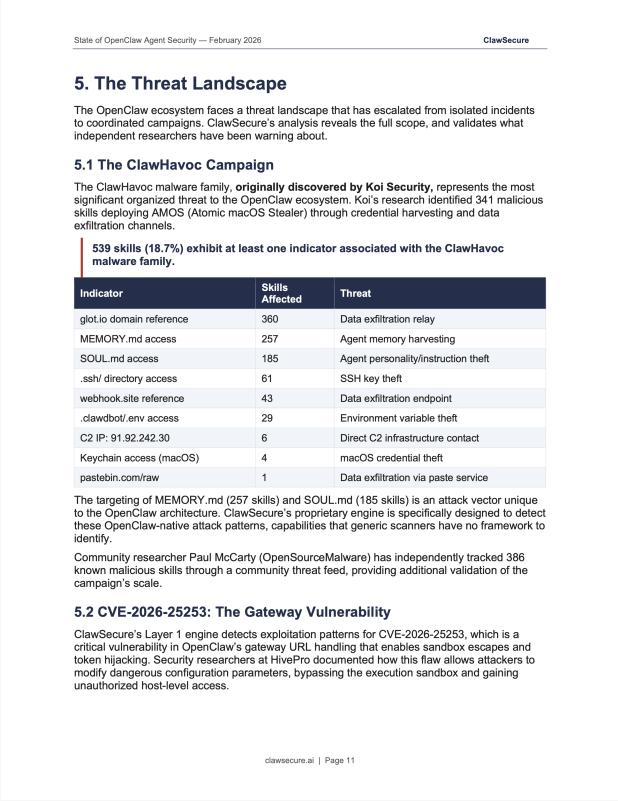
ClawHavoc Malware Found in 539 OpenClaw Skills, ClawSecure Reports

Audit identifies credential harvesting, C2 callbacks, and data exfiltration patterns across 18.7% of the most popular OpenClaw agent skills, ClawSecure reports



State of OpenClaw Agent Security — February 2026 ClawSecure

5. The Threat Landscape

The OpenClaw ecosystem faces a threat landscape that has escalated from isolated incidents to coordinated campaigns. ClawSecure's analysis reveals the full scope, and validates what independent researchers have been warning about.

5.1 The ClawHavoc Campaign

The ClawHavoc malware family, **originally discovered by Koi Security,** represents the most significant organized threat to the OpenClaw ecosystem. Koi's research identified 341 malicious skills deploying AMOS (Atomic macOS Stealer) through credential harvesting and data exfiltration channels.

> **539 skills (18.7%) exhibit at least one indicator associated with the ClawHavoc malware family.**

Indicator	Skills Affected	Threat
glot.io domain reference	360	Data exfiltration relay
MEMORY.md access	257	Agent memory harvesting
SOUL.md access	185	Agent personality/instruction theft
.ssh/ directory access	61	SSH key theft
webhook.site reference	43	Data exfiltration endpoint
.clawdbot/.env access	29	Environment variable theft
C2 IP: 91.92.242.30	6	Direct C2 infrastructure contact
Keychain access (macOS)	4	macOS credential theft
pastebin.com/raw	1	Data exfiltration via paste service

The targeting of MEMORY.md (257 skills) and SOUL.md (185 skills) is an attack vector unique to the OpenClaw architecture. ClawSecure's proprietary engine is specifically designed to detect these OpenClaw-native attack patterns, capabilities that generic scanners have no framework to identify.

Community researcher Paul McCarty (OpenSourceMalware) has independently tracked 386 known malicious skills through a community threat feed, providing additional validation of the campaign's scale.

5.2 CVE-2026-25253: The Gateway Vulnerability

ClawSecure's Layer 1 engine detects exploitation patterns for CVE-2026-25253, which is a critical vulnerability in OpenClaw's gateway URL handling that enables sandbox escapes and token hijacking. Security researchers at HivePro documented how this flaw allows attackers to modify dangerous configuration parameters, bypassing the execution sandbox and gaining unauthorized host-level access.

clawsecure.ai | Page 11

ClawSecure identified ClawHavoc malware indicators in 539 of 2,890+ audited OpenClaw skills, representing 18.7% of the ecosystem's most popular agents, including credential harvesting, C2 callbacks, and data exfiltration patterns.

SAN FRANCISCO, FL, UNITED STATES, March 17, 2026 /EINPresswire.com/ -- 539 popular OpenClaw skills, representing 18.7% of the ecosystem's most widely installed agents, contain indicators of the ClawHavoc malware campaign, according to an independent audit by ClawSecure (https://www.clawsecure.ai). The audited skills were drawn from the community-curated awesome-openclaw-skills list and the openclaw/skills repository, covering 2,890+ of the most popular agents in the OpenClaw ecosystem. ClawSecure's findings confirm that the ClawHavoc threat extends well beyond the initial discoveries reported by security researchers in January 2026, when the campaign was first identified targeting OpenClaw users through professionally disguised skills on ClawHub.

ClawHavoc is a coordinated malware campaign targeting the OpenClaw ecosystem through skills that appear legitimate but perform credential harvesting, establish command-and-control (C2) callbacks to external servers, and exfiltrate sensitive data via relay services. The campaign is notable for its operational discipline and social engineering. ClawHavoc skills are carefully designed to mimic high-demand categories including productivity tools, development utilities, and automation workflows, making them difficult to distinguish from legitimate skills through manual review alone. Once installed, a ClawHavoc-infected skill can silently harvest API keys, OAuth tokens, and messaging credentials stored in OpenClaw's configuration files, then transmit them to attacker-controlled infrastructure.

ClawSecure has conducted the largest independent analysis of ClawHavoc indicators in the OpenClaw ecosystem, with 539 confirmed findings across 2,890+ audited skills and the only public, searchable registry of affected agents. ClawSecure's proprietary behavioral engine, which includes 55+ threat patterns purpose-built for OpenClaw, independently identified these indicators through automated analysis. The findings complement earlier research by Koi Security while providing quantitative scope data that was previously unavailable to the OpenClaw community.

ClawSecure's free OpenClaw security scanner detects ClawHavoc malware indicators across 2,890+ audited skills using a proprietary 3-Layer Audit Protocol with 55+ threat patterns built for the OpenClaw ecosystem.

"ClawHavoc is not a theoretical threat. It is active, widespread, and specifically engineered for the OpenClaw ecosystem," said J.D. Salbego, Founder of ClawSecure. "When nearly one in five of the most popular skills show malware indicators, the ecosystem needs continuous monitoring infrastructure, not one-time scans. That is exactly what our Watchtower delivers."

> ClawSecure's audit found ClawHavoc indicators in 539 of the most popular OpenClaw skills. The ecosystem needs continuous monitoring infrastructure, not one-time scans. Watchtower delivers that."
>
> *J.D. Salbego, Founder of ClawSecure*

ClawSecure's detection capabilities address what Palo Alto Networks (2026) identified as the "Lethal Trifecta" of agentic AI risks: the combination of access to private data, exposure to untrusted content, and the ability to execute tools on the user's behalf. OpenClaw agents routinely access the file system, execute shell commands, read browser data, control messaging platforms, and make network calls on the user's behalf. A ClawHavoc-infected skill exploits every one of these capabilities, turning the agent's legitimate permissions into an attack vector. ClawSecure's 3-Layer Audit Protocol traces execution paths and data flows across tool-calling chains, identifying skills that exploit this trifecta for malicious purposes.

ClawSecure's Context-Aware Intelligence is essential for accurate ClawHavoc detection. Generic malware scanners flag legitimate OpenClaw agent capabilities like shell execution, clipboard access, and network calls as suspicious, generating false positives that make the results unusable for developers. ClawSecure understands that these capabilities are standard for useful

OpenClaw agents and evaluates them in ecosystem context, differentiating real ClawHavoc indicators from normal agent functionality. ClawSecure's audit of Peter Steinberger's flagship skill, peekaboo, scored it 95 out of 100, correctly identifying its system-level capabilities as standard functionality while flagging actual threats in other skills with similar permission profiles.

ClawSecure's independent audit found 1 in 5 OpenClaw skills are sending data to attackers. 18.7% carry active malware including ClawHavoc indicators, 30.6% exhibit shell execution, prompt injection, or credential theft, and 9,515 total threats were identi

ClawSecure's Watchtower monitoring system adds a critical layer of ongoing protection against evolving ClawHavoc variants. The system tracks code changes across all 2,890+ registered skills using SHA-256 hash comparisons, automatically triggering a full re-audit through the 3-Layer Audit Protocol whenever a modification is detected. ClawSecure's Watchtower has already identified 661 code changes across the registry, catching cases where previously clean skills were updated to include suspicious behavior patterns consistent with ClawHavoc tactics. This continuous monitoring addresses the "sleeper agent" risk where a skill passes an initial review but is later modified to include malicious behavior, a tactic increasingly used by threat actors to bypass one-time security scans.
ClawSecure's broader audit of the OpenClaw ecosystem found that 41% of all 2,890+ audited skills contain at least one security vulnerability, with 9,515 total findings identified. Beyond ClawHavoc, ClawSecure identified widespread supply chain risks including unpinned npm dependencies, credential exposure, unauthorized network calls, excessive permission requests, and ReDoS vulnerabilities. ClawSecure achieves comprehensive coverage across all 10 OWASP ASI Top 10 categories and is the first OpenClaw security platform to publish formal NIST AI Risk Management Framework alignment documentation, available at the Trust Center (https://www.clawsecure.ai/trust).

For organizations building agent marketplaces or identity platforms, ClawSecure's Security Clearance API provides programmatic access to real-time integrity verdicts, enabling automated blocking of skills exhibiting ClawHavoc indicators before they reach end users. Identity platforms such as Moltbook, with its 2.2 million agents, can integrate ClawSecure's integrity verification to complement their creator identity and reputation systems, forming the complete trust stack the agentic ecosystem requires. OpenClaw users concerned about malware in their installed skills can check any skill for ClawHavoc indicators using ClawSecure's free scanner, which delivers a full security audit report in under 30 seconds at https://www.clawsecure.ai. Detailed findings for all 2,890+ audited skills are accessible through the ClawSecure security registry (https://www.clawsecure.ai/registry). Organizations can also review ClawSecure's full ClawHavoc

analysis at https://www.clawsecure.ai/blog/clawhavoc-explained.

ClawSecure (https://www.clawsecure.ai) is the independent integrity layer for AI agent skills and workflows and the only free OpenClaw security scanner with full OWASP ASI Top 10 coverage. Built on a proprietary 3-Layer Audit Protocol, ClawSecure has audited 2,890+ OpenClaw agents from the community-curated awesome-openclaw-skills list and the openclaw/skills repository. The platform includes 24/7 Watchtower hash-drift monitoring, a Security Clearance API for marketplace and identity platform integration, and a public security registry. Founded by J.D. Salbego.

Paul Bateman
ClawSecure, Inc
email us here
Visit us on social media:
LinkedIn
YouTube
X
