O primeiro malware ‘MalTerminal’ alimentado por IA usa OpenAI GPT-4 pa

Data: 2025-09-21 04:23:02

Autor: Inteligência Against Invaders

O malware orientado por IA chamado ‘MalTerminal’ utiliza o GPT-4 da OpenAI para criar códigos prejudiciais como ransomware e shells reversos, indicando uma grande mudança na criação e implantação de ameaças. A descoberta fez parte da pesquisa “LLM-Enabled Malware In the Wild” da SentinelLABS apresentada na conferência de segurança LABScon 2025.

PromptLock: uma prova de conceito acadêmica:

Em agosto de 2025, a ESET ***fundar*** PromptLock, inicialmente identificado como o primeiro ransomware com inteligência artificial. Mais tarde, foi revelado que era uma prova de conceito por pesquisadores da NYU para mostrar os riscos de tais ameaças.

O PromptLock, escrito em Golang, opera localmente na máquina da vítima usando a API Ollama, ao contrário do MalTerminal, que usa uma API baseada em nuvem.

Com base em prompts predefinidos, o PromptLock gera scripts Lua maliciosos em tempo real, tornando-o compatível com Windows, Linux e macOS.

O malware detecta o tipo de sistema infectado – computador pessoal, servidor ou controlador industrial – e decide por conta própria se deseja exfiltrar ou criptografar dados com o algoritmo de criptografia SPECK de 128 bits.

MalTerminal descoberto:

Pesquisadores do SentinelLABS ***descoberto*** Malware habilitado para LLM durante o projeto de pesquisa PromptLock. Eles se concentraram em artefatos específicos da integração do LLM, em vez de código malicioso conhecido.

A equipe criou regras YARA para encontrar chaves de API codificadas e estruturas de prompt comuns em binários. Esse método detectou efetivamente scripts Python suspeitos e um executável do Windows chamado MalTerminal.exe.

A análise mostra que o malware usa um endpoint de API OpenAI desatualizado, indicando que foi desenvolvido antes de novembro de 2023, tornando-o a amostra mais antiga conhecida desse tipo.

O MalTerminal é um gerador de malware que permite aos usuários criar ‘Ransomware’ ou um ‘Reverse Shell’. Quando executado, ele solicita que a API GPT-4 gere o código Python malicioso relevante.

Esse método impede que o código malicioso seja armazenado no binário inicial, permitindo que ele evite a detecção por análise estática e ferramentas baseadas em assinatura.

A pesquisa encontrou scripts relacionados, como versões anteriores (TestMal2.py) e uma ferramenta defensiva chamada ‘FalconShield’, que parece ser um scanner de malware experimental feito pelo mesmo autor.

Malwares como MalTerminal e PromptLock apresentam um novo desafio para a segurança cibernética. Sua capacidade de criar código malicioso exclusivo para cada execução complica a detecção e a análise.

No entanto, esse tipo emergente de malware vem com suas próprias vulnerabilidades. Sua dependência de APIs externas, modelos locais e prompts codificados abre novos caminhos para os defensores explorarem.

Se uma chave de API for revogada ou um modelo for bloqueado, o malware não funcionará. Embora o malware habilitado para LLM ainda seja experimental, esses casos destacam a necessidade de os defensores se adaptarem, concentrando-se na detecção do uso malicioso de APIs e atividades incomuns de prompts.