CSAI Foundation | Frontier Ready Cybersecurity

Frontier Ready Daily

Issue 79 · Member edition

2026-11-13



CSAI FOUNDATION INITIATIVE

Frontier Ready.

Machine-speed agentic cybersecurity

cloud security alliance®

CSAI

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

In this issue

One operator or a small team ran a forked open-source agent framework against 19 organisations at once and got into 10, and Unit 42's conclusion is that the agents removed the operator-hours limit on how many targets one actor can work, not that they made the attacks more sophisticated. The other items: a rogue ScreenConnect push reached 288 endpoints in 90 minutes through an extension that skipped login and MFA; N-central servers on the hotfix we told you to apply on 11 November were mining Monero five hours after a new proof of concept appeared; MCP servers pinned to one release kept pulling new dependencies; large-enterprise teams are closing vulnerability findings they never checked; and the OCC has told bank examiners how to assess agentic AI. We examined CloudSEK's Silent Lynx case, OpenAI's Daybreak agents and Anthropic's self-hosted Managed Agents and left all three out: Silent Lynx is reported only second-hand, with medium-confidence attribution and an intrusion that took weeks, and corroborates the Cairn item without adding to it, while the other two are product launches with no independent evidence of performance or adoption.

One operator ran a forked Cairn agent framework against 19 organisations at once and reached data at 10, with single operations lasting from 20 minutes to more than 50 hours

Category: machine_speed **Significance:** major **Confidence:** VERBATIM (PROVIDER) for the 25 operations, 19 targets and 10 confirmed intrusions, the 20-minute to 50-hour durations, the model counts, the statement that "Neither Zhipu GLM nor DeepSeek refused" the authorisation preamble, and the conclusion that autonomous AI "raises the volume and coverage of exploitation, but not its complexity"; VERBATIM (PROVIDER) for the moderate-confidence China-nexus attribution; NO PROVIDER CLAIM on how many organisations detected the activity themselves; CHARACTERIZATION (CSA) for the staffing reading **Vendor interest:** Palo Alto Networks sells detection, identity and continuous offensive-testing services, and the report recommends its own products, so the volume framing is commercially favourable to it. **Source:** Unit 42 – Attackers Use Autonomous AI to Target Semiconductor, Defense and Civil Society Sectors

What changed. On 12 November 2026 Unit 42 published its account of CL-STA-1238, which it assesses with moderate confidence to be China-nexus intelligence collection. In August and September 2026 the actor ran 25 operations against at least 19 organisations in semiconductor and advanced manufacturing, US defence policy, and human rights and civil society, and confirmed unauthorised access or data

retrieval at 10. The tooling was a customised fork of the open-source Cairn pentest framework, run from a Hong Kong server with Zhipu GLM 5.2 (13 operations), DeepSeek V4 Flash (8) and DeepSeek V4 Pro (4), with Pro used against targets that resisted Flash. Operations ran from about 20 minutes to more than 50 hours with no time limit set. The results included an overwritten standards-ballot document, new administrator accounts and a persistent API token at a semiconductor company, and more than one million member records reached at a teachers' union. Unit 42 assesses that a single operator or a small team ran the whole campaign.

Why it reaches you. The entry points were internet-facing application logins, open registration and account-creation endpoints, and cloud storage that accepted unauthenticated writes. These are the routine weaknesses an exposure programme already tracks, so the change is in throughput, not technique. Unit 42 says orchestration "removes the historical limit on the number of targets an actor could pursue". Organisations that sat below a state actor's priority line because each intrusion used to cost operator time should assume that line has moved. The guardrail point also matters: two open-weight models accepted a one-line claim of authorisation, so refusal on the model side gives defenders no protection.

What to do. Exposure management owner, with identity: validate. Run a sweep this cycle for the five entry classes Unit 42 names: weak or shared passwords on internet-facing apps, account-creation and registration endpoints that issue sessions without verification, unauthenticated cloud-storage writes, credentials valid across several accounts, and VPN group passwords. Set remediation targets for these classes in weeks, not quarters, because the campaign worked 19 targets in two weeks. Urgency: validate.

CSA coverage. State-Level Tempo for Everyone: Commodity Agent Frameworks and Adversary Tiering; see also Fifty Dollars per Breach: Detecting Autonomous Agent Intrusion Campaigns.

A ScreenConnect extension that accepted a blank access key let an attacker push rogue clients to 288 endpoints in 90 minutes, bypassing login, roles and MFA on fully patched servers

Category: machine_speed **Significance:** major **Confidence:** VERBATIM (PROVIDER) for the 288 endpoints across 48 organisations, the 90-minute spread with "about one third of those in the first 10 minutes", the blank-key behaviour of RMM+ 1.0.31 and 1.0.32, and the instruction not to connect to suspect hosts over ScreenConnect; SELF-REPORTED (PROVIDER METRIC) for the scope, which counts only Huntress-protected organisations; VERBATIM (PROVIDER) for ConnectWise's statement that it "found no evidence of a compromise of ConnectWise infrastructure", as reported by The Register; CHARACTERIZATION (CSA) for the parallel-authorisation-path reading **Vendor interest:** Huntress sells

managed endpoint detection and response to MSPs, so a finding that RMM channels need independent endpoint monitoring is commercially favourable to it. **Source:** Huntress – SCKit Worm Installs Rogue ScreenConnect Clients and Session Cookie Relays Across Hundreds of Endpoints

What changed. From about 17:47 UTC on 11 November 2026, an unidentified operator used ScreenConnect's Run Command, issued through the third-party RMM+ extension, to install attacker-controlled ScreenConnect clients and a 16-file toolkit called SCKit as SYSTEM. Huntress saw all 288 hosts hit within 90 minutes, a third of them in the first 10. RMM+ authenticates with its own shared key, which defaulted to empty in versions 1.0.31 and 1.0.32 and was accepted when empty, so commands ran with no ScreenConnect login, role check or MFA. The victims were on current ScreenConnect builds. The toolkit then waits for a technician to connect so it can try to steal the technician's panel session cookie and push to every session the technician can reach. Huntress and Blackpoint say that step has not succeeded against clients at 26.6.5 or later. ConnectWise disabled RMM+ and four related extensions across hosted partners on 11 November.

Why it reaches you. The component is the remote-support and RMM plane, whether you run it or an MSP runs it for you. An extension added for an integration project created a second way to reach the platform's most privileged function, and that path carried none of the controls configured on the first. Patch currency on the core product did not help. The response guidance reverses normal practice: connecting to a suspect machine over ScreenConnect is the trigger SCKit waits for, so the usual triage tool becomes the propagation path.

What to do. IT operations owner, with the MSP relationship owner: escalate. Get written confirmation today that every ScreenConnect instance touching your estate, hosted or on-premises, has RMM+ and the related extensions disabled, that all clients are at 26.6.5 or later, and that Run Command history since 11 November shows no mass execution without a matching technician login. Investigate suspect hosts only through EDR live response or another out-of-band channel. Then add a standing supplier question: list every extension or integration that can issue commands without the platform's own authentication. Urgency: escalate.

CSA coverage. SCKit and Rogue ScreenConnect Clients: MSP Remote Support as Attack Path.

N-central servers already on the hotfix we told you to apply on 11 November were mining Monero within hours of a new proof of concept, 19 of the 21 victims found

Category: vuln_storm **Significance:** major **Confidence:** VERBATIM (PROVIDER) for the 03:58 UTC proof-of-concept time, the 08:30 UTC first probe and 09:16 UTC weaponised request against huntback's decoys, and the 19-of-21 build count; SELF-REPORTED (PROVIDER METRIC) for the 21-server total, which huntback derived from mining-pool rig names and its own HTTPS checks and which does not

publish victim addresses; VERBATIM (PROVIDER) for N-able's statement that it is "restricting certain N-central features and integrations", as reported by BleepingComputer; NO PROVIDER CLAIM from N-able of a complete fix; CHARACTERIZATION (CSA) for the remediation-verification reading **Vendor interest:** huntback operates deception sensors and sells threat-intelligence research, so a finding that patching alone did not protect is commercially favourable to it. **Source:** huntback – 4EverAble2Shell is already mining Monero on live N-central servers

What changed. On 12 November 2026 at 03:58 UTC, researcher Sean Bales published 4EverAble2Shell and Agent2Shell, new exploit chains that work against N-central 2026.4 Hotfix 2, which N-able had shipped on 10 November in response to the earlier Able2Shell disclosure. Huntback's decoys took a first probe at 08:30 UTC, four and a half hours after publication, and a weaponised request at 09:16. Huntback then found 21 production N-central servers mining Monero, and 19 of them were running build 2026.4.1.26, which is Hotfix 2. N-able is restricting features, has disabled the Intune integration and Legacy API on hosted instances, tells customers that upgrading Agents and Probes to 2026.4 HF1 or later is "a MUST", and says remediation is still in development.

Why it reaches you. The component is the RMM server that holds SYSTEM-level reach into every managed endpoint. Our 11 November item told you to confirm N-central was on 2026.4 HF2 or later. That was correct then and is no longer sufficient. The interval that matters is now the one between a vendor's partial fix and the next public exploit against it, and here that was two days. The payload so far is a miner, but the chain gives root, so a miner on the host means anyone else can get root there too. A patch-status report showing green did not reflect exposure.

What to do. IT operations owner, with the MSP relationship owner: escalate. This supersedes our 11 November advice. Take N-central off direct internet exposure and keep it reachable only through a VPN or an allow-listed path until N-able ships a fix it states is complete. Upgrade Agents and Probes as N-able directs, confirm Legacy API is disabled, and check every server for unexpected CPU load, `/tmp` remounted with execute permission, and outbound mining-pool connections. Record the gap between each RMM hotfix and the first exploit that beats it as a tracked interval. Urgency: escalate.

CSA coverage. Able2Shell: N-central Zero-Days and the RMM Kill Switch; see also When Patching Is the Attack Window: WordPress CVE-2026-119977.

Close to half of large-enterprise security teams close or deprioritise vulnerability findings without verifying them, and about a quarter hold more than 10,000 untriaged findings

Category: vuln_storm **Significance:** notable **Confidence:** SELF-REPORTED (PROVIDER METRIC) for all survey figures, including the close-to-half closing findings unverified, the three in four reporting backlog growth, the quarter above 10,000 untriaged findings, and the nearly half saying AI adds findings; NO PROVIDER CLAIM on sample size, survey dates or methodology, none of which the report coverage gives; LINK ONLY – VERIFY AT SOURCE for the Mandiant time-to-exploit figures quoted alongside; CHARACTERIZATION (CSA) for the bottleneck reading **Vendor interest:** ProjectDiscovery sells vulnerability-scanning and exploitability-validation tooling, so a finding that teams cannot verify findings is commercially favourable to it. **Source:** Help Net Security – Nearly half of big firms shelve or close vulnerability findings they haven't confirmed

What changed. On 11 November 2026 Help Net Security reported a ProjectDiscovery survey of organisations with more than 5,000 employees. Close to half of security teams close or deprioritise findings without fully verifying them, citing lack of time. Among those teams, 57% say at least three in four of the findings they do check turn out to be real. Three in four say their backlog grew in the past year, most hold more than 1,000 untriaged findings, and about a quarter hold more than 10,000. Teams drawing on five or more finding sources are nearly twice as likely to close findings unverified. Nearly half say AI is adding findings to their workload, more than half expect AI-equipped attackers to reach backlog items first, and half say more hiring would not fix the problem.

Why it reaches you. The component is the vulnerability intake and triage queue that sits between scanners, bug bounties and AI discovery on one side and remediation on the other. A finding closed unverified is recorded as handled, so the denominator your board sees (findings resolved) overstates the share actually dealt with. More than half of the teams that skip verification say at least three in four of the findings they do check are real. If the skipped findings look similar, unverified closure takes mostly real issues out of the queue. The figures need independent confirmation: a published sample size and method, or a second survey with comparable questions.

What to do. Vulnerability management owner: validate. Split "closed" in your own metrics into verified-fixed, verified-not-exploitable and closed-unverified, and report the third category separately. If it is material, sample 50 closed-unverified findings and verify them. The real-finding rate in that sample is the number to put in front of the CISO. Urgency: validate.

CSA coverage. The Disclosure Bottleneck: When AI Discovery Outpaces Patch Consumption; see also VulnOps: Vulnerability Management in the Age of AI.

An MCP server pinned to one release ran with 15 different dependency sets in five weeks, and 23 of 35 servers reviewed pull fresh dependencies from a public registry on launch

Category: agentic_surface **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for the 15 dependency sets from 19 builds, the 23 of 69 packages that changed version, the 23 of 35 servers whose recommended install resolves dependencies fresh, and the 457 starts with unchanged package lists after freezing; SELF-REPORTED (PROVIDER METRIC) for the author's statement, reported by Help Net Security, that 100% of administrators he has spoken to run at least one affected server; NO PROVIDER CLAIM of any malicious package delivered this way; CHARACTERIZATION (CSA) for the reading that version pinning is not a supply-chain control here **Vendor interest:** Corelight sells network detection and response and publishes its own MCP servers, so a finding that favours vendor-maintained, centrally reviewed servers is commercially favourable to it. **Source:** Corelight — The supply-chain risk hiding in your MCP servers

What changed. On 10 November 2026 Corelight published measurements from MCP servers launched by AI coding assistants on an administrator's workstation. One community server, configured to a single release tag and commit, was rebuilt 19 times and ran with 15 distinct dependency sets over five weeks. 23 of its 69 packages changed version, including the cryptography library, a JSON Web Token library and the MCP SDK. The configuration never changed. The cause is the package-runner launcher, which installs from the package's declared dependencies, ignores the project's hash-pinned lock file, and resolves about 60 packages fresh from PyPI on every environment rebuild. Of 35 servers Corelight reviewed from two vendor catalogues and its lab, 23 had a recommended install with the same behaviour. Five servers frozen from a hash-pinned lock file started 457 times over two weeks with no change to their package lists.

Why it reaches you. The component is the MCP server set an AI coding assistant or agent runtime launches with the user's credentials, often next to tokens for the firewall, SIEM and other management systems. In Corelight's test those servers made more than 1,600 connections to 11 management systems in one week. A reviewer who approves "server X at version Y" has approved one tree of dependencies, while the runtime executes whichever tree the registry serves that day. A hijacked dependency that is live for only a few hours, as in the Shai-Hulud PyPI wave against MCP-themed packages in June, reaches every assistant that relaunches during that window.

What to do. AI platform owner, with application security: validate. Inventory MCP entries started through a Python or npm package runner, or tagged `latest`. For servers that touch production or security tooling, install once from a hash-pinned lock file, launch the installed environment directly, and turn off

update checks. Then compare each server's installed package list against its install-day list on a schedule. Freezing ages the code, so pair it with a named owner who reviews and rebuilds deliberately. Urgency: validate.

CSA coverage. New — no prior CSA coverage of MCP dependency drift; see also AI Coding Agents as Attack Surface: MCP, Poisoning, and Miasma.

The OCC told national bank examiners to scale scrutiny of agentic AI by consequence and reversibility, and to treat agents as privileged non-human identities

Category: security_operating_model **Significance:** notable **Confidence:** VERBATIM (PROVIDER) for the bulletin's statement that the framework "does not establish supervisory expectations for banks", the six pillars and sixteen components, and the framework's treatment of agent credentials "in a manner comparable to privileged users"; NO PROVIDER CLAIM on when examinations applying the framework begin or how findings will be recorded; CHARACTERIZATION (CSA) for the reading that it sets the de facto evidence list for banks and their AI vendors **Source:** Office of the Comptroller of the Currency — Super Intelligence: Supervisory Framework (OCC Bulletin 2026-54)

What changed. On 4 November 2026 the OCC issued Bulletin 2026-54 and an accompanying framework of about 16 pages telling examiners how to assess national banks' use of generative and agentic AI, which it calls "Super Intelligence" following Executive Order 14434. The framework does not treat AI as a separate risk category. It groups 16 components into six pillars (purpose, risk, control, oversight, resilience, assurance) and scales control intensity by decision consequence, reversibility, autonomy and data sensitivity rather than by model sophistication. An autonomous customer-facing payment agent sits at the top tier however well it tests. It also expects banks to detect and contain an agent that stays available while taking harmful actions, using circuit breakers, rate limits, least privilege and kill switches. It brings vendors into scope, covering provider agents' authority, AI incident definitions, liability and notice when vendors add AI features to existing products. It fills the gap left in April when the revised model-risk guidance put generative and agentic AI out of scope.

Why it reaches you. The component is the governance and evidence layer around production agents: the agent inventory, the authority envelope per agent, the containment controls and the third-party contract terms. The OCC says the document is not a self-assessment standard, but it is the list of questions examiners will ask, and banks will pass those questions on to their AI vendors. Non-bank enterprises get a usable reference model from a primary regulator. The framework ties control intensity to consequence and reversibility and treats a misbehaving but available agent as a resilience event, which most agent programmes do not yet test for.

What to do. CISO, with model risk and third-party risk: validate. For each production agent, record its tier on consequence and reversibility, the credentials it holds and whether they are governed as privileged access, and a tested way to halt it without taking the service down. Add the framework's vendor questions (scope of provider-agent authority, AI incident definition, notice of new AI features) to AI supplier reviews. Banks should do this before their next examination cycle, and other regulated firms should do it as a benchmark. Urgency: validate.

CSA coverage. OCC Bulletin 2026-54: Examiner Framework for Agentic AI in Banking.

Watchlist

- **OpenAI reward-hacking postmortem — downstream response** — No change. No frontier lab disclosed an eval-to-production escape, and there was no new regulatory or contractual development. *(opened 2026-08-27)*
- **VM/hypervisor containment hardening for cyber-capable agents** — No change. No provider or enterprise moved agent sandboxes to hardened microVMs. We saw secondary reports of Xen guest-to-guest memory flaws, but they do not concern QEMU/KVM/libslirp and we have not verified them. *(opened 2026-08-27)*
- **Claude Code Auto Mode prompt-injection ASR discrepancy** — No change. There is still no Anthropic patch or public response. *(opened 2026-08-27)*
- **AI defensive-triage guardrail evasion** — No change. No new campaign triggering refusals to blind triage tooling was reported. Unit 42 found that GLM and DeepSeek accepted a one-line authorisation preamble in the Cairn campaign (item 1), but that is an offensive guardrail bypass, not a defensive one. *(opened 2026-08-31)*
- **AI account session hijacking at scale** — No change. No AI provider disclosed a session-hijacking campaign or shipped device-bound session tokens. Anthropic's self-hosted Managed Agents beta keeps tool credentials outside the agent sandbox, but it covers API credentials, not user sessions. *(opened 2026-08-31)*

Opened this issue

- **RMM hotfixes beaten by follow-on exploits** (vuln_storm) — Watching for N-able to ship an N-central fix it describes as complete, for whether the Agent and Probe upgrade stops 4EverAble2Shell, for exploitation beyond cryptomining, and for the same hotfix-then-bypass pattern at other RMM vendors after N-central and ScreenConnect. *(opened 2026-11-13)*