we run your business

PromptLock – Erste Ransomware mit echter KI-Unterstützung

Einleitung

Ein besorgniserregender Meilenstein in der Cybersecurity: Die Sicherheitsforscher von **ESET** haben Ende August 2025 **PromptLock** entdeckt – eine bisher einzigartige Ransomware, die mithilfe einer **lokalen KI** bösartige Aktionen automatisiert generiert.

Funktionsweise von PromptLock

- **Lokale KI-Nutzung**
  PromptLock nutzt OpenAIs Open-Weight-Modell **gpt-oss:20b**, das über die **Ollama API** lokal auf dem infizierten Rechner ausgeführt wird – eine cloud-freie, autarke Ausführung.
- **On-the-Fly Code-Generierung**
  Die Malware, geschrieben in **Go (Golang)**, enthält **vordefinierte Textprompts**, über die das LLM **Lua-Skripte generiert**, die dann bösartige Funktionen übernehmen – wie Systemprüfung, Dateierkennung, Datendiebstahl und Verschlüsselung.
- **Plattformübergreifend & flexibel**
  Dank Lua lassen sich die generierten Skripte auf **Windows, Linux und macOS** ausführen – eine deutlich höhere Angriffsfähigkeit.
- **Verschlüsselung & Tarnung**
  Für die Verschlüsselung verwendet PromptLock den **leichtgewichtigen SPECK-128-Bit-Blockcipher**, was ungewöhnlich ist, aber der Flexibilität dient.
- **Proof-of-Concept (PoC), nicht aktiv im Einsatz**
  Die Analyse deutet darauf hin, dass PromptLock ein **Prototyp oder Machbarkeitsnachweis** ist, noch nicht in realen Angriffen genutzt.
- **Eigenartige Signatur**
  Im Prompt ist eine **Bitcoin-Adresse eingebettet**, die mit **Satoshi Nakamoto** assoziiert wird – wohl eher symbolisch oder irreführend, nicht als tatsächliche Zahlungsadresse.

Bedeutung und Risiken

Handlungsempfehlungen für IT-Verantwortliche

1. **Aktivität von Lua-Skripten überwachen**
   Insbesondere solche, die atypische Operationen wie Systemauslesung oder Dateienverschlüsselung durchführen.
2. **Netzwerk-Logs prüfen**
   Suche nach ungewöhnlichen Verbindungen zu lokalen Ollama-API-Ports oder internen LLM-Systemen.
3. **IoCs einpflegen und überwachen**
   Nutze bekannte **SHA-Hashes** frühzeitig im Monitoring.
4. **Härtung gemäß Standards**
   Strikte Zugangskontrollen, Netzwerksegmentierung und Schutz gegen LLM-Prompt-Missbrauch einführen – beispielsweise durch Prompt-Guardrails.
5. **Statische Signaturen ergänzen**
   Erkenne die Grenzen signaturbasierter Tools – investiere in **verhaltensbasierte Erkennung** und **Sandbox-Analysen**.
6. **Weiterbildung & Awareness**
   Sichere Grundmaßnahmen nach BSI-Empfehlungen nicht vernachlässigen: Business Continuity, ISMS oder BSI-Grundschutz etablieren.

Fazit

PromptLock ist mehr als ein technisches Experiment – es ist ein **Weckruf**: LLMs können Malware dynamisch erzeugen, schwerer zu erkennen sein und plattformübergreifend agieren. Die zukünftige Verteidigung muss **ebenfalls adaptiv** sein – mit einem Fokus auf Verhalten, proaktiver Überwachung und robusten IT-Sicherheitsprozessen.

Quellen:

- Cybersecurity News: First AI Ransomware Uses GPT Model to Generate Malicious Scripts
  https://cybersecuritynews.com/first-ai-ransomware/
- ITPro: Researchers identify first AI-powered ransomware strain
  https://www.itpro.com/security/ransomware/security-researchers-have-just-identified-what-could-be-the-first-ai-powered-ransomware-strain-and-it-uses-openais-gpt-oss-20b-model
- TechRadar: First AI-powered ransomware discovered – why we should all be worried
  https://www.techradar.com/pro/security/the-first-ai-powered-ransomware-has-been-spotted-and-heres-why-we-should-all-be-worried
- The Hacker News: AI-Driven PromptLock Ransomware Emerges
  https://thehackernews.com/2025/08/someone-created-first-ai-powered.html
- GlobeNewswire: ESET discovers PromptLock
  https://www.globenewswire.com/news-release/2025/08/27/3140207/0/en/ESET-discovers-PromptLock-the-first-AI-powered-ransomware.html