Investigadores en ciberseguridad han identificado una vulnerabilidad crítica en el proyecto *Model Context Protocol (MCP) Inspector* de la empresa de inteligencia artificial Anthropic, la cual podría permitir la ejecución remota de código (RCE) y dar acceso total al sistema afectado.

La vulnerabilidad, registrada como CVE-2025-49596, tiene una puntuación CVSS de 9.4 sobre 10, indicando una severidad muy alta.

> *"Se trata de una de las primeras fallas críticas de ejecución remota en el ecosistema MCP de Anthropic, lo que revela una nueva clase de ataques basados en navegador dirigidos a herramientas de desarrollo de IA"*, declaró Avi Lumelsky de Oligo Security en un informe publicado la semana pasada.
>
> *"Al obtener ejecución de código en el equipo de un desarrollador, los atacantes pueden robar información, instalar puertas traseras y moverse lateralmente por redes —lo que plantea riesgos importantes para equipos de IA, proyectos de código abierto y empresas que utilizan MCP."*

Presentado por Anthropic en noviembre de 2024, MCP es un protocolo abierto que estandariza cómo las aplicaciones basadas en modelos de lenguaje (LLM) integran y comparten datos con herramientas o fuentes externas.

MCP Inspector es una herramienta para desarrolladores que permite probar y depurar servidores MCP, los cuales exponen capacidades específicas mediante el protocolo, facilitando que un sistema de IA acceda a información más allá de su entrenamiento.

Está compuesto por dos partes: un cliente con interfaz interactiva para pruebas y depuración, y un servidor proxy que actúa como puente entre la interfaz web y diversos servidores MCP.

Sin embargo, es fundamental tener presente que este servidor no debe estar expuesto a

redes no confiables, ya que tiene permiso para ejecutar procesos locales y conectarse con cualquier servidor MCP especificado.

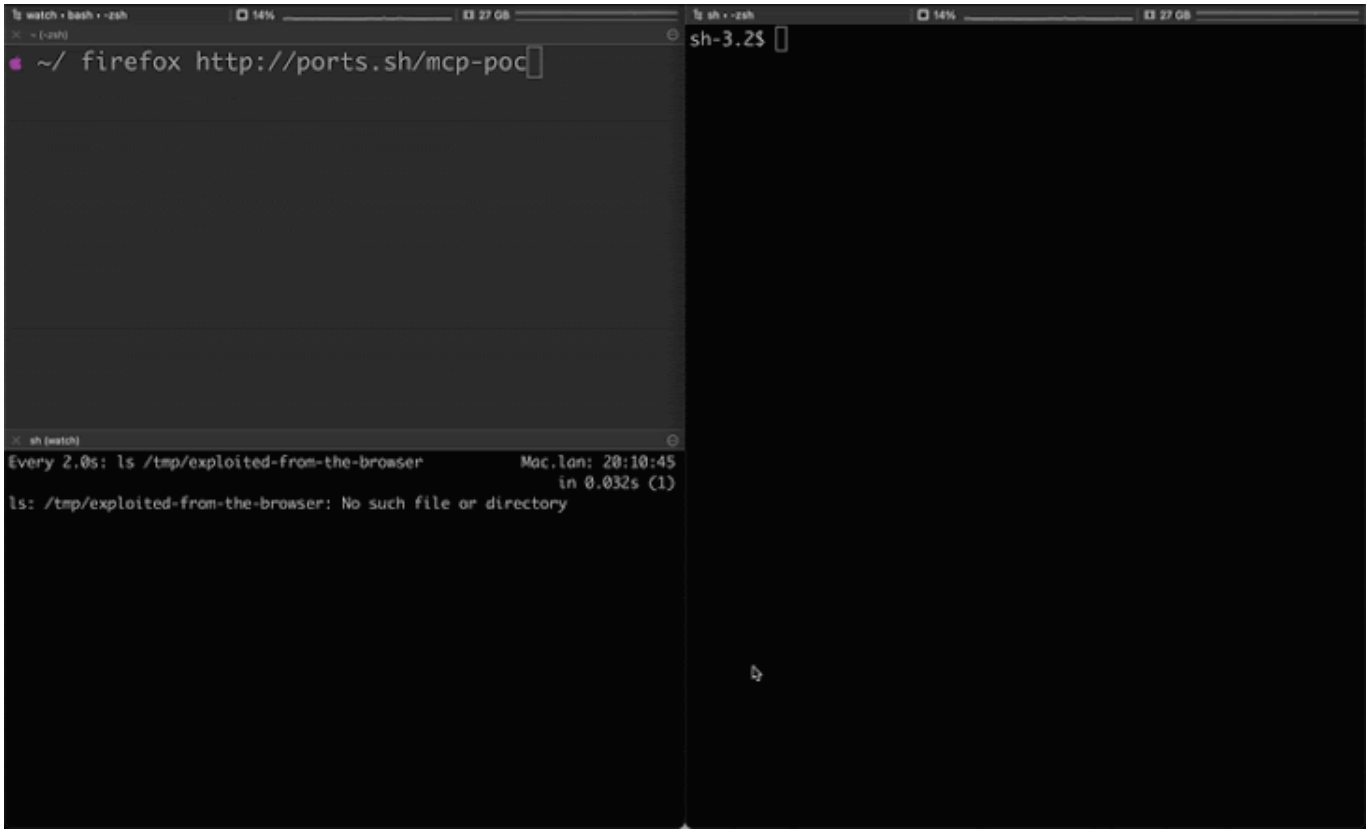
Este detalle, sumado al hecho de que los desarrolladores suelen iniciar la herramienta con configuraciones predeterminadas que carecen de autenticación y cifrado, genera graves riesgos de seguridad, según Oligo.

> *"Esta mala configuración crea una superficie de ataque considerable, ya que cualquier persona en la red local o en internet podría interactuar y explotar estos servidores",* advirtió Lumelsky.

El ataque aprovecha la combinación de una falla ya conocida en navegadores modernos, llamada 0.0.0.0 Day, junto con una vulnerabilidad CSRF en Inspector (CVE-2025-49596), permitiendo la ejecución de código simplemente al visitar un sitio web malicioso.

> *"Las versiones de MCP Inspector anteriores a la 0.14.1 son vulnerables a ejecución remota de código debido a la falta de autenticación entre el cliente Inspector y el proxy, permitiendo solicitudes no autenticadas que ejecutan comandos MCP vía stdio",* indicaron los desarrolladores en el aviso sobre CVE-2025-49596.

0.0.0.0 Day es una vulnerabilidad con 19 años de antigüedad presente en navegadores actuales que puede ser usada por sitios maliciosos para acceder a redes locales. Se basa en el manejo inseguro de la IP 0.0.0.0, que puede resultar en ejecución de código.



> *"Los atacantes pueden explotar esta falla mediante una página web diseñada para enviar peticiones a servicios locales corriendo en un servidor MCP, logrando así ejecutar comandos arbitrarios en el equipo del desarrollador",* explicó Lumelsky.
>
> *"El hecho de que la configuración por defecto exponga estos servidores a tales ataques significa que muchos desarrolladores podrían estar abriendo sin saberlo una puerta trasera a sus sistemas."*

En concreto, la prueba de concepto (PoC) utiliza el endpoint Server-Sent Events (SSE) para enviar una solicitud maliciosa desde una página web controlada por el atacante, con el objetivo de ejecutar código en la máquina donde se ejecuta la herramienta, incluso si solo escucha en localhost (127.0.0.1).

Esto es posible porque la dirección 0.0.0.0 indica al sistema operativo que debe escuchar en todas las interfaces IP del equipo, incluyendo la interfaz de bucle local (localhost).

En un escenario de ataque, un atacante podría crear una página web falsa y engañar a un desarrollador para que la visite. En ese momento, un script malicioso embebido en la página enviaría una petición a 0.0.0.0:6277 (puerto por defecto del proxy), instruyendo al servidor MCP Inspector a ejecutar comandos arbitrarios.

Además, el ataque puede incorporar técnicas de DNS rebinding para generar registros DNS falsos que apunten a 0.0.0.0:6277 o 127.0.0.1:6277, eludiendo controles de seguridad y logrando ejecución remota de código.

Tras una divulgación responsable en abril de 2025, los responsables del proyecto corrigieron el fallo el 13 de junio, lanzando la versión 0.14.1, que incorpora un token de sesión para el proxy y validación de origen para bloquear el vector de ataque.

> *"Aunque los servicios en localhost parezcan seguros, muchas veces están expuestos a internet debido a las capacidades de enrutamiento en navegadores y clientes MCP"*, indicó Oligo.
>
> *"La solución agrega autorización (que antes faltaba por defecto), además de validar los encabezados Host y Origin en las peticiones HTTP, asegurando que el cliente proviene de un dominio confiable. Ahora, por defecto, el servidor bloquea ataques de DNS rebinding y CSRF."*