ANÁLISE SEMANAL · SEMANA DE 20 A 27 DE OUTUBRO DE 2026

Sete servidores abertos, 8.257 mensagens e 46 criptografadores prontos: a bancada de IA de um afiliado de ransomware

SOCRadar STRU, 20 de outubro · 7 servidores, 133 sessões, 8.257 mensagens · 46 criptografadores para 6 vítimas · 62 organizações listadas, 36 nos EUA · US$ 23.920 em consumo de modelo

Em 20 de outubro a unidade de pesquisa da SOCRadar publicou o inventário de sete servidores expostos por afiliados das operações INC Ransomware e The Gentlemen. Dentro havia o histórico de 133 sessões de um agente de IA, somando 8.257 mensagens, a base de conhecimento dos operadores, o código de um arcabouço próprio chamado Phantom, 46 criptografadores compilados sob medida para seis vítimas, e uma lista de 62 organizações. A conta de uso dos modelos passava de US$ 23.920. O achado interessa menos pelo que diz sobre IA e mais pelo que diz sobre como esse tipo de operação realmente roda.

O que esta análise sustenta: a IA não foi a falha. Onde o material descreve como os atacantes entraram, a porta é a de sempre: equipamento de borda sem correção, ferramenta de acesso remoto exposta e credencial reaproveitada. O que a IA mudou foi o custo de operar depois de entrar, e a escala que uma equipe pequena consegue manter. Os US$ 23.920 de consumo de modelo não são orçamento de grupo grande.

O segundo achado é arquitetural e incomoda mais. O que tornou a operação eficiente foi a combinação de três coisas: acesso a ferramenta por protocolo padronizado, memória persistente entre sessões e um laço de aprovação humana antes de cada passo. É exatamente o desenho que times de defesa estão montando neste momento, com os mesmos componentes. A diferença entre o agente que responde a incidente e o agente que conduz a intrusão está no que ele tem permissão de fazer, não no que ele é.

Um diretório aberto em servidor de comando expôs o que normalmente ninguém vê: não o resultado de uma intrusão, mas o caderno de trabalho de quem a conduz. A unidade de pesquisa de ameaças da SOCRadar publicou o material em 20 de outubro, e há leituras independentes desde o dia seguinte. Os números a seguir vêm do registro dos próprios operadores, o que significa que são bons para entender método e ruins para medir impacto. Ninguém verificou se as seis implantações planejadas aconteceram.

O que estava nos servidores

Item	Quantidade	Observação
Servidores expostos	7	Ligados a afiliados do INC e, com menos certeza, do The Gentlemen
Sessões de agente	133	Interface OpenCode, com base de conhecimento persistente
Mensagens	8.257	Planejamento, geração de comando, depuração e redação
Criptografadores compilados	46	Sob medida para seis vítimas, com onda prevista para 1º de outubro

Item	Quantidade	Observação
Organizações listadas	62	36 nos Estados Unidos; indústria e manufatura em 16,1%, saúde e farmacêutica em 14,5%
Consumo de modelo	US$ 23.920	Modelos comerciais, entre eles Claude Opus 4.6 e Kimi K3
Servidores MCP	32	Conectados a um conjunto de ferramentas montado para a operação

O agente executa, o humano assina

O desenho é mais conservador do que a manchete sugere. Cada campanha começa com um prompt específico da vítima, e o agente espera autorização antes de agir. Uma das instruções recuperadas diz, em tradução direta, coordene cada ação comigo. O humano permanece no laço porque a operação depende disso: errar o alvo custa acesso.

O que o agente faz é trabalho de analista. Planeja reconhecimento, gera comando, depura ferramenta que quebrou, triagem de credencial roubada, separa o que presta no volume de dado exfiltrado e escreve o texto da extorsão. Nada disso é capacidade nova. Todas são tarefas que um afiliado experiente já fazia, e que agora cabem em menos gente e menos tempo.

A recusa do modelo não é controle de acesso

Quando o modelo comercial recusava um pedido, os operadores não trocavam de modelo nem buscavam versão sem restrição. Reescreviam o pedido como solicitação de teste autorizado, no formato de um pedido de red team com autorização declarada. Funcionou o suficiente para a operação continuar com modelo comercial.

Esse é o achado mais desconfortável do material, e vale separá-lo da discussão sobre alinhamento. A recusa opera por mensagem, uma de cada vez, e avalia o texto que recebe. Ela não tem como verificar se a autorização alegada existe, não enxerga a sessão inteira e não sabe o que aconteceu nas demais. Detecção de abuso do lado do provedor enfrenta o mesmo problema por outro ângulo: o tráfego parecia atividade de desenvolvedor, distribuída entre sessões e contas.

A entrada continuou analógica

O registro dos operadores traz contagens de exploração que não foram verificadas de forma independente e devem ser lidas como alegação. Ainda assim, o padrão é claro e combina com o que se vê em resposta a incidente.

- Uma falha em painel de hospedagem, com mais de 49 invasões de provedores reivindicadas.
- Uma falha conhecida em ferramenta de suporte remoto, com varredura de 4.922 servidores que, segundo os registros, rendeu 427 senhas de administrador e 242 endereços distintos listados como afetados.
- Uma falha em gerenciador de endpoint corporativo, com 2.236 alvos, 108 contas de administrador criadas e 4.140 estações sob gestão.

Nenhuma dessas entradas tem relação com IA. São equipamento de borda e ferramenta de administração remota sem correção, mais credencial reaproveitada. Uma captura de tela de 30 de setembro, de um console de suporte remoto, mostra criptografador em execução em estação de vítima, que é a evidência de que pelo menos um caso saiu do papel.

Três evidências para levar ao conselho

Primeira: o custo de operar caiu, não a dificuldade de entrar. Os US$ 23.920 de consumo de modelo e sete servidores sustentaram uma lista de 62 alvos. Qualquer raciocínio de risco baseado em quantas equipes capazes existem no mundo precisa ser refeito, porque a capacidade deixou de depender do tamanho da equipe.

Segunda: o perímetro que falhou é o mesmo de sempre. Correção de equipamento de borda e de ferramenta de acesso remoto, restrição de acesso de RMM e de prestador, credencial única com segundo fator e cópia de segurança imutável continuam sendo o que separa a lista de alvos da lista de vítimas. Não há controle novo a comprar por causa deste relatório.

Terceira, e é a que ainda não tem resposta pronta: se a sua equipe está montando agente com acesso a ferramenta por MCP e memória persistente, você está construindo a mesma arquitetura, com as mesmas lacunas de autorização. A recomendação que sai do material é tratar ação de impacto alto fora do modelo, com autorização em outra camada, limitar a lista de ferramentas por escopo, registrar e correlacionar atividade por conta e não por mensagem, e proteger a memória do agente como repositório de dado sensível. Vale perguntar, na próxima revisão, qual dessas quatro já existe.

Uma nota de método para fechar. Este texto usa os números do inventário dos atacantes porque não há outros, e eles descrevem intenção e volume de trabalho, não dano confirmado. Onde a fonte diz alegado, mantivemos alegado. É uma distinção chata e é justamente ela que separa análise de propaganda.

Três controles que o material cobra

01 · AUTORIZAÇÃO

A permissão precisa viver fora do modelo

A recusa do modelo avalia uma mensagem e aceita a autorização que o texto alega. Ação de impacto alto, como apagar, cifrar, mover volume grande ou desligar serviço de segurança, precisa de verificação em camada separada, que consulte quem pediu e com qual direito, e não o que o prompt afirma. Enquanto isso não existir, filtro de conteúdo será confundido com controle de acesso.

02 · CORRELAÇÃO

O sinal aparece na conta, não na mensagem

O tráfego da operação passou por parecer atividade de desenvolvedor espalhada entre sessões e contas. Detecção por mensagem isolada não encontra isso. O que encontra é correlação por conta ao longo do tempo: volume anômalo, sequência de tarefa incompatível com o perfil declarado e uso concentrado em janelas estranhas. Isso vale tanto para o provedor de modelo quanto para quem paga a fatura corporativa.

03 · MEMÓRIA

A memória do agente é repositório de dado sensível e raramente está classificada assim

Base de conhecimento persistente foi o que permitiu retomar campanha entre sessões sem recontar o contexto. Do lado defensivo, a memória do agente acumula topologia de rede, nome de conta, caminho de sistema e trecho de investigação. Se ela não está no inventário de dado, não está no plano de retenção nem no escopo de revisão de acesso, e quem comprometer o agente leva o mapa junto.

Fontes

Hackread · INC and The Gentlemen Ransomware Affiliates Expose 8,257 AI Messages, com os sete servidores, as 133 sessões, os 46 criptografadores, os 62 alvos, o consumo de US$ 23.920 e as contagens de exploração alegadas pelos operadores.

Seif Hateb · Deep Dive into an AI-Assisted Ransomware Attack, com os 32 servidores MCP, o arcabouço Phantom, a análise sobre recusa de modelo como filtro por mensagem e as recomendações para plataforma de IA.

Publicado em latamsec.com.br/recursos/bancada-de-ia-afiliados-inc-gentlemen/ · 2026-10-27