TECHNICAL WHITEPAPER WP-02 · VIRTUAL MCP (VMCP), TOKEN ECONOMY & ZERO-TRUST AGENT RUNTIME

Virtual MCP (vMCP) & Token Economy Architecture: Compressing Enterprise APIs by Up to 99.7% with Progressive MCP 2026-07-28 & Arazzo 1.1

Arcverin overcomes the 50-tool LLM accuracy cliff by compiling Arazzo 1.1 workflows and Virtual MCP (vMCP) profiles. Every call runs behind encrypted Vault credential injection and SHA-256 hash-chained policy journals.

137k -> 477
-99.7% PROMPT TOKEN BLOAT
Exact o200k_base BPE compression on 520-endpoint enterprise catalog.

O(1) Context
PROGRESSIVE MCP 1-2-3
arcverin_discover -> arcverin_describe -> arcverin_invoke meta-tools.

23 Scenarios
12-MODE FAULT SIMULATOR
Proves HTTP 504 retry idempotency & async 202 polling locally.

Zero Secrets
IN LLM PROMPT CONTEXT
XChaCha20-Poly1305 Vault injection + SHA-256 hash-chained OPA journal.

1. Shattering the 50-Tool Accuracy Cliff: Reproducible Synthetic Reference Benchmarks

Even with provider prompt caching, sending hundreds of raw endpoints degrades tool selection and increases latency. The five synthetic, vendor-neutral OpenAPI 3.1 fixtures below (downloadable at arcverin.com/docs/fixtures/reference-benchmarks.zip) measure exact o200k_base MCP tool-description payloads-not full prompts, cache discounts, or customer invoices:

Reference API Profile (OpenAPI 3.1)	Raw 1:1 MCP (o200k_base)	Full-Catalog Progressive 1-2-3	Single-Tool Curated & Baseline
Enterprise Cloud Fabric (520 ops)	137,012 tokens/turn	~380 upfront + 477 on-demand (-99.4% · all 520 ops)	477 tokens (-99.7%) · 42/100 AR0 baseline
Storage Control Plane (132 ops)	37,885 tokens/turn	~380 upfront + 502 on-demand (-97.7% · all 132 ops)	502 tokens (-98.7%) · 43/100 AR0 baseline
Transactional Mail (46 ops)	3,473 tokens/turn	~380 upfront + 383 on-demand (-78.0% · all 46 ops)	383 tokens (-89.0%) · 67/100 AR1 baseline
Checkout & Settlement (14 ops)	1,123 tokens/turn	~380 upfront + 223 on-demand (-46.3% · all 14 ops)	223 tokens (-80.1%) · 65/100 AR1 baseline
Telemetry Alerts Service (4 ops)	333 tokens/turn	Direct curated optimal (micro profile)	233 tokens (-30.0%) · 84/100 AR1 baseline

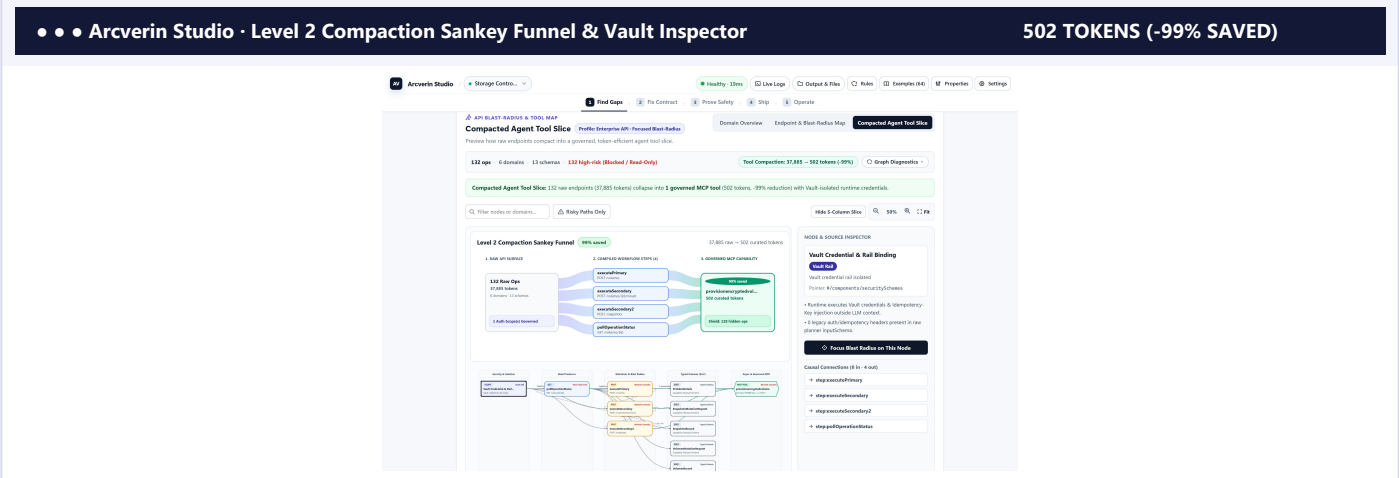


Figure 1: Level 2 Compaction Sankey Funnel collapsing 132 raw Storage Control Plane operations (37,885 tokens) into 1 governed MCP capability (502 tokens, -99%) while shielding 128 hidden operations and isolating Vault credentials.

2. Progressive MCP 1-2-3 Discovery & 23-Scenario Stateful Fault Simulation

Arcverin Runner (--mode progressive) serves three constant-size meta-tools over MCP 2026-07-28 Streamable HTTP (/mcp) and stdio (with --compat-mcp for Claude Desktop, Cursor, and VS Code). In Enterprise, serve-vmcp combines multiple providers under strictest-tier autonomy inheritance:

Progressive Meta-Tool	Payload Complexity	Architectural Role in the Agent Execution Loop
1. arcverin_discover	O(1) compact index (~380 upfront tokens)	Returns capability slugs, business intent summaries, and Autonomy Tier badges matching the agent query.
2. arcverin_describe	On-demand single-tool schema	Loads the strict JSON Schema, preconditions, and blast-radius limits only for the selected capability.
3. arcverin_invoke	Validated execution envelope	Enforces input credential guards, external OPA policy checks, and runtime grants before invocation.

- Proven Retry Safety: When a 504 drops a committed charge response, naive agents double-bill. Arcverin's Arazzo 1.1 workflow runs a pre-retry GET lookup to guarantee exactly-once execution.

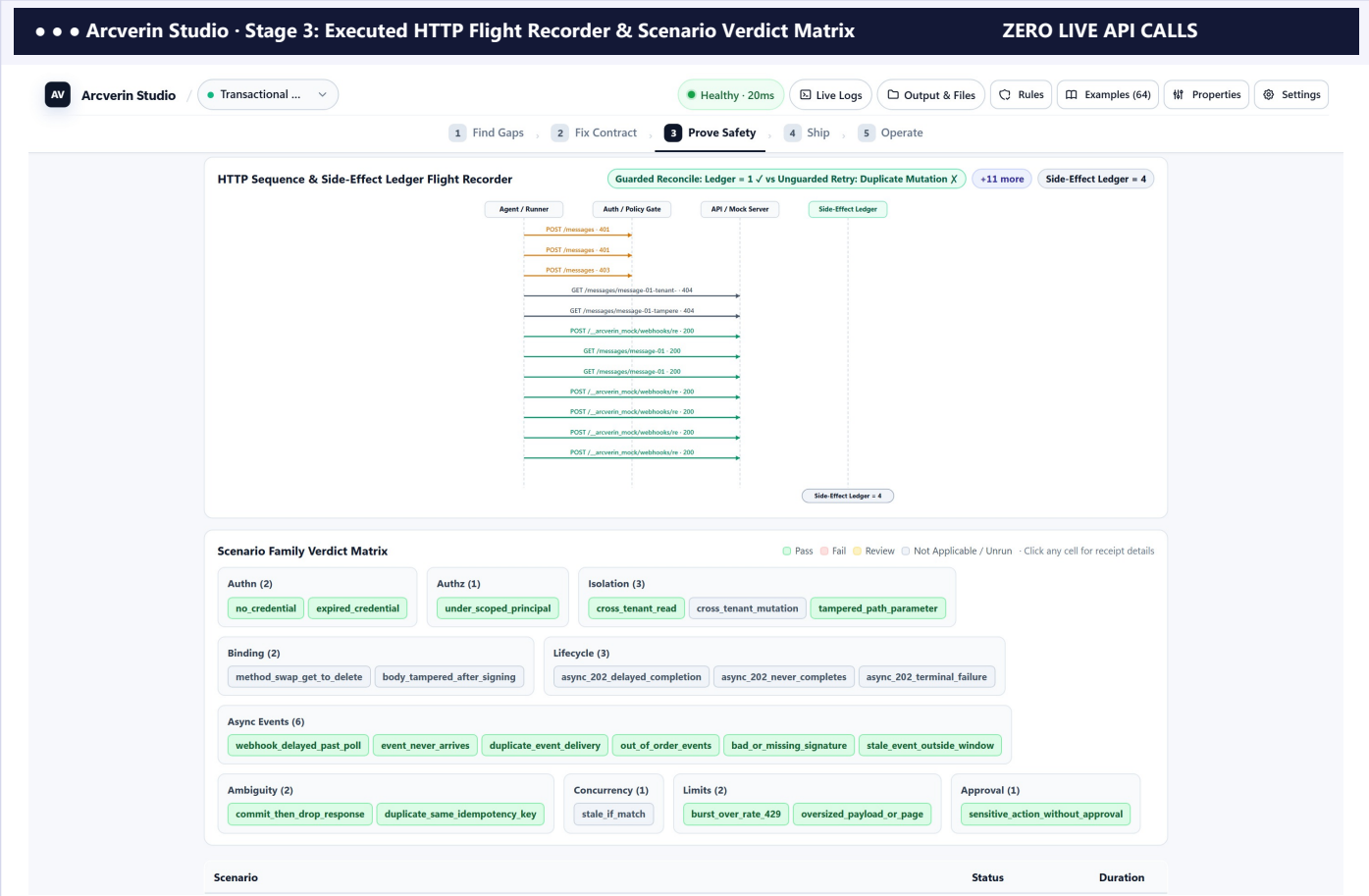


Figure 2: Stage 3 (3. Prove Safety) testing Arazzo 1.1 and vMCP capabilities across 23 stateful scenarios, 10 scenario families, and 12 fault modes (504 timeouts, 429 rate limits, async 202 polling, and cross-tenant denials).

3. Zero-Trust Data-Plane Runtime, Inline Vault & 1-Click AI Agent Onboarding

In Stage 5 (5. Operate), host-locked worker packages (.dwpack) follow a 3-Step Readiness Checklist: 1. Connect Service & Bind Vault -> 2. Check Readiness & Run -> 3. Publish to AI Agents. Secrets inject inside arcverin-runtime from the Argon2id + XChaCha20-Poly1305 encrypted Vault (.arcverinvault) and never enter LLM context. Every call passes external OPA checks and logs to policy-journal.jsonl (verified via arcverin-runner verify-journal).

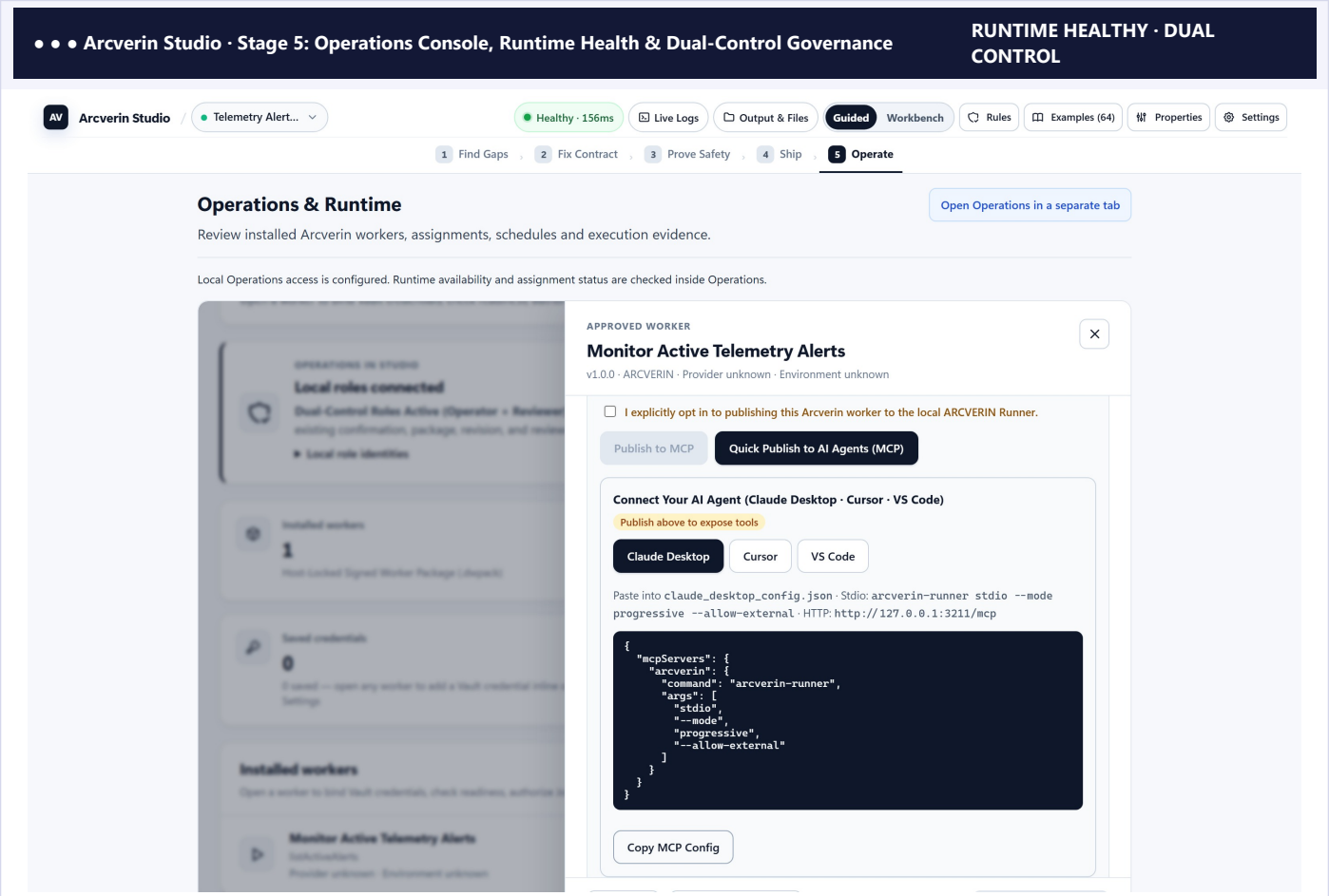


Figure 3: Stage 5 (5. Operate) Enterprise Operations console running promoted capabilities with dual-control Operator and Reviewer governance, Vault credential bindings, and cryptographic execution receipts.

Explore the Arcverin Technical Documentation & CLI Reference at https://www.arcverin.com/docs/ →

https://www.arcverin.com/docs/

Protocol & measurement note: Arcverin Runner supports stateless MCP 2026-07-28 Streamable HTTP (/mcp) and stdio. Token metrics measure MCP tool-description payloads on synthetic reference specs, not provider invoices or audited customer savings. Evaluate on one diagnostic workflow offline with a $99 30-Day Personal Pass and model prompt caching at https://www.arcverin.com/pricing/#breakeven.