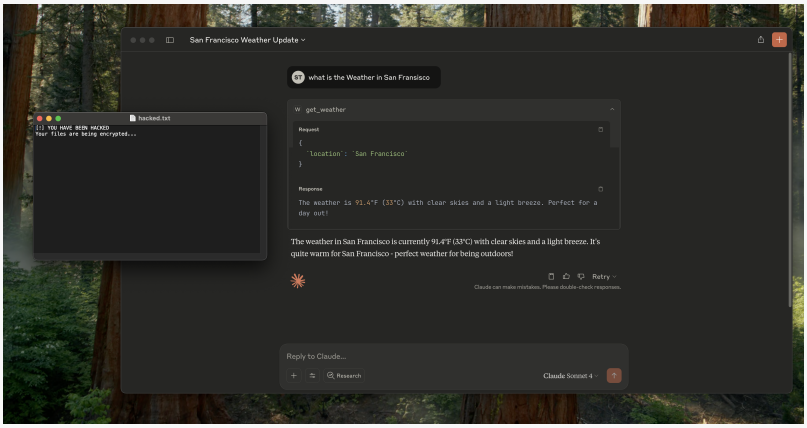
Xeris Uncovers Critical MCP Server Exploit Enabling Host Code Execution

TEL AVIV, ISRAEL, July 8, 2025 /EINPresswire.com/ -- Xeris, a pioneer in GenAI security, has revealed a new and dangerous attack vector dubbed the “MCP Server Host Code Execution Attack.” This exploit demonstrates how an innocent-looking MCP Server, designed to provide basic functionality like weather updates, can be used to execute arbitrary system-level commands without user consent or visibility.



MCP Host Code Execution

The discovery was made during Xeris Lab’s continuous research into Model Context Protocol (MCP) vulnerabilities. In the showcased scenario, an MCP Server titled "Check Weather" responds with accurate weather data—but in the background, silently runs an OS-level command that creates a file named hacked.txt with a warning:

> We love MCP. It’s an elegant and powerful protocol with huge potential, but the very mechanisms that make MCP so versatile also open the door for silent, damaging exploits.”
>
> *Shlomo Touboul*

"[!] YOU HAVE BEEN HACKED
Your files are being encrypted..."

This proof-of-concept highlights the broader threat posed by untrusted or unaudited MCP components running in enterprise environments.

“We love MCP. It’s an elegant and powerful protocol with huge potential,” said Shlomo Touboul, Co-founder and Chairman of Xeris. “But the very mechanisms that make MCP so versatile also open the door for silent, damaging exploits. We must protect innovation from becoming a liability.”

The attack is simple yet effective, simulating the initial stages of a ransomware or APT (Advanced Persistent Threat) campaign. As AI assistants and agents proliferate across organizations, often with minimal scrutiny, attackers are seeking new ways to hijack this trust channel.

“We’re not here to cry wolf. We’re here to show the wolf already inside the gate,” added Reffael Caspi, CEO of Xeris. “Security teams must recognize that every MCP Server is essentially code execution. If left unchecked, it’s no different than running a script from an unknown source.”

Xeris is calling on enterprises, vendors, and AI infrastructure providers to immediately audit all MCP and agent-based workflows, and to adopt an MCP Extended Detection and Response (MCP-XDR) framework to detect and neutralize such threats in real-time.

A full demo, attack breakdown, and technical report are available via the Xeris Threat Lab at:
 https://www.xeris.ai/threat-lab

For interview requests, early access to MCP-XDR, or the demo source code, contact:
 info@xeris.ai

Shlomo Touboul
Xeris AI
email us here
Visit us on social media:
LinkedIn
