

Availability & Validity

Protocol Specification

Contents

0.1	Introduction	4
0.2	Preliminaries	5
0.2.1	Extra Validation Data	5
0.3	Overall process	7
0.4	Candidate Selection	7
0.5	Candidate Backing	9
0.5.1	Inclusion of candidate receipt on the relay chain	12
0.6	PoV Distribution	13
0.6.1	Primary Validation Disagreement	13
0.7	Availability	13
0.8	Distribution of Chunks	16
0.9	Announcing Availability	16
0.9.1	Processing on-chain availability data	17
0.10	Publishing Attestations	18
0.11	Secondary Approval checking	18
0.11.1	Approval Checker Assignment	18
0.11.2	VRF computation	18
0.11.3	One-Shot Approval Checker Assignemnt	18
0.11.4	Extra Approval Checker Assigment	19
0.11.5	Additional Checking in Case of Equivocation	19
0.12	The Approval Check	20
0.12.1	Verification	21
0.12.2	Process validity and invalidity messages	21
0.12.3	Invalidity Escalation	22

0.1 Introduction

Validators are responsible for guaranteeing the validity and availability of PoV blocks. There are two phases of validation that takes place in the AnV protocol.

The primary validation check is carried out by parachain validators who are assigned to the parachain which has produced the PoV block as described in Section 0.4. Once parachain validators have validated a parachain's PoV block successfully, they have to announce that according to the procedure described in Section 0.5 where they generate a statement that includes the parachain header with the new state root and the XCMP message root. This candidate receipt and attestations, which carries signatures from other parachain validators is put on the relay chain.

As soon as the proposal of a PoV block is on-chain, the parachain validators break the PoV block into erasure-coded chunks as described in Section 19 and distribute them among all validators. See Section 0.8 for details on how this distribution takes place.

Once validators have received erasure-coded chunks for several PoV blocks for the current relay chain block (that might have been proposed a couple of blocks earlier on the relay chain), they announce that they have received the erasure coded chunks on the relay chain by voting on the received chunks, see Section 0.9 for more details.

As soon as $> 2/3$ of validators have made this announcement for any parachain block we *act on* the parachain block. Acting on parachain blocks means we update the relay chain state based on the candidate receipt and considered the parachain block to have happened on this relay chain fork.

After a certain time, if we did not collect enough signatures approving the availability of the parachain data associated with a certain candidate receipt we decide this parachain block is unavailable and allow alternative blocks to be built on its parent parachain block, see ??.

The secondary check described in Section 0.11, is done by one or more randomly assigned validators to make sure colluding parachain validators may not get away with validating a PoV block that is invalid and not keeping it available to avoid the possibility of being punished for the attack.

During any of the phases, if any validator announces that a parachain block is invalid then all validators obtain the parachain block and check its validity, see Section ?? for more details.

All validity and invalidity attestations go onto the relay chain, see Section 0.10 for details. If a parachain block has been checked at least by certain number of validators, the rest of the validators continue with voting on that relay chain block in the GRANDPA protocol. Note that the block might be challenged later.

0.2 Preliminaries

Definition 1 *The Polkadot project uses the **SCALE** codec to encode common data types such as integers, byte arrays, varying data types as well as other data structure. The SCALE codec is defined in a separate document known as "The Polkadot Host - Protocol Specification". **TODO: @fabio: link to document***

Definition 2 *In the remainder of this chapter we assume that ρ is a Polkadot Parachain and B is a block which has been produced by ρ and is supposed to be approved to be ρ 's next block. By R_ρ we refer to the **validation code** of parachain ρ as a WASM blob, which is responsible for validating the corresponding Parachain's blocks.*

Definition 3 *The witness proof of block B , denoted by π_B , is the set of all the external data which has gathered while the ρ runtime executes block B . The data suffices to re-execute R_ρ against B and achieve the final state indicated in the $H(B)$.*

This witness proof consists of light client proofs of state data that are generally Merkle proofs for the parachain state trie. We need this because validators do not have access to the parachain state, but only have the state root of it.

Definition 4 *Accordingly we define the **proof of validity block** or **PoV** block in short, PoV_B , to be the tuple:*

$$(B, \pi_B)$$

*A PoV block is an extracted Merkle subtree, attached to the block. **TODO: @fabio: clarify this***

0.2.1 Extra Validation Data

Validators must submit extra validation data to Runtime R_ρ in order to build candidates, to fully validate those and to vote on their availability. Depending on the context, different types of information must be used.

Parachain validators get this extra validation data from the current relay chain state. Note that a PoV block can be paired with different extra validation data depending on when and which relay chain fork it is included in. Future validators would need this extra validation data because since the candidate receipt as defined in Definition 13 was included on the relay chain the needed relay chain state may have changed.

Definition 5 R_ρ^{up} is an varying data type (Def. 1) which implies whether the parachain is allowed to upgrade its validation code.

$$R_\rho^{up} := Option(H_i(B_{chain}^{relay}) + n)$$

TODO: @fabio: adjust formula?

*If this is Some, it contains the number of the minimum relay chain height at which the upgrade will be applied, assuming an upgrade is currently signaled **TODO: @fabio: where is this signaled?**. A parachain should enact its side of the upgrade at the end of the first parachain block executing in the context of a relay-chain block with at least this height. This may be equal to the current perceived*

relay-chain block height, in which case the code upgrade should be applied at the end of the signaling block.

Definition 6 The **validation parameters**, v_B^{VP} , is an extra input to the validation function, i.e. additional data from the relay chain state that is needed. It's a tuple of the following format:

$$vp_B := (B, \text{head}(B_p), v_B^{GVS}, R_p^{up})$$

where each value represents:

- B : the parachain block itself.
- $\text{head}(B_p)$: the parent head data (Def. 12) of block B .
- v_B^{GVP} : the global validation parameters (7).
- R_p^{up} : implies whether the parachain is allowed to upgrade its validation code (Def. 5).

Definition 7 The **global validation parameters**, v_B^{GVP} , defines global data that apply to all candidates in a block.

$$v_B^{GVS} := (Max_{size}^R, Max_{size}^{head}, H_i(B_{chain}^{relay}))$$

where each value represents:

- Max_{size}^R : the maximum amount of bytes of the parachain Wasm code permitted.
- Max_{size}^{head} : the maximum amount of bytes of the head data (Def. 12) permitted.
- $H_i(B_{chain}^{relay})$: the relay chain block number this is in the context of.

Definition 8 The **local validation parameters**, v_B^{LVP} , defines parachain-specific data required to fully validate a block. It is a tuple of the following format:

$$v_B^{LVP} := (\text{head}(B_p), \text{UINT128}, \text{Blake2b}(R_p), R_p^{up})$$

where each value represents:

- $\text{head}(B_p)$: the parent head data (Def. 12) of block B .
- UINT128 : the balance of the parachain at the moment of validation.
- $\text{Blake2b}(R_p)$: the Blake2b hash of the validation code used to execute the candidate.
- R_p^{up} : implies whether the parachain is allowed to upgrade its validation code (Def. 5).

Definition 9 The **validation result**, r_B , is returned by the validation code R_p if the provided candidate is valid. It is a tuple of the following format:

$$r_B := (\text{head}(B), \text{Option}(P_p^B), (\text{Msg}_0, \dots, \text{Msg}_n), \text{UINT32})$$

$$\text{Msg} := (\mathbb{O}, \text{Enc}_{SC}(b_0, \dots, b_n))$$

where each value represents:

- $head(B)$: the new head data (Def. 12) of block B .
- $Option(P_\rho^B)$: a varying data (Def. 1) containing an update to the validation code that should be scheduled in the relay chain.
- Msg : parachain "upward messages" to the relay chain. \mathbb{O} identifies the origin of the messages and is a varying data type (Def. 1) and can be one of the following values:

$$\mathbb{O} = \begin{cases} 0, & \text{Signed} \\ 1, & \text{Parachain} \\ 2, & \text{Root} \end{cases}$$

TODO: @fabio: define the concept of "origin"

The following SCALE encoded array, $Enc_{SC}(b_0, ..b_n)$, contains the raw bytes of the message which varies in size.

- $UINT32$: number of downward messages that were processed by the Parachain. It is expected that the Parachain processes them from first to last.

Definition 10 Accordingly we define the **erasure-encoded blob** or **blob** in short, \bar{B} , to be the tuple:

$$(B, \pi_B, v_B^{GVP}, v_B^{LVP})$$

where each value represents:

- B : the parachain block.
- π_B : the witness data.
- v_B^{GVP} : the global validation parameters (Def. 7).
- v_B^{LVP} : the local validation parameters (Def. 8).

Note that in the code the blob is referred to as "AvailableData".

0.3 Overall process

The Figure 0.3 demonstrates the overall process of assuring availability and validity in Polkadot
TODO: complete the Diagram.

0.4 Candidate Selection

Collators produce candidates (Def. 11) and send those to validators. Validators verify the validity of the received candidates (Algo. 1) by executing the validation code, R_ρ , and issue statements (Def. 16) about the candidates to connected peers. The validator ensures that every candidate considered for inclusion has at least one other validator backing it. Candidates without backing are discarded.

```

Restricted shell escape. PlantUML cannot be called. Start pdflatex/lualatex with -shell-escape.
@startuml
(*) -i "math_i Parachain Collator C_rho Generates B and PoV_B/math_i" -i "math_i C_rho sends PoV_B
to rho's validator V_rho/math_i" -i "math_i V_rho runs rho's runtime on PoV/math_i" if "math_i PoV_B is
valid/math_i" then -i[true] if "math_i V_rho have seen the CandidateReceipt for PoV_B/math_i" then
-i[true] Sign CandidateReceipt -i[Ending process] (*)
else -i [False] "Gerenate CandiateReceipt" -i[Ending process] (*)
endif else -i[false] "math_i Broadcast message of invalidity for PoV_B/math_i" end if
-i[Ending process] (*)
@enduml

```

Figure 1: Overall process to acheive availability and validity in Polkadot

The validator must keep track of which candidates were submitted by collators, including which validators back those candidates in order to penalize bad behavior. This is described in more detail in section 0.5

Definition 11 A *candidate*, $C_{coll}(PoV_B)$, is issues by collators and contains the PoV block and enough data in order for any validator to verify its validity. A candidate is a tuple of the following format:

$$C_{coll}(PoV_B) := (id_p, h_b(B_{relay_parent}), id_C, Sig_{SR25519}^{Collator}, head(B), h_b(PoV_B))$$

where each value represents:

- id_p : the Parachain Id this candidate is for.
- $h_b(B_{relay_parent})$: the hash of the relay chain block that this candidate should be executed in the context of.
- id_C : the Collator relay-chain account ID as defined in Definition *TODO: @fabio*.
- $Sig_{SR25519}^{Collator}$: the signature on the 256-bit Blake2 hash of the block data by the collator.
- $head(B)$: the head data (Def. 12) of block B .
- $h_b(PoV_B)$: the 32-byte Blake2 hash of the PoV block.

Definition 12 The *head data*, $head(B)$, of a parachain block is a tuple of the following format:

$$head(B) := (H_i(B), H_p(B), H_r(B))$$

Where $H_i(B)$ is the block number of parachain block B , $H_p(B)$ is the 32-byte Blake2 hash of the parent block header and $H_r(B)$ represents the root of the post-execution state. *TODO: @fabio: clarify if H_p is the hash of the header or full block TODO: @fabio: maybe define those symbols at the start (already defined in the Host spec)?*

Algorithm 1 PRIMARYVALIDATION**Input:** B, π_B , relay chain parent block B_{parent}^{relay}

- 1: Retrieve v_B from the relay chain state at B_{parent}^{relay}
- 2: Run Algorithm 2 using B, π_B, v_B

Algorithm 2 VALIDATEBLOCK**Input:** B, π_B, v_B

- 1: retrieve the runtime code R_ρ that is specified by v_B from the relay chain state.
- 2: check that the initial state root in π_B is the one claimed in v_B
- 3: Execute R_ρ on B using π_B to simulate the state.
- 4: If the execution fails, return fail.
- 5: Else return success, the new header data h_B and the outgoing messages M . **TODO: @fabio: same as head data?**

0.5 Candidate Backing

Validators back the validity respectively the invalidity of candidates by extending those into candidate receipts as defined in Definition 13 and communicate those receipts by issuing statements as defined in Definition 16. Validator v needs to perform Algorithm 3 to announce the statement of primary validation to the Polkadot network. If the validator receives a statement from another validator, the candidate is confirmed based on algorithm 4.

As algorithm 3 and 4 clarifies, the validator should blacklist collators which send invalid candidates and announce this misbehavior. If another validator claims that an invalid candidates is actually valid, that misbehavior must be announced, too. **TODO: @fabio**

The validator tries to back as many candidates as it can, but does not attempt to prioritize specific candidates. Each validator decides on its own - on whatever metric - which candidate will ultimately get included in the block.

Definition 13 A *candidate receipt*, $C_{receipt}(PoV_B)$, is an extension of a candidate as defined in Definition 11 which includes additional information about the validator which verified the PoV block. The candidate receipt is communicated to other validators by issuing a statement as defined in Definition 16.

This type is a tuple of the following format:

$$C_{receipt}(PoV_B) := (id_p, h_b(B_{parent}^{relay}), head(B), id_C, Sig_{SR25519}^{Collator}, h_b(PoV_B), Blake2b(CC(PoV_B)))$$

where each value represents:

- id_p : the Parachain Id this candidate is for.
- $h_b(B_{parent}^{relay})$: the hash of the relay chain block that this candidate should be executed in the context of.

- $\text{head}(B)$: the head data (Def. 12) of block B . *TODO: @fabio (collator module relevant?)*.
- id_C : the collator relay-chain account ID as defined in Definition *TODO: @fabio*.
- $\text{Sig}_{\text{SR25519}}^{\text{Collator}}$: the signature on the 256-bit Blake2 hash of the block data by the collator.
- $h_b(\text{PoV}_B)$: the hash of the PoV block.
- $\text{Blake2b}(\text{CC}(\text{PoV}_B))$: The hash of the commitments made as a result of validation, as defined in Definition 14.

Definition 14 *Candidate commitments*, $\text{CC}(\text{PoV}_B)$, are results of the execution and validation of parachain (or parathread) candidates whose produced values must be committed to the relay chain. A candidate commitments is represented as a tuple of the following format:

$$\begin{aligned}\text{CC}(\text{PoV}_B) &:= (\mathbb{F}, \text{Enc}_{SC}(\text{Msg}_0, \dots, \text{Msg}_n), H_r(B), \text{Option}(R_\rho)) \\ \text{Msg} &:= (\mathbb{O}, \text{Enc}_{SC}(b_0, \dots, b_n))\end{aligned}$$

where each value represents:

- \mathbb{F} : fees paid from the chain to the relay chain validators.
- Msg : parachain messages to the relay chain. \mathbb{O} identifies the origin of the messages and is a varying data type (Def. 1) and can be one of the following values:

$$\mathbb{O} = \begin{cases} 0, & \text{Signed} \\ 1, & \text{Parachain} \\ 2, & \text{Root} \end{cases}$$

TODO: @fabio: define the concept of "origin"

The following SCALE encoded array, $\text{Enc}_{SC}(b_0, \dots, b_n)$, contains the raw bytes of the message which varies in size.

- $H_r(B)$: the root of a block's erasure encoding Merkle tree *TODO: @fabio: use different symbol for this?*.
- $\text{Option}(R_\rho)$: A varying datatype (Def. 1) containing the new runtime code for the parachain. *TODO: @fabio: clarify further*

Definition 15 A *Gossip PoV block* is a tuple of the following format:

$$(h_b(B_{\text{relay}_{\text{parent}}}), h_b(\text{C}_{\text{coll}}(\text{PoV}_B)), \text{PoV}_B)$$

where $h_b(B_{\text{relay}_{\text{parent}}})$ is the block hash of the relay chain being referred to and $h_b(\text{C}_{\text{coll}}(\text{PoV}_B))$ is the hash of some candidate localized to the same Relay chain block.

Definition 16 A **statement** notifies other validators about the validity of a PoV block. This type is a tuple of the following format:

$$(Stmt, id_V, Sig_{SR25519}^{Validator})$$

where $Sig_{SR25519}^{Validator}$ is the signature of the validator and id_V refers to the index of validator according to the authority set. *TODO: @fabio: define authority set (specified in the Host spec).* Stmt refers to a statement the validator wants to make about a certain candidate. Stmt is a varying data type (Def. 1) and can be one of the following values:

$$Stmt = \begin{cases} 0, & \text{Seconded, followed by: } C_{receipt}(PoV_B) \\ 1, & \text{Validity, followed by: } Blake2(C_{coll}(PoV_B)) \\ 2, & \text{Invalidity, followed by: } Blake2(C_{coll}(PoV_B)) \end{cases}$$

The main semantic difference between ‘Seconded’ and ‘Valid’ comes from the fact that every validator may second only one candidate per relay chain block; this places an upper bound on the total number of candidates whose validity needs to be checked. A validator who seconds more than one parachain candidate per relay chain block is subject to slashing.

Validation does not directly create a seconded statement, but is rather upgraded by the validator when it choses to back a valid candidate as described in Algorithm 3.

Algorithm 3 PRIMARYVALIDATIONANNOUNCEMENT

Input: PoV_B

```

1: Init Stmt;
2: if VALIDATEBLOCK( $PoV_B$ ) is valid then
3:   Stmt  $\leftarrow$  SETVALID( $PoV_B$ )
4: else
5:   Stmt  $\leftarrow$  SETINVALID( $PoV_B$ )
6:   BLACKLISTCOLLATOROF( $PoV_B$ )
7: end if
8: PROPAGATE(Stmt)

```

- VALIDATEBLOCK: Validates PoV_B as defined in Algorithm 2.
- SETVALID: Creates a valid statement as defined in Definition 16.
- SETINVALID: Creates an invalid statement as defined in Definition 16.
- BLACKLISTCOLLATOROF: blacklists the collator which sent the invalid PoV block, preventing any new PoV blocks from being received. The amount of time for blacklisting is unspecified.
- PROPAGATE: sends the statement to the connected peers.

Algorithm 4 CONFIRMCANDIDATERECEIPT

Input: $Stmt_{peer}$

```

1: Init  $Stmt$ ;
2:  $PoV_B \leftarrow \text{RETRIEVE}(Stmt_{peer})$ 
3: if  $\text{VALIDATEBLOCK}(PoV_B)$  is valid then
4:   if  $\text{ALREADYSECONDED}(B_{chain}^{relay})$  then
5:      $Stmt \leftarrow \text{SETVALID}(PoV_B)$ 
6:   else
7:      $Stmt \leftarrow \text{SETSECONDED}(PoV_B)$ 
8:   end if
9: else
10:   $Stmt \leftarrow \text{SETINVALID}(PoV_B)$ 
11:   $\text{ANNOUNCEMISBEHAVIOROF}(PoV_B)$ 
12: end if
13:  $\text{PROPAGATE}(Stmt)$ 

```

- $Stmt_{peer}$: a statement received from another validator.
- **RETRIEVE**: Retrieves the PoV block from the statement (16).
- **VALIDATEBLOCK**: Validates PoV_B as defined in Algorithm 2.
- **ALREADYSECONDED**: Verifies if a parachain block has already been seconded for the given Relay Chain block. Validators that second more than one (1) block per Relay chain block are subject to slashing. More information is available in Definition 16.
- **SETVALID**: Creates a valid statement as defined in Definition 16.
- **SETSECONDED**: Creates a seconded statement as defined in Definition 16. Seconding a block should ensure that the next call to **ALREADYSECONDED** reliably affirms this action.
- **SETINVALID**: Creates an invalid statement as defined in Definition 16.
- **BLACKLISTCOLLATOROF**: blacklists the collator which sent the invalid PoV block, preventing any new PoV blocks from being received. The amount of time for blacklisting is unspecified.
- **ANNOUNCEMISBEHAVIOROF**: announces the misbehavior of the validator who claimed a valid statement of invalid PoV block as described in algorithm **TODO: @fabio**.
- **PROPAGATE**: sends the statement to the connected peers.

0.5.1 Inclusion of candidate receipt on the relay chain

TODO: @fabio: should this be a subsection?

Definition 17 *Parachain Block Proposal*, noted by P_ρ^B is a candidate receipt for a parachain block B for a parachain ρ along with signatures for at least $2/3$ of \mathcal{V}_ρ .

A block producer which observe a Parachain Block Proposal as defined in definition 17 may/should include the proposal in the block they are producing according to Algorithm 5 during block production procedure.

Algorithm 5 INCLUDEPARACHAINPROPOSAL(P_ρ^B)

Input:

1: TBS

0.6 PoV Distribution

TODO: @fabio

0.6.1 Primary Validation Disagreement

Parachain validators need to keep track of candidate receipts (see Definition 13) and validation failure messages of their peers. In case, there is a disagreement among the parachain validators about \bar{B} , all parachain validators must invoke Algorithm 6

Algorithm 6 PRIMARYVALIDATIONDISAGREEMENT

Input:

1: TBS

0.7 Availability

Backed candidates must be widely available for the entire, elected validators set without requiring each of those to maintain a full copy. PoV blocks get broken up into erasure-encoded chunks and each validators keep track of how those chunks are distributed among the validator set. When a validator has to verify a PoV block, it can request the chunk for one of its peers.

Definition 18 The **erasure encoder/decoder** $encode_{k,n}/decoder_{k,n}$ is defined to be the Reed-Solomon encoder defined in [?].

Algorithm 7 ERASURE-ENCODE

Input: \bar{B} : blob defined in Definition 10

```

1: Init  $Shards \leftarrow \text{MAKE-SHARDS}(\mathcal{V}_\rho, v_B)$ 

   // Create a trie from the shards in order generate the trie nodes
   // which are required to verify each chunk with a Merkle root.
2: Init  $Trie$ 
3: Init  $index = 0$ 
4: for  $shard \in Shards$  do
5:    $\text{INSERT}(Trie, index, \text{BLAKE2}(shard))$ 
6:    $index = index + 1$ 
7: end for

   // Insert individual chunks into collection (def. 19).
8: Init  $Er_B$ 
9: Init  $index = 0$ 
10: for  $shard \in Shards$  do
11:   Init  $nodes \leftarrow \text{GET-NODES}(Trie, index)$ 
12:    $\text{ADD}(Er_B, (shard, index, nodes))$ 
13:    $index = index + 1$ 
14: end for

15: return  $Er_B$ 

```

- $\text{MAKE-SHARDS}(\cdot)$: return shards for each validator as described in algorithm 8. Return value is defined as $(\mathbb{S}_0, \dots, \mathbb{S}_n)$ where $\mathbb{S} := (b_0, \dots, b_n)$
- $\text{INSERT}(trie, key, val)$: insert the given *key* and *value* into the *trie*.
- $\text{GET-NODES}(trie, key)$: based on the *key*, return all required *trie* nodes in order to verify the corresponding value for a (unspecified) Merkle root. Return value is defined as $(\mathbb{N}_0, \dots, \mathbb{N}_n)$ where $\mathbb{N} := (b_0, \dots, b_n)$.
- $\text{ADD}(sequence, item)$: add the given *item* to the *sequence*.

Algorithm 8 MAKE-SHARDS

Input: \mathcal{V}_ρ, v_B
 // Calculate the required values for Reed-Solomon.
 // Calculate the required lengths.
 1: **Init** $Shard_{data} = \frac{(|\mathcal{V}_\rho|-1)}{3} + 1$
 2: **Init** $Shard_{parity} = |\mathcal{V}_\rho| - \frac{(|\mathcal{V}_\rho|-1)}{3} - 1$

 3: **Init** $base_{len} = \begin{cases} 0 & \text{if } |\mathcal{V}_\rho| \bmod Shard_{data} = 0 \\ 1 & \text{if } |\mathcal{V}_\rho| \bmod Shard_{data} \neq 0 \end{cases}$
 4: **Init** $Shard_{len} = base_{len} + (base_{len} \bmod 2)$

 // Prepare shards, each padded with zeroes.
 // $Shards := (\mathbb{S}_0, \dots, \mathbb{S}_n)$ where $\mathbb{S} := (b_0, \dots, b_n)$
 5: **Init** $Shards$
 6: **for** $n \in (Shard_{data} + Shard_{parity})$ **do**
 7: ADD($Shards, (0_0, \dots, 0_{Shard_{len}})$)
 8: **end for**

 // Copy shards of v_b into each shard.
 9: **for** $(chunk, shard) \in (TAKE(Enc_{SC}(v_B), Shard_{len}), Shards)$ **do**
 10: **Init** $len \leftarrow \text{MIN}(Shard_{len}, |chunk|)$
 11: $shard \leftarrow \text{COPY-FROM}(chunk, len)$
 12: **end for**

 // $Shards$ contains split shards of v_B .
 13: **return** $Shards$

- ADD($sequence, item$): add the given $item$ to the $sequence$.
- TAKE($sequence, len$): iterate over len amount of bytes from $sequence$ on each iteration. If the $sequence$ does not provide len number of bytes, then it simply uses what's available.
- MIN($num1, num2$): return the minimum value of $num1$ or $num2$.
- COPY-FROM($source, len$): return len amount of bytes from $source$.

Definition 19 The collection of erasure-encoded chunks of \bar{B} , denoted by:

$$Er_B := (e_1, \dots, e_n)$$

is defined to be the output of the Algorithm 7. Each chunk is a tuple of the following format:

$$\begin{aligned} e &:= (\mathbb{S}, I, (\mathbb{N}_0, \dots, \mathbb{N}_n)) \\ \mathbb{S} &:= (b_0, \dots, b_n) \\ \mathbb{N} &:= (b_0, \dots, b_n) \end{aligned}$$

where each value represents:

- \mathbb{S} : a byte array containing the erasure-encoded shard of data.
- I : the unsigned 32-bit integer representing the index of this erasure-encoded chunk of data.
- $(\mathbb{N}_0, \dots, \mathbb{N}_n)$: an array of inner byte arrays, each containing the nodes of the Trie in order to verify the chunk based on the Merkle root.

0.8 Distribution of Chunks

Following the computation of Er_B , v must construct the \bar{B} Availability message defined in Definition 20. And distribute them to target validators designated by the Availability Networking Specification [?].

Definition 20 *PoV erasure chunk message $M_{PoV_{\bar{B}}}(i)$ is TBS*

0.9 Announcing Availability

When validator v receives its designated chunk for \bar{B} it needs to broadcast Availability vote message as defined in Definition 21

Definition 21 *Availability vote message M_{PoV}^{Avail, v_i} TBS*

Some parachains have blocks that we need to vote on the availability of, that is decided by $> 2/3$ of validators voting for availability. For 100 parachain and 1000 validators this will involve putting 100k items of data and processing them on-chain for every relay chain block, hence we want to use bit operations that will be very efficient. We describe next what operations the relay chain runtime uses to process these availability votes.

Definition 22 *An **availability bitfield** is signed by a particular validator about the availability of pending candidates. It's a tuple of the following format:*

(u32, ...)

TODO: @fabio

For each parachain, the relay chain stores the following data:

1) availability status, 2) candidate receipt, 3) candidate relay chain block number where availability status is one of {no candidate, to be determined, unavailable, available} .

For each block, each validator v signs a message

Sign(bitfield b_v , block hash h_b)

where the i th bit of b_v is 1 if and only if

1. the availability status of the candidate receipt is "to be determined" on the relay chain at block hash h_b **and**
2. v has the erasure coded chunk of the corresponding parachain block to this candidate receipt.

These signatures go into a relay chain block.

0.9.1 Processing on-chain availability data

This section explains how the availability attestations stored on the relay chain, as described in Section ??, are processed as follows:

Algorithm 9 Relay chain's signature processing

- 1: The relay chain stores the last vote from each validator on chain. For each new signature, the relay chain checks if it is for a block in this chain later than the last vote stored from this validator. If it is the relay chain updates the stored vote and updates the bitfield b_v and block number of the vote.
 - 2: For each block within the last t blocks where t is some timeout period, the relay chain computes a bitmask bm_n (n is block number). This bitmask is a bitfield that represents whether the candidate considered in that block is still relevant. That is the i th bit of bm_n is 1 if and only if for the i th parachain, (a) the availability status is to be determined and (b) candidate block number $\leq n$
 - 3: The relay chain initialises a vector of counts with one entry for each parachain to zero. After executing the following algorithm it ends up with a vector of counts of the number of validators who think the latest candidates is available.
 1. The relay chain computes b_v and bm_n where n is the block number of the validator's last vote
 2. For each bit in b_v and bm_n
 - add the i th bit to the i th count.
 - 4: For each count that is $> 2/3$ of the number of validators, the relay chain sets the candidates status to "available". Otherwise, if the candidate is at least t blocks old, then it sets its status to "unavailable".
 - 5: The relay chain acts on available candidates and discards unavailable ones, and then clears the record, setting the availability status to "no candidate". Then the relay chain accepts new candidate receipts for parachains that have "no candidate: status and once any such new candidate receipts is included on the relay chain it sets their availability status as "to be determined".
-

Based on the result of Algorithm ?? the validator node should mark a parachain block as either available or eventually unavailable according to definitions 23 and ??

Definition 23 *Parachain blocks for which the corresponding blob is noted on the relay chain to be available, meaning that the candidate receipt has been voted to be available by $2/3$ validators.*

After a certain time-out in blocks since we first put the candidate receipt on the relay chain if there is not enough votes of availability the relay chain logic decides that a parachain block is unavailable, see 9.

Definition 24 *An unavailable parachain block is TBS*

/syedSo to be clear we are not announcing unavailability we just keep it for grand pa vote

0.10 Publishing Attestations

We have two type of attestations, primary and secondary. Primary attestations are signed by the parachain validators and secondary attestations are signed by secondary checkers and include the VRF that assigned them as a secondary checker into the attestation. Both types of attestations are included in the relay chain block as a transaction. For each parachain block candidate the relay chain keeps track of which validators have attested to its validity or invalidity.

0.11 Secondary Approval checking

Once a parachain block is acted on we carry the secondary validity/availability checks as follows. A scheme assigns every validator to one or more PoV blocks to check its validity, see Section 0.11.3 for details. An assigned validator acquires the PoV block (see Section ??) and checks its validity by comparing it to the candidate receipt. If validators notices that an equivocation has happened an additional validity/availability assignments will be made that is described in Section 0.11.5.

0.11.1 Approval Checker Assignment

Validators assign themselves to parachain block proposals as defined in Definition 17. The assignment needs to be random. Validators use their own VRF to sign the VRF output from the current relay chain block as described in Section 0.11.2. Each validator uses the output of the VRF to decide the block(s) they are revalidating as a secondary checker. See Section ?? for the detail.

In addition to this assignment some extra validators are assigned to every PoV block which is described in Section ??.

0.11.2 VRF computation

Every validator needs to run Algorithm 10 for every Parachain ρ to determines assignments. **TODO: Fix this. It is incorrect so far.**

Algorithm 10 VRF-FOR-APPROVAL(B, z, s_k)

Input: B : the block to be approved

z : randomness for approval assignment

s_k : session secret key of validator planning to participate in approval

1: $(\pi, d) \leftarrow \text{VRF}(H_h(B), sk(z))$

2: **return** (π, d)

Where VRF function is defined in [?].

0.11.3 One-Shot Approval Checker Assignemnt

Every validator v takes the output of this VRF computed by 10 mod the number of parachain blocks that we were decided to be available in this relay chain block according to Definition 23 and

executed. This will give them the index of the PoV block they are assigned to and need to check. The procedure is formalised in 11.

Algorithm 11 ONESHOTASSIGNMENT

Input:

1: TBS

0.11.4 Extra Approval Checker Assignment

Now for each parachain block, let us assume we want $\#VCheck$ validators to check every PoV block during the secondary checking. Note that $\#VCheck$ is not a fixed number but depends on reports from collators or fishermen. Lets us $\#VDefault$ be the minimum number of validator we want to check the block, which should be the number of parachain validators plus some constant like 2. We set

$$\#VCheck = \#VDefault + c_f * \text{total fishermen stake}$$

where c_f is some factor we use to weight fishermen reports. Reports from fishermen about this

Now each validator computes for each PoV block a VRF with the input being the relay chain block VRF concatenated with the parachain index.

For every PoV block, every validator compares $\#VCheck - \#VDefault$ to the output of this VRF and if the VRF output is small enough than the validator checks this PoV blocks immediately otherwise depending on their difference waits for some time and only perform a check if it has not seen $\#VCheck$ checks from validators who either 1) parachain validators of this PoV block 2) or assigned during the assignment procedure or 3) had a smaller VRF output than us during this time.

More fisherman reports can increase $\#VCheck$ and require new checks. We should carry on doing secondary checks for the entire fishing period if more are required. A validator need to keep track of which blocks have $\#VCheck$ smaller than the number of higher priority checks performed. A new report can make us check straight away, no matter the number of current checks, or mean that we need to put this block back into this set. If we later decide to prune some of this data, such as who has checked the block, then we'll need a new approach here.

Algorithm 12 ONESHOTASSIGNMENT

Input:

1: TBS

0.11.5 Additional Checking in Case of Equivocation

In the case of a relay chain equivocation, i.e. a validator produces two blocks with the same VRF, we do not want the secondary checkers for the second block to be predictable. To this end we use the block hash as well as the VRF as input for secondary checkers VRF. So each secondary checker is going to produce twice as many VRFs for each relay chain block that was equivocated. If either

of these VRFs is small enough then the validator is assigned to perform a secondary check on the PoV block. The process is formalized in Algorithm 13

Algorithm 13 EQUIVOCATEDASSIGNMENT

Input:

1: TBS

0.12 The Approval Check

Once a validator has a VRF which tells them to check a block, they announce this VRF and attempt to obtain the block. It is unclear yet whether this is best done by requesting the PoV block from parachain validators or by announcing that they want erasure-encoded chunks.

Retrieval

There are two fundamental ways to retrieve a parachain block for checking validity. One is to request the whole block from any validator who has attested to its validity or invalidity. Assigned approval checker v sends RequestWholeBlock message specified in Definition ?? to parachain validator in order to receive the specific parachain block. Any parachain validator receiving must reply with PoVBlockResponse message defined in Definition 25

Definition 25 *PoV Block Response Message TBS*

The second method is to retrieve enough erasure-encoded chunks to reconstruct the block from them. In the latter cases an announcement of the form specified in Definition has to be gossiped to all validators indicating that one needs the erasure-encoded chunks.

Definition 26 *Erasure-coded chunks request message TBS*

On their part, when a validator receive a erasure-coded chunks request message it response with the message specified in Definition 27.

Definition 27 *Erasure-coded chunks response message TBS*

Assigned approval checker v must retrieve enough erasure-encoded chunks of the block they are verifying to be able to reconstruct the block and the erasure chunks tree.

Reconstruction

After receiving $2f + 1$ of erasure chunks every assigned approval checker v needs to recreate the entirety of the erasure code, hence every v will run Algorithm ?? to make sure that the code is complete and the subsequently recover the original \bar{B} .

Algorithm 14 RECONSTRUCT-POV-ERASURE(S_{Er_B})

Input: $S_{Er_B} := (e_{j_1}, m_{j_1}), \dots, (e_{j_k}, m_{j_k})$ such that $k > 2f$

```

1:  $\bar{B} \rightarrow \text{ERASURE-DECODER}(e_{j_1}, \dots, e_{j_k})$ 
2: if ERASURE-DECODER failed then
3:   ANNOUNCE-FAILURE
4:   return
5: end if
6:  $Er_B \rightarrow \text{ERASURE-ENCODER}(\bar{B})$ 
7: if VERIFY-MERKLE-PROOF( $S_{Er_B}, Er_B$ ) failed then
8:   ANNOUNCE-FAILURE
9:   return
10: end if
11: return  $\bar{B}$ 

```

0.12.1 Verification

Once the parachain block has been obtained or reconstructed the secondary checker needs to execute the PoV block. We declare a candidate receipt as invalid if one of the following three conditions hold: 1) While reconstructing if the erasure code does not have the claimed Merkle root, 2) the validation function says that the PoV block is invalid, or 3) the result of executing the block is inconsistent with the candidate receipt on the relay chain.

The procedure is formalized in Algorithm

Algorithm 15 REVALIDATINGRECONSTRUCTEDPOV

Input:1: TBS

If everything checks out correctly, we declare the block is valid. This means gossiping an attestation, including a reference that identifies candidate receipt and our VRF as specified in Definition 28.

Definition 28 *Secondary approval attestation message TBS*

0.12.2 Process validity and invalidity messages

When a Block produced receive a Secondary approval attestation message, it execute Algorithm 16 to verify the VRF and may need to judge when enough time has passed.

Algorithm 16 VERIFYAPPROVALATTESTATION

Input:1: TBS

These attestations are included in the relay chain as a transaction specified in

Definition 29 *Approval Attestation Transaction TBS*

Collators reports of unavailability and invalidity specified in Definition **TODO: Define these messages** also go onto the relay chain as well in the format specified in Definition

Definition 30 *Collator Invalidity Transaction TBS*

Definition 31 *Collator unavailability Transaction*

0.12.3 Invalidity Escalation

When for any candidate receipt, there are attestations for both its validity and invalidity, then all validators acquire and validate the blob, irrespective of the assignments from section by executing Algorithm ?? and 15.

We do not vote in GRANDPA for a chain were the candidate receipt is executed until its vote is resolved. If we have n validators, we wait for $> 2n/3$ of them to attest to the blob and then the outcome of this vote is one of the following:

If $> n/3$ validators attest to the validity of the blob and $\leq n/3$ attest to its invalidity, then we can vote on the chain in GRANDPA again and slash validators who attested to its invalidity.

If $> n/3$ validators attest to the invalidity of the blob and $\leq n/3$ attest to its validity, then we consider the blob as invalid. If the relay chain block where the corresponding candidate receipt was executed was not finalised, then we never vote on it or build on it. We slash the validators who attested to its validity.

If $> n/3$ validators attest to the validity of the blob and $> n/3$ attest to its invalidity then we consider the blob to be invalid as above but we do not slash validators who attest either way. We want to leave a reasonable length of time in the first two cases to slash anyone to see if this happens.