Hive Pro

Threat Level

Amber

Hiveforce Labs

THREAT ADVISORY

ATTACK REPORT

Experimental AI Ransomware PromptLock Sparks Security Concerns

Date of Publication	Admiralty Code	TA Number
September 4, 2025	A1	TA2025267

Summary

First Seen: August 2025
Malware: PromptLock Ransomware
Affected OS: Windows, macOS, and Linux

Attack: PromptLock ransomware, the first known AI-powered ransomware written in Golang, showcases how large language models can dynamically generate malicious code, making detection far more challenging. Unlike traditional attacks, this proof-of-concept uses Lua scripts, OpenAI's gpt-oss-20b model via the Ollama API, and the rare SPECK 128-bit encryption to demonstrate a future where malware constantly evolves, pressuring defenders to focus on behavior-based security over static signatures.

Attack Details

#1 PromptLock ransomware, written in Golang, is the first known example of AI-powered ransomware. Unlike traditional threats, PromptLock was created as a proof of concept rather than a fully deployed piece of malware. It leverages Lua scripts to steal and encrypt data across Windows, macOS, and Linux systems.

#2 The ransomware relies on hard-coded prompts to dynamically generate malicious Lua scripts. These scripts perform tasks such as filesystem enumeration, file inspection, data exfiltration, and encryption. To achieve this, PromptLock runs OpenAI's gpt-oss-20b model locally through the Ollama API, enabling real-time creation of harmful Lua code. For encryption, it employs the lightweight SPECK 128-bit algorithm, an uncommon choice in ransomware attacks.

#3 Once files are encrypted, PromptLock automatically generates ransom notes. These notes can include details such as a ransom demand and a Bitcoin address, specifically, the first Bitcoin address ever created. Since it is only a proof of concept, no actual data or transactions are involved.

#4 Importantly, PromptLock does not require downloading the full AI model onto a victim's system. Attackers can simply establish a proxy or tunnel from the compromised network to a server running the Ollama API with the gpt-oss-20b model.

#5 What makes PromptLock notable is its ability to evolve its code each time it runs. While polymorphic malware and fileless attack techniques have demonstrated similar adaptability in the past, here the large language model performs the core work. This capability significantly accelerates the process, forcing defenders to shift focus away from static signatures and instead prioritize behavioral analysis and anomaly detection.

Recommendations

Strengthen Behavioral Monitoring: Track and flag unusual Lua script executions. Detect anomalies in data access patterns. Employ intelligent event correlation to uncover hidden attack chains.

Enhance Network Controls: Continuously monitor for connections to unauthorized LLM services. In this case, watch for activity on **port 11434**, which may be exploited if attackers deploy Ollama locally. Implement stronger segmentation of critical networks to contain threats.

Next-Generation Security Tools: Implement AI-powered security platforms to detect and counter evolving threats, advanced behavioral detection systems to identify anomalies beyond static signatures, and automated intelligent response solutions to reduce reaction times during incidents.

Backup & Recovery Preparedness: Maintain offline, immutable, and regularly tested backups. Ensure recovery time objectives (RTOs) and recovery point objectives (RPOs) meet business continuity requirements in the event of ransomware deployment.

Implement the 3-2-1 Backup Rule: Maintain three total copies of your data, with two backups stored on different devices and one backup kept offsite or in the cloud. This ensures redundancy and protects against data loss from ransomware attacks.

Potential MITRE ATT&CK TTPs

TA0002 Execution	**TA0005** Defense Evasion	**TA0007** Discovery	**TA0009** Collection
TA0011 Command and Control	**TA0010** Exfiltration	**TA0040** Impact	**TA0042** Resource Development
T1588 Obtain Capabilities	**T1588.007** Artificial Intelligence	**T1059** Command and Scripting Interpreter	**T1027** Obfuscated Files or Information
T1620 Reflective Code Loading	**T1083** File and Directory Discovery	**T1005** Data from Local System	**T1119** Automated Collection
T1041 Exfiltration Over C2 Channel	**T1090** Proxy	**T1486** Data Encrypted for Impact	**T1490** Inhibit System Recovery
T1491.001 Internal Defacement	**T1020** Automated Exfiltration	**T1587** Develop Capabilities	**T1587.001** Malware

Indicators of Compromise (IOCs)

TYPE	VALUE
SHA1	24bf7b72f54aa5b93c6681b4f69e579a47d7c102, ad223fe2bb4563446aee5227357bbfdc8ada3797, bb8fb75285bcd151132a3287f2786d4d91da58b8, f3f4c40c344695388e10cbf29ddb18ef3b61f7ef, 639dbc9b365096d6347142fcae64725bd9f73270, 161cdcdb46fb8a348aec609a86ff5823752065d2
SHA256	2755e1ec1e4c3c0cd94ebe43bd66391f05282b6020b2177ee3b939fdd33216f6, 1612ab799df51a7f1169d3f47ea129356b42c8ad81286d05b0256f80c17d4089, b43e7d481c4fdc9217e17908f3a4efa351a1dab867ca902883205fe7d1aab5e7,

TYPE	VALUE
SHA256	09bf891b7b35b2081d3ebca8de715da07a70151227ab55aec1da26eb769c006f, e24fe0dd0bf8d3943d9c4282f172746af6b0787539b371e6626bdb86605ccd70, 1458b6dc98a878f237bfb3c3f354ea6e12d76e340cefe55d6a1c9c7eb64c9aee

References

https://www.welivesecurity.com/en/ransomware/first-known-ai-powered-ransomware-uncovered-eset-research/

https://openai.com/index/introducing-gpt-oss/

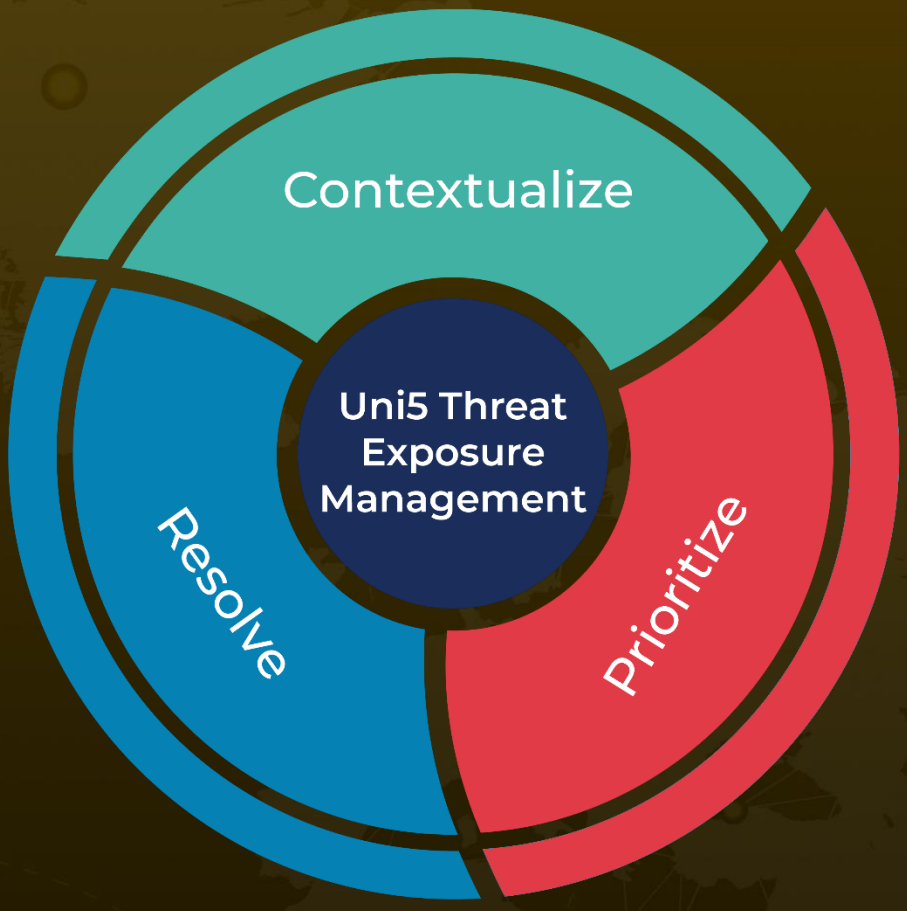
https://docs.runpod.io/tutorials/pods/run-ollama

https://arxiv.org/pdf/2508.20444

What Next?

At **Hive Pro**, it is our mission to detect the most likely threats to your organization and to help you prevent them from happening.

Book a free demo with **HivePro Uni5**: Threat Exposure Management Platform.



REPORT GENERATED ON

September 4, 2025 • 4:30 AM

More at www.hivepro.com