ZERO TRUST CONSULTING
ZTC RESEARCH / INVESTIGACIÓN INDEPENDIENTE

LA NUEVA ASIMETRÍA DE LA CIBERSEGURIDAD

La arquitectura de contención como límite verificable del daño en la era de la IA

APORTE METODOLÓGICO / CCA-01

Una barrera puede cortar una ruta de ataque y también interrumpir una operación esencial.

CCA-01 integra en una sola regla la alcanzabilidad residual de activos críticos, la continuidad de funciones esenciales y la suficiencia del inventario de dependencias. Señala rutas abiertas y fallas operativas; con evidencia insuficiente, la decisión es no estimable.

DIRECCIÓN DE INVESTIGACIÓN

Dr. Juan Pablo Ybarra Llamas
ZTC Research / Zero Trust Consulting

VERSIÓN 1.0 2 de noviembre de 2026 | México
ZTCR-AI-VM-2026-01 zerotrust.consulting/research

00 Síntesis ejecutiva y contribución

La aceleración de la investigación de vulnerabilidades apoyada por IA cambia el costo de descubrir y analizar fallas, pero no permite inferir que todos los CVE se exploten de inmediato. El problema de gobierno es más preciso: durante el intervalo entre descubrir una exposición y corregirla, ¿qué rutas de privilegio, datos o servicio siguen abiertas y qué funciones esenciales permanecerían disponibles ante un aislamiento?

El aporte metodológico de ZTC Research es CCA-01: una regla de decisión que vincula tres cuestiones que no pueden resolverse por separado: (i) qué rutas adversarias permiten todavía alcanzar activos o datos críticos; (ii) qué funciones esenciales deben seguir operando bajo los mismos límites de defensa; y (iii) si el modelo de dependencias es suficientemente completo para sostener el resultado. Una arquitectura que impide la propagación pero inutiliza un proceso esencial no supera el criterio conjunto; tampoco una que conserva el servicio y mantiene rutas críticas abiertas. Ante evidencia material insuficiente, el resultado es no estimable, no un cumplimiento supuesto.

UNA DECISIÓN DISTINTA

Un inventario de parches responde cuándo corregir un defecto. CCA-01 aspira a responder qué consecuencias pueden producirse mientras existe ese defecto y qué debe seguir operando. La ausencia de información no cuenta como protección; el bloqueo de una ruta que destruye la función crítica tampoco cuenta como éxito.

Los datos de la obra incluyen series públicas de divulgación y explotación con sus denominadores y dos bancos sintéticos separados: 12 escenarios del índice ponderado B, y 16 escenarios del vector PDC de daño y continuidad. Son ilustraciones reproducibles de reglas lógicas, no mediciones empíricas de empresas, proveedores o infraestructuras. Su valor está en identificar las condiciones en que un evaluador debe fallar o abstenerse.

La propuesta se ofrece a la discusión de comunidades técnicas, responsables de marcos de seguridad, autoridades financieras y organizaciones que gestionan funciones críticas. NIST, ISO/IEC, PCI SSC y CPMI-IOSCO constituyen referencias pertinentes por sus respectivos ámbitos de competencia; el G7 Cyber Expert Group es un foro de diálogo sobre resiliencia financiera, no un emisor universal de estándares. Se plantea examinar si la evidencia de contención y continuidad de CCA-01 aporta valor adicional a los instrumentos existentes, sin solicitar su adopción inmediata ni atribuirle carácter normativo.

International abstract

AI-assisted vulnerability research is changing the economics of discovery and analysis. ZTC Research proposes CCA-01, an auditable architectural containment criterion that jointly examines the reachability of critical assets after an assumed initial compromise and the continuity of essential legitimate functions. An incomplete dependency model must not be mistaken for effective protection. The study aligns the proposal with existing NIST, ISO/IEC, PCI SSC and CPMI-IOSCO frameworks while carefully distinguishing documentary evidence, constructed synthetic examples and operational validation. We invite standards bodies, financial authorities and critical-sector organizations to assess the proportional incorporation of this outcome-oriented evidence into their own assurance and governance practices.

Estructura de la investigación

Evidencia y fuentes públicas · Arquitectura y consecuencias · Protocolo CCA-01 · Ensayos sintéticos y refutación · Correspondencia normativa · Invitación a evaluación técnica · Límites y recomendaciones.

01 Pregunta, hipótesis y método



Figura 1. Secuencia de decisión desde publicación hasta impacto. Es un mapa de relaciones, no una inferencia causal universal.

Hipótesis contrastables

H1. Las divulgaciones de CVE crecen aceleradamente. H2. La IA contribuye al descubrimiento validado, sin que pueda asignársele toda la variación mundial. H3. El riesgo depende de la diferencia entre entrada de hallazgos aplicables y capacidad efectiva de contención/cierre. H4. La arquitectura reduce la alcanzabilidad y el radio de impacto de defectos explotables, hipótesis que requiere pruebas de aislamiento, privilegios y exfiltración por activo.

Se aplica una revisión crítica de fuentes públicas, comparación de series cuantitativas y análisis de casos técnicos. No se realizó auditoría de código, explotación, escaneo activo ni investigación forense propia. Se distinguen sistemáticamente hecho registrado, declaración del proveedor, inferencia y escenario de trabajo.

Qué mide cada fuente

Fuente	Unidad observada	Uso y límite
CVETodo	CVE publicados por mes; NVD/CVE/KEV	Serie consolidada 2025–2026, no fuente primaria de CVE. No identifica universalmente el método de descubrimiento.
CVE Tools	CVE mensuales y altas mensuales a KEV	Corroboración con resultados cercanos, pero no idénticos. KEV registra otra categoría y otra fecha.
GTIG	Divulgación, riesgo interno y explotación observada	Datos de inteligencia con cobertura y criterios propios; no son una auditoría universal de todos los incidentes.
IBM / Unit 42 / Securin	Muestras y campañas atribuidas a IA	Prueba de capacidad técnica; tienen sesgos de selección, diversidad de definiciones y procedencia.
FIRST	Pronósticos de divulgación	Proyecciones hechas en fechas previas; comparar con observaciones posteriores exige respetar métodos y cortes.

Regla metodológica. El término “descubrimiento por IA” se reserva a resultados con declaración explícita de la metodología en fuente primaria o atribución pública contrastable. “CVE divulgado” no es sinónimo de “vulnerabilidad descubierta por IA”; “alta en KEV” no equivale a “explotación originada ese mes”.

Evidencia y niveles de confianza

Nivel	Criterio	Ejemplo
Alto	Advisory del mantenedor, datos y	PHP Group; GTIG para sus propios

Nivel	Criterio	Ejemplo
	metodología explícitos, registros técnicos de una entidad primaria	datos; CVETodo para su serie
Medio	Estudios propios de proveedor con cifras verificables pero reproducibilidad parcial	NOVA; muestra IBM; tracker Securin
Exploratorio	Opinión, tweet, pronóstico o escenario sin comprobación empírica externa	Horizonte de 18 meses; modelos de saturación

02 El cambio observado en el volumen de vulnerabilidades

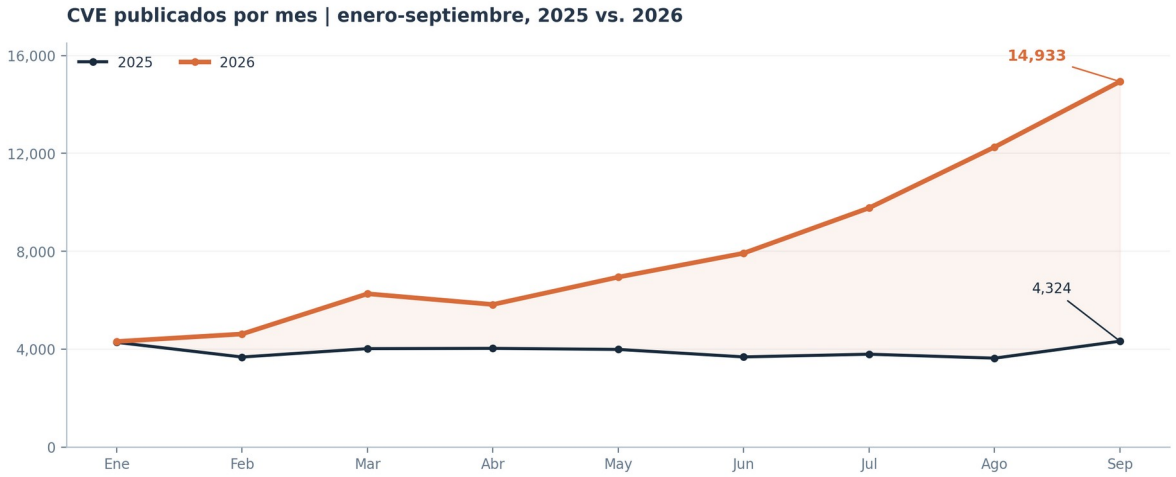


Figura 2. Serie mensual de CVE publicados, enero-septiembre de 2025 y 2026. Elaboración ZTC Research con datos publicados por CVETodo [01].

La diferencia entre enero y septiembre de 2026 es particularmente relevante: 4,311 frente a 14,933 publicaciones, un crecimiento de 246.4% dentro del año. La comparación interanual septiembre-septiembre arroja 245.4%. En los primeros nueve meses se observan 72,833 frente a 35,415 (+105.7%). [01]

El cambio se produce en meses, no sólo en un ciclo anual. Eso desafía procesos de remediación organizados como proyectos trimestrales aislados. No obstante, se requiere calcular cuántos CVE son realmente aplicables al inventario de cada empresa, cuáles poseen rutas de ataque viables y cuáles alcanzan servicios esenciales antes de traducir esta curva en riesgo.

Validación cruzada y diferencias legítimas

La consulta de CVE Tools registra 15,180 CVE en septiembre de 2026 y 4,664 en septiembre de 2025, frente a 14,933 y 4,324 en CVETodo. Los resultados confirman dirección y magnitud del salto, pero no son intercambiables. Las diferencias pueden provenir del corte temporal, las sincronizaciones, reglas de inclusión y actualización posterior de registros. [01–02]

GTIG registró 5,045 divulgaciones en enero de 2026 y 10,740 en agosto, según su propia cobertura; el incremento no es idéntico al de CVETodo. FIRST revisó en junio su proyección anual central de aproximadamente 59,427 a 66,000 CVE. Esas proyecciones son hechos históricos sobre la expectativa de entonces, no recuentos finales. [03–05]

Volumen de publicaciones frente a explotación observada

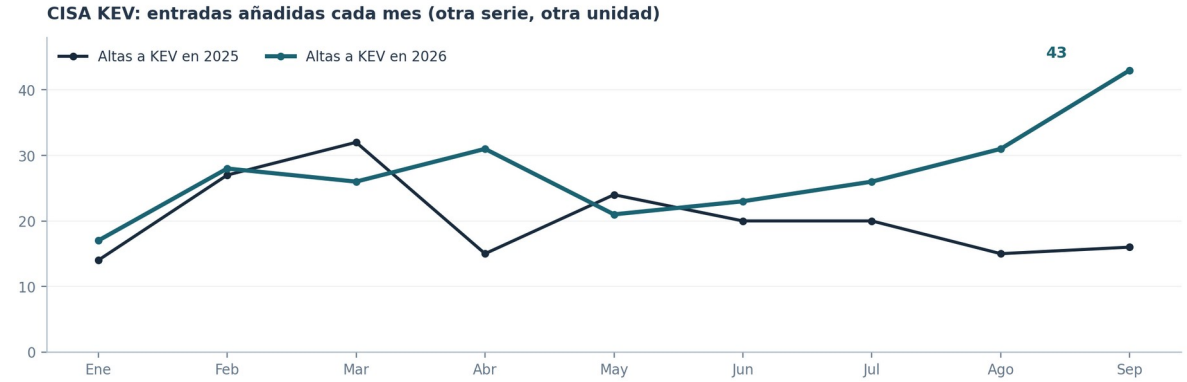


Figura 3. Altas mensuales al catálogo CISA KEV según CVE Tools: enero-septiembre de 2025 y 2026. La fecha corresponde a la inclusión en KEV, no a la publicación original del CVE. [02]

No procede dividir sin más las entradas de KEV por los CVE publicados en el mismo mes y llamar al resultado “tasa real de explotación”: ambos flujos pueden corresponder a poblaciones y fechas distintas. El contraste sirve para ilustrar órdenes de magnitud y la importancia de contextualizar. [02]

GTIG, con un método de observación distinto, reportó 141 vulnerabilidades divulgadas y explotadas activamente entre enero y agosto de 2026, unas 18 al mes; en 2025 la media fue de 10.5. Para esa ventana de 2026, la fracción observada explotada fue aproximadamente 0.23% del total divulgado. Es una medición acotada por visibilidad: no demuestra ausencia de explotación en los demás registros. [03]

El desajuste medido entre riesgo y corrección

El Data Breach Investigations Report 2026 de Verizon informa que, para su conjunto de datos de 2025, únicamente el 26% de las vulnerabilidades incluidas en CISA KEV estaban completamente corregidas; la mediana para la resolución completa fue de 43 días, frente a 32 días en el periodo anterior. No son métricas globales de todas las empresas ni tiempos desde la publicación de cada CVE. Se refieren a la cobertura y unidad de análisis definidas por Verizon. [29]

Ese rezago ayuda a explicar el valor de la arquitectura: una organización que no puede corregir de inmediato necesita reducir alcanzabilidad, accesos y exposición de datos mientras corrige. No sería correcto restar directamente los 43 días de Verizon a los 3 días de una obligación federal: poblaciones, relojes y criterios difieren. La decisión razonable es medir por separado el tiempo a protección, el tiempo a cierre y las instancias que siguen expuestas. [20, 29]

03 Lo que sí podemos atribuir a inteligencia artificial

Evidencia de capacidad: no equivale a porcentaje del universo

NOVA (Unit 42). El grupo informó haber examinado 3,915 proyectos abiertos y confirmado 14,090 hallazgos en dos meses, de los cuales 99.4% no estaban reportados previamente. El 39.7% obtuvo clasificación alta o crítica bajo CVSS v4.0. Son hallazgos de una campaña de investigación y validación, no 14,090 CVE emitidos ni 14,090 incidentes. El estudio procede del desarrollador del sistema; la validación externa completa de la muestra no está publicada en la misma extensión. [07]

Muestra IBM X-Force. En 223 vulnerabilidades públicamente reportadas como descubiertas con modelos de frontera, IBM observó 55.2% de severidad alta/crítica, 91% accesibles remotamente y 25.6% compatibles con ejecución de código. Las cifras describen esa muestra seleccionada; no estiman las proporciones del universo de CVE ni prueban que la IA encuentre exclusivamente fallas graves. [06]

Tracker Securin. Reúne 597 CVE relacionados con actores o sistemas de IA, pero diferencia la calidad de la atribución: 356 tienen evidencia pública explícita de descubrimiento con IA (166 autónomos y 190 con

asistencia humana); otros 178 sólo están ligados a personas de un laboratorio, 14 carecen de método declarado y 49 no presentaban evidencia pública de atribución localizada. La precisión de esta separación es tan importante como el recuento bruto. [08]

Antecedente verificable. Google Project Zero y DeepMind publicaron en 2024 un hallazgo real identificado mediante Big Sleep en SQLite; en 2025 se documentó CVE-2025-6965 con apoyo del mismo programa. El avance técnico precede, por tanto, a la aceleración de registros de 2026. [13, 17]

Por qué falla una inferencia causal fácil

El número de CVE puede crecer por investigación de IA, ampliación del programa CVE, automatización del etiquetado, cambios en las autoridades de numeración, mayor transparencia, oleadas de divulgación de proveedores, revisión masiva de proyectos antiguos o mejoras de sincronización. GTIG destacó alrededor de 5,000 CVE vinculados a descripciones de Linux Kernel entre enero y agosto de 2026 y señaló que no observó zero-days explotados activamente para ese subconjunto. [03]

Para calcular el efecto causal de IA sería necesario vincular cada hallazgo al método de descubrimiento, deduplicar exposiciones derivadas, controlar proveedores, tipos de software y ciclos de divulgación, y comparar con un contrafactual. La información pública disponible no permite hacerlo con robustez. Por eso, el crecimiento observado de CVE es un hecho; atribuirlo enteramente a IA sería una conclusión no demostrada.

04 PHP2Shell: un caso emblemático y sus límites

El 21 de octubre de 2026, PHP anunció correcciones de CVE-2026-109618 y CVE-2026-109619. Ambas pueden permitir ejecución remota de código sin autenticación en contextos afectados. La primera involucra manejo inadecuado de memoria ante agotamiento; la segunda, estado interno conservado después de liberar memoria. El proyecto señaló mayor riesgo para mod_php y advirtió que la actualización es la protección completa disponible. [09]

OpenAI informó que Calif identificó una de las fallas con un modelo de frontera y que Wiz descubrió y notificó ambas de forma independiente. Los mantenedores de PHP encabezaron la corrección y coordinaron la respuesta con otros actores. La evidencia documenta una contribución real de herramientas de IA, con investigación humana y mantenimiento de software en el centro del proceso. [10]

Familia PHP	Versión corregida mínima	Lectura operativa
8.2	8.2.35	Planificar migración antes de fin de soporte; verificar binario desplegado.
8.3	8.3.36	Actualizar y comprobar procesos efectivos.
8.4	8.4.27	Actualizar y verificar todas las instancias/containers.
8.5	8.5.12	Actualizar; en Windows validar reconstrucciones posteriores cuando corresponda.

Matiz de alcance. La presencia de PHP en un gran porcentaje de sitios no significa que ese porcentaje esté vulnerable, ni que exista explotación confirmada a escala similar. Se requiere identificar versión y distribución efectiva, modo de ejecución, ruta de exposición, controles, actualizaciones y, cuando sea pertinente, registros forenses. [09]

El 23 de octubre, PHP describió además un escenario específico vinculado a `session.upload_progress.enabled` y cambios dinámicos de `session.serialize_handler`, bajo el identificador

CVE-2026-115560. Lo trató como una combinación inusual pendiente de corrección en el ciclo regular, con mitigación de configuración recomendada. No debe confundirse con los dos CVE principales. [09]

Lección arquitectónica. La vulnerabilidad de un runtime puede materializarse antes de que la aplicación ejecute su lógica ordinaria de autenticación. Los controles de identidad de una API no sustituyen la higiene del runtime; aislamiento, mínimos privilegios, restricciones de salida, segmentación y protección perimetral reducen consecuencias, pero no reemplazan la actualización cuando el fabricante señala que no existen mitigaciones completas.

05 Los próximos 18 meses: escenario de presión, no fecha de colapso

Quién es Jason D. Clinton y por qué importa su diagnóstico

Jason D. Clinton es Deputy Chief Information Security Officer de Anthropic y fue su primer responsable máximo de seguridad. Anteriormente dirigió la seguridad de la infraestructura de Google Chrome; su trayectoria incluye ChromeOS, Android Pay y protección frente a amenazas persistentes avanzadas. Su perfil biográfico es público en RSAC Conference. Por ello, su opinión procede de experiencia en infraestructura de software ampliamente distribuido y en el entorno de desarrollo de IA avanzada, sin convertir sus estimaciones en evidencia prospectiva demostrada. [25]

En DEF CON 34, durante agosto de 2026, advirtió sobre un periodo posible de 18 a 24 meses excepcionalmente intenso para el sector. El 22 de octubre acortó su referencia operativa a unos 18 meses y recomendó reforzar los controles básicos. También había defendido el aislamiento de infraestructura crítica cuando el parcheo no pueda seguir la velocidad del descubrimiento. Son declaraciones atribuidas, no resultados de una predicción cuantitativa. [15, 26]

Su razonamiento introduce una tensión importante: los mismos sistemas capaces de encontrar errores podrían abaratar su eliminación en el futuro. La idea de que modelos más accesibles terminarán reduciendo la introducción de defectos es una expectativa de Clinton, todavía no un resultado probado. [26]

La advertencia pública de Clinton del 22 de octubre de 2026 es un llamado a reforzar capacidades fundamentales antes de aumentar herramientas. Su horizonte de unos 18 meses sirve para ensayos de estrés, no para fijar una fecha de crisis o para afirmar que cualquier sector quedará necesariamente comprometido. [15, 26]

Tres mecanismos plausibles de presión

Entrada más rápida: los agentes de investigación amplían cobertura de repositorios, generan hipótesis, verifican casos y reducen el costo de encontrar fallas difíciles.

Ventana ofensiva potencialmente más breve: la difusión de detalles técnicos, diferencias de parches y automatización de pruebas puede agilizar la reproducción de ataques, aunque no hay una reducción universal del tiempo a explotación probada para todas las familias.

Cierre lento por dependencias: los hallazgos aplicables exigen aprobación de propietarios, validación de versiones, pruebas, ventanas de cambio, coordinación con terceros y verificación de que el parche corre realmente en producción.

Modelo ilustrativo de saturación

Definimos D_t como el número de hallazgos aplicables pendientes al término del mes t; λ_t representa entradas nuevas verificadas y μ_t cierres verificados. La identidad operativa básica es $D_t = \max(0, D_{t-1} + \lambda_t - \mu_t)$. No incluye severidad, riesgo ni dependencias, que deben medirse por separado. No es un pronóstico de ZTC ni una cifra tomada del mercado.

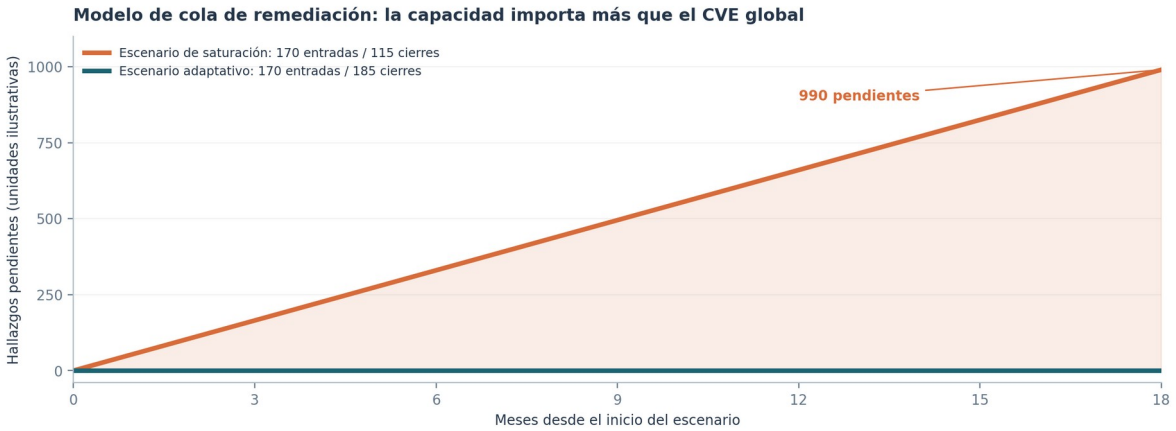


Figura 4. Simulación didáctica de 18 meses con datos hipotéticos, no observados. En saturación, la deuda crece 55 casos al mes y alcanza 990; con capacidad suficiente no acumula pendientes.

La utilidad del modelo es ubicar una decisión que un consejo puede adoptar: aumentar capacidad de cierre, disminuir exposición de activos, automatizar verificación o reducir trabajo que nunca fue relevante. Añadir herramientas que generen alertas sin generar cierres podría agravar el volumen pendiente sin reducir riesgo.

Indicadores de alerta temprana

Métrica	Definición operacional	Señal relevante
Cobertura de inventario	% de activos expuestos con propietario, versión, dependencia y criticidad	Crecimiento de activos sin propietario o sin versión.
Tiempo a clasificación	Horas desde alerta hasta decisión de aplicabilidad y prioridad	Se prolonga aun para servicios expuestos.
Tiempo a protección	Horas hasta control temporal medido o parche funcional	Supera la ventana de actividad adversaria observada.
Tiempo a cierre verificado	Horas/días hasta comprobar versión y ausencia de exposición	Cierre administrativo sin prueba técnica.
Backlog relevante	Casos aplicables abiertos en activos críticos, ponderados por edad	Entradas > cierres durante varios ciclos.
Cobertura de terceros	Servicios y librerías que dependen de proveedor con SLA probado	Sin evidencia de actualización ni aislamiento alterno.

La estadística no puede construirse con un caso extremo

Una demostración de compromiso en minutos puede describir una capacidad real de atacante bajo condiciones específicas, pero no establece el tiempo medio de explotación de Internet, de acceso a privilegios administrativos ni de afectación a todas las empresas. Es indispensable conocer el entorno, el punto de partida, el universo, el reloj utilizado y el criterio de éxito. [29–30]

Mandiant estimó para su población de explotación investigada un tiempo promedio de -7 días entre la explotación inicial y la divulgación relevante, indicador de que parte de los casos fue explotada antes de hacerse pública. Tampoco equivale al tiempo medio de compromiso de cualquier organización. La convergencia de métricas distintas en una sola curva de aceleración produciría una precisión engañosa. [30]

06 Modelo de defensa orientado al tiempo de exposición

Secuencia de decisión por vulnerabilidad

El criterio de respuesta debe combinar aplicabilidad técnica + exposición alcanzable + evidencia de explotación + impacto empresarial + restricciones de cambio. CVSS informa severidad técnica, EPSS estima probabilidad de explotación observada en los siguientes 30 días bajo su modelo y KEV identifica explotación confirmada bajo los criterios de CISA. Ninguno sustituye por sí solo el contexto propio. [11–12]

Etapa	Evidencia mínima para darla por terminada
1. Descubrir	Inventario y versión comprobada; SBOM o evidencia equivalente si aplica.
2. Determinar alcance	Rutas de entrada, exposición pública, privilegios, dependencia y propietario del activo.
3. Decidir respuesta	Registro de KEV/EPSS/CVSS, impacto, excepciones, medida y responsable.
4. Contener	Aislamiento/limitación de acceso/salida o regla temporal con pruebas de efectividad y efecto colateral.
5. Corregir	Nueva imagen/paquete publicado, despliegue verificable, regresión funcional y rollback preparado.
6. Comprobar	Lectura de versión efectiva, test negativo reproducible, telemetría y acta de cierre.

Controles que reducen la asimetría

Inventario y mapeo de dependencias: activos, contenedores, runtimes, API, componentes de terceros, cuentas de servicio, versión ejecutada y propietario; incorporar SBOM y VEX cuando existan datos confiables.

Aislamiento por arquitectura: mínimos privilegios, separación entre planos de administración y datos, microsegmentación, redes privadas y rutas de salida restringidas; validar que la exposición no se recomponga por una integración secundaria.

Gestión de cambios preparada: entornos representativos, canary, automatización de despliegue, rollback y pruebas funcionales; la protección debe medirse en el artefacto realmente ejecutado.

Telemetría de explotación: registro de actividad útil para identificar intentos y confirmar compromisos; activar análisis forense cuando haya señales de intrusión previa al parche.

Inteligencia y priorización: KEV, EPSS, superficie accesible, criticidad de procesos, datos sensibles, capacidad de movimiento lateral y dependencia operativa.

Coordinación con terceros: plazos contractuales para respuesta, comunicación de dependencias, mantenimiento de sistemas fuera de soporte y evidencia de actualización.

Priorización en cuatro decisiones

Decisión	Disparadores indicativos	Respuesta
EMERGENCIA	Explotación confirmada y ruta accesible hacia activo crítico; o RCE remota crítica con señal confiable de actividad.	Contener de inmediato, investigar exposición histórica, parchear y verificar.
ACELERADA	Alta probabilidad y exposición real, dependencia esencial o impacto alto.	Despliegue preferente con tiempo objetivo aprobado y pruebas.
PROGRAMADA	Aplicable con restricciones y controles compensatorios.	Cambio planificado con vencimiento, aceptación explícita y seguimiento.

Decisión	Disparadores indicativos	Respuesta
NO APLICABLE	Producto ausente o condición de ataque descartada con evidencia verificable.	Documentar descartes; reabrir si cambia arquitectura o inteligencia.

Estos estados son un modelo propuesto de gestión, no equivalen a una categoría normativa de PCI DSS, ISO 27001 o CISA. Las obligaciones regulatorias, contratos y normas aplicables deben comprobarse separadamente según país, sector y fecha.

Arquitectura resiliente: desacoplar el crecimiento de CVE del daño

El número de CVE pertenece al ecosistema; la exposición efectiva pertenece a la arquitectura de cada organización. NIST SP 800-207 desplaza el foco desde confiar en la ubicación de red hacia proteger recursos con autenticación y autorización explícitas; NIST SP 800-160 v2 define resiliencia como la capacidad de anticipar, resistir, recuperarse y adaptarse aun bajo compromiso. [23–24]

Hipótesis de arquitectura (H4). Si dos sistemas comparten la misma versión vulnerable, el que mantiene rutas públicas limitadas, identidades independientes, privilegios mínimos, segmentación por función, egreso restringido y datos segregados debería tener menos caminos viables hacia activos críticos y menor impacto potencial. La afirmación es condicional: debe validarse en pruebas negativas y telemetría; ninguna barrera es perfecta. [23–24]

Una cadena de intrusión tiene varios puntos de ruptura

Una vulnerabilidad puede representar un defecto en un paquete. Para convertirlo en pérdida se requiere, normalmente, una cadena: alcanzar la función, activar el error, obtener una capacidad útil, moverse o acceder a secretos y completar una acción dañina. Controles de diseño independientes deben limitar eslabones distintos. No se debe asumir que un WAF, un proxy de autenticación o un entorno privado bloquean una explotación que pudiera originarse desde un proceso permitido. [23, 24, 27]

Controles falsables, no promesas de invulnerabilidad

Capa	Decisión arquitectónica	Evidencia verificable
Exposición	Inventario de servicios y dependencias; cierre de rutas de administración pública.	Reconocimiento externo autorizado: sin ruta directa a consola o runtime interno.
Identidad	Identidad independiente por carga, credenciales efímeras, autorización por recurso.	Comprometer una cuenta de servicio no permite acceder a otras funciones o privilegios.
Aislamiento	Segmentación este-oeste, políticas denegación por defecto, aislamiento entre clientes.	Pruebas negativas entre servicios, redes y tenants; registrar rechazos y bypass.
Ejecución	Procesos sin privilegios, sandbox, sistemas de archivos de solo lectura y endurecimiento.	Un RCE de ensayo queda restringido al contenedor, sin acceso al host o secretos.
Datos	Minimización, tokenización y custodia externa cuando proceda; control de uso de claves.	El servicio periférico no puede leer datos originales ni ampliar su autorización.
Egreso	Salida explícita por destino, protocolo y finalidad; monitoreo de anomalías.	Intento de exfiltración hacia un destino no autorizado es rechazado.
Resiliencia	Despliegue inmutable, canary/blue-green, reversión y restauración probada.	Simulacro permite contener, actualizar y restaurar sin comprometer la integridad.

Escenario de prueba: RCE en un servicio público

Arquitectura plana: una RCE en un componente web que comparte credenciales amplias y red con las bases de datos puede convertirse en acceso lateral y exposición de varios clientes. Arquitectura segmentada: la misma RCE puede comprometer el proceso inicial, pero identidades acotadas, aislamiento y custodia separada limitan lo que puede alcanzar. Esta comparación es conceptual; no representa una prueba ejecutada sobre una organización concreta.

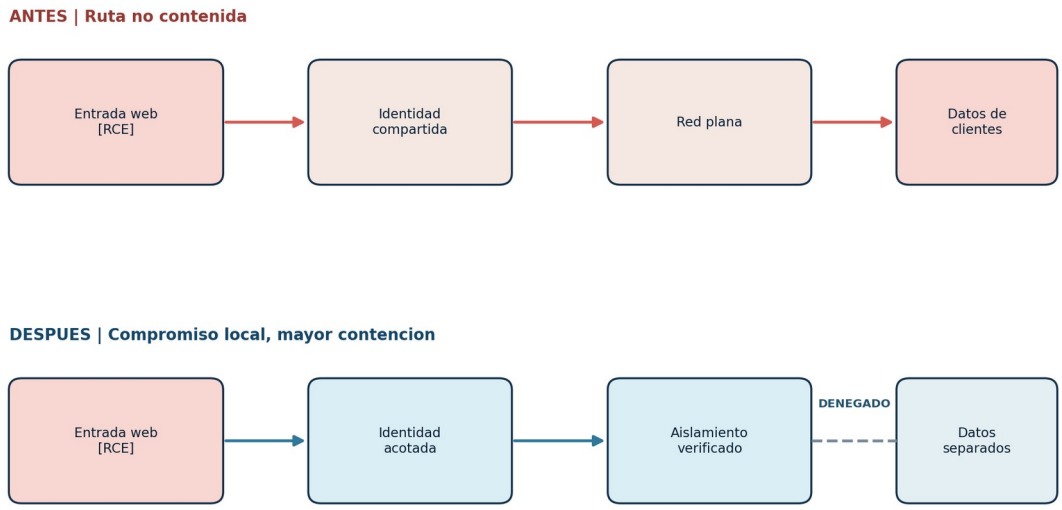


Figura 5. Dos topologías conceptuales bajo el mismo defecto de software. La arquitectura no elimina la RCE; busca reducir sus efectos observables.

Cómo medir la eficacia sin inventar reducciones porcentuales

Proponemos registrar por cada activo: (a) versión vulnerable confirmada; (b) alcanzabilidad efectiva; (c) privilegios utilizables tras compromiso; (d) destinos laterales accesibles; (e) datos consultables; (f) rutas de egreso; (g) tiempo hasta protección verificable. La comparación antes/después debe ejecutarse con escenarios repetibles. No es correcto publicar, por ejemplo, que la segmentación reduce el riesgo 80% sin una base empírica de explotabilidad, accesos e impacto.

Indicadores propuestos: cobertura de pruebas negativas en límites críticos; porcentaje de activos públicos sin acceso administrativo directo; rutas laterales autorizadas por carga; servicios con identidad compartida; horas-activo de exposición efectiva ponderadas por criticidad; y tiempo p90 hasta contención y cierre. Un CVE no aplicable se documenta, no se borra del inventario.

Limitaciones. El compromiso de un plano de identidad, la exfiltración por canales permitidos, las configuraciones erróneas y las relaciones de confianza de terceros pueden anular defensas aparentemente sólidas. La criptografía y la tokenización protegen datos dentro de su alcance; no corrigen fallas de memoria ni sustituyen permisos, segregación o parcheo. [23–24, 27]

Cuatro límites de resiliencia medibles

La especialidad arquitectónica se expresa en cuatro límites, cada uno sujeto a una prueba de fallo seguro. El objetivo no es proclamar que el sistema es invulnerable; es demostrar qué recurso continúa inaccesible cuando el componente anterior ya se considera comprometido. [23, 24, 31]

Límite	Diseño	Prueba negativa
Entrada	Acceso público mínimo y aislamiento administrativo	RCE en un servicio periférico no abre consola ni planos internos
Identidad	Privilegio por carga, caducidad, autorización por recurso	Token comprometido no permite elevar permisos ni cruzar clientes

Límite	Diseño	Prueba negativa
Datos	Custodia separada, minimización y control de uso	Servicio web comprometido no recupera originales fuera de su necesidad
Salida y recuperación	Egreso explícito, reversión, restauración con evidencia	Exfiltración no autorizada bloqueada; cambio reversible y verificado

Las pruebas deben incluir credenciales de servicio válidas pero insuficientes, rutas laterales y accesos entre clientes denegados, intentos de egreso no autorizado y fallos de despliegue. Un WAF, una segmentación dibujada o la simple existencia de un token no demuestran por sí solos que esos límites funcionen. [23–24, 31]

07 Implicaciones para servicios financieros y proveedores críticos

En servicios de pago, banca y plataformas con información personal, una vulnerabilidad no es relevante solamente por permitir entrada a un servidor. La pregunta es si rompe límites de confianza, accede a información sensible, compromete identidades de servicio o alcanza a terceros con permisos permanentes. Por eso una revisión técnicamente buena debe reconstruir rutas completas, no limitarse a preguntar si el CVE figura en el inventario.

- Plano público: API gateways, portales, reverse proxies, runtimes y librerías expuestas; separar explotación de aplicación de compromiso del entorno de ejecución.
- Plano de identidad: credenciales efímeras, autorización por transacción, MFA resistente al phishing donde aplique, permisos de servicio y segregación de tareas.
- Plano de datos: limitar exposición de PAN/PII y datos sensibles, reducir copias, usar tokenización cuando sea apropiado y probar restricciones en lectura y salida.
- Plano de proveedores: registrar bibliotecas críticas, responsabilidades de actualización, acuerdos de divulgación y límites de confianza en integraciones.
- Plano de resiliencia: pruebas negativas de aislamiento, rollback operativo, restauración comprobada, retest y análisis de registros de explotación antes/después del cierre.

Observación sobre cumplimiento. Acreditar cumplimiento de un estándar no demuestra inmunidad frente a CVE futuros; del mismo modo, una lista de escáner no sustituye la comprobación de alcance, controles y evidencia. La continuidad del negocio exige conectar gestión de vulnerabilidades con arquitectura y operación, no tratarlas como disciplinas aisladas.

Preguntas de gobierno que deberían tener respuesta documentada

- ¿Cuántos servicios expuestos carecen hoy de propietario, versión de runtime y dependencias identificadas?
- ¿Cuáles de los CVE nuevos afectan activos nuestros y cuáles disponen de camino de ataque y privilegios suficientes para producir impacto?
- ¿Cuál es la mediana y percentil 90 entre la alerta, el aislamiento, el despliegue efectivo y el cierre técnicamente verificado?
- ¿Qué ocurre si un proveedor esencial tiene una vulnerabilidad crítica y no entrega parche durante días?
- ¿Qué pruebas demuestran que los límites entre clientes, servicios y redes sobreviven al compromiso de un componente periférico?

08 Plazos actuales de vulnerabilidades, corrección y mitigación

La pregunta normativa no es si todos los marcos autorizan esperar treinta días; no lo hacen. Algunos definen máximos por criticidad, otros frecuencias de proceso o condiciones por activo. Confundir una frecuencia de

escaneo con un plazo de corrección, o una recomendación con una obligación, distorsiona el análisis. [11, 19–22]

Comparación de requisitos y orientaciones vigentes

Marco	Plazo o frecuencia documentada	Lectura correcta
PCI DSS 4.0.1	Req. 6.3.3: parches/actualizaciones para vulnerabilidades críticas dentro de un mes de publicación. Los demás, según riesgo definido por la entidad.	Obligación en sistemas dentro del alcance PCI; no significa esperar todo el mes. [19]
CISA BOD 26-04	Máximos condicionales de 3, 14 o 60 días naturales, o en siguiente actualización mayor; nivel de 3 días puede exigir triage forense. Cuatro factores por activo.	Mandato a agencias federales cubiertas. Fase de aplicación plena desde 7 dic 2026. [11, 20]
CIS Controls v8.1	7.2 revisión de estrategia de remediación mensual o más frecuente; 7.3/7.4 parcheo automatizado mensual o más frecuente; 7.7 remediación mensual o más frecuente.	Frecuencias recomendadas; no son plazo universal concedido a cada CVE. [21]
DORA / RTS 2024/1774	Art. 10: escaneo al menos semanal de activos que soportan funciones críticas/importantes; exige procedimientos de emergencia y plazos de instalación y escalamiento definidos.	No fija en ese artículo un SLA universal de 3, 7 o 30 días para todas las fallas. [22]
NIST SP 800-40 Rev. 4	Parches como mantenimiento preventivo: identificar, priorizar, adquirir, instalar y verificar; diseño de estrategia empresarial de riesgo.	Guía técnica, sin fecha universal por CVE. [28]

Precisión de cómputo. El plazo PCI para vulnerabilidades críticas se cuenta desde la publicación del parche/actualización aplicable. Para BOD 26-04, el reloj se activa por la primera de dos condiciones normativas: inclusión en KEV o identificación del CVE en un activo y su reporte de acuerdo con el sistema federal; los plazos pueden acortarse cuando cambian exposición o inteligencia. [19, 20]

Qué deberían modificar los marcos en la era de descubrimiento asistido por IA

Propuesta de investigación, no regla en vigor: complementar los plazos categóricos por severidad con objetivos dependientes de explotación real, automatabilidad, exposición pública, privilegios y alcance hacia datos o servicios esenciales. La verificación de alcanzabilidad debe ser obligatoria y quedar auditable.

Segundo: establecer dos relojes distintos. El primero mide tiempo desde noticia creíble hasta protección efectiva; el segundo, hasta parche o eliminación verificados. El cumplimiento de una contención temporal no debe cerrar el expediente mientras subsista la falla.

Tercero: exigir pruebas periódicas de segmentación, aislamiento entre clientes, integridad de identidades y restricciones de egreso, en vez de asumir que su mera configuración constituye protección. Cuarto: requerir rutas de cambio de emergencia, reversión y obligaciones verificables de proveedores críticos. [23–24, 27]

La evolución normativa podría exigir que los tiempos cambien cuando cambien los hechos: exposición de un activo, confirmación de explotación, automatización de un exploit o efecto sobre un plano de identidad. CISA BOD 26-04 ya es un ejemplo verificable de priorización contextual; la iniciativa aquí propuesta va más allá al pedir objetivos explícitos de contención en horas, pruebas arquitectónicas negativas y reporte de riesgo residual. Esto es una propuesta de ZTC Research, no una obligación actual. [20, 32]

09 CCA-01: contención en horas y cierre verificable en días

Los objetivos siguientes son un perfil voluntario y más exigente para activos de negocio relevantes. No sustituyen leyes, contratos ni requisitos PCI/CISA aplicables. Cuando rija un plazo legal más estricto, prevalece. Un control compensatorio aceptado debe ser probado, reversible y tener vencimiento; no equivale automáticamente a remediación definitiva.

Umbrales propuestos por riesgo contextual

Nivel	Disparador contextual	Objetivo interno propuesto
A / Emergencia	Explotación activa aplicable; KEV alcanzable; RCE remota automatizable con alto impacto y ruta externa.	Contención <= 4 h; evaluación forense inmediata; cierre objetivo <= 72 h cuando exista corrección segura.
B / Alta exposición	Ruta externa demostrada y explotación plausible / prueba funcional / servicio crítico.	Protección <= 24 h; cierre objetivo <= 7 días.
C / Expuesta sin explotación	Alta/critica aplicable, acceso público, defensas verificadas y sin actividad confirmada.	Protección <= 72 h; corrección objetivo <= 14 días.
D / Interna acotada	No expuesta ni alcanzable desde zonas no confiables, con verificación y alcance limitado.	Revisar en <= 7 días; corregir objetivo <= 30 días.
E / Baja relevancia	No crítica, no explotada, inventario y ruta documentados, sin señales de cambio.	Ventana según análisis de riesgo; revisión continua, no cierre por silencio.

En el nivel A, si no existe parche o la aplicación inmediata introduce un riesgo superior para la continuidad, la primera decisión puede consistir en retirar exposición, restringir uso o aislar temporalmente el sistema con evidencia. La excepción debe elevarse de inmediato, incorporar medidas de continuidad y conservar investigación forense. No se debe reportar la vulnerabilidad como subsanada por el mero hecho de haberla aislado. [20, 23–24]

Métricas que pueden auditarse

MTP (tiempo a protección): horas entre detección creíble y verificación de un control que interrumpe el camino de ataque observado. MTC (tiempo a cierre): horas/días hasta demostrar eliminación del defecto en el activo afectado. EAEC (exposición de activos esenciales): suma de horas de activos vulnerables con una ruta de ataque relevante, ponderada por criticidad de negocio bajo una escala interna documentada. Cobertura de barreras: proporción de pruebas negativas aprobadas en límites de identidad, red, datos, tenants y egreso.

Los indicadores operativos se reportan por niveles y percentiles (mediana, p90), con muestras y exclusiones. Los conteos de CVE no pueden sustituir al inventario de instancias afectadas: un único CVE en cientos de activos puede crear más trabajo y exposición que cientos de CVE que no aplican.

Límites de costo y continuidad

La reducción del tiempo de exposición debe financiarse junto con automatización de pruebas y despliegues, capacidad de respuesta, redundancia y reversión. Una política de parcheo agresiva sin pruebas puede aumentar indisponibilidad y pérdida operacional. La evaluación económica debe comparar horas-activo de exposición, costo de controles, frecuencia de intervenciones y resultados de continuidad, sin atribuir ahorros hipotéticos como ganancias confirmadas.

Prioridad de inversión y rendición de cuentas

La secuencia de inversión recomendada inicia en límites de confianza que reduzcan daños posibles, continúa con actualización y retest automatizados y termina en observabilidad y respuesta sostenibles. Cada iniciativa requiere propietario, costo total, tiempo de implantación, riesgo residual y prueba de eficacia. Donde no haya datos para estimar pérdidas, se informará la incertidumbre: evitar un incidente no constituye por sí mismo un ahorro financiero realizado.

La continuidad operativa se valida mediante ensayos de aislamiento de servicios, reversión, restauración y verificación de integridad antes de imponer plazos agresivos de despliegue. La ejecución no debe transferir a la operación un riesgo de caída mayor que la exposición que se pretende reducir.

Tres relojes y un criterio de cierre

Reloj 1 — Reconocimiento: comienza con la noticia creíble o detección verificable y termina con una decisión registrada de aplicabilidad. Reloj 2 — Protección: concluye solo cuando un control interrumpe y demuestra la ruta de ataque relevante. Reloj 3 — Cierre: termina cuando la corrección definitiva ha sido desplegada en las instancias afectadas, comprobada y sometida a retest funcional y de seguridad.

Para vulnerabilidades explotadas antes de su publicación, el reloj de detección organizacional no captura la totalidad del tiempo de exposición histórica. Debe complementarse con revisión retrospectiva de eventos, investigación de intrusiones y límites de retención. Un cierre administrativo sin prueba de protección real o sin revisión de compromiso previo puede dar una falsa sensación de seguridad. [29–30]

10 Arquitectura de contención CCA-01: doble frontera y decisión verificable

La propuesta parte de una premisa limitada pero operacionalmente decisiva: una vulnerabilidad no determina por sí sola el daño, porque el resultado depende también de las rutas de confianza, identidades, permisos, flujos de datos y funciones esenciales que conectan el punto de compromiso con los objetivos críticos. El criterio CCA-01 no es un algoritmo criptográfico, una técnica de segmentación nueva ni una garantía de seguridad; es un contrato de evidencia arquitectónica diseñado para evitar conclusiones favorables basadas en controles aislados.

Antecedentes y novedad delimitada

Los grafos de ataque, la alcanzabilidad, los principios Zero Trust y la ingeniería de sistemas ciberresilientes son antecedentes consolidados. CCA-01 no reclama su invención. Su novedad acotada es convertir la alcanzabilidad adversaria, la continuidad de funciones legítimas y la suficiencia de dependencias en una misma adjudicación, con reglas explícitas de aceptación, rechazo y abstención. La diferencia práctica es verificable en el modelo: aislar puede reducir el daño potencial y, simultáneamente, inutilizar una función crítica. El protocolo obliga a identificar esa contradicción antes de concluir y conecta el resultado con la gestión de exposición previa a la remediación. [23, 24, 33–35]

Frontera de daño: el grafo de capacidad adversaria

Representar el escenario con un grafo dirigido G_a de transiciones plausibles de capacidad adversaria. Se declara un punto inicial s, un conjunto K de objetivos críticos, reglas de privilegio y condiciones de paso. Una arista no describe simplemente una llamada normal entre servicios: debe representar una transición factible bajo el adversario supuesto. Cualquier camino desde s hasta un $k \in K$ afecta el resultado de contención del escenario; una ruta de egreso hacia un destino no autorizado se registra independientemente.

Frontera de continuidad: el grafo de función legítima

Modelar en un grafo separado G_m las dependencias necesarias para completar una misión legítima explícita. Un aislamiento que suprime todas las rutas adversarias pero interrumpe liquidación, atención, pagos o recuperaciones esenciales no constituye éxito conjunto. Los objetivos críticos de contención y la misión que debe subsistir deben fijarse antes de observar el resultado, para evitar seleccionar ex post solo aquellos objetivos favorables.

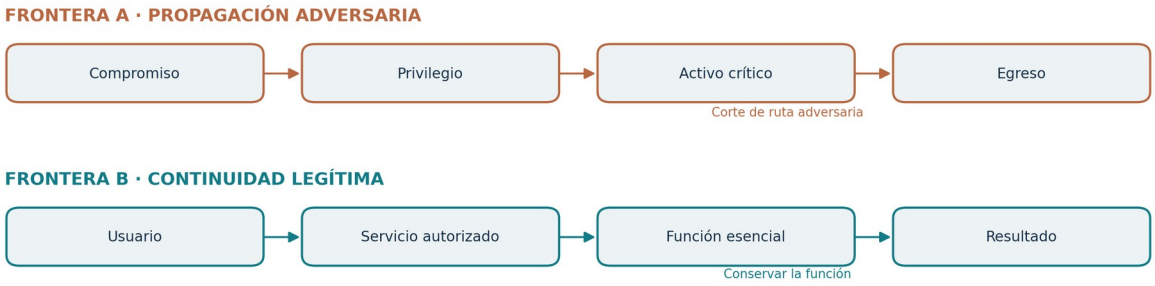


Figura 6. Dos modelos vinculados por una única decisión. Las rutas son conceptuales y no representan sistemas existentes.

Vector transparente de daño y continuidad

Para los escenarios no ponderados de escritorio se propone PDC = (Rc, E, M, C). Rc representa la fracción de objetivos críticos alcanzables; E indica la presencia de una ruta de egreso no controlado; M la posibilidad modelada de completar una misión legítima; C la cobertura de las dependencias relevantes. Rc no representa probabilidad de incidente, severidad CVSS ni pérdida económica. El índice B del ensayo E-01 es ponderado y distinto de Rc: sus cifras no son intercambiables.

Variable	Definición	Interpretación admisible
Rc	$\lvert Reach(s,G_a) \cap K \rvert / \lvert K \rvert$	Proporción de objetivos críticos alcanzables, según el escenario supuesto
E	0: ruta de egreso no alcanzable; 1: alcanzable	No equivale a exfiltración observada
M	1: la misión legítima conserva su ruta; 0: se interrumpe	No constituye medición de disponibilidad real
C	1: universo de dependencias completo para el escenario; 0: incompleto	Una completitud declarada no equivale a inventario auditado

Regla de decisión y abstención

En el ensayo de escritorio se utiliza un umbral estricto de demostración: PASS solo cuando Rc = 0, E = 0, M = 1 y C = 1; FAIL cuando existe un incumplimiento conocido con alcance suficiente; NE cuando faltan relaciones necesarias para adjudicar. Los umbrales son convenciones locales del modelo sintético, no parámetros regulatorios ni requisitos universales de producción. Una evaluación real necesitaría umbrales basados en riesgo, calidad de datos y tolerancia de misión.

PRINCIPIO DE NO REDUNDANCIA

Antes de recomendar una nueva evaluación, identificar qué prueba o evidencia ya existe en PCI DSS, ISO, NIST, SOC 2 u otros programas. Reutilizar su resultado cuando sea equivalente. CCA-01 solo aporta valor incremental si descubre una ruta, dependencia o conflicto de continuidad no cubierto y cambia una decisión concreta.

Relojes de exposición, protección y recuperación

Para una aplicación operativa futura, la medición requiere fijar cuatro instantes distintos: T0 (alerta confiable y aplicable), TD (detección y declaración del compromiso si difieren), TP (protección efectiva sostenida tras una medida) y TR (remediación definitiva comprobada). El tiempo a protección es MTP = TP − T0; el tiempo a cierre o remediación es MTR = TR − T0. Una modificación de política registrada no demuestra protección sostenida por sí sola; una vulnerabilidad que no se corrige dentro del horizonte observado no se contabiliza como éxito por desaparición de la muestra.

Cuando existan datos autorizados, los informes deberían declarar horizonte y población, mediana y percentiles de MTP/MTR, eventos censurados, fallos, incidentes de continuidad, regresiones y fallas de frontera. En la investigación presente estas métricas son definiciones de un protocolo de evaluación: los únicos minutos comunicados pertenecen al ensayo ficticio E-01A. No se ha medido tiempo de contención en servicios reales ni se propone sustituir objetivos regulatorios de recuperación. [24, 28, 34, 41]

Condiciones de refutación de la propuesta

CCA-01 debe considerarse limitado o refutado en una aplicación específica si: una ruta residual relevante no aparece en el modelo; una decisión de PASS se obtiene con inventario incompleto; una política de aislamiento interrumpe la función esencial definida; o la evidencia incremental no mejora ninguna decisión sobre la que ya se pronunciaron pruebas existentes. El método tiene que tratar esos resultados como evidencia desfavorable, nunca subsumirlos en una media global favorable.

11 Demostración sintética: dos instrumentos, una tesis arquitectónica

Se documentan dos series sintéticas distintas, elaboradas para evaluar la consistencia y el poder de rechazo de las reglas. Ninguna representa clientes, sistemas de pago, redes de terceros o plataformas ZTC. No deben agregarse para estimar un porcentaje de eficacia. No hubo pentesting, pruebas de campo ni instrumentación de entornos reales.

Instrumento E-01A: doce pares ficticios de índice B

En E-01A, ΔB = B(referencia) − B(delimitada). El índice de referencia y el posterior a un control supuesto fueron diseñados deliberadamente para ensayar cálculos y estados. La media de ΔB es +0.3083 puntos de índice; ese valor NO representa 30.83 % menos riesgo. La aprobación de un escenario depende, además, de continuidad, egreso, fronteras y suficiencia del registro.

Caso	Escenario ficticio	B ref.	B delim.	ΔB	Decisión
S01	Sesión periférica	0.90	0.40	+0.50	PASS
S02	Solicitud entre ámbitos	0.60	0.00	+0.60	PASS
S03	Permiso amplio	1.00	0.30	+0.70	PASS
S04	Dependencia encadenada	0.50	0.20	+0.30	PASS
S05	Salida no autorizada	0.40	0.00	+0.40	PASS
S06	Identidad elevada	1.00	1.00	+0.00	FAIL
S07	Registro incompleto	0.10	0.10	+0.00	NE

Caso	Escenario ficticio	B ref.	B delim.	ΔB	Decisión
S08	Continuidad degradada	0.30	0.10	+0.20	FAIL
S09	Permiso caducado	0.50	0.10	+0.40	PASS
S10	Propagación lateral	0.80	0.30	+0.50	PASS
S11	Cambio concurrente	0.50	0.90	-0.40	FAIL
S12	Dependencia externa	0.70	0.20	+0.50	PASS

La regla rechaza S06 por persistencia de rutas; S08 porque el descenso de B coincide con ruptura de continuidad; S11 por regresión de alcanzabilidad; y mantiene S07 en NE por registro incompleto. En los ocho PASS sintéticos el MTP programado tuvo mediana de 40 minutos (rango 16–192); son tiempos declarados del ejercicio, no tiempos medidos en productos o servicios.

Instrumento E-01B: dieciséis modelos de doble frontera

E-01B modela tres objetivos críticos por caso y condiciones separadas de egreso, continuidad y suficiencia del inventario. La muestra contiene arquitectura plana, controles parciales, configuración combinada y excepciones adversas. Cuatro casos cumplen, once fallan y uno queda no estimable. Los resultados comprueban la adjudicación del evaluador ante entradas construidas; no establecen que una arquitectura comercial alcance esas tasas.

Caso	Condición simulada	Críticos alcanzables	E	M	C	Estado
S01	Confianza plana	3/3	1	1	1	FAIL
S02	Solo corte de salida	3/3	0	1	1	FAIL
S03	Solo separacion administrativa	2/3	1	1	1	FAIL
S04	Solo corte este-oeste	1/3	1	1	1	FAIL
S05	Contencion combinada	0/3	0	1	1	PASS
S06	Contencion con archivo auxiliar	0/3	0	1	1	PASS
S07	Credencial de servicio reutilizada	2/3	0	1	1	FAIL
S08	Acceso de emergencia persistente	1/3	0	1	1	FAIL
S09	Acceso cruzado a datos	1/3	0	1	1	FAIL
S10	Puente de tercero	2/3	0	1	1	FAIL
S11	Contencion que interrumpe servicio	0/3	0	0	1	FAIL
S12	Inventario de dependencias	0/3	0	1	0	NE

Caso	Condición simulada	Críticos alcanzables	E	M	C	Estado
	incompleto					
S13	Acceso directo a custodia	1/3	0	1	1	FAIL
S14	Aislamiento con auditoria operativa	0/3	0	1	1	PASS
S15	Control de continuidad redundante	0/3	0	1	1	PASS
S16	Excepcion de facturacion	1/3	0	1	1	FAIL

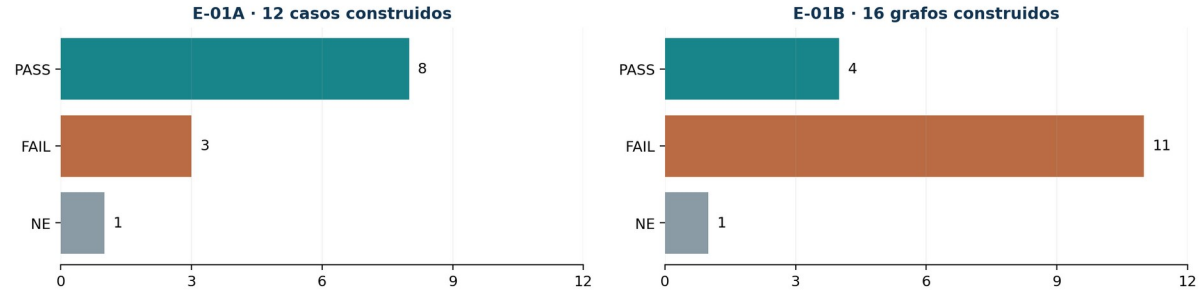


Figura 7. Decisiones sintéticas por instrumento; los denominadores y las reglas son diferentes. No es una estadística de eficacia operativa.

Ensayos adversos y conclusiones admisibles

Los fallos seleccionados incluyen: S07 (credencial reutilizada, dos objetivos alcanzables), S08 (acceso de emergencia persistente), S10 (puente de proveedor), S11 (contención que interrumpe una misión) y S12 (dependencias desconocidas: NE). El evaluador no valida si las hipótesis describen fielmente una red real. Su alcance es la integridad lógica del tratamiento de excepciones, límites y ausencia de evidencia.

La evidencia reproducible permite afirmar que el algoritmo clasifica estos casos construidos bajo reglas conocidas. No permite inferir menor incidencia de ataques, reducción de pérdidas monetarias, cumplimiento PCI DSS, conformidad ISO/IEC 27001, ni eficacia sobre activos de una organización. Tampoco demuestra revisión independiente por pares.

12 Correspondencia crítica con marcos internacionales vigentes

El principio de contención arquitectónica debe dialogar con controles ya consolidados, no caricaturizarlos. Las recomendaciones que siguen son una propuesta técnica de ZTC Research, no interpretación oficial ni modificación adoptada por los emisores. Las equivalencias son funcionales, condicionadas al alcance de cada auditoría, región y versión del marco.

Organismo y marco	Cobertura existente relevante	Aporte evaluativo sugerido por CCA-01
NIST · CSF 2.0; SP 800-207; SP 800-160; SP 800-61r3	Gestión de riesgo, arquitectura Zero Trust, sistemas ciberresilientes, respuesta y recuperación.	Ficha de evidencia sobre rutas críticas residuales, incertidumbre del modelo y continuidad tras compromiso supuesto. [23,24,33–35]
ISO/IEC · 27001:2022; 27002:2022; 27005:2022	Sistema de gestión de seguridad, tratamiento del riesgo y orientación sobre controles.	Relacionar límites de daño, dependencias, excepciones y salvaguardas con activos, propietarios y riesgos aceptados. [36–38]
PCI SSC · PCI DSS v4.0.1	Controles de red, corrección de fallas,	Complementar, no sustituir, evidencia de

Organismo y marco	Cobertura existente relevante	Aporte evaluativo sugerido por CCA-01
	respuesta a incidentes y pruebas de segmentación cuando apliquen.	aislamiento del CDE con rutas transversales y continuidad fuera del alcance ya probado. [19,39–40]
CPMI-IOSCO / BIS · Guía 2016 y toolkit consultivo 2026	Resiliencia de infraestructuras de mercado, dependencia de terceros y recuperación crítica.	Documentar límites de propagación entre dependencias y funciones de liquidación sin equiparar MTP a objetivos regulatorios RTO. [41–43]
G7 Cyber Expert Group · declaración IA 2025	Discusión y cooperación frente a amenazas de IA en servicios financieros; no impone nuevas obligaciones.	Evaluar el lenguaje de contención como referencia para discusión transfronteriza y colaboración público-privada. [44]
ISO/IEC · 42001:2023	Gobierno y gestión de sistemas de IA en proveedores y usuarios.	Considerar exposiciones y flujos autorizados cuando componentes de IA participen en servicios críticos. [45]
Unión Europea · DORA y RTS aplicables	Resiliencia operacional digital y requisitos de gestión de riesgos TIC dentro del ámbito regulado.	Usar inventarios y análisis de dependencias para evidenciar limitación de consecuencias sin alterar deberes legales. [22]
AICPA / SOC 2; CIS Controls v8.1	Criterios de servicios de confianza e higiene de gestión de vulnerabilidades según alcance.	Reutilizar evidencia existente cuando cubra el mismo resultado; evitar duplicidad en encargos y auditorías. [21]

PCI DSS: complementariedad, no sustitución de pruebas obligatorias

PCI DSS v4.0.1 exige actividades concretas y, cuando se utiliza segmentación para aislar el entorno de datos de tarjetas (CDE), pruebas de penetración de sus controles de segmentación conforme a los requisitos aplicables. Una revisión de escritorio de CCA-01 NO satisface por sí misma ese deber, ni reemplaza un ejercicio autorizado ni la valoración de un QSA. Su aportación sería una matriz adicional de consecuencia residual y continuidad en aquellos dominios y relaciones que no queden ya cubiertos por evidencia equivalente. [39–40]

La guía del PCI SSC de junio de 2026 distingue controles compensatorios y enfoque personalizado, preserva responsabilidades de entidad y asesor y pone énfasis en documentar correctamente eficacia. CCA-01 puede informar el diseño documental o la evaluación de objetivos cuando sea pertinente, pero no constituye un mecanismo de cumplimiento, certificación o sustitución de requisitos. [40]

ISO/IEC: integrarse al SGSI sin imponer un control inexistente

ISO/IEC 27001:2022 define requisitos de gestión de seguridad; ISO/IEC 27002:2022 orienta sobre controles e ISO/IEC 27005:2022 sobre gestión del riesgo. CCA-01 propone una forma de describir consecuencias y dependencias ya contempladas por una evaluación de riesgo bien diseñada, añadiendo una regla explícita de abstención ante inventarios incompletos y una comprobación cruzada entre reducción de alcance adversario y continuidad. No se atribuye numeración oficial de control ni conformidad ISO.

BIS/CPMI-IOSCO: distinguir contención de recuperación

La guía de resiliencia cibernética de 2016 para infraestructuras del mercado financiero contempla el objetivo de permitir la reanudación segura de operaciones críticas en dos horas y completar la liquidación al cierre de la jornada, sujeto a sus condiciones y contexto. Ese objetivo de recuperación no es un plazo universal de aplicación de parches ni equivale al tiempo de protección MTP de CCA-01. La consulta CPMI-IOSCO del 8 de septiembre de 2026 ofrece una vía contemporánea para discutir herramientas complementarias de evidencia y dependencias de terceros; sus instrumentos se presentan como voluntarios y no vinculantes. [41–43]

13 Control arquitectónico propuesto: texto y evidencias de referencia

La siguiente redacción es una propuesta de ZTC Research susceptible de evaluación por mesas técnicas, operadores y emisores de marcos. No constituye redacción oficial de NIST, PCI SSC, ISO/IEC ni CPMI-IOSCO; su eventual incorporación debe respetar las competencias y procedimientos de cada organización emisora.

CCA-01 · PROPUESTA DE CONTROL

La organización identificará, para cada función esencial y exposición crítica aplicable, los componentes, relaciones de confianza, permisos transitivos, rutas de entrada y egreso, dependencias y excepciones relevantes. Ante un compromiso inicial supuesto definirá el conjunto de objetivos críticos potencialmente alcanzables y las funciones legítimas que deben preservarse. Registrará las restricciones arquitectónicas previstas, la evidencia disponible y las incertidumbres que impidan concluir sobre el alcance del daño. Ninguna configuración será declarada satisfactoria con base únicamente en contención aparente si compromete una función crítica o desconoce dependencias materiales.

Criterios de implementación proporcional

Dimensión	Evidencia mínima recomendada	Regla de decisión orientativa
Alcance	Unidad, servicio, identidad, función crítica, versión y límites del modelo	Si el universo es insuficiente: NE, no PASS
Rutas adversarias	Puntos iniciales supuestos, permisos, transiciones, excepciones, rutas de egreso	Toda ruta residual material requiere tratamiento explícito
Continuidad	Dependencias de la función, alternativa de servicio, degradación permitida	Una contención que destruye la función no es éxito integral
Tiempo	T0, tiempo a protección (MTP), tiempo a cierre (MTC), tiempo a recuperación si aplica	Relojes diferentes; no usar un solo SLA como evidencia integral
Gobierno	Propietario, aceptación de riesgo, caducidad de excepciones, revisión ante cambios	Riesgos y excepciones deben quedar atribuidos y versionados
Comparabilidad	Mismo compromiso supuesto, horizonte y universo antes/después	No comparar configuraciones incompatibles o parciales
Reutilización	Pruebas y evaluaciones vigentes de PCI/ISO/NIST/SOC 2 u otras	No repetir cuando la evidencia existente ya resuelve la pregunta

Métricas para evaluación de escritorio y operación futura

Publicar y mantener separadas cuatro clases de métricas: (i) exposición y aplicabilidad de vulnerabilidades, (ii) fracción de objetivos críticos alcanzables según el modelo y su cobertura, (iii) continuidad de funciones legítimas y egreso, y (iv) tiempos observables de protección, cierre y recuperación cuando existan instrumentos autorizados. En una evaluación puramente documental, tiempo y eficacia real quedan no observados. Ningún valor hipotético debe transformarse en declaración de cumplimiento.

Políticas internas: cláusula de actualización

La política de seguridad de la información y la gestión del riesgo cibernético deberían incorporar una evaluación recurrente de arquitectura de contención para exposiciones críticas, roles con privilegios elevados,

integraciones de terceros y procesos de negocio esenciales. La revisión debe realizarse cuando cambien dependencias, permisos, salidas, topologías o funciones vitales y debe elevar incertidumbres materiales a gobierno de riesgo. La política puede implementarse inicialmente mediante ejercicios de escritorio y fuentes documentales sin operaciones activas sobre plataformas.

Registro mínimo de evidencia para diálogo con emisores de estándares

Campo	Registro verificable esperado
Identidad	Código de escenario, versión del protocolo, fecha, propietario y alcance
Contexto	Función esencial, identidad/punto inicial, objetivos críticos y dependencias
Aseguramiento preexistente	Estándares aplicables, pruebas previas y excepciones ya documentadas
Modelo de daño	Grafo de capacidades, egreso, privilegios transitivos y rutas de proveedores
Modelo de continuidad	Flujo legítimo, degradación aceptable y objetivos de recuperación aplicables
Decisión y límites	PASS, FAIL o NE justificado; incertidumbres y potencial falsador
Gobierno	Riesgo aceptado, fecha de reexamen, responsable y razón de la decisión

Este formato se propone como punto de partida de cooperación técnica, con evidencia proporcional y reutilización de material existente. No es una nueva plantilla de auditoría oficial ni modifica los procedimientos de acreditación de un QSA, un organismo de certificación ISO o una autoridad supervisora.

14 Invitación internacional a la evaluación técnica de la arquitectura de contención

ZTC Research propone abrir una discusión técnica sobre la incorporación proporcional de evidencia de contención arquitectónica en la gestión de riesgos cibernéticos. La investigación invita a quienes desarrollan marcos y orientaciones, a las autoridades con funciones de supervisión y a las organizaciones responsables de servicios críticos a examinar si el criterio CCA-01 permite describir mejor el daño residual y la continuidad legítima mientras se corrige una vulnerabilidad. No se solicita crear de inmediato un requisito obligatorio ni sustituir controles consolidados: la propuesta está abierta a refutación, adaptación y validación independiente.

Para los organismos que elaboran marcos y estándares técnicos

En sus ámbitos respectivos, NIST, ISO/IEC JTC 1/SC 27, PCI Security Standards Council y otros grupos especializados pueden encontrar útil examinar una matriz que conecte compromiso supuesto, objetivos críticos, rutas residuales, egreso, continuidad, incertidumbre y responsables. La aportación propuesta es un patrón de evidencia complementaria susceptible de discusión en perfiles, ejemplos, guías técnicas o futuras revisiones, mediante sus procesos ordinarios de consulta. CCA-01 no es un control aprobado por esas entidades ni reemplaza pruebas reglamentarias, requisitos PCI DSS o evaluaciones de conformidad ISO/IEC.

Para el diálogo de resiliencia financiera y supervisión

La declaración de 2025 del G7 Cyber Expert Group reconoce riesgos cibernéticos vinculados con IA y alienta la actualización de marcos y el diálogo entre sectores, sin fijar obligaciones regulatorias. Por su parte, CPMI-

IOSCO abrió en septiembre de 2026 una consulta sobre herramientas voluntarias de resiliencia para infraestructuras de mercado financiero, publicada a través del BIS. En esos espacios, CCA-01 podría someterse a consideración como insumo metodológico, especialmente respecto de interdependencias, propagación residual y preservación de operaciones críticas. Los bancos centrales y supervisores conservarían plena discrecionalidad dentro de sus facultades; esta investigación no representa una solicitud formal, una postura institucional ni un aval de tales entidades. [42–44]

Para organizaciones financieras, operadores y proveedores críticos

Las organizaciones pueden evaluar si sus políticas internas de arquitectura, segregación, gestión de vulnerabilidades, respuesta a incidentes, continuidad y gestión de terceros documentan el alcance potencial del daño después de un compromiso supuesto. Se recomienda preservar excepciones, incertidumbres, responsables y plazos de remediación, reutilizando la evidencia ya generada por los marcos aplicables. La utilización de CCA-01, en su estado actual, es voluntaria y metodológica; no sustituye pruebas de penetración, auditorías, certificaciones, evaluaciones regulatorias ni requisitos contractuales.

A la comunidad científica y profesionales de seguridad

Se abre la propuesta a revisión, reproducción y refutación técnica. Se solicita comparar su valor incremental con resultados de auditorías y marcos ya existentes, evitando contabilizar como innovación pruebas duplicadas. Un resultado adverso, un inventario incompleto o una incompatibilidad con continuidad son hallazgos útiles; ocultarlos debilitaría la investigación. Los datos de escritorio publicados permiten cuestionar la lógica del evaluador sin exponer infraestructura o información de terceros.

INVITACIÓN FORMAL A LA ADOPCIÓN EVALUADA

ZTC Research propone que la arquitectura de contención verificable CCA-01 sea considerada por organismos de normalización, autoridades y organizaciones como referencia complementaria para revisar marcos, políticas internas y prácticas de aseguramiento ante amenazas aceleradas por IA. Su adopción debería estar sujeta a consulta, pruebas apropiadas, proporcionalidad de riesgo y resguardo de obligaciones vigentes. Ningún organismo mencionado ha aprobado, certificado ni adoptado la propuesta por el hecho de citarse en este estudio.

15 Objeciones, alternativas y límites de inferencia

Objeción 1: más CVE no significa software peor. Es correcta. Una mayor cobertura de auditoría puede revelar defectos antiguos y producir más registros sin aumentar proporcionalmente la tasa de fallos introducidos. La mejora de identificación es una señal favorable para la defensa; la presión se desplaza al proceso de corrección. [03–05]

Objeción 2: la IA también aumenta la ventaja defensiva. Es correcta en varios casos documentados. El programa de coordinación de PHP y las investigaciones de Big Sleep muestran que descubrir, comunicar y corregir antes de la explotación puede ampliar la ventana del defensor. El beneficio depende de que las correcciones lleguen a los entornos reales. [10, 13, 17]

Objeción 3: escáneres y proveedores pueden sobredimensionar la amenaza. La distinción entre hallazgo, CVE, instancia afectada, exploit funcional, intento de ataque y compromiso confirmado es obligatoria. La cifra de NOVA y las muestras atribuidas a IA no deben incorporarse al total mundial como vulnerabilidades nuevas independientes. [06–08]

Objeción 4: 18 meses es una predicción arbitraria. Es un horizonte de opinión, útil para ejercicios de resiliencia, pero no para publicar una fecha de crisis o de recuperación global. Sería metodológicamente improcedente afirmar que un colapso de Internet está previsto. [15–16]

Objeción 5: la proporción de explotación observada es pequeña. Sí, pero una fracción reducida de fallas en sistemas muy extendidos puede concentrar daño. A la inversa, miles de CVE no aplicables pueden no exigir

trabajo operativo en una empresa concreta. El punto decisivo es la convergencia de criticidad, exposición y velocidad de respuesta. [03, 18]

16 Hoja de ruta de decisión y mejora continua

Fase	Objetivo	Evidencia de resultado
0–4 horas	Nombrar equipo de guardia; localizar activos críticos de acceso público y activar contención de emergencia.	Decisión registrada, responsables y ruta de aislamiento comprobable.
4–24 horas	Validar CVE aplicables y vías de ataque; aplicar protecciones temporales y vigilar actividad.	Registro de alcance y pruebas de bloqueo / compromisos previos.
24–72 horas	Desplegar correcciones urgentes seguras, retest, canary y rollback; revisar accesos laterales.	Versiones efectivas, pruebas negativas y cierre de casos máximos.
Días 4–7	Generalizar política de privilegio mínimo, inventario de dependencias y restricciones de egreso en sistemas esenciales.	Matriz de rutas y propietarios; pruebas de aislamiento entre servicios/tenants.
Días 8–14	Ejecutar ejercicios de explotación controlada y recuperación; formalizar objetivos por riesgo.	Cobertura de verificación, p90 de protección, excepciones con vencimiento.
Días 15–30	Automatizar cadena de actualización y pruebas; contratos de respuesta con terceros.	Balance entradas/cierres y evidencia de reducción de exposición.
Continuo: 1–18 meses	Sostener rediseño por límites de confianza; medir tendencias y validar independientemente.	Escenarios de compromiso contenidos y evidencia recurrente de resiliencia.

Gobierno de ciclo corto. Durante los primeros catorce días, revisar diariamente las alertas que afecten a activos esenciales; después, mantener supervisión continua con reporte semanal de excepciones y tendencias. La dirección evaluará al menos cuatro cifras: activos críticos sin cobertura; hallazgos relevantes de alto impacto aún expuestos; tiempo p90 hasta protección efectiva; y capacidad de cierre frente a nuevas entradas. La inversión en IA defensiva sólo se considera exitosa si reduce exposición y mejora cierre, no si incrementa alertas.

17 Alcance, incertidumbre y validación futura

El alcance de esta investigación es documental, comparativo y arquitectónico. Sus conclusiones distinguen datos observados, inferencias razonadas y propuestas que todavía requieren validación experimental. Explicitar estas fronteras permite reproducir los hallazgos, reconocer dónde termina la evidencia y precisar qué resultado obligaría a revisar las hipótesis.

Límites metodológicos y medidas de control

Dimensión	Limitación	Control de validez
Series CVE y KEV	Los repositorios difieren por ingesta, fechas, cambios retroactivos y criterios de inclusión. Publicación no equivale a descubrimiento.	Se comparan tendencias dentro de cada fuente, con fecha de corte; se evita construir un total mundial único o una tasa KEV/CVE artificial. [01–05]

Dimensión	Limitación	Control de validez
Atribución a IA	No existe un campo universal verificable que indique cómo se descubrió cada CVE. Las muestras divulgadas por proveedores pueden tener sesgos de selección.	Se distinguen casos documentados de series agregadas. No se estima un porcentaje causal mundial atribuible a IA. [06–10]
Explotación real	CISA KEV y la telemetría de empresas de seguridad captan universos distintos; ausencia de observación no demuestra ausencia de ataques.	Se separan explotación observada, exploit disponible, intentos y compromiso. Se declara el denominador de cada afirmación. [02–03]
Plazos de corrección	Las normas expresan obligaciones o frecuencias para poblaciones distintas; los tiempos de campo usan relojes, activos y muestras no siempre comparables.	Se diferencian reconocimiento, protección y cierre; las medianas observadas no se confunden con un plazo regulatorio universal. [19–22, 29]
Eficacia arquitectónica	El modelo de límites de confianza se apoya en principios técnicos; no se ejecutaron pruebas activas ni comparaciones controladas en infraestructuras de clientes.	Se presentan mecanismos y pruebas negativas propuestas, sin atribuir porcentajes de reducción de incidentes ni inmunidad. [23–24, 31]
Alcance sectorial y económico	Las fuentes internacionales no constituyen una muestra representativa de México o Latinoamérica. No se dispone de costos e impactos medidos por entidad.	Se evita extrapolar tasas sectoriales y retornos financieros; se proponen métricas de exposición, continuidad y costo para validación local.
Horizonte y vigencia	La advertencia de dieciocho meses no es una predicción estadística. Los registros y la dinámica de capacidades de IA pueden cambiar tras el corte de 2026.	El horizonte se usa como escenario de estrés, no como fecha cierta; las cifras deberán reconfirmarse en futuras ediciones. [15–16]

Qué permiten afirmar estos resultados

La evidencia respalda tres afirmaciones: las divulgaciones aumentaron marcadamente en las series examinadas; existen descubrimientos relevantes realizados con asistencia de IA; y las decisiones de arquitectura pueden restringir rutas de ataque mediante controles cuya eficacia es posible probar. La magnitud de ese beneficio depende de los activos, la configuración, los adversarios y la continuidad de los controles.

La evidencia no permite afirmar que la IA explique por sí sola el crecimiento mundial de los CVE, que cada registro tenga explotación efectiva, que toda organización enfrente idénticos plazos, que los dieciocho meses constituyan una previsión cuantificada o que una topología particular reduzca el riesgo en un porcentaje universal. Los escenarios y objetivos propuestos son hipótesis de operación sujetas a prueba; no son resultados empíricos de una implantación.

Evidencia adicional necesaria para avanzar

Reproducibilidad temporal: publicar capturas fechadas y consultas de cada serie, reglas de deduplicación y registro de correcciones posteriores.

Atribución independiente: documentar para cada hallazgo el papel concreto de la IA, la revisión humana, la prueba de concepto y el momento de divulgación.

Validación arquitectónica: comparar, en laboratorio autorizado y con datos sintéticos, el mismo componente vulnerable bajo diseños diferentes; medir alcanzabilidad, escalada, acceso cruzado entre clientes y egreso.

Seguimiento longitudinal: registrar por activo los tres relojes (reconocimiento, protección y cierre), percentiles, fallas de contención, efectos sobre disponibilidad y costos verificables.

Criterio de revisión y posible refutación

La hipótesis arquitectónica debe limitarse o reformularse si una prueba controlada demuestra que el compromiso de un componente permite cruzar límites declarados, acceder a datos fuera de autorización o mantener egreso indebido. También deben revisarse los objetivos de remediación si la contención en horas no puede verificarse o produce daños operativos superiores al riesgo que reduce. Una investigación seria conserva estas condiciones de refutación abiertas y documentadas.

Estado de las hipótesis y próximo criterio de comprobación

Hipótesis	Estado de la evidencia	Prueba decisiva pendiente
H1 · CVE	Crecimiento observado en las series examinadas.	Reconstruir capturas de los repositorios e identificar efectos de cambios de registro. [01–05]
H2 · IA	Casos reales demostrados; causalidad global no cuantificada.	Construir una cohorte con atribución del método de descubrimiento y contraste independiente. [06–10]
H3 · Respuesta	Desajuste plausible; magnitud depende del entorno.	Medir nuevas entradas, cierre definitivo, tiempo a protección y deuda por activo esencial. [29]
H4 · Arquitectura	Mecanismo técnicamente sustentado; efecto empírico no medido aquí.	Enfrentar una misma falla en dos diseños y documentar rutas bloqueadas y rutas que persisten. [23–24,31]

18 Conclusiones e implicaciones estratégicas

La evidencia disponible documenta un cambio importante en el volumen de divulgaciones de vulnerabilidades y casos concretos de automatización asistida por IA. La magnitud exacta del efecto causal agregado de la IA no puede estimarse de modo universal con las fuentes públicas utilizadas; toda estadística debe conservar su unidad de observación y ventana temporal.

La arquitectura no reemplaza el parche. Una organización puede necesitar corregir un defecto durante días, mientras su exposición crítica exige una decisión sobre el daño posible desde el primer momento. El resultado relevante no es una promesa de invulnerabilidad, sino la delimitación de consecuencias bajo hipótesis transparentes, evidencia documentada y actualización continua.

La contribución CCA-01 consiste en transformar la contención arquitectónica en una decisión conjunta sobre daño residual, continuidad legítima y suficiencia de evidencia. El primer conjunto sintético verifica reglas del índice B; el segundo somete el criterio dual a casos favorables, adversos y no estimables. Estos resultados evidencian que una segmentación declarada, una media favorable o una topología incompleta no justifican por sí mismas aprobar una arquitectura. El alcance de la demostración es metodológico: la eficacia operativa requiere otro tipo de evidencia.

La agenda de gobernanza propuesta es concreta: (1) describir rutas de compromiso y objetivos críticos; (2) declarar fronteras y excepciones; (3) verificar, mediante la evidencia que corresponda, las consecuencias residuales; (4) preservar funciones críticas y tiempos sectoriales de recuperación; (5) integrar el resultado a políticas, procesos y responsabilidades; y (6) exigir trazabilidad al adoptar o modificar los marcos relevantes.

Se invita a responsables de marcos técnicos, investigadores, entidades financieras y autoridades pertinentes a examinar el valor incremental de CCA-01. Los debates del G7 Cyber Expert Group y CPMI-IOSCO aportan contexto, sin implicar acogida, patrocinio ni validación. La propuesta permanece abierta a mejora y refutación; los ensayos sintéticos no acreditan eficacia operativa.

CONCLUSIÓN PRINCIPAL

La pregunta que merece incorporarse a las políticas y a los marcos de seguridad es verificable: “Si hoy se compromete un componente, ¿qué objetivos críticos puede alcanzar el adversario y qué funciones del negocio seguirán operando de forma segura antes de la corrección definitiva?”. La arquitectura debe poder responderla con evidencia y con incertidumbre explícita.

A. Datos y series utilizadas en figuras

Mes	CVE 2025	CVE 2026	Var. anual	KEV 2025	KEV 2026
Ene	4,276	4,311	+0.8%	14	17
Feb	3,676	4,617	+25.6%	27	28
Mar	4,019	6,261	+55.8%	32	26
Abr	4,030	5,823	+44.5%	15	31
May	3,987	6,942	+74.1%	24	21
Jun	3,681	7,920	+115.2%	20	23
Jul	3,792	9,770	+157.6%	20	26
Ago	3,630	12,256	+237.6%	15	31
Sep	4,324	14,933	+245.4%	16	43

La serie CVE procede de CVETodo [01] y la serie KEV de CVE Tools [02]; se muestran próximas para comparación visual pero no se calculan cocientes interpretados como riesgo de explotación. Los valores de publicaciones pueden modificarse por revisiones, cargas tardías y diferencias de sincronización. Fechas de publicación y de incorporación a KEV no necesariamente coinciden.

B. Glosario de variables y términos

Término	Definición de trabajo
CVE	Identificador de una vulnerabilidad pública; no es un incidente ni prueba de que un activo esté expuesto.
CWE	Categoría de debilidad de software; agrupa clases de defectos.
KEV	Catálogo de vulnerabilidades cuya explotación en el mundo real está confirmada según los criterios de CISA.
EPSS	Estimación estadística de probabilidad de explotación observada en los siguientes 30 días; no calcula el impacto empresarial.
CVD	Divulgación coordinada para dar espacio a correcciones y medidas de protección antes de exposición masiva.
RCE	Ejecución remota de código; efecto potencial cuyo contexto, condiciones y privilegios deben investigarse.
SBOM / VEX	Inventario de componentes de software / declaración de afectación o no afectación con evidencia.
Cierre verificado	Comprobación técnica de que una medida reduce o elimina la exposición en el sistema realmente desplegado.

C. Trazabilidad, reproducibilidad y falsabilidad

El conjunto E-01A fue publicado como un ejercicio del evaluador genérico y documenta 12 casos y valores B construidos, con 8 PASS, 3 FAIL y 1 NE. El conjunto E-01B contiene 16 escenarios sintéticos, evaluados mediante Python 3 en un entorno local, con 4 PASS, 11 FAIL y 1 NE. Los archivos de reproducción correspondientes deben conservarse junto con el estudio; sus hashes y versiones se registran fuera del PDF público en el expediente técnico de publicación.

La falsabilidad del método exige analizar contraejemplos: rutas ocultas, credenciales reutilizadas, excepciones de emergencia, dependencias de terceros, cambios concurrentes, pérdida de misión y ausencia de inventario. En una evaluación de escritorio no se mide latencia de control, compromiso real ni fiabilidad de una implementación. Para un estudio externo posterior, los criterios deberían registrarse antes de observar resultados y compararse con evidencia ya existente a fin de medir el valor incremental.

Tipo de evidencia	Admisible en este estudio	No se infiere
Documento primario	Versiones, fechas, afirmaciones y alcance de la fuente	Causalidad universal ni conformidad de entidades no evaluadas
Grafo sintético	Regla lógica y resultado sobre nodos ficticios	Seguridad, segmentación o recuperación de una plataforma
Prueba unitaria local	Integridad limitada de funciones de simulación	Validación de campo o aval académico
Correspondencia de marcos	Relación temática y propuesta complementaria	Nuevo requisito vinculante, aprobación de NIST/ISO/PCI/BIS
Hipótesis de arquitectura	Consecuencia que podría probarse con información pertinente	Reducción estadística de riesgo ni efecto monetario

D. Ruta de replicación y valor incremental

Un estudio posterior, enteramente voluntario y distinto de esta publicación, podría comparar dos configuraciones documentadas bajo escenarios emparejados, con condiciones de compromiso, dependencias, permisos, controles y funciones esenciales fijadas antes del análisis. En ausencia de autorización para evaluación activa, el mismo diseño puede aplicarse a grafos documentales y pruebas de consistencia; ese modo solo valida el modelo, no la infraestructura.

Una comparación metodológica rigurosa debe identificar por cada control: evidencia que ya exigen o generan PCI DSS, ISO/IEC, NIST o programas equivalentes; pregunta material aún no resuelta; dato incremental solicitado por CCA-01; costo de obtenerlo; y decisión de seguridad efectivamente modificada. Si no aparece una decisión adicional, la reutilización de evidencia preexistente tiene prioridad sobre cualquier prueba nueva.

Para replicación independiente de escritorio deben preservarse el codebook de escenarios, las definiciones de frontera, el conjunto fijo de nodos y aristas, las reglas de adjudicación, el software de evaluación, los casos adversos y sus resultados. La documentación adicional puede incluir un registro de cambios, descripción del tratamiento NE y una comparación con las técnicas previas de attack graphs. Las ejecuciones sintéticas publicadas permiten verificar la lógica; no se atribuyen revisiones humanas o arbitradas que no constan en el expediente.

E. Naturaleza institucional y declaración de intereses

ZTC Research es la iniciativa de investigación de Zero Trust Consulting (ZTC), firma privada de consultoría especializada en arquitectura, ciberseguridad, criptografía y cumplimiento. La afinidad entre su actividad profesional y la materia de este estudio se declara por transparencia institucional; tal afinidad no demuestra, por sí misma, un conflicto de interés ni invalida el método. La independencia metodológica de la investigación debe examinarse a partir de sus reglas de análisis, la trazabilidad de las fuentes, la distinción entre hechos e inferencias y la presentación de resultados favorables y adversos.

La propuesta CCA-01 puede resultar pertinente para servicios de asesoría en arquitectura, riesgo y resiliencia, incluidos los que presta ZTC. Esa posible convergencia de intereses profesionales se informa de forma expresa, sin equipararla a un interés indebido ni a una evaluación externa independiente. El estudio publica sus supuestos, criterios y escenarios sintéticos con resultados favorables, fallidos y no estimables. Sus inferencias se circunscriben a la evidencia documental y al modelo construido, por lo que no acreditan eficacia de controles en entornos operativos ni adopción por terceros.

Referencias y fuentes

Referencias de la investigación principal [01–32] y fuentes de correspondencia normativa [33–45]. Las referencias a organismos se ofrecen como fundamento de sus marcos vigentes, no como atribución de aceptación de CCA-01. Fecha de actualización documental: 2 de noviembre de 2026.

[01] CVETodo — CVE Statistics 2026; serie mensual 2025-2026. Consulta: 31 oct 2026. | Enlace

[02] CVE Tools — Monthly CVE trends; serie de snapshots y adiciones a CISA KEV. Consulta: 31 oct 2026. | Enlace

[03] Google Threat Intelligence Group — Vulnerability Discovery and Exploitation Trends in the AI Era, 30 sep 2026. | Enlace

[04] FIRST — 2026 Vulnerability Forecast, 11 feb 2026. | Enlace

[05] FIRST — Mid-Year Vulnerability Forecast, 15 jun 2026. | Enlace

[06] IBM X-Force — AI can find serious vulnerabilities at unprecedented speed. Can defenders keep up?, oct 2026. | Enlace

[07] Palo Alto Networks Unit 42 — The Frontier AI Vulnerability Burst: Industrializing Autonomous Zero-Day Discovery, 4 ago 2026. | Enlace

[08] Securin — AI-Discovered CVE Tracker; base atribuida con niveles de evidencia, 27 oct 2026. | Enlace

[09] PHP Group — Security advisory y versiones corregidas de PHP2Shell, 21 y 23 oct 2026. | Enlace

[10] OpenAI — Patch the Planet: global response coordination as agents uncover critical vulnerabilities, 21 oct 2026. | Enlace

[11] CISA — BOD 26-04: Prioritizing Security Updates Based on Risk, jun 2026. | Enlace

[12] FIRST — EPSS Frequently Asked Questions: alcance, usos y limitaciones. | Enlace

[13] Google Project Zero — From Naptime to Big Sleep, 1 nov 2024. | Enlace

[14] Jon Hernández — La IA está rompiendo la seguridad de Internet, 26 oct 2026. | Enlace

[15] Jason D. Clinton — Publicación en X, 22 oct 2026; contenido corroborado mediante reproducción de terceros. | Enlace

[16] Mantic — Black October and the future of cybersecurity, 27 oct 2026. | Enlace

[17] Google — Big Sleep y descubrimiento de CVE-2025-6965, 2025. | Enlace

[18] CISA — Vulnerability Review: controles básicos, priorización y explotación, 2026. | Enlace

[19] PCI Security Standards Council — FAQ 1597, distinción de 6.3.3 y 11.3; actualizada mayo 2025 | Enlace

[20] CISA — BOD 26-04, Directiva oficial, 10 jun 2026; plazos condicionales | Enlace

[21] Center for Internet Security — CIS Controls v8.1, Control 7, frecuencias de proceso y remediación | Enlace

[22] Comisión Europea — Reglamento delegado (UE) 2024/1774, artículo 10, DORA | Enlace

[23] NIST — SP 800-207, Zero Trust Architecture, ed. 2020 | Enlace

[24] NIST — SP 800-160 Vol. 2 Rev. 1, cyber resiliency engineering, dic 2021 | Enlace

[25] RSAC Conference — Jason Clinton, perfil biográfico y trayectoria verificada | Enlace

[26] Clarín / DEF CON 34 — Declaración de Jason Clinton sobre periodo de 18 a 24 meses, ago 2026 | Enlace

[27] Microsoft Learn — Zero Trust, límites de confianza y defensa en profundidad; guía de arquitectura | Enlace

[28] NIST — SP 800-40 Rev. 4, planificación de gestión de parches, abr 2022 | Enlace

[29] Verizon — 2026 Data Breach Investigations Report; cobertura KEV y mediana de cierre (datos de 2025). | Enlace

[30] Google Cloud / Mandiant — M-Trends 2026; tiempo a explotación estimado de -7 días en la muestra. | Enlace

[31] NIST — SP 800-207A, identidad y controles de acceso por servicio en cloud-native. | Enlace
[32] FedRAMP — Respuesta institucional a CISA BOD 26-04 y calendario de implementación. | Enlace
[33] NIST. The NIST Cybersecurity Framework (CSF) 2.0. CSWP 29 (2024). | Enlace
[34] NIST. SP 800-61 Revision 3, Incident Response Recommendations (2025). | Enlace
[35] NIST. SP 800-207, Zero Trust Architecture (2020). | Enlace
[36] ISO/IEC. 27001:2022, Information Security Management Systems - Requirements. | Enlace
[37] ISO/IEC. 27002:2022, Information security controls. | Enlace
[38] ISO/IEC. 27005:2022, Guidance on managing information security risks. | Enlace
[39] PCI SSC. Document Library: PCI DSS v4.0.1 (June 2024). | Enlace
[40] PCI SSC. Guidance for Compensating Controls and the Customized Approach (10 June 2026). | Enlace
[41] CPMI-IOSCO / BIS. Guidance on cyber resilience for financial market infrastructures (2016). | Enlace
[42] CPMI-IOSCO / BIS. Cyber resilience toolkit: practical considerations for FMIs, consultation (8 Sep 2026). | Enlace
[43] CPMI-IOSCO / BIS. FMIs reliance on third-party service providers: discussion paper (8 Sep 2026). | Enlace
[44] G7 Cyber Expert Group. Statement on AI and Cybersecurity (Sep 2025; published Oct 2025). | Enlace
[45] ISO/IEC. 42001:2023, Artificial Intelligence Management System. | Enlace

Cita sugerida

ZTC Research / Zero Trust Consulting (2026). La nueva asimetría de la ciberseguridad: la arquitectura de contención como límite verificable del daño en la era de la inteligencia artificial. ZTCR-AI-VM-2026-01, versión 1.0, 2 de noviembre de 2026. México.

© 2026 Zero Trust Consulting. Se permite la cita con atribución del estudio y conservación de su alcance metodológico. Los escenarios sintéticos no podrán presentarse como evidencia de protección operativa, certificación o aval de terceros.