KILLSWITCH · ISSUE #17 · 11 NOV 2026

Localhost Was the Password: The GPUStack Flaw That Turned "127.0.0.1" Into Admin

GPUStack treated any request that looked local as the administrator. Behind a reverse proxy, every request looks local.

Killswitch rating: Patch and check your AI servers

By Emmanuel Aigbehi Walstra

Localhost was the password

Before 2.2.0, GPUStack treated any request that looked local as the built-in administrator, by default [1].

- Behind a same-host reverse proxy, "every remote request" looks local [1].
- GPUStack reports exploitation against a 2.1.1 deployment [1]. CISA KEV, 9 Nov [2][3].

Critical, exploited, fixed in June

- Authentication bypass, CVSS 3.1 score 9.8 [1][4].
- Gives users, models, admin API keys, worker and cloud secrets, code on workers [1][3].
- CISA: forensic triage required for US federal agencies; deadline 12 Nov [3].
- First fixed in 2.2.0 (18 Jun); upgrade to the latest, 2.3.0 [1][6][12].

03 · TIMELINE

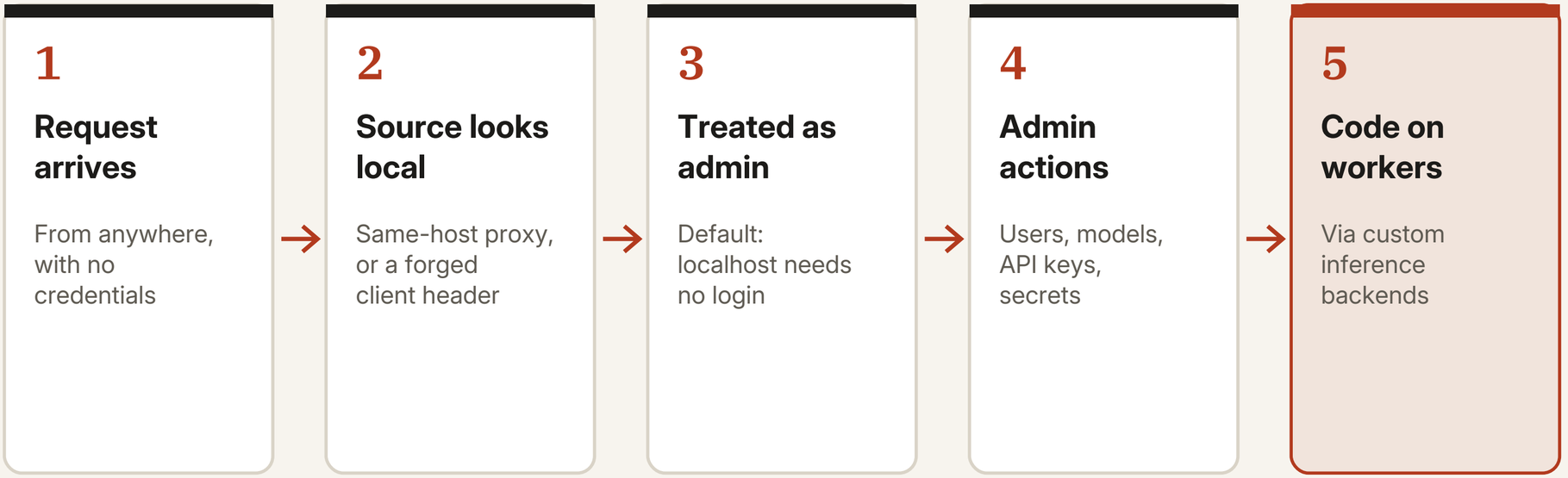
The fix was out for months before the alarm



GPUStack [1][5][6][9], NVD [4], CISA [2][3]. All dates 2026.

144 days from the fixed release to the KEV listing, by our count.

How "it looks local" became admin



Per GPUStack's advisory [1] and CISA's KEV entry [3]. Described at design level only.

The network source is not a credential

- GPUStack: "Trusting the network source of a request … as an authentication signal" [1].
- A same-host proxy makes every request arrive from 127.0.0.1 [1].
- In some gateway set-ups, a client can forge the real-IP header [1].
- Keys minted as admin stay valid "from any source IP thereafter" [1].

Below 2.2.0, behind a proxy, reachable

- Any GPUStack version before 2.2.0 [1].
- Highest risk: same-host reverse proxy, reachable API, force_auth_localhost never set [1].
- The 2.2.0 release notes do not mention the fix, as far as we can find [6]. The PR was titled a refactor; the commit says "fix" [5].

07 · DETECTION

What to look for if it was reachable

1. **Admin API keys**
   Any key you did not create, especially with full scope
2. **Custom inference backends**
   Any backend added from YAML that you do not recognise
3. **Users and models**
   New users, role changes, models you did not deploy
4. **Workers**
   Unexpected processes or outbound connections
5. **Cloud and provider secrets**
   Use in provider logs since the server was exposed

From the impacts GPUStack [1] and CISA [3] list. Preserve evidence first [8].

Seven moves, in this order

1 Preserve

If reachable: memory, then logs, database, keys, backends

2 Upgrade

Latest release (2.3.0); 2.2.0 is the first with this fix

3 Until then

force_auth_localhost=true; strip the real-IP header at the edge

4 Close the door

API off untrusted networks; VPN or edge authentication

5 Check for a visitor

Keys, backends, users, workers, secrets

6 Rotate

Admin keys, worker and cloud credentials, from a clean machine

7 Inventory

Every self-hosted AI server and how it authenticates

Evidence before patching, as in CISA's BOD 26-04 steps [8]. Fix and workarounds: GPUStack [1][12]. The inventory is a Killswitch addition.

2.2.0 is the first fixed release; 2.2.0 and 2.2.1 have a cluster-takeover flaw fixed in 2.2.2 [12].

Where this lives in your frameworks

Prevent	A.8.5, A.5.15, A.8.20, A.8.9 · NIS2 21(2)(i), (j) · Evidence: auth settings and proxy config
Patch	A.8.8, A.8.32 · NIS2 21(2)(e) · Evidence: version record and patch date
Detect and respond	A.8.15, A.8.16, A.5.26, A.5.28 · NIS2 21(2)(b) · Evidence: triage report
Know	A.5.9, A.5.23 · NIS2 21(2)(d), (i) · Evidence: AI platform inventory

Killswitch's own mapping. Check ISO titles against your licensed copy.

ISO/IEC 27001:2022 [10] · NIS2 Art. 21(2), Art. 23 [11]. Killswitch mapping.

10 · STILL OPEN

What we don't know yet

- No victim, actor or scale is public; KEV ransomware use: "Unknown" [1][3].
- Whether the in-the-wild note was in the advisory on day one or added later [1].
- Header hardening: in our reading, PR #5310 (in 2.2.0); not named by GPUStack [1][13].

This week

- Find every GPUStack server and its version.
- If reachable, preserve evidence first [8].
- Upgrade to the latest (2.3.0), not just 2.2.0 [12]; until then force_auth_localhost=true and strip the header [1].
- Take the API off untrusted networks [1].
- Check keys, backends, users, workers; keep a triage record [3][8].
- Rotate keys and secrets from a clean machine.

Sources

[1] GPUStack: Security advisory GHSA-gj9m-f8ph-pmhr, authentication bypass via trusted loopback source grants unauthenticated admin (published 31 Jul 2026, updated 29 Sep 2026)

[2] CISA: CISA Adds Four Known Exploited Vulnerabilities to Catalog (9 Nov 2026)

[3] CISA: Known Exploited Vulnerabilities Catalog, CVE-2026-102895 (added 9 Nov 2026, due 12 Nov 2026, forensic triage required)

[4] NIST NVD: CVE-2026-102895 (published 4 Nov 2026)

[5] GPUStack: pull request #5346, "refactor: remove localhost auto-admin bypass" (merged 19 May 2026)

[6] GPUStack: release v2.2.0 (18 Jun 2026)

[7] GPUStack Docs: CLI reference, gpustack start (--force-auth-localhost, --trusted-hosts)

[8] CISA: BOD 26-04 Implementation Guidance, Prioritizing Security Updates Based on Risk (forensic triage steps)

[9] GPUStack: release v2.1.1 (26 Mar 2026)

[10] ISO/IEC 27001:2022 Information security management systems

[11] NIS2 Directive (EU) 2022/2555, Articles 21 and 23

[12] GPUStack: release v2.2.2, Security section (24 Jul 2026)

[13] GPUStack: pull request #5310, "Fix client ip getter" (merged 13 May 2026; commit 15767c7d)

Facts are from GPUStack's advisory, pull request, release and documentation, CISA's alert, KEV entry and BOD 26-04 guidance, and the NVD record, as published by 10 Nov 2026. The framework mapping, the inventory control and the tabletop are Killswitch's own.

FROM THE PUBLISHERS OF KILLSWITCH

Know which AI servers you run, who can reach them, and how they authenticate

GRCOMPLIANCE

Keep an inventory of your AI platforms and their access controls, link the evidence, and map controls to ISO 27001 and NIS2.

Try the live demo →

demo.grcompliance.com

Self-host free: github.com/Sectutor/GRCompliance-Core

SECTUTOR

Security and GRC books, training and audit kits.

sectutor.com

Emmanuel Aigbehi Walstra is a CISO who also serves as a virtual CISO (vCISO) for several organisations, with 30+ years in cybersecurity, 10,000+ people trained, and author of the CCSP Excellence books.