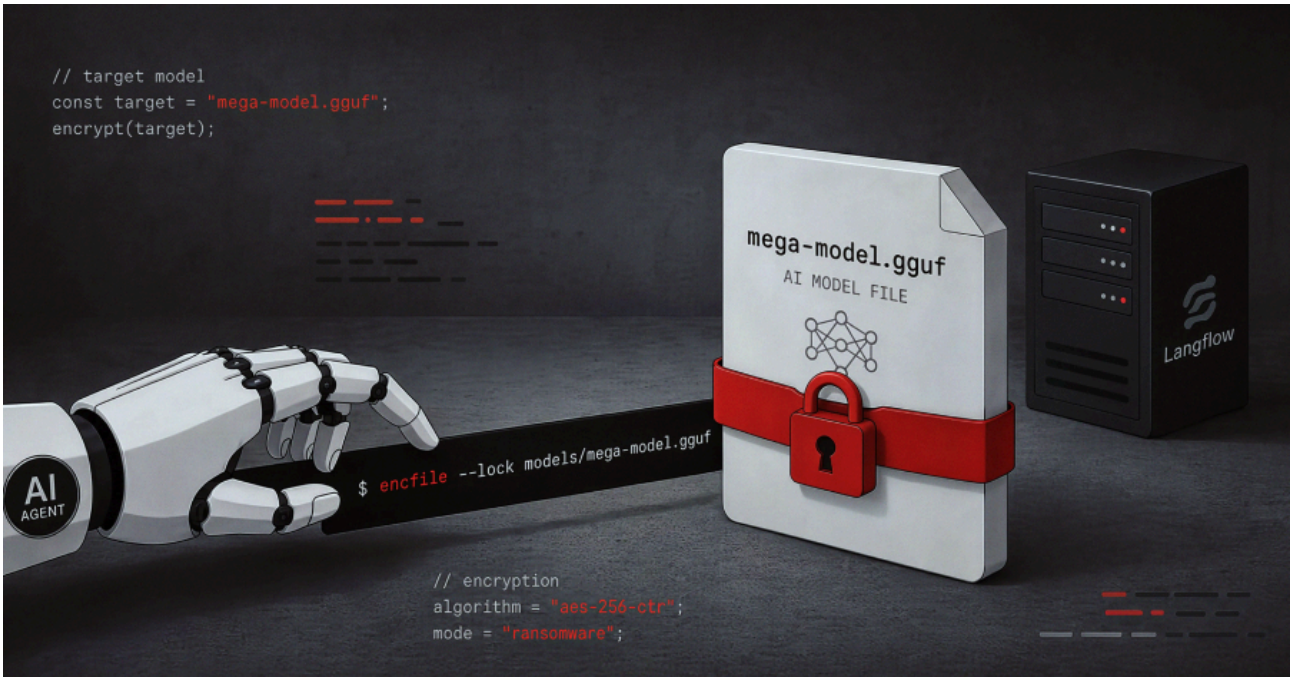
New ENCFORGE Ransomware Targets AI Model Files in Langflow RCE Attack

By The Hacker News

Published: 2026-07-21 · Archived: 2026-07-28 02:00:16 UTC



Researchers at Sysdig have linked a second attack on the same Langflow server to JADEPUFFER, the AI-agent-driven operator it first documented earlier this month.

The same operator has now been spotted deploying **ENCFORGE**, a new compiled Go ransomware designed to encrypt model weights, vector indexes, training datasets, and other AI infrastructure files across the host filesystem.

The entry point did not change. Langflow versions before 1.3.0 expose the `/api/v1/validate/code` endpoint without authentication, allowing any remote attacker to execute arbitrary Python on the server. The flaw, `CVE-2025-3248`, carries a CVSS score of 9.8 and has been in CISA's Known Exploited Vulnerabilities catalog since May 5, 2025.

As The Hacker News reported earlier this month, the prior operation used throwaway Python code and MySQL's `AES_ENCRYPT()` function to encrypt and destroy data in Nacos (Alibaba's configuration server) and production databases.

The new ENCFORGE payload replaces those improvised scripts with compiled tooling aimed at the model stores, vector databases, and training pipelines the first campaign swept for credentials.

The ENCFORGE Payload

Researchers retrieved the binary from the attacker's command-and-control server, where it was hidden as `/.lockd`; a direct request to `/lockd` returns 404, and the leading dot keeps it off a plain directory listing. The file is a UPX 5.20-packed static Go 1.22.12 ELF.

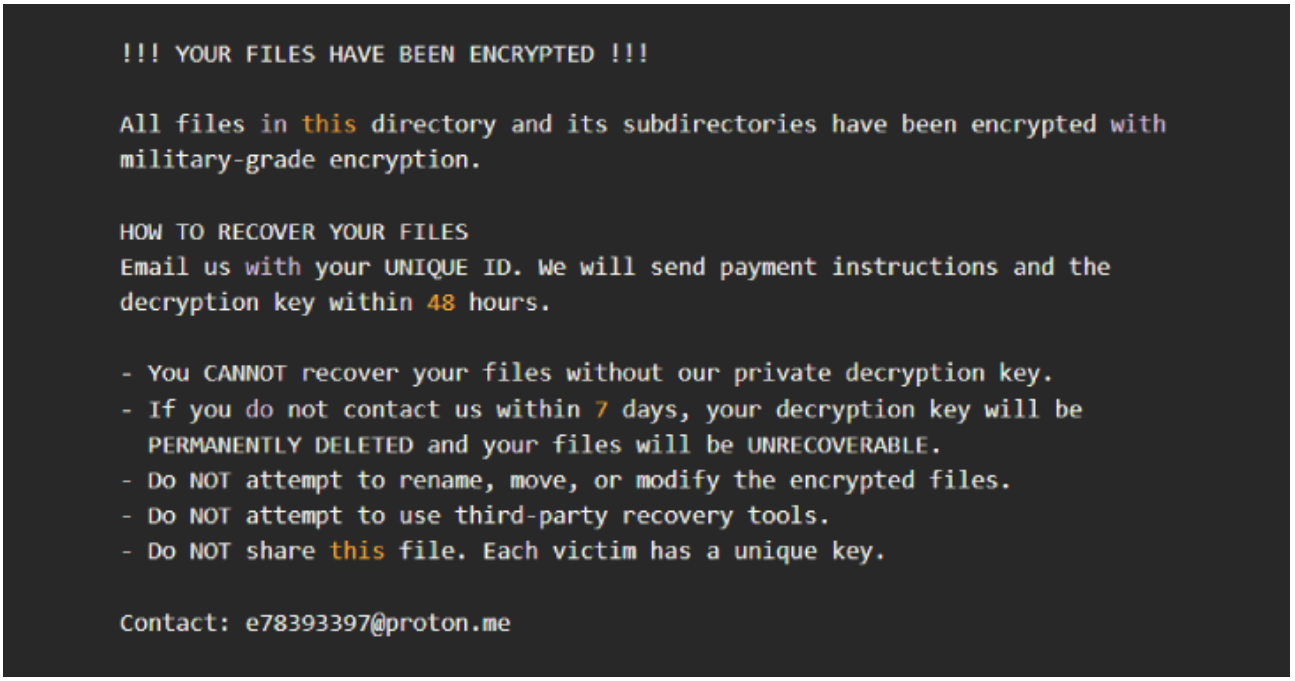
Threat intelligence platforms returned no detections on either the packed or unpacked hash at the time of Sysdig's analysis. The internal project name is `encfile`; the binary's error text references a companion keygen tool named `keyforge`. Both strings survive recompilation of the same codebase and serve as stable detection anchors.

Its default extension list covers PyTorch and TensorFlow checkpoints, Hugging Face SafeTensors, ONNX interchange format, GGUF (the current standard for locally deployed LLMs) and its predecessor GGML, FAISS vector indexes, Parquet and Arrow training datasets, NumPy arrays, and TensorFlow records.

An `--include` flag lets the operator append additional file globs; the built-in help text uses LoRA fine-tuning adapters and legacy GGML weights as examples. The full list runs to approximately 180 extensions. Those examples point directly at AI environments; a generic file locker would have little reason to name LoRA adapters or legacy GGML weights. Researchers read the choice as deliberate targeting, not incidental coverage.

ENCFORGE uses AES-256-CTR for file data, with the per-run symmetric key wrapped under an embedded RSA-2048 public key compiled into this build. Rather than encrypting whole files, it encrypts selected regions, the same speed optimization LockBit and BlackCat-class lockers use.

Each processed file is renamed with a `.locked` extension. The binary kills processes holding files open before encrypting, handles restarts without re-encrypting completed files, drops ransom notes as `README`, `HOW_TO_DECRYPT`, and `README_DECRYPT`, and deletes itself after running.



The recovered ENCFORGE binary has no exfiltration capability. It carries no networking code, cloud storage client, or staging mechanism, and researchers found no evidence of data exfiltration, a leak site, or a Tor payment portal during the session it observed. Its only leverage is the encrypted data.

The extortion contact embedded in the ransom notes is `e78393397@proton.me`, the same Proton Mail address used in the prior campaign. Sysdig describes this as the strongest attribution link between the two operations.

Researchers disclosed one observed session, did not name the organization, and provided no victim count or evidence of another ENCFORGE deployment.

From Langflow to the Host

After confirming code execution, JADEPUFFER swept the container for credentials and found the Docker socket at `/var/run/docker.sock`. Its first attempt to pull ENCFORGE from the GCP command-and-control server failed. Rather than stopping, it adapted.

Over five minutes and 24 seconds, the operator created and revised six Python scripts through the same Langflow RCE channel until it had a working path to the host. The first script was built one line at a time, keeping any single request inert for signature-based inspection.

Starting with the second, the operator encoded each full script in base64 and decoded it inside an `exec()` call, avoiding shell-level searches for commands such as `base64 -d`. The final version used the Docker API to spin up a privileged container with the host PID namespace and root filesystem mounted, located the target process, copied ENCFORGE through `/proc/<pid>/root`, then ran it on the host via `nsenter`.

Across all iterations, the containers were created with `Privileged: true`, `PidMode: host`, `NetworkMode: host`, and the root filesystem bind-mounted read-write. That is host root.

Before the live run, the operator launched `--try-run` to scan the filesystem, then `--lock` for the live encryption pass. The final script checked process status, read the lock log, and counted files ending in `.locked`.

Researchers did not publish the resulting count; the disclosed evidence establishes a live encryption attempt, not how many model or dataset files were successfully encrypted.

Sysdig assessed the `--task-id gcp_h1` flag as evidence the operator was tracking this host as a GCP target within a broader campaign; a try-run earlier in the session used task ID `gcp_test`. The report disclosed no additional victims or deployment sites.

Researchers documented the prior JADEPUFFER campaign correcting a failed Nacos login in 31 seconds. The same pattern held here against a harder problem: the operator built a host breakout through the exposed Docker socket when its preferred delivery route was blocked.

Patch Langflow, Then Protect the Models

Researchers estimate that rebuilding a production AI model once it has been encrypted could cost between $75,000 and $500,000 per model in cloud GPU compute and engineering time.

Production environments often run multiple specialized variants on shared storage, so a single ENCFORGE run could encrypt multiple variants stored on the same accessible filesystem. If training data sits on the same host, the organization has to reconstruct it before any retraining can start.

Binary SHA-256: packed `8cb0c223b018cecef1d990ec81c67b826eb3c30d54f06193cf69969e9a8baea2`; unpacked `ea7822eac6cecef7746c606b862b4d3034856caf754c4cf69533662637905328`.

Sysdig has published the source and C2 addresses, the embedded RSA-2048 key fingerprint, and a YARA rule in its full report.

- Upgrade Langflow to 1.9.1 or a current supported release. Version 1.3.0 closed `CVE-2025-3248`, the entry vector for this campaign, but CISA has since added two more Langflow vulnerabilities to its KEV catalog: `CVE-2026-33017`, an unauthenticated RCE flaw fixed in 1.9.0, added to KEV March 25, 2026; and `CVE-2026-55255`, a cross-user authorization bypass fixed in 1.9.1, added July 7, 2026.
- Rotate AI provider keys, cloud credentials, database secrets, and any other tokens accessible to the Langflow process. Patching does not revoke credentials already harvested through a vulnerable instance.
- Remove `/var/run/docker.sock` from any container that does not require it. Where socket access is unavoidable, scope it through a narrowly configured proxy; a standard Langflow deployment generally has no need to create containers, and unrestricted Docker socket access should be treated as a misconfiguration.
- Alert on application processes calling Docker container-creation APIs, containers launched with `Privileged: true` or `PidMode: host`, host-root bind mounts, and `nsenter` execution from inside a container.

- Keep model weights, vector indexes, and training datasets in offline or immutable snapshots. Monitor those directories for mass `.locked` file creation.

The Hacker News contacted Sysdig's Threat Research Team for further detail on the fleet campaign scope and attribution confidence; Sysdig had not responded by publication.

Model artifacts now belong in the same recovery tier as source code and production databases. An organization that can rebuild the application but cannot restore its weights, indexes, or training state has no clean route back.

Found this article interesting? Follow us on Google News, Twitter and LinkedIn to read more exclusive content we post.
