Estado de cyberdeception

Revisión inicial · 15 de septiembre de 2026

Este informe conecta investigación académica, herramientas abiertas, productos y experiencia operativa. Su alcance es la información pública verificada en el catálogo. Las conclusiones distinguen lo que sostienen los autores, lo que declaran los proveedores y lo que observaron terceros. Las fichas basadas solo en metadatos no se usan para inferir resultados científicos.

Definición y límites del campo

Cyberdeception introduce señales o recursos diseñados deliberadamente para que un adversario los vea, los use o tome decisiones a partir de ellos. Un honeypot simula un sistema o servicio; un honeytoken es un dato señuelo cuya utilización produce una señal; otros activos pueden representar credenciales, archivos, aplicaciones o rutas de movimiento. El objetivo puede ser detectar, observar, desviar o imponer un costo a la actividad adversaria. MITRE Engage ofrece un marco para planificar deception, denial y adversary engagement.

La terminología cambia según la disciplina. Una revisión de cyberdeception de 2024 propone una taxonomía transversal y analiza líneas con y sin IA. Una revisión basada en teoría de juegos distingue perturbación, moving target defense, obfuscation, mixing, honey-x y attacker engagement. Por eso este Atlas registra técnica, entorno y objetivo como dimensiones separadas. Moving target defense puede formar parte de una estrategia de engaño, pero no toda implementación debe aparecer automáticamente como cyberdeception.

Qué existe hoy

El software abierto cubre varios niveles de interacción. OpenCanary implementa un honeypot de red multiprotocolo orientado a detectar actividad en redes internas. Cowrie se centra en SSH y Telnet. T-Pot reúne varios honeypots y herramientas de análisis. Proyectos como Galah exploran interacciones web basadas en modelos de lenguaje. Sus fichas registran documentación y mantenimiento; figurar aquí no implica una evaluación comparativa de rendimiento.

El mercado comercial combina señuelos, datos trampa e integración con operaciones de seguridad. Ejemplos documentados son Thinkst Canary, FortiDeceptor, Acalvio ShadowPlex, CounterCraft The Platform, Zscaler Deception y Tracebit. Las capacidades de esas fichas proceden, por ahora, de los proveedores. Precio, facilidad de despliegue y resultados necesitan documentación pública o pruebas independientes para poder compararse.

Los servicios incluyen diseño de estrategia, instalación, operación gestionada, pruebas y capacitación. Su clasificación merece cuidado: un fabricante que ofrece soporte no equivale necesariamente a un proveedor de operación gestionada. El catálogo registra cada servicio solo cuando encuentra una oferta verificable.

Evidencia de uso y resultados

El NCSC británico informó en diciembre de 2025 que su programa incluyó 121 organizaciones, 14 proveedores comerciales y 10 pruebas de productos en entornos diversos. Sus hallazgos señalan potencial para detección e inteligencia, pero también dificultades de terminología, falta de métricas basadas en resultados, necesidad de orientación imparcial y riesgos de configuración. El NCSC subraya que los señuelos requieren estrategia y contexto operativo.

Los relatos de clientes ayudan a identificar problemas de adopción, aunque su procedencia cambia el peso de la evidencia. La experiencia de Riot Games publicada por Tracebit describe el interés en trasladar enfoques de deception a cloud; el Atlas la muestra como caso publicado por el proveedor. Una evaluación independiente, si existe, deberá aparecer como fuente adicional.

Investigación y tendencias

La literatura explora mecanismos de red y host, modelos de decisión, selección de señuelos y medición. La revisión de Lu y colaboradores organiza esquemas estratégicos, de red, de host y criptográficos. Un trabajo sobre simulación de ataques propone evaluar honeypots y moving target defense con un modelo de red. Otro survey estudia requisitos de red para implementar cyberdeception eficaz.

Los honeypots basados en modelos de lenguaje son una línea visible, aún heterogénea. LLM Honeypot describe ajuste de un modelo con comandos de atacantes y una evaluación de respuestas. HoneyGPT presenta un honeypot de terminal y una evaluación en campo. Estos resultados pertenecen a los trabajos citados; no prueban que todos los modelos de lenguaje sean seguros, económicos o más eficaces que señuelos tradicionales.

Qué falta medir

Una evaluación útil debe describir ubicación de señuelos, tráfico legítimo, ataques observables, calidad de las alertas, trabajo de mantenimiento y consecuencias de una interacción. Las métricas pueden incluir tiempo hasta detección, eventos útiles por señuelo, tasa de alertas que requieren investigación, profundidad de interacción y costo operativo. El NCSC identificó la ausencia de métricas de resultado como una brecha. Contar alertas sin contexto no permite concluir que una implementación haya reducido riesgo.

El catálogo también debe registrar resultados negativos, ataques que identifican señuelos, proyectos discontinuados, cambios de nombre y datos no publicados. Esa información es necesaria para investigadores que buscan reproducibilidad y para equipos que comparan inversiones.

Cobertura y próximas revisiones

Esta versión inicia con las fuentes abiertas y las fichas piloto. La cobertura por sector, región, idioma y técnica se publicará junto al catálogo. No se deducirá una distribución global del mercado a partir de una muestra incompleta. La siguiente revisión incorporará lectura de textos completos, experiencias operativas adicionales y pruebas reproducibles, identificando qué conclusiones cambian.