Deterministic Replay: Monoid Homomorphism and Induction Proof

Definition 0.1 (State space). *Let* $\mathbb{B} = \{0, 1\}$. *Let* Σ *denote the system state space (finite or otherwise well-typed).*

Definition 0.2 (Events, event monoid). *Let* $\mathcal{T}$ *be a finite set of* transition labels *and let* I *be the set of* input tokens *(including recorded external responses, binding/router identifiers, timestamps, etc.). Define the event set*

$$\mathcal{E} := \mathcal{T} \times I \times \mathcal{M},$$

where $\mathcal{M}$ *is an environment metadata space. The free monoid* $\mathcal{E}^*$ *consists of all finite sequences (strings) of events with concatenation* $\cdot$ *and empty word* ε.

Definition 0.3 (Endomorphism monoid). *Let* $\mathrm{End}(\Sigma)$ *denote the monoid of endomorphisms on* Σ *under composition* $\circ$, *with identity* id_Σ.

Definition 0.4 (Per-event semantics). *For each transition label* $\tau \in \mathcal{T}$ *define a* pure *transition function*

$$\phi_\tau : \Sigma \times I \times R \times \mathcal{M} \to \Sigma,$$

where R *denotes a deterministic random-output domain (the PRNG output space). For an event* $e = (\tau, i, \eta) \in \mathcal{E}$ *and a PRNG sample* $r \in R$ *the per-event endomorphism is*

$$F(e; r) : \Sigma \to \Sigma, \qquad F(e; r)(\sigma) := \phi_\tau(\sigma, i, r, \eta).$$

Definition 0.5 (Telemetry record). *A telemetry record is a tuple* $R = (g, s, w, \Delta)$ *where*

- g *is a GUID trace identifier,*
- s *is a cryptographic seed (integer),*
- $w = e_1 e_2 \dots e_n \in \mathcal{E}^*$ *is the ordered event log,*
- Δ *is an optional set of snapshot state(s) (checkpoint(s)).*

Associated to seed s *and a deterministic PRNG is a sequence of outputs* $(r_1, r_2, \dots)$.

Assumption 0.6 (Completeness of recorded nondeterminism)**.** *All sources of nondeterminism present in the original execution are either:*

1. *explicitly recorded in w or Δ, or*
2. *deterministically derivable from the seed s via the PRNG.*

Assumption 0.7 (Purity of replay semantics)**.** *During replay, the binding implementations used to compute ϕ_τ are pure functions of their explicit inputs: prior state, input token, PRNG output, and environment metadata. External side effects are replaced by recorded responses or deterministic stubs.*

Definition 0.8 (Extension of F to sequences)**.** *Fix a record $R = (g, s, w, \Delta)$ and PRNG outputs $(r_1, r_2, \ldots, r_n)$. For event e_j define $F(e_j) := F(e_j; r_j) \in \mathrm{End}(\Sigma)$. Extend F multiplicatively to sequences by*

$$F(e_1 e_2 \cdots e_n) := F(e_n) \circ F(e_{n-1}) \circ \cdots \circ F(e_1).$$

By convention, $F(\varepsilon) := \mathrm{id}_\Sigma$.

Lemma 0.9 (Monoid homomorphism)**.** *The map*

$$F : (\mathcal{E}^*, \cdot, \varepsilon) \longrightarrow (\mathrm{End}(\Sigma), \circ, \mathrm{id}_\Sigma)$$

defined above is a monoid homomorphism: for all $u, v \in \mathcal{E}^*$,

$$F(u \cdot v) = F(v) \circ F(u), \qquad F(\varepsilon) = \mathrm{id}_\Sigma.$$

Proof. Immediate from the definition. The empty word maps to identity by definition. Concatenation $u \cdot v$ corresponds to the sequential application of event endomorphisms; by definition of composition ordering we have $F(u \cdot v) = F(v) \circ F(u)$. Hence F preserves the monoid structure. □

Theorem 0.10 (Deterministic Replay — formal statement)**.** *Let $R = (g, s, w, \Delta)$ be a telemetry record satisfying the stated assumptions. Let $w = e_1 e_2 \cdots e_n$ and let $(r_1, \ldots, r_n)$ be the PRNG outputs deterministically generated from seed s according to the agreed indexing convention. Then there exists a deterministic replay procedure* $\mathrm{Replay}(R)$ *that reconstructs the unique state sequence*

$$\sigma_0 \xrightarrow{e_1} \sigma_1 \xrightarrow{e_2} \cdots \xrightarrow{e_n} \sigma_n,$$

where $\sigma_j = F(e_1 \cdots e_j)(\sigma_0)$ and σ_0 equals the snapshot in Δ (if provided) or the canonical initial state.

Proof. We prove by induction on $n = |w|$ that replay reconstructs the same states as the original execution.

Base case ($n = 0$). If $w = \varepsilon$ then $F(w) = \mathrm{id}_\Sigma$ and the replayed state sequence is the singleton $\{\sigma_0\}$ where σ_0 is the snapshot in Δ (or the canonical initial state). Equality with the original trivially holds.

Inductive hypothesis. Assume for some $k \geq 0$ that for any telemetry record whose event log length is k the replay procedure reconstructs the original states $\sigma_0, \sigma_1, \ldots, \sigma_k$ satisfying

$$\sigma_j = F(e_1 \cdots e_j)(\sigma_0) \quad \text{for } 0 \leq j \leq k.$$

Inductive step ($k \to k+1$). Consider a record with $w = e_1 \cdots e_k e_{k+1}$ and associated PRNG outputs $(r_1, \ldots, r_{k+1})$. By the inductive hypothesis, replaying the prefix $e_1 \cdots e_k$ produces the state $\sigma_k = F(e_1 \cdots e_k)(\sigma_0)$. For step $k+1$ the replay computes

$$\sigma'_{k+1} := F(e_{k+1}; r_{k+1})(\sigma_k),$$

where $F(e_{k+1}; r_{k+1})$ is the per-event endomorphism computed with the recorded input token, environment metadata, and the PRNG output r_{k+1}. By the assumptions:

- the PRNG output r_{k+1} equals the original run's r_{k+1} (determinism of PRNG given seed s),
- the input token and environment metadata match the original values (they are recorded),
- the per-event function $\phi_{\tau_{k+1}}$ is pure and deterministic for the given arguments.

Thus $\sigma'_{k+1} = \sigma_{k+1}$, where σ_{k+1} is the state the original execution produced after event e_{k+1}. This completes the inductive step.

Therefore, by mathematical induction, replay reconstructs the same state sequence for all n. $\square$

Corollary 0.11 (Closure under concatenation). *If $R_1 = (g_1, s_1, w_1, \Delta_1)$ and $R_2 = (g_2, s_2, w_2, \Delta_2)$ are replayable traces and the concatenation $w_1 \cdot w_2$ is provided with a consistent PRNG indexing scheme (or equivalent combined seed), then the concatenated trace is replayable and*

$$F(w_1 \cdot w_2) = F(w_2) \circ F(w_1).$$

Remark 0.12 (Practical caveats)**.** *The theorem is only valid under the stated assumptions. In practice, ensure:*

- *explicit recording of external responses and wall-clock values if they influence ϕ_τ;*
- *deterministic handling (or recording) of concurrency interleavings (vector clocks, global counters, or total-order logs);*
- *PRNG outputs are indexed and/or recorded so the mapping event index $\mapsto r_j$ is unambiguous.*

Failure to observe these requirements yields partial or no reproducibility.