). By  $E_1$ , we know that  $B_w$  gains at least  $2S/3$  weight while a rival block gains at most  $S/3$  weight during this slot, which means that the weight differential preferring  $B_w$  changes by at least  $(2S/3 - S/3) = S/3$ . This means by the end of slot  $jC + 1$ ,  $B_w$  is now winning with at least weight  $S/3 + S/3 = 2S/3$ .

Now consider  $E_2$  and the next 2 slots,  $jC+2$  and  $jC+3$ . Between them, we know  $h_2 + h_3 < 2S/3$ , so even if all the byzantine actors conspire, they cannot destroy the winning differential of  $B_w$ , which by the end of these 2 next slots will be winning with at least weight  $(2S/3 - 2S/3 + 4S/3) = 4S/3$ .

Inductively, this logic continues for all remaining slots in the epoch (specifically, all honest validators attest to  $B_w$  or a descendant. Here the structure of Algorithm 4.2 is important because while honest attestors may (and probably will) attest to new blocks, the weight of their attestations are added to that of  $B_w$  as well.

We have thus concluded that during all remaining slots,  $B_w$  accumulates all attestations from honest validators. After slot  $jC$ , it has at least  $2S/3$  votes with probability  $r$ . Then, we know it picks up weight at least  $h_1 + h_2 + \dots + h_{C-1}$ , for a total weight of

$$\frac{2}{3}S + \sum_{i=1}^L \frac{2^i}{3}S = \frac{2}{3}S + \frac{2}{3}(2^L - 1)S = \frac{2^{L+1}}{3}S = \frac{2}{3}N,$$

so we indeed achieve enough weight for a supermajority link.

To prove  $B_J$  is different from  $B_w$ , we need at least one honest validator to propose a new block on top of  $B_J$  before or in the first slot of epoch  $\text{aep}(B_w)$ . Since  $B_J$  receives more than  $\frac{2}{3}$  of votes in epoch  $\text{aep}(B_J)$ , there are more than  $\frac{2}{3}$  honest validators in  $\text{aep}(B_J)$ . The chance of **no** honest validator proposing a block on top of  $B_J$  in  $\text{aep}(B_J)$  is then bounded above by  $(\frac{1}{3})^{C-1}$ , which is vanishingly small.  $\square$

### 7.3 Probabilistic Justification Leads to Probabilistic Finalization

The main theorem is the following:

**Theorem 7.5** (Finalization Probabilistic Liveness). Assume the probability of justifying a block (as in Proposition 7.2) is independently  $p \geq 1/2$  for each epoch, the probability of failing to finalize a block in the next  $n$  epochs approaches 0 exponentially as a function of  $n$ .

*Proof.* Recall that we expect most finalization to be 1-finalization, when one justified block justifies an adjacent epoch boundary block. For this proof, it suffices to show that the probability of failing to even 1-finalize approaches 0.

We consider an epoch a “success” if a block is justified in the epoch, and a “failure” otherwise. Thus, if 2 adjacent epochs are “successful,” we finalize a block. Therefore, the probability of not getting a  $(k = 1)$  finalization in  $n$  epochs is the probability that no 2 adjacent epochs out of the next  $n$  are successful. Using the independence assumption, this has the probability

$$\sum_{i \geq \frac{n}{2}} \binom{i+1}{n-i} (1-p)^i p^{n-i},$$

because for a particular  $i \geq n/2$ , there are  $\binom{i+1}{n-i}$  ways to select  $i$  failures.



Since  $p \geq 0.5$ , we know  $p \geq (1 - p)$ , so the sum is bounded above by

$$\left( \sum_{i=n}^{\lfloor \frac{n}{2} \rfloor} \binom{i+1}{n-i} \right) (1-p)^{n/2} p^{n/2}$$

It is well-known (by e.g. induction) that:

$$\sum_{i=n}^{\lfloor \frac{n}{2} \rfloor} \binom{i+1}{n-i} = \binom{n+1}{0} + \binom{n}{1} + \binom{n-1}{2} + \dots = F_n,$$

the  $n$ -th Fibonacci number, which we know is of the form

$$F_n = \frac{1}{\sqrt{5}} \left[ \frac{(1 + \sqrt{5})}{2} \right]^n - \frac{1}{\sqrt{5}} \left[ \frac{(1 - \sqrt{5})}{2} \right]^n$$

The second term vanishes as  $n \rightarrow \infty$ , so our desired quantity is bounded above by (times a constant)

$$\left[ \frac{(1 + \sqrt{5}) \sqrt{p(1-p)}}{2} \right]^n$$

We have the bound  $\sqrt{p(1-p)} \leq 1/2$  (by, e.g. AM-GM inequality), so the failure rate is bounded above by

$$\frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{4} \right)^n,$$

which goes to 0 as  $n \rightarrow \infty$ . Finally, recall that we are only looking at 1-finalization, so the chances of finalization is strictly higher (though in practice the other types of finalization should occur very rarely).

□

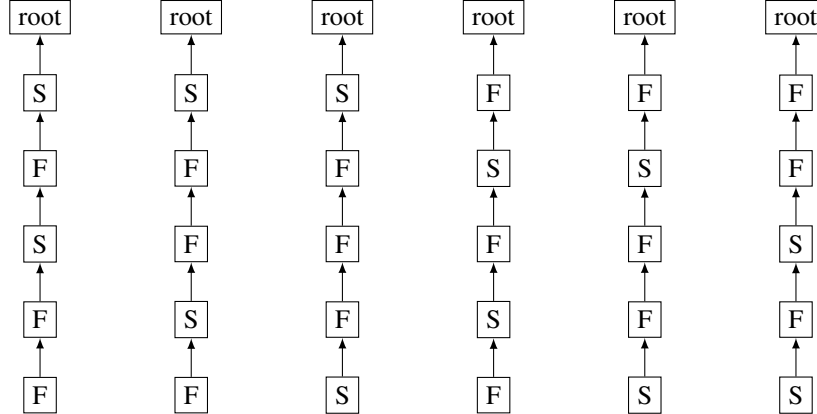


Figure 7: The cases where we **fail** to finalize a block in  $n = 5$  epochs with 3 failing epochs. “S” denotes success and “F” denotes failure (with respect to Proposition 7.2).

Theorem 7.5 is “agnostic” of the justification probability  $p$  and how we obtained it. This means if we tweak the protocol, change the assumptions, etc. and obtain different bounds/estimates for  $p$  than we do from the current Theorem 7.4, our result still holds. For this reason, we use  $p$  as an **input** to Theorem 7.5 instead of reusing our actual bound in Theorem 7.4. A secondary effect of abstracting away justification liveness is that **the techniques can be used to prove finalization probabilistic liveness for general 2-stage protocols, not just Gasper** (it could even be extended easily to protocols with 3 or more stages, though the bounding would be different). Regardless of the

protocol, as  $n$  increases, as long as  $p \geq 0.5$ , the probability of finalizing a block increases rapidly, getting around 99% even for  $p = 0.5$  after  $n = 20$  epochs. See Table 1 for some computations.

Table 1: The probabilities of **not** finalizing blocks in  $n$  epochs, with  $p$  the probability of justification within each epoch.

n	p	probability of not finalizing any block in n epochs
2	0.5	0.75
5	0.5	0.40625
7	0.5	0.265625
10	0.5	0.140625
20	0.5	0.016890525817871094
2	0.66	0.5644
5	0.66	0.18460210239999997
7	0.66	0.08322669164799996
10	0.66	0.025351233503186934
20	0.66	0.0004854107646743359

**Remark** (Relationship with Plausible Liveness). It may feel like we technically get plausible liveness “for free” from probabilistic liveness. So why do we treat them separately in this paper?

For one, plausible liveness is an immediate consequence from the rules of the protocol Gasper and requires very few assumptions, while probabilistic liveness is dependent on probabilistic assumptions (such as network synchrony) and is more fragile with slight changes to the protocol. In particular, our probabilistic liveness treatment makes the very strong assumption that all validators have the same stake. Thus, we choose to present plausible liveness separately to emphasize that even if these assumptions are not satisfied “in real life,” plausible liveness remains.

For another, the emphasis of plausible liveness is that “honest validators will never be required by the protocol to voluntarily slash themselves to continue” and the emphasis of probabilistic liveness is “new blocks will probably be justified/finalized quickly;” these are different takeaways.

## 8 Practice versus Theory

Ethereum 2.0’s practical implementation in [11] contains different design decisions from Gasper, which is meant to be a “clean” and more mathematically tractable protocol that captures the theoretical core of the beacon chain design. In this section, we consider the proposed implementation’s differences from Gasper, without getting lost in the much messier analyses a rigorous study of combining all of these details would require.

### 8.1 Sharding

Gasper is motivated by the Ethereum 2.0 *beacon chain*, which is the “main” blockchain in the Ethereum 2.0 design that stores and manages the registry of validators. In this implementation, a *validator* is a registered participant in the beacon chain. Individuals can become a validator of the beacon chain by sending Ether into the Ethereum 1.0 deposit contract. As in Gasper, validators create and attest to blocks in the beacon chain. Attestations are simultaneously proof-of-stake votes for a beacon block (as in our design) but also availability votes for a “shard block,” which contains data in a different “shard chain.” This concept of *sharding* creates interesting engineering and mathematical questions outside the scope of our paper, which is limited to the beacon chain.

### 8.2 Implementing the View

In Gasper, we can treat views and related concepts as abstract mathematical concepts and validators as perfectly reasoning agents with infinite computational power. In practice, validators will not be directly reasoning with a graphical data structure of a view; instead, they will use software to parse the view given to them and follow the protocol. Thus, in the actual implementation [11], validators run a program that updates the “store”, which is basically a representation of the view. The store, as the input for LMD GHOST fork choice rule, is updated whenever a block or an attestation is

received. The beacon chain also keeps track of a “state,” a derived data structure from the view that tracks stake-related data.

Obviously, mistakes when interpreting these structures and their updates may cause issues with safety and liveness not related to those coming from Gasper itself. Even though in our work we limit our analysis to the mathematical parts of the protocol, we remind the reader these other issues are also important; security holes arising from a carelessly implemented protocol at the software level are not protected by the mathematical guarantees of Gasper.

### 8.3 Attestation Inclusion Delay

In Gasper, when a validator proposes a block, he/she includes all new attestations in his/her view. In the planned implementation, there is an integer parameter for the *attestation inclusion delay* (say  $n$ ) such that when a validator proposes a block, he/she only includes attestations that were made at least  $n$  slots ago.

The purpose of the attestation inclusion delay is to prevent centralization and reward advantage. If the slot length is “short” (meaning the latency of attestation propagation is high for normal nodes in the network), then highly-connected nodes might be able to get attestation data faster and be able to publish them faster, which may correspond to various advantages depending on the reward incentives.

The enforced delay allows attestations to disseminate more widely before they can be included in blocks. Thus, nodes have more of an equal opportunity to capture the attestation inclusion rewards for proposing blocks. The plan is to tune the attestation inclusion delay (in the range of 1 to 4 slots) depending on real-world network data. Smaller values improve the transaction processing speed, and larger values improve decentralization. Note increasing the slot time is an alternative method to promote decentralization.

### 8.4 Attestation Consideration Delay

In Gasper, when a validator is supposed to attest at slot  $N$ , the validator runs the HLMD GHOST fork-choice rule considering all attestations in the validator’s view. In the planned implementation, he/she only considers attestations that are at least 1 slot old. Specifically, the view used as input to the fork-choice rule contains all valid blocks up to slot  $N$  but only contains valid attestations up through slot  $N - 1$ .

This one-slot delay protects validators from a certain class of timing attacks in which byzantine validators eagerly broadcast slot  $N$  attestations rather than waiting for the  $N + 1/2$  time when they are “supposed” to attest. This allows them to theoretically take advantage of the network latency to split the vote of the honest validators to keep the chain in a state of equivocation. This is an important attack to counter as publishing attestations at the “wrong” time in a block is not a slashable offense (since the adversary can theoretically fake timestamps and we did not assume Gasper has mechanisms that make faking timestamps harder).

We cover the one-shot delay attack in our analysis of the *equivocation game* (see Appendix B), and conclude it is plausible for the attack to be effective in very pessimistic regimes. However, the probability byzantine validators succeed in this particular type of attack is greatly lowered even in pessimistic regimes if we implement something like the attestation consideration delay.

Specifically, this delay improves the parameters for probabilistic liveness as analyzed in Theorem 7.4. As long as an honest validator is selected to propose a block (with probability at least  $2/3$  in the worst case)  $B$ , then all honest validators will vote for  $B$  in the first slot since they will ignore all the attestations made in the first slot, including those made by the dishonest validators. In other words, the attestation consideration delay neutralizes the 1-st slot equivocation game as a point of attack, meaning the probabilistic liveness of the proposed implementation comes with even better guarantees than raw Gasper. A rigorous analysis of the probabilistic liveness for this setup may make for interesting future work.

## 8.5 A Four-Case Finalization Rule

While Gasper’s Definition 4.9 captures the “general” version of the mathematical idea of finalization, the proposed implementation uses a “reduced” version only looking at the last 4 epochs for practicality. In particular, let  $B_1, B_2, B_3, B_4$  be epoch boundary blocks for consecutive epochs, with  $B_4$  being the most recent epoch boundary block.

1. If  $B_1, B_2$  and  $B_3$  are justified and the attestations  $\alpha$  that justified  $B_3$  have  $LJ(\alpha) = B_1$ , we finalize  $B_1$ .
2. If  $B_2, B_3$  are justified and the attestations  $\alpha$  that justified  $B_3$  have  $LJ(\alpha) = B_2$ , we finalize  $B_2$ .
3. If  $B_2, B_3$  and  $B_4$  are justified and the attestations  $\alpha$  that justified  $B_4$  have  $LJ(\alpha) = B_2$ , we finalize  $B_2$ .
4. If  $B_3, B_4$  are justified and the attestations  $\alpha$  that justified  $B_4$  have  $LJ(\alpha) = B_3$ , we finalize  $B_3$ .

These are special cases of Definition 4.9 for cases  $k = 1$  (the first two) and  $k = 2$ . The idea here is Ethereum 2.0 will only honor attestations for up to 2 epochs. This means only epoch boundary blocks up to 2 epochs in the past can be newly justified (and thus finalized), which gives the 4 cases. See Figure 8.

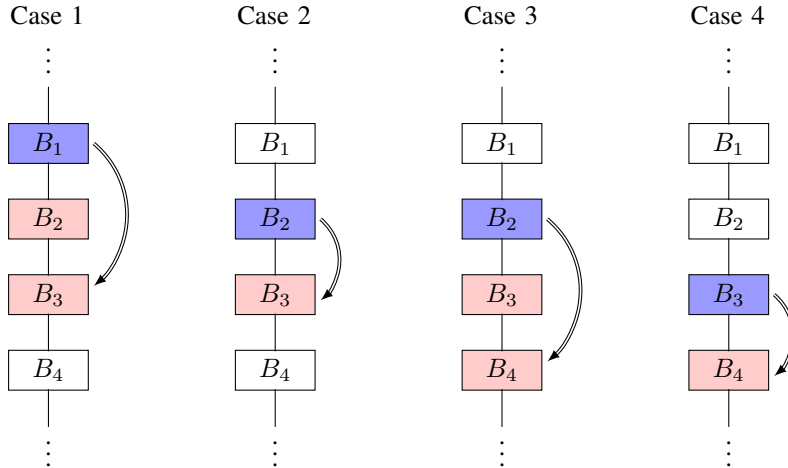


Figure 8: From left to right, example of cases 1 to 4 respectively. Both red and blue colors indicate justified blocks, where the blue blocks are now finalized because of the observed justification, shown by double arrows.

## 8.6 Safety - Dynamic Validator Sets

In Gasper’s safety analysis (Section 5), we assumed *static* validator sets, meaning the set of validators cannot change over time. Recall our main result, Theorem 5.2, tells us we are able to catch  $N/3$  weight worth of validators violating the slashing conditions if safety is broken. However, in practice, we would like to support *dynamic* validator sets, which means validators are allowed to *activate* (enter) and *exit* the validator set  $V$  over time. This means byzantine activators can act maliciously, but then leave to avoid their stake being slashed. This setup reduces the number of “active” validators we can punish for violating slashing conditions.

**Definition 8.1.** Assume there are 2 validator sets,  $\mathcal{V}_1$  and  $\mathcal{V}_2$ , where  $\mathcal{V}_2$  is a validator set later in time compared to  $\mathcal{V}_1$ . We define  $A(\mathcal{V}_1, \mathcal{V}_2)$ , the validators who *activated* (from  $\mathcal{V}_1$  to  $\mathcal{V}_2$ ), to be the set of validators who are not in  $\mathcal{V}_1$  but are in  $\mathcal{V}_2$ , and  $E(\mathcal{V}_1, \mathcal{V}_2)$ , the validators who *exited* (from  $\mathcal{V}_1$  to  $\mathcal{V}_2$ ), to be the set of validators who are in  $\mathcal{V}_1$  but not in  $\mathcal{V}_2$ .

First, we note that our key ideas from Section 5 still hold:

**Lemma 8.2.** Suppose we allow dynamic validator sets. In a view  $G$ , if  $(B_1, f_1)$  and  $(B_2, f_2)$  in  $F(G)$  conflict, then the blockchain must be  $(1/3)$ -slashable. Specifically, there must exist 2 justified pairs  $(B_L, j_L)$  and  $(B_R, j_R)$  in  $G$  and 2 subsets  $\mathcal{V}_1 \subset \mathcal{V}(B_L), \mathcal{V}_2 \subset \mathcal{V}(B_R)$ , each with weight at least  $2/3$  of the stake of its corresponding attestation epoch, such that their intersection  $\mathcal{V}_1 \cap \mathcal{V}_2$  violates (S1) or (S2).

*Proof.* The proof is essentially identical to that of Lemma 5.1 and the reduction of Theorem 5.2 to it. The only difference is that we must rephrase the conditions of Lemma 5.1 as conditions about the validator sets at the time of their attestations.

As in the proof of Theorem 5.2, if we have 2 conflicting blocks  $(B_1, f_1)$  and  $(B_2, f_2)$ , then either  $f_1 = f_2$  or  $f_1 \neq f_2$ . If  $f_1 = f_2$ , then we can set  $(B_L, j_L) = (B_1, f_1)$  and  $(B_R, j_R) = (B_2, f_2)$  to satisfy our claim, as their intersection violates (S1). If  $f_1 \neq f_2$ , without loss of generality,  $f_1 < f_2$ . Since  $(B_2, f_2)$  is finalized, it is also justified, and we now satisfy the setup for Lemma 5.1 where  $(B_F, f) = (B_1, f_1)$  and  $(B_J, j) = (B_2, f_2)$ . Applying the Lemma, we conclude that either  $B_2$  is a descendent of  $B_1$  (a contradiction as they are conflicting), or there must be some pair of pairs (not necessarily  $(B_F, f)$  or  $(B_J, j)$  themselves), each of whose attestations have  $2N/3$  stake, and whose intersections violate (S1) or (S2).  $\square$

We now present our main theorem. The main idea is to first take a block  $B_0$  (most naturally, we can take the common parent of the conflicting blocks, but this choice is not necessary) published at a “reference time” with a “reference validator set”  $\mathcal{V}_0$ . Then, we can upper bound the total weight of activations and exits between the validator sets of the justified blocks  $B_L$  and  $B_R$  (from Lemma 8.2) and  $\mathcal{V}_0$  as a function of the elapsed time since the referenced time. This allows us to lower bound the size of the slashable intersection of the quorums of those blocks.

**Theorem 8.3.** Suppose we have a view  $G$  which contains two conflicting finalized blocks, then  $G$  has enough evidence to slash validators with total weight at least

$$\max(w(\mathcal{V}_L) - a_L - e_R, w(\mathcal{V}_R) - a_R - e_L) - w(\mathcal{V}_L)/3 - w(\mathcal{V}_R)/3,$$

where

1. there is a block  $B_0 \in G$  with validator set  $\mathcal{V}_0$ ;
2. blocks  $B_L$  and  $B_R$  exist by Lemma 8.2, with validator sets  $\mathcal{V}_L$  and  $\mathcal{V}_R$  respectively.
3.  $a_L = w(A(\mathcal{V}_0, \mathcal{V}_L))$  and  $e_L = w(E(\mathcal{V}_0, \mathcal{V}_L))$ ; similarly define  $a_R$  and  $e_R$ .

*Proof.* By assumption, the pair of (unnamed) conflicting finalized blocks give rise to  $B_L$  and  $B_R$  with Lemma 8.2, which in turn have quorums  $Q_L \subset \mathcal{V}_L$  and  $Q_R \subset \mathcal{V}_R$ , both with at least  $2/3$  of the weight of their corresponding validator sets  $\mathcal{V}_L$  and  $\mathcal{V}_R$ .

We denote the different sets created by the overlap of our 3 validator sets to be  $A, B, C, D, E, F, G$ , as in Figure 9.

Next, we wish to bound the intersection of the quorums  $Q_L$  and  $Q_R$  as depicted in Figure 10. Let  $\mathcal{V}_{LR} = \mathcal{V}_L \cap \mathcal{V}_R = C \cup F$ , we have:

$$\begin{aligned} w(Q_L \cap Q_R) &\geq w(Q_L \cap \mathcal{V}_{LR}) + w(Q_R \cap \mathcal{V}_{LR}) - w(C \cup F) \\ &\geq 2w(\mathcal{V}_L)/3 - w(B \cup E) + 2w(\mathcal{V}_R)/3 - w(D \cup G \cup C \cup F) \\ &= 2w(\mathcal{V}_L)/3 + 2w(\mathcal{V}_R)/3 - (w(B \cup C \cup D \cup E \cup F \cup G)) \\ &= w(C) + w(F) - w(\mathcal{V}_L)/3 - w(\mathcal{V}_R)/3. \end{aligned}$$

Let  $X = w(C) + w(F)$ . We know

$$\begin{aligned} X &= w(\mathcal{V}_L) - w(B \cup E) \\ &\geq w(\mathcal{V}_L) - w(A \cup B \cup E \cup F) \\ &= w(\mathcal{V}_L) - a_L - e_R, \end{aligned}$$

and similarly

$$X \geq w(\mathcal{V}_R) - a_R - e_L. \quad \square$$

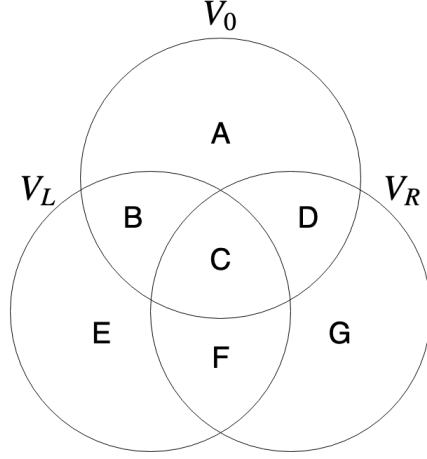


Figure 9: A Venn diagram of the initial validator set  $\mathcal{V}_0$  and two validator sets  $\mathcal{V}_L$  and  $\mathcal{V}_R$ , the letters A through G pictorially represent how many validators are in each overlap.

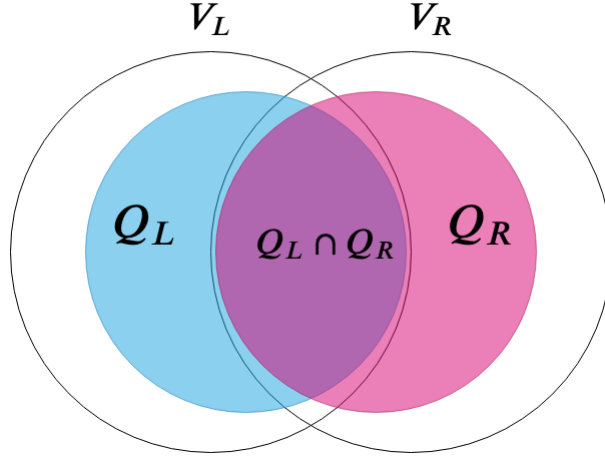


Figure 10: The quorums  $Q_L$  and  $Q_R$  have overlapping validators between the conflicting blocks  $B_L$  and  $B_R$ .

As a sanity check, note all the  $a$ 's and  $e$ 's are zero in the extreme case of no exits or activations, so we recover the original bound of  $w(\mathcal{V}_L)/3 = N/3$ . Also, we can avoid negative signs by e.g. using a linear combination of the two bounds for  $X$  to get the bound

$$w(\mathcal{V}_L)/3 - (2a_L/3 + 2e_R/3 + a_R/3 + e_L/3).$$

The power of the bound depends on our policies on activating and exiting. Suppose  $B_0$ , the “reference block” was published at time  $T_0$ , and  $B_L$  and  $B_R$  are published at  $T_{now}$  (more precisely, we should define  $T_{now}$  to be the maximum of the two blocks’ publication times). Then we can control the  $a_*$  and  $e_*$  as a function of the time difference and the policies. Here are some examples and observations:

- If we allow a constant stake of  $k$  worth of new validators to activate per epoch, then we can bound

$$a_L, a_R \leq k(T_{now} - T_0).$$

We can do the same with exiting and obtain bounds on  $e_L$  and  $e_R$ , possibly with a different constant in the same protocol.

- If we allow activating proportional to validator stake, e.g., up to  $k \cdot w(\mathcal{V})$  can activate in one epoch, then the bounds become exponential:

$$a_L, a_R \leq w(\mathcal{V})(1 + k)^{T_{now} - T_0};$$

In the limit this becomes an exponential bound.

- As the difference between  $T_{now}$  and  $T_0$  becomes longer, the  $w()$  terms will get big and the bound becomes less useful, with the right-hand-side eventually becoming negative. This is unavoidable; one can imagine a “ship of Thesesus” situation where the validator sets have changed completely, in which case it would be impossible to slash anyone still in the system.
- The bounds can also take care of latency; if we are prepared for up to time  $\delta$  of latency, then we effectively add  $\delta$  to computations where  $(T_{now} - T_0)$  occur in our bounds of  $a$ ’s and  $e$ ’s, which we can then feed back into Theorem 8.3.

These bounds show **a tradeoff between the flexibility of entering and exiting the blockchain and the ability to catch malicious actors**. Our static result, Theorem 5.2, has the intuition “if our protocol breaks safety, we can provably slash 1/3 worth of validators” to imply safeness of our protocol as long as we have a strong enough belief in the validators. Our dynamic result, Theorem 8.3, is of the more nuanced form “if our protocol breaks safety, then as long as the attackers cannot create a fork with long branch lengths, we can provably slash a little less than 1/3 worth of validators.” This means pure Gasper is susceptible to attacks such as malicious actors shutting down the whole network for many<sup>8</sup> epochs and then suddenly appearing with a conflicting fork, such that we cannot slash much stake worth of validators since they have already exited during this time. Such situations can be safeguarded in practice by simply not including old attestations into one’s view; the current Ethereum 2.0 spec, for example, does not accept attestations 2 or more epochs old, which avoids this attack.

## 8.7 Extreme Cases; Hard Forks

Even though we have fairly unconditional plausible liveness with a  $\frac{2}{3}$  majority of honest stakeholders and probabilistic liveness under “good” conditions, it is important to take into account of worst-case scenarios. In the case of extended forking and lack of finality due to lots of *non-live* (non-participating) validators who are not necessarily malicious (and thus not slashed), the Ethereum 2.0 beacon chain has a mechanism by which *live* (participating) validators on a fork retain their stake, whereas the non-live validators “bleed” stake such that the live validators eventually become a  $\frac{2}{3}$  majority, in time on the order of magnitude of weeks. Such a case may happen if, e.g., the global internet is partitioned such that 50% of validators are on one side and 50% on the other.

A different extreme case is where a chain “flip-flops” continuously such that there is no distinct partition, and instead the majority just keeps switching forks each epoch. The ultimate backup plan for this situation or more severe cases is making *hard forks* (or *manual forks*), points at which the community running a blockchain chooses to alter the consensus rules (famous such cases include Bitcoin vs. Bitcoin Cash, or ETH vs. ETH Classic). This might be due to upgrading or adding features, fixing critical issues in production, or addressing a fractured community or political split. Usually, this results in a new set of protocol rules being run starting at some particular point in the blockchain. It is an important engineering problem, again outside of our mathematical scope, that the state transition and fork-choice functions can handle these changes smoothly. In most considerations the HLMD GHOST fork-choice rule would be untouched. However, when the alterations are “deep” inside the logic, considerations beyond the built-in forking mechanism would have to be taken and manual forks might be implemented.

## 9 Conclusion

We presented an abstract protocol, Gasper, that combines LMD GHOST and Casper FFG for a full proof-of-stake based blockchain design. This is the first formal proof-of-concept of Casper FFG to

<sup>8</sup>For reference, it takes roughly 2.5 months in the concrete protocol to turn over 1/3 of the validator set, given the queuing mechanism in the Ethereum implementation, which is on the order of 20000 epochs.

a complete blockchain protocol. Our goal was to separate the “mathematically clean” part of the Ethereum 2.0 design from the implementation details. Some of the analytical techniques are general enough to be useful for studying other protocols, such as the proofs of probabilistic liveness and the equivocation game construction. In practice, much of the worst-case worlds from the analysis can be patched by ideas such as in Section 8.

For transparency, it is important to note that it is possible for “patches” or even “implementation details” to change safety/liveness assumptions, or even for 2 otherwise innocuous “patches” to combine in undesirable ways. However, it is impossible to predict all the potential implementation changes the protocol will eventually receive, or which of the current proposed patches would be included or removed from implementation for possibly unrelated reasons. Thus, we wrote this paper to analyze the main theoretical foundations of the protocol and we simply sketched an analysis of each implementation detail as an isolated change instead of in conjunction with others. As the protocol solidifies, possibly with real-life testing, it would be meaningful to reassess the effects of patches that would become “core” to the protocol. In particular, some version (possibly significantly differently from what we proposed) of dynamic validator sets would be indispensable; once its design is finalized, it would be important to revisit the safety/liveness implications.

There exist many other proof-of-stake based protocols in the cryptocurrency space. Some examples include:

- Tendermint [4]: a very “pure” design that is a simplification of PBFT applied to a blockchain proof-of-stake context. This design favors safety over liveness. All validators participate in every consensus round and no consensus happens faster than those rounds, with progress only possible if  $\geq 2/3$  stake worth of validators are online.
- Casper CBC [5]: an alternative proposal to Casper FFG, focusing on mathematical correctness by construction. Casper CBC does not have fixed in-protocol thresholds, based on an emergent approach to safety arising from nodes following the majority of what other nodes have done in the past and making finality inferences about blocks. Besides being a proposed protocol, Casper CBC is also meant to be a general framework for analyzing consensus designs.
- Hotstuff (v6) [21]: has many similar properties as Casper FFG, with a flexible mathematical framework; one key difference is that Hotstuff uses an exponential backoff mechanism to be able to make progress under arbitrary network delays.
- Ouroboros [14]: a protocol under the “synchronous” school which assumes synchrony for fault-tolerance (and thus gets stronger 50% bounds for fault-tolerance). It uses the longest-chain rule and reward mechanisms to incentivize participation and prevent passive attacks. In Ouroboros Praos [14], the protocol is updated in response to vulnerabilities against “message delay” attacks. In Ouroboros Genesis in [1], the protocol is further expanded to allow for the ability for a participant to bootstrap the consensus from the genesis block (hence the name), proving globally universally composable (GUC) security against a 50% adversary.
- Snow White [9]: a protocol where the committee leader can extend the blockchain with a block that includes a reference to the previous block, similar to block dependencies, and a nonce that can be used to verify the block’s validity against some a priori difficulty constant. Valid timestamps must increase and “any timestamp in the future will cause a chain to be rejected”. Participants only accept incoming chains that have not been modified too far in the past, similar to accepting views within a certain number of epochs.
- Nxt [16]: a proof-of-stake protocol that has a finite number of tokens. Nxt uses a stake-based probability for the right to generate a block, but it imposes transaction fees to circulate currency since it does not generate new tokens. Nxt contrasts themselves from Peercoin, noting that Peercoin’s algorithm grants more power to miners who have had coins on the network for a long time. The protocol also describes certain restrictions on block creation and transferring tokens to prevent standard attacks and moving stake between accounts.
- Thunderella [18]: a proof-of-stake protocol that aims to achieve optimal transaction verification relative to message delay, assuming “good conditions” that the leader of a committee is honest and has a super-majority of honest validators. A new leader will be implemented to build upon a back-up network in the event of a stall. The protocol also allows for dynamic validators with cool-down periods.



- Dfinity [12]: beacon-notarization protocol that works with both proof-of-work and proof-of-stake. The system uses a random beacon that selects block proposers, and a decentralized notary chooses the highest-ranked block based on a criteria described in [12]. Their analogues of *validators* and *committees* are called *replicas* and *groups* respectively, and the decentralized notary depends on the same Byzantine fault-tolerance to notarize blocks and reach consensus. They only have a passive notion of finalization, but the notarization process is quite fast, and blocks do get published in a timely manner. Replicas do need permission to leave the chain.

Our goal is not to show that our design is strictly better than any of these other designs. All of these protocols contain tradeoffs based on different design goals and assumptions about the network. Our design is guided by a balance between simplicity, understandability, and practicality, with a mixed emphasis on safety and liveness. For example, if the context of a proposed blockchain prioritizes safety over speed, then one may, e.g., select or modify a proposal more in the direction of Tendermint.

## Acknowledgments

We thank the Ethereum Foundation and the San Jose State University Research Foundation for support for this program. This project is made possible by the SJSU Math and Statistics Department’s CAMCOS (Center for Applied Mathematics, Computation, and Statistics) program. We also thank Aditya Asgaonkar, Carl Beekhuizen, Dankrad Feist, Brian Gu, Brice Huang, Ryuya Nakamura, Juan Sanchez, Yi Sun, Mayank Varia, and Sebastien Zany for helpful comments. We thank Runtime Verification (Musab Alturki, Elaine Li, Daejun Park) for helpful comments and proof verification with Coq, building off work by Karl Palmskog, Lucas Pena and Brandon Moore. Author Yan X Zhang is the director of CAMCOS; authors Vitalik Buterin and Danny Ryan are members of the Ethereum Foundation.

## References

- [1] C. Badertscher, P. Gaži, A. Kiayias, A. Russell, and V. Zikas. Ouroboros genesis: Composable proof-of-stake blockchains with dynamic availability. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 913–930. ACM, 2018.
- [2] L. Baird, M. Harmon, and P. Madsen. Hedera: A public hashgraph network & governing council. 2018. <https://www.hedera.com/hh-whitepaper-v1.4-181017.pdf>.
- [3] G. Bracha and S. Toueg. Asynchronous consensus and broadcast protocols. *Journal of the Association for Computing Machinery (JACM)*, 32(4):824–840, 1985.
- [4] E. Buchman, J. Kwon, and Z. Milosevic. The latest gossip on BFT consensus. *CoRR*, abs/1807.04938, 2018.
- [5] V. Buterin. A CBC casper tutorial. 2018. [https://vitalik.ca/general/2018/12/05/cbc\\_casper.html](https://vitalik.ca/general/2018/12/05/cbc_casper.html).
- [6] V. Buterin. Sharding. [https://vitalik.ca/files/Ithaca201807\\_Sharding.pdf](https://vitalik.ca/files/Ithaca201807_Sharding.pdf), 2018.
- [7] V. Buterin and V. Griffith. Casper the friendly finality gadget. *CoRR*, abs/1710.09437, 2017.
- [8] M. Castro, B. Liskov, et al. Practical byzantine fault tolerance. In *Operating Systems Design and Implementation*, volume 99, pages 173–186, 1999.
- [9] P. Daian, R. Pass, and E. Shi. Snow white: Robustly reconfigurable consensus and applications to provably secure proofs of stake. *Cryptology ePrint Archive*, 2017.
- [10] C. Dwork, N. Lynch, and L. Stockmeyer. Consensus in the presence of partial synchrony. *Journal of the Association for Computing Machinery (JACM)*, 35(2):288–323, 1988.
- [11] Ethereum Developers. Ethereum 2.0 phase 0: The beacon chain. 2019. <https://github.com/ethereum/eth2.0-specs/blob/dev/specs/phase1/beacon-chain.md>.
- [12] T. Hanke, M. Movahedi, and D. Williams. DFINITY technology overview series, consensus system. *CoRR*, abs/1805.04548, 2018.

- [13] A. Jain, S. Arora, Y. Shukla, T. Patil, and S. Sawant-Patil. Proof of stake with casper the friendly finality gadget protocol for fair validation consensus in ethereum. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 3(3):291–298, 2018.
- [14] A. Kiayias, A. Russell, B. David, and R. Oliynykov. Ouroboros: A provably secure proof-of-stake blockchain protocol. In *Annual International Cryptology Conference*, pages 357–388. Springer, 2017.
- [15] S. Nakamoto et al. Bitcoin: A peer-to-peer electronic cash system. 2008.
- [16] Nxt Community. Nxt whitepaper. 2018. <https://nxtwiki.org/wiki/Whitepaper:Nxt>.
- [17] K. Palmskog, M. Gligoric, L. Peña, B. Moore, and G. Rosu. Verification of casper in the coq proof assistant. Technical report, 2018.
- [18] R. Pass and E. Shi. Thunderella: Blockchains with optimistic instant confirmation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 3–33. Springer, 2018.
- [19] R. J. Serfling. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, 2(1):39–48, 01 1974.
- [20] Y. Sompolinsky and A. Zohar. Secure high-rate transaction processing in bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 507–527. Springer, 2015.
- [21] M. Yin, D. Malkhi, M. K. Reiter, G. G. Gueta, and I. Abraham. Hotstuff: Bft consensus in the lens of blockchain. *CoRR*, abs/1803.05069, 2018.
- [22] V. Zamfir. Casper the friendly ghost: a ”correct-by-construction” blockchain consensus protocol. 2017. <https://github.com/vladzamfir/research/blob/master/papers/CasperTFG/CasperTFG.pdf>.

## A Technicalities of Views

In Section 2, we defined *view* in a streamlined treatment. We provide a more detailed treatment in this section that offers further intuition.

First, we give a different definition from *view* that should seem more intuitive to some readers. We define the *full view* of a validator  $V$  at a given time  $T$  to be the set of all messages (and timestamps) that have been seen by  $V$  by time  $T$ . We can denote the full view as  $\text{fv}(V, T)$ . Similarly to the network view  $\text{view}(\text{NW}, T)$ , we also define a “God’s-eye-view”  $\text{fv}(\text{NW}, T)$ , the *full network view*, as the collection of all messages any validator has broadcast at any time to the network. As with views, for any validator  $V$  and any given time  $T$ ,  $\text{fv}(\text{NW}, T)$  includes all the messages for any  $\text{fv}(V, T)$ , though the timestamps may be mismatched. With this definition, our definitions  $\text{view}(V, T)$  and  $\text{view}(\text{NW}, T)$  respectively are just subsets of the **accepted** messages in  $\text{fv}(V, T)$  and  $\text{view}(\text{NW}, T)$  respectively.

Now, it is completely possible to describe everything in terms of full views instead of views. However, there are many reasons why we work with views instead of full views in this paper (and why we leave the concept of full views to the appendix). To give a few:

- We can visualize views as a connected tree of blocks and attestations (with edges corresponding to dependencies). To visualize full views, we necessarily have to use a possibly disconnected graph.
- Motivated by real-life concerns, a protocol working with full views would have to contain instructions that differ on how to handle messages a validator  $V$  has seen (but shouldn’t act on, as it depends on a possibly nonexistent message  $V$  has not seen) versus a message a validator has accepted and fully “understands”. As an example, a cryptocurrency protocol may have to contain statements of the sort “when a full view sees a transaction about a coin  $B$ , if the parent blocks of  $B$  exist (recursively) and are all signed correctly, then consider the transaction valid.” This is very clunky, whereas if parent-child relationships were dependencies then using views abstracts away the recursive-checking logic, and we just need to say “when a view sees a transaction about a coin  $B$ , consider the transaction valid.”
- It is always possible to have a protocol that works on views: if message  $B$  depends on message  $A$  but we receive  $B$  first, then working with views is equivalent to ignoring the existence of message  $B$  when making choices as a validator until  $A$  (and other dependencies) come.

To summarize: even though the *full view* is somewhat intuitively easier to understand than the *view*, since all the dependencies are met in a view by construction, what the view sees forms a “coherent” state where the validator can reason about any message with no ambiguity, whereas reasoning about the full view will usually come with caveats about checking that all dependencies of the messages involved in the reasoning are met (probably recursively).

## B The Equivocation Game

Recall that in Section 7, we abstracted a single slot of Gasper into a one-shot game called the *equivocation game*, parametrized by  $(\mathcal{V}, a, \epsilon_1, \epsilon_2)$ , where  $\mathcal{V}$  votes on 2 options  $O_1$  and  $O_2$ . In this Appendix, we give some further details and perform some simulations.

First, we explicitly discuss how time and synchrony conditions are modelled in this game:

1. There is a single time period in this game, formalized as the real interval  $[0, 1]$ . This interval corresponds to the 1 slot of time in Gasper.
2. Honest validators follow the following protocol: “vote at [your]  $t = 0.5$  for the option with more total stake voted in your view; in the case of a tie, vote  $O_1$ .” This protocol corresponds to the instruction in Gasper that attestors for slot  $i$  are supposed to attest at time  $i + 1/2$  (and tiebreakers are broken by a hash, which is an arbitrary but fixed value).
3. Synchrony model:

- (a) We assume that all validators have perfectly synced clocks, but each validator attempting to vote at time  $t$  actually votes at “real” time  $t + X$  (rounding to inside the interval  $[0, 1]$ ), where  $X$  is a uniform random variable with support  $[-\epsilon_1, \epsilon_1]$ , independently re-sampled for each message. We can think of  $\epsilon_1$  as a “timing error” bound, accounting for clock differences, client-side timing issues, etc.
- (b) When a validator votes at “real” time  $t'$ , all other validators obtain the vote at time  $t' + a + Y$ , where  $Y$  is a uniformly distributed variable with support  $[-\epsilon_2, \epsilon_2]$ . We can think of  $a$  as the average delay per message, and  $\epsilon_2$  a noise on top of the delay. Combining with the previous point, we have that a validator attempting to vote at time  $t$  actually has his/her vote received by another validator at  $t + a + X + Y$ , where the  $X$  and  $Y$  are independently re-sampled for each event.

For our analysis in this section, we make the following additional assumptions and notations:

- Every validator has exactly 1 unit of stake. Barring further knowledge (such as empirical data) about validator stake distributions, this is the most intuitive choice for our toy model.
- We define  $N_h \geq 2N/3$  to be the total number of honest validators (which is equivalent to their total amount of stake). In reality (and in Section 7), we expect  $N_h$  to be a bit bigger at  $2N/3 + \epsilon N$  for small  $\epsilon$ .
- Similarly, we define  $N_b \leq N/3$  to be the number of byzantine validators. We use  $p = \frac{N_b}{N}$  to be the proportion of byzantine validators.

Finally, for our computer simulations, we additionally set the parameters to  $\mathcal{V} = 111$ ,  $N_h = 74$ ,  $N_b = 37$ . The total number of validators of 111 is based on [6], as a heuristic lower bound for making the committees safe. Erring on the conservative side, we pick the maximally pessimistic parameter of  $p = 1/3$  fraction of byzantine validators. Having more honest validators than this assumption significantly improves our chances of winning.

### B.1 The Pessimistic Regime - High Latency

This pessimistic regime is a mental experiment to show that under certain conditions fair to both sides, the power of collusion is enough to make the game favor the dishonest side.

We assume that  $a$  (the delay for messages from one validator to another) is big compared to  $\epsilon_1$  (the error in timing a vote). In this situation, we argue that the dishonest validators can perform the following “smoke bomb attack”: at time close to  $0.5 - a$ , all the dishonest validators vote in a way that the votes are split between the 2 options, for a total of  $pN/2$  votes for each option, all with fake timestamps for time close to 0.5.

When this attack activates, from the view of each honest validator, at time 0.5 they will see some random votes from the dishonest validators. As  $a$  is big compared to  $\epsilon_1$ , they almost never see other votes from honest validators. Thus, the dishonest validators and honest validators both end up splitting the vote: the dishonest validators end up splitting  $pN$  of the votes, and the honest validators split the remaining  $(1 - p)N$  probabilistically.

Assuming the honest validators all have 1 stake, the honest validator’s votes can be modeled by  $(1 - p)N$  Bernoulli trials between the two outcomes, ending up close to a normal distribution around a tie, with a standard deviation proportional to  $\sqrt{(1 - p)N}$  votes. In particular, for large  $N$ , it is very unlikely for one of the options to get anywhere near  $2N/3$  stake worth of votes.

**Remark.** As an extension of this example, we can also consider the case where the protocol tiebreaker is “flip a coin,” the situation becomes even worse: the dishonest validators can simply wait until the end to vote. Because  $a$  is big compared to  $\epsilon_1$ , the honest validators are still essentially making coin flips since they have not seen any other votes. Our tiebreaker rule (vote for option  $O_1$  if there is a tie) solves this problem. The main takeaway here is that these protocols (including attacks on them) are very sensitive to very small implementation details.

In Figure 11, we have four cases that demonstrate different outcomes of the pessimistic regime with  $a = 0.15$ ,  $\epsilon_1 = 0.05$ , and  $\epsilon_2 = 0.15$ . This ensures that messages sent at time  $t$  are received uniformly randomly in the interval  $[t, t + 0.3]$ . Note that having smaller  $\epsilon_2$  would be even more pessimistic, almost ensuring that honest validators cannot see each other’s votes.

In Case 1, dishonest validators vote much earlier than honest validators do, so most of the attestations made by the former are visible to the latter when they vote. Thus, the honest validators are likely to vote for the same option. For Cases 2 and 3, dishonest validators vote closer and closer to when the honest validators do, which increases the power of the attack; in Case 3 dishonest validators win at 42% of the time. In Case 4, dishonest validators vote at the same time as honest validators; because of the delay, the dishonest votes do not confuse the honest validators, so honest validators almost always win (because they are likely to vote for the default option  $O_1$ , barring seeing any votes).

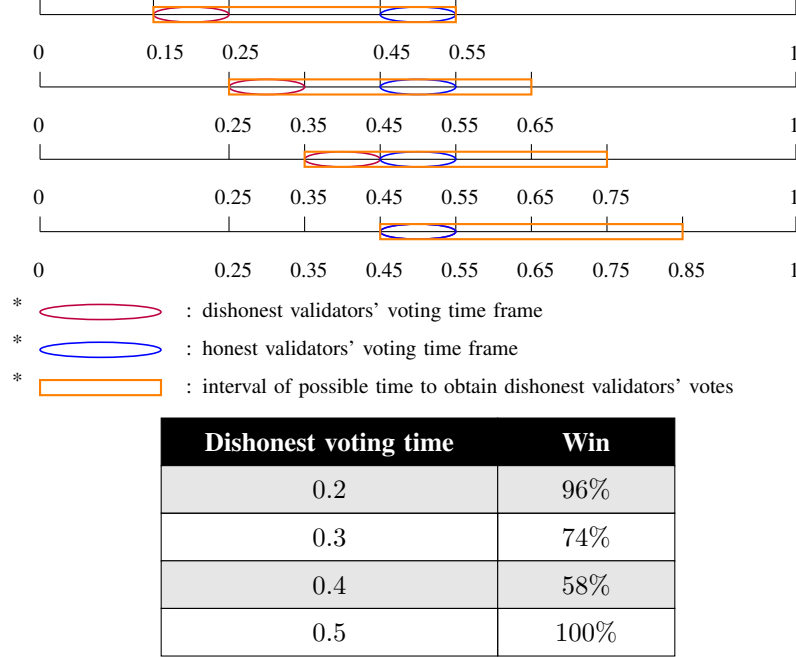


Figure 11: Pessimistic regime simulations with  $a = 0.15$ ,  $\epsilon_1 = 0.05$ ,  $\epsilon_2 = 0.15$  for the 4 cases above in order. Each row has a different coordinated dishonest voting time, which affects the winning rates of honest validators.

## B.2 The Optimistic Regime - Low Latency

For this regime, we go to the other extreme and assume perfect synchrony ( $a = \epsilon_2 = 0$ ). This means all decisions are immediately propagated to all other validators on the network.

In this game, the moment one choice has a lead (by default,  $O_1$ ), everybody is aware of the lead and all honest validators will vote for the choice in the lead. Thus, the optimal strategy for dishonest validators is to keep the voting exactly at a tie (in which case the next honest validator will vote for  $O_1$ ) or such that  $O_2$  has one extra vote (in which case the next validator will vote  $O_2$ ). However, as it takes more votes to get the count to under  $O_1$ , the dishonest validators really can do no better than as if they voted at the very end in the wrong direction, which still gives the honest validators  $(1 - p)N$  votes in the correct direction. In our situation where  $p = 1/3$ , this gives a total vote differential of  $(1 - 2p)N = N/3$ , which is good for us. One can see a simulated outcome of this case in Figure 12.

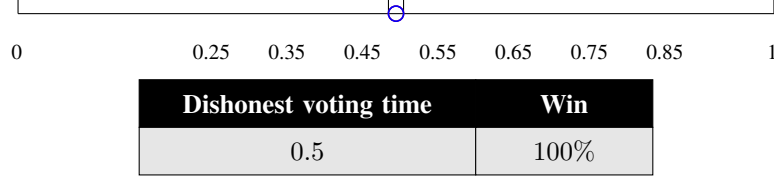


Figure 12: Optimistic regime simulation outcome with  $a = 0$ ,  $\epsilon_1 = 0.05$ ,  $\epsilon_2 = 0$  with both dishonest and honest validators voting in the same time frame. Honest validators always win.

### B.3 An Example Inbetween

We make a specific set of assumptions somewhere between the extremes presented in sections B.1 and B.2. These assumptions are fairly arbitrary and hopefully presents a realistic example. It would be good to find principled “bottom truth” parameters somehow (such as after the actual blockchain is implemented).

In Figure 13, we have three cases that demonstrate the outcomes of the inbetween examples with  $a = 0.1$ ,  $\epsilon_1 = 0.05$  and  $\epsilon_2 = 0.1$ . Thus,  $a$  is comparable to  $\epsilon_1$  (avoiding the pessimistic case) but fairly big (avoiding the optimistic case). Making  $\epsilon_2$  bigger doesn’t really change the results much until  $\epsilon_2$  becomes around the order of  $a$ .

In Case 1, some dishonest validators vote earlier than honest validators do and their “smoke bomb attack” has some impact; however, the amount of votes are not distracting enough compared to the amount of votes honest validators are able to vote following the protocol. In case 2, as we earlier discussed in the pessimistic regime, dishonest validators vote closer to when the honest validators do, which boosts the effectiveness of the attack. In Case 3, dishonest validators vote at the same time as honest validators do; as in the pessimistic regime, honest validators still almost always win.

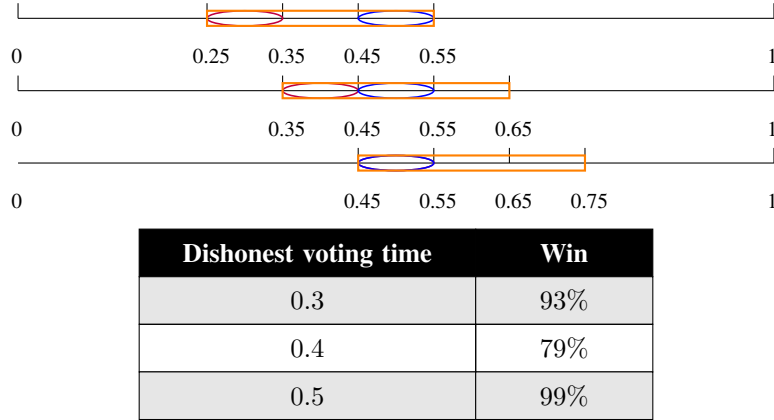


Figure 13: Simulation outcomes with  $a = 0.1$ ,  $\epsilon_1 = 0.05$ ,  $\epsilon_2 = 0.1$ .