OpenAI and Cisco release Shirudo, an open-source tool to shield unpatched Linux systems



OpenAI and Cisco released Shirudo, an open-source eBPF-based tool that applies temporary runtime "shields" to block exploitation on Linux while patches are prepared, without restarting services.

- Shirudo uses eBPF to deploy targeted temporary protections, called shields, without restarting systems.
- AI agents can help investigate vulnerabilities and draft shields, but deployment stays under human control.
- Shields are signed cryptographically, and TPMs record when they are loaded or removed.
- In testing, a custom shield blocked a Java4Shell exploit attempt against a running Elasticsearch service without interrupting it.

> - An expert from Expel cautions against letting AI agents write rules that can affect the kernel.

OpenAI and Cisco have released Shirudo, an open-source Linux security tool meant to reduce exposure during the gap between a vulnerability becoming known and a patch being deployed. It applies temporary runtime protections, called shields, without restarting services.

How it works

Shirudo relies on eBPF, a Linux mechanism that runs controlled programs inside the kernel without changing its source code or requiring a restart. Shields aim to block malicious behavior narrowly so that legitimate application operations continue. Operators can assess a shield's impact before enforcing it, and they receive hit counts showing attempted exploitation.

AI agents can investigate vulnerabilities, develop shields and validate them, while deployment remains under human control. Shields are cryptographically signed, checked by the eBPF verifier and kernel controls, and TPMs provide verifiable records of loading and removal. The developers say the tool can be activated from a command line and needs no permanently running daemon. Planned tested shields will be distributed as signed packages through an OCI registry.

Caveats

In testing, OpenAI and Cisco researchers blocked a Java4Shell exploit attempt against a running Elasticsearch service without interrupting it. Marcus Hutchins of Expel sees real potential in eBPF mitigation but questions letting AI agents write rules that could affect the kernel. He also notes that the approach is limited to Linux. A shield is a stopgap, not a substitute for patching.

OpenAI has scheduled a technical webinar on December 15.

What to do now

1. Identify Linux servers where patching is delayed by change windows or uptime requirements, and record them as candidates for compensating controls.
2. Evaluate Shirudo in a test environment before any production use, including its effect on legitimate application behavior.
3. Require human approval and change management for any AI-generated shield, in line with your change and risk procedures.
4. Treat shields as temporary mitigations and keep patch deployment tracked to closure.
5. Verify the signing and TPM-based logging of shields can feed into your audit and monitoring evidence.