Incident Response Playbook

Incident Response Playbook

Purpose: When an alert fires, what to do immediately. No research, no debugging — just triage → diagnose → escalate/fix.
Audience: John (primary), Alem (fallback), FlowForge/CodeCraft agents (delegated fixes)
Last updated: 2026-04-19 (SENTINEL Sprint)

Alert Triage Matrix

When you see this alert → do this immediately:

Alert Message	Severity	First Action	Diagnostic Commands	Escalate If
"⚠ PUBLIC SURFACE DOWN: alai.no"	P0	Verify tunnel + origin	`curl -I https://alai.no` `launchctl list | grep cloudflared` `tail -50 ~/Library/Logs/ALAI/cloudflared-error.log`	Down > 5 min → Alem directly
"⚠ PUBLIC SURFACE DOWN: lumiscare.alai.no"	P0	Check Docker containers	`docker ps | grep lumiscare` `docker logs lumiscare-web` `curl http://localhost:4001`	Container stopped → restart, if fail → Alem
"⚠ PUBLIC SURFACE DOWN: getdrop.no"	P0	Check Vercel deployment	`curl -I https://getdrop.no` `vercel ls drop-landing` Vercel dashboard	Vercel outage or DNS → Alem
"⚠ PUBLIC SURFACE DOWN: docs/vault/sign.alai.no"	P0	Check Azure VM + Docker	`ssh alai-admin@4.223.110.181` `docker ps` `systemctl status docker`	VM down or out of disk → Alem
"⚠ PUBLIC SURFACE DOWN: snowit.ba"	P1	Check DNS + domain expiry	`dig snowit.ba` `whois snowit.ba | grep -i expiry`	Domain lapsed → Alem (billing decision)

Alert Message	Severity	First Action	Diagnostic Commands	Escalate If
"[SENTINEL ALERT] ops-watchdog"	P1	Check which service died	`launchctl list | grep -E "alai|john"` View plist logs: `tail -50 ~/Library/Logs/ALAI/<service>.log`	Critical service down > 10 min → escalate
"Slack bot DOWN — email fallback active"	P0	Restart slack-bot	`launchctl kickstart -k gui/$(id -u)/com.john.slack-bot` `node ~/system/tools/slack.js send ops "test after restart"`	Restart fails → Alem (all alerts via email until fixed)
"Email DLQ size > 5 entries"	P1	Check vault + bw CLI	`bw unlock --check` `curl -I https://vault.alai.no` `wc -l ~/system/logs/email-dlq.jsonl`	Vault down > 1 hr OR DLQ > 20 → Alem
"TLS cert expiry: in 7 days"	P1	Verify cert date + renew	`echo | openssl s_client -connect <domain>:443 -servername <domain> 2>/dev/null | openssl x509 -noout -enddate` Cloudflare dashboard → SSL/TLS	Cert renew fails → Alem (public outage risk)
"[HM-ALERT] agent: "	P2	Check HiveMind source	`sqlite3 ~/system/databases/hivemind.db "SELECT * FROM events WHERE kind='alert' ORDER BY timestamp DESC LIMIT 5"`	Agent loop detected OR repeated fail → investigate
"[INTAKE] source: "	P2	Review MC task auto-created	`node ~/system/tools/mc.js list --status pending` Check intake source (email/form/Slack)	Spam OR malformed intake → tune classification
"[NO-EVIDENCE] Task # done"	P3	Check sidecar + re-validate	`tail ~/system/logs/task-outcomes-pending-evidence.jsonl` `node ~/system/tools/mc.js show <id>`	Builder repeatedly skips evidence → Proveo re-validation

Common Incidents (From 30-Day Ledger)

1. Drop Landing Page 502 (Happened: Apr 7, 9)

Symptoms: BetterStack alert "Drop Landing Page DOWN" (HTTP 502 or DNS timeout)

Diagnosis:

```
# 1. Check Vercel deployment status
curl -I https://getdrop.no
vercel ls drop-landing

# 2. Check DNS
dig getdrop.no

# 3. Check Vercel dashboard
# Open: https://vercel.com/basic-as/drop-landing
# Look for: "Deployment Failed" or "Domain Configuration Error"
```

Fix:

- If Vercel deployment failed → redeploy: `cd ~/projects/drop-landing && vercel --prod`
- If DNS misconfigured → Cloudflare dashboard → DNS records → verify CNAME points to cname.vercel-dns.com
- If Vercel platform outage → check https://www.vercel-status.com → notify Alem (no fix available, wait)

Escalate if: Down > 10 min AND revenue event (customer trying to pay) → Alem directly via phone +47 404 74 251

Post-incident: Update Drop incident log at ~/system/evidence/drop-incidents.md

---

2. LumisCare 502 (Happened: Apr 19 — silent for hours)

Symptoms: "⚠ PUBLIC SURFACE DOWN: lumiscare.alai.no" (HTTP 502 — connection refused :4001)

Diagnosis:

```
# 1. Check Docker containers
docker ps | grep lumiscare
# Expected: lumiscare-web (port 4001), lumiscare-api (port 8090), lumiscare-ollama (port 4003)
```

```
# 2. If missing, check stopped containers
docker ps -a | grep lumiscare

# 3. Check logs
docker logs lumiscare-web --tail 50
docker logs lumiscare-api --tail 50
```

Fix:

```
# If containers stopped, restart
cd ~/projects/lumiscare
docker compose up -d

# Verify
curl -I http://localhost:4001
curl -I http://localhost:8090

# Check cloudflared tunnel routing
curl -I https://lumiscare.alai.no
```

Escalate if: Container restart fails with error OR OOM killed repeatedly → Alem (may need Azure migration for LumisCare)

Root cause notes: LumisCare Docker containers were stopped on Apr 19 for unknown reason (no crash logs, Mac uptime 47d). Possibly manual `docker stop` or OOM. Needs Docker health check monitoring.

---

3. Slack Bot SIGKILL (Happened: unknown date — killed ALL alerts)

Symptoms: No alerts in #ops for days, launchctl shows `com.john.slack-bot` with exit -9, email fallback activates

Diagnosis:

```
# 1. Check if bot is dead
launchctl list | grep slack-bot
# If PID = "-" and Status = "-9" → killed
```

```
# 2. Check memory usage history (if available)
# OOM kill leaves no direct trace, but check system.log
log show --predicate 'eventMessage contains "slack-bot"' --info --last 1h

# 3. Test Slack API reachability
curl -I https://slack.com/api/api.test
```

Fix:

```
# 1. Restart bot
launchctl kickstart -k gui/$(id -u)/com.john.slack-bot

# 2. Verify alive
launchctl list | grep slack-bot
# Should show non-zero PID, LastExit = 0

# 3. Test alert delivery
node ~/system/tools/slack.js send ops "sentinel: slack-bot restarted after SIGKILL"

# 4. Check if alert appears in #ops within 5 sec
```

Escalate if: Restart fails OR bot dies again within 1 hour → Alem (memory leak investigation needed, may need rewrite)

Prevention: After sprint, ops-watchdog monitors slack-bot itself. If bot dies, email fallback activates automatically.

---

4. Email Intake Pipeline Dead (Happened: Feb 25 — silent 53 days)

Symptoms: "Email DLQ size > 5 entries" OR manual discovery (email-agent.log not updated in days)

Diagnosis:

```
# 1. Check email-agent daemon
launchctl list | grep email-agent
# If LastExit != 0 → daemon crashed

# 2. Check vault connectivity
```

```
bw unlock --check
# If fails → vault session expired or Vaultwarden down

# 3. Check Vaultwarden Docker (Azure VM)
ssh alai-admin@4.223.110.181
docker ps | grep vaultwarden
# If missing → container stopped

# 4. Check DLQ size
wc -l ~/system/logs/email-dlq.jsonl
```

Fix:

```
# If vault session expired (ETIMEDOUT):
# 1. Restart Vaultwarden on Azure VM
ssh alai-admin@4.223.110.181 "cd ~/docker/vaultwarden && docker compose up -d"

# 2. Unlock vault locally
bw unlock
# Enter master password (from Alem or ~/system/config/.vault-session if cached)

# 3. Restart email-agent
launchctl kickstart -k gui/$(id -u)/com.john.email-agent

# 4. Replay DLQ
bash ~/system/tools/email-dlq-replay.sh

# 5. Verify DLQ cleared
wc -l ~/system/logs/email-dlq.jsonl
# Should be 0 or 1
```

Escalate if: Vaultwarden container won't start OR bw unlock fails with password error → Alem (may need Bitwarden master password reset)

Prevention: After sprint, email-agent writes failed emails to DLQ. Alert fires if DLQ > 5 entries. Vault downtime no longer causes silent email loss.

---

5. MC Dashboard 502 (Happened: Apr 19)

Symptoms: "⚠ PUBLIC SURFACE DOWN: mc.alai.no" (HTTP 502 — connection refused :3030)

Diagnosis:

```
# 1. Check mc-dashboard daemon
launchctl list | grep mc-dashboard
# If LastExit = 1 → daemon crashed

# 2. Check local port
curl -I http://localhost:3030
# If connection refused → service not running

# 3. Check logs
tail -50 ~/system/logs/mc-dashboard.log
```

Fix:

```
# 1. Restart daemon
launchctl kickstart -k gui/$(id -u)/com.john.mc-dashboard

# 2. Verify local
curl -I http://localhost:3030
# Should return 200

# 3. Verify public (through cloudflared tunnel)
curl -I https://mc.alai.no
```

Escalate if: Restart fails with "missing node_modules" OR "port 3030 in use" → CodeCraft fix (dependency or port conflict issue)

---

6. Cloudflared Tunnel Down (SPOF — ALL 26 hostnames die)

Symptoms: Multiple BetterStack alerts simultaneously (alai.no + lumiscare.alai.no + docs + vault + sign + getdrop all down within 1 min)

Diagnosis:

```
# 1. Check cloudflared daemon
launchctl list | grep cloudflared
# If PID = "-" → tunnel dead
```

```
# 2. Check error log
tail -100 ~/Library/Logs/ALAI/cloudflared-error.log

# 3. Check Cloudflare Zero Trust dashboard
# Open: https://one.dash.cloudflare.com
# Navigate: Networks → Tunnels → "alai-main-tunnel"
# Look for: "Tunnel Disconnected" or "No Healthy Connectors"
```

Fix:

```
# 1. Restart tunnel
launchctl kickstart -k gui/$(id -u)/com.john.cloudflared

# 2. Wait 10 seconds for reconnect

# 3. Verify public endpoints
for url in https://alai.no https://lumiscare.alai.no https://getdrop.no; do
  echo -n "$url: "
  curl -sfL --max-time 10 -o /dev/null -w '%{http_code}\n' "$url"
done
```

Escalate if:

- Restart fails → Alem immediately (ALL public surfaces down)
- Mac Studio hardware issue (power, network) → Alem (may need physical reboot or Azure failover)
- Tunnel reconnects but hostnames still down → check Cloudflare dashboard for DNS propagation delay (can take 2-5 min)

CRITICAL: This is the single biggest SPOF in ALAI infrastructure. Phase 2 sprint (deferred) will add secondary tunnel on Azure VM.

---

7. Azure VM SSH Timeout (Happened: Apr 19)

Symptoms: `ssh alai-admin@4.223.110.181` hangs or "Connection timed out"

Diagnosis:

```
# 1. Check VM reachability
ping -c 3 4.223.110.181
```

```
# 2. Check Azure portal
# Open: https://portal.azure.com
# Navigate: Resource groups → alai-support → vm-alai-support
# Look for: "VM Status: Stopped" or "Networking issues"

# 3. Check NSG rules
# Azure portal → vm-alai-support → Networking → Inbound port rules
# Verify: Port 22 (SSH) is allowed from your IP
```

Fix:

- If VM stopped → Azure portal → vm-alai-support → Start
- If NSG blocking → Add inbound rule: Port 22, Protocol TCP, Source: Your IP, Priority 100
- If VM running but SSH hangs → Restart VM (Azure portal → Restart)

Escalate if: VM won't start OR restart fails → Alem (Azure billing issue OR quota exceeded)

Impact: If vm-alai-support is down, these services die: BookStack (docs.alai.no), Vaultwarden (vault.alai.no), Documenso (sign.alai.no). BetterStack will fire 3 simultaneous alerts.

---

8. TLS Cert Expiry Warning (bilko-demo expires Jun 22, 2026)

Symptoms: "TLS cert expiry: bilko-demo.basicconsulting.no in 7 days" (alert fires 7 days before lapse)

Diagnosis:

```
# 1. Verify cert expiry date
echo | openssl s_client -connect bilko-demo.basicconsulting.no:443 -servername bilko-
demo.basicconsulting.no 2>/dev/null | openssl x509 -noout -enddate

# 2. Check Cloudflare SSL settings
# Open: https://dash.cloudflare.com
# Select domain: basicconsulting.no
# Navigate: SSL/TLS → Edge Certificates
# Look for: "Universal SSL" status + expiry date
```

Fix:

- If Cloudflare Universal SSL → automatic renewal (no action needed, Cloudflare renews 30 days before expiry)
- If custom cert (uploaded to Cloudflare) → renew manually:
  1. Generate new cert via Let's Encrypt: `certbot certonly --manual -d bilko-demo.basicconsulting.no`
  2. Upload to Cloudflare: SSL/TLS → Edge Certificates → Upload Custom Certificate
  3. Verify: `curl -I https://bilko-demo.basicconsulting.no` (check `Expires:` header in cert)

Escalate if: Cloudflare renewal fails OR custom cert upload fails → Alem (public outage imminent within 7 days)

Escalation Path

Incident Type	Escalate To	When	Contact Method
Public surface down > 5 min	Alem	Immediately	Slack DM + Phone +47 404 74 251
Revenue event (Drop payment failing)	Alem	Immediately	Phone first, Slack second
Security breach or suspicious activity	Alem + Securion	Immediately	Slack #ops + Email alembasic@gmail.com
PI licenca revoked or legal issue	Alem	Within 1 hour	Phone + Email
Azure VM / billing / quota issue	Alem	Within 30 min	Slack + Email (needs Azure portal access)
Mac Studio hardware (power/network)	Alem	Immediately	Phone (may need physical access)
Cloudflared tunnel down > 10 min	Alem	Immediately	ALL public surfaces offline
Builder agent repeated failures (3+ in 1 hour)	Petter Graff (specialist)	Within 1 hour	Slack #ops → delegate fix
Slack bot down (messenger dead)	John (self-fix)	Within 5 min	Email fallback active, restart bot
Daemon down (non-critical)	John (self-fix)	Within 15 min	Investigate + restart or ticket for agent

CRITICAL: If John (orchestrator) is offline, all P0 alerts route to Alem via email (alembasic@gmail.com). Check inbox every 15 min during incidents.

Runbook References

For step-by-step daemon restart procedures, see:

- SENTINEL Reliability Sprint Overview — System architecture after sprint
- Alert Routing — Channel routing table (Slack #ops vs email vs digest)
- Email Intake Revival — Vault ETIMEDOUT fix + DLQ replay
- BetterStack Setup — How to add new monitors

For safe daemon unload/reload:

```
# Unload (stop daemon, keep plist)
launchctl unload -w ~/Library/LaunchAgents/com.john.<service>.plist

# Load (start daemon from plist)
launchctl load -w ~/Library/LaunchAgents/com.john.<service>.plist

# Kickstart (restart without unload/load)
launchctl kickstart -k gui/$(id -u)/com.john.<service>
```

Playbook maintained by: Skillforge (SENTINEL Task 7)
Last incident review: 2026-04-19 (30-day ledger: 17 incidents, 2 with alerts, 15 silent)
Next review: After Phase 2 sprint (secondary tunnel + 12 dead daemons fixed)

Revision #5
Created 2026-04-19 08:32:02 UTC by John
Updated 2026-06-21 20:03:11 UTC by John