Bilko Security & Engineering Decisions (Observability Program)

Bilko Security & Engineering Decisions (Observability Program)

This page covers security findings, live fixes, engineering decisions, and arming prerequisites for the Bilko observability/self-healing program. It is a companion to the Program Overview (MC #103328) page.

---

1. CRITICAL Security Finding F7 and Fix (MC #103369 to #103371)

What Securion found (MC #103369, 2026-06-10)

Securion reviewed `POST /auth/test/session` on **bilko-demo-api.alai.no** (live trial API). Verdict: **MOVE_OFF_PROD**. Source: /tmp/evidence-103369/verification.json

ID	Severity	Finding
F7	CRITICAL	createTestSession() accepted arbitrary email. No whitelist. Leaked secret mints owner JWT for any registered prospect in bilko-demo-db.

ID	Severity	Finding
F6	HIGH	Endpoint on live customer trial surface (app.bilko.cloud / bilko-demo-api.alai.no).
F3	HIGH	Generic auth bucket 200 req/min on demo - no endpoint-specific rate-limit.
F2	MEDIUM	Non-constant-time string compare (Kotlin !=). Timing side-channel defect.
F5	MEDIUM	RLS isolates E2E tenant but F7 expanded blast radius to all demo users.
F1	HIGH	Endpoint always registered at startup; 404 only when secret absent.
F4	LOW	Secret strength operator-dependent; no enforced entropy or rotation schedule.

Fix deployed (MC #103371, 2026-06-10)

PR #330 merged. Deploy run 27272876257 success. Revision bilko-api-demo-00179-wdz at 100% traffic. Source: /tmp/evidence-103371/verification.json

Remediation	What changed
Email whitelist (F7 closed)	testEmail hard-checked against BILKO_E2E_TEST_EMAIL env var. Non-matching returns HTTP 403 BILKO-AUTH-003. DB lookup never reached.
Constant-time compare (F2)	Replaced Kotlin != with MessageDigest.isEqual() on both testEmail and secret.
Dedicated rate-limit (F3)	5 req/min sub-bucket for /auth/test/session, independent of AUTH_RATE_LIMIT_PER_MINUTE.
Sentry audit (F3)	Sentry.captureMessage on any secret mismatch - structured event, not just warn log.
PERMANENT gate	F7-WHITELIST-GATE test added to ciam-auth-lifecycle.spec.ts (PR #332). Deploy pipeline now 3/3 tests; blocks on whitelist regression.

Live proof (Proveo independent verification)

Source: /tmp/verify-103371/verification.json and /tmp/evidence-103371/verification.json

Probe	Expected	Got	Result

valid secret + non-whitelisted email (attacker@example.com)	403	403	PASS
valid secret + seeded E2E email	200	200	PASS
wrong secret + seeded email	401	401	PASS

Deploy run 27274186928 (post-gate PR): success, 3/3 passed, F7-WHITELIST-GATE active.

Residual open findings (fix before first real paying customers)

- **F4:** No enforced entropy minimum or rotation schedule for BILKO_E2E_TOKEN_SECRET.
- **F1/F6:** Endpoint permanently registered when secret is set; on customer trial surface. Migrate E2E to ephemeral no-traffic revision before meaningful real-customer volume.

---

2. CIAM E2E Blocking Gate (MC #103365)

Design

A Playwright spec (apps/e2e/tests/ciam-auth-lifecycle.spec.ts) runs as a mandatory blocking step in the GCP deploy pipeline (continue-on-error: false). Targets bilko-demo-api.alai.no specifically because stage masks RLS bugs (documented lesson). Source: /tmp/evidence-103365/verification.json

Token-seed pattern: spec calls POST /auth/test/session with the 64-char BILKO_E2E_TOKEN_SECRET from GCP Secret Manager to obtain a real Bilko JWT, then exercises 9 authenticated steps:

1. POST /auth/test/session - 200 (JWT minted)
2. GET /auth/me - 200 (email + RLS identity confirmed)
3. GET /settings/users - 200 (tenant isolation: 1 user in org)
4. PUT /settings {vatNumber} - 200 (supplier OIB seeded)
5. POST /contacts - 201 (authenticated write)
6. POST /invoices - 201 (invoice create)
7. GET /invoices/{id} - 200 (RLS tenant read-back)
8. POST /auth/logout - 204 (refresh token revoked)

9. POST /auth/mobile/refresh (stale token) - 401 (revocation proven)

Two-sided proof (Proveo)

- **Green:** all 9 steps pass in 1800ms. Deploy proceeds.
- **Red:** bad secret returns 401. Gate cannot be bypassed.

Current state post-#103371: 3/3 tests (original 9-step spec + F7-WHITELIST-GATE). Permanent blocking gate in every future deploy.

---

3. First Real Incident (503, 2026-06-10)

Root cause (verified, /tmp/evidence-incident-503/finding.json)

Transient Cloud Run scale-from-zero + revision cutover blips. NOT a code bug. No outage.

- bilko-api-demo has minScale=0 (scale-to-zero). Revisions 00186 and 00187 deployed during the 20:36-20:43 UTC window (PR #330/#332 merges via CD).
- 503 latency: 11-16ms (immediate infra-level reject, no Kotlin stack trace).
- An interleaved 200 on the same endpoint at 20:36:44 confirmed service was otherwise healthy.
- Health endpoint bilko-demo-api.alai.no/api/v1/health returned 200 throughout.
- Alert pipeline worked: error-tracking alert fired and was investigated correctly.

Actions taken

- Alert threshold tuned: >0 errors to **>3 errors/5 min**. Single deploy-cutover blips no longer page.
- min-instances=1 deferred until real paying customers (cost trade-off).

Sentinel blind-spot exposed and fixed (MC #103420)

The incident revealed the error-count policy (ID 2342970117877340710, "Bilko API Demo - Backend ERROR log rate") was not in the Sentinel's hardcoded ALERT_POLICIES list. The Sentinel was silently blind to it. Source: /tmp/evidence-103420/verification.json

Fix: live gcloud alpha monitoring policies list dynamic discovery with 5-minute on-disk cache (~/system/state/bilko-sentinel-policy-cache.json) and embedded fallback. Sentinel now evaluates 9 policies / 13 conditions per cycle. Fallback WARN log emitted if gcloud fails (no silent blind spots). Cache hit log confirms the target policy on every cycle.

---

4. Engineering Decisions

John (AI Director) on CEO delegation. Decision 2 grounded in Kelsey Hightower (SRE) advisory consult. Canonical file: /Users/makinja/business/ALAI-Holding-AS/products/Bilko/docs/infrastructure/DECISIONS-observability-2026-06-10.md

Decision 1 - E2E test-session endpoint location

ACCEPT-WITH-HARDENING. Keep on demo. Overrides pre-fix MOVE_OFF_PROD verdict.

Why: F7 (the CRITICAL basis of MOVE_OFF_PROD) is fully remediated. Residual controls are strong (64-char secret, constant-time compare, 5/min rate-limit, Sentry audit, F7-WHITELIST-GATE). Demo is where RLS coverage lives; moving to stage forfeits the gate's purpose. No real customers yet; LOW residual risk.

Future trigger: before meaningful real-customer volume, migrate to a dedicated ephemeral no-traffic Cloud Run revision. Rotate secret on schedule.

Decision 2 - Tier-1 auto-remediation

Do NOT enable Tier-1 now. Stay Tier-0 (read-only). Earn promotion via the bar below.

Grounding: recent agent-caused production incidents (IAM wipe; F7 hole introduced by an agent's own change); Tier-0 is hours old, zero signal calibration; Cloud Run rollback is not migration-aware on a financial system.

Kelsey Hightower (SRE) consult

Independent conclusion: automated remediation is justified only after (a) calibrated track record of correct proposals, (b) migration-safe rollbacks, and (c) action set signed off by a human engineer. The promotion bar operationalizes this.

Promotion bar: Tier-0 to Tier-1 (ALL must be true)

1. 30+ days Tier-0 live AND 20+ evaluated proposals (extend window until 20 proposals).
2. Proposal false-positive rate below 5% (human verdict within 24h; "root cause wrong" or "fix would worsen" = FP).
3. ZERO proposals that would have caused a secondary incident if auto-executed.

4. At least 1 ground-truth case: Tier-0 diagnosed correctly, human executed that exact fix, it resolved the incident.
5. Schema-deploy coupling audit complete + deploy manifest records migrations per revision (rollback safety).
6. Synthetic Entra-CIAM auth probe added to observability (bad rollback can break auth silently).
7. Revisions that are themselves rollbacks are tagged (never roll back to a known-bad revision).
8. Tier-1 action set signed off by a human engineer (not just CEO).

Tier-1 permitted actions (enforced, not advisory)

- Permitted only: roll back to N-1, scale min-instances 0 to 1, Slack escalation.
- Never-automate (must live at IAM, not just code): any IAM/policy, any Cloud SQL op, any secret, any DNS/LB/network, rollback older than N-1, action during in-flight deploy or protected business window.
- Pre-fire (ALL must be true): alert firing 5+ min; LLM confidence above calibrated threshold; target revision healthy 10+ min; no migration in bad revision; no prior action in last 60 min; 3-min human-ack window elapsed.
- Circuit breakers: max 2 actions/24h; auto-disable after any failed remediation; pre-action IAM-diff vs known-good snapshot; single-writer lock; audit log written before execution.

---

5. Tier-1 Arming Prerequisites (MC #103439, Securion MC #103436)

Source: /tmp/evidence-103436/verification.json. Securion verdict: HARDENING_REQUIRED.

Shadow is structurally inert today. Dual barrier confirmed: (1) handleIncident() returns before execution block when MODE==='shadow'; (2) executeRollback and executeScaleFloor throw as their first statement in shadow. Two independent mechanisms. No mutation path exists in shadow.

Hard blockers before ack or auto mode

Finding	Severity	Required fix before arming
F5 - Ledger has no integrity protection	HIGH	bilko-sentinel-tier1-ledger.jsonl is writable and unsigned. Forged human_verdict entries could satisfy the promotion bar. Fix: HMAC-sign each row; verify on read. BLOCK ack/auto until done.

Finding	Severity	Required fix before arming
F7 - SA actual GCP IAM roles unverified	MEDIUM	alai-cli-deployer SA project-level bindings not verified at review time. In shadow: must hold only monitoring.viewer + logging.viewer + run.viewer. For auto: roles/run.developer scoped to bilko-api-demo and bilko-web-demo only (resource-level condition). Must NOT hold cloudsql.*, iam.*, secretmanager.*, or dns.* roles.
F4 - Ack poll unwired; future identity risk	INFO (future MEDIUM)	When ack poll is wired: approver must be verified against hardcoded Slack user ID allowlist. Channel membership not sufficient. Require thread_ts match to prevent cross-incident approval.

Additional findings (non-blocking for shadow)

Finding	Severity	Summary
F6 - IAM snapshot bootstrap window	MEDIUM	Snapshot can be deleted to force re-baseline. Seal after first write; alert on deletion/recreation.
F2 - Object.freeze({MODE}) is a no-op	MEDIUM	Misleading call; remove or replace with comment. MODE is immutable by JS const semantics in strict mode.
F8 - Gate 8 inconsistent with Gate 4	LOW	Gate 8 warns-and-passes when deploy manifest absent; Gate 4 blocks. Align to block.
F3 - Module integrity not checked at load	LOW	Add SHA-256 startup integrity check for Tier-1 module path.
F9 - Tier-1 missing execute bit	LOW	chmod +x /Users/makinja/system/tools/bilko-sentinel-tier1.js (cosmetic, no runtime impact).

Current state: Tier-1 running in shadow mode. All proposals logged to ~/system/logs/bilko-sentinel-tier1-ledger.jsonl. Calibration clock started. Review at 30 days / 20 proposals.

Revision #3
Created 2026-06-11 19:49:25 UTC by John
Updated 2026-07-12 20:03:46 UTC by John