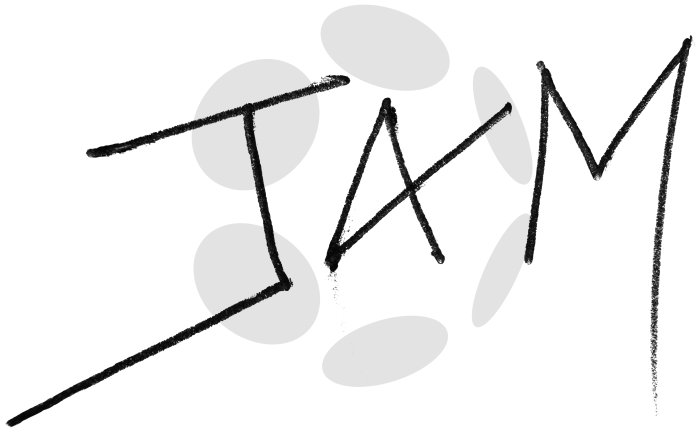

JOIN-ACCUMULATE MACHINE: A SEMI-COHERENT SCALABLE TRUSTLESS VM

DRAFT 0.3.4 August 9, 2024

DR. GAVIN WOOD
FOUNDER, POLKADOT & ETHEREUM
GAVIN@PARITY.IO

Abstract. We present a comprehensive and formal definition of JAM, a protocol combining elements of both *Polkadot* and *Ethereum*. In a single coherent model, JAM provides a global singleton permissionless object environment—much like the smart-contract environment pioneered by Ethereum—paired with secure sideband computation parallelized over a scalable node network, a proposition pioneered by Polkadot.

JAM introduces a decentralized hybrid system offering smart-contract functionality structured around a secure and scalable in-core/on-chain dualism. While the smart-contract functionality implies some similarities with Ethereum's paradigm, the overall model of the service offered is driven largely by underlying architecture of Polkadot.

JAM is permissionless in nature, allowing anyone to deploy code as a service on it for a fee commensurate with the resources this code utilizes and to induce execution of this code through the procurement and allocation of *core-time*, a metric of resilient and ubiquitous computation, somewhat similar to the purchasing of gas in Ethereum. We already envision a Polkadot-compatible *CoreChains* service.

1. Introduction

1.1. **Nomenclature.** In this paper, we introduce a decentralized, crypto-economic protocol to which the Polkadot Network could conceivably transition itself in a major revision. Following this eventuality (which must not be taken for granted since Polkadot is a decentralized network) this protocol might also become known as *Polkadot* or some derivation thereof. However, at this stage this is not the case, therefore our proposed protocol will for the present be known as JAM.

An early, unrefined, version of this protocol was first proposed in Polkadot Fellowship RFC31, known as *CoreJam*. CoreJam takes its name after the collect/refine/join/accumulate model of computation at the heart of its service proposition. While the CoreJam RFC suggested an incomplete, scope-limited alteration to the Polkadot protocol, JAM refers to a complete and coherent overall blockchain protocol.

1.2. **Driving Factors.** Within the realm of blockchain and the wider Web3, we are driven by the need first and foremost to deliver resilience. A proper Web3 digital system should honor a declared service profile—and ideally meet even perceived expectations—regardless of the desires, wealth or power of any economic actors including individuals, organizations and, indeed, other Web3 systems. Inevitably this is aspirational, and we must be pragmatic over how perfectly this may really be delivered. Nonetheless, a Web3 system should aim to provide such radically strong guarantees that, for practical purposes, the system may be described as *unstoppable*.

While Bitcoin is, perhaps, the first example of such a system within the economic domain, it was not general purpose in terms of the nature of the service it offered. A rules-based service is only as useful as the generality of the rules which may be conceived and placed within it. Bitcoin's rules allowed for an initial use-case, namely a fixed-issuance token, ownership of which is well-approximated and autonomously enforced through knowledge of a secret, as well as some further elaborations on this theme.

Later, Ethereum would provide a categorically more general-purpose rule set, one which was practically Turing complete.[1] In the context of Web3 where we are aiming to deliver a massively multiuser application platform, generality is crucial, and thus we take this as a given.

Beyond resilience and generality, things get more interesting, and we must look a little deeper to understand

[1]The gas mechanism did restrict what programs can execute on it by placing an upper bound on the number of steps which may be executed, but some restriction to avoid infinite-computation must surely be introduced in a permissionless setting.

what our driving factors are. For the present purposes, we identify three additional goals:

(1) Resilience: highly resistant from being stopped, corrupted and censored.
(2) Generality: able to perform Turing-complete computation.
(3) Performance: able to perform computation quickly and at low cost.
(4) Coherency: the causal relationship possible between different elements of state and thus how well individual applications may be composed.
(5) Accessibility: negligible barriers to innovation; easy, fast, cheap and permissionless.

As a declared Web3 technology, we make an implicit assumption of the first two items. Interestingly, items 3 and 4 are antagonistic according to an information theoretic principle which we are sure must already exist in some form but are nonetheless unaware of a name for it. For argument's sake we shall name it *size-synchrony antagonism*.

1.3. **Scaling under Size-Synchrony Antagonism.** Size-synchrony antagonism is a simple principle implying that as the state-space of information systems grow, then the system necessarily becomes less synchronous. The argument goes:

(1) The more state a system utilizes for its data-processing, the greater the amount of space this state must occupy.
(2) The more space used, then the greater the mean and variance of distances between state-components.
(3) As the mean and variance increase, then interactions become slower and subsystems must manage the possibility that distances between interdependent components of state could be materially different, requiring asynchrony.

This assumes perfect coherency of the system's state. Setting the question of overall security aside for a moment, we can avoid this rule by applying the *divide and conquer* maxim and fragmenting the state of a system, sacrificing its coherency. We might for example create two independent smaller-state systems rather than one large-state system. This pattern applies a step-curve to the principle; intra-system processing has low size and high synchrony, inter-system processing has high size but low synchrony. It is the principle behind meta-networks such as Polkadot, Cosmos and the predominant vision of a scaled Ethereum (all to be discussed in depth shortly).

The present work explores a middle-ground in the antagonism, avoiding the persistent fragmentation of state-space of the system as with existing approaches. We do this by introducing a new model of computation which pipelines a highly scalable element to a highly synchronous element. Asynchrony is not avoided, but we do open the possibility for a greater degree of granularity over how it is traded against size. In particular fragmentation can be made ephemeral rather than persistent, drawing upon a coherent state and fragmenting it only for as long as it takes to execute any given piece of processing on it.

Unlike with SNARK-based L2-blockchain techniques for scaling, this model draws upon crypto-economic mechanisms and inherits their low-cost and high-performance profiles and averts a bias toward centralization.

1.4. **Document Structure.** We begin with a brief overview of present scaling approaches in blockchain technology in section 2. In section 3 we define and clarify the notation from which we will draw for our formalisms.

We follow with a broad overview of the protocol in section 4 outlining the major areas including the Polka Virtual Machine (PVM), the consensus protocols Safrole and GRANDPA, the common clock and build the foundations of the formalism.

We then continue with the full protocol definition split into two parts: firstly the correct on-chain state-transition formula helpful for all nodes wishing to validate the chain state, and secondly, in sections 14 and 19 the honest strategy for the off-chain actions of any actors who wield a validator key.

The main body ends with a discussion over the performance characteristics of the protocol in section 20 and finally conclude in section 21.

The appendix contains various additional material important for the protocol definition including the PVM in appendices A & B, serialization and Merklization in appendices C & D and cryptography in appendices E, G & H. We finish with an index of terms which includes the values of all simple constant terms used in the work in appendix I, and close with the bibliography.

2. Previous Work and Present Trends

In the years since the initial publication of the Ethereum *YP*, the field of blockchain development has grown immensely. Other than scalability, development has been done around underlying consensus algorithms, smart-contract languages and machines and overall state environments. While interesting, these latter subjects are mostly out scope of the present work since they generally do not impact underlying scalability.

2.1. **Polkadot.** In order to deliver its service, JAM co-opts much of the same game-theoretic and cryptographic machinery as Polkadot known as ELVES and described by Jeff Burdges, Cevallos, et al. 2024. However, major differences exist in the actual service offered with JAM, providing an abstraction much closer to the actual computation model generated by the validator nodes its economy incentivizes.

It was a major point of the original Polkadot proposal, a scalable heterogeneous multichain, to deliver high-performance through partition and distribution of the workload over multiple host machines. In doing so it took an explicit position that composability would be lowered. Polkadot's constituent components, parachains are, practically speaking, highly isolated in their nature. Though a message passing system (XCMP) exists it is asynchronous, coarse-grained and practically limited by its reliance on a high-level slowly evolving interaction language XCM.

As such, the composability offered by Polkadot between its constituent chains is lower than that of Ethereum-like smart-contract systems offering a single and universal object environment and allowing for the kind of agile and innovative integration which underpins

their success. Polkadot, as it stands, is a collection of independent ecosystems with only limited opportunity for collaboration, very similar in ergonomics to bridged blockchains though with a categorically different security profile. A technical proposal known as SPREE would utilize Polkadot's unique shared-security and improve composability, though blockchains would still remain isolated.

Implementing and launching a blockchain is hard, time-consuming and costly. By its original design, Polkadot limits the clients able to utilize its service to those who are both able to do this and raise a sufficient deposit to win an auction for a long-term slot, one of around 50 at the present time. While not permissioned per se, accessibility is categorically and substantially lower than for smart-contract systems similar to Ethereum.

Enabling as many innovators to participate and interact, both with each other and each other's user-base, appears to be an important component of success for a Web3 application platform. Accessibility is therefore crucial.

2.2. **Ethereum.** The Ethereum protocol was formally defined in this paper's spiritual predecessor, the *Yellow Paper*, by Wood 2014. This was derived in large part from the initial concept paper by Buterin 2013. In the decade since the *YP* was published, the *de facto* Ethereum protocol and public network instance have gone through a number of evolutions, primarily structured around introducing flexibility via the transaction format and the instruction set and "precompiles" (niche, sophisticated bonus instructions) of its scripting core, the Ethereum virtual machine (EVM).

Almost one million crypto-economic actors take part in the validation for Ethereum.[2] Block extension is done through a randomized leader-rotation method where the physical address of the leader is public in advance of their block production.[3] Ethereum uses Casper-FFG introduced by Buterin and Griffith 2019 to determine finality, which with the large validator base finalizes the chain extension around every 13 minutes.

Ethereum's direct computational performance remains broadly similar to that with which it launched in 2015, with a notable exception that an additional service now allows 1MB of *commitment data* to be hosted per block (all nodes to store it for a limited period). The data cannot be directly utilized by the main state-transition function, but special functions provide proof that the data (or some subsection thereof) is available. According to Ethereum Foundation 2024b, the present design direction is to improve on this over the coming years by splitting responsibility for its storage amongst the validator base in a protocol known as *Dank-sharding*.

According to Ethereum Foundation 2024a, the scaling strategy of Ethereum would be to couple this data availability with a private market of *roll-ups*, sideband computation facilities of various design, with ZK-SNARK-based roll-ups being a stated preference. Each vendor's roll-up design, execution and operation comes with its own implications.

One might reasonably assume that a diversified market-based approach for scaling via multivendor roll-ups will allow well-designed solutions to thrive. However, there are potential issues facing the strategy. A research report by Sharma 2023 on the level of decentralization in the various roll-ups found a broad pattern of centralization, but notes that work is underway to attempt to mitigate this. It remains to be seen how decentralized they can yet be made.

Heterogeneous communication properties (such as datagram latency and semantic range), security properties (such as the costs for reversion, corruption, stalling and censorship) and economic properties (the cost of accepting and processing some incoming message or transaction) may differ, potentially quite dramatically, between major areas of some grand patchwork of roll-ups by various competing vendors. While the overall Ethereum network may eventually provide some or even most of the underlying machinery needed to do the sideband computation it is far from clear that there would be a "grand consolidation" of the various properties should such a thing happen. We have not found any good discussion of the negative ramifications of such a fragmented approach.[4]

2.2.1. *Snark Roll-ups.* While the protocol's foundation makes no great presuppositions on the nature of roll-ups, Ethereum's strategy for sideband computation does centre around SNARK-based rollups and as such the protocol is being evolved into a design that makes sense for this. SNARKs are the product of an area of exotic cryptography which allow proofs to be constructed to demonstrate to a neutral observer that the purported result of performing some predefined computation is correct. The complexity of the verification of these proofs tends to be sub-linear in their size of computation to be proven and will not give away any of the internals of said computation, nor any dependent witness data on which it may rely.

ZK-SNARKs come with constraints. There is a trade-off between the proof's size, verification complexity and the computational complexity of generating it. Non-trivial computation, and especially the sort of general-purpose computation laden with binary manipulation which makes smart-contracts so appealing, is hard to fit into the model of SNARKs.

To give a practical example, RISC-zero (as assessed by Bögli 2024) is a leading project and provides a platform for producing SNARKs of computation done by a RISC-V virtual machine, an open-source and succinct RISC machine architecture well-supported by tooling. A recent benchmarking report by PolkaVM Project 2024 showed that compared to RISC-zero's own benchmark, proof generation alone takes over 61,000 times as long as simply recompiling and executing even when executing on 32 times as many cores, using 20,000 times as much RAM and an additional state-of-the-art GPU. According to hardware rental agents `https://cloud-gpus.com/`, the cost multiplier of proving

[2]Practical matters do limit the level of real decentralization. Validator software expressly provides functionality to allow a single instance to be configured with multiple key sets, systematically facilitating a much lower level of actual decentralization than the apparent number of actors, both in terms of individual operators and hardware. Using data collated by Dune and hildobby 2024 on Ethereum 2, one can see one major node operator, Lido, has steadily accounted for almost one-third of the almost one million crypto-economic participants.

[3]Ethereum's developers hope to change this to something more secure, but no timeline is fixed.

[4]Some initial thoughts on the matter resulted in a proposal by Sadana 2024 to utilize Polkadot technology as a means of helping create a modicum of compatibility between roll-up ecosystems!

using RISC-zero is 66,000,000x of the cost[5] to execute using our RISC-V recompiler.

Many cryptographic primitives become too expensive to be practical to use and specialized algorithms and structures must be substituted. Often times they are otherwise suboptimal. In expectation of the use of SNARKs (such as PLONK as proposed by Gabizon, Williamson, and Ciobotaru 2019), the prevailing design of the Ethereum project's Dank-sharding availability system uses a form of erasure coding centered around polynomial commitments over a large prime field in order to allow SNARKs to get acceptably performant access to subsections of data. Compared to alternatives, such as a binary field and Merklization in the present work, it leads to a load on the validator nodes orders of magnitude higher in terms of CPU usage.

In addition to their basic cost, SNARKs present no great escape from decentralization and the need for redundancy, leading to further cost multiples. While the need for some benefits of staked decentralization is averted through their verifiable nature, the need to incentivize multiple parties to do much the same work is a requirement to ensure that a single party not form a monopoly (or several not form a cartel). Proving an incorrect state-transition should be impossible, however service integrity may be compromised in other ways; a temporary suspension of proof-generation, even if only for minutes, could amount to major economic ramifications for real-time financial applications.

Real-world examples exist of the pit of centralization giving rise to monopolies. One would be the aforementioned SNARK-based exchange framework; while notionally serving decentralized exchanges, it is in fact centralized with Starkware itself wielding a monopoly over enacting trades through the generation and submission of proofs, leading to a single point of failure—should Starkware's service become compromised, then the liveness of the system would suffer.

It has yet to be demonstrated that SNARK-based strategies for eliminating the trust from computation will ever be able to compete on a cost-basis with a multi-party crypto-economic platform. All as-yet proposed SNARK-based solutions are heavily reliant on crypto-economic systems to frame them and work around their issues. Data availability and sequencing are two areas well understood as requiring a crypto-economic solution.

We would note that SNARK technology is improving and the cryptographers and engineers behind them do expect improvements in the coming years. In a recent article by Thaler 2023 we see some credible speculation that with some recent advancements in cryptographic techniques, slowdowns for proof generation could be as little as 50,000x from regular native execution and much of this could be parallelized. This is substantially better than the present situation, but still several orders of magnitude greater than would be required to compete on a cost-basis with established crypto-economic techniques such as ELVES.

2.3. **Fragmented Meta-Networks.** Directions for general-purpose computation scalability taken by other projects broadly centre around one of two approaches; either what might be termed a *fragmentation* approach or alternatively a *centralization* approach. We argue that neither approach offers a compelling solution.

The fragmentation approach is heralded by projects such as Cosmos (proposed by Kwon and Buchman 2019) and Avalanche (by Tanana 2019). It involves a system fragmented by networks of a homogenous consensus mechanic, yet staffed by separately motivated sets of validators. This is in contrast to Polkadot's single validator set and Ethereum's declared strategy of heterogeneous roll-ups secured partially by the same validator set operating under a coherent incentive framework. The homogeneity of said fragmentation approach allows for reasonably consistent messaging mechanics, helping to present a fairly unified interface to the multitude of connected networks.

However, the apparent consistency is superficial. The networks are trustless only by assuming correct operation of their validators, who operate under a crypto-economic security framework ultimately conjured and enforced by economic incentives and punishments. To do twice as much work with the same levels of security and no special coordination between validator sets, then such systems essentially prescribe forming a new network with the same overall levels of incentivization.

Several problems arise. Firstly, there is a similar downside as with Polkadot's isolated parachains and Ethereum's isolated roll-up chains: a lack of coherency due to a persistently sharded state preventing synchronous composability.

More problematically, the scaling-by-fragmentation approach, proposed specifically by Cosmos, provides no homogenous security—and therefore trustlessness—guarantees. Validator sets between networks must be assumed to be independently selected and incentivized with no relationship, causal or probabilistic, between the Byzantine actions of a party on one network and potential for appropriate repercussions on another. Essentially, this means that should validators conspire to corrupt or revert the state of one network, the effects may be felt across other networks of the ecosystem.

That this is an issue is broadly accepted, and projects propose for it to be addressed in one of two ways. Firstly, to fix the expected cost-of-attack (and thus level of security) across networks by drawing from the same validator set. The massively redundant way of doing this, as proposed by Cosmos Project 2023 under the name *replicated security*, would be to require each validator to validate on all networks and for the same incentives and punishments. This is economically inefficient in the cost of security provision as each network would need to independently provide the same level of incentives and punishment-requirements as the most secure with which it wanted to interoperate. This is to ensure the economic proposition remain unchanged for validators and the security proposition remained equivalent for all networks. At the present time, replicated security is not a readily available permissionless service. We might speculate that these punishing economics have something to do with it.

The more efficient approach, proposed by the OmniLedger team, Kokoris-Kogias et al. 2017, would be to make the validators non-redundant, partitioning them between different networks and periodically, securely and

[5]In all likelihood actually substantially more as this was using low-tier "spare" hardware in consumer units, and our recompiler was unoptimized.

randomly repartitioning them. A reduction in the cost to attack over having them all validate on a single network is implied since there is a chance of having a single network accidentally have a compromising number of malicious validators even with less than this proportion overall. This aside it presents an effective means of scaling under a basis of weak-coherency.

Alternatively, as in Elves by Jeff Burdges, Cevallos, et al. 2024, we may utilize non-redundant partitioning, combine this with a proposal-and-auditing game which validators play to weed out and punish invalid computations, and then require that the finality of one network be contingent on all causally-entangled networks. This is the most secure and economically efficient solution of the three, since there is a mechanism for being highly confident that invalid transitions will be recognized and corrected before their effect is finalized across the ecosystem of networks. However, it requires substantially more sophisticated logic and their causal-entanglement implies some upper limit on the number of networks which may be added.

2.4. **High-Performance Fully Synchronous Networks.** Another trend in the recent years of blockchain development has been to make "tactical" optimizations over data throughput by limiting the validator set size or diversity, focusing on software optimizations, requiring a higher degree of coherency between validators, onerous requirements on the hardware which validators must have, or limiting data availability.

The Solana blockchain is underpinned by technology introduced by Yakovenko 2018 and boasts theoretical figures of over 700,000 transactions per second, though according to Ng 2024 the network is only seen processing a small fraction of this. The underlying throughput is still substantially more than most blockchain networks and is owed to various engineering optimizations in favor of maximizing synchronous performance. The result is a highly-coherent smart-contract environment with an API not unlike that of *YP* Ethereum (albeit using a different underlying VM), but with a near-instant time to inclusion and finality which is taken to be immediate upon inclusion.

Two issues arise with such an approach: firstly, defining the protocol as the outcome of a heavily optimized codebase creates structural centralization and can undermine resilience. Jha 2024 writes "since January 2022, 11 significant outages gave rise to 15 days in which major or partial outages were experienced". This is an outlier within the major blockchains as the vast majority of major chains have no downtime. There are various causes to this downtime, but they are generally due to bugs found in various subsystems.

Ethereum, at least until recently, provided the most contrasting alternative with its well-reviewed specification, clear research over its crypto-economic foundations and multiple clean-room implementations. It is perhaps no surprise that the network very notably continued largely unabated when a flaw in its most deployed implementation was found and maliciously exploited, as described by Hertig 2016.

The second issue is concerning ultimate scalability of the protocol when it provides no means of distributing workload beyond the hardware of a single machine.

In major usage, both historical transaction data and state would grow impractically. Solana illustrates how much of a problem this can be. Unlike classical blockchains, the Solana protocol offers no solution for the archival and subsequent review of historical data, crucial if the present state is to be proven correct from first principle by a third party. There is little information on how Solana manages this in the literature, but according to Solana Foundation 2023, nodes simply place the data onto a centralized database hosted by Google.[6]

Solana validators are encouraged to install large amounts of RAM to help hold its large state in memory (512 GB is the current recommendation according to Solana Labs 2024). Without a divide-and-conquer approach, Solana shows that the level of hardware which validators can reasonably be expected to provide dictates the upper limit on the performance of a totally synchronous, coherent execution model. Hardware requirements represent barriers to entry for the validator set and cannot grow without sacrificing decentralization and, ultimately, transparency.

3. Notational Conventions

Much as in the Ethereum Yellow Paper, a number of notational conventions are used throughout the present work. We define them here for clarity. The Ethereum Yellow Paper itself may be referred to henceforth as the *YP*.

3.1. **Typography.** We use a number of different typefaces to denote different kinds of terms. Where a term is used to refer to a value only relevant within some localized section of the document, we use a lower-case roman letter e.g. x, y (typically used for an item of a set or sequence) or e.g. i, j (typically used for numerical indices). Where we refer to a Boolean term or a function in a local context, we tend to use a capitalized roman alphabet letter such as A, F. If particular emphasis is needed on the fact a term is sophisticated or multidimensional, then we may use a bold typeface, especially in the case of sequences and sets.

For items which retain their definition throughout the present work, we use other typographic conventions. Sets are usually referred to with a blackboard typeface, e.g. $\mathbb{N}$ refers to all natural numbers including zero. Sets which may be parameterized may be subscripted or be followed by parenthesized arguments. Imported functions, used by the present work but not specifically introduced by it, are written in calligraphic typeface, e.g. $\mathcal{H}$ the Blake2 cryptographic hashing function. For other non-context dependent functions introduced in the present work, we use upper case Greek letters, e.g. denotes the state transition function.

Values which are not fixed but nonetheless hold some consistent meaning throughout the present work are denoted with lower case Greek letters such as σ, the state identifier. These may be placed in bold typeface to denote that they refer to an abnormally complex value.

[6] Earlier node versions utilized Arweave network, a decentralized data store, but this was found to be unreliable for the data throughput which Solana required.

3.2. **Functions and Operators.** We define the precedes relation to indicate that one term is defined in terms of another. E.g. $y \quad x$ indicates that y may be defined purely in terms of x:

$$(1) \qquad y \quad x \qquad f \quad y = f(x)$$

The substitute-if-nothing function U is equivalent to the first argument which is not , or if no such argument exists:

$$(2) \qquad \mathrm{U}(a_0, \dots a_n) \equiv a_x \quad (a_x \neq \qquad x = n), \quad a_i =$$

Thus, e.g. U(,1, ,2) = 1 and U(,) = .

3.3. **Sets.** Given some set $\mathbf{s}$, its power set and cardinality are denoted as the usual $\mathbf{s}$ and $\mathbf{s}$. When forming a power set, we may use a numeric subscript in order to restrict the resultant expansion to a particular cardinality. E.g. $\{1,2,3\}_2 = \{\{1,2\},\{1,3\},\{2,3\}\}$.

Sets may be operated on with scalars, in which case the result is a set with the operation applied to each element, e.g. $\{1,2,3\} + 3 = \{4,5,6\}$

We denote set-disjointness with the relation . Formally:

$$A \cap B = \qquad A \quad B$$

We commonly use to indicate that some term is validly left without a specific value. Its cardinality is defined as zero. We define the operation ? such that $A? \equiv A \quad \{ \}$ indicating the same set but with the addition of the element.

The term is utilized to indicate the unexpected failure of an operation or that a value is invalid or unexpected. (We try to avoid the use of the more conventional here to avoid confusion with Boolean false, which may be interpreted as some successful result in some contexts.)

3.4. **Numbers.** $\mathbb{N}$ denotes the set of naturals including zero whereas $\mathbb{N}_n$ implies a restriction on that set to values less than n. Formally, $\mathbb{N} = \{0, 1, \dots\}$ and $\mathbb{N}_n = \{x \quad x \quad \mathbb{N}, x < n\}$.

$\mathbb{Z}$ denotes the set of integers. We denote $\mathbb{Z}_{a:::b}$ to be the set of integers within the interval $[a, b)$. Formally, $\mathbb{Z}_{a:::b} = \{x \quad x \quad \mathbb{Z}, a \le x < b\}$. E.g. $\mathbb{Z}_{2:::5} = \{2,3,4\}$. We denote the offset/length form of this set as $\mathbb{Z}_{a \ +b}$, a short form of $\mathbb{Z}_{a:::a+b}$.

It can sometimes be useful to represent lengths of sequences and yet limit their size, especially when dealing with sequences of octets which must be stored practically. Typically, these lengths can be defined as the set $\mathbb{N}_{2^{32}}$. To improve clarity, we denote $\mathbb{N}_L$ as the set of lengths of octet sequences and is equivalent to $\mathbb{N}_{2^{32}}$.

We denote the % operator as the modulo operator, e.g. $5 \% 3 = 2$. Furthermore, we may occasionally express a division result as a quotient and remainder with the separator | , e.g. $5 \div 3 = 1 \mid 2$.

3.5. **Dictionaries.** A *dictionary* is a possibly partial mapping from some domain into some co-domain in much the same manner as a regular function. Unlike functions however, with dictionaries the total set of pairings are necessarily enumerable, and we represent them in some data structure as the set of all (*key* *value*) pairs. (In such data-defined mappings, it is common to name the values within the domain a *key* and the values within the co-domain a *value*, hence the naming.)

Thus, we define the formalism $\mathbb{D}$ K → V to denote a dictionary which maps from the domain K to the range V. We define a dictionary as a member of the set of all dictionaries $\mathbb{D}$ and a set of pairs $p = (k \quad v)$:

$$(3) \qquad \mathbb{D} \quad \{\{(k \quad v)\}\}$$

A dictionary's members must associate at most one unique value for any key k:

$$(4) \qquad \mathbf{d} \quad \mathbb{D} \quad (k \quad v) \quad \mathbf{d} \quad !v' \quad (k \quad v') \quad \mathbf{d}$$

This assertion allows us to unambiguously define the subscript and subtraction operator for a dictionary d:

$$(5) \qquad \mathbf{d} \quad \mathbb{D} \quad \mathbf{d}[k] \equiv \begin{cases} v & \text{if } k \quad (k \quad v) \quad \mathbf{d} \\ & \text{otherwise} \end{cases}$$

$$(6) \qquad \mathbf{d} \quad \mathbb{D}, \mathbf{s} \quad K \quad \mathbf{d} \quad \mathbf{s} \equiv \{(k \quad v) \quad (k \quad v) \quad \mathbf{d}, k \,/\, \mathbf{s}\}$$

Note that when using a subscript, it is an implicit assertion that the key exists in the dictionary. Should the key not exist, the result is undefined and any block which relies on it must be considered invalid.

It is typically useful to limit the sets from which the keys and values may be drawn. Formally, we define a typed dictionary $\mathbb{D}\ K \to V$ as a set of pairs p of the form $(k \quad v)$:

$$(7) \qquad \mathbb{D}\ K \to V \quad \mathbb{D}$$

$$(8) \qquad \mathbb{D}\ K \to V \ \equiv \{\{(k \quad v) \quad k \quad K \quad v \quad V\}\}$$

To denote the active domain (i.e. set of keys) of a dictionary $\mathbf{d} \quad \mathbb{D}\ K \to V$, we use $\mathrm{K}(\mathbf{d}) \quad K$ and for the range (i.e. set of values), $\mathrm{V}(\mathbf{d}) \quad V$. Formally:

$$(9) \qquad \mathrm{K}(\mathbf{d} \quad \mathbb{D}) \equiv \{\ k \quad v \quad (k \quad v) \quad \mathbf{d}\ \}$$

$$(10) \qquad \mathrm{V}(\mathbf{d} \quad \mathbb{D}) \equiv \{\ v \quad k \quad (k \quad v) \quad \mathbf{d}\ \}$$

Note that since the co-domain of V is a set, should different keys with equal values appear in the dictionary, the set will only contain one such value.

3.6. **Tuples.** Tuples are groups of values where each item may belong to a different set. They are denoted with parentheses, e.g. the tuple t of the naturals 3 and 5 is denoted $t = (3,5)$, and it exists in the set of natural pairs sometimes denoted $\mathbb{N} \times \mathbb{N}$, but denoted in the present work as $(\mathbb{N}, \mathbb{N})$.

We have frequent need to refer to a specific item within a tuple value and as such find it convenient to declare a name for each item. E.g. we may denote a tuple with two named natural components a and b as $T = \quad a \quad \mathbb{N}, b \quad \mathbb{N}$. We would denote an item $t \quad T$ through subscripting its name, thus for some $t = \quad a \quad 3, b \quad 5$, $t_a = 3$ and $t_b = 5$.

3.7. **Sequences.** A sequence is a series of elements with particular ordering not dependent on their values. The set of sequences of elements all of which are drawn from some set T is denoted T, and it defines a partial mapping $\mathbb{N} \to T$. The set of sequences containing exactly n elements each a member of the set T may be denoted T_n and accordingly defines a complete mapping $\mathbb{N}_n \to T$. Similarly, sets of sequences of at most n elements and at least n elements may be denoted $T_{\ n}$ and $T_{\ n}$ respectively.

Sequences are subscriptable, thus a specific item at index i within a sequence $\mathbf{s}$ may be denoted $\mathbf{s}[i]$, or where unambiguous, $\mathbf{s}_i$. A range may be denoted using an ellipsis for example: $[0,1,2,3]_{::2} = [0,1]$ and $[0,1,2,3]_{1\ +2} = [1,2]$. The length of such a sequence may be denoted $\mathbf{s}$.

We denote modulo subscription as $\mathbf{s}[i]^{\circ} \equiv \mathbf{s}[i \bmod |\mathbf{s}|]$. We denote the final element x of a sequence $\mathbf{s} = [..., x]$ through the function $\text{last}(\mathbf{s}) \equiv x$.

3.7.1. *Construction.* We may wish to define a sequence in terms of incremental subscripts of other values: $[\mathbf{x}_0, \mathbf{x}_1, \dots]_n$ denotes a sequence of n values beginning $\mathbf{x}_0$ continuing up to $\mathbf{x}_{n-1}$. Furthermore, we may also wish to define a sequence as elements each of which are a function of their index i; in this case we denote $[f(i) \mid i \in \mathbb{N}_n] \equiv [f(0), f(1), \dots, f(n-1)]$. Thus, when the ordering of elements matters we use $\in$ rather than the unordered notation $\in$. The latter may also be written in short form $[f(i \in \mathbb{N}_n)]$. This applies to any set which has an unambiguous ordering, particularly sequences, thus $[i^2 \mid i \in [1,2,3]] = [1,4,9]$. Multiple sequences may be combined, thus $[i \cdot j \mid i \in [1,2,3], j \in [2,3,4]] = [2,6,12]$.

We use explicit notation $f^{\#}$ to denote a function mapping over all items of a sequence. Thus given some function $y = f(x)$:

$$[f(\mathbf{x}_0), f(\mathbf{x}_1), \dots] = f^{\#}([\mathbf{x}_0, \mathbf{x}_1, \dots]) \tag{11}$$

Sequences may be constructed from sets or other sequences whose order should be ignored through sequence ordering notation $[i_k \mid i \in X]$, which is defined to result in the set or sequence of its argument except that all elements i are placed in ascending order of the corresponding value i_k.

The key component may be elided in which case it is assumed to be ordered by the elements directly; i.e. $[i \mid i \in X] \equiv [i \mid i \in X]$. $[i_k \mid i \in X]$ does the same, but excludes any duplicate values of i. E.g. assuming $\mathbf{s} = [1,3,2,3]$, then $[i \mid i \in \mathbf{s}] = [1,2,3]$ and $[-i \mid i \in \mathbf{s}] = [3,3,2,1]$.

Sets may be constructed from sequences with the regular set construction syntax, e.g. assuming $\mathbf{s} = [1,2,3,1]$, then $\{a \mid a \in \mathbf{s}\}$ would be equivalent to $\{1,2,3\}$.

Sequences of values which themselves have a defined ordering have an implied ordering akin to a regular dictionary, thus $[1,2,3] < [1,2,4]$ and $[1,2,3] < [1,2,3,1]$.

3.7.2. *Editing.* We define the sequence concatenation operator $\frown$ such that $[\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{y}_0, \mathbf{y}_1, \dots] \equiv \mathbf{x} \frown \mathbf{y}$. For sequences of sequences, we define a unary concatenate-all operator: $\frown \mathbf{x} \equiv \mathbf{x}_0 \frown \mathbf{x}_1 \frown \dots$. Further, we denote element concatenation as $x \doubleplus i \equiv x \frown [i]$. We denote the sequence made up of the first n elements of sequence $\mathbf{s}$ to be $\overleftarrow{\mathbf{s}}^{n} \equiv [\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_{n-1}]$, and only the final elements as $\overrightarrow{\mathbf{s}}^{n}$.

We define $^{\mathsf{T}}\mathbf{x}$ as the transposition of the sequence-of-sequences $\mathbf{x}$, fully defined in equation 316. We may also apply this to sequences-of-tuples to yield a tuple of sequences.

We denote sequence subtraction with a slight modification of the set subtraction operator; specifically, some sequence $\mathbf{s}$ excepting the left-most element equal to v would be denoted $\mathbf{s} \setminus \{v\}$.

3.7.3. *Boolean values.* $\mathbb{B}_s$ denotes the set of Boolean strings of length s, thus $\mathbb{B}_s = \{\text{false}, \text{true}\}_s$. When dealing with Boolean values we may assume an implicit equivalence mapping to a bit whereby true $= 1$ and false $= 0$, thus $\mathbb{B} = \mathbb{N}_2$. We use the function $\text{bits}(\mathbb{Y}) \to \mathbb{B}$ to denote the sequence of bits, ordered with the least significant first, which represent the octet sequence $\mathbb{Y}$, thus $\text{bits}([5,0]) = [1,0,1,0,0,\dots]$.

3.7.4. *Octets and Blobs.* $\mathbb{Y}$ denotes the set of octet strings ("blobs") of arbitrary length. As might be expected, $\mathbb{Y}_x$ denotes the set of such sequences of length x. $\mathbb{Y}_{\$}$ denotes the subset of $\mathbb{Y}$ which are ASCII-encoded strings. Note that while an octet has an implicit and obvious bijective relationship with natural numbers less than 256, and we may implicitly coerce between octet form and natural number form, we do not treat them as exactly equivalent entities. In particular for the purpose of serialization, an octet is always serialized to itself, whereas a natural number may be serialized as a sequence of potentially several octets, depending on its magnitude and the encoding variant.

3.7.5. *Shuffling.* We define the sequence-shuffle function F, originally introduced by Fisher and Yates 1938, with an efficient in-place algorithm described by Wikipedia 2024. This accepts a sequence and some entropy and returns a sequence of the same length with the same elements but in an order determined by the entropy. The entropy may be provided as either an indefinite sequence of naturals or a hash. For a full definition see appendix F.

3.8. **Cryptography.**

3.8.1. *Hashing.* $\mathbb{H}$ denotes the set of 256-bit values typically expected to be arrived at through a cryptographic function, equivalent to $\mathbb{Y}_{32}$, with $\mathbb{H}^0$ being equal to $[0]_{32}$. We assume a function $\mathsf{H}(m \in \mathbb{Y}) \in \mathbb{H}$ denoting the Blake2b 256-bit hash introduced by Saarinen and Aumasson 2015 and a function $\mathsf{H}_K(m \in \mathbb{Y}) \in \mathbb{H}$ denoting the Keccak 256-bit hash as proposed by Bertoni et al. 2013 and utilized by Wood 2014.

We may sometimes wish to take only the first x octets of a hash, in which case we denote $\mathsf{H}_x(m) \in \mathbb{Y}_x$ to be the first x octets of $\mathsf{H}(m)$. The inputs of a hash function are generally assumed to be serialized with our codec $\mathsf{E}(x) \in \mathbb{Y}$, however for the purposes of clarity or unambiguity we may also explicitly denote the serialization. Similarly, we may wish to interpret a sequence of octets as some other kind of value with the assumed decoder function $\mathsf{E}^{-1}(x \in \mathbb{Y})$. In both cases, we may subscript the transformation function with the number of octets we expect the octet sequence term to have. Thus, $r = \mathsf{E}_4(x \in \mathbb{N})$ would assert $x \in \mathbb{N}_{2^{32}}$ and $r \in \mathbb{Y}_4$, whereas $s = \mathsf{E}_8^{-1}(y)$ would assert $y \in \mathbb{Y}_8$ and $s \in \mathbb{N}_{2^{64}}$.

3.8.2. *Signing Schemes.* $\mathbb{E}_k\langle m \in \mathbb{Y}\rangle \in \mathbb{Y}_{64}$ is the set of valid Ed25519 signatures, defined by Josefsson and Liusvaara 2017, made through knowledge of a secret key whose public key counterpart is $k \in \mathbb{Y}_{32}$ and whose message is m. To aid readability, we denote the set of valid public keys $k \in \mathbb{H}_E$.

We use $\mathbb{Y}_{BLS} \in \mathbb{Y}_{144}$ to denote the set of public keys for the BLS signature scheme, described by Boneh, Lynn, and Shacham 2004, on curve BLS12-381 defined by Hopwood et al. 2020.

We denote the set of valid Bandersnatch public keys as $\mathbb{H}_B$, defined in appendix G. $\mathbb{F}_k\langle m \in \mathbb{Y}\rangle_{\mathbb{H}_B}(x \in \mathbb{Y}) \in \mathbb{Y}_{96}$ is the set of valid singly-contextualized signatures of utilizing the secret counterpart to the public key k, some context x and message m.

$\mathbb{F}_r\langle m \in \mathbb{Y}\rangle_{\mathbb{Y}_R}(x \in \mathbb{Y}) \in \mathbb{Y}_{784}$, meanwhile, is the set of valid Bandersnatch RingVRF deterministic singly-contextualized proofs of knowledge of a secret within some set of secrets

identified by some root in the set of valid *roots* $\mathbb{Y}_R \quad \mathbb{Y}_{144}$. We denote $\mathsf{O}(\mathbf{s} \quad \mathbb{H}_B \ ) \quad \mathbb{Y}_R$ to be the root specific to the set of public key counterparts $\mathbf{s}$. A root implies a specific set of Bandersnatch key pairs, knowledge of one of the secrets would imply being capable of making a unique, valid—and anonymous—proof of knowledge of a unique secret within the set.

Both the Bandersnatch signature and RingVRF proof strictly imply that a member utilized their secret key in combination with both the context x and the message m; the difference is that the member is identified in the former and is anonymous in the latter. Furthermore, both define a VRF *output*, a high entropy hash influenced by x but not by m, formally denoted $\mathrm{Y}(\mathbb{F}_r^m \ x \ ) \quad \mathbb{H}$ and $\mathrm{Y}(\mathbb{F}_k^m \ x \ ) \quad \mathbb{H}$.

We define the function S as the signature function, such that $\mathsf{S}_k(m) \quad \mathbb{F}_k^m \ [] \quad \mathbb{E}_k \ m$. We assert that the ability to compute a result for this function relies on knowledge of a secret key.

4. Overview

As in the Yellow Paper, we begin our formalisms by recalling that a blockchain may be defined as a pairing of some initial state together with a block-level state-transition function. The latter defines the posterior state given a pairing of some prior state and a block of data applied to it. Formally, we say:

$$\sigma' \equiv \quad (\sigma, \mathbf{B}) \tag{12}$$

Where σ is the prior state, σ' is the posterior state, B is some valid block and is our block-level state-transition function.

Broadly speaking, JAM (and indeed blockchains in general) may be defined simply by specifying and some *genesis state* σ^0.[7] We also make several additional assumptions of agreed knowledge: a universally known clock, and the practical means of sharing data with other systems operating under the same consensus rules. The latter two were both assumptions silently made in the *YP*.

4.1. The Block.

To aid comprehension and definition of our protocol, we partition as many of our terms as possible into their functional components. We begin with the block B which may be restated as the header H and some input data external to the system and thus said to be *extrinsic*, $\mathbf{E}$:

$$\mathbf{B} \equiv (\mathbf{H}, \mathbf{E}) \tag{13}$$

$$\mathbf{E} \equiv (\mathbf{E}_T, \mathbf{E}_D, \mathbf{E}_P, \mathbf{E}_A, \mathbf{E}_G) \tag{14}$$

The header is a collection of metadata primarily concerned with cryptographic references to the blockchain ancestors and the operands and result of the present transition. As an immutable known *a priori*, it is assumed to be available throughout the functional components of block transition. The extrinsic data is split into its several portions:

tickets: Tickets, used for the mechanism which manages the selection of validators for the permissioning of block authoring. This component is denoted $\mathbf{E}_T$.

judgements: Votes, by validators, on dispute(s) arising between them presently taking place. This is denoted $\mathbf{E}_D$.

preimages: Static data which is presently being requested to be available for workloads to be able to fetch on demand. This is denoted $\mathbf{E}_P$.

availability: Assurances by each validator concerning which of the input data of workloads they have correctly received and are storing locally. This is denoted $\mathbf{E}_A$.

reports: Reports of newly completed workloads whose accuracy is guaranteed by specific validators. This is denoted $\mathbf{E}_G$.

4.2. The State.

Our state may be logically partitioned into several largely independent segments which can both help avoid visual clutter within our protocol description and provide formality over elements of computation which may be simultaneously calculated (i.e. parallelized). We therefore pronounce an equivalence between σ (some complete state) and a tuple of partitioned segments of that state:

$$\sigma \equiv (\alpha, \beta, \gamma, \delta, \eta, \iota, \kappa, \lambda, \rho, \tau, \varphi, \chi, \psi, \pi) \tag{15}$$

In summary, δ is the portion of state dealing with *services*, analogous in JAM to the Yellow Paper's (smart contract) *accounts*, the only state of the *YP*'s Ethereum. The identities of services which hold some privileged status are tracked in χ.

Validators, who are the set of economic actors uniquely privileged to help build and maintain the JAM chain, are identified within κ, archived in λ and enqueued from ι. All other state concerning the determination of these keys is held within γ. Note this is a departure from the *YP* proof-of-work definitions which were mostly stateless, and this set was not enumerated but rather limited to those with sufficient compute power to find a partial hash-collision in the SHA2-256 cryptographic hash function. An on-chain entropy pool is retained in η.

Our state also tracks two aspects of each core: α, the authorization requirement which work done on that core must satisfy at the time of being reported on-chain, together with the queue which fills this, φ; and ρ, each of the cores' currently assigned *report*, the availability of whose *work-package* must yet be assured by a super-majority of validators.

Finally, details of the most recent blocks and time are tracked in β and τ respectively and, judgements are tracked in ψ and validator statistics are tracked in π.

4.2.1. *State Transition Dependency Graph.*

Much as in the *YP*, we specify as the implication of formulating all items of posterior state in terms of the prior state and block. To aid the architecting of implementations which parallelize this computation, we minimize the depth of the dependency graph where possible. The overall dependency graph is specified here:

$$\tau' \quad \mathbf{H} \tag{16}$$

$$\beta^\dagger \quad (\mathbf{H}, \beta) \tag{17}$$

$$\beta' \quad (\mathbf{H}, \mathbf{E}_G, \beta^\dagger, \mathbf{C}) \tag{18}$$

[7] Practically speaking, blockchains sometimes make assumptions of some fraction of participants whose behavior is simply *honest*, and not provably incorrect nor otherwise economically disincentivized. While the assumption may be reasonable, it must nevertheless be stated apart from the rules of state-transition.

$$\gamma' \quad (\mathbf{H}, \tau, \mathbf{E}_T, \gamma, \iota, \eta', \kappa') \tag{19}$$

$$\eta' \quad (\mathbf{H}, \tau, \eta) \tag{20}$$

$$\kappa' \quad (\mathbf{H}, \tau, \kappa, \gamma, \psi') \tag{21}$$

$$\lambda' \quad (\mathbf{H}, \tau, \lambda, \kappa) \tag{22}$$

$$\psi' \quad (\mathbf{E}_D, \psi) \tag{23}$$

$$\delta^\dagger \quad (\mathbf{E}_P, \delta, \tau') \tag{24}$$

$$\rho^\dagger \quad (\mathbf{E}_D, \rho) \tag{25}$$

$$\rho^\ddagger \quad (\mathbf{E}_A, \rho^\dagger) \tag{26}$$

$$\rho' \quad (\mathbf{E}_G, \rho^\ddagger, \kappa, \tau') \tag{27}$$

$$\delta'$$

$$\chi'$$

$$\iota' \quad (\mathbf{E}_A, \rho', \delta^\dagger, \chi, \iota, \varphi) \tag{28}$$

$$\varphi'$$

$$\mathbf{C}$$

$$\alpha' \quad (\mathbf{E}_G, \varphi', \alpha) \tag{29}$$

$$\pi' \quad (\mathbf{E}_G, \mathbf{E}_P, \mathbf{E}_A, \mathbf{E}_T, \tau, \tau', \pi, \mathbf{H}) \tag{30}$$

The only synchronous entanglements are visible through the intermediate components superscripted with a dagger and defined in equations 17, 24 and 26. The latter two mark a merge and join in the dependency graph and, concretely, imply that the preimage lookup extrinsic must be folded into state before the availability extrinsic may be fully processed and accumulation of work happen.

4.3. **Which History?** A blockchain is a sequence of blocks, each cryptographically referencing some prior block by including a hash of its header, all the way back to some first block which references the genesis header. We already presume consensus over this genesis header $\mathbf{H}^0$ and the state it represents already defined as σ^0.

By defining a deterministic function for deriving a single posterior state for any (valid) combination of prior state and block, we are able to define a unique *canonical* state for any given block. We generally call the block with the most ancestors the *head* and its state the *head state*.

It is generally possible for two blocks to be valid and yet reference the same prior block in what is known as a *fork*. This implies the possibility of two different heads, each with their own state. While we know of no way to strictly preclude this possibility, for the system to be useful we must nonetheless attempt to minimize it. We therefore strive to ensure that:

(1) It be generally unlikely for two heads to form.
(2) When two heads do form they be quickly resolved into a single head.
(3) It be possible to identify a block not much older than the head which we can be extremely confident will form part of the blockchain's history in perpetuity. When a block becomes identified as such we call it *finalized* and this property naturally extends to all of its ancestor blocks.

These goals are achieved through a combination of two consensus mechanisms: *Safrole*, which governs the (not-necessarily forkless) extension of the blockchain; and *Grandpa*, which governs the finalization of some extension into canonical history. Thus, the former delivers point 1, the latter delivers point 3 and both are important for delivering point 2. We describe these portions of the protocol in detail in sections 6 and 19 respectively.

While Safrole limits forks to a large extent (through cryptography, economics and common-time, below), there may be times when we wish to intentionally fork since we have come to know that a particular chain extension must be reverted. In regular operation this should never happen, however we cannot discount the possibility of malicious or malfunctioning nodes. We therefore define such an extension as any which contains a block in which data is reported which *any other* block's state has tagged as invalid (see section 10 on how this is done). We further require that Grandpa not finalize any extension which contains such a block. See section 19 for more information here.

4.4. **Time.** We presume a pre-existing consensus over time specifically for block production and import. While this was not an assumption of Polkadot, pragmatic and resilient solutions exist including the NTP protocol and network. We utilize this assumption in only one way: we require that blocks be considered temporarily invalid if their timeslot is in the future. This is specified in detail in section 6.

Formally, we define the time in terms of seconds passed since the beginning of the JAM *Common Era*, 1200 UTC on January 1, 2024.[8] Midday CET is selected to ensure that all significant timezones are on the same date at any exact 24-hour multiple from the beginning of the common era. Formally, this value is denoted T.

4.5. **Best block.** Given the recognition of a number of valid blocks, it is necessary to determine which should be treated as the "best" block, by which we mean the most recent block we believe will ultimately be within of all future JAM chains. The simplest and least risky means of doing this would be to inspect the Grandpa finality mechanism which is able to provide a block for which there is a very high degree of confidence it will remain an ancestor to any future chain head.

However, in reducing the risk of the resulting block ultimately not being within the canonical chain, Grandpa will typically return a block some small period older than the most recently authored block. (Existing deployments suggest around 1-2 blocks in the past under regular operation.) There are often circumstances when we may wish to have less latency at the risk of the returned block not ultimately forming a part of the future canonical chain. E.g. we may be in a position of being able to author a block, and we need to decide what its parent should be. Alternatively, we may care to speculate about the most recent state for the purpose of providing information to a downstream application reliant on the state of JAM.

In these cases, we define the best block as the head of the best chain, itself defined in section 19.

4.6. **Economics.** The present work describes a crypto-economic system, i.e. one combining elements of both cryptography and economics and game theory to deliver a self-sovereign digital service. In order to codify and manipulate economic incentives we define a token which is

[8] 1,704,110,400 seconds after the Unix Epoch.

native to the system, which we will simply call *tokens* in the present work.

A value of tokens is generally referred to as a *balance*, and such a value is said to be a member of the set of balances, $\mathbb{N}_B$, which is exactly equivalent to the set of naturals less than 2^{64} (i.e. 64-bit unsigned integers in coding parlance). Formally:

$$\mathbb{N}_B \equiv \mathbb{N}_{2^{64}} \tag{31}$$

Though unimportant for the present work, we presume that there be a standard named denomination for 10^9 tokens. This is different to both Ethereum (which uses a denomination of 10^{18}), Polkadot (which uses a denomination of 10^{10}) and Polkadot's experimental cousin Kusama (which uses 10^{12}).

The fact that balances are constained to being less than 2^{64} implies that there may never be more than around 18×10^9 tokens (each divisible into portions of 10^{-9}) within JAM. We would expect that the total number of tokens ever issued will be a substantially smaller amount than this.

We further presume that a number of constant *prices* stated in terms of tokens are known. However we leave the specific values to be determined in following work:

- B_I: the additional minimum balance implied for a single item within a mapping.
- B_L: the additional minimum balance implied for a single octet of data within a mapping.
- B_S: the minimum balance implied for a service.

4.7. **The Virtual Machine and Gas.** In the present work, we presume the definition of a *Polka Virtual Machine* (PVM). This virtual machine is based around the RISC-V instruction set architecture, specifically the RV32EM variant, and is the basis for introducing permissionless logic into our state-transition function.

The PVM is comparable to the EVM defined in the Yellow Paper, but somewhat simpler: the complex instructions for cryptographic operations are missing as are those which deal with environmental interactions. Overall it is far less opinionated since it alters a pre-existing general purpose design, RISC-V, and optimizes it for our needs. This gives us excellent pre-existing tooling, since PVM remains essentially compatible with RISC-V, including support from the compiler toolkit LLVM and languages such as Rust and C++. Furthermore, the instruction set simplicity which RISC-V and PVM share, together with the register size (32-bit), active number (13) and endianness (little) make it especially well-suited for creating efficient recompilers on to common hardware architectures.

The PVM is fully defined in appendix A, but for contextualization we will briefly summarize the basic invocation function which computes the resultant state of a PVM instance initialized with some registers ($\mathbb{N}_{R\ 13}$) and RAM ($\mathbb{M}$) and has executed for up to some amount of gas ($\mathbb{N}_G$), a number of approximately time-proportional computational steps:

$$\begin{matrix}\mathbb{Y},\ \mathbb{N}_R,\ \mathbb{N}_G,\\ \mathbb{N}_{R\ 13},\ \mathbb{M}\end{matrix} \rightarrow \begin{matrix}\{\ ,\ ,\infty\}\ \ \{\mathrm{S},\hbar\}\times\mathbb{N}_R,\\ \mathbb{N}_R,\ \mathbb{Z}_G,\ \ \mathbb{N}_{R\ 13},\ \mathbb{M}\end{matrix} \tag{32}$$

We refer to the time-proportional computational steps as *gas* (much like in the *YP*) and limit it to a 64-bit quantity. We may use either $\mathbb{N}_G$ or $\mathbb{Z}_G$ to bound it, the first as a prior argument since it is known to be positive, the latter as a result where a negative value indicates an attempt to execute beyond the gas limit. Within the context of the PVM, ξ $\mathbb{N}_G$ is typically used to denote gas.

$$\mathbb{Z}_G \equiv \mathbb{Z}_{-2^{63}\,2^{63}}\ ,\quad \mathbb{N}_G \equiv \mathbb{N}_{2^{64}}\ ,\quad \mathbb{N}_R \equiv \mathbb{N}_{2^{32}} \tag{33}$$

It is left as a rather important implementation detail to ensure that the amount of time taken while computing the function $(\dots,\xi,\dots)$ has a maximum computation time approximately proportional to the value of ξ regardless of other operands.

The PVM is a very simple RISC *register machine* and as such has 13 registers, each of which is a 32-bit quantity, denoted as $\mathbb{N}_R$, a natural less than 2^{32}.[9] Within the context of the PVM, ω $\mathbb{N}_{R\ 13}$ is typically used to denote the registers.

$$\mathbb{M} \equiv \ \mathbf{V}\ \ \mathbb{Y}_{2^{32}},\ \mathbf{A}\ \ \{\mathrm{W},\mathrm{R},\ \}_{\ 2^{32}} \tag{34}$$

The PVM assumes a simple pageable RAM of 32-bit addressable octets where each octet may be either immutable, mutable or inaccessible. The RAM definition $\mathbb{M}$ includes two components: a value $\mathbf{V}$ and access $\mathbf{A}$. If the component is unspecified while being subscripted then the value component may be assumed. Within the context of the virtual machine, μ $\mathbb{M}$ is typically used to denote RAM.

$$\mathbb{V}\ \equiv\{i\ \ \mu_{\mathbf{A}}[i]\neq\ \}\qquad \mathbb{V}\ \equiv\{i\ \ \mu_{\mathbf{A}}[i]=\mathrm{W}\} \tag{35}$$

We define two sets of indices for the RAM μ: $\mathbb{V}$ is the set of indices which may be read from; and $\mathbb{V}$ is the set of indices which may be written to.

Invocation of the PVM has an exit-reason as the first item in the resultant tuple. It is either:

- Regular program termination caused by an explicit halt instruction, .
- Irregular program termination caused by some exceptional circumstance, .
- Exhaustion of gas, ∞.
- A page fault (attempt to access some address in RAM which is not accessible), S. This includes the address at fault.
- An attempt at progressing a host-call, $\hbar$. This allows for the progression and integration of a context-dependent state-machine beyond the regular PVM.

The full definition follows in appendix A.

4.8. **Epochs and Slots.** Unlike the *YP* Ethereum with its proof-of-work consensus system, JAM defines a proof-of-authority consensus mechanism, with the authorized validators presumed to be identified by a set of public keys and decided by a *staking* mechanism residing within some system hosted by JAM. The staking system is out of scope for the present work; instead there is an API which may be utilized to update these keys, and we presume that whatever logic is needed for the staking system will be introduced and utilize this API as needed.

The Safrole mechanism subdivides time following genesis into fixed length *epoch*s with each epoch divided into

[9] This is three fewer than RISC-V's 16, however the amount that program code output by compilers uses is 13 since two are reserved for operating system use and the third is fixed as zero

$E = 600$ time*slot*s each of uniform length $P = 6$ seconds, given an epoch period of $E \cdot P = 3600$ seconds or one hour.

This six-second slot period represents the minimum time between JAM blocks, and through Safrole we aim to strictly minimize forks arising both due to contention within a slot (where two valid blocks may be produced within the same six-second period) and due to contention over multiple slots (where two valid blocks are produced in different time slots but with the same parent).

Formally when identifying a timeslot index, we use a natural less than 2^{32} (in compute parlance, a 32-bit unsigned integer) indicating the number of six-second timeslots from the JAM Common Era. For use in this context we introduce the set $\mathbb{N}_T$:

$$\mathbb{N}_T \equiv \mathbb{N}_{2^{32}} \tag{36}$$

This implies that the lifespan of the proposed protocol takes us to mid-August of the year 2840, which with the current course that humanity is on should be ample.

4.9. **The Core Model and Services.** Whereas in the Ethereum Yellow Paper when defining the state machine which is held in consensus amongst all network participants, we presume that all machines maintaining the full network state and contributing to its enlargement—or, at least, hoping to—evaluate all computation. This "everybody does everything" approach might be called the *on-chain consensus model*. It is unfortunately not scalable, since the network can only process as much logic in consensus that it could hope any individual node is capable of doing itself within any given period of time.

4.9.1. *In-core Consensus.* In the present work, we achieve scalability of the work done through introducing a second model for such computation which we call the *in-core consensus model*. In this model, and under normal circumstances, only a subset of the network is responsible for actually executing any given computation and assuring the availability of any input data it relies upon to others. By doing this and assuming a certain amount of computational parallelism within the validator nodes of the network, we are able to scale the amount of computation done in consensus commensurate with the size of the network, and not with the computational power of any single machine. In the present work we expect the network to be able to do upwards of 300 times the amount of computation *in-core* as that which could be performed by a single machine running the virtual machine at full speed.

Since in-core consensus is not evaluated or verified by all nodes on the network, we must find other ways to become adequately confident that the results of the computation are correct, and any data used in determining this is available for a practical period of time. We do this through a crypto-economic game of three stages called *guaranteeing*, *assuring*, *auditing* and, potentially, *judging*. Respectively, these attach a substantial economic cost to the invalidity of some proposed computation; then a sufficient degree of confidence that the inputs of the computation will be available for some period of time; and finally, a sufficient degree of confidence that the validity of the computation (and thus enforcement of the first guarantee) will be checked by some party who we can expect to be honest.

All execution done in-core must be reproducible by any node synchronized to the portion of the chain which has been finalized. Execution done in-core is therefore designed to be as stateless as possible. The requirements for doing it include only the refinement code of the service, the code of the authorizer and any preimage lookups it carried out during its execution.

When a work-report is presented on-chain, a specific block known as the *lookup-anchor* is identified. Correct behavior requires that this must be in the finalized chain and reasonably recent, both properties which may be proven and thus are acceptable for use within a consensus protocol.

We describe this pipeline in detail in the relevant sections later.

4.9.2. *On Services and Accounts.* In *YP* Ethereum, we have two kinds of accounts: *contract accounts* (whose actions are defined deterministically based on the account's associated code and state) and *simple accounts* which act as gateways for data to arrive into the world state and are controlled by knowledge of some secret key. In JAM, all accounts are *service accounts*. Like Ethereum's contract accounts, they have an associated balance, some code and state. Since they are not controlled by a secret key, they do not need a nonce.

The question then arises: how can external data be fed into the world state of JAM? And, by extension, how does overall payment happen if not by deducting the account balances of those who sign transactions? The answer to the first lies in the fact that our service definition actually includes *multiple* code entry-points, one concerning *refinement* and the other concerning *accumulation*. The former acts as a sort of high-performance stateless processor, able to accept arbitrary input data and distill it into some much smaller amount of output data. The latter code is more stateful, providing access to certain on-chain functionality including the possibility of transferring balance and invoking the execution of code in other services. Being stateful this might be said to more closely correspond to the code of an Ethereum contract account.

To understand how JAM breaks up its service code is to understand JAM's fundamental proposition of generality and scalability. All data extrinsic to JAM is fed into the refinement code of some service. This code is not executed *on-chain* but rather is said to be executed *in-core*. Thus, whereas the accumulator code is subject to the same scalability constraints as Ethereum's contract accounts, refinement code is executed off-chain and subject to no such constraints, enabling JAM services to scale dramatically both in the size of their inputs and in the complexity of their computation.

While refinement and accumulation take place in consensus environments of a different nature, both are executed by the members of the same validator set. The JAM protocol through its rewards and penalties ensures that code executed *in-core* has a comparable level of crypto-economic security to that executed *on-chain*, leaving the primary difference between them one of scalability versus synchroneity.

As for managing payment, JAM introduces a new abstraction mechanism based around Polkadot's Agile Coretime. Within the Ethereum transactive model, the mechanism of account authorization is somewhat combined with the mechanism of purchasing blockspace, both relying on a cryptographic signature to identify a single "transactor" account. In JAM, these are separated and there is no such concept of a "transactor".

In place of Ethereum's gas model for purchasing and measuring blockspace, JAM has the concept of *coretime*, which is prepurchased and assigned to an authorization agent. Coretime is analogous to gas insofar as it is the underlying resource which is being consumed when utilizing JAM. Its procurement is out of scope in the present work and is expected to be managed by a system parachain operating within a parachains service itself blessed with a number of cores for running such system services. The authorization agent allows external actors to provide input to a service without necessarily needing to identify themselves as with Ethereum's transaction signatures. They are discussed in detail in section 8.

5. The Header

We must first define the header in terms of its components. The header comprises a parent hash and prior state root ($\mathbf{H}_p$ and $\mathbf{H}_r$), an extrinsic hash $\mathbf{H}_x$, a time-slot index $\mathbf{H}_t$, the epoch, winning-tickets, verdicts and offenders markers $\mathbf{H}_e$, $\mathbf{H}_w$, $\mathbf{H}_j$ and $\mathbf{H}_o$, a Bandersnatch block author index $\mathbf{H}_i$ and two Bandersnatch signatures; the entropy-yielding VRF signature $\mathbf{H}_v$ and a block seal $\mathbf{H}_s$. Headers may be serialized to an octet sequence with and without the latter seal component using E and E_U respectively. Formally:

$$\mathbf{H} \equiv (\mathbf{H}_p, \mathbf{H}_r, \mathbf{H}_x, \mathbf{H}_t, \mathbf{H}_e, \mathbf{H}_w, \mathbf{H}_j, \mathbf{H}_o, \mathbf{H}_i, \mathbf{H}_v, \mathbf{H}_s) \tag{37}$$

Blocks considered invalid by this rule may become valid as T advances.

The blockchain is a sequence of blocks, each cryptographically referencing some prior block by including a hash derived from the parent's header, all the way back to some first block which references the genesis header. We already presume consensus over this genesis header $\mathbf{H}^0$ and the state it represents defined as σ^0.

Excepting the Genesis header, all block headers $\mathbf{H}$ have an associated parent header, whose hash is $\mathbf{H}_p$. We denote the parent header $\mathbf{H}^- = P(\mathbf{H})$:

$$\mathbf{H}_p \in \mathbb{H}, \quad \mathbf{H}_p \equiv \mathrm{H}(\mathrm{E}(P(\mathbf{H}))) \tag{38}$$

P is thus defined as being the mapping from one block header to its parent block header. With P, we are able to define the set of ancestor headers $\mathbf{A}$:

$$h \in \mathbf{A} \Leftrightarrow h = \mathbf{H} \quad (\ i \in \mathbf{A} \quad h = P(i)) \tag{39}$$

We only require implementations to store headers of ancestors which were authored in the previous $\mathrm{L} = 24$ hours of any block $\mathbf{B}$ they wish to validate.

The extrinsic hash is the hash of the block's extrinsic data. Given any block $\mathbf{B} = (\mathbf{H}, \mathrm{E})$, then formally:

$$\mathbf{H}_x \in \mathbb{H}, \quad \mathbf{H}_x \equiv \mathrm{H}(\mathrm{E}(\mathbf{E})) \tag{40}$$

A block may only be regarded as valid once the time-slot index $\mathbf{H}_t$ is in the past. It is always strictly greater than that of its parent. Formally:

$$\mathbf{H}_t \in \mathbb{N}_T, \quad P(\mathbf{H})_t < \mathbf{H}_t \qquad \mathbf{H}_t \ \mathrm{P} \le \mathrm{T} \tag{41}$$

The parent state root $\mathbf{H}_r$ is the root of a Merkle trie composed by the mapping of the *prior* state's Merkle root, which by definition is also the parent block's posterior state. This is a departure from both Polkadot and the Yellow Paper's Ethereum, in both of which a block's header contains the *posterior* state's Merkle root. We do this to facilitate the pipelining of block computation and in particular of Merklization.

$$\mathbf{H}_r \in \mathbb{H}, \quad \mathbf{H}_r \equiv \mathrm{M}\ (\sigma) \tag{42}$$

We assume the state-Merklization function M is capable of transforming our state σ into a 32-octet commitment. See appendix D for a full definition of these two functions.

All blocks have an associated public key to identify the author of the block. We identify this as an index into the posterior current validator set κ'. We denote the Bandersnatch key of the author as $\mathbf{H}_a$ though note that this is merely an equivalence, and is not serialized as part of the header.

$$\mathbf{H}_i \in \mathbb{N}_{\mathrm{V}}, \quad \mathbf{H}_a \equiv \kappa'[\mathbf{H}_i] \tag{43}$$

5.1. **The Markers.** If not , then the epoch marker specifies key and entropy relevant to the following epoch in case the ticket contest does not complete adequately (a very much unexpected eventuality). Similarly, the winning-tickets marker, if not , provides the series of 600 slot sealing "tickets" for the next epoch (see the next section). Finally, the verdicts and offenders markers are the sequence of report hashes newly judged as not good and the sequence of Ed25519 keys of newly misbehaving validators, to be fully explained in section 10. Formally:

$$\mathbf{H}_e \in \mathbb{H}, \ \mathbb{H}_{B\ \mathrm{V}}? \qquad \mathbf{H}_w \in \mathbb{C}_{\ \mathrm{E}}? \tag{44}$$

$$\mathbf{H}_j \in \mathbb{H} \qquad \mathbf{H}_o \in \mathbb{H}_E \tag{45}$$

The terms are fully defined in sections 6.6 and 10.

6. Block Production and Chain Growth

As mentioned earlier, JAM is architected around a hybrid consensus mechanism, similar in nature to that of Polkadot's BABE/GRANDPA hybrid. JAM's block production mechanism, termed Safrole after the novel Sassafras production mechanism of which it is a simplified variant, is a stateful system rather more complex than the Nakamoto consensus described in the *YP*.

The chief purpose of a block production consensus mechanism is to limit the rate at which new blocks may be authored and, ideally, preclude the possibility of "forks": multiple blocks with equal numbers of ancestors.

To achieve this, Safrole limits the possible author of any block within any given six-second timeslot to a single key-holder from within a prespecified set of *validators*. Furthermore, under normal operation, the identity of the key-holder of any future timeslot will have a very high degree of anonymity. As a side effect of its operation, we can generate a high-quality pool of entropy which may be used by other parts of the protocol and is accessible to services running on it.

Because of its tightly scoped role, the core of Safrole's state, γ, is independent of the rest of the protocol. It interacts with other portions of the protocol through ι and

κ, the prospective and active sets of validator keys respectively; τ, the most recent block's timeslot; and η, the entropy accumulator.

The Safrole protocol generates, once per epoch, a sequence of E *sealing keys*, one for each potential block within a whole epoch. Each block header includes its timeslot index $\mathbf{H}_t$ (the number of six-second periods since the JAM Common Era began) and a valid seal signature $\mathbf{H}_s$, signed by the sealing key corresponding to the timeslot within the aforementioned sequence. Each sealing key is in fact a pseudonym for some validator which was agreed the privilege of authoring a block in the corresponding timeslot.

In order to generate this sequence of sealing keys in regular operation, and in particular to do so without making public the correspondence relation between them and the validator set, we use a novel cryptographic structure known as a RingVRF, utilizing the Bandersnatch curve. Bandersnatch RingVRF allows for a proof to be provided which simultaneously guarantees the author controlled a key within a set (in our case validators), and secondly provides an output, an unbiasable deterministic hash giving us a secure verifiable random function (VRF). This anonymous and secure random output is a *ticket* and validators' tickets with the best score define the new sealing keys allowing the chosen validators to exercise their privilege and create a new block at the appropriate time.

6.1. **Timekeeping.** Here, τ defines the most recent block's slot index, which we transition to the slot index as defined in the block's header:

$$\tau \in \mathbb{N}_T, \quad \tau' \equiv \mathbf{H}_t \tag{46}$$

We track the slot index in state as τ in order that we are able to easily both identify a new epoch and determine the slot at which the prior block was authored. We denote e as the prior's epoch index and m as the prior's slot phase index within that epoch and e' and m' are the corresponding values for the present block:

$$\text{let } e \mid m = \frac{\tau}{\mathrm{E}}, \quad e' \mid m' = \frac{\tau'}{\mathrm{E}} \tag{47}$$

6.2. **Safrole Basic State.** We restate γ into a number of components:

$$\gamma \equiv (\gamma_{\mathbf{k}}, \gamma_z, \gamma_{\mathbf{s}}, \gamma_{\mathbf{a}}) \tag{48}$$

γ_z is the epoch's root, a Bandersnatch ring root composed with the one Bandersnatch key of each of the next epoch's validators, defined in $\gamma_{\mathbf{k}}$ (itself defined in the next section).

$$\gamma_z \in \mathbb{Y}_R \tag{49}$$

Finally, $\gamma_{\mathbf{a}}$ is the ticket accumulator, a series of highest-scoring ticket identifiers to be used for the next epoch. $\gamma_{\mathbf{s}}$ is the current epoch's slot-sealer series, which is either a full complement of E tickets or, in the case of a fallback mode, a series of E Bandersnatch keys:

$$\gamma_{\mathbf{a}} \in \mathbb{C}_{\mathrm{E}}, \quad \gamma_{\mathbf{s}} \in \mathbb{C}_{\mathrm{E}} \cup \mathbb{H}_{B\,\mathrm{E}} \tag{50}$$

Here, $\mathbb{C}$ is used to denote the set of *tickets*, a combination of a verifiably random ticket identifier $\mathbf{y}$ and the ticket's entry-index r:

$$\mathbb{C} \equiv (\mathbf{y} \in \mathbb{H}, r \in \mathbb{N}_{\mathrm{N}}) \tag{51}$$

As we state in section 6.4, Safrole requires that every block header $\mathbf{H}$ contain a valid seal $\mathbf{H}_s$, which is a Bandersnatch signature for a public key at the appropriate index m of the current epoch's seal-key series, present in state as $\gamma_{\mathbf{s}}$.

6.3. **Key Rotation.** In addition to the active set of validator keys κ and staging set ι, internal to the Safrole state we retain a pending set $\gamma_{\mathbf{k}}$. The active set is the set of keys identifying the nodes which are currently privileged to author blocks and carry out the validation processes, whereas the pending set $\gamma_{\mathbf{k}}$, which is reset to ι at the beginning of each epoch, is the set of keys which will be active in the next epoch and which determine the Bandersnatch ring root which authorizes tickets into the sealing-key contest for the next epoch.

$$\iota \in \mathbb{K}_{\mathrm{V}}, \quad \gamma_{\mathbf{k}} \in \mathbb{K}_{\mathrm{V}}, \quad \kappa \in \mathbb{K}_{\mathrm{V}}, \quad \lambda \in \mathbb{K}_{\mathrm{V}} \tag{52}$$

We must introduce $\mathbb{K}$, the set of validator key tuples. This is a combination of a set of cryptographic public keys and metadata which is an opaque octet sequence, but utilized to specify practical identifiers for the validator, not least a hardware address.

The set of validator keys itself is equivalent to the set of 336-octet sequences. However, for clarity, we divide the sequence into four easily denoted components. For any validator key k, the Bandersnatch key is denoted k_b, and is equivalent to the first 32-octets; the Ed25519 key, k_e, is the second 32 octets; the BLS key denoted k_{BLS} is equivalent to the following 144 octets, and finally the metadata k_m is the last 128 octets. Formally:

$$\mathbb{K} \equiv \mathbb{Y}_{336} \tag{53}$$

$$k \in \mathbb{K} \Rightarrow k_b \in \mathbb{H}_B \equiv k_{0\,+32} \tag{54}$$

$$k \in \mathbb{K} \Rightarrow k_e \in \mathbb{H}_E \equiv k_{32\,+32} \tag{55}$$

$$k \in \mathbb{K} \Rightarrow k_{BLS} \in \mathbb{Y}_{BLS} \equiv k_{64\,+144} \tag{56}$$

$$k \in \mathbb{K} \Rightarrow k_m \in \mathbb{Y}_{128} \equiv k_{208\,+128} \tag{57}$$

With a new epoch under regular conditions, validator keys get rotated and the epoch's Bandersnatch key root is updated into γ'_z:

$$(\gamma'_{\mathbf{k}}, \kappa', \lambda', \gamma'_z) \equiv \begin{cases} (\iota, \gamma_{\mathbf{k}}, \kappa, z) & \text{if } e' > e \\ (\gamma_{\mathbf{k}}, \kappa, \lambda, \gamma_z) & \text{otherwise} \end{cases} \tag{58}$$

$$\text{where } z = \mathrm{O}([k_b \mid k \in \gamma'_{\mathbf{k}}])$$

$$\mathbf{k}) \equiv [\begin{cases} [0,0,\dots] & \text{if } k_e \in \psi'_{\mathbf{o}} \\ k & \text{otherwise} \end{cases} \mid k \in \mathbf{k}] \tag{59}$$

Note that on epoch changes the posterior queued validator key set γ'_k is defined such that incoming keys belonging to the offenders $\psi'_{\mathbf{o}}$ are replaced with a null key containing only zeroes. The origin of the offenders is explained in section 10.

6.4. **Sealing and Entropy Accumulation.** The header must contain a valid seal and valid VRF output. These are two signatures both using the current slot's seal key; the message data of the former is the header's serialization omitting the seal component $\mathbf{H}_s$, whereas the latter is used as a bias-resistant entropy source and thus its message must already have been fixed: we use the entropy stemming from the VRF of the seal signature. Formally:

let $i = \gamma'_{\mathbf{s}}[\mathbf{H}_t]$ ▫

$$(60)\qquad \gamma_{\mathbf{s}}' \quad \mathbb{C} \quad \begin{array}{l} i_{\mathbf{y}} = \mathrm{Y}(\mathrm{H}_s), \\ \mathrm{H}_s \quad \mathbb{F}^{\mathrm{E}_U(\mathbf{H})}_{\mathbf{H}_a} \ \mathrm{X}_T \quad \eta_3 +\!\!+ i_r \ , \\ \mathrm{T} = 1 \\ i = \mathrm{H}_a \, , \end{array}$$

$$(61)\qquad \gamma_{\mathbf{s}}' \quad \mathbb{H}_B \quad \begin{array}{l} \mathrm{H}_s \quad \mathbb{F}^{\mathrm{E}_U(\mathbf{H})}_{\mathbf{H}_a} \ \mathrm{X}_F \quad \eta_3 \ , \\ \mathrm{T} = 0 \end{array}$$

$$(62)\qquad \mathrm{H}_v \quad \mathbb{F}^{[]}_{\mathbf{H}_a} \ \mathrm{X}_E \quad \mathrm{Y}(\mathrm{H}_s)$$

$$(63)\qquad \mathrm{X}_E = \texttt{Ma9h¬Oj§}\}\texttt{nq}'$$

$$(64)\qquad \mathrm{X}_F = \texttt{Ma9h¬T9ee@9Hc¬ O9e}$$

$$(65)\qquad \mathrm{X}_T = \texttt{Ma9h¬§\_HcO§¬ O9e}$$

Sealing using the ticket is of greater security, and we utilize this knowledge when determining a candidate block on which to extend the chain, detailed in section 19. We thus note that the block was sealed under the regular security with the boolean marker T. We define this only for the purpose of ease of later specification.

In addition to the entropy accumulator η_0, we retain three additional historical values of the accumulator at the point of each of the three most recently ended epochs, η_1, η_2 and η_3. The second-oldest of these η_2 is utilized to help ensure future entropy is unbiased (see equation 74) and seed the fallback seal-key generation function with randomness (see equation 69). The oldest is used to regenerate this randomness when verifying the seal above (see equations 61 and 60).

$$(66)\qquad \eta \quad \mathbb{H}_4$$

η_0 defines the state of the randomness accumulator to which the provably random output of the VRF, the signature over some unbiasable input, is combined each block. η_1 and η_2 meanwhile retain the state of this accumulator at the end of the two most recently ended epochs in order.

$$(67)\qquad \eta_0' \equiv \mathrm{H}(\eta_0 \quad \mathrm{Y}(\mathrm{H}_v))$$

On an epoch transition (identified as the condition $e' > e$), we therefore rotate the accumulator value into the history η_1, η_2 and η_3:

$$(68)\qquad (\eta_1', \eta_2', \eta_3') \equiv \begin{cases} (\eta_0, \eta_1, \eta_2) & \text{if } e' > e \\ (\eta_1, \eta_2, \eta_3) & \text{otherwise} \end{cases}$$

6.5. **The Slot Key Sequence.** The posterior slot key sequence $\gamma_{\mathbf{s}}'$ is one of three expressions depending on the circumstance of the block. If the block is not the first in an epoch, then it remains unchanged from the prior $\gamma_{\mathbf{s}}$. If the block signals the next epoch (by epoch index) and the previous block's slot was within the closing period of the previous epoch, then it takes the value of the prior ticket accumulator $\gamma_{\mathbf{a}}$. Otherwise, it takes the value of the fallback key sequence. Formally:

$$(69)\qquad \gamma_{\mathbf{s}}' \equiv \begin{cases} Z(\gamma_{\mathbf{a}}) & \text{if } e' = e + 1 \quad m \geq \mathrm{Y} \quad \gamma_{\mathbf{a}} = \mathrm{E} \\ \gamma_{\mathbf{s}} & \text{if } e' = e \\ F(\eta_2', \kappa') & \text{otherwise} \end{cases}$$

Here, we use Z as the outside-in sequencer function, defined as follows:

$$(70)\qquad Z \ \{ \ \begin{array}{l} \mathbb{C}_{\mathrm{E}} \to \mathbb{C}_{\mathrm{E}} \\ \mathbf{s} \quad [\mathbf{s}_0, \mathbf{s}_{\mathbf{s}-1}, \mathbf{s}_1, \mathbf{s}_{\mathbf{s}-2}, \ldots] \end{array}$$

Finally, F is the fallback key sequence function which selects an epoch's worth of validator Bandersnatch keys ($\mathbb{H}_{B\ \mathrm{E}}$) from the validator key set $\mathbf{k}$ using the entropy collected on-chain r:

$$(71)\qquad F \quad \begin{array}{l} \mathbb{H}, \ \mathbb{K} \quad \to \quad \mathbb{H}_{B\ \mathrm{E}} \\ r, \mathbf{k} \quad [\mathbf{k}[\mathrm{E}^{-1}(\mathrm{H}_4(r \quad \mathrm{E}_4(i)))]^{\square}_b \quad i \quad \mathbb{N}_{\mathrm{E}}] \end{array}$$

6.6. **The Markers.** The epoch and winning-tickets markers are information placed in the header in order to minimize data transfer necessary to determine the validator keys associated with any given epoch. They are particularly useful to nodes which do not synchronize the entire state for any given block since they facilitate the secure tracking of changes to the validator key sets using only the chain of headers.

As mentioned earlier, the header's epoch marker H_e is either empty or, if the block is the first in a new epoch, then a tuple of the epoch randomness and a sequence of Bandersnatch keys defining the Bandersnatch validator keys (k_b) beginning in the next epoch. Formally:

$$(72)\qquad \mathrm{H}_e \equiv \begin{cases} (\eta_1', [k_b \quad k \in \gamma_{\mathbf{k}}']) & \text{if } e' > e \\ \text{otherwise} \end{cases}$$

The winning-tickets marker H_w is either empty or, if the block is the first after the end of the submission period for tickets and if the ticket accumulator is saturated, then the final sequence of ticket identifiers. Formally:

$$(73)\qquad \mathrm{H}_w \equiv \begin{cases} Z(\gamma_{\mathbf{a}}) & \text{if } e' = e \quad m < \mathrm{Y} \leq m' \quad \gamma_{\mathbf{a}} = \mathrm{E} \\ \text{otherwise} \end{cases}$$

6.7. **The Extrinsic and Tickets.** The extrinsic E_T is a sequence of proofs of valid tickets; a ticket implies an entry in our epochal "contest" to determine which validators are privileged to author a block for each timeslot in the following epoch. Tickets specify an entry index together with a proof of ticket's validity. The proof implies a ticket identifier, a high-entropy unbiasable 32-octet sequence, which is used both as a score in the aforementioned contest and as input to the on-chain VRF.

Towards the end of the epoch (i.e. Y slots from the start) this contest is closed implying successive blocks within the same epoch must have an empty tickets extrinsic. At this point, the following epoch's seal key sequence becomes fixed.

We define the extrinsic as a sequence of proofs of valid tickets, each of which is a tuple of an entry index (a natural number less than N) and a proof of ticket validity. Formally:

$$(74)\qquad \mathrm{E}_T \quad r \quad \mathbb{N}_{\mathrm{N}}, p \quad \mathbb{F}^{[]}_z \ \mathrm{X}_T \quad \eta_2' +\!\!+ r$$

$$(75)\qquad \mathrm{E}_T \ \leq \ \begin{cases} \mathrm{K} & \text{if } m' < \mathrm{Y} \\ 0 & \text{otherwise} \end{cases}$$

We define $\mathbf{n}$ as the set of new tickets, with the ticket identifier, a hash, defined as the output component of the Bandersnatch RingVRF proof:

$$(76)\qquad \mathbf{n} \equiv [\ \mathbf{y} \ \mathrm{Y}(i_p), r \ i_r \quad i \in \mathrm{E}_T]$$

The tickets submitted via the extrinsic must already have been placed in order of their implied identifier. Duplicate identifiers are never allowed lest a validator submit the same ticket multiple times:

$$(77)\qquad \mathbf{n} = [x_{\mathbf{y}} \quad x \quad \mathbf{n}]$$

$$(78)\qquad \{x_{\mathbf{y}} \quad x \quad \mathbf{n}\} \quad \{x_{\mathbf{y}} \quad x \quad \gamma_{\mathbf{a}}\}$$

The new ticket accumulator $\gamma_{\mathbf{a}}'$ is constructed by merging new tickets into the previous accumulator value (or the empty sequence if it is a new epoch):

$$\gamma_{\mathbf{a}}' \equiv \begin{cases} x_{\mathbf{y}} \quad x \quad \mathbf{n} & \text{if } e' >^{\to \mathrm{E}} e \\ \gamma_{\mathbf{a}} & \text{otherwise} \end{cases} \tag{79}$$

The maximum size of the ticket accumulator is E. On each block, the accumulator becomes the lowest items of the sorted union of tickets from prior accumulator $\gamma_{\mathbf{a}}$ and the submitted tickets. It is invalid to include useless tickets in the extrinsic, so all submitted tickets must exist in their posterior ticket accumulator. Formally:

$$\mathbf{n} \quad \gamma_{\mathbf{a}}' \tag{80}$$

Note that it can be shown that in the case of an empty extrinsic $\mathbf{E}_T = []$, as implied by $m' \geq \mathrm{Y}$, then $\gamma_{\mathbf{a}}' = \gamma_{\mathbf{a}}$.

7. Recent History

We retain in state information on the most recent H blocks. This is used to preclude the possibility of duplicate or out of date work-reports from being submitted.

$$\beta \quad h \quad \mathbb{H}, \mathbf{b} \quad \mathbb{H}?, s \quad \mathbb{H}, \mathbf{p} \quad \mathbb{H}_{\mathrm{C}\ \mathrm{H}} \tag{81}$$

For each recent block, we retain its header hash, its state root, its accumulation-result MMR and the hash of each work-report made into it which is no more than the total number of cores, $\mathrm{C} = 341$.

During the accumulation stage, a value with the partial transition of this state is provided which contains the update for the newly-known roots of the parent block:

$$\beta^\dagger \equiv \beta \quad \text{except} \quad \beta^\dagger[\beta - 1]_s = \mathbf{H}_r \tag{82}$$

We define an item n comprising the new block's header hash, its accumulation-result Merkle tree root and the set of work-reports made into it (for which we use the guarantees extrinsic, $\mathbf{E}_G$). Note that the accumulation-result tree root r is derived from $\mathbf{C}$ (defined in section 12) using the basic binary Merklization function $\mathbf{M}_B$ (defined in appendix E) and appending it using the MMR append function A (defined in appendix E.2) to form a Merkle mountain range.

$$\begin{aligned} &\text{let } r = \mathbf{M}_B([s \quad \mathrm{E}_4(s) \quad \mathrm{E}(h) \quad (s, h) \quad \mathbf{C}], \mathrm{H}_K) \\ &\text{let } \mathbf{b} = \mathrm{A}(\text{last}([[]] \quad [x_{\mathbf{b}} \quad x \lessdot \beta]), r) \\ &\text{let } n = \mathbf{p} \ [((g_w)_s)_h \quad g \lessdot \mathbf{E}_G], h \ \ \mathrm{H}(\mathbf{H}), \mathbf{b}, s \ \ \mathbb{H}^0 \end{aligned} \tag{83}$$

The state-trie root is as being the zero hash, $\mathbb{H}^0$ which while inaccurate at the end state of the block β', it is nevertheless safe since β' is not utilized except to define the next block's $\beta^\dagger$, which contains a corrected value for this.

The final state transition is then:

$$\beta' \equiv \overleftarrow{\beta^\dagger} \mathrel{+\!\!+}^{\mathrm{H}} n \tag{84}$$

8. Authorization

We have previously discussed the model of work-packages and services in section 4.9, however we have yet to make a substantial discussion of exactly how some *core-time* resource may be apportioned to some work-package and its associated service. In the *YP* Ethereum model, the underlying resource, gas, is procured at the point of introduction on-chain and the purchaser is always the same agent who authors the data which describes the work to be done (i.e. the transaction). Conversely, in Polkadot the underlying resource, a parachain slot, is procured with a substantial deposit for typically 24 months at a time and the procurer, generally a parachain team, will often have no direct relation to the author of the work to be done (i.e. a parachain block).

On a principle of flexibility, we would wish JAM capable of supporting a range of interaction patterns both Ethereum-style and Polkadot-style. In an effort to do so, we introduce the *authorization system*, a means of disentangling the intention of usage for some coretime from the specification and submission of a particular workload to be executed on it. We are thus able to disassociate the purchase and assignment of coretime from the specific determination of work to be done with it, and so are able to support both Ethereum-style and Polkadot-style interaction patterns.

8.1. Authorizers and Authorizations.

The authorization system involves two key concepts: *authorizers* and *authorizations*. An authorization is simply a piece of opaque data to be included with a work-package. An authorizer meanwhile, is a piece of pre-parameterized logic which accepts as an additional parameter an authorization and, when executed within a VM of prespecified computational limits, provides a Boolean output denoting the veracity of said authorization.

Authorizations are identified as the hash of their logic (specified as the VM code) and their pre-parameterization. The process by which work-packages are determined to be authorized (or not) is not the competence of on-chain logic and happens entirely in-core and as such is discussed in section 14.3. However, on-chain logic must identify each set of authorizers assigned to each core in order to verify that a work-package is legitimately able to utilize that resource. It is this subsystem we will now define.

8.2. Pool and Queue.

We define the set of authorizers allowable for a particular core c as the *authorizer pool* $\alpha[c]$. To maintain this value, a further portion of state is tracked for each core: the core's current *authorizer queue* $\varphi[c]$, from which we draw values to fill the pool. Formally:

$$\alpha \quad \mathbb{H}_{\mathrm{O}\ \mathrm{C}}, \quad \varphi \quad \mathbb{H}_{\mathrm{Q}\ \mathrm{C}} \tag{85}$$

Note: The portion of state φ may be altered only through an exogenous call made from the accumulate logic of an appropriately privileged service.

The state transition of a block involves placing a new authorization into the pool from the queue:

$$c \quad \mathbb{N}_{\mathrm{C}} \quad \alpha'[c] \equiv \overleftarrow{F(c)} \mathrel{+\!\!+}^{\mathrm{O}} \varphi'[c][\mathbf{H}_t]^{\square} \tag{86}$$

$$F(c) \equiv \begin{cases} \alpha[c] \quad \{(g_w)_a\} & \text{if } g \quad \mathbf{E}_G \quad (g_w)_c = c \\ \alpha[c] & \text{otherwise} \end{cases} \tag{87}$$

Since α' is dependent on φ', practically speaking, this step must be computed after accumulation, the stage in which φ' is defined. Note that we utilize the guarantees extrinsic $\mathbf{E}_G$ to remove the oldest authorizer which has been used to justify a guaranteed work-package in the current block. This is further defined in equation 138.

9. Service Accounts

As we already noted, a service in JAM is somewhat analogous to a smart contract in Ethereum in that it includes amongst other items, a code component, a storage component and a balance. Unlike Ethereum, the code is split over two isolated entry-points each with their own environmental conditions; one, *refinement*, is essentially stateless and happens in-core, and the other, *accumulation*, which is stateful and happens on-chain. It is the latter which we will concern ourselves with now.

Service accounts are held in state under δ, a partial mapping from a service identifier $\mathbb{N}_S$ into a tuple of named elements which specify the attributes of the service relevant to the JAM protocol. Formally:

$$\mathbb{N}_S \equiv \mathbb{N}_{2^{32}} \tag{88}$$

$$\delta \quad \mathbb{D}\ \mathbb{N}_S \to \mathbb{A} \tag{89}$$

The service account is defined as the tuple of storage dictionary $\mathbf{s}$, preimage lookup dictionaries $\mathbf{p}$ and $\mathbf{l}$, code hash c, and balance b as well as the two code gas limits g & m. Formally:

$$\mathbb{A} \equiv \begin{array}{l} \mathbf{s} \quad \mathbb{D}\ \mathbb{H} \to \mathbb{Y}\ , \qquad \mathbf{p} \quad \mathbb{D}\ \mathbb{H} \to \mathbb{Y}\ , \\ \mathbf{l} \quad \mathbb{D} \quad \mathbb{H}, \mathbb{N}_L \ \to \ \mathbb{N}_T \ _3\ , \\ c \quad \mathbb{H}\ , \quad b \quad \mathbb{N}_B\ , \quad g \quad \mathbb{N}_G\ , \quad m \quad \mathbb{N}_G \end{array} \tag{90}$$

Thus, the balance of the service of index s would be denoted $\delta[s]_b$ and the storage item of key k $\mathbb{H}$ for that service is written $\delta[s]_{\mathbf{s}}[k]$.

9.1. Code and Gas.

The code c of a service account is represented by a hash which, if the service is to be functional, must be present within its preimage lookup (see section 9.2). We thus define the actual code $\mathbf{c}$:

$$\mathbf{a} \quad \mathbb{A} \quad \mathbf{a_c} \equiv \begin{cases} \mathbf{a_p}[\mathbf{a}_c] & \text{if } \mathbf{a}_c \quad \mathbf{a_p} \\ & \text{otherwise} \end{cases} \tag{91}$$

There are three entry-points in the code:

- **0** `}OT_jO:` Refinement, executed in-core and stateless.[10]
- **1** `9HH«h«e9§O:` Accumulation, executed on-chain and stateful.
- **2** `nj¬§}9j TO}:` Transfer handler, executed on-chain and stateful.

Whereas the first, executing in-core, is described in more detail in section 14.3, the latter two are defined in the present section.

As stated in appendix A, execution time in the JAM virtual machine is measured deterministically in units of *gas*, represented as a natural number less than 2^{64} and formally denoted $\mathbb{N}_G$. We may also use $\mathbf{Z}_G$ to denote the set $\mathbf{Z}_{-(2^{32}):::2^{32}}$ if the quantity may be negative. There are two limits specified in the account, g, the minimum gas required in order to execute the *Accumulate* entry-point of the service's code, and m, the minimum required for the *On Transfer* entry-point.

9.2. Preimage Lookups.

In addition to storing data in arbitrary key/value pairs available only on-chain, an account may also solicit data to be made available also in-core, and thus available to the Refine logic of the service's code. State concerning this facility is held under the service's $\mathbf{p}$ and $\mathbf{l}$ components.

There are several differences between preimage-lookups and storage. Firstly, preimage-lookups act as a mapping from a hash to its preimage, whereas general storage maps arbitrary keys to values. Secondly, preimage data is supplied extrinsically, whereas storage data originates as part of the service's accumulation. Thirdly preimage data, once supplied, may not be removed freely; instead it goes through a process of being marked as unavailable, and only after a period of time may it be removed from state. This ensures that historical information on its existence is retained. The final point especially is important since preimage data is designed to be queried in-core, under the Refine logic of the service's code, and thus it is important that the historical availability of the preimage is known.

We begin by reformulating the portion of state concerning our data-lookup system. The purpose of this system is to provide a means of storing static data on-chain such that it may later be made available within the execution of any service code as a function accepting only the hash of the data and its length in octets.

During the on-chain execution of the *Accumulate* function, this is trivial to achieve since there is inherently a state which all validators verifying the block necessarily have complete knowledge of, i.e. σ. However, for the in-core execution of *Refine*, there is no such state inherently available to all validators; we thus name a historical state, the *lookup anchor* which must be considered recently finalized before the work result may be accumulated hence providing this guarantee.

By retaining historical information on its availability, we become confident that any validator with a recently finalized view of the chain is able to determine whether any given preimage was available at any time within the period where auditing may occur. This ensures confidence that judgements will be deterministic even without consensus on chain state.

Restated, we must be able to define some *historical* lookup function which determines whether the preimage of some hash h was available for lookup by some service account $\mathbf{a}$ at some timeslot t, and if so, provide its preimage:

$$\{ \begin{array}{l} (\mathbb{A}, \mathbb{N}_{\mathbf{H}_t - C_D :::\mathbf{H}_t}, \mathbb{H}) \to \mathbb{Y}? \\ \qquad (\mathbf{a}, t, \mathrm{H}(\mathbf{p})) \quad v \ \ v \quad \{\mathbf{p}, \ \ \} \end{array} \tag{92}$$

This function is defined shortly below in equation 94.

The preimage lookup for some service of index s is denoted $\delta[s]_{\mathbf{p}}$ is a dictionary mapping a hash to its corresponding preimage. Additionally, there is metadata associated with the lookup denoted $\delta[s]_{\mathbf{l}}$ which is a dictionary mapping some hash and presupposed length into historical information.

9.2.1. *Invariants.* The state of the lookup system naturally satisfies a number of invariants. Firstly, any preimage value must correspond to its hash, equation 93. Secondly, a preimage value being in state implies that its hash and length pair has some associated status, also in

[10] Technically there is some small assumption of state, namely that some modestly recent instance of each service's preimages. The specifics of this are discussed in section 14.3.

equation 93. Formally:

$$(93)\quad a \quad \mathbb{A}, (h \quad \mathbf{p}) \quad a_{\mathbf{p}} \quad h = \mathrm{H}(\mathbf{p}) \quad h,\ \mathbf{p} \quad \mathrm{K}(a_{\mathbf{l}})$$

9.2.2. *Semantics.* The historical status component h $[\mathbb{N}_T]_3$ is a sequence of up to three time slots and the cardinality of this sequence implies one of four modes:

- $h = []$: The preimage is *requested*, but has not yet been supplied.
- $h \quad \mathbb{N}_{T\ 1}$: The preimage is *available* and has been from time h_0.
- $h \quad \mathbb{N}_{T\ 2}$: The previously available preimage is now *unavailable* since time h_1. It had been available from time h_0.
- $h \quad \mathbb{N}_{T\ 3}$: The preimage is *available* and has been from time h_2. It had previously been available from time h_0 until time h_1.

The historical lookup function may now be defined as:

$$(\mathbb{A}, \mathbb{N}_T, \mathbb{H}) \to \mathbb{Y}?$$

$$(\mathbf{a}, t, h) \equiv \begin{cases} \mathbf{a_p}[h] & \text{if } h \quad \mathrm{K}(\mathbf{a_p}) \quad I(\mathbf{a_l}[h,\ \mathbf{a_p}[h]\ ], t) \\ & \text{otherwise} \end{cases} \tag{94}$$

$$\text{where } I(\mathbf{l}, t) = \begin{cases} & \text{if } [] = \mathbf{l} \\ x \le t & \text{if } [x] = \mathbf{l} \\ x \le t < y & \text{if } [x, y] = \mathbf{l} \\ x \le t < y \quad z \le t & \text{if } [x, y, z] = \mathbf{l} \end{cases}$$

9.3. **Account Footprint and Threshold Balance.** We define the dependent values i and l as the storage footprint of the service, specifically the number of items in storage and the total number of octets used in storage. They are defined purely in terms of the storage map of a service, and it must be assumed that whenever a service's storage is changed, these change also.

Furthermore, as we will see in the account serialization function in section C, these are expected to be found explicitly within the Merklized state data. Because of this we make explicit their set.

We may then define a second dependent term t, the minimum, or *threshold*, balance needed for any given service account in terms of its storage footprint.

$$(95)\quad a \quad \mathrm{V}(\delta) \qquad \begin{aligned} a_i \quad \mathbb{N}_{2^{32}} \quad &\equiv 2 \quad a_{\mathbf{l}} + a_{\mathbf{s}} \\ a_l \quad \mathbb{N}_{2^{64}} \quad &\equiv \sum_{(h;z)\ \mathrm{K}(a_{\mathbf{l}})} 81 + z \\ &\quad + \sum_{x\ \mathrm{V}(a_{\mathbf{s}})} 32 + x \\ a_t \quad \mathbb{N}_B \quad &\equiv \mathrm{B}_S + \mathrm{B}_I \quad a_i + \mathrm{B}_L \quad a_l \end{aligned}$$

9.4. **Service Privileges.** Up to three services may be recognized as privileged. The portion of state in which this is held is denoted χ and has three components, each a service index. m is the index of the *manager* service, the service able to effect an alteration of χ from block to block. a and v are each the indices of services able to alter φ and ι from block to block. Formally:

$$(96)\qquad \chi \equiv \chi_m \quad \mathbb{N}_S, \chi_a \quad \mathbb{N}_S, \chi_v \quad \mathbb{N}_S$$

10. Disputes, Verdicts and Judgements

JAM provides a means of recording *judgements*: consequential votes amongst most of the validators over the validity of a *work-report* (a unit of work done within JAM, see section 11). Such collections of judgements are known as *verdicts*. JAM also provides a means of registering *offenses*, judgements and guarantees which dissent with an established *verdict*. Together these form the *disputes* system.

The registration of a verdict is not expected to happen very often in practice, however it is an important security backstop for removing and banning invalid work-reports from the processing pipeline as well as removing troublesome keys from the validator set where there is consensus over their malfunction. It also helps coordinate nodes to revert chain-extensions containing invalid work-reports and provides a convenient means of aggregating all offending validators for punishment in a higher-level system.

Judgement statements come about naturally as part of the auditing process and are expected to be positive, further affirming the guarantors' assertion that the work-report is valid. In the event of a negative judgement, then all validators audit said work-report and we assume a verdict will be reached. Auditing and guaranteeing are off-chain processes properly described in sections 14 and 17.

A judgement against a report implies that the chain is already reverted to some point prior to the accumulation of said report, usually forking at the block immediately prior to that at which accumulation happened. The specific strategy for chain selection is described fully in section 19. Authoring a block with a non-positive verdict has the effect of cancelling its imminent accumulation, as can be seen in equation 111.

Registering a verdict also has the effect of placing a permanent record of the event on-chain and allowing any offending keys to be placed on-chain both immediately or in forthcoming blocks, again for permanent record.

Having a persistent on-chain record of misbehavior is helpful in a number of ways. It provides a very simple means of recognizing the circumstances under which action against a validator must be taken by any higher-level validator-selection logic. Should JAM be used for a public network such as *Polkadot*, this would imply the slashing of the offending validator's stake on the staking parachain.

As mentioned, recording reports found to have a high confidence of invalidity is important to ensure that said reports are not allowed to be resubmitted. Conversely, recording reports found to be valid ensures that additional disputes cannot be raised in the future of the chain.

10.1. **The State.** The *disputes* state includes four items, three of which concern verdicts: a good-set ($\psi_{\mathbf{g}}$), a bad-set ($\psi_{\mathbf{b}}$) and a wonky-set ($\psi_{\mathbf{w}}$) containing the hashes of all work-reports which were respectively judged to be correct, incorrect or that it appears impossible to judge. The fourth item, the punish-set ($\psi_{\mathbf{o}}$), is a set of Ed25519 keys representing validators which were found to have misjudged a work-report.

$$(97)\qquad \psi \equiv \psi_{\mathbf{g}}, \psi_{\mathbf{b}}, \psi_{\mathbf{w}}, \psi_{\mathbf{o}}$$

10.2. **Extrinsic.** The disputes extrinsic, $\mathbf{E}_D$, may contain one or more verdicts $\mathbf{v}$ as a compilation of judgements coming from exactly two-thirds plus one of either the active validator set or the previous epoch's validator set, i.e. the Ed25519 keys of κ or λ. Additionally, it may contain proofs of the misbehavior of one or more validators, either by guaranteeing a work-report found to be invalid

(*culprits*, $\mathbf{c}$), or by signing a judgement found to be contradiction to a work-report's validity (*faults*, $\mathbf{f}$). Both are considered a kind of *offense*. Formally:

$$\mathbf{E}_D \equiv (\mathbf{v}, \mathbf{c}, \mathbf{f})$$

$$\text{(98)}\quad \text{where } \mathbf{v} \quad \mathbb{H}, \frac{\tau}{\mathrm{E}} \; -\mathbb{N}_2, \quad \{\ ,\ \}, \mathbb{N}_V, \mathbb{E}_{\ 2/3\mathrm{V}\ +1}$$
$$\text{and } \mathbf{c} \quad \mathbb{H}, \mathbb{H}_E, \mathbb{E}\ , \quad \mathbf{f} \quad \mathbb{H}, \{\ ,\ \}, \mathbb{H}_E, \mathbb{E}$$

The signatures of all judgements must be valid in terms of one of the two allowed validator key-sets, identified by the verdict's second term which must be either the epoch index of the prior state or one less. Formally:

$$(r, a, \mathbf{j}) \quad \mathbf{v}, \quad (v, i, s) \quad \mathbf{j} \quad s \quad \mathbb{E}_{\mathbf{k}[i]_e}\ \mathrm{X}_v \quad r$$

$$\text{(99)}\qquad \text{where } \mathbf{k} = \begin{cases} \kappa & \text{if } a = \dfrac{\tau}{\mathrm{E}} \\ \lambda & \text{otherwise} \end{cases}$$

$$\text{(100)}\qquad \mathrm{X}\ \equiv \texttt{Ma9h¬®9e\_L}\,,\ \mathrm{X}\ \equiv \texttt{Ma9h¬\_j®9e\_L}$$

Offender signatures must be similarly valid and reference work-reports with judgements and may not report keys which are already in the punish-set:

$$\text{(101)}\qquad (r, k, s) \quad \mathbf{c} \qquad \begin{array}{l} r \quad \psi'_{\mathbf{b}}\,, \\ k \quad (\lambda \quad \kappa) \quad \psi_{\mathbf{o}}\,, \\ s \quad \mathbb{E}_k\ \mathrm{X}_G \quad r \\ r \quad \psi'_{\mathbf{b}} \Leftrightarrow r \,/\, \psi'_{\mathbf{g}} \Leftrightarrow v\,, \end{array}$$

$$\text{(102)}\qquad (r, v, k, s) \quad \mathbf{f} \qquad \begin{array}{l} k \quad (\lambda \quad \kappa) \quad \psi_{\mathbf{o}}\,, \\ s \quad \mathbb{E}_k\ \mathrm{X}_v \quad r \end{array}$$

Verdicts $\mathbf{v}$ must be ordered by report hash. Offender signatures $\mathbf{c}$ and $\mathbf{f}$ must each be ordered by the validator's Ed25519 key. There may be no duplicate report hashes within the extrinsic, nor amongst any past reported hashes. Formally:

$$\text{(103)}\qquad \mathbf{v} = [r \quad r, a, \mathbf{j} \quad \mathbf{v}]$$

$$\text{(104)}\qquad \mathbf{c} = [k \quad r, k, s \quad \mathbf{c}]\,, \quad \mathbf{f} = [k \quad r, v, k, s \quad \mathbf{f}]$$

$$\text{(105)}\qquad \{r \quad r, a, \mathbf{j} \quad \mathbf{v}\} \quad \psi_{\mathbf{g}} \quad \psi_{\mathbf{b}} \quad \psi_{\mathbf{w}}$$

The judgements of all verdicts must be ordered by validator index and there may be no duplicates:

$$\text{(106)}\qquad (r, a, \mathbf{j}) \quad \mathbf{v} \quad \mathbf{j} = [i \quad v, i, s \quad \mathbf{j}]$$

We define $\mathbf{V}$ to derive from the sequence of verdicts introduced in the block's extrinsic, containing only the report hash and the sum of positive judgements. We require this total to be either exactly two-thirds-plus-one, zero or one-third of the validator set indicating, respectively, that the report is good, that it's bad, or that it's wonky.[11] Formally:

$$\text{(107)}\qquad \mathbf{V} \quad \mathbb{H}, \{0,\ 1/3\mathrm{V}\ ,\ 2/3\mathrm{V}\ +1\}$$

$$\text{(108)}\qquad \mathbf{V} = \quad r, \sum_{v;i;s}\ v_{\ \mathbf{j}} \qquad r, a, \mathbf{j} \ \prec \mathbf{v}$$

There are some constraints placed on the composition of this extrinsic: any verdict containing solely valid judgements implies the same report having at least one valid entry in the faults sequence $\mathbf{f}$. Any verdict containing solely invalid judgements implies the same report having at least two valid entries in the culprits sequence $\mathbf{c}$. Formally:

$$\text{(109)}\qquad (r,\ 2/3\mathrm{V}\ +1) \quad \mathbf{V} \quad (r, \dots) \quad \mathbf{f}$$

$$\text{(110)}\qquad (r, 0) \quad \mathbf{V} \quad \{(r, \dots) \quad \mathbf{c}\} \geq 2$$

We clear any work-reports which we judged as uncertain or invalid from their core:

$$\text{(111)}\qquad c \quad \mathbb{N}_C \quad \rho^\dagger[c] = \begin{cases} \text{if } \{(\rho[c]_r, t) \quad \mathbf{V}, t < \ 2/3\mathrm{V}\ \} \\ \rho_c \quad \text{otherwise} \end{cases}$$

The state's good-set, bad-set and wonky-set assimilate the hashes of the reports from each verdict. Finally, the punish-set accumulates the keys of any validators who have been found guilty of offending. Formally:

$$\text{(112)}\qquad \psi'_{\mathbf{g}} \equiv \psi_{\mathbf{g}} \quad \{r \quad r,\ 2/3\mathrm{V}\ +1 \quad \mathbf{V}\}$$

$$\text{(113)}\qquad \psi'_{\mathbf{b}} \equiv \psi_{\mathbf{b}} \quad \{r \quad r, 0 \quad \mathbf{V}\}$$

$$\text{(114)}\qquad \psi'_{\mathbf{w}} \equiv \psi_{\mathbf{w}} \quad \{r \quad r,\ 1/3\mathrm{V} \quad \mathbf{V}\}$$

$$\text{(115)}\qquad \psi'_{\mathbf{o}} \equiv \psi_{\mathbf{o}} \quad \{k \quad (r, k, s) \quad \mathbf{c}\} \quad \{k \quad (r, v, k, s) \quad \mathbf{f}\}$$

10.3. **Header.** The verdicts and offenders markers must contain exactly the sequence of report hashes of all new bad & wonky verdicts and keys of all new offenders, respectively. Formally:

$$\text{(116)}\qquad \mathbf{H}_j \equiv [r \quad r, t \ \prec \mathbf{V}, t \neq \ 2/3\mathrm{V}\ +1]$$

$$\text{(117)}\qquad \mathbf{H}_o \equiv [k \quad (r, k, s) \quad \mathbf{c}] \quad [k \quad (r, v, k, s) \quad \mathbf{f}]$$

11. Reporting and Assurance

Reporting and assurance are the two on-chain processes we do to allow the results of in-core computation to make its way into the service state singleton, δ. A *work-package*, which comprises several *work items*, is transformed by validators acting as *guarantors* into its corresponding *work-report*, which similarly comprises several *work outputs* and then presented on-chain within the *guarantees* extrinsic. At this point, the work-package is erasure coded into a multitude of segments and each segment distributed to the associated validator who then attests to its availability through an *assurance* placed on-chain. After enough assurances the work-report is considered *available*, and the work outputs transform the state of their associated service by virtue of accumulation, covered in section 12. The report may also be *timed-out*, implying it may be replaced by another report without accumulation.

From the perspective of the work-report, therefore, the guarantee happens first and the assurance afterwards. However, from the perspective of a block's state-transition, the assurances are best processed first since each core may only have a single work-report pending its package becoming available at a time. Thus, we will first cover the transition arising from processing the availability assurances followed by the work-report guarantees. This synchroneity can be seen formally through the requirement of an intermediate state $\rho^\ddagger$, utilized later in equation 144.

[11] This requirement may seem somewhat arbitrary, but these happen to be the decision thresholds for our three possible actions and are acceptable since the security assumptions include the requirement that at least two-thirds-plus-one validators are live (Jeff Burdges, Cevallos, et al. 2024 discusses the security implications in depth).

11.1. **State.** The state of the reporting and availability portion of the protocol is largely contained within ρ, which tracks the work-reports which have been reported but not yet accumulated and the identities of the guarantors who reported them and the time at which it was reported. As mentioned earlier, only one report may be assigned to a core at any given time. Formally:

(118) $\rho \quad w \quad \mathbb{W}, t \quad \mathbb{N}_T \quad ?_{\mathbb{C}}$

As usual, intermediate and posterior values ($\rho^\dagger$, $\rho^\ddagger$, ρ') are held under the same constraints as the prior value.

11.1.1. *Work Report.* A work-report, of the set $\mathbb{W}$, is defined as a tuple of the work-package specification s, the refinement context x, and the core-index (i.e. on which the work is done) as well as the authorizer hash a and output $\mathtt{o}$ and finally the results of the evaluation of each of the items in the package $\mathtt{r}$, which is always at least one item and may be no more than I items. Formally:

(119) $\mathbb{W} \equiv s \quad \mathbb{S}, x \quad \mathbb{X}, c \quad \mathbb{N}_{\mathbb{C}}, a \quad \mathbb{H}, \mathtt{o} \quad \mathbb{Y}, \mathtt{r} \quad \mathbb{L}_{1\,\mathrm{I}}$

The total serialized size of a work-report may be no greater than $\mathbf{W}_R$ bytes:

(120) $w \quad \mathbb{W} \quad \mathrm{E}(w) \le \mathbf{W}_R$

11.1.2. *Refinement Context.* A *refinement context*, denoted by the set $\mathbb{X}$, describes the context of the chain at the point that the report's corresponding work-package was evaluated. It identifies two historical blocks, the *anchor*, header hash a along with its associated posterior state-root s and posterior Beefy root b; and the *lookup-anchor*, header hash l and of timeslot t. Finally, it identifies the hash of an optional prerequisite work-package p. Formally:

(121) $\mathbb{X} \equiv \begin{matrix} a \quad \mathbb{H}, & s \quad \mathbb{H}, & b \quad \mathbb{H}, \\ l \quad \mathbb{H}, & t \quad \mathbb{N}_T, & p \quad \mathbb{H}? \end{matrix}$

11.1.3. *Availability.* We define the set of *availability specifications*, $\mathbb{S}$, as the tuple of the work-package's hash h, an auditable work bundle length l (see section 14.4.1 for more clarity on what this is), together with an erasure-root u and a segment-root e. Work-results include this availability specification in order to ensure they are able to correctly reconstruct and audit the purported ramifications of any reported work-package. Formally:

(122) $\mathbb{S} \equiv h \quad \mathbb{H}, l \quad \mathbb{N}_L, u \quad \mathbb{H}, e \quad \mathbb{H}$

The *erasure-root* (u) is the root of a binary Merkle tree which functions as a commitment to all data required for the auditing of the report and for use by later work-packages should they need to retrieve any data yielded. It is thus used by assurers to verify the correctness of data they have been sent by guarantors, and it is later verified as correct by auditors. It is discussed fully in section 14.

The *segment-root* (e) is the root of a constant-depth, left-biased and zero-hash-padded binary Merkle tree committing to the hashes of each of the exported segments of each work-item. These are used by guarantors to verify the correctness of any reconstructed segments they are called upon to import for evaluation of some later work-package. It is also discussed in section 14.

11.1.4. *Work Result.* We finally come to define a *work result*, $\mathbb{L}$, which is the data conduit by which services' states may be altered through the computation done within a work-package.

(123) $\mathbb{L} \equiv (s \quad \mathbb{N}_S, c \quad \mathbb{H}, l \quad \mathbb{H}, g \quad \mathbb{N}_G, o \quad \mathbb{Y} \quad \mathbb{J})$

Work results are a tuple comprising several items. Firstly s, the index of the service whose state is to be altered and thus whose refine code was already executed. We include the hash of the code of the service at the time of being reported c, which must be accurately predicted within the work-report according to equation 153;

Next, the hash of the payload (l) within the work item which was executed in the refine stage to give this result. This has no immediate relevance, but is something provided to the accumulation logic of the service. We follow with the gas prioritization ratio g used when determining how much gas should be allocated to execute of this item's accumulate.

Finally, there is the output or error of the execution of the code o, which may be either an octet sequence in case it was successful, or a member of the set $\mathbb{J}$, if not. This latter set is defined as the set of possible errors, formally:

(124) $\mathbb{J} \quad \{\infty, \ , \mathtt{?8K}, \mathtt{?\text{\^{}}W}\}$

The first two are special values concerning execution of the virtual machine, ∞ denoting an out-of-gas error and denoting an unexpected program termination. Of the remaining two, the first indicates that the service's code was not available for lookup in state at the posterior state of the lookup-anchor block. The second indicates that the code was available but was beyond the maximum size allowed S.

11.2. **Package Availability Assurances.** We first define $\rho^\ddagger$, the intermediate state to be utilized next in section 11.4 as well as $\mathbf{W}$, the set of available work-reports, which will we utilize later in section 12. Both require the integration of information from the assurances extrinsic $\mathbf{E}_A$.

11.2.1. *The Assurances Extrinsic.* The assurances extrinsic is a sequence of *assurance* values, at most one per validator. Each assurance is a sequence of binary values (i.e. a bitstring), one per core, together with a signature and the index of the validator who is assuring. A value of 1 (or , if interpreted as a Boolean) at any given index implies that the validator assures they are contributing to its availability.[12] Formally:

(125) $\mathbf{E}_A \quad a \quad \mathbb{H}, f \quad \mathbb{B}_{\mathbb{C}}, v \quad \mathbb{N}_{\mathbb{V}}, s \quad \mathbb{E}_{\mathbb{V}}$

The assurances must all be anchored on the parent and ordered by validator index:

(126) $a \quad \mathbf{E}_A \quad a_a = \mathbf{H}_p$

(127) $i \quad \{1 \ldots \mathbf{E}_A\} \quad \mathbf{E}_A[i-1]_v < \mathbf{E}_A[i]_v$

The signature must be one whose public key is that of the validator assuring and whose message is the serialization of the parent hash $\mathbf{H}_p$ and the aforementioned bitstring:

(128) $a \quad \mathbf{E}_A \quad a_s \quad \mathbb{E}_{'[a_v]_e} \ \mathrm{X}_A \quad \mathrm{H}(\mathbf{H}_p, a_f)$

(129) $\mathrm{X}_A \equiv \mathtt{Ma9h\neg9@9\_e9@eO}$

[12] This is a "soft" implication since there is no consequence on-chain if dishonestly reported. For more information on this implication see section 16.

A bit may only be set if the corresponding core has a report pending availability on it:

$$(130) \qquad a \quad \mathbf{E}_A \quad c \quad \mathbb{N}_C \quad a_v[c] \qquad \rho^\dagger[c] \neq$$

11.2.2. *Available Reports.* A work-report is said to become *available* if and only if there are a clear 2/3 supermajority of validators who have marked its core as set within the block's assurance extrinsic. Formally, we define the series of available work-reports $\mathbf{W}$ as:

$$(131) \qquad \mathbf{W} \equiv \quad \rho^\dagger[c]_w \quad c \in \mathbb{N}_C, \quad \sum_{a \ \mathbf{E}_A} a_v[c] > 2/3\,\mathrm{V}$$

This value is utilized in the definition of both δ' and $\rho^\ddagger$ which we will define presently as equivalent to $\rho^\dagger$ except for the removal of items which are now available:

$$(132) \qquad c \quad \mathbb{N}_C \quad \rho^\ddagger[c] \equiv \begin{cases} \text{if } \rho[c]_w \quad \mathbf{W} \\ \rho^\dagger[c] \quad \text{otherwise} \end{cases}$$

11.3. **Guarantor Assignments.** Every block, each core has three validators uniquely assigned to guarantee work-reports for it. This is borne out with $\mathrm{V} = 1,023$ validators and $\mathrm{C} = 341$ cores, since $\mathrm{V}/\mathrm{C} = 3$. The core index assigned to each of the validators, as well as the validators' Ed25519 keys are denoted by $\mathbf{G}$:

$$(133) \qquad \mathbf{G} \quad (\ \mathbb{N}_{C\ N_V}, \ \mathbb{H}_{K\ N_V}\ )$$

We determine the core to which any given validator is assigned through a shuffle using epochal entropy and a periodic rotation to help guard the security and liveness of the network. We use η_2 for the epochal entropy rather than η_1 to avoid the possibility of fork-magnification where uncertainty about chain state at the end of an epoch could give rise to two established forks before it naturally resolves.

We define the permute function P, the rotation function R and finally the guarantor assignments $\mathbf{G}$ as follows:

$$(134) \quad R(\mathbf{c}, n) \equiv [(x + n) \bmod \mathrm{C} \quad x \in \mathbf{c}]$$

$$(135) \quad P(e,t) \equiv R(\mathrm{F}([\ \frac{\mathrm{C}\ i}{\mathrm{V}} \quad i \in \mathbb{N}_\mathrm{V}], e),\ \frac{t \bmod \mathrm{E}}{\mathrm{R}}\ )$$

$$(136) \qquad \mathbf{G} \equiv (P(\eta_2', \tau'),\ (\kappa'))$$

We also define $\mathbf{G}^*$, which is equivalent to the value $\mathbf{G}$ as it would have been under the previous rotation:

$$(137) \qquad \text{let } (e, \mathbf{k}) = \begin{cases} (\eta_2', \kappa') & \text{if } \frac{\tau' - \mathrm{R}}{\mathrm{E}} = \frac{\tau'}{\mathrm{E}} \\ (\eta_3', \lambda') & \text{otherwise} \end{cases}$$
$$\mathbf{G}^* \equiv (P(e, \tau' - \mathrm{R}),\ (\mathbf{k}))$$

11.4. **Work Report Guarantees.** We begin by defining the guarantees extrinsic, $\mathbf{E}_G$, a series of *guarantees*, at most one for each core, each of which is a tuple of a core index, *work-report*, a credential a and its corresponding timeslot t. The core index of each guarantee must be unique and guarantees must be in ascending order of this. Formally:

$$(138) \qquad \mathbf{E}_G \quad w \quad \mathbb{W}, t \quad \mathbb{N}_T, a \quad \mathbb{N}_V, \mathbb{E}_{\ 2\,3\ \ \mathrm{C}}$$

$$(139) \qquad \mathbf{E}_G = [(g_w)_c \quad g \quad \mathbf{E}_G]$$

The credential is a sequence of two or three tuples of a unique validator index and a signature. Credentials must be ordered by their validator index:

$$(140) \qquad g \quad \mathbf{E}_G \quad g_a = [v \quad v, s \quad g_a]$$

The signature must be one whose public key is that of the validator identified in the credential, and whose message is the serialization of the hash of the work-report. The signing validators must be assigned to the core in question in either this block $\mathbf{G}$ if the timeslot for the guarantee is in the same rotation as this block's timeslot, or in the most recent previous set of assignments, $\mathbf{G}^*$:

$$(141) \quad \begin{array}{l} (w,t,a) \quad \mathbf{E}_G, \quad s \quad \mathbb{E}_{(\mathbf{k}_v)_E}\ \mathrm{X}_G \quad \mathrm{H}(\mathrm{E}(w)) \\ (v,s) \quad a \qquad \mathbf{c}_v = w_c \quad \mathrm{R}(\ \lfloor \tau'/\mathrm{R} \rfloor - 1) \le t \le \tau' \end{array}$$
$$k \quad \mathbf{R} \Leftrightarrow \quad (w,t,a) \quad \mathbf{E}_G, \ (v,s) \quad a \quad k = (\mathbf{k}_v)_E$$
$$\text{where } (\mathbf{c}, \mathbf{k}) = \begin{cases} \mathbf{G} & \text{if } \frac{\tau'}{\mathrm{R}} = \frac{t}{\mathrm{R}} \\ \mathbf{G}^* & \text{otherwise} \end{cases}$$

$$(142) \quad \mathrm{X}_G \equiv \texttt{Ma9h¬X«9}\}\texttt{9j§OO}$$

We note that the Ed25519 key of each validator whose signature is in a credential is placed in the *reporters* set $\mathbf{R}$. This is utilized by the validator activity statistics bookkeeping system section 13.

We denote $\mathbf{w}$ to be the set of work-reports in the present extrinsic $\mathbf{E}$:

$$(143) \qquad \text{let } \mathbf{w} = \{g_w \quad g \quad \mathbf{E}_G\}$$

No reports may be placed on cores with a report pending availability on it unless it has timed out. In the latter case, $\mathrm{U} = 5$ slots must have elapsed after the report was made. A report is valid only if the authorizer hash is present in the authorizer pool of the core on which the work is reported. Formally:

$$(144) \qquad w \quad \mathbf{w} \quad \begin{array}{l} \rho^\ddagger[w_c] = \\ w_a \quad \alpha[w_c] \end{array} \quad \mathrm{H}_t \ge \rho^\ddagger[w_c]_t + \mathrm{U}\ ,$$

We specify the maximum total accumulation gas requirement a work-report may imply as G_A, and we require the sum of all services' minimum gas requirements to be no greater than this:

$$(145) \qquad w \quad \mathbf{w} \quad \sum_{s \ (w_r)_s} \delta[s]_g \ \le \ \mathrm{G}_A$$

11.4.1. *Contextual Validity of Reports.* For convenience, we define two equivalences $\mathbf{x}$ and $\mathbf{p}$ to be, respectively, the set of all contexts and work-package hashes within the extrinsic:

$$(146) \qquad \text{let } \mathbf{x} \equiv \{w_x \quad w \quad \mathbf{w}\}\ , \quad \mathbf{p} \equiv \{(w_s)_h \quad w \quad \mathbf{w}\}$$

There must be no duplicate work-package hashes (i.e. two work-reports of the same package). Therefore, we require the cardinality of $\mathbf{p}$ to be the length of the work-report sequence $\mathbf{w}$:

$$(147) \qquad \mathbf{p} = \mathbf{w}$$

We require that the anchor block be within the last H blocks and that its details be correct by ensuring that it appears within our most recent blocks β:

$$(148) \quad x \quad \mathbf{x} \quad y \quad \beta \quad x_a = y_h \quad x_s = y_s \quad x_b = \mathrm{H}_K(\mathrm{E}_M(y_\mathbf{b}))$$

We require that each lookup-anchor block be within the last L timeslots:

$$(149) \qquad x \quad \mathbf{x} \quad x_t \ge \mathrm{H}_t - \mathrm{L}$$

We also require that we have a record of it; this is one of the few conditions which cannot be checked purely with

on-chain state and must be checked by virtue of retaining the series of the last L headers as the ancestor set $\mathbf{A}$. Since it is determined through the header chain, it is still deterministic and calculable. Formally:

$$x \quad \mathbf{x} \quad h \quad \mathbf{A} \quad h_t = x_t \quad \mathrm{H}(h) = x_h \tag{150}$$

We require that the work-package of the report not be the work-package of some other report made in the past. Since the work-package implies the anchor block, and the anchor block is limited to the most recent blocks, we need only ensure that the work-package not appear in our recent history:

$$p \quad \mathbf{p}, \quad x \quad \beta \quad p \,/\, x_{\mathbf{p}} \tag{151}$$

We require that the prerequisite work-package, if present, be either in the extrinsic or in our recent history:

$$\begin{array}{ll} w & \mathbf{w}, (w_x)_p \neq \\ (w_x)_p & \mathbf{p} \quad \{x \quad x \quad b_{\mathbf{p}}, b \quad \beta\} \end{array} \tag{152}$$

We require that all work results within the extrinsic predicted the correct code hash for their corresponding service:

$$w \quad \mathbf{w}, \quad r \quad w_r \quad r_c = \delta[r_s]_c \tag{153}$$

11.5. **Transitioning for Reports.** We define ρ' as being equivalent to $\rho^\ddagger$, except where the extrinsic replaced an entry. In the case an entry is replaced, the new value includes the present time τ' allowing for the value may be replaced without respect to its availability once sufficient time has elapsed (see equation 144).

$$c \quad \mathbb{N}_\mathrm{C} \quad \rho'[c] \equiv \begin{cases} w, t \ \tau' & \text{if} \quad c, w, a \quad \mathbf{E}_G \\ \rho^\ddagger[c] & \text{otherwise} \end{cases} \tag{154}$$

This concludes the section on reporting and assurance. We now have a complete definition of ρ' together with $\mathbf{W}$ to be utilized in section 12, describing the portion of the state transition happening once a work-report is guaranteed and made available.

12. Accumulation

Accumulation may be defined as some function whose arguments are $\mathbf{W}$ and δ together with selected portions of (at times partially transitioned) state and which yields the posterior service state δ' together with additional state elements ι', φ' and χ'.

The proposition of accumulation is in fact quite simple: we merely wish to execute the *Accumulate* logic of the service code of each of the services which has at least one work output, passing to it the work outputs and useful contextual information. However, there are three main complications. Firstly, we must define the execution environment of this logic and in particular the host functions available to it. Secondly, we must define the amount of gas to be allowed for each service's execution. Finally, we must determine the nature of transfers within Accumulate which, as we will see, leads to the need for a second entry-point, *on-transfer*.

12.1. **Preimage Integration.** Prior to accumulation, we must first integrate all preimages provided in the lookup extrinsic. The lookup extrinsic is a sequence of pairs of service indices and data. These pairs must be ordered and without duplicates (equation 156 requires this). The data must have been solicited by a service but not yet be provided. Formally:

$$\mathbf{E}_P \quad \mathbb{N}_S, \mathbb{Y} \tag{155}$$

$$\mathbf{E}_P = [i \quad i \quad \mathbf{E}_P] \tag{156}$$

$$s, \mathrm{p} \quad \mathbf{E}_P \quad \{ \begin{array}{l} \mathrm{K}(\delta[s]_{\mathbf{p}}) \ / \ \mathrm{H}(\mathrm{p}) \ , \\ \delta[s]_{\mathbf{l}}[\ \mathrm{H}(\mathrm{p}), \ \mathrm{p} \ ] = [] \end{array} \tag{157}$$

We define $\delta^\dagger$ as the state after the integration of the preimages:

$$\delta^\dagger = \delta \ \text{ex.} \quad s, \mathrm{p} \quad \mathbf{E}_P \quad \begin{array}{l} \delta^\dagger[s]_{\mathbf{p}}[\mathrm{H}(\mathrm{p})] = \mathrm{p} \\ \delta^\dagger[s]_{\mathbf{l}}[\mathrm{H}(\mathrm{p}), \ \mathrm{p} \ ] = [\tau'] \end{array} \tag{158}$$

12.2. **Gas Accounting.** We define $\mathbf{S}$, the set of all services which will be accumulated in this block; this is all services which have at least one work output within $\mathbf{W}$, together with all privileged services, χ. Formally:

$$\mathbf{S} \equiv \{\mathrm{r}_s \quad w \quad \mathbf{W}, \mathrm{r} \quad w_{\mathbf{r}}\} \quad \{\chi_m, \chi_a, \chi_v\} \tag{159}$$

We calculate the gas attributable for each service as the sum of each of the service's work outputs' share of their report's elective accumulation gas together with the subtotal of minimum gas requirements:

$$\mathbb{N}_S \to \mathbb{N}_G$$

$$G \quad s \quad \sum_{w \ \mathbf{W}\ \mathrm{r}} \ \sum_{w_{\mathbf{r}}; \mathbf{r}_s = s} \delta^\dagger[s]_g + \ \mathrm{r}_g \quad \frac{\mathrm{G}_A - \sum_{\mathbf{r}\ w_{\mathbf{r}}} \delta^\dagger[\mathrm{r}_s]_g}{\sum_{\mathbf{r}\ w_{\mathbf{r}}} \mathrm{r}_g} \tag{160}$$

12.3. **Wrangling.** We finally define the results which will be given as an operand into the accumulate function for each service in $\mathbf{S}$. This is a sequence of operand tuples $\mathbb{O}$, one sequence for each service in $\mathbf{S}$. Each sequence contains one element per work output (or error) to be accumulated for that service, together with said work output's payload hash, package hash and authorization output. The tuples are sequenced in the same order as they appear in $\mathbf{W}$. Formally:

$$\mathbb{O} \equiv o \quad \mathbb{Y} \quad \mathbb{J}, l \quad \mathbb{H}, k \quad \mathbb{H}, a \quad \mathbb{Y} \tag{161}$$

$$\mathbb{N}_S \to \mathbb{O}$$

$$M \quad s \quad \begin{array}{ll} o \ \mathrm{r}_o, \ l \ \mathrm{r}_p, & w \quad \mathbf{W}, \\ a \ w_o, k \ (w_s)_h & \mathrm{r} \quad w_{\mathbf{r}}, \\ & \mathrm{r}_s = s \end{array} \tag{162}$$

12.4. **Invocation.** Within this section, we define A, the function which conducts the accumulation of a single service. Formally speaking, A assumes omnipresence of timeslot $\mathbf{H}_t$ and some prior state components $\delta^\dagger$, ν, $\mathbf{W}_{\mathbf{d}}$, and takes as specific arguments the service index $s \quad \mathbf{S}$ (from which it may derive the wrangled results $M(s)$ and gas limit $G(s)$) and yields values for $\delta^\dagger[s]$ and staging assignments into φ, ι together with a series of lookup solicitations/forgets, a series of deferred transfers and $\mathbf{C}$ mapping from service index to Beefy commitment hashes.

We first denote the set of deferred transfers as $\mathbb{T}$, noting that a transfer includes a memo component m of 64 octets, together with the service index of the sender s,

the service index of the receiver d, the amount of tokens to be transferred a and the gas limit g for the transfer. Formally:

$$\mathbb{T} \equiv s \in \mathbb{N}_S, d \in \mathbb{N}_S, a \in \mathbb{N}_B, m \in \mathbb{Y}_M, g \in \mathbb{N}_G \tag{163}$$

We may then define A, the mapping from the index of accumulated services to the various components in terms of which we will be imminently defining our posterior state:

$$A \in \begin{matrix} \mathbb{N}_S \to \{ & \mathbf{s} \in \mathbb{A}?, & \mathbf{v} \in \mathbb{K}_{\mathrm{V}}, & \mathbf{t} \in \mathbb{T}, & r \in \mathbb{H}?, \\ & \mathbf{c} \in \mathbb{H}_{\mathrm{Q}\,\mathrm{C}}, & & \mathbf{n} \in \mathbb{D}\,\mathbb{N}_S \to \mathbb{A}, & \\ & p \in \mathbb{N}_S, & m \in \mathbb{N}_S, & a \in \mathbb{N}_S, & v \in \mathbb{N}_S \end{matrix} \}$$
$$s \in \Psi_A(\delta^\dagger, s, M(s), G(s)) \tag{164}$$

As can be seen plainly, our accumulation mapping A combines portions of the prior state into arguments for a virtual-machine invocation. Specifically the service accounts $\delta^\dagger$ together with the index of the service in question s and its wrangled refine-results $M(s)$ and gas limit $G(s)$ are arranged to create the arguments for Ψ_A, itself using a virtual-machine invocation as defined in appendix B.4.

The Beefy commitment map is a function mapping all accumulated services to their accumulation result (the r component of the result of A). This is utilized in determining the accumulation-result tree root for the present block, useful for the Beefy protocol:

$$\mathbf{C} \equiv \{(s, A(s)_r) \mid s \in \mathbf{S}, A(s)_r \neq \emptyset\} \tag{165}$$

Given our mapping A, which may be calculated exhaustively from the vm invocations of each accumulated service $\mathbf{S}$, we may define the posterior state δ', χ', φ' and ι' as the result of integrating A into our state.

12.4.1. *Privileged Transitions.* The staging core assignments, and validator keys and privileged service set are each altered based on the effects of the accumulation of each of the three privileged services:

$$\chi' \equiv A(\chi_m)_p\,, \quad \varphi' \equiv A(\chi_a)_{\mathbf{c}}\,, \quad \iota' \equiv A(\chi_v)_{\mathbf{v}} \tag{166}$$

12.4.2. *Service Account Transitions.* Finally, we integrate all changes to the service accounts into state.

We note that all newly added service indices, defined as $\mathrm{K}(A(s)_{\mathbf{n}})$ for any accumulated service s, must not conflict with the indices of existing services or newly added services. This should never happen, since new indices are explicitly selected to avoid such conflicts, but in the unlikely event it happens, the block would be invalid. Formally:

$$\begin{aligned} &\forall s \in \mathbf{S} : \mathrm{K}(A(s)_{\mathbf{n}}) \cap \mathrm{K}(\delta^\dagger) = \emptyset\,, \\ &\forall t \in \mathbf{S} \setminus \{s\} : \mathrm{K}(A(s)_{\mathbf{n}}) \cap \mathrm{K}(A(t)_{\mathbf{n}}) = \emptyset \end{aligned} \tag{167}$$

We first define $\delta^\ddagger$, an intermediate state after main accumulation but before the transfers have been credited and handled:

$$\mathrm{K}(\delta^\ddagger) \equiv (\mathrm{K}(\delta^\dagger) \setminus \bigcup_{s \in \mathbf{S}} \mathrm{K}(A(s)_{\mathbf{n}})) \setminus \{s \mid s \in \mathbf{S}, s_{\mathbf{s}} = \emptyset\}$$
$$\delta^\ddagger[s] \equiv \begin{cases} A(s)_{\mathbf{s}} & \text{if } s \in \mathbf{S} \\ A(t)_{\mathbf{n}}[s] & \text{if } \neg\exists t \in \mathbf{S} : s \in \mathrm{K}(A(t)_{\mathbf{n}}) \\ \delta^\dagger[s] & \text{otherwise} \end{cases} \tag{168}$$

We denote $R(s)$ the sequence of transfers received by a given service of index s, in order of them being sent from services of ascending index. (If some service s received no transfers or simply does not exist then $R(s)$ would be validly defined as the empty sequence.) Formally:

$$R \in \{\mathbb{N}_S \to \mathbb{T} \mid d \in [t \in \mathbf{S} \mid s \in \mathbf{S}, t \in A(s)_{\mathbf{t}}, t_d = d]\} \tag{169}$$

The posterior state δ' may then be defined as the intermediate state with all the deferred effects of the transfers applied:

$$\delta' = \{s \in \mathbf{S} \mid \Psi_T(\delta^\ddagger, a, R(a)) \in (s \in \mathbf{S} \mid a \in \mathbf{a})\} \tag{170}$$

Note that Ψ_T is defined in appendix B.5 such that it results in $\delta^\ddagger[d]$, i.e. no difference to the account's intermediate state, if $R(d) = []$, i.e. said account received no transfers.

13. Validator Activity Statistics

The Jam chain does not explicitly issue rewards—we leave this as a job to be done by the staking subsystem (in Polkadot's case envisioned as a system parachain—hosted without fees—in the current imagining of a public Jam network). However, much as with validator punishment information, it is important for the Jam chain to facilitate the arrival of information on validator activity in to the staking subsystem so that it may be acted upon.

Such performance information cannot directly cover all aspects of validator activity; whereas block production, guarantor reports and availability assurance can easily be tracked on-chain, Grandpa, Beefy and auditing activity cannot. In the latter case, this is instead tracked with validator voting activity: validators vote on their impression of each other's efforts and a median may be accepted as the truth for any given validator. With an assumption of 50% honest validators, this gives an adequate means of oraclizing this information.

The validator statistics are made on a per-epoch basis and we retain one record of completed statistics together with one record which serves as an accumulator for the present epoch. Both are tracked in π, which is thus a sequence of two elements, with the first being the accumulator and the second the previous epoch's statistics. For each epoch we track a performance record for each validator:

$$\pi \in \mathbf{b} \in \mathbb{N}, \mathbf{t} \in \mathbb{N}, \mathbf{p} \in \mathbb{N}, \mathbf{d} \in \mathbb{N}, \mathbf{g} \in \mathbb{N}, \mathbf{a} \in \mathbb{N}_{\mathrm{V}}^2 \tag{171}$$

The six statistics we track are:

- b: The number of blocks produced by the validator.
- t: The number of tickets introduced by the validator.
- p: The number of preimages introduced by the validator.
- d: The total number of octets across all preimages introduced by the validator.
- g: The number of reports guaranteed by the validator.
- a: The number of availability assurances made by the validator.

The objective statistics are updated in line with their description, formally:

$$\text{let } e = \frac{\tau}{\mathrm{E}}\,, \quad e' = \frac{\tau'}{\mathrm{E}} \tag{172}$$

$$(\mathbf{a}, \pi'_1) \equiv \begin{cases} (\pi_0, \pi_1) & \text{if } e' = e \\ ([0, \dots, [0, \dots], \dots], \pi_0) & \text{otherwise} \end{cases} \tag{173}$$

$$
\begin{aligned}
\pi'_0[v]_b &\equiv \mathbf{a}[v]_b + (v = \mathbf{H}_i) \\
\pi'_0[v]_t &\equiv \mathbf{a}[v]_t + \begin{cases} \mathbf{E}_T & \text{if } v = \mathbf{H}_i \\ 0 & \text{otherwise} \end{cases} \\
\pi'_0[v]_p &\equiv \mathbf{a}[v]_p + \begin{cases} \mathbf{E}_P & \text{if } v = \mathbf{H}_i \\ 0 & \text{otherwise} \end{cases} \\
\pi'_0[v]_d &\equiv \mathbf{a}[v]_d + \begin{cases} \sum_{d} \mathbf{E}_P\ d & \text{if } v = \mathbf{H}_i \\ 0 & \text{otherwise} \end{cases} \\
\pi'_0[v]_g &\equiv \mathbf{a}[v]_g + (\kappa'_v \quad \mathbf{R}) \\
\pi'_0[v]_a &\equiv \mathbf{a}[v]_a + (\ a \quad \mathbf{E}_A \quad a_v = v)
\end{aligned}
\qquad (174) \quad v \quad \mathbb{N}_V
$$

Note that $\mathbf{R}$ is the *Reporters* set, as defined in equation 141.

14. Work Packages and Work Reports

14.1. **Honest Behavior.** We have so far specified how to recognize blocks for a correctly transitioning JAM blockchain. Through defining the state transition function and a state Merklization function, we have also defined how to recognize a valid header. While it is not especially difficult to understand how a new block may be authored for any node which controls a key which would allow the creation of the two signatures in the header, nor indeed to fill in the other header fields, readers will note that the contents of the extrinsic remain unclear.

We define not only correct behavior through the creation of correct blocks but also *honest behavior*, which involves the node taking part in several *off-chain* activities. This does have analogous aspects within *YP* Ethereum, though it is not mentioned so explicitly in said document: the creation of blocks along with the gossiping and inclusion of transactions within those blocks would all count as off-chain activities for which honest behavior is helpful. In JAM's case, honest behavior is well-defined and expected of at least $2/3$ of validators.

Beyond the production of blocks, incentivized honest behavior includes:

- the guaranteeing and reporting of work-packages, along with chunking and distribution of both the chunks and the work-package itself, discussed in section 15;
- assuring the availability of work-packages after being in receipt of their data;
- determining which work-reports to audit, fetching and auditing them, and creating and distributing judgements appropriately based on the outcome of the audit;
- submitting the correct amount of auditing work seen being done by other validators, discussed in section 13.

14.2. **Segments and the Manifest.** Our basic erasure-coding segment size is $\mathsf{W}_C = 684$ octets, derived from the fact we wish to be able to reconstruct even should almost two-thirds of our 1023 participants be malicious or incapacitated, the 16-bit Galois field on which the erasure-code is based and the desire to efficiently support encoding data of close to, but no less than, 4KB.

Work-packages are generally small to ensure guarantors need not invest a lot of bandwidth in order to discover whether they can get paid for their evaluation into a work-report. Rather than having much data inline, they instead *reference* data through commitments. The simplest commitments are extrinsic data.

Extrinsic data are blobs which are being introduced into the system alongside the work-package itself generally by the work-package builder. They are exposed to the Refine logic as an argument. We commit to them through including each of their hashes in the work-package.

Work-packages have two other types of external data associated with them: A cryptographic commitment to each *imported* segment and finally the number of segments which are *exported*.

14.2.1. *Segments, Imports and Exports.* The ability to communicate large amounts of data from one work-package to some subsequent work-package is a key feature of the JAM availability system. An export segment, defined as the set $\mathbb{G}$, is an octet sequence of fixed length $\mathsf{W}_S\mathsf{W}_C = 4104$. It is the smallest datum which may individually be imported from—or exported to—the long-term *Imports DA* during the Refine function of a work-package. Being an exact multiple of the erasure-coding piece size ensures that the data segments of work-package can be efficiently placed in the DA system.

$$\mathbb{G} \equiv \mathbb{Y}_{\mathsf{W}_S\mathsf{W}_C} \qquad (175)$$

Exported segments are data which are *generated* through the execution of the Refine logic and thus are a side effect of transforming the work-package into a work-report. Since their data is deterministic based on the execution of the Refine logic, we do not require any particular commitment to them in the work-package beyond knowing how many are associated with each Refine invocation in order that we can supply an exact index.

On the other hand, imported segments are segments which were exported by previous work-packages. In order to them to be easily fetched and verified they are referenced not by hash but rather the root of a Merkle tree which includes any other segments introduced at the time, together with an index into this sequence. This allows for justifications of correctness to be generated, stored, included alongside the fetched data and verified. This is described in depth in the next section.

14.2.2. *Data Collection and Justification.* It is the task of a guarantor to reconstitute all imported segments through fetching said segments' erasure-coded chunks from enough unique validators. Reconstitution alone is not enough since corruption of the data would occur if one or more validators provided an incorrect chunk. For this reason we ensure that the import segment specification (a Merkle root and an index into the tree) be a kind of cryptographic commitment capable of having a justification applied to demonstrate that any particular segment is indeed correct.

Justification data must be available to any node over the course of its segment's potential requirement. At around 350 bytes to justify a single segment, justification data is too voluminous to have all validators store all data. We therefore use the same overall availability framework for hosting justification metadata as the data itself.

The guarantor is able to use this proof to justify to themselves that they are not wasting their time on incorrect behavior. We do not force auditors to go through the same process. Instead, guarantors build an *Auditable Work Package*, and place this in the Audit DA system.

This is the original work-package, its extrinsic data, its imported data and a concise proof of correctness of that imported data. This tactic routinely duplicates data between the Imports DA and the Audits DA, however it is acceptable in order to reduce the bandwidth cost for auditors who must justify the correctness as cheaply as possible as auditing happens on average 30 times for each work-package whereas guaranteeing happens only twice or thrice.

14.3. **Packages and Items.** We begin by defining a *work-package*, of set $\mathbb{P}$, and its constituent *work items*, of set $\mathbb{I}$. A work-package includes a simple blob acting as an authorization token $\mathbf{j}$, the index of the service which hosts the authorization code h, an authorization code hash c and a parameterization blob $\mathbf{p}$, a context $\mathbf{x}$ and a sequence of work items $\mathbf{i}$:

$$\mathbb{P} \equiv \begin{matrix} \mathbf{j} & \mathbb{Y}, & h & \mathbb{N}_S, & c & \mathbb{H}, \\ \mathbf{p} & \mathbb{Y}, & \mathbf{x} & \mathbb{X}, & \mathbf{i} & \mathbb{I}_{\ 1\mathrm{I}} \end{matrix} \tag{176}$$

A work item includes: s the identifier of the service to which it relates, the code hash of the service at the time of reporting c (whose preimage must be available from the perspective of the lookup anchor block), a payload blob y, a gas limit g, and the three elements of its manifest, a sequence of imported data segments $\mathbf{i}$ identified by the root of the *segments tree* and an index into it, $\mathbf{x}$, a sequence of hashed of blob hashes and lengths to be introduced in this block (and which we assume the validator knows) and e the number of data segments exported by this work item:

$$\mathbb{I} \equiv \begin{matrix} s & \mathbb{N}_S, & c & \mathbb{H}, & \mathbf{y} & \mathbb{Y}, & g & \mathbb{N}_G, \\ \mathbf{i} & \mathbb{H}, \mathbb{N} & , & \mathbf{x} & (\mathbb{H}, \mathbb{N}) & , & e & \mathbb{N} \end{matrix} \tag{177}$$

We limit the total number of exported items to $\mathrm{W}_M = 2^{11}$. We also place the same limit on the total number of imported items:

$$p \in \mathbb{P} \quad \sum_{i \in p_{\mathbf{i}}} i_e \le \mathrm{W}_M \qquad \sum_{i \in p_{\mathbf{i}}} i_{\mathbf{i}} \le \mathrm{W}_M \tag{178}$$

We make an assumption that the preimage to each extrinsic hash in each work-item is known by the guarantor. In general this data will be passed to the guarantor alongside the work-package.

We limit the encoded size of a work-package plus the total size of the implied import and extrinsic items to 12MB in order to allow for around 2MB/s/core data throughput:

$$p \in \mathbb{P} \quad \sum_{i \in p_{\mathbf{i}}} i_{\mathbf{i}} \ \mathrm{W}_S\mathrm{W}_C + \sum_{i \in p_{\mathbf{i}}} \sum_{(h;l) \in i_{\mathbf{x}}} l \ \le \mathrm{W}_P \tag{179}$$

$$\mathrm{W}_P = 12 \ 2^{20} \tag{180}$$

We define the item-to-result function C as:

$$C \in \begin{cases} (\mathbb{I}, \mathbb{Y} \ \mathbb{J}) \to \mathbb{L} \\ ((s, c, \mathbf{y}, g), o) \quad (s, c, \mathrm{H}(\mathbf{y}), g, o) \end{cases} \tag{181}$$

We define the work-package's implied authorizer as $\mathbf{p}_a$, the hash of the concatenation of the authorization code and the parameterization. We define the authorization code as $\mathbf{p}_\mathbf{c}$ and require that it be available at the time of the lookup anchor block from the historical lookup of service h. Formally:

$$\mathbf{p} \in \mathbb{P} \quad \begin{matrix} \mathbf{p}_a \equiv \mathrm{H}(\mathbf{p}_\mathbf{c} \quad \mathbf{p}_\mathbf{p}) \\ \mathbf{p}_\mathbf{c} \equiv (\delta[\mathbf{p}_h], (\mathbf{p}_x)_t, \mathbf{p}_c) \\ \mathbf{p}_\mathbf{c} \in \mathbb{Y} \end{matrix} \tag{182}$$

(is the historical lookup function defined in equation 94.)

14.3.1. *Exporting.* Any of a work-package's work-items may *export* segments and a *segments-root* is placed in the work-report committing to these, ordered according to the work-item which is exporting. It is formed as the root of a constant-depth binary Merkle tree as defined in equation 300:

Guarantors are required to erasure-code and distribute two data sets: one blob, the auditable work-package containing the encoded work-package, extrinsic data and self-justifying imported segments which is placed in the short-term Audit DA store and a second set of exported-segments data together with the *Paged-Proofs* metadata. Items in the first store are short-lived; assurers are expected to keep them only until finality of the block which included the work-result. Items in the second, meanwhile, are long-lived and expected to be kept for a minimum of 28 days (672 complete epochs) following the reporting of the work-report.

We define the paged-proofs function P which accepts a series of exported segments $\mathbf{s}$ and defines some series of additional segments placed into the Imports DA system via erasure-coding and distribution. The function evaluates to pages of hashes, together with subtree proofs, such that justifications of correctness based on a segments-root may be made from it:

$$P \in \begin{cases} \mathbb{G} \to \mathbb{G} \\ \mathbf{s} \quad [\mathrm{P}_{\mathrm{W}_S}(\mathrm{J}_6(\mathbf{s}, i) \quad \mathbf{s}_{i\ +64}) \quad i \in 64 \ \mathbb{N}_{:\mathbf{s}:/64}] \end{cases} \tag{183}$$

Note: in the case that $\mathbf{s}$ is not a multiple of 64, then the term $\mathbf{s}_{i\ +64}$ will correctly refer to fewer than 64 elements if it is the final page.

14.4. **Computation of Work Results.** We now come to the work result computation function . This forms the basis for all utilization of cores on JAM. It accepts some work-package $\mathbf{p}$ for some nominated core c and results in either an error or the work result and series of exported segments. This function is deterministic and requires only that it be evaluated within eight epochs of a recently finalized block thanks to the historical lookup functionality. It can thus comfortably be evaluated by any node within the auditing period, even allowing for practicalities of imperfect synchronization.

Formally:

$$(\mathbb{P}, \mathbb{N}_\mathrm{C}) \to \mathbb{W}$$
$$(\mathbf{p}, c) \quad \begin{cases} \text{if } \mathbf{o} \in \mathbb{Y} \\ a \ \mathbf{p}_a, \mathbf{o}, x \ \mathbf{p}_x, s, \mathbf{r} \quad \text{otherwise} \end{cases} \tag{184}$$

where:

$$\mathbf{o} = {}_I(\mathbf{p}, c)$$
$$(\mathbf{r}, \overline{\mathbf{e}}) = {}^{\mathrm{T}}[(C(\mathbf{p}_\mathbf{i}[j], r), \mathbf{e}) \quad (r, \mathbf{e}) = I(\mathbf{p}, j), j \in \mathbb{N}_{\mathbf{p}_\mathbf{i}}]$$
$$I(\mathbf{p}, j) \equiv R(\mathbf{p}, \mathbf{p}_\mathbf{i}[j], \sum_{k<j} \mathbf{p}_\mathbf{i}[k]_e)$$
$$R(\mathbf{p}, i, \ell) \equiv {}_R(i_c, i_g, i_s, \mathrm{H}(\mathbf{p}), i_\mathbf{y}, \mathbf{p}_\mathbf{x}, \mathbf{p}_a, M(i), X(i), \ell)$$

The definition here is staged over several functions for ease of reading. The first term to be introduced, $\mathbf{o}$ is the authorization output, the result of the Is-Authorized function. The second term, $(\mathbf{r}, \overline{\mathbf{e}})$ is the sequence of results for

each of the work-items in the work-package together with all segments exported by each work-item.

The third and forth definition are helper terms for this, with the third I performing an ordered accumulation (i.e. counter) in order to ensure that the Refine function has access to the total number of exports made from the work-package up to the current work-item. The fourth term, R, is essentially just a call to the Refine function, marshalling the relevant data from the work-package and work-item.

The above relies on two functions, M and X which, respectively, define the import segment data and the extrinsic data for some work-item argument i:

$$\begin{aligned} M(i \quad \mathbb{I}) &\equiv [\mathbf{s}[n] \quad (\mathbf{M}(\mathbf{s}), n) \in i_{\mathbf{i}}] \\ X(i \quad \mathbb{I}) &\equiv [\mathbf{d} \quad (\mathbf{H}(\mathbf{d}), \mathbf{d}) \in i_{\mathbf{x}}] \end{aligned} \tag{185}$$

We may then define s as the data availability specification of the package using these two functions together with the yet to be defined *Availability Specifier* function A (see section 14.4.1):

$$\begin{aligned} s = A(\mathbf{H}(\mathbf{p}), \mathbf{E}(\updownarrow \mathbf{p}, \mathbf{x}, \mathbf{i}, \mathbf{j}), \overline{\mathbf{e}}) \\ \text{where } \mathbf{x} = [[\updownarrow \mathbf{E}(\mathbf{d}) \quad \mathbf{d} \quad X(i)] \quad i \in \mathbf{p}_{\mathbf{i}}] \\ \text{and } \mathbf{i} = [M(i) \quad i \in \mathbf{p}_{\mathbf{i}}] \\ \text{and } \mathbf{j} = [\mathrm{J}(\mathbf{s}, n) \quad (\mathbf{M}(\mathbf{s}), n) \in i_{\mathbf{i}}, i \in \mathbf{p}_{\mathbf{i}}] \end{aligned}$$

Note that while M and $\mathbf{j}$ are both formulated using the term $\mathbf{s}$ (all segments exported by all work-packages exporting a segment to be imported) such a vast amount of data is not generally needed as the justification can be derived through a single paged-proof. This reduces the worst case data fetching for a guarantor to two segments for every one to be imported. In the case that contiguously exported segments are imported (which we might assume is a fairly common situation), then a single proof-page should be sufficient to justify many imported segments.

The Is-Authorized logic it references is first executed to ensure that the work-package warrants the needed core-time. Next, the guarantor should ensure that all segment-tree roots which form imported segment commitments are known and have not expired. Finally, the guarantor should ensure that they can fetch all preimage data referenced as the commitments of extrinsic segments.

Once done, then imported segments must be reconstructed. This process may in fact be lazy as the Refine function makes no usage of the data until the *import* host-call is made. Fetching generally implies that, for each imported segment, erasure-coded chunks are retrieved from enough unique validators (342, including the guarantor) and is described in more depth in appendix H. (Since we specify systematic erasure-coding, its reconstruction is trivial in the case that the correct 342 validators are responsive.) Chunks must be fetched for both the data itself and for justification metadata which allows us to ensure that the data is correct.

Validators, in their role as availability assurers, should index such chunks according to the index of the segments-tree whose reconstruction they facilitate. Since the data for segment chunks is so small at 12 bytes, fixed communications costs should be kept to a bare minimum. A good network protocol (out of scope at present) will allow guarantors to specify only the segments-tree root and index together with a Boolean to indicate whether the proof chunk need be supplied. Since we assume at least 341 other validators are online and benevolent, we can assume that the guarantor can compute $\mathbf{i}$ and $\mathbf{j}$ above with confidence, based on the general availability of data committed to with $\mathbf{s}^{\clubsuit}$, which is specified below.

14.4.1. *Availability Specifier.* We define the availability specifier function A, which creates an availability specifier from the package hash, an octet sequence of the audit-friendly work-package bundle (comprising the work-package itself, the extrinsic data and the concatenated import segments along with their proofs of correctness), and the sequence of exported segments:

$$\begin{aligned} A \in \{ & \mathbb{H}, \mathbb{Y}, \mathbb{G} \to \mathbb{S} \\ & h, \mathbf{b}, \mathbf{s} \quad h, l \quad \mathbf{b}, u, e \quad \mathbf{M}(\mathbf{s}) \end{aligned} \tag{186}$$

$$\begin{aligned} \text{where } u &= \mathbf{M}_B([\mathbf{x} \quad \mathbf{x} \quad ^{\mathrm{T}}[\mathbf{b}^{\clubsuit}, \mathbf{s}^{\clubsuit}]]) \\ \text{and } \mathbf{b}^{\clubsuit} &= \mathrm{H}^{\#}(\mathrm{C}_{\circ \mathbf{b} \wr \mathrm{W}_C}(\mathrm{P}_{\mathrm{W}_C}(\mathbf{b}))) \\ \text{and } \mathbf{s}^{\clubsuit} &= \mathbf{M}_B^{\#}(^{\mathrm{T}}\mathrm{C}_6^{\#}(\mathbf{s} \quad P(\mathbf{s}))) \end{aligned}$$

The paged-proofs function P, defined earlier in equation 183, accepts a sequence of segments and returns a sequence of paged-proofs sufficient to justify the correctness of every segment. There are exactly $1/64$ paged-proof segments as the number of yielded segments, each composed of a page of 64 hashes of segments, together with a Merkle proof from the root to the subtree-root which includes those 64 segments.

The functions $\mathbf{M}$ and $\mathbf{M}_B$ are the fixed-depth and simple binary Merkle root functions, defined in equations 300 and 299. The function C is the erasure-coding function, defined in appendix H.

And P is the zero-padding function to take an octet array to some multiple of n in length:

$$\begin{aligned} \mathrm{P}_n \in \{ & \mathbb{Y} \to \mathbb{Y}_{k\,n} \\ & \mathbf{x} \quad \mathbf{x} \quad [0, 0, \ldots]_{((x+n-1) \bmod n)+1 \ldots n} \end{aligned} \tag{187}$$

Validators are incentivized to distribute each newly erasure-coded data chunk to the relevant validator, since they are not paid for guaranteeing unless a work-report is considered to be *available* by a super-majority of validators. Given our work-package $\mathbf{p}$, we should therefore send the corresponding work-package bundle chunk and exported segments chunks to each validator whose keys are together with similarly corresponding chunks for imported, extrinsic and exported segments data, such that each validator can justify completeness according to the work-report's *erasure-root*. In the case of a coming epoch change, they may also maximize expected reward by distributing to the new validator set.

We will see this function utilized in the next sections, for guaranteeing, auditing and judging.

15. Guaranteeing

Guaranteeing work-packages involves the creation and distribution of a corresponding *work-report* which requires certain conditions to be met. Along with the report, a signature demonstrating the validator's commitment to its correctness is needed. With two guarantor signatures, the work-report may be distributed to the forthcoming Jam chain block author in order to be used in the $\mathbf{E}_G$, which leads to a reward for the guarantors.

We presume that in a public system, validators will be punished severely if they malfunction and commit to a

report which does not faithfully represent the result of applied on a work-package. Overall, the process is:

(1) Evaluation of the work-package's authorization, and cross-referencing against the authorization pool in the most recent JAM chain state.
(2) Creation and publication of a work-package report.
(3) Chunking of the work-package and each of its extrinsic and exported data, according to the erasure codec.
(4) Distributing the aforementioned chunks across the validator set.
(5) Providing the work-package, extrinsic and exported data to other validators on request is also helpful for optimal network performance.

For any work-package p we are in receipt of, we may determine the work-report, if any, it corresponds to for the core c that we are assigned to. When JAM chain state is needed, we always utilize the chain state of the most recent block.

For any guarantor of index v assigned to core c and a work-package p, we define the work-report r simply as:

$$r = \quad (p, c) \tag{188}$$

Such guarantors may safely create and distribute the payload (s, v). The component s may be created according to equation 141; specifically it is a signature using the validator's registered Ed25519 key on a payload l:

$$l = \mathrm{H}(c, r) \tag{189}$$

To maximize profit, the guarantor should require the work result meets all expectations which are in place during the guarantee extrinsic described in section 11.4. This includes contextual validity, inclusion of the authorization in the authorization pool, and ensuring total gas is at most G_A. No doing so does not result in punishment, but will prevent the block author from including the package and so reduces rewards.

Advanced nodes may maximize the likelihood that their reports will be includable on-chain by attempting to predict the state of the chain at the time that the report will get to the block author. Naive nodes may simply use the current chain head when verifying the work-report. To minimize work done, nodes should make all such evaluations *prior* to evaluating the $_R$ function to calculate the report's work results.

Once evaluated as a reasonable work-package to guarantee, guarantors should maximize the chance that their work is not wasted by attempting to form consensus over the core. To achieve this they should send the work-package to any other guarantors on the same core which they do not believe already know of it.

In order to minimize the work for block authors and thus maximize expected profits, guarantors should attempt to construct their core's next guarantee extrinsic from the work-report, core index and set of attestations including their own and as many others as possible.

In order to minimize the chance of any block authors disregarding the guarantor for anti-spam measures, guarantors should sign an average of no more than two work-reports per timeslot.

16. Availability Assurance

Validators should issue a signed statement, called an *assurance*, when they are in possession of all of their corresponding erasure-coded chunks for a given work-report which is currently pending availability. For any work-report to gain an assurance, there are four classes of data a validator must have:

Firstly, their erasure-coded chunk for this report. The validity of this chunk can be trivially proven through the work-report's work-package erasure-root and a Merkle-proof of inclusion in the correct location. The proof should be included from the guarantor. The chunk should be retained for 28 days and provided to any validator on request.

Secondly, the two manifests for the two classes of data items; *required* and *provided*. These have commitments as binary Merkle roots in the work-report. They must be provided alongside the following data but are only needed to verify its validity and completeness and need not be retained after the work-report is considered audited. Until then, it should be provided on request to validators.

Thirdly, the validator should already have in hand their corresponding erasure-coded chunk for each of the items in the *required* manifest. These chunks may be proven in a similar manner as for the work-package, with a Merkle proof on the root included in the manifest.[13] All such items must not be scheduled for expiry for another 4,800 timeslots. (If they are then the work-report should not be considered available.)

Finally, the validator should have in hand the corresponding erasure-coded chunk for each of the items in the *provided* manifest. Much as the work-package chunk these should be retained for 28 days and provided to any validator on request.

17. Auditing and Judging

The auditing and judging system is theoretically equivalent to that in Elves, introduced by Jeff Burdges, Cevallos, et al. 2024. For a full security analysis of the mechanism, see this work. There is a difference in terminology, where the terms *backing*, *approval* and *inclusion* there refer to our guaranteeing, auditing and accumulation, respectively.

17.1. **Overview.** The auditing process involves each node requiring themselves to fetch, evaluate and issue judgement on a random but deterministic set of work-reports from each JAM chain block in which the work-report becomes available (i.e. from $\mathbf{W}$). Prior to any evaluation, a node declares and proves its requirement. At specific common junctures in time thereafter, the set of work-reports which a node requires itself to evaluate from each block's $\mathbf{W}$ may be enlarged if any declared intentions are not matched by a positive judgement in a reasonable time or in the event of a negative judgement being seen. These enlargement events are called tranches.

If all declared intentions for a work-report are matched by a positive judgement at any given juncture, then the work-report is considered *audited*. Once all of any given block's newly available work-reports are audited, then we consider the block to be *audited*. One prerequisite of a

[13] If it appears their own availability system is incomplete for the last 28 days of blocks, then helpful validators will make some effort to reconstruct their chunk by making requests from other validators.

node finalizing a block is for it to view the block as audited. Note that while there will be eventual consensus on whether a block is audited, there may not be consensus at the time that the block gets finalized. This does not affect the crypto-economic guarantees of this system.

In regular operation, no negative judgements will ultimately be found for a work-report, and there will be no direct consequences of the auditing stage. In the unlikely event that a negative judgement is found, then one of several things happens; if there are still more than $^2/_3$V positive judgements, then validators issuing negative judgements may receive a punishment for time-wasting. If there are greater than $^1/_3$V negative judgements, then the block which includes the work-report is ban-listed. It and all its descendants are disregarded and may not be built on. In all cases, once there are enough votes, a judgement extrinsic can be constructed by a block author and placed on-chain to denote the outcome. See section 10 for details on this.

All announcements and judgements are published to all validators along with metadata describing the signed material. On receipt of sure data, validators are expected to update their perspective accordingly (later defined as J and A).

17.2. **Data Fetching.** For each work-report to be audited, we use its erasure-root to request erasure-coded chunks from enough assurers. From each assurer we fetch three items (which with a good network protocol should be done under a single request) corresponding to the work-package super-chunks, the self-justifying imports super-chunks and the extrinsic segments super-chunks.

We may validate the work-package reconstruction by ensuring its hash is equivalent to the hash includes as part of the work-package specification in the work-report. We may validate the extrinsic segments through ensuring their hashes are each equivalent to those found in the relevant work-item.

Finally, we may validate each imported segment as a justification must follow the concatenated segments which allows verification that each segment's hash is included in the referencing Merkle root and index of the corresponding work-item.

Exported segments need not be reconstructed in the same way, but rather should be determined in the same manner as with guaranteeing, i.e. through the execution of the Refine logic.

All items in the work-package specification field of the work-report should be recalculated from this now known-good data and verified, essentially retracing the guarantors steps and ensuring correctness.

17.3. **Selection of Reports.** Each validator shall perform auditing duties on each valid block received. Since we are entering off-chain logic, and we cannot assume consensus, we henceforth consider ourselves a specific validator of index v and assume ourselves focused on some block $\mathbf{B}$ with other terms corresponding, so σ' is said block's posterior state, $\mathbf{H}$ is its header &c. Practically, all considerations must be replicated for all blocks and multiple blocks' considerations may be underway simultaneously.

We define the sequence of work-reports which we may be required to audit as $\mathbf{Q}$, a sequence of length equal to the number of cores, which functions as a mapping of core index to a work-report pending which has just become available, or if no report became available on the core. Formally:

(190) $\mathbf{Q} \quad \mathbb{W}?_{\mathbb{C}}$

(191) $\mathbf{Q} \equiv [\begin{cases} \rho[c]_w & \text{if } \rho[c]_w \quad \mathbf{W} \\ & \text{otherwise} \end{cases} \quad c \in \mathbb{N}_C]$

We define our initial audit tranche in terms of a verifiable random quantity s_0 created specifically for it:

(192) $s_0 \quad \mathbb{F}^{[]}_{[v]_b} \; \mathrm{X}_U \quad \mathrm{Y}(\mathrm{H}_v)$

(193) $\mathrm{X}_U = \texttt{Ma9h¬9«L\_§}$

We may then define $\mathbf{a}_0$ as the non-empty items to audit through a verifiably random selection of ten cores:

(194) $\mathbf{a}_0 = \{ \; c, w \quad c, w \quad \mathbf{p}_{+10}, w \neq \; \}$

(195) where $\mathbf{p} = \mathrm{F}([\; c, \mathbf{Q}_c \quad c \quad \mathbb{N}_C], r)$

(196) and $r = \mathrm{Y}(s_0)$

Every $\mathrm{A} = 8$ seconds following a new time slot, a new tranche begins, and we may determine that additional cores warrant an audit from us. Such items are defined as $\mathbf{a}_n$ where n is the current tranche. Formally:

(197) $$\text{let } n = \frac{\mathrm{T} - \mathbf{P} \quad \tau'}{\mathrm{A}}$$

New tranches may contain items from $\mathbf{Q}$ stemming from one of two reasons: either a negative judgement has been received; or the number of judgements from the previous tranche is less than the number of announcements from said tranche. In the first case, the validator is always required to issue a judgement on the work-report. In the second case, a new special-purpose VRF must be constructed to determine if an audit and judgement is warranted from us.

In all cases, we publish a signed statement of which of the cores we believe we are required to audit (an *announcement*) together with evidence of the VRF signature to select them and the other validators' announcements from the previous tranche unmatched with a judgement in order that all other validators are capable of verifying the announcement. *Publication of an announcement should be taken as a contract to complete the audit regardless of any future information.*

Formally, for each tranche n we ensure the announcement statement is published and distributed to all other validators along with our validator index v, evidence s_n and all signed data. Validator's announcement statements must be in the set:

(198) $\mathbb{E}_{[v]_e} \; \mathrm{X}_I \mathbin{+\!\!+} n \quad \mathrm{E}([\mathrm{E}_2(c) \quad \mathrm{H}(w) \quad c, w \quad \mathbf{a}_0])$

(199) $\mathrm{X}_I = \texttt{Ma9h¬9jjn«jHO}$

We define A_n as our perception of which validator is required to audit each of the work-reports (identified by their associated core) at tranche n. This comes from each other validators' announcements (defined above). It cannot be correctly evaluated until n is current. We have absolute knowledge about our own audit requirements.

(200) $A_n \quad \mathbb{W} \to \quad \mathbb{N}_V$

(201) $(c, w) \quad \mathbf{a}_0 \quad v \quad q_0(w)$

We further define J and J to be the validator indices who we know to have made respectively, positive and negative, judgements mapped from each work-report's core. We don't care from which tranche a judgement is made.

(202) $J_{\{\ ;\ \}}\ \mathbb{W} \to \ \mathbb{N}_V$

We are able to define $\mathbf{a}_n$ for tranches beyond the first on the basis of the number of validators who we know are required to conduct an audit yet from whom we have not yet seen a judgement. It is possible that the late arrival of information alters $\mathbf{a}_n$ and nodes should reevaluate and act accordingly should this happen.

We can thus define $\mathbf{a}_n$ beyond the initial tranche through a new VRF which acts upon the set of *no-show* validators.

$n > 0$

(203) $s_n(w)\ \ \mathbb{F}^{[]}_{[v]_b}\ \mathrm{X}_U\ \ \mathrm{Y}(\mathrm{H}_v)\ \ \mathrm{H}(w) + n$

(204) $\mathbf{a}_n \equiv \{w\ \ \mathbf{Q}\ \ \frac{\mathrm{F}}{256\mathrm{V}}\mathrm{Y}(s_n(w))_0 < A_{n-1}(w)\ \ J\ (w)\ \}$

We define our bias factor $\mathrm{F} = 2$, which is the expected number of validators which will be required to issue a judgement for a work-report given a single no-show in the tranche before. Modeling by Jeff Burdges, Cevallos, et al. 2024 shows that this is optimal.

Later audits must be announced in a similar fashion to the first. If audit requirements lesson on the receipt of new information (i.e. a positive judgement being returned for a previous *no-show*), then any audits already announced are completed and judgements published. If audit requirements raise on the receipt of new information (i.e. an additional announcement being found without an accompanying judgement), then we announce the additional audit(s) we will undertake.

As n increases with the passage of time $\mathbf{a}_n$ becomes known and defines our auditing responsibilities. We must attempt to reconstruct all work-packages and their requisite data corresponding to each work-report we must audit. This may be done through requesting erasure-coded chunks from one-third of the validators. It may also be short-cutted through asking a cooperative third-party (e.g. an original guarantor) for the preimages.

Thus, for any such work-report w we are assured we will be able to fetch some candidate work-package encoding $F(w)$ which comes either from reconstructing erasure-coded chunks verified through the erasure coding's Merkle root, or alternatively from the preimage of the work-package hash. We decode this candidate blob into a work-package.

In addition to the work-package, we also assume we are able to fetch all manifest data associated with it through requesting and reconstructing erasure-coded chunks from one-third of validators in the same way as above.

We then attempt to reproduce the report on the core to give e_n, a mapping from cores to evaluations:

(205) $$e_n(w)\ \ \mathbf{a}_n\ \ \Leftrightarrow\ \ w = \begin{cases} (p,c)\ \text{if}\ \ p\ \ \mathbb{P}\ \ \mathrm{E}(p) = F(w) \\ \text{otherwise} \end{cases}$$

Note that a failure to decode implies an invalid work-report.

From this mapping the validator issues a set of judgements $\mathbf{j}_n$:

(206) $\mathbf{j}_n = \{\mathrm{S}\ _{[v]_e}(\mathrm{X}_{e(w)}\ \ \mathrm{H}(w))\ \ (c,w)\ \ \mathbf{a}_n\}$

All judgements $\mathbf{j}$ should be published to other validators in order that they build their view of J and in the case of a negative judgement arising, can form an extrinsic for $\mathbf{E}_D$.

We consider a work-report as audited under two circumstances. Either, when it has no negative judgements and there exists some tranche in which we see a positive judgement from all validators who we believe are required to audit it; or when we see positive judgements for it from greater than two-thirds of the validator set.

(207) $$U(w)\ \Leftrightarrow\ \ \begin{cases} J\ (w) = \quad n\ \ A_n(w)\ \ J\ (w) \\ J\ (w)\ > {}^2/_3\mathrm{V} \end{cases}$$

Our block $\mathbf{B}$ may be considered audited, a condition denoted $\mathbf{U}$, when all the work-reports which were made available are considered audited. Formally:

(208) $\mathbf{U} \Leftrightarrow\ \ w\ \ \mathbf{W}\ \ U(w)$

For any block we must judge it to be audited (i.e. $\mathbf{U} =\ $) before we vote for the block to be finalized in GRANDPA. See section 19 for more information here.

Furthermore, we pointedly disregard chains which include the accumulation of a report which we know at least $1/3$ of validators judge as being invalid. Any chains including such a block are not eligible for authoring on. The *best block*, i.e. that on which we build new blocks, is defined as the chain with the most regular Safrole blocks which does *not* contain any such disregarded block. Implementation-wise, this may require reversion to an earlier head or alternative fork.

As a block author, we include a judgement extrinsic which collects judgement signatures together and reports them on-chain. In the case of a non-valid judgement (i.e. one which is not two-thirds-plus-one of judgements confirming validity) then this extrinsic will be introduced in a block in which accumulation of the non-valid work-report is about to take place. The non-valid judgement extrinsic removes it from the pending work-reports, ρ. Refer to section 10 for more details on this.

18. BEEFY DISTRIBUTION

For each finalized block $\mathbf{B}$ which a validator imports, said validator shall make a BLS signature on the BLS12-381 curve, as defined by Hopwood et al. 2020, affirming the Keccak hash of the block's most recent BEEFY MMR. This should be published and distributed freely, along with the signed material. These signatures may be aggregated in order to provide concise proofs of finality to third-party systems. The signing and aggregation mechanism is defined fully by Jeff Burdges, Ciobotaru, et al. 2022.

Formally, let $\mathbf{F}_v$ be the signed commitment of validator index v which will be published:

(209) $\mathbf{F}_v \equiv \mathrm{S}\ _v(\mathrm{X}_B\ \ \mathrm{H}_K(\mathrm{E}_M(\mathrm{last}(\beta)_\mathbf{b}])))$

(210) $\mathrm{X}_B = \texttt{Ma9h¬@OOT´}$

19. GRANDPA AND THE BEST CHAIN

Nodes take part in the GRANDPA protocol as defined by Stewart and Kokoris-Kogia 2020.

We define the latest finalized block as $\mathbf{B}$. All associated terms concerning block and state are similarly superscripted. We consider the *best block*, $\mathbf{B}$ to be that which

is drawn from the set of acceptable blocks of the following criteria:

- Has the finalized block as an ancestor.
- Contains no unfinalized blocks where we see an equivocation (two valid blocks at the same timeslot).
- Is considered audited.

Formally:

$$\mathbf{A}(\mathbf{H}) \quad \mathbf{H} \tag{211}$$

$$\mathbf{U} \equiv \tag{212}$$

$$/\ \mathbf{H}^A, \mathbf{H}^B \quad \begin{array}{l} \mathbf{H}^A \neq \mathbf{H}^B \\ \mathbf{H}^A_T = \mathbf{H}^B_T \\ \mathbf{H}^A \quad \mathbf{A}(\mathbf{H}) \\ \mathbf{H}^A\ /\ \mathbf{A}(\mathbf{H}) \end{array} \tag{213}$$

Of these acceptable blocks, that which contains the most ancestor blocks whose author used a seal-key ticket, rather than a fallback key should be selected as the best head, and thus the chain on which the participant should make GRANDPA votes.

Formally, we aim to select $\mathbf{B}$ to maximize the value m where:

$$m = \sum_{\mathbf{H}^A\ \mathbf{A}} \mathbf{T}^A \tag{214}$$

20. Discussion

20.1. **Technical Characteristics.** In total, with our stated target of 1,023 validators and three validators per core, along with requiring a mean of ten audits per validator per timeslot, and thus 30 audits per work-report, JAM is capable of trustlessly processing and integrating 341 work-packages per timeslot.

We assume node hardware is a modern 16 core CPU with 64GB RAM, 1TB secondary storage and 0.5Gbe networking.

Our performance models assume a rough split of CPU time as follows:

	Proportion
Audits	10/16
Merklization	1/16
Block execution	2/16
GRANDPA and BEEFY	1/16
Erasure coding	1/16
Networking & misc	1/16

Estimates for network bandwidth requirements are as follows:

	Upload Mb/s	*Download* Mb/s
Guaranteeing	30	40
Assuring	60	56
Auditing	200	200
Block publication	42	42
GRANDPA and BEEFY	4	4
Total	**336**	**342**

Thus, a connection able to sustain 500Mb/s should leave a sufficient margin of error and headroom to serve other validators as well as some public connections, though the burstiness of block publication would imply validators are best to ensure that peak bandwidth is higher.

Under these conditions, we would expect an overall network-provided data availability capacity of 2PB, with each node dedicating at most 6TB to availability storage.

Estimates for memory usage are as follows:

	GB	
Auditing	20	2 × 10 PVM instances
Block execution	2	1 PVM instance
State cache	40	
Misc	2	
Total	**64**	

As a rough guide, each parachain has an average footprint of around 2MB in the Polkadot Relay chain; a 40GB state would allow 20,000 parachains' information to be retained in state.

What might be called the "virtual hardware" of a JAM core is essentially a regular CPU core executing at somewhere between 25% and 50% of regular speed for the whole six-second portion and which may draw and provide 2.5MB/s average in general-purpose I/O and utilize up to 2GB in RAM. The I/O includes any trustless reads from the JAM chain state, albeit in the recent past. This virtual hardware also provides unlimited reads from a semi-static preimage-lookup database.

Each work-package may occupy this hardware and execute arbitrary code on it in six-second segments to create some result of at most 90KB. This work result is then entitled to 10ms on the same machine, this time with no "external" I/O beyond said result, but instead with full and immediate access to the JAM chain state and may alter the service(s) to which the results belong.

20.2. **Illustrating Performance.** In terms of pure processing power, the JAM machine architecture can deliver extremely high levels of homogeneous trustless computation. However, the core model of JAM is a classic parallelized compute architecture, and for solutions to be able to utilize the architecture well they must be designed with it in mind to some extent. Accordingly, until such use-cases appear on JAM with similar semantics to existing ones, it is very difficult to make direct comparisons to existing systems. That said, if we indulge ourselves with some assumptions then we can make some crude comparisons.

20.2.1. *Comparison to Polkadot.* Pre-asynchronous backing, Polkadot validates around 50 parachains, each one utilizing approximately 250ms of native computation (i.e. half a second of Wasm execution time at around a 50% overhead) and 5MB of I/O for every twelve seconds of real time which passes. This corresponds to an aggregate compute performance of around parity with a native CPU core and a total 24-hour distributed availability of around 20MB/s. Accumulation is beyond Polkadot's capabilities and so not comparable.

Post asynchronous-backing and estimating that Polkadot is at present capable of validating at most 80 parachains each doing one second of native computation in every six, then the aggregate performance is increased to around 13x native CPU and the distributed availability increased to around 67MB/s.

For comparison, in our basic models, JAM should be capable of attaining around 85x the computation load of a single native CPU core and a distributed availability of 852MB/s.

20.2.2. *Simple Transfers.* We might also attempt to model a simple transactions-per-second amount, with each transaction requiring a signature verification and the modification of two account balances. Once again, until there are clear designs for precisely how this would work we must make some assumptions. Our most naive model would be to use the JAM cores (i.e. refinement) simply for transaction verification and account lookups. The JAM chain would then hold and alter the balances in its state. This is unlikely to give great performance since almost all the needed I/O would be synchronous, but it can serve as a basis.

A 15MB work-package can hold around 125k transactions at 128 bytes per transaction. However, a 90KB work-result could only encode around 11k account updates when each update is given as a pair of a 4 byte account index and 4 byte balance, resulting in a limit of 5.5k transactions per package, or 312k TPS in total. It is possible that the eight bytes could typically be compressed by a byte or two, increasing maximum throughput a little. Our expectations are that state updates, with highly parallelized Merklization, can be done at between 500k and 1 million reads/write per second, implying around 250k-350k TPS, depending on which turns out to be the bottleneck.

A more sophisticated model would be to use the JAM cores for balance updates as well as transaction verification. We would have to assume that state and the transactions which operate on them can be partitioned between work-packages with some degree of efficiency, and that the 15MB of the work-package would be split between transaction data and state witness data. Our basic models predict that a 4bn 32-bit account system paginated into 2^{10} accounts/page and 128 bytes per transaction could, assuming only around 1% of oraclized accounts were useful, average upwards of 1.7mTPS depending on partitioning and usage characteristics. Partitioning could be done with a fixed fragmentation (essentially sharding state), a rotating partition pattern or a dynamic partitioning (which would require specialized sequencing).

Interestingly, we expect neither model to be bottlenecked in computation, meaning that transactions could be substantially more sophisticated, perhaps with more flexible cryptography or smart contract functionality, without a significant impact on performance.

20.2.3. *Computation Throughput.* The TPS metric does not lend itself well to measuring distributed systems' computational performance, so we now turn to another slightly more compute-focussed benchmark: the EVM. The basic *YP* Ethereum network, now approaching a decade old, is probably the best known example of general purpose decentralized computation and makes for a reasonable yardstick. It is able to sustain a computation and I/O rate of 1.25M gas/sec, with a peak throughput of twice that. The EVM gas metric was designed to be a time-proportional metric for predicting and constraining program execution. Attempting to determine a concrete comparison to PVM throughput is non-trivial and necessarily opinionated owing to the disparity between the two platforms including word size, endianness and stack/register architecture and memory model. However, we will attempt to determine a reasonable range of values.

EVM gas does not directly translate into native execution as it also combines state reads and writes as well as transaction input data, implying it is able to process some combination of up to 595 storage reads, 57 storage writes and 1.25M gas as well as 78KB input data in each second, trading one against the other.[14] We cannot find any analysis of the typical breakdown between storage I/O and pure computation, so to make a very conservative estimate, we assume it does all four. In reality, we would expect it to be able to do on average 1/4 of each.

Our experiments[15] show that on modern, high-end consumer hardware with a modern EVM implementation, we can expect somewhere between 100 and 500 gas/µs in throughput on pure-compute workloads (we specifically utilized Odd-Product, Triangle-Number and several implementations of the Fibonacci calculation). To make a conservative comparison to PVM, we propose transcompilation of the EVM code into PVM code and then re-execution of it under the PolkaVM prototype.[16]

To help estimate a reasonable lower-bound of EVM gas/µs, e.g. for workloads which are more memory and I/O intensive, we look toward real-world permissionless deployments of the EVM and see that the Moonbeam network, after correcting for the slowdown of executing within the recompiled WebAssembly platform on the somewhat conservative Polkadot hardware platform, implies a throughput of around 100 gas/µs. We therefore assert that in terms of computation, 1µs EVM gas approximates to around 100-500 gas on modern high-end consumer hardware.[17]

Benchmarking and regression tests show that the prototype PVM engine has a fixed preprocessing overhead of around 5ns/byte of program code and, for arithmetic-heavy tasks at least, a marginal factor of 1.6-2% compared to EVM execution, implying an asymptotic speedup of around 50-60x. For machine code 1MB in size expected to take of the order of a second to compute, the compilation cost becomes only 0.5% of the overall time. [18] For code not inherently suited to the 256-bit EVM ISA, we would expect substantially improved relative execution times on PVM, though more work must be done in order to gain confidence that these speed-ups are broadly applicable.

[14] The latest "proto-danksharding" changes allow it to accept 87.3KB/s in committed-to data though this is not directly available within state, so we exclude it from this illustration, though including it with the input data would change the results little.

[15] This is detailed at `\§§q I¤¤\9HchLu_n¤>±±±kGgo«~~G¯-iS|³9~?UX¤[`9}¦ª\`8` and intended to be updated as we get more information.

[16] It is conservative since we don't take into account that the source code was originally compiled into EVM code and thus the PVM machine code will replicate architectural artifacts and thus is very likely to be pessimistic. As an example, all arithmetic operations in EVM are 256-bit and 32-bit native PVM is being forced to honor this even if the source code only actually required 32-bit values.

[17] We speculate that the substantial range could possibly be caused in part by the major architectural differences between the EVM ISA typical modern hardware.

[18] As an example, our odd-product benchmark, a very much pure-compute arithmetic task, execution takes 58s on EVM, and 1.04s within our PVM prototype, including all preprocessing.

If we allow for preprocessing to take up to the same component within execution as the marginal cost (owing to, for example, an extremely large but short-running program) and for the PVM metering to imply a safety overhead of 2x to execution speeds, then we can expect a JAM core to be able to process the equivalent of around 1,500 EVM gas/µs. Owing to the crudeness of our analysis we might reasonably predict it to be somewhere within a factor of three either way—i.e. 500-5,000 EVM gas/µs.

JAM cores are each capable of 2.5MB/s bandwidth, which must include any state I/O and data which must be newly introduced (e.g. transactions). While writes come at comparatively little cost to the core, only requiring hashing to determine an eventual updated Merkle root, reads must be witnessed, with each one costing around 640 bytes of witness conservatively assuming a one-million entry binary Merkle trie. This would result in a maximum of a little under 4k reads/second/core, with the exact amount dependent upon how much of the bandwidth is used for newly introduced input data.

Aggregating everything across JAM, excepting accumulation which could add further throughput, numbers can be multiplied by 341 (with the caveat that each one's computation cannot interfere with any of the others' except through state oraclization and accumulation). Unlike for *roll-up chain* designs such as Polkadot and Ethereum, there is no need to have persistently fragmented state. Smart-contract state may be held in a coherent format on the JAM chain so long as any updates are made through the 15KB/core/sec work results, which would need to contain only the hashes of the altered contracts' state roots.

Under our modelling assumptions, we can therefore summarize:

	Eth. L1	JAM Core	JAM
Compute (EVM gas/µs)	$1.25^{\dagger}$	500-5,000	0.15-1.5M
State writes (s^{-1})	$57^{\dagger}$	n/a	n/a
State reads (s^{-1})	$595^{\dagger}$	$4K^{\ddagger}$	$1.4M^{\ddagger}$
Input data (s^{-1})	$78KB^{\dagger}$	$2.5MB^{\ddagger}$	$852MB^{\ddagger}$

What we can see is that JAM's overall predicted performance profile implies it could be comparable to many thousands of that of the basic Ethereum L1 chain. The large factor here is essentially due to three things: spacial parallelism, as JAM can host several hundred cores under its security apparatus; temporal parallelism, as JAM targets continuous execution for its cores and pipelines much of the computation between blocks to ensure a constant, optimal workload; and platform optimization by using a VM and gas model which closely fits modern hardware architectures.

It must however be understood that this is a provisional and crude estimation only. It is included for only the purpose of expressing JAM's performance in tangible terms and is not intended as a means of comparing to a "full-blown" Ethereum/L2-ecosystem combination. Specifically, it does not take into account:

- that these numbers are based on real performance of Ethereum and performance modelling of JAM (though our models are based on real-world performance of the components);
- any L2 scaling which may be possible with either JAM or Ethereum;
- the state partitioning which uses of JAM would imply;
- the as-yet unfixed gas model for the PVM;
- that PVM/EVM comparisons are necessarily imprecise;
- (†) all figures for Ethereum L1 are drawn from the same resource: on average each figure will be only 1/4 of this maximum.
- (‡) the state reads and input data figures for JAM are drawn from the same resource: on average each figure will be only 1/2 of this maximum.

We leave it as further work for an empirical analysis of performance and an analysis and comparison between JAM and the aggregate of a hypothetical Ethereum ecosystem which included some maximal amount of L2 deployments together with full Dank-sharding and any other additional consensus elements which they would require. This, however, is out of scope for the present work.

21. Conclusion

We have introduced a novel computation model which is able to make use of pre-existing crypto-economic mechanisms in order to deliver major improvements in scalability without causing persistent state-fragmentation and thus sacrificing overall cohesion. We call this overall pattern collect-refine-join-accumulate. Furthermore, we have formally defined the on-chain portion of this logic, essentially the join-accumulate portion. We call this protocol the JAM chain.

We argue that the model of JAM provides a novel "sweet spot", allowing for massive amounts of computation to be done in secure, resilient consensus compared to fully-synchronous models, and yet still have strict guarantees about both timing and integration of the computation into some singleton state machine unlike persistently fragmented models.

21.1. Further Work.

While we are able to estimate theoretical computation possible given some basic assumptions and even make broad comparisons to existing systems, practical numbers are invaluable. We believe the model warrants further empirical research in order to better understand how these theoretical limits translate into real-world performance. We feel a proper cost analysis and comparison to pre-existing protocols would also be an excellent topic for further work.

We can be reasonably confident that the design of JAM allows it to host a service under which Polkadot *parachains* could be validated, however further prototyping work is needed to understand the possible throughput which a PVM-powered metering system could support. We leave such a report as further work. Likewise, we have also intentionally omitted details of higher-level protocol elements including cryptocurrency, coretime sales, staking and regular smart-contract functionality.

A number of potential alterations to the protocol described here are being considered in order to make practical utilization of the protocol easier. These include:

- Synchronous calls between services in accumulate.
- Restrictions on the `transfer` function in order to allow for substantial parallelism over accumulation.

- The possibility of reserving substantial additional computation capacity during accumulate under certain conditions.
- Introducing Merklization into the Work Package format in order to obviate the need to have the whole package downloaded in order to evaluate its authorization.

The networking protocol is also left intentionally undefined at this stage and its description must be done in a follow-up proposal.

Validator performance is not presently tracked on-chain. We do expect this to be tracked on-chain in the final revision of the JAM protocol, but its specific format is not yet certain and it is therefore omitted at present.

22. Acknowledgements

Much of this present work is based in large part on the work of others. The Web3 Foundation research team and in particular Alistair Stewart and Jeff Burdges are responsible for Elves, the security apparatus of Polkadot which enables the possibility of in-core computation for JAM. The same team is responsible for Sassafras, Grandpa and Beefy.

Safrole is a mild simplification of Sassafras and was made under the careful review of Davide Galassi and Alistair Stewart.

The original CoreJam RFC was refined under the review of Bastian Köcher and Robert Habermeier and most of the key elements of that proposal have made their way into the present work.

The PVM is a formalization of a partially simplified *PolkaVM* software prototype, developed by Jan Bujak. Cyrill Leutwiler contributed to the empirical analysis of the PVM reported in the present work.

The *PolkaJam* team and in particular Arkadiy Paronyan, Emeric Chevalier and Dave Emmet have been instrumental in the design of the lower-level aspects of the JAM protocol, especially concerning Merklization and I/O.

Numerous contributors to the repository since publication have helped correct errors. Thank you to all.

And, of course, thanks to the awesome Lemon Jelly, a.k.a. Fred Deakin and Nick Franglen, for three of the most beautiful albums ever produced, the cover art of the first of which was inspiration for this paper's background art.

Appendix A. Polka Virtual Machine

A.1. **Basic Definition.** We declare the general PVM function . We assume a single-step invocation function define $_1$ and define the full PVM recursively as a sequence of such mutations up until the single-step mutation results in a halting condition.

$$(\mathbb{Y}, \mathbb{N}_R, \mathbb{N}_G, \ \mathbb{N}_{R\ 13}, \mathbb{M}) \to (\{\ ,\ ,\infty\}\ \ \{\mathsf{S}\} \times \mathbb{N}_R\ \ \{\hbar\} \times \mathbb{N}_R, \mathbb{N}_R, \mathbb{Z}_G, \ \mathbb{N}_{R\ 13}, \mathbb{M})$$

$$(\mathbf{p}, \imath, \xi, \omega, \mu) \quad \begin{cases} (\mathbf{p}, \imath', \xi', \omega', \mu') & \text{if } \varepsilon = \\ (\infty, \imath', \xi', \omega', \mu') & \text{if } \xi' < 0 \\ (\varepsilon, \imath', \xi', \omega', \mu') & \text{otherwise} \end{cases} \tag{215}$$

$$\text{where } (\varepsilon, \imath', \xi', \omega', \mu') = \ _1(\mathbf{c}, \mathbf{k}, \mathbf{j}, \imath, \xi, \omega, \mu)$$

$$\text{and } \mathbf{p} = \mathrm{E}(\mathbf{j})\ \ \mathrm{E}_1(z)\ \ \mathrm{E}(\mathbf{c})\ \ \mathrm{E}_z(\mathbf{j})\ \ \mathrm{E}(\mathbf{c})\ \ \mathrm{E}(\mathbf{k}),\ \ \mathbf{k} = \mathbf{c}$$

If the latter condition cannot be satisfied, then $(\ , \imath, \xi, \omega, \mu)$ is the result.

The PVM exit reason ε $\{\ ,\ ,\infty\}$ $\{\mathsf{S}, \hbar\} \times \mathbb{N}_R$ may be one of regular halt , panic or out-of-gas ∞, or alternatively a host-call $\hbar$, in which the host-call identifier is associated, or page-fault S in which case the address into RAM is associated.

A.2. **Instructions, Opcodes and Skip-distance.** The program blob $\mathbf{p}$ is split into a series of octets which make up the *instruction data* $\mathbf{c}$ and the *opcode bitmask* $\mathbf{k}$ as well as the *dynamic jump table*, $\mathbf{j}$. The former two imply an instruction sequence, and by extension a *basic-block sequence*, itself a sequence of indices of the instructions which follow a *block-termination* instruction.

The latter, dynamic jump table, is a sequence of indices into the instruction data blob and is indexed into when dynamically-computed jumps are taken. It is encoded as a sequence of natural numbers (i.e. non-negative integers) each encoded with the same length in octets. This length, term z above, is itself encoded prior.

The PVM counts instructions in octet terms (rather than in terms of instructions) and it is thus convenient to define which octets represent the beginning of an instruction, i.e. the opcode octet, and which do not. This is the purpose of $\mathbf{k}$, the instruction-opcode bitmask. We assert that the length of the bitmask is no smaller than the length of the instruction blob (and in fact is simply rounded to the nearest multiple of eight for ease of octet-encoding).

We define the Skip function skip which provides the number of octets, minus one, to the next instruction's opcode, given the index of instruction's opcode index into $\mathbf{c}$ (and by extension $\mathbf{k}$):

$$\text{skip} \begin{cases} \mathbb{N} \to \mathbb{N} \\ i \quad \min(24,\ j \quad \mathbb{N}\ \ \mathbf{k}_{i+1+j} = 1) \end{cases} \tag{216}$$

Given some instruction-index i, its opcode is readily expressed as $\mathbf{c}_i$ and the distance in octets to move forward to the next instruction is $1 + \text{skip}(i)$. However, each instruction's "length" (defined as the number of contiguous octets starting with the opcode which are needed to fully define the instruction's semantics) is left implicit though limited to being at most 16.

We define ζ as being equivalent to the instructions $\mathbf{c}$ except with an indefinite sequence of zeroes suffixed to ensure that no out-of-bounds access is possible. This effectively defines any otherwise-undefined arguments to the final instruction and ensures that a trap will occur if the program counter passes beyond the program code. Formally:

$$\zeta \equiv \mathbf{c}\quad [0, 0, \dots] \tag{217}$$

A.3. **Basic Blocks and Termination Instructions.** Instructions of the following opcodes are considered basic-block termination instructions; other than `§}9q` & `T9ee§\}n«X\`, they correspond to instructions which may define the instruction-counter to be something other than its prior value plus the instruction's skip amount:

- Trap and fallthrough: `§}9q` , `T9ee§\}n«X\`
- Jumps: `a«hq` , `a«hq¬_jL`
- Load-and-Jumps: `en9L¬_hh¬a«hq` , `en9L¬_hh¬a«hq¬_jL`
- Branches: `@}9jH\¬Ox` , `@}9jH\¬jO` , `@}9jH\¬XO¬«` , `@}9jH\¬XO¬` , `@}9jH\¬e§¬«` , `@}9jH\¬e§¬` , `@}9jH\¬Ox¬_hh` , `@}9jH\¬jO¬_hh`
- Immediate branches: `@}9jH\¬e§¬«¬_hh` , `@}9jH\¬e§¬ ¬_hh` , `@}9jH\¬eO¬«¬_hh` , `@}9jH\¬eO¬ ¬_hh` , `@}9jH\¬XO¬«¬_hh` , `@}9jH\¬XO¬ ¬_hh` , `@}9jH\¬X§¬«¬_hh` , `@}9jH\¬X§¬ ¬_hh`

We denote this set, as opcode indices rather than names, as T. We define the instruction opcode indices denoting the beginning of basic-blocks as ϖ:

$$\varpi \equiv [0]\quad [n + 1 + \text{skip}(n)\ \ n \in \mathbb{N}_{\mathbf{c}}\ \ \mathbf{k}_n = 1\ \ \mathbf{c}_n\ \ T] \tag{218}$$

A.4. **Single-Step State Transition.** We must now define the single-step PVM state-transition function $_1$:

$$_1 \begin{cases} (\mathbb{Y},\ \mathbb{N}_R\ , \mathbb{N}_R, \mathbb{N}_G,\ \mathbb{N}_{R\ 13}, \mathbb{M}) \to (\{\ ,\ ,\ \}\ \ \{\mathsf{S}, \hbar\} \times \mathbb{N}_R, \mathbb{Z}_G,\ \mathbb{N}_{R\ 13}, \mathbb{M}) \\ (\mathbf{c}, \mathbf{j}, \imath, \xi, \omega, \mu)\quad (\varepsilon, \imath', \xi', \omega', \mu') \end{cases} \tag{219}$$

We define ε together with the posterior values (denoted as prime) of each of the items of the machine state as being in accordance with the table below.

In general, when transitioning machine state for an instruction a number of conditions hold true and instructions are defined essentially by their exceptions to these rules. Specifically, the machine does not halt, the instruction counter

increments by one, the gas remaining is reduced by the amount corresponding to the instruction type and RAM & registers are unchanged. Formally:

$$\varepsilon = \quad , \quad \imath' = \imath + 1 + \text{skip}(\imath), \quad \xi' = \xi - \xi \quad , \quad \omega' = \omega, \quad \mu' = \mu \text{ except as indicated} \tag{220}$$

Where RAM must be inspected and yet access is not possible, then machine state is unchanged, and the exit reason is a fault with the lowest address to be read which is inaccessible. More formally, let $\mathbf{a}$ be the set of indices in to which μ must be subscripted in order to calculate the result of $\quad_1$. If $\mathbf{a} \not\in \mathbb{V} \quad$ then let $\varepsilon = \mathsf{S} \times \min(\mathbf{a} \quad \mathbb{V} \quad)$.

Similarly, where RAM must be mutated and yet mutable access is not possible, then machine state is unchanged, and the exit reason is a fault with the lowest address to be read which is inaccessible. More formally, let $\mathbf{a}$ be the set of indices in to which μ' must be subscripted in order to calculate the result of $\quad_1$. If $\mathbf{a} \not\in \mathbb{V} \quad$ then let $\varepsilon = \mathsf{S} \times \min(\mathbf{a} \quad \mathbb{V} \quad)$.

We define signed/unsigned transitions for various octet widths:

$$\mathbb{N}_{2^{8n}} \to \mathbb{Z}_{-2^{8n-1}\,2^{8n-1}}$$

$$\mathsf{Z}_{n\ \mathrm{N}} \quad a \quad \begin{cases} a & \text{if } a < 2^{8n-1} \\ a - 2^{8n} & \text{otherwise} \end{cases} \tag{221}$$

$$\mathbb{Z}_{-2^{8n-1}\,2^{8n-1}} \to \mathbb{N}_{2^{8n}}$$

$$\mathsf{Z}^{-1}_{n\ \mathrm{N}} \quad a \quad (2^{8n} + a) \bmod 2^{8n} \tag{222}$$

$$\mathbb{N}_{2^{8n}} \to \mathbb{B}_{8n}$$

$$\mathsf{B}_{n\ \mathrm{N}} \quad x \quad \mathbf{y} \quad i \quad \mathbb{N}_{2^{8n}} \quad \mathbf{y}[i] \Leftrightarrow \frac{x}{2^i} \mod 2 \tag{223}$$

$$\mathbb{B}_{8n} \to \mathbb{N}_{2^{8n}}$$

$$\mathsf{B}^{-1}_{n\ \mathrm{N}} \quad \mathbf{x} \quad y \quad \sum_{i\ \mathrm{N}_{2^{8n}}} \mathbf{x}_i \ 2^i \tag{224}$$

Immediate arguments are encoded in little-endian format with the most-significant bit being the sign bit. They may be compactly encoded by eliding more significant octets. Elided octets are assumed to be zero if the MSB of the value is zero, and 255 otherwise. This allows for compact representation of both positive and negative encoded values. We thus define the signed extension function operating on an input of n octets as X_n:

$$\mathbb{N}_{2^{8n}} \to \mathbb{N}_{2^{32}}$$

$$\mathsf{X}_{n\ \{0;1;2;3;4\}} \quad x \quad x + \frac{x}{2^{8n-1}} \ (2^{32} - 2^{8n}) \tag{225}$$

Any alterations of the program counter stemming from a static jump, call or branch must be to the start of a basic block or else a panic occurs. Hypotheticals are not considered. Formally:

$$\texttt{@}\}\texttt{9jH}\backslash(b, C) \qquad (\varepsilon, \imath') = \begin{cases} (\ , \imath) & \text{if } \neg C \\ (\ , \imath) & \text{otherwise if } b \not\in \varpi \\ (\ , b) & \text{otherwise} \end{cases} \tag{226}$$

Jumps whose next instruction is dynamically computed must use an address which may be indexed into the jump-table $\mathbf{j}$. Through a quirk of tooling[19], we define the dynamic address required by the instructions as the jump table index incremented by one and then multiplied by our jump alignment factor $\mathsf{Z}_A = 4$.

As with other irregular alterations to the program counter, target code index must be the start of a basic block or else a panic occurs. Formally:

$$\texttt{La«hq}(a) \qquad (\varepsilon, \imath') = \begin{cases} (\ , \imath) & \text{if } a = 2^{32} - 2^{16} \\ (\ , \imath) & \text{otherwise if } a = 0 \quad a > \mathbf{j} \quad \mathsf{Z}_A \quad a \bmod \mathsf{Z}_A \neq 0 \quad \mathbf{j}_{(a/\mathsf{Z}_A)-1} \not\in \varpi \\ (\ , \mathbf{j}_{(a/\mathsf{Z}_A)-1}) & \text{otherwise} \end{cases} \tag{227}$$

A.5. **Instruction Tables.** Note that in the case that the opcode is not defined in the following tables then the instruction is considered invalid, and it results in a panic; $\varepsilon = \quad$.

We assume the skip length ℓ is well-defined:

$$\ell \equiv \text{skip}(\imath) \tag{228}$$

A.5.1. *Instructions without Arguments.*

ℓ	i 9 h O		g «§9§_n j
0	§}9q	0	$\varepsilon =$
17	T9ee§\}n«X\	0	

[19]The popular code generation backend LLVM requires and assumes in its code generation that dynamically computed jump destinations always have a certain memory alignment. Since at present we depend on this for our tooling, we must acquiesce to its assumptions.

A.5.2. *Instructions with Arguments of One Immediate.*

$$\text{let } l_X = \min(4, \ell), \quad \nu_X \equiv \mathrm{X}_{l_X}(\mathrm{E}^{-1}_{l_X}(\zeta_{\imath+1\ +l_X})) \tag{229}$$

ℓ	i 9 h 0		g «§9§_nj
78	OH9ee_	0	$\varepsilon = \hbar \times \nu_X$

A.5.3. *Instructions with Arguments of Two Immediates.*

$$\begin{aligned} &\text{let } l_X = \min(4, \zeta_{\imath+1} \bmod 8), && \nu_X \equiv \mathrm{X}_{l_X}(\mathrm{E}^{-1}_{l_X}(\zeta_{\imath+2\ +l_X})) \\ &\text{let } l_Y = \min(4, \max(0, \ell - l_X - 1)), && \nu_Y \equiv \mathrm{X}_{l_Y}(\mathrm{E}^{-1}_{l_Y}(\zeta_{\imath+2+l_X\ +l_Y})) \end{aligned} \tag{230}$$

ℓ	i 9 h 0		g «§9§_nj
62	§n}O¬_hh¬«P	0	$\mu'_{\square_X} = \nu_Y \bmod 2^8$
79	§n}O¬_hh¬«o£	0	$\mu'_{\square_{X\ +2}} = \mathrm{E}_2(\nu_Y \bmod 2^{16})$
38	§n}O¬_hh¬«¨©	0	$\mu'_{\square_{X\ +4}} = \mathrm{E}_4(\nu_Y)$

A.5.4. *Instructions with Arguments of One Offset.*

$$\text{let } l_X = \min(4, \ell), \quad \nu_X \equiv \imath + \mathrm{Z}_{l_X}(\mathrm{E}^{-1}_{l_X}(\zeta_{\imath+1\ +l_X})) \tag{231}$$

ℓ	i 9 h 0		g «§9§_nj
5	a«hq	0	@}9jH\(ν_X,)

A.5.5. *Instructions with Arguments of One Register & One Immediate.*

$$\begin{aligned} &\text{let } r_A = \min(12, \zeta_{\imath+1} \bmod 16), && \omega_A \equiv \omega_{r_A}, \quad \omega'_A \equiv \omega'_{r_A} \\ &\text{let } l_X = \min(4, \max(0, \ell - 1)), && \nu_X \equiv \mathrm{X}_{l_X}(\mathrm{E}^{-1}_{l_X}(\zeta_{\imath+2\ +l_X})) \end{aligned} \tag{232}$$

ℓ	i 9 h 0		g «§9§_nj
19	a«hq¬_jL	0	$\mathrm{La}«\mathrm{hq}((\omega_A + \nu_X) \bmod 2^{32})$
4	en9L¬_hh	0	$\omega'_A = \nu_X$
60	en9L¬«P	0	$\omega'_A = \mu_{\ X}$
74	en9L¬_P	0	$\omega'_A = \mathrm{Z}_4^{-1}(\mathrm{Z}_1(\mu_{\ X}))$
76	en9L¬«o£	0	$\omega'_A = \mathrm{E}_2^{-1}(\mu_{\square_{X\ +2}})$
66	en9L¬_o£	0	$\omega'_A = \mathrm{Z}_4^{-1}(\mathrm{Z}_2(\mathrm{E}_2^{-1}(\mu_{\square_{X\ +2}})))$
10	en9L¬«¨©	0	$\omega'_A = \mathrm{E}_4^{-1}(\mu_{\square_{X\ +4}})$
71	§n}O¬«P	0	$\mu'_{\square_X} = \omega_A \bmod 2^8$
69	§n}O¬«o£	0	$\mu'_{\square_{X\ +2}} = \mathrm{E}_2(\omega_A \bmod 2^{16})$
22	§n}O¬«¨©	0	$\mu'_{\square_{X\ +4}} = \mathrm{E}_4(\omega_A)$

A.5.6. *Instructions with Arguments of One Register & Two Immediates.*

$$\begin{aligned} &\text{let } r_A = \min(12, \zeta_{\imath+1} \bmod 16), && \omega_A \equiv \omega_{r_A}, \quad \omega'_A \equiv \omega'_{r_A} \\ &\text{let } l_X = \min(4, \frac{\zeta_{\imath+1}}{16} \bmod 8), && \nu_X = \mathrm{X}_{l_X}(\mathrm{E}^{-1}_{l_X}(\zeta_{\imath+2\ +l_X})) \\ &\text{let } l_Y = \min(4, \max(0, \ell - l_X - 1)), && \nu_Y = \mathrm{X}_{l_Y}(\mathrm{E}^{-1}_{l_Y}(\zeta_{\imath+2+l_X\ +l_Y})) \end{aligned} \tag{233}$$

ℓ	i 9 h 0		g «§9§_nj
26	§n}O¬_hh¬_jL¬«P	0	$\mu'_{\square_{!_{A+}\ X}} = \nu_Y \bmod 2^8$
54	§n}O¬_hh¬_jL¬«o£	0	$\mu'_{\square_{!_{A+}\ X\ +2}} = \mathrm{E}_2(\nu_Y \bmod 2^{16})$

13	§n}O¬_hh¬_jL¬«¨©	0	$\mu'_{\square_{A+\ X\ +4}} = \mathrm{E}_4(\nu_Y)$

A.5.7. *Instructions with Arguments of One Register, One Immediate and One Offset.*

$$
\begin{aligned}
&\text{let } r_A = \min(12, \zeta_{\iota+1} \bmod 16), && \omega_A \equiv \omega_{r_A}, \quad \omega'_A \equiv \omega'_{r_A} \\
&\text{let } l_X = \min(4, \frac{\zeta_{\iota+1}}{16} \bmod 8), && \nu_X = \mathrm{X}_{l_X}(\mathrm{E}^{-1}_{l_X}(\zeta_{\iota+2\ +l_X})) \\
&\text{let } l_Y = \min(4, \max(0, \ell - l_X - 1)), && \nu_Y = \imath + \mathrm{Z}_{l_Y}(\mathrm{E}^{-1}_{l_Y}(\zeta_{\iota+2+l_X\ +l_Y}))
\end{aligned} \tag{234}
$$

6	en9L¬_hh¬a«hq	0	@}9jH\(ν_Y,), $\omega'_A = \nu_X$
7	@}9jH\¬Ox¬_hh	0	@}9jH\($\nu_Y, \omega_A = \nu_X$)
15	@}9jH\¬jO¬_hh	0	@}9jH\($\nu_Y, \omega_A \neq \nu_X$)
44	@}9jH\¬e§¬«¬_hh	0	@}9jH\($\nu_Y, \omega_A < \nu_X$)
59	@}9jH\¬eO¬«¬_hh	0	@}9jH\($\nu_Y, \omega_A \leq \nu_X$)
52	@}9jH\¬XO¬«¬_hh	0	@}9jH\($\nu_Y, \omega_A \geq \nu_X$)
50	@}9jH\¬X§¬«¬_hh	0	@}9jH\($\nu_Y, \omega_A > \nu_X$)
32	@}9jH\¬e§¬ ¬_hh	0	@}9jH\($\nu_Y, \mathrm{Z}_4(\omega_A) < \mathrm{Z}_4(\nu_X)$)
46	@}9jH\¬eO¬ ¬_hh	0	@}9jH\($\nu_Y, \mathrm{Z}_4(\omega_A) \leq \mathrm{Z}_4(\nu_X)$)
45	@}9jH\¬XO¬ ¬_hh	0	@}9jH\($\nu_Y, \mathrm{Z}_4(\omega_A) \geq \mathrm{Z}_4(\nu_X)$)
53	@}9jH\¬X§¬ ¬_hh	0	@}9jH\($\nu_Y, \mathrm{Z}_4(\omega_A) > \mathrm{Z}_4(\nu_X)$)

A.5.8. *Instructions with Arguments of Two Registers.*

$$
\begin{aligned}
&\text{let } r_D = \min(12, (\zeta_{\iota+1}) \bmod 16), && \omega_D \equiv \omega_{r_D}, \quad \omega'_D \equiv \omega'_{r_D} \\
&\text{let } r_A = \min(12, \frac{\zeta_{\iota+1}}{16}), && \omega_A \equiv \omega_{r_A}, \quad \omega'_A \equiv \omega'_{r_A}
\end{aligned} \tag{235}
$$

82	hn®O¬}OX	0	$\omega'_D = \omega_A$
87	@}c	0	$\omega'_D \equiv \min(x \quad \mathbb{N}_R)$ $x \geq h$ $\mathbb{N}_{x\ +!_A}$ / $\mathbb{V}$ $\mathbb{N}_{x\ +!_A}$ $\mathbb{V}$,

Note, the term h above refers to the beginning of the heap, the second major segment of memory as defined in equation 246 as $2\mathrm{Z}_Q + Q(\mathrm{o})$. If @}c instruction is invoked on a PVM instance which does not have such a memory layout, then $h = 0$.

A.5.9. *Instructions with Arguments of Two Registers & One Immediate.*

$$
\begin{aligned}
&\text{let } r_A = \min(12, (\zeta_{\iota+1}) \bmod 16), && \omega_A \equiv \omega_{r_A}, \quad \omega'_A \equiv \omega'_{r_A} \\
&\text{let } r_B = \min(12, \frac{\zeta_{\iota+1}}{16}), && \omega_B \equiv \omega_{r_B}, \quad \omega'_B \equiv \omega'_{r_B} \\
&\text{let } l_X = \min(4, \max(0, \ell - 1)), && \nu_X \equiv \mathrm{X}_{l_X}(\mathrm{E}^{-1}_{l_X}(\zeta_{\iota+2\ +l_X}))
\end{aligned} \tag{236}
$$

16	§n}O¬_jL¬«P	0	$\mu'_{\square_{B+\ X}} = \omega_A \bmod 2^8$
29	§n}O¬_jL¬«o£	0	$\mu'_{\square_{B+\ X\ +2}} = \mathrm{E}_2(\omega_A \bmod 2^{16})$
3	§n}O¬_jL¬«¨©	0	$\mu'_{\square_{B+\ X\ +4}} = \mathrm{E}_4(\omega_A)$
11	en9L¬_jL¬«P	0	$\omega'_A = \mu_{!_{B+\ X}}$

ℓ	i 9 h O		g «§9§_nj
21	en9L¬_jL¬_P	0	$\omega'_A = \mathrm{Z}_4^{-1}(\mathrm{Z}_1(\mu_{!_B+\ X}))$
37	en9L¬_jL¬«o£	0	$\omega'_A = \mathrm{E}_2^{-1}(\mu_{!_{B+\ X\ +2}})$
33	en9L¬_jL¬_o£	0	$\omega'_A = \mathrm{Z}_4^{-1}(\mathrm{Z}_2(\mathrm{E}_2^{-1}(\mu_{!_{B+\ X\ +2}})))$
1	en9L¬_jL¬«¨©	0	$\omega'_A = \mathrm{E}_4^{-1}(\mu_{!_{B+\ X\ +4}})$
2	9LL¬_hh	0	$\omega'_A = (\omega_B + \nu_X) \bmod 2^{32}$
18	9jL¬_hh	0	$i \in \mathbb{N}_{32} \quad \mathrm{B}_4(\omega'_A)_i = \mathrm{B}_4(\omega_B)_i \quad \mathrm{B}_4(\nu_X)_i$
31	²n}¬_hh	0	$i \in \mathbb{N}_{32} \quad \mathrm{B}_4(\omega'_A)_i = \mathrm{B}_4(\omega_B)_i \quad \mathrm{B}_4(\nu_X)_i$
49	n}¬_hh	0	$i \in \mathbb{N}_{32} \quad \mathrm{B}_4(\omega'_A)_i = \mathrm{B}_4(\omega_B)_i \quad \mathrm{B}_4(\nu_X)_i$
35	h«e¬_hh	0	$\omega'_A = (\omega_B \quad \nu_X) \bmod 2^{32}$
65	h«e¬«qqO}¬ ¬ ¬_hh	0	$\omega'_A = \mathrm{Z}_4^{-1}(\ (\mathrm{Z}_4(\omega_B)\ \ \mathrm{Z}_4(\nu_X)) \div 2^{32}\ )$
63	h«e¬«qqO}¬«¬«¬_hh	0	$\omega'_A = \ (\omega_B \quad \nu_X) \div 2^{32}$
27	O§¬e§¬«¬_hh	0	$\omega'_A = \omega_B < \nu_X$
56	O§¬e§¬ ¬_hh	0	$\omega'_A = \mathrm{Z}_4(\omega_B) < \mathrm{Z}_4(\nu_X)$
9	\en¬e¬_hh	0	$\omega'_A = (\omega_B \quad 2^{\ x \bmod 32}) \bmod 2^{32}$
14	\en¬}¬_hh	0	$\omega'_A = \ \omega_B \div 2^{\ x \bmod 32}$
25	\9}¬}¬_hh	0	$\omega'_A = \mathrm{Z}_4^{-1}(\ \mathrm{Z}_4(\omega_B) \div 2^{\ x \bmod 32}\ )$
40	jOX¬9LL¬_hh	0	$\omega'_A = (\nu_X + 2^{32} - \omega_B) \bmod 2^{32}$
39	O§¬X§¬«¬_hh	0	$\omega'_A = \omega_B > \nu_X$
61	O§¬X§¬ ¬_hh	0	$\omega'_A = \mathrm{Z}_4(\omega_B) > \mathrm{Z}_4(\nu_X)$
75	\en¬e¬_hh¬9e§	0	$\omega'_A = (\nu_X \quad 2^{!_B \bmod 32}) \bmod 2^{32}$
72	\en¬}¬_hh¬9e§	0	$\omega'_A = \ \nu_X \div 2^{!_B \bmod 32}$
80	\9}¬}¬_hh¬9e§	0	$\omega'_A = \mathrm{Z}_4^{-1}(\ \mathrm{Z}_4(\nu_X) \div 2^{!_B \bmod 32}\ )$
85	Hhn®¬_¶¬_hh	0	$\omega'_A = \begin{cases} \nu_X & \text{if } \omega_B = 0 \\ \omega_A & \text{otherwise} \end{cases}$
86	Hhn®¬j¶¬_hh	0	$\omega'_A = \begin{cases} \nu_X & \text{if } \omega_B \neq 0 \\ \omega_A & \text{otherwise} \end{cases}$

A.5.10. *Instructions with Arguments of Two Registers & One Offset.*

$$
\begin{aligned}
&\text{let } r_A = \min(12, (\zeta_{\ell+1}) \bmod 16), && \omega_A \equiv \omega_{r_A}, \quad \omega'_A \equiv \omega'_{r_A} \\
&\text{let } r_B = \min(12, \ \frac{\zeta_{\ell+1}}{16}\ ), && \omega_B \equiv \omega_{r_B}, \quad \omega'_B \equiv \omega'_{r_B} \\
&\text{let } l_X = \min(4, \max(0, \ell - 1)), && \nu_X \equiv \imath + \mathrm{Z}_{l_X}(\mathrm{E}_{l_X}^{-1}(\zeta_{\ell+2\ +l_X}))
\end{aligned}
\tag{237}
$$

ℓ	i 9 h O		g «§9§_nj
24	@}9jH\¬Ox	0	@}9jH\$(\nu_X, \omega_A = \omega_B)$
30	@}9jH\¬jO	0	@}9jH\$(\nu_X, \omega_A \neq \omega_B)$
47	@}9jH\¬e§¬«	0	@}9jH\$(\nu_X, \omega_A < \omega_B)$
48	@}9jH\¬e§¬	0	@}9jH\$(\nu_X, \mathrm{Z}_4(\omega_A) < \mathrm{Z}_4(\omega_B))$
41	@}9jH\¬XO¬«	0	@}9jH\$(\nu_X, \omega_A \geq \omega_B)$
43	@}9jH\¬XO¬	0	@}9jH\$(\nu_X, \mathrm{Z}_4(\omega_A) \geq \mathrm{Z}_4(\omega_B))$

A.5.11. *Instruction with Arguments of Two Registers and Two Immediates.*

$$
\begin{aligned}
&\text{let } r_A = \min(12, (\zeta_{\imath+1}) \bmod 16), && \omega_A \equiv \omega_{r_A}, \quad \omega'_A \equiv \omega'_{r_A} \\
&\text{let } r_B = \min(12, \tfrac{\zeta_{\imath+1}}{16}), && \omega_B \equiv \omega_{r_B}, \quad \omega'_B \equiv \omega'_{r_B} \\
&\text{let } l_X = \min(4, \zeta_{\imath+2} \bmod 8), && \nu_X = \mathrm{X}_{l_X}(\mathrm{E}^{-1}_{l_X}(\zeta_{\imath+3\ +l_X})) \\
&\text{let } l_Y = \min(4, \max(0, \ell - l_X - 2)), && \nu_Y = \mathrm{X}_{l_Y}(\mathrm{E}^{-1}_{l_Y}(\zeta_{\imath+3+l_X\ +l_Y}))
\end{aligned} \tag{238}
$$

$\imath$	i 9 h O		g «§9§_nj
42	en9L¬_hh¬a«hq¬_jL	0	La«hq$((\omega_B + \nu_Y) \bmod 2^{32})$, $\omega'_A = \nu_X$

A.5.12. *Instructions with Arguments of Three Registers.*

$$
\begin{aligned}
&\text{let } r_A = \min(12, (\zeta_{\imath+1}) \bmod 16), && \omega_A \equiv \omega_{r_A}, \quad \omega'_A \equiv \omega'_{r_A} \\
&\text{let } r_B = \min(12, \tfrac{\zeta_{\imath+1}}{16}), && \omega_B \equiv \omega_{r_B}, \quad \omega'_B \equiv \omega'_{r_B} \\
&\text{let } r_D = \min(12, \zeta_{\imath+2}), && \omega_D \equiv \omega_{r_D}, \quad \omega'_D \equiv \omega'_{r_D}
\end{aligned} \tag{239}
$$

$\imath$	i 9 h O		g «§9§_nj
8	9LL	0	$\omega'_D = (\omega_A + \omega_B) \bmod 2^{32}$
20	«@	0	$\omega'_D = (\omega_A + 2^{32} - \omega_B) \bmod 2^{32}$
23	9jL	0	$i \in \mathbb{N}_{32} \quad \mathrm{B}_4(\omega'_D)_i = \mathrm{B}_4(\omega_A)_i \quad \mathrm{B}_4(\omega_B)_i$
28	²n}	0	$i \in \mathbb{N}_{32} \quad \mathrm{B}_4(\omega'_D)_i = \mathrm{B}_4(\omega_A)_i \quad \mathrm{B}_4(\omega_B)_i$
12	n}	0	$i \in \mathbb{N}_{32} \quad \mathrm{B}_4(\omega'_D)_i = \mathrm{B}_4(\omega_A)_i \quad \mathrm{B}_4(\omega_B)_i$
34	h«e	0	$\omega'_D = (\omega_A \ \ \omega_B) \bmod 2^{32}$
67	h«e¬«qqO}¬ ¬	0	$\omega'_D = \mathrm{Z}_4^{-1}(\ (\mathrm{Z}_4(\omega_A) \ \ \mathrm{Z}_4(\omega_B)) \div 2^{32}\ )$
57	h«e¬«qqO}¬«¬«	0	$\omega'_D = \ (\omega_A \ \ \omega_B) \div 2^{32}$
81	h«e¬«qqO}¬ ¬«	0	$\omega'_D = \mathrm{Z}_4^{-1}(\ (\mathrm{Z}_4(\omega_A) \ \ \omega_B) \div 2^{32}\ )$
68	L_®¬«	0	$\omega'_D = \begin{cases} 2^{32}-1 & \text{if } \omega_B = 0 \\ \omega_A \div \omega_B & \text{otherwise} \end{cases}$
64	L_®¬	0	$\omega'_D = \begin{cases} 2^{32}-1 & \text{if } \omega_B = 0 \\ \omega_A & \text{if } \mathrm{Z}_4(\omega_A) = -2^{31} \quad \mathrm{Z}_4(\omega_B) = -1 \\ \mathrm{Z}_4^{-1}(\ \mathrm{Z}_4(\omega_A) \div \mathrm{Z}_4(\omega_B)\ ) & \text{otherwise} \end{cases}$
73	}Oh¬«	0	$\omega'_D = \begin{cases} \omega_A & \text{if } \omega_B = 0 \\ \omega_A \bmod \omega_B & \text{otherwise} \end{cases}$
70	}Oh¬	0	$\omega'_D = \begin{cases} \omega_A & \text{if } \omega_B = 0 \\ 0 & \text{if } \mathrm{Z}_4(\omega_A) = -2^{31} \quad \mathrm{Z}_4(\omega_B) = -1 \\ \mathrm{Z}_4^{-1}(\mathrm{Z}_4(\omega_A) \bmod \mathrm{Z}_4(\omega_B)) & \text{otherwise} \end{cases}$
36	O§¬e§¬«	0	$\omega'_D = \omega_A < \omega_B$
58	O§¬e§¬	0	$\omega'_D = \mathrm{Z}_4(\omega_A) < \mathrm{Z}_4(\omega_B)$
55	\en¬e	0	$\omega'_D = (\omega_A \ \ 2^{\imath_B \bmod 32}) \bmod 2^{32}$
51	\en¬}	0	$\omega'_D = \ \omega_A \div 2^{\imath_B \bmod 32}$
77	\9}¬}	0	$\omega'_D = \mathrm{Z}_4^{-1}(\ \mathrm{Z}_4(\omega_A) \div 2^{\imath_B \bmod 32}\ )$
83	Hhn®¬_¶	0	$\omega'_D = \begin{cases} \omega_A & \text{if } \omega_B = 0 \\ \omega_D & \text{otherwise} \end{cases}$
84	Hhn®¬j¶	0	$\omega'_D = \begin{cases} \omega_A & \text{if } \omega_B \neq 0 \\ \omega_D & \text{otherwise} \end{cases}$

A.6. **Host Call Definition.** An extended version of the PVM invocation which is able to progress an inner *host-call* state-machine in the case of a host-call halt condition is defined as Ψ_H:

(240)
$$(\mathbb{Y}, \mathbb{N}_R, \mathbb{N}_G, \mathbb{N}_{R_{13}}, \mathbb{M}, \Psi_X, X) \to (\{\blacktriangleright, \infty, \blacktriangleleft\} \times \{\mathbb{S}\} \times \mathbb{N}_R, \mathbb{Z}_G, \mathbb{N}_{R_{13}}, \mathbb{M}, X)$$

$$\Psi_H(\mathbf{c}, \imath, \xi, \omega, \mu, f, \mathbf{x}) \equiv \begin{cases} (\blacktriangleright, \mathbb{S} \times a, \imath', \xi', \omega', \mu', \mathbf{x}') & \text{if } \begin{cases} \varepsilon = \hbar \times h \\ \mathbb{S} \times a = f(h, \xi', \omega', \mu', \mathbf{x}) \end{cases} \\ \Psi_H(\mathbf{c}, \imath' + 1 + \text{skip}(\imath'), \xi'', \omega'', \mu'', f, \mathbf{x}'') & \text{if } \begin{cases} \varepsilon = \hbar \times h \\ (\xi'', \omega'', \mu'', \mathbf{x}'') = f(h, \xi', \omega', \mu', \mathbf{x}) \end{cases} \\ (\varepsilon, \imath', \xi', \omega', \mu', \mathbf{x}) & \text{otherwise} \end{cases}$$

$$\text{where } (\varepsilon, \imath', \xi', \omega', \mu') = \Psi_1(\mathbf{c}, \imath, \xi, \omega, \mu)$$

On exit, the instruction counter $\imath'$ references the instruction *which caused the exit*. Should the machine be invoked again using this instruction counter and code, then the same instruction which caused the exit would be executed. This is sensible when the instruction is one which necessarily needs re-executing such as in the case of an out-of-gas or page fault reason.

However, when the exit reason to Ψ_H is a host-call $\hbar$, then the resultant instruction-counter has a value of the host-call instruction and resuming with this state would immediately exit with the same result. Re-invoking would therefore require both the post-host-call machine state *and* the instruction counter value for the instruction following the one which resulted in the host-call exit reason. This is always one greater plus the relevant argument skip distance. Resuming the machine with this instruction counter will continue beyond the host-call instruction.

We use both values of instruction-counter for the definition of Ψ_H since if the host-call results in a page fault we need to allow the outer environment to resolve the fault and re-try the host-call. Conversely, if we successfully transition state according to the host-call, then on resumption we wish to begin with the instruction directly following the host-call.

A.7. **Standard Program Initialization.** The software programs which will run in each of the four instances where the PVM is utilized in the main document have a very typical setup pattern characteristic of an output of a compiler and linker. This means sections for program-specific read-only data, read-write (heap) data and the stack. An adjunct to this, very typical of our usage patterns is an extra read-only segment via which invocation-specific data may be passed (i.e. arguments). It thus makes sense to define this properly in a single initializer function.

We thus define the standard program code format $\mathbf{p}$, which includes not only the instructions and jump table (previously represented by the term $\mathbf{c}$), but also information on the state of the RAM and registers at program start. Given some $\mathbf{p}$ which is appropriately encoded together with some argument data $\mathbf{a}$, we can define program code $\mathbf{c}$, registers ω and RAM μ through the standard initialization decoder function Y:

(241)
$$Y(\mathbf{p}) \to (\mathbb{Y}, \mathbb{N}_{R_{13}}, \mathbb{M})?$$

Where:

(242)
$$\text{let } \mathbf{E}_3(\mathbf{o}), \mathbf{E}_3(\mathbf{w}), \mathbf{E}_2(z), \mathbf{E}_3(s), \mathbf{o}, \mathbf{w}, \mathbf{E}_4(\mathbf{c}), \mathbf{c} = \mathbf{p}$$

(243)
$$Z_P = 2^{14}, \quad Z_Q = 2^{16}, \quad Z_I = 2^{24}$$

(244)
$$\text{let } P(x \in \mathbb{N}) \equiv Z_P \left\lceil \frac{x}{Z_P} \right\rceil, \quad Q(x \in \mathbb{N}) \equiv Z_Q \left\lceil \frac{x}{Z_Q} \right\rceil$$

(245)
$$5Z_Q + Q(|\mathbf{o}|) + Q(|\mathbf{w}| + zZ_P) + Q(s) + Z_I \le 2^{32}$$

If the above cannot be satisfied, then $x = \nabla$, otherwise $x = (\mathbf{c}, \omega, \mu)$ with $\mathbf{c}$ as above and ω, μ where:

(246)
$$i \in \mathbb{N}_R \quad \mu_i = \begin{cases} \forall \mathbf{o}_{i-Z_Q}, \mathbb{A} & \text{if } Z_Q \le i < Z_Q + |\mathbf{o}| \\ (0, R) & \text{if } Z_Q + |\mathbf{o}| \le i < Z_Q + P(|\mathbf{o}|) \\ (\mathbf{w}_{i-(2Z_Q+Q(|\mathbf{o}|))}, W) & \text{if } 2Z_Q + Q(|\mathbf{o}|) \le i < 2Z_Q + Q(|\mathbf{o}|) + |\mathbf{w}| \\ (0, W) & \text{if } 2Z_Q + Q(|\mathbf{o}|) + |\mathbf{w}| \le i < 2Z_Q + Q(|\mathbf{o}|) + P(|\mathbf{w}|) + zZ_P \\ (0, W) & \text{if } 2^{32} - 2Z_Q - Z_I - P(s) \le i < 2^{32} - 2Z_Q - Z_I \\ (\mathbf{a}_{i-(2^{32}-Z_Q-Z_I)}, R) & \text{if } 2^{32} - Z_Q - Z_I \le i < 2^{32} - Z_Q - Z_I + |\mathbf{a}| \\ (0, R) & \text{if } 2^{32} - Z_Q - Z_I + |\mathbf{a}| \le i < 2^{32} - Z_Q - Z_I + P(|\mathbf{a}|) \\ (0, \nabla) & \text{otherwise} \end{cases}$$

(247)
$$i \in \mathbb{N}_{16} \quad \omega_i = \begin{cases} 2^{32} - 2^{16} & \text{if } i = 1 \\ 2^{32} - 2Z_Q - Z_I & \text{if } i = 2 \\ 2^{32} - Z_Q - Z_I & \text{if } i = 10 \\ |\mathbf{a}| & \text{if } i = 11 \\ 0 & \text{otherwise} \end{cases}$$

A.8. **Argument Invocation Definition.** The four instances where the PVM is utilized each expect to be able to pass argument data in and receive some return data back. We thus define the common PVM program-argument invocation function Ψ_M:

$$\Psi_M \colon (\mathbb{Y}, \mathbb{N}, \mathbb{N}_G, \mathbb{Y}^{Z_I}, \Omega_X, X) \to ((\mathbb{N}_G, \mathbb{Y}) \cup \{\text{☇}, \infty\}, X) \tag{248}$$

$$\Psi_M(\mathbf{p}, \imath, \xi, \mathbf{a}, f, \mathbf{x}) \equiv \begin{cases} (\text{☇}, \mathbf{x}) & \text{if } Y(\mathbf{p}) = \text{☇} \\ R(\Psi_H(\mathbf{c}, \imath, \xi, \omega, \mu, f, \mathbf{x})) & \text{if } Y(\mathbf{p}) = (\mathbf{c}, \omega, \mu) \end{cases}$$

$$\text{where } R(\varepsilon, \imath', \xi', \omega', \mu', \mathbf{x}) \equiv \begin{cases} (\varepsilon, \mathbf{x}) & \text{if } \varepsilon = \infty \\ (\xi', \mu'^{\,\mathbf{x}'_{10} \ldots +\mathbf{x}'_{11}}) & \text{if } \varepsilon = \text{☇} \\ (\xi', []) & \text{if } \varepsilon = \text{⧺} \\ (\text{☇}, \mathbf{x}) & \text{otherwise} \end{cases} \tag{249}$$

Appendix B. Virtual Machine Invocations

B.1. **Host-Call Result Constants.**

$\mathsf{NONE} = 2^{32} - 1$**:** The return value indicating an item does not exist.
$\mathsf{OOB} = 2^{32} - 2$**:** The return value for when a memory index is provided for reading/writing which is not accessible.
$\mathsf{WHO} = 2^{32} - 3$**:** Index unknown.
$\mathsf{FULL} = 2^{32} - 4$**:** Storage full.
$\mathsf{CORE} = 2^{32} - 5$**:** Core index unknown.
$\mathsf{CASH} = 2^{32} - 6$**:** Insufficient funds.
$\mathsf{LOW} = 2^{32} - 7$**:** Gas limit too low.
$\mathsf{HIGH} = 2^{32} - 8$**:** Gas limit too high.
$\mathsf{WHAT} = 2^{32} - 9$**:** Name unknown.
$\mathsf{HUH} = 2^{32} - 10$**:** The item is already solicited or cannot be forgotten.
$\mathsf{OK} = 0$**:** The return value indicating general success.

Inner PVM invocations have their own set of result codes:

$\mathsf{HALT} = 0$**:** The invocation completed and halted normally.
$\mathsf{PANIC} = 2^{32} - 12$**:** The invocation completed with a panic.
$\mathsf{FAULT} = 2^{32} - 13$**:** The invocation completed with a page fault.
$\mathsf{HOST} = 2^{32} - 14$**:** The invocation completed with a host-call fault.

Note return codes for a host-call-request exit are any non-zero value less than $2^{32} - 13$.

B.2. **Is-Authorized Invocation.** The Is-Authorized invocation is the first and simplest of the four, being totally stateless. It provides only a single host-call function, Ω_G for determining the amount of gas remaining. It accepts as arguments the work-package as a whole, $\mathbf{p}$ and the core on which it should be executed, c. Formally, it is defined as Ψ_I:

$$\Psi_I \colon (\mathbb{P}, \mathbb{N}_C) \to \mathbb{Y} \cup \mathbb{J}$$

$$\Psi_I(\mathbf{p}, c) \equiv \begin{cases} \mathbf{r} & \text{otherwise if } \mathbf{r} \in \{\infty, \text{☇}\} \\ \mathbf{o} & \text{otherwise if } \mathbf{r} = g, \mathbf{o} \\ & \text{where } (\mathbf{r}, \_) = \Psi_M(\mathbf{p}_\mathbf{c}, 0, \mathbf{G}_I, \mathbf{E}(\mathbf{p}, c), F, \_) \end{cases} \tag{250}$$

$$F \colon (\mathbb{N}, \mathbb{N}_G, \mathbb{N}_{R_6}, \mathbb{M}) \to (\mathbb{Z}_G, \mathbb{N}_{R_2}, \mathbb{M})$$

$$F(n, \xi, \omega, \mu) \equiv \begin{cases} \Omega_G(\xi, \omega, \mu) & \text{if } n = \mathsf{gas} \\ (\xi - 10, [\mathsf{WHAT}, \omega_1, \ldots], \mu) & \text{otherwise} \end{cases} \tag{251}$$

Note for the Is-Authorized host-call dispatch function F in equation 251, we elide the host-call context since, being essentially stateless, it is always $\_$.

B.3. **Refine Invocation.** We define the Refine service-account invocation function as Ψ_R. It has no general access to the state of the JAM chain, with the slight exception being the ability to make a historical lookup. Beyond this it is able to create inner instances of the PVM and dictate pieces of data to export.

The historical-lookup host-call function, Ω_H, is designed to give the same result regardless of the state of the chain for any time when auditing may occur (which we bound to be less than two epochs from being accumulated). The lookup anchor may be up to L timeslots before the recent history and therefore adds to the potential age at the time of audit. We therefore set $\mathsf{D} = 4,800$, a safe amount of eight hours.

The inner PVM invocation host-calls, meanwhile, depend on an integrated PVM type, which we shall denote $\mathbf{M}$. It holds some program code, instruction counter and RAM:

$$\mathbf{M} \equiv (\mathbf{p} \in \mathbb{Y}, \mathbf{u} \in \mathbb{M}, i \in \mathbb{N}_R) \tag{252}$$

The Export host-call depends on two pieces of context; one sequence of segments (blobs of length W_S) to which it may append, and the other an argument passed to the invocation function to dictate the number of segments prior which

may assumed to have already been appended. The latter value ensures that an accurate segment index can be provided to the caller.

The Refine invocation function also implicitly draws upon some recent head state δ and explicitly accepts the work payload, $\mathbf{y}$, together with the service index which is the subject of refinement s, the prediction of the hash of that service's code c at the time of reporting, the hash of the containing work-package p, the refinement context $\mathbf{c}$, the authorizer hash a and its output $\mathbf{o}$, and an export segment offset ς, the import and extrinsic data blobs (both just concatenated segments) as dictated by the work-item, $\mathbf{i}$ and $\mathbf{x}$. It results in either some error $\blacktriangle$ or a pair of the refinement output blob and the export sequence. Formally:

$$\Big(\begin{matrix}\mathbb{H},\mathbb{N}_G,\mathbb{N}_S,\mathbb{H},\mathbb{Y},\mathbb{X},\\ \mathbb{H},\mathbb{Y},\ \mathbb{G}\ ,\ \mathbb{Y}\ ,\mathbb{N}\end{matrix}\Big) \to (\mathbb{Y}\ \cup\ \blacktriangle,\ \mathbb{Y}\ )$$

$$\Psi_R(c,g,s,p,\mathbf{y},\mathbf{c},a,\mathbf{o},\mathbf{i},\mathbf{x},\varsigma) \equiv \begin{cases}(\texttt{?8K},[]) & \text{if } s \notin \mathrm{K}(\delta)\\ (\texttt{?\^{}W},[]) & \text{otherwise if } (\delta[s],\mathbf{c}_t,c) > \mathrm{S}\\ (\mathbf{r},(\mathbf{m},\mathbf{e})) & \text{otherwise}\end{cases} \quad \text{(253)}$$

$$\text{let } a = \mathrm{E}(s,\mathbf{y},p,\mathbf{c},a,\mathbf{o},\updownarrow\mathbf{x}),$$
$$\text{and } (\mathbf{r},(\mathbf{m},\mathbf{e})) = \Psi_M(\ldots(\delta[s],\mathbf{c}_t,c),1,g,a,F,(\ldots,[]))$$

$$\begin{cases}(\mathbf{r},[]) & \text{if } \mathbf{r} \in \{\infty,\ldots\}\\ (\mathbf{u},\mathbf{e}) & \text{if } \mathbf{r} = g,\mathbf{u}\end{cases}$$

$$F(n,\xi,\omega,\mu,(\mathbf{m},\mathbf{e})) \equiv \begin{cases}\Omega_H(\xi,\omega,\mu,(\mathbf{m},\mathbf{e}),s,\delta,\mathbf{c}_t) & \text{if } n = \texttt{lookup}\\ \Omega_Y(\xi,\omega,\mu,(\mathbf{m},\mathbf{e}),\mathbf{i}) & \text{if } n = \texttt{import}\\ \Omega_Z(\xi,\omega,\mu,(\mathbf{m},\mathbf{e}),\varsigma) & \text{if } n = \texttt{export}\\ \Omega_G(\xi,\omega,\mu,(\mathbf{m},\mathbf{e})) & \text{if } n = \texttt{gas}\\ \Omega_M(\xi,\omega,\mu,(\mathbf{m},\mathbf{e})) & \text{if } n = \texttt{machine}\\ \Omega_P(\xi,\omega,\mu,(\mathbf{m},\mathbf{e})) & \text{if } n = \texttt{peek}\\ \Omega_O(\xi,\omega,\mu,(\mathbf{m},\mathbf{e})) & \text{if } n = \texttt{poke}\\ \Omega_K(\xi,\omega,\mu,(\mathbf{m},\mathbf{e})) & \text{if } n = \texttt{invoke}\\ \Omega_X(\xi,\omega,\mu,(\mathbf{m},\mathbf{e})) & \text{if } n = \texttt{expunge}\\ (\xi-10,[\texttt{WHAT},\omega_1,\ldots],\mu) & \text{otherwise}\end{cases} \quad \text{(254)}$$

B.4. **Accumulate Invocation.** Since this is a transition which can directly affect a substantial amount of on-chain state, our invocation context is accordingly complex. It is a tuple with elements for each of the aspects of state which can be altered through this invocation and beyond the account of the service itself includes the deferred transfer list and several dictionaries for alterations to preimage lookup state, core assignments, validator key assignments, newly created accounts and alterations to account privilege levels.

Formally, we define our result context to be $\mathbf{X}$, and our invocation context to be a pair of these contexts, $\mathbf{X} \times \mathbf{X}$, with one dimension being the regular dimension and generally named $\mathbf{x}$ and the other being the exceptional dimension and being named $\mathbf{y}$. The only function which actually alters this second dimension is `checkpoint`, Ω_C and so it is rarely seen.

We track both regular and exceptional dimensions within our context mutator, but collapse the result of the invocation to one or the other depending on whether the termination was regular or exceptional (i.e. out-of-gas or panic).

$$\mathbf{X} \equiv \begin{matrix}\mathbf{s} & \mathbb{A}?, & \mathbf{c} & \mathbb{H}_{\mathrm{Q}\,\mathrm{C}}, & \mathbf{v} & \mathbb{K}_{\mathrm{V}}, & i & \mathbb{N}_S,\\ \mathbf{t} & \mathbb{T}\ , & \mathbf{n} & \mathbb{D}\,\mathbb{N}_S \to \mathbb{A}\ , & p & m & \mathbb{N}_S, a & \mathbb{N}_S, v & \mathbb{N}_S\end{matrix} \quad \text{(255)}$$

We define Ψ_A, the Accumulation invocation function as:

$$\Psi_A(\delta^\dagger,s,g,\mathbf{o}) \equiv \begin{cases}\delta^\dagger[s] & \text{if } \delta^\dagger[s]_{\mathbf{c}} = \mathbf{o} = []\\ C(\Psi_M(\delta^\dagger[s]_{\mathbf{c}},2,g,\mathrm{E}(\mathbf{o}),F,I(\delta^\dagger[s],s))) & \text{otherwise}\end{cases} \quad \text{(256)}$$

$$(\mathbb{D}\,\mathbb{N}_S \to \mathbb{A}\ ,\mathbb{N}_S,\mathbb{N}_G,\ \mathbb{O}\ ) \to \mathbf{X} \times r\ \mathbb{H}?$$

$$I(\mathbf{a} \in \mathbb{A}, s \in \mathbb{N}_S) \equiv (\mathbf{x},\mathbf{y}) \text{ where } \begin{matrix}\mathbf{x} = \mathbf{y} = \mathbf{s}\ \mathbf{a}, \mathbf{t}\ [], i, p\ \chi, \mathbf{c}\ \varphi, \mathbf{v}\ \iota, \mathbf{n}\ \ldots,\\ i = \mathrm{check}((\mathrm{E}_4^{-1}(\mathrm{H}(s,\eta_0',\mathbf{H}_t)) \bmod (2^{32}-2^9)) + 2^8)\end{matrix} \quad \text{(257)}$$

$$F(n, \xi, \omega, \mu, (\mathbf{x}, \mathbf{y})) \equiv \begin{cases} G(\ _R(\xi, \omega, \mu, \mathbf{x}_\mathbf{s}, s, \delta^\dagger), (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{read} \\ G(\ _W(\xi, \omega, \mu, \mathbf{x}_\mathbf{s}), (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{write} \\ G(\ _L(\xi, \omega, \mu, s, \delta^\dagger), (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{lookup} \\ G(\ _G(\xi, \omega, \mu), (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{gas} \\ G(\ _I(\xi, \omega, \mu, \mathbf{x}_\mathbf{s}, s, \delta^\dagger), (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{info} \\ \ _E(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{empower} \\ \ _A(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{assign} \\ \ _D(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{designate} \\ \ _C(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{checkpoint} \\ \ _N(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y})) & \text{if } n = \texttt{new} \\ \ _U(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), s) & \text{if } n = \texttt{upgrade} \\ \ _T(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), s, \delta^\dagger) & \text{if } n = \texttt{transfer} \\ \ _Q(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), s) & \text{if } n = \texttt{quit} \\ \ _S(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), \mathbf{H}_t) & \text{if } n = \texttt{solicit} \\ \ _F(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), \mathbf{H}_t) & \text{if } n = \texttt{forget} \\ (\xi - 10, [\texttt{WHAT}, \omega_1, \dots], \mu, \mathbf{x}) & \text{otherwise} \end{cases} \tag{258}$$

$$G((\xi', \omega', \mu', \mathbf{s}'), (\mathbf{x}, \mathbf{y})) \equiv (\xi', \omega', \mu', (\mathbf{x}, \mathbf{y})) \text{ where } \mathbf{x} = \mathbf{x}' \text{ except } \mathbf{x}_\mathbf{s} = \mathbf{s}' \tag{259}$$

$$C(\mathbf{o} \in \mathbb{Y} \cup \{\infty, \}, (\mathbf{x} \in \mathbf{X}, \mathbf{y} \in \mathbf{X})) \equiv \begin{cases} \mathbf{x} \times r \subseteq \mathbf{o} & \text{if } \mathbf{o} \in \mathbb{H} \\ \mathbf{x} \times r \subseteq \mathbf{o} & \text{if } \mathbf{o} \in \mathbb{Y} \setminus \mathbb{H} \\ \mathbf{y} \times r \subseteq \mathbf{o} & \text{if } \mathbf{o} \in \{\infty, \} \end{cases} \tag{260}$$

The mutator F governs how this context will alter for any given parameterization, and the collapse function C selects one of the two dimensions of context depending on whether the virtual machine's halt was regular or exceptional.

The initializer function I maps some service account $\mathbf{s}$ along with its index s to yield a mutator context such that no alterations to state are implied (beyond those already inherent in $\mathbf{s}$) in either exit scenario. Note that the component a utilizes the random accumulator η_0 and the block's timeslot $\mathbf{H}_t$ to create a deterministic sequence of identifiers which are extremely likely to be unique.

Concretely, we create the identifier from the Blake2 hash of the identifier of the creating service, the current random accumulator η_0 and the block's timeslot. Thus, within a service's accumulation it is almost certainly unique, but it is not necessarily unique across all services, nor at all times in the past. We utilize a *check* function to find the first such index in this sequence which does not already represent a service:

$$\text{check}(i \in \mathbb{N}_S) \equiv \begin{cases} i & \text{if } i \notin \text{K}(\delta^\dagger) \\ \text{check}((i - 2^8 + 1) \bmod (2^{32} - 2^9) + 2^8) & \text{otherwise} \end{cases} \tag{261}$$

NB In the highly unlikely event that a block executes to find that a single service index has inadvertently been attached to two different services, then the block is considered invalid. Since no service can predict the identifier sequence ahead of time, they cannot intentionally disadvantage the block author.

B.5. **On-Transfer Invocation.** We define the On-Transfer service-account invocation function as Ψ_T; it is somewhat similar to the Accumulation Invocation except that the only state alteration it facilitates are basic alteration to the storage of the subject account. No further transfers may be made, no privileged operations are possible, no new accounts may be created nor other operations done on the subject account itself. The function is defined as:

$$\Psi_T(\delta^\ddagger, s, \mathbf{t}) \in (\mathbb{D}\langle \mathbb{N}_S \to \mathbb{A}, \mathbb{N}_S, \mathbb{T} \rangle) \to \mathbb{A} \tag{262}$$

$$\Psi_T(\delta^\ddagger, s, \mathbf{t}) \equiv \begin{cases} \mathbf{s} & \text{if } \mathbf{s}_\mathbf{c} = \emptyset \wedge \mathbf{t} = [] \\ \Psi_M(\mathbf{s}_\mathbf{c}, 3, \sum_{r \in \mathbf{t}}(r_g), \text{E}(\mathbf{t}), F, \mathbf{s}) & \text{otherwise} \end{cases}$$

$$\text{where } \mathbf{s} = \delta^\ddagger[s] \text{ except } \mathbf{s}_b = \delta^\ddagger[s]_b + \sum_{r \in \mathbf{t}} r_a \tag{263}$$

$$F(n, \xi, \omega, \mu, \mathbf{s}) \equiv \begin{cases} \ _L(\xi, \omega, \mu, \mathbf{s}, s, \delta^\ddagger) & \text{if } n = \texttt{lookup} \\ \ _R(\xi, \omega, \mu, \mathbf{s}, s, \delta^\ddagger) & \text{if } n = \texttt{read} \\ \ _W(\xi, \omega, \mu, \mathbf{s}) & \text{if } n = \texttt{write} \\ \ _G(\xi, \omega, \mu) & \text{if } n = \texttt{gas} \\ \ _I(\xi, \omega, \mu, \mathbf{s}, s, \delta^\ddagger) & \text{if } n = \texttt{info} \\ (\xi - 10, [\texttt{WHAT}, \omega_1, \dots], \mu, \mathbf{s}) & \text{otherwise} \end{cases} \tag{264}$$

NB When considering the mutator functions Ω_R and Ω_I, the final arguments passed are both the post-accumulation accounts state, $\delta^\ddagger$. Within the function, this parameter however is denoted simply $\mathbf{d}$. This is intentional and avoids potential confusion since the functions are also utilized for the Accumulation Invocation where the argument is $\delta^\dagger$.

B.6. **General Functions.** This defines a number of functions broadly of the form $(\xi', \mathbb{Z}_G, \omega', \mathbb{N}_{R\ 2}, \mu', \mathbf{s}') = (\xi, \mathbb{N}_G, \omega, \mathbb{N}_{R\ 6}, \mu, \mathbb{M}, \mathbf{s}, \mathbf{A}, \dots)$. Functions which have a result component which is equivalent to the corresponding argument may have said components elided in the description. Functions may also depend upon particular additional parameters.

Unlike the Accumulate functions in appendix B.7, these do not mutate an accumulation context, but merely a service account $\mathbf{s}$.

The gas function, Ω_G has a parameter list suffixed with an ellipsis to denote that any additional parameters may be taken and are provided transparently into its result. This allows it to be easily utilized in multiple PVM invocations.

(265) $$\xi' \equiv \xi - g$$

(266) $$(\omega', \mu', \mathbf{s}') \equiv \begin{cases} (\omega, \mu, \mathbf{s}) & \text{if } \xi < g \\ (\omega, \mu, \mathbf{s}) \text{ except as indicated below} & \text{otherwise} \end{cases}$$

S« jH§_nj ^L0j§_»0} W9 « 9X0	g «§9§_nj
$\Omega_G(\xi, \omega, \dots)$ gas = 0 $g = 10$	$\omega'_0 \equiv \xi' \bmod 2^{32}$ $\omega'_1 \equiv \xi' \div 2^{32}$
$\Omega_L(\xi, \omega, \mu, \mathbf{s}, s, \mathbf{d})$ lookup = 1 $g = 10$	let $\mathbf{a} = \begin{cases} \mathbf{s} & \text{if } \omega_0 \quad \{s, 2^{32}-1\} \\ \mathbf{d}[\omega_0] & \text{otherwise} \end{cases}$ let $[h_o, b_o, b_z] = \omega_{1::4}$ let $h = \begin{cases} \mathrm{H}(\mu_{h_o\ +32}) & \text{if } \mathbb{Z}_{h_o\ +32} \quad \mathbb{V} \\ & \text{otherwise} \end{cases}$ let $\mathbf{v} = \begin{cases} \mathbf{a_p}[h] & \text{if } \mathbf{a} \neq \quad h \quad \mathrm{K}(\mathbf{a_p}) \\ & \text{otherwise} \end{cases}$ $i \quad \mathbb{N}_{\min(b_z;\, \mathbf{v})} \quad \mu'_{b_o+i} \equiv \begin{cases} \mathbf{v}_i & \text{if } \mathbf{v} \neq \quad \mathbb{Z}_{b_o\ +b_z} \quad \mathbb{V} \\ \mu_{b_o+i} & \text{otherwise} \end{cases}$ $\omega'_0 \equiv \begin{cases} \text{NONE} & \text{if } \mathbf{v} = \\ \mathbf{v} & \text{otherwise} \} \text{ if } k \neq \quad \mathbb{Z}_{b_o\ +b_z} \quad \mathbb{V} \\ \text{OOB} & \text{otherwise} \end{cases}$
$\Omega_R(\xi, \omega, \mu, \mathbf{s}, s, \mathbf{d})$ read = 2 $g = 10$	let $\mathbf{a} = \begin{cases} \mathbf{s} & \text{if } \omega_0 \quad \{s, 2^{32}-1\} \\ \mathbf{d}[\omega_0] & \text{otherwise if } \omega_0 \quad \mathrm{K}(\mathbf{d}) \\ & \text{otherwise} \end{cases}$ let $[k_o, k_z, b_o, b_z] = \omega_{1::5}$ let $k = \begin{cases} \mathrm{H}(\mathrm{E}_4(s) \quad \mu_{k_o\ +k_z}) & \text{if } \mathbb{Z}_{k_o\ +k_z} \quad \mathbb{V} \\ & \text{otherwise} \end{cases}$ let $\mathbf{v} = \begin{cases} \mathbf{a_s}[k] & \text{if } \mathbf{a} \neq \quad k \quad \mathrm{K}(\mathbf{a_s}) \\ & \text{otherwise} \end{cases}$ $i \quad \mathbb{N}_{\min(b_z;\, \mathbf{v})} \quad \mu'_{b_o+i} \equiv \begin{cases} \mathbf{v}_i & \text{if } \mathbf{v} \neq \quad \mathbb{Z}_{b_o\ +b_z} \quad \mathbb{V} \\ \mu_{b_o+i} & \text{otherwise} \end{cases}$ $\omega'_0 \equiv \begin{cases} \text{NONE} & \text{if } \mathbf{v} = \\ \mathbf{v} & \text{otherwise} \} \text{ if } k \neq \quad \mathbb{Z}_{b_o\ +b_z} \quad \mathbb{V} \\ \text{OOB} & \text{otherwise} \end{cases}$

S« jH§_nj ^LOj§_»0} W9 « 9X0	g «§9§_nj
$\Psi_W(\xi, \omega, \mu, \mathbf{s})$ write $= 3$ $g = 10$	let $[k_o, k_z, v_o, v_z] = \omega_{0::4}$ let $k = \begin{cases} \mathrm{H}(\mathrm{E}_4(s) \quad \mu_{k_o \ +k_z}) & \text{if } \mathbb{Z}_{k_o \ +k_z} \quad \mathbb{V} \\ \text{otherwise} \end{cases}$ let $\mathbf{a} = \begin{cases} \mathbf{s} \text{ except } \begin{cases} \mathrm{K}(\mathbf{a_s}) = \mathrm{K}(\mathbf{a_s}) \quad \{k\} & \text{if } v_z = 0 \\ \mathbf{a_s}[k] = \mu_{v_o \ +v_z} & \text{otherwise} \end{cases} & \text{if } \mathbb{Z}_{v_o \ +v_z} \quad \mathbb{V} \\ & \text{otherwise} \end{cases}$ let $l = \begin{cases} \mathbf{s_s}[k] & \text{if } k \quad \mathrm{K}(\mathbf{s_s}) \\ \mathtt{NONE} & \text{otherwise} \end{cases}$ $(\omega_0', \mathbf{s}') \equiv \begin{cases} (l, \mathbf{a}) & \text{if } k \neq \quad \mathbf{a} \neq \quad \mathbf{a}_t \leq \mathbf{a}_b \\ (\mathtt{FULL}, \mathbf{s}) & \text{if } \mathbf{a}_t > \mathbf{a}_b \\ (\mathtt{OOB}, \mathbf{s}) & \text{otherwise} \end{cases}$
$\Psi_I(\xi, \omega, \mu, \mathbf{s}, s, \mathbf{d})$ info $= 4$ $g = 10$	let $\mathbf{t} = \begin{cases} \mathbf{s} & \text{if } \omega_0 \quad \{s, 2^{32} - 1\} \\ (\mathbf{d} \quad \mathbf{x_n})[\omega_0] & \text{otherwise} \end{cases}$ let $o = \omega_1$ let $\mathbf{m} = \begin{cases} \mathrm{E}(\mathbf{t}_c, \mathbf{t}_b, \mathbf{t}_t, \mathbf{t}_g, \mathbf{t}_m, \mathbf{t}_l, \mathbf{t}_i) & \text{if } \mathbf{t} \neq \\ & \text{otherwise} \end{cases}$ $i \quad \mathbb{N}_{\mathbf{m}} \quad \mu'_{o+i} \equiv \begin{cases} \mathbf{m}_i & \text{if } \mathbf{m} \neq \quad \mathbb{Z}_{o \ +\mathbf{m}} \quad \mathbb{V} \\ \mu_{b_o+i} & \text{otherwise} \end{cases}$ $\omega_0' \equiv \begin{cases} \mathtt{OK} & \text{if } \mathbf{m} \neq \quad \mathbb{Z}_{o \ +\mathbf{m}} \quad \mathbb{V} \\ \mathtt{NONE} & \text{if } \mathbf{m} = \\ \mathtt{OOB} & \text{otherwise} \end{cases}$

B.7. **Accumulate Functions.** This defines a number of functions broadly of the form $(\xi' \quad \mathbb{Z}_G, \omega' \quad \mathbb{N}_{R \ 2}, \mu', (\mathbf{x}', \mathbf{y}')) = (\xi \quad \mathbb{N}_G, \omega \quad \mathbb{N}_{R \ 6}, \mu \quad \mathbb{M}, (\mathbf{x} \quad \mathbf{X}, \mathbf{y} \quad \mathbf{X}), \ldots)$. Functions which have a result component which is equivalent to the corresponding argument may have said components elided in the description. Functions may also depend upon particular additional parameters.

(267) $$\xi' \equiv \xi - g$$

(268) $$(\omega', \mu', \mathbf{x}', \mathbf{y}') \equiv \begin{cases} (\omega, \mu, \mathbf{x}, \mathbf{y}) & \text{if } \xi < g \\ (\omega, \mu, \mathbf{x}, \mathbf{y}) \text{ except as indicated below} & \text{otherwise} \end{cases}$$

S« jH§_nj ^LOj§_»0} W9 « 9X0	g «§9§_nj
$\Psi_E(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}))$ empower $= 5$ $g = 10$	$(\mathbf{x}'_p)_m = \omega_0$ $(\mathbf{x}'_p)_a = \omega_1$ $(\mathbf{x}'_p)_v = \omega_2$
$\Psi_A(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}))$ assign $g = 10$	let $o = \omega_1$ let $\mathbf{c} = \begin{cases} [\mu_{o+32i \ +32} \quad i \Subset \mathbb{N}_{\mathrm{Q}}] & \text{if } \mathbb{Z}_{o \ +32\mathrm{Q}} \quad \mathbb{V} \\ & \text{otherwise} \end{cases}$ $(\omega_0', \mathbf{x}') = \begin{cases} (\mathtt{OK}, \mathbf{x} \text{ except } \mathbf{x}'_{\mathbf{c}}[\omega_0] = \mathbf{c}) & \text{if } \omega_0 < \mathrm{C} \quad \mathbf{c} \neq \\ (\mathtt{OOB}, \mathbf{x}) & \text{if } \mathbf{c} = \\ (\mathtt{CORE}, \mathbf{x}) & \text{otherwise} \end{cases}$

S« jH§_n j ^LOj§_»O} W9 « 9XO	g «§9§_n j
$_D(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}))$ designate $= 6$ $g = 10$	let $o = \omega_0$ let $\mathbf{v} = \begin{cases} [\mu_{o+176i \ +176} \quad i \in \mathbb{N}_{\mathbf{V}}] & \text{if } \mathbb{Z}_{o \ +176\mathbf{V}} \quad \mathbb{V} \\ \text{otherwise} \end{cases}$ $(\omega_0', \mathbf{x}') = \begin{cases} (\mathsf{OK}, \mathbf{x} \text{ except } \mathbf{x}'_{\mathbf{v}} = \mathbf{v}) & \text{if } \mathbf{v} \neq \\ (\mathsf{OOB}, \mathbf{x}) & \text{otherwise} \end{cases}$
$_C(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}))$ checkpoint $= 7$ $g = 10$	$\mathbf{y}' \equiv \mathbf{x}$ $\omega_0' \equiv \xi' \bmod 2^{32}$ $\omega_1' \equiv \ \xi' \div 2^{32}$
$_N(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}))$ new $g = 10$	let $[o, l, g_l, g_h, m_l, m_h] = \omega_{0::6}$ let $c = \begin{cases} \mu_{o \ +32} & \text{if } \mathbb{N}_{o \ +32} \quad \mathbb{V} \\ \text{otherwise} \end{cases}$ let $g = 2^{32} \ g_h + g_l$ let $m = 2^{32} \ m_h + m_l$ let $\mathbf{a} \ \mathbb{A} \ \{ \ \} = \begin{cases} (c, \mathbf{s} \ \{\}, \mathbf{l} \ \{(c, l) \quad []\}, b \ \mathbf{a}_t, g, m) & \text{if } c \neq \\ & \text{otherwise} \end{cases}$ let $b = (\mathbf{x_s})_b - \mathbf{a}_t$ $(\omega_0', \mathbf{x}_i', \mathbf{x}_{\mathbf{n}}', (\mathbf{x}_{\mathbf{s}}')_b) \equiv \begin{cases} (\mathbf{x}_i, \text{check}(\text{bump}(\mathbf{x}_i)), \mathbf{x_n} \quad \{\mathbf{x}_i \quad \mathbf{a}\}, b) & \text{if } \mathbf{a} \neq \quad b \geq (\mathbf{x_s})_t \\ (\mathsf{OOB}, \mathbf{x}_T) & \text{if } c = \\ (\mathsf{CASH}, \mathbf{x}_T) & \text{otherwise} \end{cases}$ where $\text{bump}(i \quad \mathbb{N}_S) = 2^8 + (i - 2^8 + 42) \bmod (2^{32} - 2^9)$
$_U(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), s)$ upgrade $= 8$ $g = 10$	let $[o, g_h, g_l, m_h, m_l] = \omega_{0::5}$ let $c = \begin{cases} \mu_{o \ +32} & \text{if } \mathbb{N}_{o \ +32} \quad \mathbb{V} \\ \text{otherwise} \end{cases}$ let $g = 2^{32} \ g_h + g_l$ let $m = 2^{32} \ m_h + m_l$ $(\omega_0', \mathbf{x}'[s]_c, \mathbf{x}'[s]_g, \mathbf{x}'[s]_m) \equiv \begin{cases} (\mathsf{OK}, c, g, m) & \text{if } c \neq \\ (\mathsf{OOB}, \mathbf{x}[s]_c, \mathbf{x}[s]_g, \mathbf{x}[s]_m) & \text{otherwise} \end{cases}$
$_T(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), s, \delta)$ transfer $= 9$ $g = 10 + \omega_1 + 2^{32} \ \omega_2$	let $(d, a_l, a_h, g_l, g_h, o) = \omega_{0::6}$, let $a = 2^{32} \ a_h + a_l$ let $g = 2^{32} \ g_h + g_l$ let $\mathbf{t} \ \mathbb{T} \ \{ \ \} = \begin{cases} (s, d, a, m, g) \quad m = \mathrm{E}^{-1}(\mu_{o \ +\mathrm{M}}) & \text{if } \mathbb{N}_{o \ +\mathrm{M}} \quad \mathbb{V} \\ \text{otherwise} \end{cases}$ let $b = (\mathbf{x_s})_b - a$ $(\omega_0', \mathbf{x}_{\mathbf{t}}', (\mathbf{x}_{\mathbf{s}}')_b) \equiv \begin{cases} (\mathsf{OOB}, \mathbf{x_t}, (\mathbf{x_s})_b) & \text{if } t = \\ (\mathsf{WHO}, \mathbf{x_t}, (\mathbf{x_s})_b) & \text{otherwise if } d \ / \ \mathrm{K}(\delta \quad \mathbf{x_n}) \\ (\mathsf{LOW}, \mathbf{x_t}, (\mathbf{x_s})_b) & \text{otherwise if } g < (\delta \quad \mathbf{x_n})[d]_m \\ (\mathsf{HIGH}, \mathbf{x_t}, (\mathbf{x_s})_b) & \text{otherwise if } \xi < g \\ (\mathsf{CASH}, \mathbf{x_t}, (\mathbf{x_s})_b) & \text{otherwise if } b < (\mathbf{x_s})_t \\ (\mathsf{OK}, \mathbf{x_t} \mathbin{+\!\!+} \mathbf{t}, b) & \text{otherwise} \end{cases}$

S« jH§_nj ^LOj§_»O} W9 « 9XO	g «§9§_nj
$_X(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), s)$ quit $= 10$ $g = 10 + \omega_1 + 2^{32} \quad \omega_2$	let $[d, o] = \omega_{0;1}$ let $a = (\mathbf{x_s})_b - (\mathbf{x_s})_t + \mathrm{B}_S$ let $g = \xi$ let t $\mathbb{T}$ { , } = (s, d, a, m, g) $m = \mathrm{E}^{-1}(\mu_{o\ +\mathrm{M}})$ if $d \quad \{s, 2^{32}-1\}$; otherwise if $\mathbb{N}_{o\ +\mathrm{M}} \quad \mathbb{V}$; otherwise $(\omega_0', \mathbf{x_s}', \mathbf{x_t}') \equiv$ (OK, , $\mathbf{x_t}$), **virtual machine halts** if t = (OOB, $\mathbf{x_t}$, $(\mathbf{x_s})_b$) otherwise if $t =$ (WHO, $\mathbf{x_t}$, $(\mathbf{x_s})_b$) otherwise if d / K($\delta \quad \mathbf{x_n}$) (LOW, $\mathbf{x_t}$, $(\mathbf{x_s})_b$) otherwise if $g < (\delta \quad \mathbf{x_n})[d]_m$ (OK, , $\mathbf{x_t} +\!\!+ \mathrm{t}$), **virtual machine halts** otherwise
$_S(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}))$ solicit $= 11$ $g = 10$	let $[o, z] = \omega_{0;1}$ let $h =$ $\mu_{o\ +32}$ if $\mathbb{Z}_{o\ +32} \quad \mathbb{V}$; otherwise let a = $\mathbf{x_s}$ except: $\mathbf{a_l}[\ h, z\ ] = []$ if $h \neq \quad (h, z) \ / \ (\mathbf{x_s})_\mathbf{l}$ $\mathbf{a_l}[\ h, z\ ] = (\mathbf{x_s})_\mathbf{l}[\ h, z\ ] +\!\!+ t$ if $(\mathbf{x_s})_\mathbf{l}[\ h, z\ ] = [x, y]$ otherwise $(\omega_0', \mathbf{x_s}') \equiv$ (OOB, $\mathbf{x_s}$) if $h =$ (HUH, $\mathbf{x_s}$) otherwise if a = (FULL, $\mathbf{x_s}$) otherwise if $\mathbf{a}_b < \mathbf{a}_t$ (OK, a) otherwise
$_F(\xi, \omega, \mu, (\mathbf{x}, \mathbf{y}), t)$ forget $= 12$ $g = 10$	let $[o, z] = \omega_{0;1}$ let $h =$ $\mu_{o\ +32}$ if $\mathbb{Z}_{o\ +32} \quad \mathbb{V}$; otherwise let a = $\mathbf{x_s}$ except: $\mathrm{K}(\mathbf{a_l}) = \mathrm{K}((\mathbf{x_s})_\mathbf{l}) \quad \{\ h, z\ \}$, $\mathrm{K}(\mathbf{a_p}) = \mathrm{K}((\mathbf{x_s})_\mathbf{p}) \quad \{h\}$ } if $(\mathbf{x_s})_\mathbf{l}[h, z] \quad \{[], [x, y]\},\ y < t - \mathrm{D}$ $\mathbf{a_l}[h, z] = (\mathbf{x_s})_\mathbf{l}[h, z] +\!\!+ t$ if $(\mathbf{x_s})_\mathbf{l}[h, z] \quad = 1$ $\mathbf{a_l}[h, z] = [(\mathbf{x_s})_\mathbf{l}[h, z]_2, t]$ if $(\mathbf{x_s})_\mathbf{l}[h, z] = [x, y, w],\ y < t - \mathrm{D}$ otherwise $(\omega_0', \mathbf{x_s}') \equiv$ (OOB, $\mathbf{x_s}$) if $h =$ (HUH, $\mathbf{x_s}$) otherwise if a = (OK, a) otherwise

B.8. **Refine Functions.** These assume some refine context pair $(\mathbf{m}, \mathbf{e}) \quad (\mathbb{D}\ \mathbb{N} \to \mathbf{M}\ ,\ \mathbb{Y}_{\mathrm{W}_S}\ )$, which are both initially empty.

S« jH§_nj ^LOj§_»O} W9 « 9XO	g «§9§_nj
$\Psi_H(\xi,\omega,\mu,(\mathbf{m},\mathbf{e}),s,\delta,t)$ historical_lookup $= 13$ $g = 10$	$\text{let } \mathbf{a} = \begin{cases} \delta[s] & \text{if } \omega_0 = 2^{32}-1 \quad s \quad \mathrm{K}(\delta) \\ \delta[\omega_0] & \text{if } \omega_0 \quad \mathrm{K}(\delta) \\ & \text{otherwise} \end{cases}$ $\text{let } [h_o, b_o, b_z] = \omega_{1::4}$ $\text{let } h = \begin{cases} \mathrm{H}(\mu_{h_o \ +32}) & \text{if } \mathbb{Z}_{h_o \ +32} \quad \mathbb{V} \\ & \text{otherwise} \end{cases}$ $\text{let } \mathbf{v} = H(\mathbf{a}, t, h)$ $i \quad \mathbb{N}_{\min(b_z;\mathbf{v})} \quad \mu'_{b_o+i} \equiv \begin{cases} \mathbf{v}_i & \text{if } \mathbf{v} \neq \quad \mathbb{Z}_{b_o \ +b_z} \quad \mathbb{V} \\ \mu_{b_o+i} & \text{otherwise} \end{cases}$ $\omega'_0 \equiv \begin{cases} \mathbf{v} & \text{if } \mathbf{v} \neq \\ \texttt{NONE} & \text{otherwise} \\ \texttt{OOB} & \end{cases} \Big\} \text{ if } k \neq \quad \mathbb{Z}_{b_o \ +b_z} \quad \mathbb{V}$ otherwise
$\Psi_Y(\xi,\omega,\mu,(\mathbf{m},\mathbf{e}),\mathbf{i})$ import $= 14$ $g = 10$	$\text{let } \mathbf{v} = \begin{cases} \mathbf{i}_{\ l\ 0} & \text{if } \omega_0 < \mathbf{i} \\ & \text{otherwise} \end{cases}$ $\text{let } o = \omega_1$ $\text{let } l = \min(\omega_2, \mathrm{W}_C \mathrm{W}_S)$ $\mu'_{o \ +l} \equiv \begin{cases} \mathbf{v} & \text{if } \mathbf{v} \neq \quad \mathbb{N}_{o \ +l} \quad \mathbb{V} \\ \mu_{o \ +l} & \text{otherwise} \end{cases}$ $\omega'_0 \equiv \begin{cases} \texttt{OOB} & \text{if } \mathbb{Z}_{o \ +l} \ / \ \mathbb{V} \\ \texttt{NONE} & \text{otherwise if } \mathbf{v} = \\ \texttt{OK} & \text{otherwise} \end{cases}$
$\Psi_Z(\xi,\omega,\mu,(\mathbf{m},\mathbf{e}),\varsigma)$ export $= 15$ $g = 10$	$\text{let } p = \omega_0$ $\text{let } z = \min(\omega_1, \mathrm{W}_C \mathrm{W}_S)$ $\text{let } \mathbf{x} = \begin{cases} \mathrm{P}_{\mathrm{W}_C \mathrm{W}_S}(\mu_{p \ +z}) & \text{if } \mathbb{N}_{p \ +z} \quad \mathbb{V} \\ & \text{otherwise} \end{cases}$ $(\omega'_0, \mathbf{e}') \equiv \begin{cases} (\texttt{OOB}, \mathbf{e}) & \text{if } \mathbf{x} = \\ (\texttt{FULL}, \mathbf{e}) & \text{otherwise if } \varsigma + \ \mathbf{e} \geq \mathrm{W}_X \\ (\varsigma + \ \mathbf{e}, \mathbf{e} +\!\!+ \mathbf{x}) & \text{otherwise} \end{cases}$
$\Psi_M(\xi,\omega,\mu,(\mathbf{m},\mathbf{e}))$ machine $= 16$ $g = 10$	$\text{let } [p_o, p_z, i] = \omega_{0:::3}$ $\text{let } \mathbf{p} = \begin{cases} \mu_{p_o \ +p_z} & \text{if } \mathbb{Z}_{p_o \ +p_z} \quad \mathbb{V} \\ & \text{otherwise} \end{cases}$ $\text{let } n = \min(n \quad \mathbb{N}, n \ / \ \mathrm{K}(\mathbf{m}))$ $\text{let } \mathbf{u} = \mathrm{V} \ [0,0,\dots], \mathrm{A} \ [\ ,\ ,\dots]$ $(\omega'_0, \mathbf{m}) \equiv \begin{cases} (\texttt{OOB}, \mathbf{m}) & \text{if } \mathbf{p} = \\ (n, \mathbf{m} \quad \{n \quad \mathbf{p}, \mathbf{u}, i\ \}) & \text{otherwise} \end{cases}$
$\Psi_P(\xi,\omega,\mu,(\mathbf{m},\mathbf{e}))$ peek $= 17$ $g = 10$	$\text{let } [n, a, b, l] = \omega_{0:::4}$ $\text{let } \mathbf{s} = \begin{cases} & \text{if } n \ / \ \mathrm{K}(\mathbf{m}) \\ & \text{if } \mathbb{N}_{b \ +i} \ / \ \mathbb{V}_{\mathbf{m}[n]_{\mathbf{u}}} \\ \mathbf{m}[n]_{\mathbf{u}_{b \ +i}} & \text{otherwise} \end{cases}$ $(\omega'_0, \mu') \equiv \begin{cases} (\texttt{OOB}, \mu) & \text{if } \mathbf{s} = \\ (\texttt{WHO}, \mu) & \text{if } \mathbf{s} = \\ (\texttt{OK}, \mu') \text{ where } \mu' = \mu \text{ except } \mu'_{a \ +l} = \mathbf{s} & \text{otherwise} \end{cases}$

S« jH§_nj ^LOj§_»O} W9 « 9XO	g «§9§_nj
$\Omega_{poke}(\xi, \omega, \mu, (\mathbf{m}, \mathbf{e}))$ poke = 18 $g = 10$	let $[n, a, b, l] = \omega_{0:::4}$ let $\mathbf{u} = \begin{cases} \mathbf{m}[n]_{\mathbf{u}} & \text{if } n \quad \mathrm{K}(\mathbf{m}) \\ & \text{otherwise} \end{cases}$ let $\mathbf{s} = \begin{cases} \mu_{a \ +i} & \text{if } \mathbb{N}_{a \ +i} \quad \mathbb{V}_{\mathbf{u}} \\ & \text{otherwise} \end{cases}$ let $\mathbf{u}' = \mathbf{u}$ except $\begin{cases} (\mathbf{u}'_{\mathbf{V}})_{b \ +l} = \mathbf{s} \\ (\mathbf{u}'_{\mathbf{A}})_{b \ +l} = [\mathrm{W}, \mathrm{W}, \ldots] \end{cases}$ $(\omega'_0, \mathbf{m}') \equiv \begin{cases} (\mathsf{OOB}, \mathbf{m}) & \text{if } \mathbf{s} = \\ (\mathsf{WHO}, \mathbf{m}) & \text{otherwise if } \mathbf{u} = \\ (\mathsf{OK}, \mathbf{m}'), \text{ where } \mathbf{m}' = \mathbf{m} \text{ except } \mathbf{m}'[n]_{\mathbf{u}} = \mathbf{u}' & \text{otherwise} \end{cases}$
$\Omega_{invoke}(\xi, \omega, \mu, (\mathbf{m}, \mathbf{e}))$ invoke = 19 $g = 10$	let $[n, o] = \omega_{0:::2}$ let $(g, \mathbf{w}) = \begin{cases} (\mathrm{E}_8^{-1}(\mu_{o \ +8}), [\mathrm{E}_4^{-1}(\mu_{o+8+4x \ +4}) \quad x \in \mathbb{N}_{13}]) & \text{if } \mathbb{N}_{o \ +60} \quad \mathbb{V} \\ (\quad, \quad) & \text{otherwise} \end{cases}$ let $(c, i', g', \mathbf{w}', \mathbf{u}') = (\mathbf{m}[n]_{\mathbf{p}}, \mathbf{m}[n]_i, g, \mathbf{w}, \mathbf{m}[n]_{\mathbf{u}})$ let $\mu \ = \mu$ except $\mu_{o \ +60} = \mathrm{E}_8(g') \quad \mathrm{E}([\mathrm{E}_4(x) \quad x \in \mathbf{w}'])$ let $\mathbf{m} \ = \mathbf{m}$ except $\begin{cases} \mathbf{m} \ [n]_{\mathbf{u}} = \mathbf{u}' \\ \mathbf{m} \ [n]_i = \begin{cases} i' + 1 & \text{if } c \quad \{\hbar\} \times \mathbb{N}_R \\ i' & \text{otherwise} \end{cases} \end{cases}$ $(\omega'_0, \omega'_1, \mu', \mathbf{m}') \equiv \begin{cases} (\mathsf{OOB}, \omega_1, \mu, \mathbf{m}) & \text{if } g = \\ (\mathsf{WHO}, \omega_1, \mu, \mathbf{m}) & \text{otherwise if } n \ / \ \mathbf{m} \\ (\mathsf{HOST}, h, \mu \ , \mathbf{m} \ ) & \text{otherwise if } c = \hbar \times h \\ (\mathsf{FAULT}, x, \mu \ , \mathbf{m} \ ) & \text{otherwise if } c = \mathbf{S} \times x \\ (\mathsf{PANIC}, \omega_1, \mu \ , \mathbf{m} \ ) & \text{otherwise if } c = \\ (\mathsf{HALT}, \omega_1, \mu \ , \mathbf{m} \ ) & \text{otherwise if } c = \end{cases}$
$\Omega_{expunge}(\xi, \omega, \mu, (\mathbf{m}, \mathbf{e}))$ expunge = 20 $g = 10$	let $n = \omega_0$ $(\omega'_0, \mathbf{m}') \equiv \begin{cases} (\mathsf{WHO}, \mathbf{m}) & \text{if } n \neq \mathrm{K}(\mathbf{m}) \\ (\mathbf{m}[n]_i, \mathbf{m} \quad n) & \text{otherwise} \end{cases}$

Appendix C. Serialization Codec

C.1. **Common Terms.** Our codec function E is used to serialize some term into a sequence of octets. We define the deserialization function $\mathrm{E}^{-1} = \mathrm{E}^{-1}$ as the inverse of E and able to decode some sequence into the original value. The codec is designed such that exactly one value is encoded into any given sequence of octets, and in cases where this is not desirable then we use special codec functions.

C.1.1. *Trivial Encodings.* We define the serialization of as the empty sequence:

$$\mathrm{E}(\quad) \equiv [] \tag{269}$$

We also define the serialization of an octet-sequence as itself:

$$\mathrm{E}(x \quad \mathbb{Y}) \equiv x \tag{270}$$

We define anonymous tuples to be encoded as the concatenation of their encoded elements:

$$\mathrm{E}(\ a, b, \ldots\ ) \equiv \mathrm{E}(a) \quad \mathrm{E}(b) \quad \ldots \tag{271}$$

Passing multiple arguments to the serialization functions is equivalent to passing a tuple of those arguments. Formally:

$$\mathrm{E}(a, b, c, \ldots) \equiv \mathrm{E}(\ a, b, c, \ldots\ ) \tag{272}$$

C.1.2. *Integer Encoding.* We first define the trivial natural number serialization functions which are subscripted by the number of octets of the final sequence. Values are encoded in a regular little-endian fashion. Formally:

$$\mathbb{N}_{2^{8l}} \to \mathbb{Y}_l$$

$$\mathcal{E}_l\ \mathrm{N}\ x \equiv \begin{cases} [] & \text{if } l = 0 \\ [x \bmod 256]\ \ \mathcal{E}_{l-1}(\ \tfrac{x}{256}\ ) & \text{otherwise} \end{cases} \tag{273}$$

We also define the variable-size prefix 29-bit natural serialization function $\mathcal{E}_4$:

$$\mathbb{N}_{2^{29}} \to \mathbb{Y}_{1\,4}$$

$$\mathcal{E}_4\ x \equiv \begin{cases} [0] & \text{if } x = 0 \\ [2^8 - 2^{8-l} + \ \tfrac{x}{2^{8l}}\ ]\ \ \mathcal{E}_l(x \bmod 2^{8l}) & \text{if } \ l\ \ \mathbb{N}_3\ \ 2^{7l} \le x < 2^{7(l+1)} \\ [2^8 - 2^5 + \ \tfrac{x}{2^{24}}\ ]\ \ \mathcal{E}_3(x \bmod 2^{24}) & \text{if } 2^{21} \le x < 2^{29} \end{cases} \tag{274}$$

We define general natural number serialization, able to encode naturals of up to 2^{64}, as:

$$\mathbb{N}_{2^{64}} \to \mathbb{Y}_{1\,9}$$

$$\mathcal{E}\ x \equiv \begin{cases} [0] & \text{if } x = 0 \\ [2^8 - 2^{8-l} + \ \tfrac{x}{2^{8l}}\ ]\ \ \mathcal{E}_l(x \bmod 2^{8l}) & \text{if } \ l\ \ \mathbb{N}_8\ \ 2^{7l} \le x < 2^{7(l+1)} \\ [2^8 - 1]\ \ \mathcal{E}_8(x) & \text{otherwise if } x < 2^{64} \end{cases} \tag{275}$$

C.1.3. *Sequence Encoding.* We define the sequence serialization function $\mathcal{E}(\ T\ )$ for any T which is itself a subset of the domain of $\mathcal{E}$. We simply concatenate the serializations of each element in the sequence in turn:

$$\mathcal{E}([i_0, i_1, ...]) \equiv \mathcal{E}(i_0)\ \ \mathcal{E}(i_1)\ \ \ldots \tag{276}$$

Thus, conveniently, fixed length octet sequences (e.g. hashes $\mathbb{H}$ and its variants) have an identity serialization.

C.1.4. *Discriminator Encoding.* When we have sets of heterogeneous items such as a union of different kinds of tuples or sequences of different length, we require a discriminator to determine the nature of the encoded item for successful deserialization. Discriminators are encoded as a natural and are encoded immediately prior to the item.

We generally use a *length discriminator* which serializing sequence terms which have variable length (e.g. general blobs $\mathbb{Y}$ or unbound numeric sequences $\mathbb{N}$) (though this is omitted in the case of fixed-length terms such as hashes $\mathbb{H}$).[20] In this case, we simply prefix the term its length prior to encoding. Thus, for some term $y \quad x \quad \mathbb{Y}, \ldots$, we would generally define its serialized form to be $\mathcal{E}(\ x\ )\ \ \mathcal{E}(x)\ \ \ldots$. To avoid repetition of the term in such cases, we define the notation $\updownarrow x$ to mean that the term of value x is variable in size and requires a length discriminator. Formally:

$$\updownarrow x \equiv\ \ x\ , x\ \ \text{thus } \mathcal{E}(\updownarrow x) \equiv \mathcal{E}(\ x\ )\ \ \mathcal{E}(x) \tag{277}$$

We also define a convenient discriminator operator $¿x$ specifically for terms defined by some serializable set in union with (generally denoted for some set S as $S?$):

$$¿x \equiv \begin{cases} 0 & \text{if } x = \\ (1, x) & \text{otherwise} \end{cases} \tag{278}$$

C.1.5. *Bit Sequence Encoding.* A sequence of bits $b \quad \mathbb{B}$ is a special case since encoding each individual bit as an octet would be very wasteful. We instead pack the bits into octets in order of least significant to most, and arrange into an octet stream. In the case of a variable length sequence, then the length is prefixed as in the general case.

$$\mathcal{E}(b \quad \mathbb{B}) \equiv \begin{cases} [] & \text{if } b = [] \\ [\ \sum_{i=0}^{\min(8;\, b)} b_i\ \ 2^i\ ]\ \ \mathcal{E}(b_{8::}) & \text{otherwise} \end{cases} \tag{279}$$

C.2. **Block Serialization.** A block $\mathbf{B}$ is serialized as a tuple of its elements in regular order, as implied in equations 13, 14 and 37. For the header, we define both the regular serialization and the unsigned serialization $\mathcal{E}_U$. Formally:

$$\mathcal{E}(\mathbf{B}) = \mathcal{E}\Big(\begin{matrix} \mathbf{H},\ \updownarrow \mathbf{E}_T,\ \updownarrow[(r, a, [(v, \mathcal{E}_2(i), s)\ \ (v, i, s) \lessdot \mathbf{v}])\ \ (r, a, \mathbf{v}) \lessdot \mathbf{j}], \updownarrow \mathbf{c}, \updownarrow \mathbf{f}, \\ \updownarrow[\ s, \updownarrow p\ \ \ \ s, p\ \lessdot \mathbf{E}_P\ ],\ \updownarrow \mathbf{E}_A,\ \updownarrow[\ c, w, \updownarrow a\ \ \ \ c, w, a\ \lessdot \mathbf{E}_G\ ] \end{matrix}\Big) \tag{280}$$

$$\text{where } \mathbf{E}_D \equiv (\mathbf{j}, \mathbf{c}, \mathbf{f})$$

$$\mathcal{E}(\mathbf{H}) = \mathcal{E}_U(\mathbf{H})\ \ \mathcal{E}(\mathbf{H}_s) \tag{281}$$

$$\mathcal{E}_U(\mathbf{H}) = \mathcal{E}(\mathbf{H}_p, \mathbf{H}_r, \mathbf{H}_x)\ \ \mathcal{E}_4(\mathbf{H}_t)\ \ \mathcal{E}(¿\mathbf{H}_e, ¿\mathbf{H}_w, \updownarrow \mathbf{H}_j, \updownarrow \mathbf{H}_o, \mathcal{E}_2(\mathbf{H}_i), \mathbf{H}_v) \tag{282}$$

$$\mathcal{E}(x \quad \mathbb{X}) \equiv \mathcal{E}(x_a, x_s, x_b, x_l)\ \ \mathcal{E}_4(x_t)\ \ \mathcal{E}(¿x_p) \tag{283}$$

$$\mathcal{E}(x \quad \mathbb{S}) \equiv \mathcal{E}(x_h)\ \ \mathcal{E}_4(x_l)\ \ \mathcal{E}(x_u, x_e) \tag{284}$$

$$\mathcal{E}(x \quad \mathbb{L}) \equiv \mathcal{E}_4(x_s)\ \ \mathcal{E}(x_c, x_l)\ \ \mathcal{E}_8(x_g)\ \ \mathcal{E}(O(x_o)) \tag{285}$$

[20] Note that since specific values may belong to both sets which would need a discriminator and those that would not then we are sadly unable to introduce a function capable of serializing corresponding to the *term*'s limitation. A more sophisticated formalism than basic set-theory would be needed, capable of taking into account not simply the value but the term from which or to which it belongs in order to do this succinctly.

(286) $$\mathrm{E}(x \quad \mathbb{W}) \equiv \mathrm{E}(x_a, x_c, \updownarrow x_o, x_x, x_s, \updownarrow x_r)$$

(287) $$\mathrm{E}(x \quad \mathbb{P}) \equiv \mathrm{E}(\updownarrow x_{\mathbf{j}}, \mathrm{E}_4(x_h), x_c, \updownarrow x_{\mathbf{p}}, x_{\mathbf{x}}, \updownarrow x_{\mathbf{i}})$$

(288) $$\mathrm{E}(x \quad \mathbb{I}) \equiv \mathrm{E}(\mathrm{E}_4(x_s), x_c, \updownarrow x_{\mathbf{y}}, \mathrm{E}_8(x_g), \updownarrow[(h, \mathrm{E}_2(i)) \quad (h,i) \lessdot x_{\mathbf{i}}], \updownarrow[(h, \mathrm{E}_4(i)) \quad (h,i) \lessdot x_{\mathbf{x}}], \mathrm{E}_2(x_e))$$

(289) $$\mathrm{E}(x \quad \mathbb{C}) \equiv \mathrm{E}(x_{\mathbf{y}}, x_r)$$

(290) $$O(o \quad \mathbb{J} \quad \mathbb{Y}) \equiv \begin{cases} (0, \updownarrow o) & \text{if } o \quad \mathbb{Y} \\ 1 & \text{if } o = \infty \\ 2 & \text{if } o = \\ 3 & \text{if } o = \texttt{?8K} \\ 4 & \text{if } o = \texttt{?\^{}W} \end{cases}$$

Note the use of O above to succinctly encode the result of a work item and the slight transformations of E_G and E_P to take account of the fact their inner tuples contain variable-length sequence terms a and p which need length discriminators.

Appendix D. State Merklization

The Merklization process defines a cryptographic commitment from which arbitrary information within state may be provided as being authentic in a concise and swift fashion. We describe this in two stages; the first defines a mapping from 32-octet sequences to (unlimited) octet sequences in a process called *state serialization*. The second forms a 32-octet commitment from this mapping in a process called *Merklization*.

D.1. **Serialization.** The serialization of state primarily involves placing all the various components of σ into a single mapping from 32-octet sequence *state-keys* to octet sequences of indefinite length. The state-key is constructed from a hash component and a chapter component, equivalent to either the index of a state component or, in the case of the inner dictionaries of δ, a service index.

We define the state-key constructor functions C as:

(291) $$C \quad \mathbb{N}_{2^8} \quad (\mathbb{N}_{2^8}, \mathbb{N}_S) \quad \mathbb{N}_S, \mathbb{Y} \quad \to \mathbb{H}$$

$$\begin{aligned} i \quad \mathbb{N}_{2^8} &\quad [i, 0, 0, \dots] \\ (i, s \quad \mathbb{N}_S) &\quad [i, n_0, n_1, n_2, n_3, 0, 0, \dots] \text{ where } n = \mathrm{E}_4(s) \\ (s, h) &\quad [n_0, h_0, n_1, h_1, n_2, h_2, n_3, h_3, h_4, h_5, \dots, h_{27}] \text{ where } n = \mathrm{E}_4(s) \end{aligned}$$

The state serialization is then defined as the dictionary built from the amalgamation of each of the components. Cryptographic hashing ensures that there will be no duplicate state-keys given that there are no duplicate inputs to C. Formally, we define T which transforms some state σ into its serialized form:

(292) $$T(\sigma) \equiv$$

$$\begin{aligned} C(1) &\quad \mathrm{E}([\updownarrow x \quad x \lessdot \alpha]) \,, \\ C(2) &\quad \mathrm{E}(\varphi) \,, \\ C(3) &\quad \mathrm{E}(\updownarrow[(h, \mathrm{E}_M(\mathsf{b}), s, \updownarrow \mathsf{p}) \quad (h, \mathsf{b}, s, \mathsf{p}) \lessdot \beta]) \,, \\ C(4) &\quad \mathrm{E}(\ \gamma_{\mathbf{k}}, \gamma_z, \{ \begin{matrix} 0 \text{ if } \gamma_{\mathbf{s}} \quad \mathbb{C}_{\ \mathrm{E}} \\ 1 \text{ if } \gamma_{\mathbf{s}} \quad \mathbb{H}_{B\ \mathrm{E}} \end{matrix} \}, \gamma_{\mathbf{s}}, \updownarrow \gamma_{\mathbf{a}} \ ) \,, \\ C(5) &\quad \mathrm{E}(\updownarrow[x \quad x \quad \psi_{\mathbf{g}}], \updownarrow[x \quad x \quad \psi_{\mathbf{b}}], \updownarrow[x \quad x \quad \psi_{\mathbf{w}}], \updownarrow[x \quad x \quad \psi_{\mathbf{o}}]) \,, \\ C(6) &\quad \mathrm{E}(\eta) \,, \\ C(7) &\quad \mathrm{E}(\iota) \,, \\ C(8) &\quad \mathrm{E}(\kappa) \,, \\ C(9) &\quad \mathrm{E}(\lambda) \,, \\ C(10) &\quad \mathrm{E}([\text{¿}(w, \mathrm{E}_4(t)) \quad (w,t) \lessdot \rho]) \,, \\ C(11) &\quad \mathrm{E}_4(\tau) \,, \\ C(12) &\quad \mathrm{E}_4(\chi) \,, \\ C(13) &\quad \mathrm{E}_4(\pi) \,, \\ C(255, s) &\quad \mathsf{a}_c \quad \mathrm{E}_8(\mathsf{a}_b, \mathsf{a}_g, \mathsf{a}_m, \mathsf{a}_l) \quad \mathrm{E}_4(\mathsf{a}_i) \,, \\ C(s, h) &\quad \mathsf{v} \,, \\ C(s, h) &\quad \mathsf{p} \,, \\ C(s, \mathrm{E}_4(l)) \quad (\neg h_4) &\quad \mathrm{E}(\updownarrow[\mathrm{E}_4(x) \quad x \lessdot \mathsf{t}]) \end{aligned}$$

$$\begin{aligned} &(s \quad \mathsf{a}) \quad \delta \\ &(s \quad \mathsf{a}) \quad \delta, (h \quad \mathsf{v}) \quad \mathsf{a}_{\mathbf{s}} \\ &(s \quad \mathsf{a}) \quad \delta, (h \quad \mathsf{p}) \quad \mathsf{a}_{\mathbf{p}} \\ &(s \quad \mathsf{a}) \quad \delta, (\ h, l \quad \mathsf{t}) \quad \mathsf{a}_{\mathbf{l}} \end{aligned}$$

Note that most rows describe a single mapping between key derived from a natural and the serialization of a state component. However, the final four rows each define sets of mappings since these items act over all service accounts and in the case of the final three rows, the keys of a nested dictionary with the service.

Also note that all non-discriminator numeric serialization in state is done in fixed-length according to the size of the term.

D.2. **Merklization.** With T defined, we now define the rest of $\mathbf{M}$ which primarily involves transforming the serialized mapping into a cryptographic commitment. We define this commitment as the root of the binary Patricia Merkle Trie with a format optimized for modern compute hardware, primarily by optimizing sizes to fit succinctly into typical memory layouts and reducing the need for unpredictable branching.

D.2.1. *Node Encoding and Trie Identification.* We identify (sub-)tries as the hash of their root node, with one exception: empty (sub-)tries are identified as the zero-hash, $\mathbb{H}^0$.

Nodes are fixed in size at 512 bit (64 bytes). Each node is either a branch or a leaf. The first bit discriminate between these two types.

In the case of a branch, the remaining 511 bits are split between the two child node hashes, using the last 255 bits of the 0-bit (left) sub-trie identity and the full 256 bits of the 1-bit (right) sub-trie identity.

Leaf nodes are further subdivided into embedded-value leaves and regular leaves. The second bit of the node discriminates between these.

In the case of an embedded-value leaf, the remaining 6 bits of the first byte are used to store the embedded value size. The following 31 bytes are dedicated to the first 31 bytes of the key. The last 32 bytes are defined as the value, filling with zeroes if its length is less than 32 bytes.

In the case of a regular leaf, the remaining 6 bits of the first byte are zeroed. The following 31 bytes store the first 31 bytes of the key. The last 32 bytes store the hash of the value.

Formally, we define the encoding functions B and L:

$$B \begin{cases} (\mathbb{H},\mathbb{H}) \to \mathbb{B}_{512} \\ (l,r) \quad [0] \quad \text{bits}(l)_{1:::} \quad \text{bits}(r) \end{cases} \tag{293}$$

$$L_{(k,v)} \begin{cases} [1,0] \quad \text{bits}(\mathsf{E}_1(v))_{:::6} \quad \text{bits}(k)_{:::248} \quad \text{bits}(v) \quad [0,0,\dots] & \text{if } v \le 32 \\ [1,1,0,0,0,0,0,0] \quad \text{bits}(k)_{:::248} \quad \text{bits}(\mathsf{H}(v)) & \text{otherwise} \end{cases} \tag{294}$$

We may then define the basic Merklization function $\mathbf{M}$ as:

$$\mathbf{M}(\sigma) \equiv M(\{(\text{bits}(k) \quad k,v) \quad (k \quad v) \quad T(\sigma)\}) \tag{295}$$

(296)

$$M(d \in \mathbb{D}\langle \mathbb{B} \to (\mathbb{H},\mathbb{Y}) \rangle) \equiv \begin{cases} \mathbb{H}^0 & \text{if } d = 0 \\ \mathsf{H}(\text{bits}^{-1}(L(k,v))) & \text{if } \mathsf{V}(d) = \{(k,v)\} \\ \mathsf{H}(\text{bits}^{-1}(B(M(l),M(r)))) & \text{otherwise, where } b,p \in (b \quad p) \quad d \Leftrightarrow (b_{1:::} \quad p) \end{cases} \quad \begin{cases} l & \text{if } b_0 = 0 \\ r & \text{if } b_0 = 1 \end{cases}$$

Appendix E. General Merklization

E.1. **Binary Merkle Trees.** The Merkle tree is a cryptographic data structure yielding a hash commitment to a specific sequence of values. It provides $O(N)$ computation and $O(\log(N))$ proof size for inclusion. This *well-balanced* formulation ensures that the maximum depth of any leaf is minimal and that the number of leaves at that depth is also minimal.

The underlying function for our Merkle trees is the *node* function N, which accepts some sequence of blobs of some length n and provides either such a blob back or a hash:

$$N \in (\mathbb{Y}_n, \mathbb{Y} \to \mathbb{H}) \to \mathbb{Y}_n \quad \mathbb{H}$$

$$N_{(\mathbf{v},H)} \begin{cases} \mathbb{H}_0 & \text{if } \mathbf{v} = 0 \\ \mathbf{v}_0 & \text{if } \mathbf{v} = 1 \\ H(\texttt{MjnLO} \quad N(\mathbf{v}_{::: \,:\mathbf{v}\mathfrak{d}/2}, H) \quad N(\mathbf{v}_{:\mathbf{v}\mathfrak{d}/2\, :::}, H)) & \text{otherwise} \end{cases} \tag{297}$$

The astute reader will realize that if our $\mathbb{Y}_n$ happens to be equivalent $\mathbb{H}$ then this function will always evaluate into $\mathbb{H}$. That said, for it to be secure care must be taken to ensure there is no possibility of preimage collision. For this purpose we include the hash prefix `MjnLO` to minimize the chance of this; simply ensure any items are hashed with a different prefix and the system can be considered secure.

We also define the *trace* function T, which returns each opposite node from top to bottom as the tree is navigated to arrive at some leaf corresponding to the item of a given index into the sequence. It is useful in creating justifications of

data inclusion.

$$T\colon (\mathbb{Y}_n, \mathbb{N}_{\mathbf{v}}, \mathbb{Y} \to \mathbb{H}) \to \mathbb{Y}_n \quad \mathbb{H}$$
$$T(\mathbf{v}, i, H) \equiv \begin{cases} [N(P(\mathbf{v}, i), H)] \quad T(P(\mathbf{v}, i), i - P_I(\mathbf{v}, i), H) & \text{if } |\mathbf{v}| > 1 \\ [\mathbf{v}_0] & \text{otherwise} \end{cases} \tag{298}$$

where

$$P^s(\mathbf{v}, i) \equiv \begin{cases} \mathbf{v}_{:::|\mathbf{v}|/2} & \text{if } (i < |\mathbf{v}|/2) = s \\ \mathbf{v}_{|\mathbf{v}|/2:::} & \text{otherwise} \end{cases}$$

$$P_I(\mathbf{v}, i) \equiv \begin{cases} 0 & \text{if } i < |\mathbf{v}|/2 \\ |\mathbf{v}|/2 & \text{otherwise} \end{cases}$$

From this we define our other Merklization functions.

E.1.1. *Well-Balanced Tree.* We define the well-balanced binary Merkle function as $\mathbf{M}_B$:

$$\mathbf{M}_B\colon (\mathbb{Y}, \mathbb{Y} \to \mathbb{H}) \to \mathbb{H}$$
$$\mathbf{M}_B(\mathbf{v}, H) \equiv \begin{cases} H(\mathbf{v}_0) & \text{if } |\mathbf{v}| = 1 \\ N(\mathbf{v}, H) & \text{otherwise} \end{cases} \tag{299}$$

This is suitable for creating proofs on data which is not much greater than 32 octets in length since it avoids hashing each item in the sequence. For sequences with larger data items, it is better to hash them beforehand to ensure proof-size is minimal since each proof will generally contain a data item.

Note: In the case that no hash function argument H is supplied, we may assume the Blake 2b hash function, H.

E.1.2. *Constant-Depth Tree.* We define the constant-depth binary Merkle function as $\mathbf{M}$ and the corresponding justification generation function as J, with the latter having an optional subscript x, which limits the justification to only those nodes required to justify inclusion of a well-aligned subtree of (maximum) size 2^x:

$$\mathbf{M}\colon (\mathbb{Y}, \mathbb{Y} \to \mathbb{H}) \to \mathbb{H}$$
$$\mathbf{M}(\mathbf{v}, H) \equiv N(C(\mathbf{v}), H) \tag{300}$$

$$\mathrm{J}\colon (\mathbb{Y}, \mathbb{N}_{\mathbf{v}}, \mathbb{Y} \to \mathbb{H}) \to \mathbb{H}$$
$$\mathrm{J}(\mathbf{v}, i, H) \equiv (T(C(\mathbf{v}), i, H)) \tag{301}$$

$$\mathrm{J}_x\colon (\mathbb{Y}, \mathbb{N}_{\mathbf{v}}, \mathbb{Y} \to \mathbb{H}) \to \mathbb{H}$$
$$\mathrm{J}_x(\mathbf{v}, i, H) \equiv (T(C(\mathbf{v}), i, H)_{:::\max(0;\, \log_2(|\mathbf{v}|) - x)}) \tag{302}$$

For the latter justification to be acceptable, we must assume the target observer knows not merely the value of the item at the given index, but also all other items within its 2^x size subtree.

As above, we may assume a default value for H of the Blake 2b hash function, H.

In all cases, a constancy preprocessor function C is applied which hashes all data items with a fixed prefix and then pads them to the next power of two with the zero hash $\mathbb{H}_0$:

$$C\colon (\mathbb{Y}, \mathbb{Y} \to \mathbb{H}) \to \mathbb{H}$$
$$C(\mathbf{v}, H) \equiv \mathbf{v}' \text{ where } \begin{cases} |\mathbf{v}'| = 2^{\log_2(|\mathbf{v}|)} \\ \mathbf{v}'_i = \begin{cases} H(\texttt{MeO9T} \ \mathbf{v}_i) & \text{if } i < |\mathbf{v}| \\ \mathbb{H}_0 & \text{otherwise} \end{cases} \end{cases} \tag{303}$$

E.2. **Merkle Mountain Ranges.** The Merkle mountain range (MMR) is an append-only cryptographic data structure which yields a commitment to a sequence of values. Appending to an MMR and proof of inclusion of some item within it are both $O(\log(N))$ in time and space for the size of the set.

We define a Merkle mountain range as being within the set $\mathbb{H}?^{}$, a sequence of peaks, each peak the root of a Merkle tree containing 2^i items where i is the index in the sequence. Since we support set sizes which are not always powers-of-two-minus-one, some peaks may be empty, rather than a Merkle root.

Since the sequence of hashes is somewhat unwieldy as a commitment, Merkle mountain ranges are themselves generally hashed before being published. Hashing them removes the possibility of further appending so the range itself is kept on the system which needs to generate future proofs.

We define the append function A as:

$$\mathrm{A}\ \begin{matrix}(\ \mathbb{H}?\ ,\mathbb{H},\mathbb{Y}\to\mathbb{H})\to\ \mathbb{H}?\\ \quad(\mathbf{r},l,H)\quad P(\mathbf{r},l,0,H)\\ (\ \mathbb{H}?\ ,\mathbb{H},\mathbb{N},\mathbb{Y}\to\mathbb{H})\to\ \mathbb{H}?\end{matrix}$$

(304) where P

$$(\mathbf{r},l,n,H)\quad\begin{cases}\mathbf{r}+\!\!+\,l & \text{if } n\geq|\mathbf{r}|\\ R(\mathbf{r},n,l) & \text{if } n<|\mathbf{r}|\quad \mathbf{r}_n=\\ P(R(\mathbf{r},n,\ ),H(\mathbf{r}_n\ \ l),n+1,H) & \text{otherwise}\end{cases}$$

and R

$$\begin{matrix}(\ T\ ,\mathbb{N},T)\to\ T\\ \quad(\mathbf{s},i,v)\quad \mathbf{s}'\text{ where }\mathbf{s}'=\mathbf{s}\text{ except }\mathbf{s}'_i=v\end{matrix}$$

We define the MMR encoding function as E_M:

(305)
$$\mathrm{E}_M\ \begin{matrix}\mathbb{H}?\ \to\mathbb{H}\\ \quad\mathbf{b}\quad \mathrm{E}(\updownarrow[\lfloor x\ \ x\in\mathbf{b}])\end{matrix}$$

Appendix F. Shuffling

The Fisher-Yates shuffle function is defined formally as:

(306)
$$T,l\ \ \mathbb{N}\ \ \mathrm{F}\quad\begin{matrix}(\ T\ _l,\ \mathbb{N}\ _l)\to\ T\ _l\\ \quad(\mathbf{s},\mathbf{r})\quad\begin{cases}[\mathbf{s}_{\mathbf{r}_0\bmod l}]\ \ \mathrm{F}([\mathbf{s}_i\ \ i\in\mathbb{N}_l\ \ \{\mathbf{r}_0\bmod l\}],[\mathbf{r}_1,\mathbf{r}_2,\ldots]) & \text{if } \mathbf{s}\neq[]\\ [] & \text{otherwise}\end{cases}\end{matrix}$$

Since it is often useful to shuffle a sequence based on some random seed in the form of a hash, we provide a secondary form of the shuffle function F which accepts a 32-byte hash instead of the numeric sequence. We define Q, the numeric-sequence-from-hash function, thus:

(307)
$$l\ \ \mathbb{N}\ \ \mathrm{Q}_l\ \begin{matrix}\mathbb{H}\to\ \mathbb{N}_l\\ \quad h\quad[\mathrm{E}_4^{-1}(\mathrm{H}(h\ \ \mathrm{E}_4(\ ^i/_8\ ))_{4i\bmod 32:::})\ \ i\ \ \mathbb{N}_l]\end{matrix}$$

(308)
$$T,l\ \ \mathbb{N}\ \ \mathrm{F}\ \{\begin{matrix}(\ T\ _l,\mathbb{H})\to\ T\ _l\\ \quad(\mathbf{s},h)\quad \mathrm{F}(\mathbf{s},\mathrm{Q}_l(h))\end{matrix}$$

Appendix G. Bandersnatch Ring VRF

The Bandersnatch curve is defined by Masson, Sanso, and Zhang 2021.

The singly-contextualized Bandersnatch Schnorr-like signatures $\mathbb{F}_k^{m\ \mathbb{Y}}{}_{\mathrm{H}_B}\ c\ \ \mathbb{H}$ are defined as a formulation under the *ietf* VRF template specified by Hosseini and Galassi 2024 (as IETF VRF) and further detailed by Goldberg et al. 2023.

(309)
$$\mathbb{F}_k^{m\ \mathbb{Y}}{}_{\mathrm{H}_B}\ c\ \ \mathbb{H}\quad\mathbb{Y}_{96}\equiv\{x\ \ x\quad Y_{96},\mathrm{verify}(k,c,m,\mathrm{decode}(x_{\ 32}),\mathrm{decode}(x_{32}))=\ \}$$

(310)
$$\mathrm{Y}(s\ \ \mathbb{F}_k^m\ c\ )\ \ \mathbb{H}\equiv\mathrm{hashed\_output}(\mathrm{decode}(x_{\ 32})\ \ x\ \ \mathbb{F}_k^m\ c\ )$$

The singly-contextualized Bandersnatch RingVRF proofs $\mathbb{F}_r^m\ c$ are a zk-SNARK-enabled analogue utilizing the Pedersen VRF, also defined by Hosseini and Galassi 2024 and further detailed by Jeffrey Burdges et al. 2023.

(311)
$$\mathrm{O}(\ \mathbb{H}_B\ )\ \ \mathbb{Y}_R\equiv\mathrm{KZG\_commitment}(\ \mathbb{H}_B\ )$$

(312)
$$\mathbb{F}_r^{m\ \mathbb{Y}}{}_{\mathbb{Y}_R}\ c\ \ \mathbb{H}\quad\mathbb{Y}_{784}\equiv\{x\ \ x\quad\mathbb{Y}_{784},\mathrm{verify}(r,c,m,\mathrm{decode}(x_{\ 32}),\mathrm{decode}(x_{32}))=\ \}$$

(313)
$$\mathrm{Y}(p\ \ \mathbb{F}_r^m\ c\ )\ \ \mathbb{H}\equiv\mathrm{hashed\_output}(\mathrm{decode}(x_{\ 32})\ \ x\ \ \mathbb{F}_r^m\ c\ )$$

Appendix H. Erasure Coding

The foundation of the data-availability and distribution system of JAM is a systematic Reed-Solomon erasure coding function in GF(16) of rate 342:1023, the same transform as done by the algorithm of Lin, Chung, and Han 2014. We use a little-endian $\mathbb{Y}_2$ form of the 16-bit GF points with a functional equivalence given by E_2. From this we may assume the encoding function $\mathrm{C}\quad\mathbb{Y}_{2\ 342}\to\ \mathbb{Y}_{2\ 1023}$ and the recovery function $\mathrm{R}\quad\mathbb{Y}_2,\mathbb{N}_{1023\ \ 342}\to\ \mathbb{Y}_{2\ 342}$. Encoding is done by extrapolating a data blob of size 684 octets (provided in C here as 342 octet pairs) into 1,023 octet pairs. Recovery is done by collecting together any distinct 342 octet pairs, together with their indices and transforming this into the original sequence of 342 octet pairs.

Practically speaking, this allows for the efficient encoding and recovery of data whose size is a multiple of 684 octets. Data whose length is not divisible by 684 must be padded (we pad with zeroes). We use this erasure-coding in two contexts within the JAM protocol; one where we encode variable sized (but typically very large) data blobs for the Audit DA and block-distribution system, and the other where we encode much smaller fixed-size data *segments* for the Import DA system.

For the Import DA system, we deal with an input size of 4,104 octets resulting in data-parallelism of order six. We may attain a greater degree of data parallelism if encoding or recovering more than one segment at a time though for recovery, we may be restricted to the requiring each segment to be formed from the same set of indices (depending on the specific algorithm).

H.1. **Blob Encoding and Recovery.** We assume some data blob $\mathbf{d}$ $\mathbb{Y}_{684k}, k$ $\mathbb{N}$. We are able to express this as a whole number of k pieces each of a sequence of 684 octets. We denote these (data-parallel) pieces $\mathbf{p}$ $\mathbb{Y}_{684}$ $= \text{split}_{684}(\mathbf{p})$. Each piece is then reformed as 342 octet pairs and erasure-coded using C as above to give 1,023 octet pairs per piece.

The resulting matrix is grouped by its pair-index and concatenated to form 1,023 *chunks*, each of k octet-pairs. Any 342 of these chunks may then be used to reconstruct the original data $\mathbf{d}$.

Formally we begin by defining two utility functions for splitting some large sequence into a number of equal-sized sub-sequences and for joining subsequences back into a single large sequence:

$$n, k \quad \mathbb{N} \quad \text{split}_n(\mathbf{d} \quad \mathbb{Y}_{k\,n}) \quad \mathbb{Y}_{n\ \ k} \equiv [\mathbf{d}_{0\ +n}, \mathbf{d}_{n\ +n}, \quad , \mathbf{d}_{(k-1)n\ +n}] \tag{314}$$

$$n, k \quad \mathbb{N} \quad \text{join}(\mathbf{c} \quad \mathbb{Y}_{n\ \ k}) \quad \mathbb{Y}_{k\,n} \equiv \mathbf{c}_0 \quad \mathbf{c}_1 \quad \ldots \tag{315}$$

We define the transposition operator hence:

$$^{\mathrm{T}}[[\mathbf{x}_{0;0}, \mathbf{x}_{0;1}, \mathbf{x}_{0;2}, \ldots], [\mathbf{x}_{1;0}, \mathbf{x}_{1;1}, \ldots], \ldots] \equiv [[\mathbf{x}_{0;0}, \mathbf{x}_{1;0}, \mathbf{x}_{2;0}, \ldots], [\mathbf{x}_{0;1}, \mathbf{x}_{1;1}, \ldots], \ldots] \tag{316}$$

We may then define our erasure-code chunking function which accepts an arbitrary sized data blob whose length divides wholly into 684 octets and results in 1,023 sequences of sequences each of smaller blobs:

$$\mathrm{C}_{k\ \mathrm{N}} \quad \begin{matrix} \mathbb{Y}_{684k} \to \mathbb{Y}_{2k\ \ 1023} \\ \mathbf{d} \quad [\text{join}(\mathbf{c}) \quad \mathbf{c} \lhd {}^{\mathrm{T}}[\mathrm{C}(\mathbf{p}) \quad \mathbf{p} \lhd \text{split}_{684}(\mathbf{d})]] \end{matrix} \tag{317}$$

The original data may be reconstructed with only 342 of the 1,023 items of said function's result, together with the items' respective indices:

$$\mathrm{R}_{k\ \mathrm{N}}\ \{ \begin{matrix} \{(\mathbb{Y}_{2k}, \mathbb{N}_{1023})\}_{342} \to \mathbb{Y}_{684k} \\ \mathbf{c} \quad \text{join}([\mathrm{R}([(\text{split}_2(\mathbf{x})_p, i) \quad (\mathbf{x}, i) \lhd \mathbf{c}]) \quad p \quad \mathbb{N}_k]) \end{matrix} \tag{318}$$

Segment encoding is just this with $k = 6$.

H.2. **Code Word representation.** For the sake of brevity we call each octet pair a *word*. The code words (including the message words) are treated as element of $\mathbb{F}_{2^{16}}$ finite field. The field is generated as an extension of $\mathbb{F}_2$ using the irreducible polynomial:

$$x^{16} + x^5 + x^3 + x^2 + 1 \tag{319}$$

Hence:

$$\mathbb{F}_{16} \equiv \frac{\mathbb{F}_2[x]}{x^{16}} + x^5 + x^3 + x^2 + 1 \tag{320}$$

We name the generator of $\frac{\mathbb{F}_{16}}{\mathbb{F}_2}$, the root of the above polynomial, α as such: $\mathbb{F}_{16} = \mathbb{F}_2(\alpha)$.

Instead of using the standard basis $\{1, \alpha, \alpha, \ldots, \alpha^{15}\}$, we opt for a representation of $\mathbb{F}_{16}$ which performs more efficiently for the encoding and the decoding process. To that aim, we name this specific representation of $\mathbb{F}_{16}$ as $\mathbb{F}_{16}$ and define it as a vector space generated by the following Cantor basis:

v_0	1
v_1	$\alpha^{15} + \alpha^{13} + \alpha^{11} + \alpha^{10} + \alpha^7 + \alpha^6 + \alpha^3 + \alpha$
v_2	$\alpha^{13} + \alpha^{12} + \alpha^{11} + \alpha^{10} + \alpha^3 + \alpha^2 + \alpha$
v_3	$\alpha^{12} + \alpha^{10} + \alpha^9 + \alpha^5 + \alpha^4 + \alpha^3 + \alpha^2 + \alpha$
v_4	$\alpha^{15} + \alpha^{14} + \alpha^{10} + \alpha^8 + \alpha^7 + \alpha$
v_5	$\alpha^{15} + \alpha^{14} + \alpha^{13} + \alpha^{11} + \alpha^{10} + \alpha^8 + \alpha^5 + \alpha^3 + \alpha^2 + \alpha$
v_6	$\alpha^{15} + \alpha^{12} + \alpha^8 + \alpha^6 + \alpha^3 + \alpha^2$
v_7	$\alpha^{14} + \alpha^4 + \alpha$
v_8	$\alpha^{14} + \alpha^{13} + \alpha^{11} + \alpha^{10} + \alpha^7 + \alpha^4 + \alpha^3$
v_9	$\alpha^{12} + \alpha^7 + \alpha^6 + \alpha^4 + \alpha^3$
v_{10}	$\alpha^{14} + \alpha^{13} + \alpha^{11} + \alpha^9 + \alpha^6 + \alpha^5 + \alpha^4 + \alpha$
v_{11}	$\alpha^{15} + \alpha^{13} + \alpha^{12} + \alpha^{11} + \alpha^8$
v_{12}	$\alpha^{15} + \alpha^{14} + \alpha^{13} + \alpha^{12} + \alpha^{11} + \alpha^{10} + \alpha^8 + \alpha^7 + \alpha^5 + \alpha^4 + \alpha^3$
v_{13}	$\alpha^{15} + \alpha^{14} + \alpha^{13} + \alpha^{12} + \alpha^{11} + \alpha^9 + \alpha^8 + \alpha^5 + \alpha^4 + \alpha^2$
v_{14}	$\alpha^{15} + \alpha^{14} + \alpha^{13} + \alpha^{12} + \alpha^{11} + \alpha^{10} + \alpha^9 + \alpha^8 + \alpha^5 + \alpha^4 + \alpha^3$
v_{15}	$\alpha^{15} + \alpha^{12} + \alpha^{11} + \alpha^8 + \alpha^4 + \alpha^3 + \alpha^2 + \alpha$

Every message word $m_i = m_{i;15} \ldots m_{i;0}$ consists of 16 bits. As such it could be regarded as binary vector of length 16:

$$m_i = (m_{i;0} \ldots m_{i;15}) \tag{321}$$

Where $m_{i;0}$ is the least significant bit of message word m_i. Accordingly we consider the field element $\widetilde{m}_i = \sum_{j=0}^{15} m_{i;j} v_j$ to represent that message word.

Similarly, we assign a unique index to each validator between 0 and 1,022 and we represent validator i with the field element:

$$\tilde{i} = \sum_{j=0}^{15} i_j v_j \tag{322}$$

where $i = i_{15} \dots i_0$ is the binary representation of i.

H.3. **The Generator Polynomial.** To erasure code a message of 342 words into 1023 code words, we represent each message as a field element as described in previous section and we interpolate the polynomial $p(y)$ of maximum 341 degree which satisfies the following equalities:

$$\begin{aligned} p(\tilde{0}) &= \tilde{m}_0 \\ p(\tilde{1}) &= \tilde{m}_1 \\ &\dots \\ p(\tilde{341}) &= \tilde{m}_{341} \end{aligned} \tag{323}$$

After finding $p(y)$ with such properties, we evaluate p at the following points:

$$\begin{aligned} \tilde{r}_{342} &= p(\tilde{342}) \\ \tilde{r}_{343} &= p(\tilde{343}) \\ &\dots \\ \tilde{r}_{1022} &= p(\tilde{1022}) \end{aligned} \tag{324}$$

We then distribute the message words and the extra code words among the validators according to their corresponding indices.

Appendix I. Index of Notation

I.1. Sets.

I.1.1. *Regular Notation.*

$\mathbb{N}$: The set of non-negative integers. Subscript denotes one greater than the maximum. See section 3.4.
- $\mathbb{N}^+$: The set of positive integers (not including zero).
- $\mathbb{N}_B$: The set of balance values. Equivalent to $\mathbb{N}_{2^{64}}$. See equation 31.
- $\mathbb{N}_G$: The set of unsigned gas values. Equivalent to $\mathbb{N}_{2^{64}}$. See equation 33.
- $\mathbb{N}_L$: The set of blob length values. Equivalent to $\mathbb{N}_{2^{32}}$. See section 3.4.
- $\mathbb{N}_S$: The set from which service indices are drawn. Equivalent to $\mathbb{N}_{2^{32}}$. See section 88.
- $\mathbb{N}_T$: The set of timeslot values. Equivalent to $\mathbb{N}_{2^{32}}$. See equation 36.

$\mathbb{Q}$: The set of rational numbers. Unused.
$\mathbb{R}$: The set of real numbers. Unused.
$\mathbb{Z}$: The set of integers. Subscript denotes range. See section 3.4.
- $\mathbb{Z}_G$: The set of signed gas values. Equivalent to $\mathbb{Z}_{-2^{63}\,2^{63}}$. See equation 33.

I.1.2. *Custom Notation.*

$\mathbb{A}$: The set of service accounts. See equation 90.
$\mathbb{B}$: The set of Boolean sequences/bitstrings. Subscript denotes length. See section 3.7.
$\mathbb{C}$: The set of seal-key tickets. See equation 51. *Not used as the set of complex numbers.*
$\mathbb{D}$: The set of dictionaries. See section 3.5.
- $\mathbb{D}\ K \to V$: The set of dictionaries making a partial bijection of domain K to range V. See section 3.5.

$\mathbb{E}$: The set of valid Ed25519 signatures. A subset of $\mathbb{Y}_{64}$. See section 3.8.
- $\mathbb{E}_K\ M$: The set of valid Ed25519 signatures of the key K and message M. A subset of $\mathbb{E}$. See section 3.8.

$\mathbb{F}$: The set of Bandersnatch signatures. A subset of $\mathbb{Y}_{64}$. See section 3.8. *NOTE: Not used as finite fields.*
- $\mathbb{F}_K^M\ C$: The set of Bandersnatch signatures of the public key K, context C and message M. A subset of $\mathbb{F}$. See section 3.8.
- $\mathbb{F}$: The set of Bandersnatch RingVRF proofs. See section 3.8.
- $\mathbb{F}_R^M\ C$: The set of Bandersnatch RingVRF proofs of the root R, context C and message M. A subset of $\mathbb{F}$. See section 3.8.

$\mathbb{G}$: The set of data segments, equivalent to octet sequences of length W_S. See equation 175.
$\mathbb{H}$: The set of 32-octet cryptographic values. A subset of $\mathbb{Y}_{32}$. $\mathbb{H}$ without a subscript generally implies a hash function result. See section 3.8. *NOTE: Not used as quaternions.*
- $\mathbb{H}_B$: The set of Bandersnatch public keys. A subset of $\mathbb{Y}_{32}$. See section 3.8 and appendix G.
- $\mathbb{H}_E$: The set of Ed25519 public keys. A subset of $\mathbb{Y}_{32}$. See section 3.8.2.

$\mathbb{I}$: The set of work items. See equation 177.
$\mathbb{J}$: The set of work execution errors.
$\mathbb{K}$: The set of validator key-sets. See equation 52.
$\mathbb{L}$: The set of work results.
$\mathbb{M}$: The set of PVM RAM states. A superset of $\mathbb{Y}_{2^{32}}$. See appendix A.
$\mathbb{O}$: The accumulation operand element, corresponding to a single work result.
$\mathbb{P}$: The set of work-packages. See equation 176.
$\mathbb{S}$: The set of work-package specifications.
$\mathbb{T}$: The set of deferred transfers.
$\mathbb{U}$: Unused.
$\mathbb{V}$: The set of validly readable indices for PVM RAM μ. See appendix A.
$\mathbb{V}$: The set of validly writable indices for PVM RAM μ. See appendix A.
$\mathbb{W}$: The set of work-reports.
$\mathbb{X}$: The set of refinement contexts.
$\mathbb{Y}$: The set of octet strings/"blobs". Subscript denotes length. See section 3.7.
- $\mathbb{Y}_{BLS}$: The set of BLS public keys. A subset of $\mathbb{Y}_{144}$. See section 3.8.2.
- $\mathbb{Y}_R$: The set of Bandersnatch ring roots. A subset of $\mathbb{Y}_{144}$. See section 3.8 and appendix G.

I.2. Functions.

: The historical lookup function. See equation 94.
: The work result computation function. See equation 184.
: The general state transition function. See equations 12, 16.
: The key-nullifier function. See equation 59.
: The whole-program PVM machine state-transition function. See equation A.
- $_1$: The single-step (PVM) machine state-transition function. See appendix A.
- $_A$: The Accumulate PVM invocation function. See appendix B.
- $_H$: The host-function invocation (PVM) with host-function marshalling. See appendix A.
- $_I$: The Is-Authorized PVM invocation function. See appendix B.

$_M$: The marshalling whole-program PVM machine state-transition function. See appendix A.
$_R$: The Refine PVM invocation function. See appendix B.
$_T$: The On-Transfer PVM invocation function. See appendix B.
$_$: Virtual machine host-call functions. See appendix B.
$_A$: Assign-core host-call.
$_C$: Checkpoint host-call.
$_D$: Designate-validators host-call.
$_E$: Empower-service host-call.
$_F$: Forget-preimage host-call.
$_G$: Gas-remaining host-call.
$_H$: Historical-lookup-preimage host-call.
$_I$: Information-on-service host-call.
$_K$: Kickoff-PVM host-call.
$_L$: Lookup-preimage host-call.
$_M$: Make-PVM host-call.
$_N$: New-service host-call.
$_O$: Poke-PVM host-call.
$_P$: Peek-PVM host-call.
$_Q$: Quit-service host-call.
$_S$: Solicit-preimage host-call.
$_R$: Read-storage host-call.
$_T$: Transfer host-call.
$_U$: Upgrade-service host-call.
$_W$: Write-storage host-call.
$_X$: Expunge-PVM host-call.
$_Y$: Import segment host-call.
$_Z$: Export segment host-call.

I.3. Utilities, Externalities and Standard Functions.

$\mathsf{A}(\dots)$: The Merkle mountain range append function. See equation 304.
$\mathsf{B}_n(\dots)$: The octets-to-bits function for n octets. Superscripted $^{-1}$ to denote the inverse. See equation 223.
$\mathsf{C}(\dots)$: The group of erasure-coding functions.
$\mathsf{C}_n(\dots)$: The erasure-coding functions for n chunks. See equation 317.
$\mathsf{E}(\dots)$: The octet-sequence encode function. Superscripted $^{-1}$ to denote the inverse. See appendix C.
$\mathsf{F}(\dots)$: The Fisher-Yates shuffle function. See equation 306.
$\mathsf{H}(\dots)$: The Blake 2b 256-bit hash function. See section 3.8.
$\mathsf{H}_K(\dots)$: The Keccak 256-bit hash function. See section 3.8.
$\mathsf{K}(\dots)$: The domain, or set of keys, of a dictionary. See section 3.5.
$\mathsf{M}(\dots)$: The constant-depth binary Merklization function. See appendix E.
$\mathsf{M}_B(\dots)$: The well-balanced binary Merklization function. See appendix E.
$\mathsf{M}\ \ (\dots)$: The state Merklization function. See appendix D.
$\mathsf{N}(\dots)$: The erasure-coding chunks function. See appendix H.
$\mathsf{O}(\dots)$: The Bandersnatch ring root function. See section 3.8 and appendix G.
$\mathsf{P}_n(\dots)$: The octet-array zero-padding function. See equation 187.
$\mathsf{Q}(\dots)$: The numeric-sequence-from-hash function. See equation 308.
R: The group of erasure-coding piece-recovery functions.
$\mathsf{S}_k(\dots)$: The general signature function. See section 3.8.
T: The current time expressed in seconds after the start of the JAM Common Era. See section 4.4.
$\mathsf{U}(\dots)$: The substitute-if-nothing function. See equation 2.
$\mathsf{V}(\dots)$: The range, or set of values, of a dictionary or sequence. See section 3.5.
$\mathsf{X}_n(\dots)$: The signed-extension function for a value in $\mathbb{N}_{2^{8n}}$. See equation 225.
$\mathsf{Y}(\dots)$: The alias/output/entropy function of a Bandersnatch VRF signature/proof. See section 3.8 and appendix G.
$\mathsf{Z}_n(\dots)$: The into-signed function for a value in $\mathbb{N}_{2^{8n}}$. Superscripted with $^{-1}$ to denote the inverse. See equation 221.
$\dots$: Power set function.

I.4. Values.

I.4.1. *Block-context Terms.* These terms are all contextualized to a single block. They may be superscripted with some other term to alter the context and reference some other block.

$\mathbf{A}$: The ancestor set of the block. See equation 39.
$\mathbf{B}$: The block. See equation 13.
$\mathbf{C}$: The service accumulation-commitment, used to form the BEEFY root. See equation 165.

$\mathbf{E}$: The block extrinsic. See equation 14.
$\mathbf{F}_v$: The Beefy signed commitment of validator v. See equation 209.
$\mathbf{G}$: The mapping from cores to guarantor keys. See section 11.3.
$\mathbf{G}$: The mapping from cores to guarantor keys for the previous rotation. See section 11.3.
$\mathbf{H}$: The block header. See equation 37.
$\mathbf{Q}$: The selection of ready work-reports which a validator determined they must audit. See equation 190.
$\mathbf{R}$: The set of Ed25519 guarantor keys who made a work-report. See equation 141.
$\mathbf{S}$: The set of indices of services which have been accumulated ("progressed") in the block. See equation 159.
$\mathbf{T}$: The ticketed condition, true if the block was sealed with a ticket signature rather than a fallback. See equations 60 and 61.
$\mathbf{U}$: The audit condition, equal to once the block is audited. See section 17.
$\mathbf{W}$: The set of work-reports which have now become available and ready for accumulation. See equation 131.

Without any superscript, the block is assumed to the block being imported or, if no block is being imported, the head of the best chain (see section 19). Explicit block-contextualizing superscripts include:

$\mathbf{B}$: The latest finalized block. See equation 19.
$\mathbf{B}$: The block at the head of the best chain. See equation 19.

I.4.2. *State components.* Here, the prime annotation indicates posterior state. Individual components may be identified with a letter subscript.

α: The core αuthorizations pool. See equation 85.
β: Information on the most recent βlocks.
γ: State concerning Safrole. See equation 48.
  $\gamma_{\mathbf{a}}$: The sealing lottery ticket accumulator.
  $\gamma_{\mathbf{k}}$: The keys for the validators of the next epoch, equivalent to those keys which constitute γ_z.
  $\gamma_{\mathbf{s}}$: The sealing-key sequence of the current epoch.
  γ_z: The Bandersnatch root for the current epoch's ticket submissions.
δ: The (prior) state of the service accounts.
  $\delta^\dagger$: The post-preimage integration, pre-accumulation intermediate state.
  $\delta^\ddagger$: The post-accumulation, pre-transfer intermediate state.
η: The eηtropy accumulator and epochal raηdomness.
ι: The validator keys and metadata to be drawn from next.
κ: The validator κeys and metadata currently active.
λ: The validator keys and metadata which were active in the prior epoch.
ρ: The ρending reports, per core, which are being made available prior to accumulation.
  $\rho^\dagger$: The post-judgement, pre-guarantees-extrinsic intermediate state.
  $\rho^\ddagger$: The post-guarantees-extrinsic, pre-assurances-extrinsic, intermediate state.
σ: The σverall state of the system. See equations 12, 15.
τ: The most recent block's τimeslot.
φ: The authorization queue.
ψ: Past judgements on work-reports and validators.
  $\psi_{\mathbf{b}}$: Work-reports judged to be incorrect.
  $\psi_{\mathbf{g}}$: Work-reports judged to be correct.
  $\psi_{\mathbf{w}}$: Work-reports whose validity is judged to be unknowable.
  $\psi_{\mathbf{o}}$: Validators who made a judgement found to be incorrect.
χ: The privileged service indices.
  χ_m: The index of the empower service.
  χ_v: The index of the designate service.
  χ_a: The index of the assign service.
π: The activity statistics for the validators.

I.4.3. *Virtual Machine components.*

ε: The exit-reason resulting from all machine state transitions.
ν: The immediate values of an instruction.
μ: The memory sequence; a member of the set $\mathbb{M}$.
ξ: The gas counter.
ω: The registers.
ζ: The instruction sequence.
ϖ: The sequence of basic blocks of the program.
$\imath$: The instruction counter.

I.4.4. *Constants.*

$\mathrm{A} = 8$: The period, in seconds, between audit tranches.
$\mathrm{B}_I = 10$: The additional minimum balance required per item of elective service state.
$\mathrm{B}_L = 1$: The additional minimum balance required per octet of elective service state.

$B_S = 100$: The basic minimum balance which all services require.
$C = 341$: The total number of cores.
$D = 28,800$: The period in timeslots after which an unreferenced preimage may be expunged.
$E = 600$: The length of an epoch in timeslots.
$F = 2$: The audit bias factor, the expected number of additional validators who will audit a work-report in the following tranche for each no-show in the previous.
G_A: The total gas allocated to a core for Accumulation.
G_I: The gas allocated to invoke a work-package's Is-Authorized logic.
G_R: The total gas allocated for a work-package's Refine logic.
$H = 8$: The size of recent history, in blocks.
$I = 4$: The maximum amount of work items in a package.
$K = 16$: The maximum number of tickets which may be submitted in a single extrinsic.
$L = 14,400$: The maximum age in timeslots of the lookup anchor.
$M = 128$: The size of a transfer memo in octets.
$N = 2$: The number of ticket entries per validator.
$O = 8$: The maximum number of items in the authorizations pool.
$P = 6$: The slot period, in seconds.
$Q = 80$: The maximum number of items in the authorizations queue.
$R = 10$: The rotation period of validator-core assignments, in timeslots.
$S = 4,000,000$: The maximum size of service code in octets.
$U = 5$: The period in timeslots after which reported but unavailable work may be replaced.
$V = 1023$: The total number of validators.
$W_C = 684$: The basic size of our erasure-coded pieces. See equation 317.
$W_M = 2^{11}$: The maximum number of entries in a work-package manifest.
$W_P = 12\ \ 2^{20}$: The maximum size of an encoded work-package together with its extrinsic data and import implications, in octets.
$W_R = 96\ \ 2^{10}$: The maximum size of an encoded work-report in octets.
$W_S = 6$: The size of an exported segment in erasure-coded pieces.
X: Context strings, see below.
$Y = 500$: The number of slots into an epoch at which ticket-submission ends.
$Z_A = 4$: The PVM dynamic address alignment factor. See equation 227.
$Z_I = 2^{24}$: The standard PVM program initialization input data size. See equation A.7.
$Z_P = 2^{14}$: The standard PVM program initialization page size. See section A.7.
$Z_Q = 2^{16}$: The standard PVM program initialization segment size. See section A.7.

I.4.5. *Signing Contexts.*

X_A = `Ma9h¬9®9_e9@eO`: *Ed25519* Availability assurances.
X_B = `Ma9h¬@OOT´`: bls Accumulate-result-root-MMR commitment.
X_E = `Ma9h¬Oj§}nq´`: On-chain entropy generation.
X_F = `Ma9h¬T9ee@9Hc¬ O9e`: *Bandersnatch* Fallback block seal.
X_G = `Ma9h¬X«9}9j§OO`: *Ed25519* Guarantee statements.
X_I = `Ma9h¬9jjn«jHO`: *Ed25519* Audit announcement statements.
X_T = `Ma9h¬§_HcO§¬ O9e`: *Bandersnatch Ringvrf* Ticket generation and regular block seal.
X_U = `Ma9h¬9«L_§`: *Bandersnatch* Audit selection entropy.
X = `Ma9h¬®9e_L`: *Ed25519* Judgements for valid work-reports.
X = `Ma9h¬_j®9e_L`: *Ed25519* Judgements for invalid work-reports.

References

Bertoni, Guido et al. (2013). "Keccak". In: *Annual international conference on the theory and applications of cryptographic techniques*. Springer, pp. 313–314.

Bögli, Roman (2024). "Assessing RISC Zero using ZKit: An Extensible Testing and Benchmarking Suite for ZKP Frameworks". PhD thesis. OST Ostschweizer Fachhochschule.

Boneh, Dan, Ben Lynn, and Hovav Shacham (2004). "Short Signatures from the Weil Pairing". In: *J. Cryptology* 17, pp. 297–319. DOI: 10.1007/s00145-004-0314-9.

Burdges, Jeff, Alfonso Cevallos, et al. (2024). *Efficient Execution Auditing for Blockchains under Byzantine Assumptions*. Cryptology ePrint Archive, Paper 2024/961. https://eprint.iacr.org/2024/961. URL: https://eprint.iacr.org/2024/961.

Burdges, Jeff, Oana Ciobotaru, et al. (2022). *Efficient Aggregatable BLS Signatures with Chaum-Pedersen Proofs*. Cryptology ePrint Archive, Paper 2022/1611. https://eprint.iacr.org/2022/1611. URL: https://eprint.iacr.org/2022/1611.

Burdges, Jeffrey et al. (2023). *Ring Verifiable Random Functions and Zero-Knowledge Continuations*. Cryptology ePrint Archive, Paper 2023/002. URL: https://eprint.iacr.org/2023/002.

Buterin, Vitalik (2013). *Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform*. URL: https://github.com/ethereum/wiki/wiki/White-Paper.

Buterin, Vitalik and Virgil Griffith (2019). *Casper the Friendly Finality Gadget*. arXiv: 1710.09437 [cs.CR].

Cosmos Project (2023). *Interchain Security Begins a New Era for Cosmos*. Fetched 18th March, 2024. URL: https://blog.cosmos.network/interchain-security-begins-a-new-era-for-cosmos-a2dc3c0be63.

Dune and hildobby (2024). *Ethereum Staking*. Fetched 18th March, 2024. URL: https://dune.com/hildobby/eth2-staking.

Ethereum Foundation (2024a). "A digital future on a global scale". In: Fetched 4th April, 2024. URL: https://ethereum.org/en/roadmap/vision/.

— (2024b). *Danksharding*. Fetched 18th March, 2024. URL: https://ethereum.org/en/roadmap/danksharding/.

Fisher, Ronald Aylmer and Frank Yates (1938). *Statistical tables for biological, agricultural and medical research*. Oliver and Boyd.

Gabizon, Ariel, Zachary J. Williamson, and Oana Ciobotaru (2019). *PLONK: Permutations over Lagrange-bases for Oecumenical Noninteractive arguments of Knowledge*. Cryptology ePrint Archive, Paper 2019/953. URL: https://eprint.iacr.org/2019/953.

Goldberg, Sharon et al. (Aug. 2023). *Verifiable Random Functions (VRFs)*. RFC 9381. DOI: 10.17487/RFC9381. URL: https://www.rfc-editor.org/info/rfc9381.

Hertig, Alyssa (2016). *So, Ethereum's Blockchain is Still Under Attack...* Fetched 18th March, 2024. URL: https://www.coindesk.com/markets/2016/10/06/so-ethereums-blockchain-is-still-under-attack/.

Hopwood, Daira et al. (2020). *BLS12-381*. URL: https://z.cash/technology/jubjub/.

Hosseini, Seyed and Davide Galassi (2024). "Bandersnatch VRF-AD Specification". In: Fetched 4th April, 2024. URL: https://github.com/davxy/bandersnatch-vrfs-spec/blob/main/specification.pdf.

Jha, Prashant (2024). *Solana outage raises questions about client diversity and beta status*. Fetched 18th March, 2024. URL: https://cointelegraph.com/news/solana-outage-client-diversity-beta.

Josefsson, Simon and Ilari Liusvaara (Jan. 2017). *Edwards-Curve Digital Signature Algorithm (EdDSA)*. RFC 8032. DOI: 10.17487/RFC8032. URL: https://www.rfc-editor.org/info/rfc8032.

Kokoris-Kogias, Eleftherios et al. (2017). *OmniLedger: A Secure, Scale-Out, Decentralized Ledger via Sharding*. Cryptology ePrint Archive, Paper 2017/406. https://eprint.iacr.org/2017/406. URL: https://eprint.iacr.org/2017/406.

Kwon, Jae and Ethan Buchman (2019). "Cosmos whitepaper". In: *A Netw. Distrib. Ledgers* 27, pp. 1–32.

Lin, Sian-Jheng, Wei-Ho Chung, and Yunghsiang S. Han (2014). "Novel Polynomial Basis and Its Application to Reed-Solomon Erasure Codes". In: *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pp. 316–325. DOI: 10.1109/FOCS.2014.41.

Masson, Simon, Antonio Sanso, and Zhenfei Zhang (2021). *Bandersnatch: a fast elliptic curve built over the BLS12-381 scalar field*. Cryptology ePrint Archive, Paper 2021/1152. URL: https://eprint.iacr.org/2021/1152.

Ng, Felix (2024). *Is measuring blockchain transactions per second stupid in 2024?* Fetched 18th March, 2024. URL: https://cointelegraph.com/magazine/blockchain-transactions-per-second-tps-stupid-big-questions/.

PolkaVM Project (2024). "PolkaVM/RISC0 Benchmark Results". In: Fetched 3rd April, 2024. URL: https://github.com/koute/risc0-benchmark/blob/master/README.md.

Saarinen, Markku-Juhani O. and Jean-Philippe Aumasson (Nov. 2015). *The BLAKE2 Cryptographic Hash and Message Authentication Code (MAC)*. RFC 7693. DOI: 10.17487/RFC7693. URL: https://www.rfc-editor.org/info/rfc7693.

Sadana, Apoorv (2024). *Bringing Polkadot tech to Ethereum*. Fetched 18th March, 2024. URL: https://ethresear.ch/t/bringing-polkadot-tech-to-ethereum/17104.

Sharma, Shivam (2023). *Ethereum's Rollups are Centralized*. URL: https://public.bnbstatic.com/static/files/research/ethereums-rollups-are-centralized-a-look-into-decentralized-sequencers.pdf.

Solana Foundation (2023). *Solana data goes live on Google Cloud BigQuery*. Fetched 18th March, 2024. URL: https://solana.com/news/solana-data-live-on-google-cloud-bigquery.

Solana Labs (2024). *Solana Validator Requirements*. Fetched 18th March, 2024. URL: https://docs.solanalabs.com/operations/requirements.

Stewart, Alistair and Eleftherios Kokoris-Kogia (2020). "Grandpa: a byzantine finality gadget". In: *arXiv preprint arXiv:2007.01560*.

Tanana, Dmitry (2019). "Avalanche blockchain protocol for distributed computing security". In: *2019 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*. IEEE, pp. 1–3.

Thaler, Justin (2023). "A technical FAQ on Lasso, Jolt, and recent advancements in SNARK design". In: Fetched 3rd April, 2024. URL: https://a16zcrypto.com/posts/article/a-technical-faq-on-lasso-jolt-and-recent-advancements-in-snark-design/.

Wikipedia (2024). *Fisher-Yates shuffle: The modern algorithm*. URL: https://en.wikipedia.org/wiki/Fisher%5C%E2%5C%80%5C%93Yates_shuffle%5C#The_modern_algorithm.

Wood, Gavin (2014). "Ethereum: A secure decentralised generalised transaction ledger". In: *Ethereum project yellow paper* 151, pp. 1–32.

Yakovenko, Anatoly (2018). "Solana: A new architecture for a high performance blockchain v0. 8.13". In.