OpenClaw Security Report: February 2026 — ClawHub Malware, Google Suspensions, and Critical Fixes

February 28, 2026 · by Mark Bartlett

Download this guide as PDF

> **Quick Answer:** February 2026 was OpenClaw's worst month — 10 CVEs patched, a supply chain attack poisoning 12% of ClawHub, Google permanently banning paid users, and the project's creator leaving for OpenAI. The ClawJacked attack proved that even local-only installations could be hijacked through a single malicious webpage. If you're running anything older than 2026.2.26, update immediately — the full timeline, every fix, and what to do next is below.

Related: OpenClaw Security Guide · January 2026 Security Report · ClawHub Security Alert · Best OpenClaw Alternatives · OpenClaw Setup Guide

Contents

February 2026 was the month everything hit at once. Seventeen security fixes across eight releases. A supply chain attack that poisoned 12% of ClawHub. Google permanently banning paid subscribers who used OpenClaw with Gemini. The project's creator leaving for OpenAI. And a new attack class — ClawJacked — that let any malicious website silently hijack local agents.

This report covers every confirmed security event from February 2026. For January's CVEs and the start of the crisis, see our January 2026 report. For the detailed technical breakdown of each fix shipped in the 2026.2.x releases, see our existing hardening guide.

Summary table

CVEs disclosed or patched in February 2026

CVE	Description	CVSS	Affected Versions	Fix Version	Disclosed
CVE-2026-25593	Unauthenticated local RCE via WebSocket config.apply	8.4	< 2026.1.20	2026.1.20	Feb 4
CVE-2026-25475	Arbitrary file read via MEDIA: path extraction	High	< 2026.1.30	2026.1.30	Feb 4
CVE-2026-26324	SSRF guard bypass via IPv4-mapped IPv6	7.5	< 2026.2.14	2026.2.14	Feb 14
CVE-2026-26319	Missing Telnyx webhook authentication	7.5	< 2026.2.14	2026.2.14	Feb 14
CVE-2026-26322	SSRF in Gateway tool via gatewayUrl param	7.6	< 2026.2.14	2026.2.14	Feb 14
CVE-2026-26329	Path traversal in browser upload	High	< 2026.2.14	2026.2.14	Feb 14
CVE-2026-28466	Exec approval bypass via unsanitized fields	8.8	< 2026.2.14	2026.2.14	Mar 5

CVE	Description	CVSS	Affected Versions	Fix Version	Disclosed
CVE-2026-28453	TAR extraction path traversal	7.5	< 2026.2.14	2026.2.14	Mar 5
CVE-2026-28478	Webhook handler DoS via unbounded buffering	7.5	< 2026.2.13	2026.2.13	Mar 5
CVE-2026-28479	Sandbox cache poisoning via SHA-1 collisions	7.5	< 2026.2.15	2026.2.15	Mar 5

Non-CVE security events

Event	Severity	Date	Status
ClawHub supply chain attack (ClawHavoc)	Critical	Feb 2 (disclosed)	Partially resolved
ClawJacked — WebSocket agent hijacking	High	Feb 26	Resolved (2026.2.25)
Google account suspensions	High (user impact)	Feb 12-14	Closed won't-fix
Steinberger joins OpenAI	Governance	Feb 15	OpenClaw moves to foundation
17 security fixes in 2026.2.x releases	Mixed	Feb 1-28	Resolved

CVE-2026-25593: Unauthenticated local RCE

Severity	High (CVSS 8.4)
CWE	Command Injection
Affected versions	All versions before 2026.1.20
Fix	Version 2026.1.20
Disclosed	February 4, 2026
Advisory	GHSA-g55j-c2v4-pjcg

What happened

The `config.apply` WebSocket endpoint accepted a `cliPath` parameter without sanitization. An attacker on the same machine — or a malicious website connecting to the local gateway via WebSocket — could inject shell commands through this parameter. No authentication required.

The gateway process executes commands with whatever privileges it runs under. On most installations, that's the user's full account.

Impact

Local privilege escalation to full RCE. Any process or website that can reach the gateway port can execute arbitrary commands. This was fixed in 2026.1.20, but many users were still on older versions throughout February.

Sources: GitHub Advisory · SentinelOne

CVE-2026-25475: File read via MEDIA: path

Severity	High
CWE	Path Traversal
Affected versions	All versions before 2026.1.30
Fix	Version 2026.1.30
Disclosed	February 4, 2026
Advisory	GHSA-r8g4-86fx-92mq

What happened

The `isValidMedia()` function in `src/media/parse.ts` accepted arbitrary file paths — absolute paths, home directory paths, and directory traversal sequences. An agent could read any file on the system by outputting `MEDIA:/path/to/file`, sending the contents to whatever channel was active.

A prompt injection attack or malicious skill could make the agent output `MEDIA:~/.ssh/id_rsa` and exfiltrate private keys to a chat channel.

Sources: GitHub Advisory · SentinelOne

CVE-2026-26324: SSRF IPv6 bypass

Severity	High (CVSS 7.5)
CWE	SSRF
Affected versions	All versions before 2026.2.14
Fix	Version 2026.2.14 (hardened further in 2026.2.23 and 2026.2.26)
Disclosed	February 14, 2026

What happened

OpenClaw's SSRF guard blocked requests to localhost and private networks, but didn't canonicalize IPv4-mapped IPv6 addresses. Feeding the browser tool a URL containing `0:0:0:0:0:ffff:7f00:1` (which is `127.0.0.1` in IPv6 notation) bypassed the guard entirely. An attacker who could influence the URLs the agent visited could reach internal services, cloud metadata endpoints (AWS `169.254.169.254`), and anything running on localhost.

The fix in 2026.2.14 converts IPv4-mapped IPv6 to standard IPv4 before checking. Version 2026.2.23 added a breaking change: the browser SSRF policy now defaults to "trusted-network" mode, blocking internal access unless explicitly opted in.

Sources: NVD · CVEDetails · GitLab Advisory

CVE-2026-26319: Telnyx webhook auth missing

Severity	High (CVSS 7.5)
Affected versions	All versions before 2026.2.14
Fix	Version 2026.2.14

The optional `@openclaw/voice-call` extension's Telnyx webhook handler had no authentication. Anyone who could reach the webhook endpoint could inject events into the voice call flow.

Sources: NVD · Infosecurity Magazine

CVE-2026-26322: Gateway SSRF

Severity	High (CVSS 7.6)
Affected versions	All versions before 2026.2.14
Fix	Version 2026.2.14

The Gateway tool accepted a tool-supplied `gatewayUrl` parameter without sufficient restrictions. A malicious tool or prompt injection could redirect the gateway to connect to arbitrary internal or external endpoints.

Sources: SentinelOne · Infosecurity Magazine

CVE-2026-26329: Browser upload path traversal

Severity	High
Affected versions	All versions before 2026.2.14
Fix	Version 2026.2.14

Authenticated attackers could read arbitrary files from the Gateway host through path traversal in the browser upload handler.

Sources: SentinelOne · Infosecurity Magazine

CVE-2026-28466: Exec approval bypass

Severity	High (CVSS 8.8)
CWE	CWE-863 (Incorrect Authorization)
Affected versions	All versions before 2026.2.14
Fix	Version 2026.2.14

Disclosed	March 5, 2026

What happened

OpenClaw's `node.invoke` parameters contained internal approval fields that weren't sanitized. An authenticated client could inject approval control fields to bypass exec gating for `system.run` commands, executing arbitrary commands on connected node hosts — developer workstations and CI runners included.

This is the highest-severity CVE from the February fix cycle. CVSS 8.8 with network attack vector, low complexity, and full impact on confidentiality, integrity, and availability.

Source: RedPacket Security

CVE-2026-28453: TAR path traversal

Severity	High (CVSS 7.5)
CWE	CWE-22 (Path Traversal)
Affected versions	All versions before 2026.2.14
Fix	Version 2026.2.14

TAR archive extraction didn't validate entry paths. Crafted archives with `../../` sequences could write files outside the intended directory. Server-side deployments processing untrusted archives — like skill installation — were most exposed.

Source: RedPacket Security

CVE-2026-28478: Webhook DoS

Severity	High (CVSS 7.5)
CWE	CWE-770 (Resource Allocation Without Limits)
Affected versions	All versions before 2026.2.13

Fix	Version 2026.2.13

Webhook handlers buffered request bodies without byte or time limits. Remote unauthenticated attackers could send oversized JSON payloads or slow uploads to exhaust memory. Impact is availability-only — no data exposure, just a crash.

Source: RedPacket Security

CVE-2026-28479: Sandbox cache poisoning

Severity	High (CVSS 7.5)
CWE	CWE-327 (Broken Cryptographic Algorithm)
Affected versions	All versions before 2026.2.15
Fix	Version 2026.2.15

OpenClaw used SHA-1 to hash sandbox identifier cache keys for Docker and browser sandbox configurations. SHA-1 is deprecated and vulnerable to collision attacks. An attacker could craft colliding cache keys to cause one sandbox configuration to be misinterpreted as another, enabling unsafe sandbox state reuse.

Source: RedPacket Security

ClawJacked: WebSocket agent hijacking

Severity	High
Discovered by	Oasis Security
Disclosed	February 26, 2026
Fix	Version 2026.2.25 (within 24 hours of disclosure)

What happened

Oasis Security discovered that malicious websites could silently hijack locally running OpenClaw agents through a three-step attack chain:

1. **WebSocket connection**: JavaScript on a malicious page opens a WebSocket to localhost on the OpenClaw gateway port. Browser cross-origin policies don't block WebSocket connections to localhost.
2. **Password brute-force**: The gateway's rate limiter exempted localhost connections entirely. The script brute-forces the gateway password at hundreds of attempts per second.
3. **Auto-pairing**: Once authenticated, the script registers as a trusted device. The gateway auto-approved device pairings from localhost with no user prompt.

The attacker then has full agent control: executing commands, reading files, dumping configuration, and exfiltrating data. The victim only needs to visit the malicious page — no clicks, no downloads, no interaction.

Impact

Every OpenClaw instance with a gateway password (which is most of them) was vulnerable. The attack works even on properly configured local-only installations because it originates from the user's own browser.

The OpenClaw team fixed it within 24 hours. Version 2026.2.25 adds origin checks for browser WebSocket clients, password brute-force throttling on localhost, and blocks auto-pairing for non-Control-UI browsers.

Sources: Oasis Security · Security Affairs · The Hacker News

ClawHub supply chain attack

Severity	Critical
Disclosed	February 2, 2026 (Koi Security report)
Scale	341 malicious skills out of 2,857 audited (12% of registry)
Status	Partially resolved — malicious skills removed, vetting still weak

What happened

Koi Security audited the entire ClawHub registry and published their findings on February 2, 2026. The numbers: 341 malicious skills, 335 from a single coordinated campaign tracked as ClawHavoc, one attacker account responsible for 314 skills with roughly 7,000 downloads.

The skills disguised themselves as crypto wallets, YouTube tools, and Google Workspace integrations. They used "ClickFix" social engineering — burying malicious shell commands inside lengthy prerequisite instructions that users would copy-paste without reading.

The primary macOS payload was Atomic macOS Stealer (AMOS), harvesting browser credentials, keychains, Telegram data, SSH keys, and cryptocurrency wallets. The campaign targeted always-on machines — exactly the kind of setup OpenClaw users run.

By mid-February, subsequent scans found the problem had grown to over 800 malicious skills across an expanded registry of 10,700+. ClawHub's vetting at the time required only a one-week-old GitHub account to publish.

For the full breakdown, see our ClawHub security alert.

Sources: The Hacker News · eSecurity Planet · CyberPress

Google account suspensions

Severity	High (user impact)
Started	February 12-14, 2026
Status	Closed won't-fix (Feb 20)

What happened

Starting around February 12, Google permanently banned Antigravity/AI Ultra subscribers who routed Gemini requests through OpenClaw's OAuth integration. No warnings. No grace period. Entire Google accounts disabled — Gmail, YouTube, Workspace, everything.

The economics explain why: an AI Ultra subscriber paying $249.99/month could generate token usage through OpenClaw equivalent to $1,000-$3,600/month in API pricing. Google's automated systems flagged the traffic as "malicious usage" and shut accounts down at scale.

The GitHub issue (#14203) drew 29 comments from affected users. The maintainers closed it as “won’t fix” on February 20, stating that users are responsible for understanding provider terms of service. OpenClaw added warning messages but did not remove the Google Antigravity OAuth integration.

What to know

If you used OpenClaw with Google’s Gemini API through Antigravity OAuth at any point, your Google account may be at risk. Google offered no appeals process and no refunds for paid subscribers. Creator Peter Steinberger called the enforcement “pretty draconian.”

Sources: GitHub Issue #14203 · Gigazine · WinBuzzer · PCWorld

Steinberger joins OpenAI

Announced	February 15, 2026
Impact	Governance — project moves to open-source foundation

What happened

On February 15, Peter Steinberger — creator of OpenClaw and the developer behind the fastest-growing open-source project in GitHub history — announced he was joining OpenAI to “drive the next generation of personal agents.”

The timing was notable: Steinberger left days after infostealers from ClawHavoc hit 1,000+ installs, and in the middle of the most intense security crisis the project had faced. He reportedly shipped roughly 40 patches before departing.

OpenAI structured it as an acqui-hire — they hired Steinberger but did not acquire OpenClaw’s code or IP. The project transitions to an independent open-source foundation. OpenAI stated they would “continue to support” the project, though what that means in practice is unclear.

For the local AI community, the question is whether security maintenance will hold up without the original developer driving fixes. The February patch velocity — eight releases in a month — was driven by Steinberger and a small core team. Whether the foundation model can sustain that pace under continuous security pressure is an open question.

Sources: TechCrunch · UCStrategies · VentureBeat

February security fixes summary

Beyond the CVEs above, the 2026.2.x release series included these security-relevant fixes (detailed in our hardening guide):

Fix	Version	Severity
Sandbox symlink/hardlink escape	2026.2.23-26	High
Telegram DM authorization bypass	2026.2.24-26	High
Workspace @-path escape	2026.2.23	High
Attachment arbitrary file read	2026.2.23	High
Anthropic OAuth PKCE verifier leak	2026.2.25	High
Node exec approval hardening	2026.2.26	High
Cross-channel reply hijacking	2026.2.24	Medium
Session key duplication	2026.2.24	Medium
WebSocket auth hardening (3 fixes)	2026.2.25	Medium
Browser temp path escape	2026.2.26	Medium
Config include escape	2026.2.26	Medium
SSRF IPv6 multicast block	2026.2.26	Medium
HTTP security headers	2026.2.23	Low
Heartbeat DM blocking (breaking change)	2026.2.24	N/A
Docker namespace isolation (breaking change)	2026.2.24	N/A
SSRF trusted-network default (breaking change)	2026.2.23	N/A

Timeline

Date	Event
Feb 2	Koi Security publishes ClawHub audit: 341 malicious skills, ClawHavoc campaign exposed.
Feb 4	CVE-2026-25593 (local RCE) and CVE-2026-25475 (file read) formally disclosed. Patches already available in 2026.1.20 and 2026.1.30.

Date	Event
Feb 11	First reports of Google banning OpenClaw users who accessed Gemini via Antigravity OAuth.
Feb 12-14	Google ban wave intensifies. Paid AI Ultra subscribers ($250/mo) permanently banned, no refunds.
Feb 13	Version 2026.2.13 ships: webhook DoS fix (CVE-2026-28478).
Feb 14	Version 2026.2.14 ships: SSRF IPv6 bypass (CVE-2026-26324), Gateway SSRF (CVE-2026-26322), Telnyx auth (CVE-2026-26319), browser path traversal (CVE-2026-26329), exec approval bypass (CVE-2026-28466), TAR traversal (CVE-2026-28453). Six CVEs patched in a single release.
Feb 15	Peter Steinberger announces he's joining OpenAI. OpenClaw moves to open-source foundation.
Feb 15	Version 2026.2.15 ships: sandbox cache SHA-1 fix (CVE-2026-28479).
Feb 20	Google ban GitHub issue closed as won't-fix. Warning messages added, integration kept.
Feb 23	Version 2026.2.23 ships: SSRF trusted-network default (breaking), workspace path escape, attachment file read, config include escape.
Feb 24	Version 2026.2.24 ships: Telegram DM auth bypass, session key duplication, cross-channel reply hijacking, heartbeat DM blocking (breaking), Docker namespace isolation (breaking).
Feb 25	Version 2026.2.25 ships: ClawJacked fix (WebSocket hijacking), Anthropic OAuth PKCE leak, WebSocket auth hardening.
Feb 26	Version 2026.2.26 ships: Sandbox symlink/hardlink escape, browser temp path escape, SSRF multicast block, node exec hardening. Oasis Security publicly discloses ClawJacked.

What to do right now

1. Update to 2026.2.26 or later

```
npm update -g openclaw
# or
brew upgrade openclaw-cli

# Verify
openclaw --version
# Must show 2026.2.26 or later
# Run diagnostics
openclaw doctor --fix
```

2. Audit ClawHub skills

List every installed skill. Remove any you didn't explicitly choose. Check remaining skills against the ClawHub security alert. If you installed anything from ClawHub in January or February 2026, assume it may be compromised until verified.

3. Check Google account status

If you used OpenClaw with Google's Antigravity/Gemini API, verify your Google account is still active. If banned, Google's response has been no appeals, no refunds. Disconnect the Antigravity OAuth integration if still active.

4. Rotate credentials

If you ran any version older than 2026.2.14 at any point in February, rotate API keys, SSH keys, and any credentials the agent had access to. Multiple CVEs in this period allowed arbitrary file reads and command execution.

5. Review breaking changes

Three defaults changed in the 2026.2.x series:

- **SSRF policy**: Browser tool now blocks local network access by default
- **Heartbeat DMs**: Agent proactive messaging blocked by default
- **Docker namespace-join**: Container network sharing blocked by default

If any of these break your workflow, see our hardening guide for the opt-in configuration.

6. Consider alternatives

If the February security load gives you pause, we maintain a comparison of lighter alternatives. Several options avoid these architectural risks entirely.

The bigger picture

February 2026 exposed the fundamental problem with OpenClaw's design: an agent with shell access, file system access, and API key access will always have a catastrophic blast radius when any vulnerability is found. Every bug becomes an RCE. Every path traversal becomes credential theft. Every WebSocket issue becomes full agent takeover.

The maintainers responded aggressively — eight releases, seventeen fixes, three breaking changes that tightened permissive defaults. Steinberger reportedly shipped 40 patches before leaving for OpenAI. But the velocity of vulnerability discovery is outpacing the velocity of fixes. Ten CVEs in a single month, a supply chain attack poisoning 12% of the skill registry, and an entirely new attack class (ClawJacked) that renders local-only deployments vulnerable through the browser.

The project now faces a governance transition on top of everything else. Whether the open-source foundation can maintain this patch velocity without its creator is the open question heading into March 2026.

Related guides

- OpenClaw Security Guide — hardening your deployment
- January 2026 Security Report — the start of the crisis
- February 2026 Hardening Guide — technical details of each 2026.2.x fix
- ClawHub Security Alert — full breakdown of the supply chain attack
- OpenClaw After Steinberger — governance changes
- Best OpenClaw Alternatives — lighter tools with smaller attack surfaces
- OpenClaw Setup Guide — safe installation from scratch

Sources: NVD CVE-2026-26324, NVD CVE-2026-26319, RedPacket Security (CVE-2026-28466), RedPacket Security (CVE-2026-28453), RedPacket Security (CVE-2026-28478), RedPacket Security (CVE-2026-28479), Oasis Security, Security Affairs, The Hacker News (ClawHub), The Hacker News (ClawJacked), eSecurity Planet, GitHub Issue #14203, Gigazine, TechCrunch, UCStrategies, Infosecurity Magazine, SentinelOne (CVE-2026-25593), SentinelOne (CVE-2026-26322), GitHub Advisory (GHSA-g55j-c2v4-pjcg), GitHub Advisory (GHSA-r8g4-86fx-92mq)

Get notified when we publish new guides.

Subscribe — free, no spam

Source: https://insiderllm.com/guides/openclaw-security-report-february-2026/

Free guides for running AI locally