

Generation publication

This specification records the implementation argument for local generation publication, installation, authenticated snapshot transfer and bounded retention. R3.3 and R3.4 are closed for the scope bound by docs/releases/2026-09-25.typ; dated observations below retain their original test boundaries. The public `sqlodin compact NODE.json` command performs serialized maintenance on a stopped voter; the network service follows published generations at startup.

Representation and assumptions

The root consensus database contains a FULL-durability catalog transaction naming one generation. The pointer name has a SHA-256 checksum, and the root identity changes to a catalog identity on first publication. Older binaries must reject that identity. Each generation has an application database and an independent, voter-local consensus database. The local base contains the complete canonical certificate and its chosen seal slot. Recovery requires the identical seal in the retained chosen suffix.

The assumptions are the existing non-Byzantine voter model, authenticated peers, SQLite's atomic FULL transactions, successful file and directory durability barriers, and storage that does not silently roll back a completed durable transaction. A checksum detects accidental corruption; it is not an adversarial signature. The process-kill tests do not certify physical power-loss behavior of the device.

Preservation argument

Let P be the certified snapshot prefix and A the source's applied prefix. A private candidate starts with the verified application image at P . Its local journal preserves every source promise, vote and chosen record above P in source order, plus the maximum global promise. The copy verifies the entire source chain before completion and constructs a fresh checked chain for the retained records. The source's reserved-ID high-water mark is copied before the candidate acquires a usable identity. No donor acceptor identity or donor promises are substituted.

Replaying the retained chosen suffix advances the candidate from P to A . Request outcomes, session fences and transaction revision come from the snapshot and that replay. A chosen suffix value cannot disappear because the copy filters only slot-scoped records at or below P . An accepted value above P remains accepted, and a lower later ballot remains fenced by the preserved promises. A stored integer prefix without its full certificate and matching chosen seal is insufficient to open the generation.

Publication occurs only after the replacement opens successfully and its file names and containing directory are synchronized. The stable root lock spans the switch. Before the catalog transaction commits, restart selects the unchanged old generation. After the commit, restart selects the new generation. Failure to validate the selected generation stops startup; recovery never tries an older generation as a fallback. The source host is poisoned before attempting the catalog commit, so an ambiguous commit result cannot allow subsequent writes on the old state.

A published host carries the active read ticket. A ticket whose marker was retired is displaced and must obtain a new quorum barrier. Pending writes with retired slot evidence retry the original request identity; the application session fence prevents duplicate effects. Later SQLite transaction revision checks remain intact. Publication retains old files; separate inventory-based retirement is described below. A request below the local retained prefix records pending snapshot work instead of treating absent retired records as storage corruption.

Executable correspondence

Model action / invariant	Implementation
Stage	build_generation, generation_copy_suffix, store_generation_base
Sync	FULL database commits, successful reopen, sync_directory before publication
Publish	write_generation catalog transaction in compact_store
Crash / Restart	open_store, read_generation, load_generation_base, validate_generation_seal, checked journal recovery
AcceptedSuffix / PromiseFence	Local suffix copy and upstream ledger_replay_fold
RequestIDs	generation_publish_identity preserves _sqlodin_ids
PublishedDurable / AcknowledgedPrefix	Complete replacement validation before catalog commit

GenerationCatalog.tla checks two generation changes, two acknowledged entries and one still-accepted future value. It deliberately assumes correct certificate and consensus evidence; their agreement arguments live in the other models. Four negative controls omit publication ordering, the accepted value, the global promise, or reserved IDs. Each must produce its named invariant violation. This finite model is an executable check of the publication argument, not a compiler refinement proof or a liveness proof for all multi-master histories.

The Odin regressions exercise two publication/restart cycles, suffix replay, reserved IDs, a future accepted vote, lower-ballot rejection, root identity refusal, and corruption of the selected base. The Linux crash probe stops at fifteen actual maintenance boundaries and the controller sends SIGKILL. Recovery checks both the acknowledged row and its retry fence, the global promise, the future vote, and the selected old/new generation. Evidence is retained under benchmarks/results/verification-20260924/.

Receiving a generation from another voter

An authenticated peer offers a normal chosen seal value and a canonical candidate image description. generation_seal_valid binds the full certificate to the local configuration and engine, checks the seal slot range, and requires the candidate's key to match. Receipt transport does not import donor promises or donor ID state.

Let the recipient's applied prefix be A and the offered certified prefix be $P > A$. Under consensus agreement, each already-applied choice through A is also in the certified prefix through P . Installing the verified image therefore preserves acknowledged effects and their outcome/session fences. Records above P , the global promise and ID reservations are copied from the recipient's own durable journal. The authenticated chosen seal is added if it is not already locally recorded; a conflicting local chosen seal is rejected. The same catalog publication order then applies. The recipient resumes above the installed prefix and fetches gaps in the retained suffix through ordinary Paxos catch-up.

The service admits one incoming image, transfers at most 1 MiB per chunk, and bounds peer frames to 2 MiB while retaining the existing client frame limit. A shared 1 MiB scratch buffer avoids allocating a new transfer buffer for every chunk. TLS calls remain bounded to the transport's 64 KiB limit. The receiver checks offsets, length and the complete SHA-256 before installation; the worker-independent installation path also verifies SQLite integrity, logical contents and the certificate key. An incoming transfer excludes starting a local snapshot capture. Available-space checks reserve

room for the received image, replacement application, local journal and 256 MiB of free space; I/O failures must still be handled because other processes can consume space after that check.

Partial, unadvertised downloads can be discarded. Once an install attempt may have published its catalog transaction, cleanup retains the image even if the outcome is uncertain. This distinction prevents disconnect/shutdown cleanup from removing an image named by a possibly committed generation.

Live build and bounded delta

`begin_compaction` pins a separate read transaction at source journal sequence q and verifies its chain head against the serialized live host. The worker owns this connection, the copied configuration/frontiers and the private output. It never reads the live host concurrently. Initial copy, integrity checking and reopening of the candidate occur on that worker while the source continues voting and applying.

After the worker finishes, each owner turn copies at most 128 journal records and 1 MiB of encoded journal payload, then applies at most one contiguous chosen entry. This yields between SQL transactions whose small journal records can expand into large data writes. The smaller local step changes scheduling, not the private application-prefix invariant or the publication predicate below. Original sequence/chain checks precede rehashing into the candidate's new chain. Every new global promise and every slot-scoped record above the certified prefix is preserved. The source cursor and candidate cursor are distinct; they must not be compared as though compaction kept their numeric sequence IDs equal.

The catalog can switch only when the verified source cursor/head equals the live source sequence/head and the candidate application has caught up. The final serialized step copies the latest reserved-ID high-water mark, restores the Paxos node from the resulting local ledger, and uses the same durable publication path. Discarding an unpublished job leaves the source authoritative. The focused test adds 140 chosen entries plus a future vote, global promise and ID reservation after the worker's pinned cut; the replacement and restart preserve them all.

Format-5 services default to automatic maintenance. The default trigger is a 256 MiB consensus tail or 15 minutes of dirty history. One reachable voter initiates the ordered barrier; every receiving voter captures that same application prefix. This scheduling preference confers no consensus authority. Every voter with a locally retained certified image can compact in the background. Configuration maintenance: "manual" explicitly disables automatic initiation/publication for operator-controlled maintenance and the offline-command test fixtures.

The generation descriptor also binds its image name, complete certificate, image description, seal slot and exact predecessor name with a checksum. It records the actually active predecessor, not a directory selected by sorting names; an aborted private build must never become the predecessor by accident. Corrupting this descriptor fails startup closed. Superseded-file deletion and the retained-history admission cap are implemented; the retirement rules below and `specs/image-retirement.typ` define their obligations.

Owned generation retirement

`create_generation_directory` records a FULL catalog inventory entry before creating a private generation directory. It refuses an existing path before inventory insertion; the stable root guard serializes supported writers. This makes interrupted private builds discoverable without treating every similarly named directory as owned storage. Inventory names carry checksums. The supported store directory is private to its owner; an external process that rewrites that directory or its catalog is outside the model.

`retire_generation` rechecks the active catalog pointer and excludes both its current generation and the exact predecessor from the checked base descriptor. It declines work during compaction or transfer. It processes at most one inventoried directory, opens directories without following symlinks, obtains both database locks, and unlinks only explicitly named database/lock files. A held lock produces backpressure. Unowned directories are never candidates; unexpected files in a retired directory remain and prevent removing that directory. Corrupt inventory stops retirement before deletion.

Filesystem deletion and parent-directory synchronization precede the FULL transaction that forgets the inventory entry. A crash before that transaction leaves an entry that can retry missing files idempotently; no active or predecessor file is eligible. `GenerationRetirement.tla` checks 889 states including abandoned builds, publication, deletion, directory durability, forgetting inventory and crashes. Four negative controls remove current, predecessor, ownership or directory-sync protection and must violate their respective invariant. The model assumes the already-checked generation contents and checks this retirement boundary; it is not a new proof of Paxos agreement.

Local tests cover multiple generations, abandoned builds, unrelated directories, preexisting-name collisions, held database locks and corrupt inventory. Linux passes three SIGKILL boundaries: after removing files, after syncing the directory, and before committing inventory removal. Three native online cycles also pass with a missing voter, its catch-up and full restart. The first crash fixture incorrectly expected an accepted slot outside the protocol window; its failures remain, and the corrected fixture verifies the in-window vote before SIGKILL. Snapshot-image retirement, original-root-file cleanup and the aggregate retained-history cap are covered by `specs/image-retirement.typ` and the closed R3.4 release evidence.

Bounded transfer progress evidence

Peer wire version 2 uses the existing lossless journal packing for complete mutation bytes, including unused tails and IEEE floating-point bits. It changes no Paxos value equality or SQL semantics. The handshake rejects incompatible wire versions. Slow-peer queue congestion drops retransmittable consensus work instead of resetting TLS; streaming starts after a receiver fetch, so an ignored offer cannot suppress traffic.

The native Linux test passes two compacted generations, a 2.28 MiB image, a 128 KiB/s test link with an injected disconnect, continued healthy-quorum writes and full restart. The earlier unpacked failures are retained. The live-maintenance test passes two online publication cycles with an absent voter, its return and full restart. Its Linux scheduling build explicitly uses a three-second test interval; the production default interval is checked separately. These are functional checks, not R6 throughput results.