

# Email-Agent Ingest Gap Postmortem (2026-05-23) — MC #101887

# Email-Agent Ingest Gap Postmortem (2026-05-23) — MC #101887

## TL;DR

Email-agent.js silently dropped SEEN-flagged messages for 9+ days (2026-05-14 → 2026-05-23) due to `HIMALAYA_DISABLED=1` forcing a fallback code path that filtered `{ seen: false }`. This caused 17 missed messages across 5 accounts, including 2 paying-client-class emails (Asmir Merdžanović SEO work, cynthia.li medical contact). Fixed by replacing SEEN filter with date-range + DB dedup. Backfilled all missed messages, added audit tool, deployed hourly monitoring LaunchAgent.

## Incident Timeline (UTC)

- **2026-05-14** → Newest alai/INBOX DB row before gap
- **2026-05-23 13:26** → Asmir Merdžanović email arrives at alai/INBOX uid=6, server already flags SEEN
- **2026-05-23 18:49 (CEST 20:49)** → John boot detects DB:0 IMAP:1 gap during inbox-pending sweep
- **2026-05-23 ~21:00** → MC #101887 created, gate cleared, ST1-ST4 dispatched
- **2026-05-23 ~21:22** → ST3 backfill complete, 17 messages ingested
- **2026-05-23 ~21:26** → ST6 (this documentation) initiated

## Root Cause

**File:** `/Users/makinja/system/daemons/email-agent.js`

**Original code (lines 638-644, pre-fix):** The `fetchUnseenLegacy` function used `{ seen: false }` as its IMAP fetch filter, which translates to an IMAP `SEARCH UNSEEN` query. Any message already flagged `\Seen` on the server (e.g., by mobile client, webmail, or Outlook auto-marking) was invisible to this query.

```
const messages = client.fetch(  
  { seen: false }, // ← PROBLEM: excludes SEEN messages  
  { uid: true, envelope: true }  
);
```

### Trigger chain:

1. LaunchAgent plist `/Users/makinja/Library/LaunchAgents/com.john.email-agent.plist` sets `HIMALAYA_DISABLED=1` as hard environment variable
2. This forces all accounts to fall back to `fetchUnseenLegacy` instead of the safer `fetchAllRecent` path (which was introduced in MC #6832 to solve exactly this class of problem)
3. When `alem@alai.no` is also accessed via mobile/web client, incoming messages are auto-flagged `\Seen` before daemon's next 5-minute cycle
4. Daemon runs every 5 minutes, sees 0 unseen, logs "alai: 0 unseen envelopes fetched", and continues — no alarm, no visibility

**Why it went undetected:** The daemon logs showed normal execution (no errors, no timeouts), just consistently 0 results for the alai account. The pattern looked like "no new email" rather than "email silently dropped."

**Fixed code (lines 638-684, post-fix):** Replaced `{ seen: false }` with date-range filter `{ since: }` + DB deduplication by UID set lookup:

```
// MC #101887 fix: SEEN filter caused 9-day gap. Switched to date-range + DB dedup.  
const lookbackDays = parseInt(process.env.EMAIL_AGENT_LOOKBACK_DAYS || '7', 10);  
const sinceDate = new Date(Date.now() - lookbackDays * 24 * 60 * 60 * 1000);  
  
// Load existing UIDs for this account from DB to enable dedup  
const db = emailInbox.getDb();  
const existingUids = new Set(  
  db.prepare("SELECT message_id FROM emails WHERE account = ?").all(boxLabel).map(r => {  
    const m = r.message_id.match(/-uid-(\d+)/);  
    return m ? parseInt(m[1], 10) : null;  
  }).filter(Boolean)  
);
```

```
// Fetch envelopes only – date-range avoids SEEN-flag blind spot
const messages = client.fetch(
  { since: sinceDate }, // ← FIX: fetch all messages in date range
  { uid: true, envelope: true }
);

for await (const msg of messages) {
  // Dedup: skip if UID already in DB
  if (existingUids.has(msg.uid)) continue;
  // ... insert logic
}
```

# Impact Assessment

- **Total missed:** 17 messages across 5 accounts in 30-day lookback window
- **Paying-client-class misses:**
  - **Asmir Merdžanović** (asmirmc@gmail.com) — "Potrebne informacije." re: 2 new SEO clients (alai/INBOX uid=6, john/INBOX uid=134)
  - **cynthia.li@jamrmed.com** (Shenzhen Jamr Medical) — "New contact-Shenzhen Jamr" (john/INBOX uid=114)
- **Informational/system misses:** 13+ messages including Google Cloud alerts, TLDR newsletters, GitHub notifications, Cloudflare alerts
- **Duration:**
  - alai account: **9 days** (2026-05-14 → 2026-05-23)
  - alem account: **11+ days** (2026-05-13 → ongoing, separate IMAP connection failure)
- **Accounts affected:** alai (1 missed), dev (3 missed), john (13 missed); info/alem had no IMAP-side new messages in window (alem broken for separate reason)

# Fix Applied

1. **Code fix:** `~/system/daemons/email-agent.js` lines 638-725 — replaced `{ seen: false }` with `{ since: }` + DB dedup via UID set lookup (idempotent, safe for overlapping runs)
2. **Backfill:** 17 missed messages ingested via `~/system/tools/email-backfill-from-audit.js` — used audit JSON as source of truth, patched subject/from metadata in 14 cases where IMAP envelope fetch failed (tool is idempotent, safe to re-run)
3. **New audit tool:** `~/system/tools/email-imap-db-audit.js` — enumerates IMAP UIDs vs DB UIDs per account+folder for configurable N-day window, outputs JSON diff with missed UID samples

4. **Monitoring LaunchAgent:** `~/Library/LaunchAgents/com.alai.email-ingest-monitor.plist`  
+ wrapper `~/system/tools/email-ingest-monitor.sh` — runs hourly, executes audit tool,  
fires Slack #exec alarm when `total_missed > 0`

## Remaining Open Items (NOT yet fixed)

- **alem@alai.no IMAP connection broken** since 2026-05-13 — credentials load OK from Vault, but server rejects connection with "Command failed" (no detailed error exposed by ImapFlow). Needs separate MC task for IMAP diagnostics + credential rotation test.
- **Monitor LaunchAgent NOT auto-loaded** — file exists at correct path, but launchctl does not auto-load new plists without manual intervention. CEO must run: `launchctl load -w ~/Library/LaunchAgents/com.alai.email-ingest-monitor.plist` (permission constraint, cannot be automated without sudo/TCC access).
- **HIMALAYA\_DISABLED env flag still active** in `com.john.email-agent.plist` — the fix made `fetchUnseenLegacy` safe, but ideally the himalaya path should be vetted and re-enabled to reduce IMAP connection load.
- **3 john/INBOX uids (61, 69, 71) backfilled with placeholder metadata** — IMAP `fetchOne` returned "Command failed" for envelope fetch, so subject/from are "(no subject)" / empty. These need separate IMAP range-fetch backfill to recover actual metadata from server.

## Reproduction / Detection Commands

```
# Detect the gap
node ~/system/tools/email-imap-db-audit.js
cat /tmp/alai/email-ingest-gap/imap-db-diff-30d.json | jq .summary

# Trigger monitor manually
launchctl kickstart -k gui/$(id -u)/com.alai.email-ingest-monitor

# Re-run backfill (idempotent)
node ~/system/tools/email-backfill-from-audit.js

# Check daemon status
launchctl list | grep email
tail -100 ~/system/logs/email-agent.log

# Test audit in verbose mode
node ~/system/tools/email-imap-db-audit.js --verbose
```

# Lessons / Preventive Actions

- **Silent skips are P0:** Any code path that filters IMAP results without an alarm when count drops to 0 unexpectedly = future incident. The daemon should have emitted a warning when alai account returned 0 unseen for >7 consecutive cycles (35+ minutes) given its historical delivery rate.
- **SEEN flag is not under our control:** Any mobile/web client can pre-read messages and set `\Seen` before the daemon polls. The ingest pipeline must not assume `UNSEEN = unread-by-us`. Date-range + DB dedup is the only reliable pattern.
- **Audit > trust:** ST2 audit revealed a 2nd unrelated paying-client miss (cynthia.li) we wouldn't have known about without full IMAP-vs-DB enumeration. Periodic audits should be part of email-agent health checks.
- **Fallback paths are production code:** The `fetchUnseenLegacy` path was treated as a temporary fallback but ran in production for weeks/months with `HIMALAYA_DISABLED=1`. All fallback paths must have equal quality gates (logging, alarms, safety checks) as primary paths.
- **Monitoring must be fail-closed:** The new monitor LaunchAgent is valuable, but it's not yet loaded (manual step required). For future daemons, the deploy checklist must verify LaunchAgent is loaded AND firing test alarms.

## Related Artifacts

- **MC:** #101887 (this fix), supersedes #101886
- **Triggering email evidence:** `/tmp/alai/john-boot-20260523T1441/asmir-search.log`
- **RCA:** `/tmp/alai/email-ingest-gap/root-cause.md`
- **Audit JSON:** `/tmp/alai/email-ingest-gap/imap-db-diff-30d.json`
- **Backfill log:** `/tmp/alai/email-ingest-gap/backfill-run.log`
- **Monitor runs:** `/tmp/alai/email-ingest-gap/monitor-runs.log`
- **Code fix:** `~/system/daemons/email-agent.js` lines 638-725
- **Tools created:**
  - `~/system/tools/email-imap-db-audit.js` (audit)
  - `~/system/tools/email-backfill-from-audit.js` (backfill)
  - `~/system/tools/email-ingest-monitor.sh` (monitor wrapper)
- **LaunchAgent:** `~/Library/LaunchAgents/com.alai.email-ingest-monitor.plist`

## Technical Details

Missed Messages Breakdown (30-day window, all accounts)

Account	Folder	Missed Count	Sample UIDs	Notes
alai	INBOX	1	6	Asmir email re: SEO clients
dev	INBOX	3	4, 7, 11	Google Cloud Logging alerts
john	INBOX	13	61, 69, 71, 72, 79, 80, 82, 83, 88, 99, 102, 114, 134	Mix: GitHub, TLDR, Cloudflare, cynthia.li, Asmir
info	INBOX	0	—	No new IMAP messages in window
alem	INBOX	N/A	—	IMAP connection broken, cannot audit

## Backfill Execution Summary

- **Total inserted:** 17 (first run)
- **Total patched:** 14 (second run — corrected subject/from metadata)
- **Total skipped:** 3 (UIDs 61, 69, 71 had no audit sample metadata, kept placeholder)
- **Tool runs:** 3 (idempotent, each run refined metadata)

## Monitor Configuration

**LaunchAgent:** `com.alai.email-ingest-monitor`

- **Schedule:** Hourly (StartCalendarInterval)
- **Command:** `~/system/tools/email-ingest-monitor.sh`
- **Output:** `~/system/logs/email-ingest-monitor.log`
- **Alarm channel:** Slack #exec
- **Trigger condition:** `total_missed > 0` in audit JSON
- **Status:** Plist exists, NOT loaded (manual load required)

## Sign-off

**Documented by:** Skillforge (ALAI agent)

**Date:** 2026-05-23

**MC Task:** #101887 ST6

**Status:** Fix deployed, backfill complete, monitoring deployed (pending manual load)

---

Revision #1

Created 2026-05-23 21:52:37 UTC by John

Updated 2026-05-23 21:52:37 UTC by John