ZERO TRUST CONSULTING

ZTC RESEARCH / INDEPENDENT METHODOLOGICAL STUDY

THE NEW ASYMMETRY OF CYBERSECURITY

Containment architecture as a verifiable bound on harm in the age of artificial intelligence

METHODOLOGICAL CONTRIBUTION / CCA-01

Contain adversarial reach. Preserve essential operations.

CCA-01 joins residual critical-asset reachability, continuity of legitimate functions, and the completeness of dependency evidence in one auditable decision. Unknown material dependencies require an indeterminate finding.

RESEARCH DIRECTOR

Dr. Juan Pablo Ybarra Llamas

ZTC Research / Zero Trust Consulting

VERSION 1.0

2 November 2026 | Mexico | ZTCR-AI-VM-2026-01

zerotrust.consulting/research

00 Executive Summary and Contribution

The acceleration of AI-assisted vulnerability research is changing the cost of discovering and analyzing flaws, but it does not establish that every CVE is immediately exploited. The governance question is more precise: during the interval between identifying an exposure and remediating it, which paths to privileges, data, or services remain open, and which essential functions would remain available under isolation?

ZTC Research contributes CCA-01: a decision rule that links three questions that cannot be answered independently: (i) which adversarial paths still permit access to critical assets or data; (ii) which essential functions must continue operating within the same defensive boundaries; and (iii) whether the dependency model is sufficiently complete to support the finding. An architecture that stops propagation but disables an essential process does not pass the combined criterion; neither does one that preserves service while leaving critical paths open. Where material evidence is insufficient, the outcome is not estimable rather than assumed compliant.

A DIFFERENT DECISION

A patch inventory tells us when to fix a defect. CCA-01 seeks to identify what consequences may arise while the defect remains, and what must continue operating. Missing information does not count as protection; blocking a path at the expense of the critical function does not count as success.

The evidence includes public disclosure and exploitation series with stated denominators and two separate synthetic datasets: 12 scenarios using the weighted B index and 16 scenarios using the PDC damage-and-continuity vector. They are reproducible illustrations of logical rules, not empirical measurements of companies, suppliers, or infrastructures. Their value lies in identifying the conditions under which an evaluator must return FAIL or abstain.

The proposal is offered for discussion among technical communities, security-framework developers, financial authorities, and organizations responsible for critical functions. NIST, ISO/IEC, PCI SSC, and CPMI-IOSCO are relevant points of reference within their respective mandates; the G7 Cyber Expert Group is a forum for dialogue on financial-sector resilience, not a universal standards-setting body. The question is whether CCA-01 evidence on containment and continuity adds value to existing instruments—not whether the proposal should be immediately adopted or treated as a binding requirement.

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 relates the proposal to existing NIST, ISO/IEC, PCI SSC, and CPMI-IOSCO frameworks, while distinguishing documentary evidence, constructed synthetic examples, and operational validation. Standards bodies, financial authorities, and critical-sector organizations are invited to assess the proportionate incorporation of this outcome-oriented evidence into their assurance and governance practices.

Structure of the Study

Public evidence and sources · Architecture and consequences · CCA-01 protocol · Synthetic tests and falsification · Framework crosswalk · Invitation to technical evaluation · Limitations and recommendations.

01 Research Question, Hypotheses, and Method



Figure 1. Decision sequence from disclosure to impact. This is a map of relationships, not a universal causal inference.

Testable Hypotheses

H1. CVE disclosures are increasing rapidly. H2. AI contributes to documented discovery, without explaining all global variation. H3. Risk depends on the difference between incoming applicable findings and effective containment/closure capacity. H4. Architecture can reduce the reachability and blast radius of exploitable defects—a hypothesis requiring asset-level tests of isolation, privileges, and exfiltration.

The method combines critical review of public sources, comparison of quantitative time series, and technical case analysis. No original code audit, exploitation, active scanning, or forensic investigation was performed. Recorded facts, vendor statements, inferences, and constructed scenarios are distinguished throughout.

What Each Source Measures

Source	Observed unit	Use and limitation
CVETodo	Monthly published CVEs; NVD/CVE/KEV	Consolidated 2025–2026 series, not a primary CVE source. Does not universally identify discovery method.
CVE Tools	Monthly CVEs and monthly KEV additions	Corroboration with similar but not identical figures. KEV represents a separate category and date.
GTIG	Disclosure, internal risk, and observed exploitation	Intelligence data with its own coverage and methods; not a universal audit of incidents.
IBM / Unit 42 / Securin	AI-attributed samples and campaigns	Evidence of technical capability, with selection bias and varied definitions and provenance.
FIRST	Disclosure forecasts	Earlier projections; comparisons with later observations must respect methods and cutoffs.

Methodological rule. “AI-discovered” is reserved for results explicitly describing their method in a primary source or supported by corroborable public attribution. A “disclosed CVE” is not synonymous with a “vulnerability discovered by AI”; “addition to KEV” does not mean “exploitation began in that month.”

Evidence and Confidence Levels

Level	Criterion	Example
High	Maintainer advisory, explicit data and method, or technical records from a primary entity	PHP Group; GTIG for its own data; CVETodo for its own compiled series

Level	Criterion	Example
Medium	Vendor-led research with checkable figures but partial reproducibility	NOVA; IBM sample; Securin tracker
Exploratory	Opinion, social-media post, forecast, or scenario without independent empirical verification	18-month horizon; backlog saturation models

02 The Observed Change in Vulnerability Volume

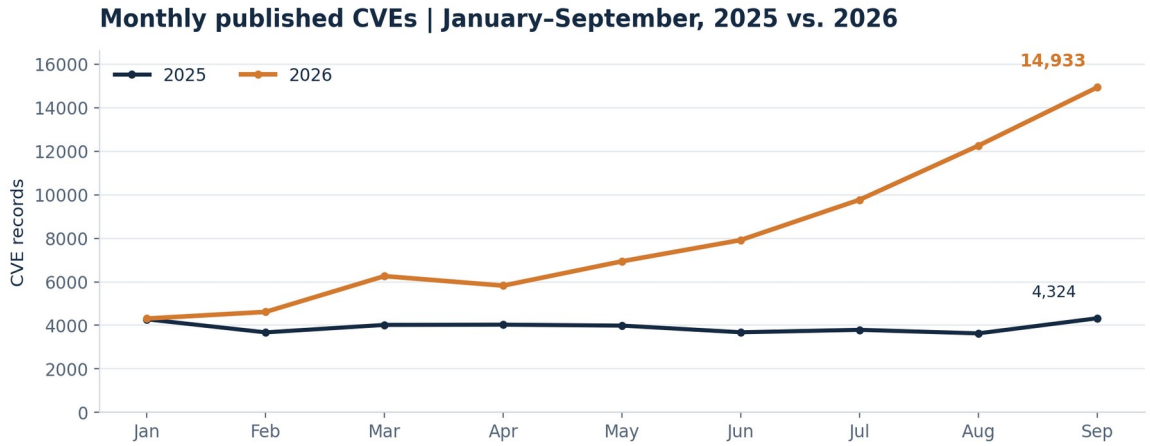


Figure 2. Monthly published CVEs, January–September 2025 and 2026. Prepared by ZTC Research from the series published by CVETodo [01].

The difference between January and September 2026 is especially notable: 4,311 versus 14,933 publications, an increase of 246.4% within the year. The September-to-September comparison yields 245.4%. Over the first nine months, the series records 72,833 versus 35,415 (+105.7%). [01]

This change occurred over months, not merely within an annual cycle. It challenges remediation processes organized as isolated quarterly projects. Nevertheless, before translating this curve into risk, one must determine which CVEs actually apply to each organization’s inventory, which have viable attack paths, and which can reach essential services.

Cross-Validation and Legitimate Differences

CVE Tools reports 15,180 CVEs in September 2026 and 4,664 in September 2025, compared with 14,933 and 4,324 in CVETodo. The results corroborate the direction and magnitude of the increase but are not interchangeable. Differences may reflect cut-off dates, synchronization, inclusion rules, and subsequent record updates. [01–02]

GTIG recorded 5,045 disclosures in January 2026 and 10,740 in August within its own coverage; this increase is not identical to CVETodo’s. In June, FIRST revised its central annual forecast from approximately 59,427 to 66,000 CVEs. These forecasts are historical records of expectations at the time, not final counts. [03–05]

Disclosure Volume Versus Observed Exploitation

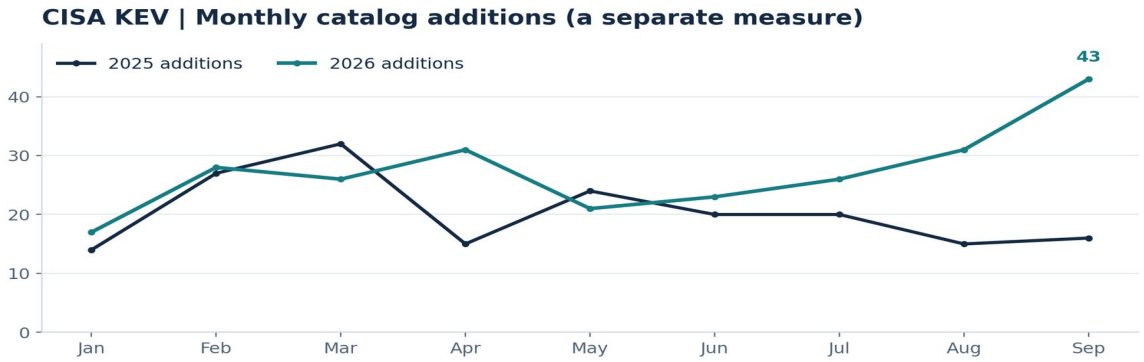


Figure 3. Monthly additions to the CISA KEV catalog according to CVE Tools, January–September 2025 and 2026. The date refers to addition to KEV, not original CVE publication. [02]

It would be inappropriate to divide KEV additions by CVEs disclosed in the same month and call the result a "real exploitation rate": the two flows can represent different populations and dates. The comparison illustrates orders of magnitude and the importance of context. [02]

Using a different observation method, GTIG reported 141 disclosed vulnerabilities actively exploited between January and August 2026, approximately 18 per month, versus a monthly average of 10.5 in 2025. In that 2026 window, the observed exploited share was approximately 0.23% of total disclosures. This measurement is constrained by visibility; it does not demonstrate that other entries were never exploited. [03]

The Measured Gap Between Risk and Remediation

Verizon's 2026 Data Breach Investigations Report states that, for its 2025 dataset, only 26% of vulnerabilities listed in CISA KEV were fully remediated; median time to full remediation was 43 days, compared with 32 days in the previous period. These are neither global metrics for all companies nor intervals measured from each CVE's publication date. They apply to the coverage and unit of analysis defined by Verizon. [29]

That lag helps explain architecture's value: an organization unable to remediate immediately must limit reachability, access, and data exposure while the fix is underway. It would be incorrect simply to subtract a three-day federal deadline from Verizon's 43 days; populations, time origins, and completion criteria differ. A sound approach measures time to protection, time to verified closure, and the number of still-exposed instances separately. [20, 29]

03 What Can Actually Be Attributed to Artificial Intelligence

Evidence of Capability Is Not a Population-Wide Rate

NOVA (Unit 42). The team reported examining 3,915 open-source projects and confirming 14,090 findings over two months, 99.4% of which had not been previously reported. Of those findings, 39.7% were rated high or critical under CVSS v4.0. These are findings from a research and validation campaign—not 14,090 issued CVEs or 14,090 incidents. The study comes from the system's developer; independently reproduced verification of the full sample was not published to an equivalent extent. [07]

IBM X-Force sample. Among 223 vulnerabilities publicly reported as discovered with frontier models, IBM observed that 55.2% were high/critical severity, 91% were remotely accessible, and 25.6% were compatible with code execution. The figures describe that selected sample; they neither estimate proportions across all CVEs nor prove that AI discovers only severe flaws. [06]

Securin tracker. It lists 597 CVEs associated with AI actors or systems but distinguishes attribution quality: 356 have explicit public evidence of AI-assisted discovery (166 autonomous and 190 human-assisted); another 178 were linked only to people associated with a laboratory, 14 had no declared method, and 49

lacked located public attribution evidence. The precision of these distinctions matters as much as the total count. [08]

Verifiable precedent. Google Project Zero and DeepMind published a real SQLite finding identified using Big Sleep in 2024; CVE-2025-6965 was documented in 2025 with assistance from the same program. These technical advances therefore predate the 2026 acceleration in registrations. [13, 17]

Why Easy Causal Inference Fails

CVE counts may grow because of AI research, expansion of the CVE program, automated tagging, changes among numbering authorities, greater transparency, waves of vendor disclosures, mass reviews of older projects, or improved synchronization. GTIG highlighted roughly 5,000 CVEs associated with Linux Kernel descriptions between January and August 2026 and stated that it did not observe actively exploited zero-days in that subset. [03]

Estimating AI's causal effect would require linking each finding to its discovery method, deduplicating related exposures, controlling for vendors, software categories, and disclosure cycles, and comparing outcomes with a counterfactual. Publicly available information is not sufficient to do so robustly. The observed CVE increase is a fact; attributing all of it to AI would be unsupported.

04 PHP2Shell: An Illustrative Case and Its Limits

On 21 October 2026, PHP announced fixes for CVE-2026-109618 and CVE-2026-109619. Both can enable unauthenticated remote code execution under affected conditions. The first concerns improper memory handling under resource exhaustion; the second concerns internal state retained after memory is freed. The project highlighted heightened risk for mod_php and warned that updating was the only complete protection available. [09]

OpenAI reported that Calif identified one of the flaws using a frontier model, and that Wiz independently discovered and reported both. PHP maintainers led remediation and coordinated the response with other parties. The evidence documents a real contribution by AI tools, with human researchers and software maintainers remaining central to the process. [10]

PHP branch	Minimum patched version	Operational interpretation
8.2	8.2.35	Plan migration before end of support; verify deployed binary.
8.3	8.3.36	Update and confirm effective running processes.
8.4	8.4.27	Update and verify all instances/containers.
8.5	8.5.12	Update; on Windows, validate subsequent rebuilds where relevant.

Scope qualification. PHP's prevalence across websites does not mean that the same proportion is vulnerable or that exploitation has been confirmed at a comparable scale. Assessment requires identifying effective version and distribution, execution mode, exposure path, controls, updates, and, where appropriate, forensic records. [09]

On 23 October, PHP also described a specific scenario involving `session.upload_progress.enabled` and dynamic changes to `session.serialize_handler`, identified as CVE-2026-115560. It treated the issue as an unusual combination awaiting a fix in the regular release cycle and recommended a configuration mitigation. This must not be confused with the two principal CVEs. [09]

Architectural lesson. A runtime flaw may be triggered before the application executes its ordinary authentication logic. API identity controls do not replace runtime hygiene; isolation, least privilege, egress restrictions, segmentation, and perimeter protection can limit consequences but cannot replace updating when the vendor states that no complete mitigations exist.

05 The Next 18 Months: A Pressure Scenario, Not a Collapse Date

Who Is Jason D. Clinton and Why Does His Assessment Matter?

Jason D. Clinton is Anthropic's Deputy Chief Information Security Officer and formerly its first head of security. He previously led security for the Google Chrome infrastructure; his experience includes ChromeOS, Android Pay, and defense against advanced persistent threats. His professional biography is publicly available through RSAC Conference. His assessment thus draws on widely deployed software infrastructure and advanced AI development, but his forecasts should not be treated as demonstrated forward-looking evidence. [25]

At DEF CON 34 in August 2026, Clinton warned of a potentially exceptional 18-to-24-month period for the field. On 22 October, he reduced his working horizon to roughly 18 months and urged stronger basic controls. He had also advocated isolating critical infrastructure when patching could not keep pace with discovery. These are attributed statements, not results from a quantitative forecast. [15, 26]

His reasoning introduces an important tension: the same systems capable of finding errors might eventually make their removal cheaper. The expectation that more widely available models will ultimately reduce the introduction of defects remains Clinton's hypothesis, not a proven outcome. [26]

Clinton's public warning of 22 October 2026 is a call to strengthen fundamentals before adding tools. His roughly 18-month horizon may inform stress scenarios; it is not a date for a predicted crisis, nor evidence that every sector will inevitably be compromised. [15, 26]

Three Plausible Pressure Mechanisms

Faster inflow: research agents broaden repository coverage, generate hypotheses, validate cases, and reduce the cost of finding difficult flaws.

Potentially shorter offensive window: technical disclosures, patch diffs, and test automation can accelerate attack reproduction, although no universal reduction in time to exploitation has been established across all vulnerability families.

Slow closure due to dependencies: applicable findings require asset-owner approval, version validation, testing, maintenance windows, coordination with third parties, and confirmation that the patch actually runs in production.

Illustrative Backlog Saturation Model

We define D_t as the number of applicable unresolved findings at the end of month t; λ_t denotes newly verified inputs and μ_t verified closures. The basic accounting identity is $D_t = \max(0, D_{t-1} + \lambda_t - \mu_t)$. This does not incorporate severity, risk, or dependencies, which must be measured separately. It is neither a ZTC forecast nor a market-derived figure.

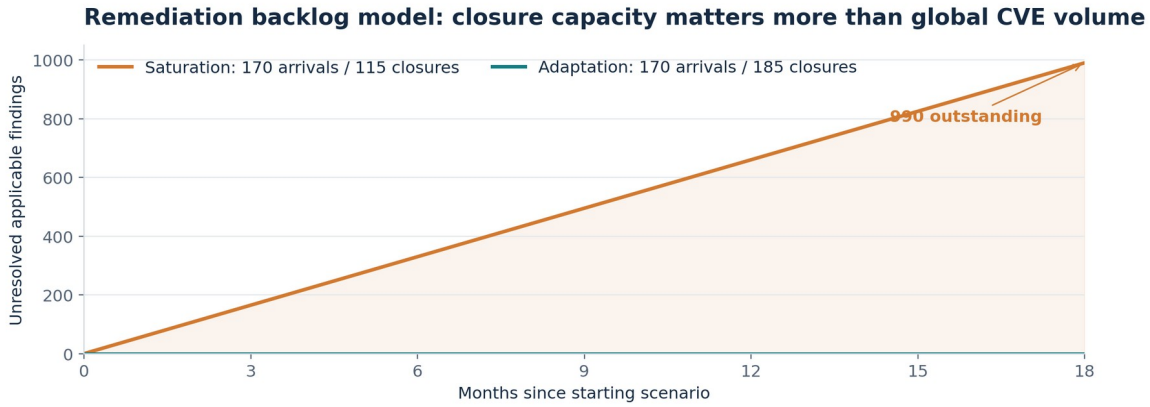


Figure 4. An 18-month didactic simulation using hypothetical, not observed, data. Under saturation, the backlog grows by 55 cases per month and reaches 990; with sufficient capacity, no backlog accumulates.

The model helps identify a board-level decision: increase closure capacity, reduce asset exposure, automate verification, or eliminate work that was never relevant. Adding tools that generate alerts without producing verified closures may aggravate the backlog without reducing risk.

Early-Warning Indicators

Metric	Operational definition	Relevant signal
Inventory coverage	Percentage of exposed assets with owner, version, dependency, and criticality	Increase in ownerless or versionless assets.
Time to classification	Hours from alert to applicability/priority decision	Long delays even for exposed services.
Time to protection	Hours until a measured temporary control or effective patch	Exceeds observed adversarial activity window.
Time to verified closure	Hours/days until version and absence of exposure are confirmed	Administrative closure without technical evidence.
Relevant backlog	Open applicable findings on critical assets, age-weighted	New cases exceed closures across cycles.
Third-party coverage	Services and libraries dependent on suppliers with tested SLA	No evidence of an update or alternative isolation.

One Extreme Case Cannot Define the Statistics

A demonstration of compromise within minutes may reflect a real attacker capability under specific conditions, but it does not establish the mean time to Internet exploitation, administrative privilege acquisition, or impact across all organizations. The environment, starting point, population, clock definition, and success criterion must be known. [29–30]

Mandiant estimated an average interval of −7 days between initial exploitation and relevant disclosure within its investigated exploitation sample, indicating that some cases were exploited before becoming public. This is not the mean time to compromise for any given organization. Collapsing different metrics into one acceleration curve would create false precision. [30]

06 A Defense Model Oriented Toward Exposure Time

Per-Vulnerability Decision Sequence

The response criterion should combine technical applicability + reachable exposure + evidence of exploitation + business impact + change constraints. CVSS describes technical severity, EPSS estimates the probability of exploitation observed over the next 30 days under its model, and KEV identifies confirmed exploitation under CISA's criteria. None substitutes for the organization's own context. [11–12]

Stage	Minimum evidence for completion
1. Discover	Verified inventory and version; SBOM or equivalent evidence where applicable.
2. Determine scope	Entry paths, public exposure, privileges, dependencies, and asset owner.
3. Select response	KEV/EPSS/CVSS record, impact, exceptions, measure, and accountable owner.
4. Contain	Isolation/access/egress restriction or temporary rule, with effectiveness and side-effect tests.
5. Remediate	New image/package released, verifiable deployment, functional regression tests, and rollback readiness.
6. Verify	Running-version readback, reproducible negative test, telemetry, and closure record.

Controls That Reduce the Asymmetry

Asset and dependency inventory: assets, containers, runtimes, APIs, third-party components, service accounts, effective versions, and owners; incorporate SBOM and VEX where trustworthy data are available.

Architecture-based isolation: least privilege, separation of management and data planes, microsegmentation, private networks, and restricted egress; verify that secondary integrations do not reopen an exposure.

Prepared change management: representative environments, canary rollout, deployment automation, rollback, and functional tests; verify protection on the artifact actually running.

Exploitation telemetry: log activity useful for identifying attempts and confirming compromise; initiate forensic analysis when there are signs of intrusion before patching.

Intelligence-led prioritization: KEV, EPSS, accessible attack surface, process criticality, sensitive data, lateral movement capability, and operational dependency.

Third-party coordination: contractual response times, dependency notifications, maintenance of unsupported systems, and evidence of updates.

Four Decision Categories for Prioritization

Decision	Indicative triggers	Response
EMERGENCY	Confirmed exploitation with a reachable path to a critical asset; or critical remote RCE with credible activity signals.	Contain immediately, investigate historical exposure, patch, and verify.
EXPEDITED	High likelihood with actual exposure, essential dependency, or major impact.	Prioritized rollout with approved target time and tests.
SCHEDULED	Applicable finding subject to restrictions and compensating controls.	Planned change with expiration date, explicit acceptance, and monitoring.

Decision	Indicative triggers	Response
NOT APPLICABLE	Product absent, or attack condition ruled out with verifiable evidence.	Document exclusions; reopen if architecture or intelligence changes.

These status labels form a proposed management model, not a category defined by PCI DSS, ISO 27001, or CISA. Applicable regulations, contracts, and standards must be assessed separately according to jurisdiction, sector, and date.

Resilient Architecture: Decoupling CVE Growth from Damage

CVE counts belong to the ecosystem; effective exposure depends on each organization's architecture. NIST SP 800-207 shifts attention away from trust in network location toward resource protection through explicit authentication and authorization; NIST SP 800-160 Vol. 2 defines resilience as the ability to anticipate, withstand, recover from, and adapt even under compromise. [23–24]

Architecture hypothesis (H4). If two systems run the same vulnerable version, the system with restricted public paths, independent identities, least privilege, function-level segmentation, restricted egress, and segregated data should have fewer viable paths to critical assets and a lower potential blast radius. The claim is conditional: it requires negative tests and telemetry; no boundary is perfect. [23–24]

An Intrusion Chain Has Multiple Breakpoints

A vulnerability may describe a defect in a software package. Turning that defect into loss usually requires a chain: reaching the relevant function, triggering the flaw, obtaining a useful capability, moving laterally or accessing secrets, and completing a harmful action. Independent design controls should constrain different links. A WAF, authentication proxy, or private environment cannot be assumed to prevent exploitation originating inside an allowed process. [23, 24, 27]

Falsifiable Controls, Not Claims of Invulnerability

Layer	Architectural decision	Verifiable evidence
Exposure	Inventory services/dependencies; close public management paths.	Authorized external reconnaissance reveals no direct path to consoles or internal runtimes.
Identity	Independent workload identities, short-lived credentials, resource-level authorization.	Compromise of one service account cannot reach other functions or privileges.
Isolation	East–west segmentation, default-deny policies, tenant separation.	Negative tests across services, networks, and tenants; log denials and bypass attempts.
Execution	Unprivileged processes, sandboxing, read-only file systems, hardening.	Test RCE remains within the container and cannot access host or secrets.
Data	Minimization, tokenization and external custody where appropriate; controlled key use.	A peripheral service cannot read original data or expand authorization.
Egress	Explicit outbound destinations, protocols, and purposes; anomaly monitoring.	Attempted exfiltration to an unauthorized destination is denied.
Resilience	Immutable rollout, canary/blue-green, rollback, tested restoration.	Exercise permits containment, update, and recovery without loss of integrity.

Test Scenario: RCE in a Public Service

Flat architecture: remote code execution in a web component that shares broad credentials and network connectivity with databases can lead to lateral access and exposure across customers. Segmented architecture: the same RCE may compromise the initial process, while scoped identities, isolation, and separate custody limit what it can reach. This is a conceptual comparison, not a test performed on a particular organization.

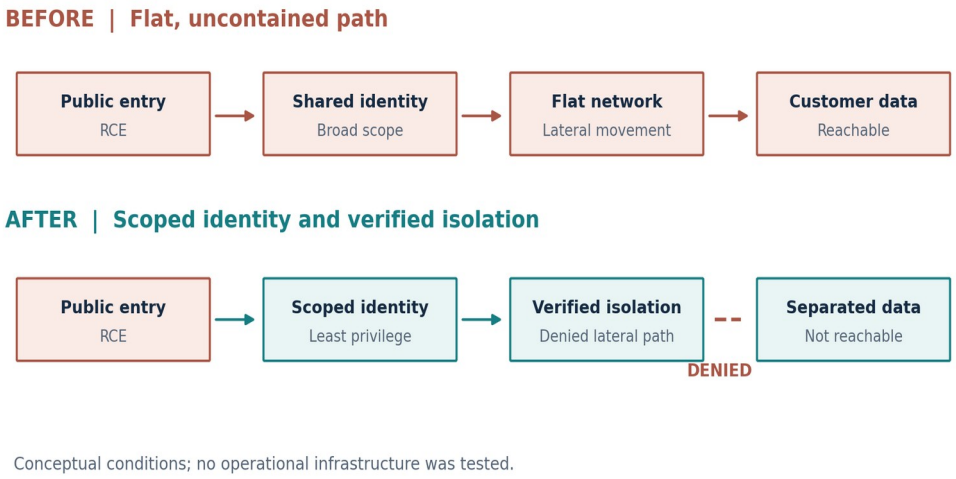


Figure 5. Two conceptual topologies exposed to the same software defect. Architecture does not eliminate the RCE; it seeks to limit its observable consequences.

Measuring Effectiveness Without Invented Percentage Reductions

For each asset, we propose recording: (a) confirmed vulnerable version; (b) effective reachability; (c) privileges usable after compromise; (d) accessible lateral destinations; (e) accessible data; (f) egress paths; and (g) time to verifiable protection. Before/after comparisons require repeatable scenarios. For example, it would be inappropriate to claim that segmentation reduces risk by 80% without an empirical basis for exploitability, access, and impact.

Suggested indicators include coverage of negative tests across critical boundaries; percentage of publicly exposed assets without direct administrative access; authorized lateral paths per workload; services sharing identities; asset-hours of effective exposure weighted by criticality; and p90 time to containment and closure. A non-applicable CVE should be documented rather than removed from the inventory.

Limitations. Compromise of an identity plane, exfiltration through allowed channels, misconfigurations, and third-party trust relationships may defeat apparently strong defenses. Cryptography and tokenization protect data within their scope; they do not correct memory flaws or replace authorization, segregation, or patching. [23–24, 27]

Four Measurable Resilience Boundaries

Architectural specialization is expressed through four boundaries, each subject to a fail-safe test. The objective is not to proclaim a system invulnerable, but to show which resource remains inaccessible after the preceding component is assumed compromised. [23, 24, 31]

Boundary	Design	Negative test
Entry	Minimal public access and administrative isolation	RCE in a peripheral service does not open consoles or internal planes.
Identity	Workload-specific privilege, expiry, resource-level authorization	A stolen token cannot escalate privileges or cross customer boundaries.

Boundary	Design	Negative test
Data	Separate custody, minimization, usage controls	A compromised web service cannot retrieve originals beyond its permitted need.
Egress and recovery	Explicit egress, rollback, evidence-backed restoration	Unauthorized exfiltration blocked; changes reversible and verified.

Tests should include valid but insufficient service credentials, denied lateral paths and cross-tenant access, unauthorized egress attempts, and deployment failures. The existence of a WAF, a segmentation diagram, or a token does not, by itself, demonstrate that these boundaries work. [23–24, 31]

07 Implications for Financial Services and Critical Providers

In payments, banking, and platforms handling personal information, a vulnerability is not significant solely because it permits entry to a server. The question is whether it crosses trust boundaries, reaches sensitive information, compromises service identities, or extends to third parties with standing permissions. A sound technical review must reconstruct complete paths rather than merely check whether a CVE appears in an inventory.

- Public-facing plane: API gateways, portals, reverse proxies, runtimes, and exposed libraries; distinguish application-level exploitation from compromise of the execution environment.
- Identity plane: short-lived credentials, per-transaction authorization, phishing-resistant MFA where applicable, service permissions, and segregation of duties.
- Data plane: limit PAN/PII and sensitive-data exposure, reduce copies, employ tokenization where appropriate, and test read and egress restrictions.
- Provider plane: document critical libraries, update responsibilities, disclosure agreements, and trust boundaries in integrations.
- Resilience plane: negative isolation tests, operational rollback, tested restoration, retesting, and analysis of exploitation records before and after closure.

Compliance observation. Demonstrating conformity with a standard is not proof of immunity from future CVEs; likewise, a scanner report does not replace verification of scope, controls, and evidence. Business continuity requires vulnerability management to be linked to architecture and operations rather than treated as isolated disciplines.

Governance Questions That Deserve Documented Answers

- How many exposed services currently lack an identified owner, runtime version, or dependency inventory?
- Which newly disclosed CVEs affect our assets, and which have a viable attack path and sufficient privileges to cause impact?
- What are the median and 90th-percentile intervals from alert to isolation, effective deployment, and technically verified closure?
- What happens if an essential provider has a critical vulnerability and cannot deliver a patch for several days?
- What evidence shows that boundaries between customers, services, and networks survive compromise of a peripheral component?

08 Existing Vulnerability, Remediation, and Mitigation Timelines

The regulatory question is not whether every framework permits a 30-day delay; they do not. Some set maximum time limits by criticality, others prescribe process frequencies or asset-specific conditions.

Confusing a scan frequency with a remediation deadline, or guidance with a binding duty, distorts the analysis. [11, 19–22]

Comparison of Existing Requirements and Guidance

Framework	Documented timeframe or frequency	Correct interpretation
PCI DSS 4.0.1	Req. 6.3.3: critical-vulnerability patches/updates within one month after release; others under entity-defined risk.	Applies to in-scope PCI systems; not permission to wait the entire month. [19]
CISA BOD 26-04	Conditional maxima of 3, 14, or 60 calendar days, or the next major update; a 3-day tier may require forensic triage. Four asset-level factors.	Binding on covered federal agencies; full implementation phase from 7 December 2026. [11, 20]
CIS Controls v8.1	7.2: review remediation strategy monthly or more often; 7.3/7.4: automate patching monthly or more often; 7.7: remediate monthly or more often.	Recommended process frequencies, not a universal grace period for every CVE. [21]
DORA / RTS 2024/1774	Art. 10: scan assets supporting critical/important functions at least weekly; requires emergency procedures and defined installation/escalation timelines.	That article does not set a universal 3-, 7-, or 30-day SLA for all flaws. [22]
NIST SP 800-40 Rev. 4	Patching as preventive maintenance: identify, prioritize, acquire, install, verify; enterprise risk-based strategy.	Technical guidance without a universal per-CVE deadline. [28]

Time-origin clarification. For critical vulnerabilities, the PCI deadline runs from publication of the applicable patch/update. Under BOD 26-04, the clock starts at the earlier of two regulatory triggers: addition to KEV or identification and reporting of the CVE on a covered federal asset; timeframes may tighten as exposure or intelligence changes. [19, 20]

How Frameworks Could Evolve in the Age of AI-Assisted Discovery

Research proposal, not an existing requirement: complement severity-based deadlines with targets informed by observed exploitation, exploit automation, public exposure, privileges, and reachability of essential data or services. Reachability evidence should be required and auditable under any future model adopting this principle.

Second, introduce two distinct clocks: one measures the time from credible notice to effective protection, and the other measures time to verified patching or elimination. Meeting a temporary containment target must not close the case while the underlying flaw persists.

Third, require periodic tests of segmentation, tenant isolation, identity integrity, and egress restrictions rather than treating configuration alone as protection. Fourth, establish emergency change paths, rollback, and verifiable obligations for critical suppliers. [23–24, 27]

Frameworks could also require target times to change when facts change: asset exposure, confirmed exploitation, exploit automation, or consequences for an identity plane. CISA BOD 26-04 already provides a verifiable example of contextual prioritization. The present proposal goes further by calling for explicit containment objectives measured in hours, negative architectural tests, and residual-risk reporting. This is a ZTC Research proposal, not a current obligation. [20, 32]

09 CCA-01: Containment in Hours and Verifiable Closure in Days

The targets below constitute a voluntary, more demanding profile for business-relevant assets. They do not replace applicable laws, contracts, or PCI/CISA requirements. A stricter statutory deadline always prevails. An accepted compensating control must be tested, reversible, and time-limited; it does not automatically constitute permanent remediation.

Proposed Contextual Risk Thresholds

Tier	Contextual trigger	Proposed internal target
A / Emergency	Applicable active exploitation; reachable KEV; automatable critical remote RCE with major impact and external path.	Containment ≤ 4 h; immediate forensic evaluation; target closure ≤ 72 h where a safe fix exists.
B / High exposure	Demonstrated external path and plausible exploitation/working PoC/critical service.	Protection ≤ 24 h; target closure ≤ 7 days.
C / Exposed, no known exploitation	Applicable high/critical finding, public access, verified defenses, and no confirmed activity.	Protection ≤ 72 h; target remediation ≤ 14 days.
D / Bounded internal	Not exposed or reachable from untrusted zones; limited scope with evidence.	Review ≤ 7 days; target remediation ≤ 30 days.
E / Low relevance	Non-critical, no known exploitation, documented inventory/path, and no change signals.	Window set by risk assessment; continuous review, no closure by silence.

At level A, if no patch exists or immediate deployment would create a greater continuity risk, the initial decision may be to remove exposure, restrict use, or temporarily isolate the system with evidence. Exceptions require immediate escalation, continuity measures, and preservation of forensic evidence. The vulnerability must not be reported as remediated merely because its host has been isolated. [20, 23–24]

Auditable Metrics

MTP (time to protection) is the number of hours between credible detection and verification of a control that interrupts the observed attack path. MTC (time to closure) is the number of hours/days until the defect is demonstrably eliminated on the affected asset. EAEC (essential-asset exposure) is the sum of vulnerable asset-hours with relevant attack paths, weighted by business criticality on a documented internal scale. Boundary coverage is the proportion of passed negative tests across identity, network, data, tenant, and egress limits.

Operational indicators should be reported by severity tier and percentile (median, p90), with samples and exclusions. CVE counts cannot replace the inventory of affected instances: one CVE affecting hundreds of assets can cause more work and exposure than hundreds of inapplicable CVEs.

Cost and Continuity Constraints

Reducing exposure time must be funded alongside test and deployment automation, response capability, redundancy, and rollback. An aggressive patching policy without adequate testing can increase downtime and operational loss. Economic assessment should compare exposure asset-hours, control costs, intervention frequency, and continuity outcomes, without treating hypothetical avoided losses as realized savings.

Investment Priorities and Accountability

Recommended investment begins with trust boundaries that reduce potential harm, continues with automated updating and retesting, and culminates in sustainable observability and response. Each initiative needs an owner, total cost, implementation time, residual-risk statement, and evidence of effectiveness. Where loss estimates lack data, uncertainty should be reported: avoiding a possible incident is not, by itself, a realized financial saving.

Operational continuity should be validated through service-isolation, rollback, restoration, and integrity tests before imposing aggressive deployment deadlines. Execution must not transfer to operations a greater outage risk than the exposure it aims to reduce.

Three Clocks and One Closure Criterion

Clock 1—Recognition: begins at credible notification or verifiable detection and ends when applicability is recorded. Clock 2—Protection: ends only when a control interrupts the relevant attack path and its effect is demonstrated. Clock 3—Closure: ends when the definitive correction is deployed on affected instances, verified, and subjected to functional and security retesting.

For vulnerabilities exploited before disclosure, an organization's detection clock does not capture the full history of exposure. It should be complemented by retrospective log review, intrusion investigation, and retention-bound analysis. Administrative closure without evidence of effective protection or a review of prior compromise can create false confidence. [29–30]

10 CCA-01 Containment Architecture: Dual Boundaries and Verifiable Decisions

The proposal begins from a limited but operationally decisive premise: a vulnerability alone does not determine damage, because outcomes also depend on trust paths, identities, permissions, data flows, and essential functions connecting the compromise point to critical objectives. CCA-01 is neither a cryptographic algorithm nor a new segmentation technique nor a security guarantee. It is an architectural evidence contract intended to prevent favorable findings based on isolated controls.

Prior Work and Delimited Novelty

Attack graphs, reachability analysis, Zero Trust principles, and cyber-resilient systems engineering are established prior art. CCA-01 does not claim to have invented them. Its limited novelty is to integrate adversarial reachability, continuity of legitimate functions, and dependency completeness into one adjudication with explicit rules for acceptance, rejection, and abstention. The practical distinction can be tested in the model: isolation may reduce potential damage while simultaneously disabling a critical function. The protocol forces that contradiction to be identified before reaching a conclusion and connects the outcome with exposure management before permanent remediation. [23, 24, 33–35]

Damage Boundary: The Adversary-Capability Graph

Represent each scenario as a directed graph G_a of plausible adversary-capability transitions. Specify an initial node s, a set K of critical objectives, privilege rules, and transition conditions. An edge is not merely a normal call between services: it must represent a feasible transition under the assumed adversary. Any path from s to some $k \in K$ affects the scenario's containment result; an egress path to an unauthorized destination is recorded independently.

Continuity Boundary: The Legitimate-Function Graph

Model, in a separate graph G_m, the dependencies required to complete an explicitly defined legitimate mission. Isolation that eliminates all adversarial paths but disrupts settlement, customer service, payments, or essential recovery does not constitute combined success. Both the critical containment objectives and the mission that must remain viable must be fixed before the outcome is observed, to prevent selecting only favorable objectives after the fact.

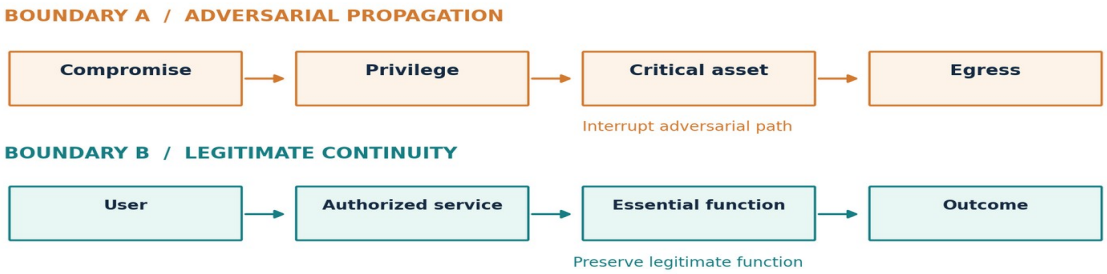


Figure 6. Two models linked by one decision. The paths are conceptual and do not depict existing systems.

A Transparent Damage-and-Continuity Vector

For unweighted desktop scenarios, we propose PDC = (Rc, E, M, C). Rc is the fraction of reachable critical objectives; E indicates an uncontrolled egress path; M denotes modeled completion of a legitimate mission; C reflects coverage of relevant dependencies. Rc is not an incident probability, CVSS severity, or monetary loss. The B index used in the E-01 exercise is weighted and distinct from Rc; its values are not interchangeable.

Variable	Definition	Permissible interpretation
Rc	$\lvert Reach(s,G_a) \cap K \rvert / \lvert K \rvert$	Fraction of reachable critical objectives under the assumed scenario.
E	0: no reachable egress path; 1: reachable	Not proof of observed exfiltration.
M	1: legitimate mission path preserved; 0: disrupted	Not a measurement of actual service availability.
C	1: dependency universe complete for scenario; 0: incomplete	Declared completeness is not an audited inventory.

Decision Rule and Abstention

The desktop exercise applies a strict demonstration threshold: PASS only if Rc = 0, E = 0, M = 1, and C = 1; FAIL if a known material breach is present; and NE (not estimable) when relationships necessary for adjudication are missing. The thresholds are local conventions of the synthetic model, not regulatory parameters or universal production requirements. A real assessment would need risk-based thresholds, data-quality rules, and mission tolerance criteria.

NON-DUPLICATION PRINCIPLE

Before recommending a new assessment, identify evidence and tests already available under PCI DSS, ISO, NIST, SOC 2, or equivalent programs. Reuse equivalent results. CCA-01 adds value only if it identifies an otherwise unaddressed path, dependency, or continuity conflict that changes a concrete decision.

Exposure, Protection, and Recovery Clocks

A future operational deployment requires four distinct timestamps: T0 (credible, applicable alert); TD (detection and declaration of compromise, if different); TP (sustained effective protection after intervention);

and TR (verified permanent remediation). Time to protection is MTP = TP − T0; time to closure/remediation is MTR = TR − T0. Merely recording a policy change does not demonstrate sustained protection; a vulnerability unresolved by the end of observation must not be counted as successful simply because it drops out of the sample.

When authorized data are available, reports should declare the time horizon and population, median and percentiles for MTP/MTR, censored observations, failures, continuity incidents, regressions, and boundary failures. In the present study, these metrics are definitions for a proposed evaluation protocol; any times expressed in minutes belong solely to the fictional E-01A exercise. No containment time has been measured on operational services, nor is the proposal intended to replace regulatory recovery objectives. [24, 28, 34, 41]

Falsification Conditions

CCA-01 should be judged limited or falsified in a specific application if a material residual path is absent from the model, PASS is reached using an incomplete inventory, an isolation policy disrupts the defined essential function, or the incremental evidence does not improve a decision already addressed by existing tests. Such outcomes must be treated as adverse evidence, never absorbed into a favorable aggregate average.

11 Synthetic Demonstration: Two Instruments, One Architectural Thesis

Two distinct synthetic series are documented to assess the consistency and rejection capability of the decision rules. Neither represents customers, payment systems, third-party networks, or ZTC platforms. The series must not be combined to calculate an effectiveness percentage. No penetration testing, field testing, or instrumentation of real environments was performed.

E-01A Instrument: Twelve Fictional Pairs of B-Index Values

In E-01A, ΔB = B(reference) − B(bounded). The baseline index and its post-control counterpart were deliberately constructed to test calculations and decision states. Mean ΔB is +0.3083 index points; this value does NOT mean 30.83% lower risk. Scenario acceptance also depends on continuity, egress, boundaries, and record completeness.

Case	Fictional scenario	B base	B bounded	ΔB	Decision
S01	Peripheral session	0.90	0.40	+0.50	PASS
S02	Cross-domain request	0.60	0.00	+0.60	PASS
S03	Broad permissions	1.00	0.30	+0.70	PASS
S04	Chained dependency	0.50	0.20	+0.30	PASS
S05	Unauthorized egress	0.40	0.00	+0.40	PASS
S06	Elevated identity	1.00	1.00	+0.00	FAIL
S07	Incomplete record	0.10	0.10	+0.00	NE
S08	Degraded continuity	0.30	0.10	+0.20	FAIL
S09	Expired permission	0.50	0.10	+0.40	PASS
S10	Lateral propagation	0.80	0.30	+0.50	PASS

Case	Fictional scenario	B base	B bounded	ΔB	Decision
S11	Concurrent change	0.50	0.90	-0.40	FAIL
S12	External dependency	0.70	0.20	+0.50	PASS

The rule rejects S06 because paths remain open; S08 because a lower B coincides with broken continuity; and S11 because reachability regresses. S07 remains NE owing to an incomplete record. Among the eight synthetic PASS scenarios, the stipulated MTP had a median of 40 minutes (range 16–192). These are exercise-defined times, not measured times for any product or service.

E-01B Instrument: Sixteen Dual-Boundary Models

E-01B models three critical objectives per case and separate conditions for egress, continuity, and inventory sufficiency. The sample includes flat architecture, partial controls, combined configurations, and adverse exceptions. Four cases pass, eleven fail, and one is not estimable. The results test evaluator adjudication on constructed inputs; they do not establish that any commercial architecture achieves these rates.

Case	Simulated condition	Reachable criticals	E	M	C	Status
S01	Flat trust	3/3	1	1	1	FAIL
S02	Egress filtering only	3/3	0	1	1	FAIL
S03	Administrative separation only	2/3	1	1	1	FAIL
S04	East–west block only	1/3	1	1	1	FAIL
S05	Combined containment	0/3	0	1	1	PASS
S06	Containment with auxiliary archive	0/3	0	1	1	PASS
S07	Reused service credential	2/3	0	1	1	FAIL
S08	Persistent emergency access	1/3	0	1	1	FAIL
S09	Cross-tenant data access	1/3	0	1	1	FAIL
S10	Third-party bridge	2/3	0	1	1	FAIL
S11	Containment disrupts service	0/3	0	0	1	FAIL
S12	Incomplete dependency inventory	0/3	0	1	0	NE
S13	Direct custody access	1/3	0	1	1	FAIL
S14	Isolation with operational audit	0/3	0	1	1	PASS

Case	Simulated condition	Reachable criticals	E	M	C	Status
S15	Redundant continuity control	0/3	0	1	1	PASS
S16	Billing exception	1/3	0	1	1	FAIL

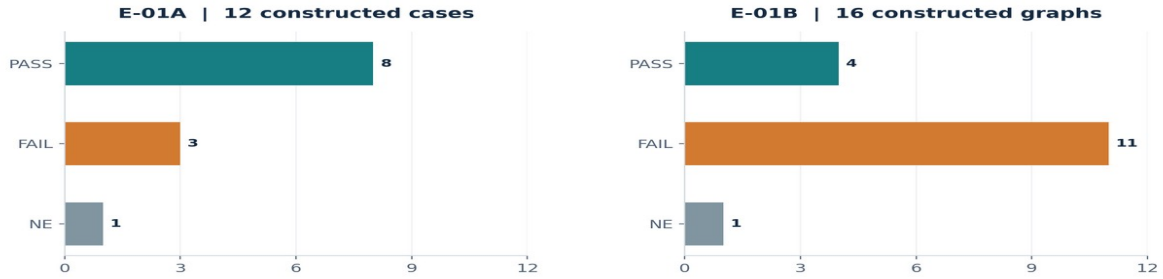


Figure 7. Synthetic decisions by instrument; denominators and rules differ. This is not a statistic of operational effectiveness.

Adverse Tests and Permissible Conclusions

Selected adverse cases include S07 (reused credentials, two reachable objectives), S08 (persistent emergency access), S10 (supplier bridge), S11 (containment that disrupts a mission), and S12 (unknown dependencies: NE). The evaluator does not verify whether the assumed conditions faithfully describe a real network. Its scope is the logical integrity of handling exceptions, boundaries, and missing evidence.

Reproducible evidence supports the assertion that the algorithm classifies these constructed cases under stated rules. It does not establish fewer attacks, lower financial loss, PCI DSS compliance, ISO/IEC 27001 conformity, or effectiveness on an organization's assets. Nor does it establish independent peer review.

12 Critical Crosswalk to Existing International Frameworks

The principle of architectural containment should engage with established controls, not caricature them. The recommendations below are a ZTC Research technical proposal, not an official interpretation or an adopted amendment by any standards body. Functional comparisons depend on the scope of each audit, jurisdiction, and framework version.

Organization and framework	Relevant existing coverage	CCA-01: proposed assessment contribution
NIST · CSF 2.0; SP 800-207; SP 800-160; SP 800-61r3	Risk management, Zero Trust architecture, cyber-resilient systems, response, and recovery.	Evidence sheet for residual critical paths, model uncertainty, and continuity after assumed compromise. [23,24,33–35]
ISO/IEC · 27001:2022; 27002:2022; 27005:2022	Information security management system, risk treatment, and control guidance.	Link damage boundaries, dependencies, exceptions, and safeguards to assets, owners, and accepted risks. [36–38]
PCI SSC · PCI DSS v4.0.1	Network controls, flaw remediation, incident response, and segmentation testing where applicable.	Supplement, not replace, tested CDE isolation evidence with cross-boundary paths and continuity beyond equivalent tested scope. [19,39–40]
CPMI-IOSCO / BIS · 2016 guidance and 2026 consultative toolkit	Market-infrastructure resilience, third-party dependency, and critical recovery.	Document propagation boundaries across dependencies and settlement functions without equating MTP with regulatory recovery targets. [41–43]
G7 Cyber Expert Group · 2025 AI statement	Dialogue and cooperation on AI threats to financial services; no new binding obligations.	Assess containment language as a reference for cross-border discussion and public–private cooperation. [44]

Organization and framework	Relevant existing coverage	CCA-01: proposed assessment contribution
ISO/IEC · 42001:2023	Governance and management of AI systems by providers and users.	Consider exposures and authorized flows when AI components participate in critical services. [45]
European Union · DORA and applicable RTS	Digital operational resilience and ICT risk-management requirements within regulated scope.	Use inventories and dependency analysis to evidence consequence limitation without altering legal obligations. [22]
AICPA / SOC 2; CIS Controls v8.1	Trust Services Criteria and vulnerability-management hygiene within applicable scope.	Reuse existing evidence where it covers the same outcome; avoid duplication across engagements and audits. [21]

PCI DSS: Complementarity, Not a Substitute for Mandatory Testing

PCI DSS v4.0.1 prescribes specific activities and, where segmentation is used to isolate the cardholder data environment (CDE), penetration tests of segmentation controls under the applicable requirements. A CCA-01 desktop review does NOT, by itself, discharge that obligation or replace an authorized exercise or a QSA's assessment. Its potential contribution is an additional matrix of residual consequences and continuity for relationships not already covered by equivalent evidence. [39–40]

PCI SSC's June 2026 guidance distinguishes compensating controls from the customized approach, preserves the respective responsibilities of entities and assessors, and emphasizes proper documentation of effectiveness. CCA-01 may inform documentation design or objective evaluation where relevant, but it is not itself a compliance mechanism, certification, or substitute for requirements. [40]

ISO/IEC: Integration with the ISMS Without Inventing a Control

ISO/IEC 27001:2022 sets information security management requirements; ISO/IEC 27002:2022 provides control guidance, and ISO/IEC 27005:2022 addresses risk management. CCA-01 proposes a way to describe consequences and dependencies already contemplated in a well-designed risk assessment, adding an explicit abstention rule for incomplete inventories and a cross-check between reduction in adversarial reach and continuity. No official control numbering or claim of ISO conformity is implied.

BIS/CPMI-IOSCO: Distinguishing Containment from Recovery

The 2016 cyber resilience guidance for financial market infrastructures includes an objective of safely resuming critical operations within two hours and completing settlement by the end of the day, subject to its conditions and context. That recovery objective is not a universal patching deadline and is not equivalent to CCA-01 time to protection (MTP). The CPMI-IOSCO consultation dated 8 September 2026 offers a contemporary forum for discussing complementary evidence tools and third-party dependencies; the instruments are presented as voluntary and nonbinding. [41–43]

13 Proposed Architectural Control: Reference Wording and Evidence

The following wording is a ZTC Research proposal that may be considered by technical working groups, operators, and framework developers. It is not official text from NIST, PCI SSC, ISO/IEC, or CPMI-IOSCO. Any incorporation would need to follow each organization's mandate and established procedures.

CCA-01 · PROPOSED CONTROL

For every essential function and applicable critical exposure, the organization shall identify relevant components, trust relationships, transitive permissions, ingress and egress paths, dependencies, and exceptions. Under an assumed initial compromise, it shall define the set of potentially reachable critical objectives and the legitimate functions that must remain available. It shall record intended architectural constraints, available evidence, and uncertainties that prevent a conclusion on the extent of damage. No configuration shall be deemed satisfactory solely because containment appears effective if it compromises a critical function or omits material dependencies.

Proportional Implementation Criteria

Dimension	Recommended minimum evidence	Indicative decision rule
Scope	Asset, service, identity, critical function, version, and model boundaries	Insufficient universe: NE, not PASS.
Adversarial paths	Assumed entry points, permissions, transitions, exceptions, and egress paths	Every material residual path requires explicit treatment.
Continuity	Function dependencies, service alternative, permitted degradation	Containment that destroys a function is not a combined success.
Time	T0, time to protection (MTP), time to closure (MTC), recovery time if applicable	Separate clocks; do not use one SLA as all-purpose evidence.
Governance	Owner, risk acceptance, exception expiry, reviews on change	Risks and exceptions must have owners and versions.
Comparability	Same assumed compromise, time horizon, and scope before/after	Do not compare incompatible or partial configurations.
Reuse	Current PCI/ISO/NIST/SOC 2 or other relevant tests and assessments	Do not repeat work where existing evidence already answers the question.

Metrics for Desktop Assessment and Future Operations

Four classes of metrics should be published and kept separate: (i) vulnerability exposure and applicability; (ii) the modeled fraction of reachable critical objectives and coverage of the model; (iii) continuity of legitimate functions and egress; and (iv) observable times to protection, closure, and recovery when authorized instrumentation exists. In a purely documentary assessment, real-world time and effectiveness remain unobserved. No hypothetical value should be transformed into a compliance claim.

Internal Policies: Update Clause

Information security and cyber-risk management policies should include recurring assessment of containment architecture for critical exposures, privileged roles, third-party integrations, and essential business processes. Review is warranted when dependencies, permissions, egress, topologies, or critical functions change, with material uncertainty escalated to risk governance. Implementation may initially use tabletop exercises and documentary evidence without active operations on platforms.

Minimum Evidence Record for Engagement with Standards Bodies

Field	Expected verifiable record
Identification	Scenario ID, protocol version, date, owner, and scope
Context	Essential function, initial identity/node, critical objectives, dependencies

Field	Expected verifiable record
Existing assurance	Applicable standards, prior tests, and documented exceptions
Damage model	Capability graph, egress, transitive privileges, supplier paths
Continuity model	Legitimate flow, acceptable degradation, relevant recovery objectives
Decision and limitations	PASS, FAIL, or NE with rationale; uncertainty and potential falsifiers
Governance	Accepted risk, reassessment date, owner, and reason for decision

This format is proposed as a starting point for technical cooperation, based on proportional evidence and reuse of existing material. It is not a new official audit template and does not modify the accreditation procedures of a QSA, an ISO certification body, or a supervisory authority.

14 International Invitation to Evaluate Containment Architecture

ZTC Research proposes a technical discussion on proportionate incorporation of architectural containment evidence into cyber-risk management. The study invites framework and guidance developers, supervisory authorities, and organizations responsible for critical services to examine whether CCA-01 better characterizes residual damage and legitimate continuity while a vulnerability remains unresolved. It does not call for an immediate binding requirement or the replacement of established controls; the proposal remains open to falsification, adaptation, and independent validation.

For Technical Framework and Standards Developers

Within their respective mandates, NIST, ISO/IEC JTC 1/SC 27, the PCI Security Standards Council, and other specialized groups may find it useful to examine a matrix connecting assumed compromise, critical objectives, residual paths, egress, continuity, uncertainty, and accountable owners. This is a proposed supplementary evidence pattern for discussion in profiles, examples, technical guidance, or future revisions through ordinary consultation procedures. CCA-01 is not a control approved by these organizations and does not replace regulatory testing, PCI DSS requirements, or ISO/IEC conformity assessments.

For Financial Resilience Dialogue and Supervision

The G7 Cyber Expert Group’s 2025 statement recognizes AI-related cyber risks and encourages framework updates and cross-sector dialogue without establishing regulatory obligations. Separately, CPMI-IOSCO opened a consultation in September 2026 on voluntary resilience tools for financial market infrastructures, published through the BIS. Within such forums, CCA-01 could be considered as a methodological input, particularly for interdependencies, residual propagation, and the preservation of critical operations. Central banks and supervisors retain full discretion within their respective powers; this study is neither a formal submission nor an institutional position or endorsement by any of them. [42–44]

For Financial Institutions, Operators, and Critical Providers

Organizations can assess whether their internal policies on architecture, segregation, vulnerability management, incident response, continuity, and third-party risk document potential damage after an assumed compromise. Exceptions, uncertainty, owners, and remediation deadlines should be preserved while reusing evidence already created under applicable frameworks. In its current form, CCA-01 is voluntary and methodological; it does not replace penetration testing, audits, certification, regulatory assessment, or contractual requirements.

For Researchers and Security Practitioners

The proposal is open to critical review, reproduction, and technical falsification. Its incremental value should be compared with existing audit and framework outcomes, without counting duplicated tests as innovation. An adverse result, an incomplete inventory, or a continuity conflict is useful evidence; concealing it would undermine the research. The published desktop data permit scrutiny of evaluator logic without exposing third-party infrastructure or information.

FORMAL INVITATION TO EVALUATED ADOPTION

ZTC Research proposes that standards organizations, authorities, and operators consider verifiable containment architecture CCA-01 as a supplementary reference when reviewing frameworks, internal policies, and assurance practices in response to AI-accelerated threats. Any adoption should follow consultation, appropriate testing, proportionate risk treatment, and preservation of existing duties. No organization cited in this study has approved, certified, or adopted the proposal merely by being mentioned.

15 Objections, Alternatives, and Limits of Inference

Objection 1: More CVEs do not necessarily mean worse software. Correct. Broader audit coverage may expose longstanding defects and create more records without a proportional increase in newly introduced failures. Better discovery can benefit defense; pressure then shifts to remediation processes. [03–05]

Objection 2: AI can also improve the defender's advantage. Correct in several documented cases. The coordinated PHP response and Big Sleep research illustrate how discovering, disclosing, and fixing flaws before exploitation can lengthen the defender's window. The benefit depends on updates reaching actual environments. [10, 13, 17]

Objection 3: Scanners and vendors may overstate threats. Findings, CVEs, affected instances, working exploits, attempted attacks, and confirmed compromises are distinct categories. NOVA's counts and AI-attributed samples must not be added to worldwide totals as independent new vulnerabilities. [06–08]

Objection 4: Eighteen months is an arbitrary prediction. It is an opinion-based horizon useful for resilience exercises, not a basis for declaring a date for crisis or global recovery. Claiming that an Internet collapse has been forecast would be methodologically unjustified. [15–16]

Objection 5: The observed exploitation share is small. Yes, but a small fraction of flaws in very widely deployed systems may concentrate substantial harm. Conversely, thousands of inapplicable CVEs may require no operational work at a given company. What matters is the combination of criticality, exposure, and response speed. [03, 18]

16 Decision Roadmap and Continuous Improvement

Phase	Objective	Evidence of outcome
0–4 hours	Assign an on-call team; identify critical public-facing assets; activate emergency containment.	Recorded decision, owners, and demonstrable isolation path.
4–24 hours	Validate applicable CVEs and attack paths; apply temporary protection; monitor activity.	Scope register and evidence of blocks/earlier compromise.
24–72 hours	Safely deploy urgent fixes with retesting, canary, and rollback; review lateral access.	Effective versions, negative tests, and closure of highest-tier cases.

Phase	Objective	Evidence of outcome
Days 4–7	Extend least privilege, dependency inventory, and egress restrictions across essential systems.	Path/owner matrix and isolation tests across services/tenants.
Days 8–14	Run controlled exploitation and recovery exercises; set risk-based targets.	Verification coverage, protection p90, expiring exceptions.
Days 15–30	Automate update-and-test chain and response contracts with suppliers.	Incoming/closed case balance and evidence of reduced exposure.
Continuous: months 1–18	Maintain trust-boundary redesign, measure trends, and validate independently.	Contained compromise scenarios and recurring resilience evidence.

Short-cycle governance. During the first fourteen days, review daily any alerts affecting essential assets; thereafter, maintain continuous oversight with weekly reporting of exceptions and trends. Management should monitor at least four figures: uncovered critical assets; relevant high-impact findings still exposed; p90 time to effective protection; and closure capacity versus incoming findings. Defensive AI investment succeeds only if it reduces exposure and improves closure, not merely if it increases alert volume.

17 Scope, Uncertainty, and Future Validation

The scope of this research is documentary, comparative, and architectural. Its conclusions distinguish observed data, reasoned inferences, and proposals that still require experimental validation. Making these boundaries explicit allows others to reproduce the findings, identify where evidence ends, and specify what outcomes would require the hypotheses to be revised.

Methodological Limitations and Validity Measures

Dimension	Limitation	Validity measure
CVE and KEV series	Repositories vary in ingestion, dates, retrospective edits, and inclusion rules. Disclosure is not discovery.	Compare within each source using cut-off dates; do not create a single global total or artificial KEV/CVE exploitation rate. [01–05]
Attribution to AI	No universally verifiable field records how each CVE was discovered. Vendor-released samples may be selection-biased.	Separate documented cases from aggregate series; do not estimate a worldwide causal percentage attributable to AI. [06–10]
Real exploitation	CISA KEV and security-firm telemetry capture different populations; no observation does not mean no attacks.	Distinguish observed exploitation, available exploit, attempts, and confirmed compromise; state denominators. [02–03]
Remediation deadlines	Requirements/frequencies apply to different populations; field durations often use noncomparable clocks, assets, and samples.	Distinguish recognition, protection, and closure; do not confuse observed medians with universal regulatory deadlines. [19–22,29]
Architectural effectiveness	Trust-boundary model follows technical principles; no active tests or controlled comparisons were performed on client infrastructure.	Present mechanisms and proposed negative tests; do not claim incident reductions or immunity. [23–24,31]

Dimension	Limitation	Validity measure
Sectoral and economic scope	International sources are not a representative Mexico/LATAM sample; no per-entity costs or impacts are measured.	Avoid extrapolating sector rates and financial returns; propose local exposure, continuity, and cost metrics.
Time horizon and currency	The 18-month warning is not a statistical forecast; records and AI capabilities may change after the 2026 cutoff.	Use the horizon as a stress scenario, not a fixed deadline; reconfirm figures in future editions. [15–16]

What the Results Support

The evidence supports three statements: disclosures increased sharply in the examined series; relevant discoveries have been made with AI assistance; and architectural decisions can constrain attack paths through controls whose effectiveness is testable. The magnitude of any benefit depends on assets, configuration, adversaries, and the continuity of controls.

The evidence does not support claims that AI alone explains worldwide CVE growth, that every record has been exploited, that all organizations face identical deadlines, that the eighteen-month horizon is a quantified forecast, or that a particular topology produces a universal percentage reduction in risk. Proposed scenarios and targets are operational hypotheses subject to testing, not empirical results from a deployment.

Additional Evidence Needed

Temporal reproducibility: publish dated snapshots and queries for each series, deduplication rules, and a record of subsequent corrections.

Independent attribution: document AI's specific role in each finding, the human review, proof of concept, and disclosure timeline.

Architectural validation: in an authorized laboratory using synthetic data, compare the same vulnerable component under different designs; measure reachability, escalation, cross-tenant access, and egress.

Longitudinal monitoring: record the three clocks by asset (recognition, protection, and closure), percentiles, containment failures, availability effects, and verifiable costs.

Review and Possible Falsification Criteria

The architectural hypothesis must be limited or reformulated if controlled testing shows that compromise of a component can cross declared boundaries, reach unauthorized data, or maintain improper egress. Remediation targets also need revisiting if containment within hours cannot be verified or creates operational damage exceeding the risk it mitigates. Serious research keeps such falsification conditions open and documented.

Hypothesis Status and Next Verification Criterion

Hypothesis	Evidence status	Decisive test still needed
H1 · CVE	Growth observed in the examined series.	Reconstruct repository snapshots and identify registration-change effects. [01–05]
H2 · AI	Documented real cases; global causality not quantified.	Build a cohort with discovery-method attribution and independent comparison. [06–10]
H3 · Response	Plausible mismatch; magnitude depends on environment.	Measure arrivals, permanent closures, time to protection, and backlog per essential asset. [29]

Hypothesis	Evidence status	Decisive test still needed
H4 · Architecture	Technically grounded mechanism; empirical effect not measured here.	Test the same flaw in two designs and document blocked and persistent paths. [23–24,31]

18 Conclusions and Strategic Implications

Available evidence documents a major change in vulnerability disclosure volume and concrete cases of AI-assisted automation. The precise aggregate causal effect of AI cannot be estimated universally from the public sources used; each statistic must retain its unit of observation and time window.

Architecture does not replace patching. An organization may need days to correct a defect while critical exposure demands an immediate decision about potential harm. The relevant outcome is not a promise of invulnerability, but a boundary around consequences under transparent assumptions, documented evidence, and continuous updating.

CCA-01 contributes a way to turn architectural containment into a combined decision on residual damage, legitimate continuity, and sufficiency of evidence. The first synthetic dataset verifies rules of the B index; the second subjects the dual criterion to favorable, adverse, and non-estimable cases. These results show that declared segmentation, a favorable average, or an incomplete topology cannot justify approving an architecture by themselves. The demonstration is methodological; operational effectiveness requires different evidence.

The proposed governance agenda is concrete: (1) describe compromise paths and critical objectives; (2) declare boundaries and exceptions; (3) verify residual consequences using appropriate evidence; (4) preserve critical functions and sector-specific recovery objectives; (5) integrate the finding into policies, processes, and accountability; and (6) require traceability when adopting or amending relevant frameworks.

Framework developers, researchers, financial institutions, and relevant authorities are invited to consider the incremental value of CCA-01. G7 Cyber Expert Group and CPMI-IOSCO discussions provide context, not evidence of acceptance, sponsorship, or validation. The proposal remains open to refinement and falsification; synthetic exercises do not establish operational effectiveness.

MAIN CONCLUSION

The question that deserves a place in security policies and frameworks is verifiable: “If a component were compromised today, which critical objectives could the adversary reach, and which business functions would continue operating safely before definitive remediation?” Architecture should answer with evidence and explicit uncertainty.

A. Data and Series Used in the Figures

Month	CVE 2025	CVE 2026	YoY change	KEV 2025	KEV 2026
Jan	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
Apr	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

Month	CVE 2025	CVE 2026	YoY change	KEV 2025	KEV 2026
Jul	3,792	9,770	+157.6%	20	26
Aug	3,630	12,256	+237.6%	15	31
Sep	4,324	14,933	+245.4%	16	43

The CVE series comes from CVETodo [01] and the KEV series from CVE Tools [02]. They are displayed together for visual comparison but are not divided to imply an exploitation-risk ratio. Published counts may change because of revisions, late submissions, and synchronization differences. Disclosure and KEV addition dates do not necessarily coincide.

B. Glossary of Variables and Terms

Term	Working definition
CVE	Public vulnerability identifier; not an incident or proof that a specific asset is exposed.
CWE	Software weakness category grouping classes of flaws.
KEV	Catalog of vulnerabilities with confirmed exploitation in the wild under CISA criteria.
EPSS	Statistical estimate of exploitation probability over the next 30 days; not a measure of business impact.
CVD	Coordinated vulnerability disclosure, allowing time for fixes and protection before mass exposure.
RCE	Remote code execution; a potential outcome whose context, preconditions, and privileges must be examined.
SBOM / VEX	Software component inventory / evidence-backed affected-or-not-affected statement.
Verified closure	Technical demonstration that a measure reduces or eliminates exposure on the actually deployed system.

C. Traceability, Reproducibility, and Falsifiability

E-01A was published as an exercise of the generic evaluator, documenting 12 constructed cases and B values, with 8 PASS, 3 FAIL, and 1 NE. E-01B contains 16 synthetic scenarios evaluated using Python 3 in a local environment, with 4 PASS, 11 FAIL, and 1 NE. Reproduction files should accompany the study; their hashes and versions are recorded outside the public PDF in the technical release dossier.

Falsifiability requires consideration of counterexamples: hidden paths, reused credentials, emergency exceptions, third-party dependencies, concurrent changes, mission failure, and missing inventories. Desktop assessment does not measure control latency, real compromise, or implementation reliability. For any subsequent external study, criteria should be registered before outcomes are observed and compared against evidence already available to quantify incremental value.

Evidence type	Admissible in this study	Cannot be inferred
Primary document	Source versions, dates, assertions, and stated scope	Universal causality or compliance of unevaluated organizations
Synthetic graph	Logical rule and output on fictitious nodes	Security, segmentation, or recovery of a real platform

Evidence type	Admissible in this study	Cannot be inferred
Local unit test	Limited integrity of simulator functions	Field validation or academic endorsement
Framework crosswalk	Topic alignment and complementary proposal	New binding requirement or NIST/ISO/PCI/BIS approval
Architecture hypothesis	Potential consequence testable with appropriate information	Statistical risk reduction or monetary effect

D. Replication Path and Incremental Value

A subsequent study, entirely voluntary and separate from this publication, could compare two documented configurations under paired scenarios, fixing compromise conditions, dependencies, permissions, controls, and essential functions before the analysis. Without authorization for active assessment, the same design can be applied to documentary graphs and consistency tests; that mode validates the model, not the infrastructure.

A rigorous methodological comparison should identify, for each control: evidence already required or generated by PCI DSS, ISO/IEC, NIST, or equivalent programs; the material question still unanswered; the additional datum requested by CCA-01; the cost of obtaining it; and the security decision actually changed. If no additional decision results, reuse of existing evidence takes priority over new testing.

Independent desktop replication requires preservation of the scenario codebook, boundary definitions, fixed node and edge sets, adjudication rules, evaluator software, adverse cases, and outputs. Supporting documentation may include a change log, rules for NE, and comparisons with prior attack-graph techniques. Published synthetic executions allow the logic to be verified; no human or peer review is claimed unless documented.

E. Institutional Nature and Declaration of Interests

ZTC Research is the research initiative of Zero Trust Consulting (ZTC), a private consulting firm specializing in architecture, cybersecurity, cryptography, and compliance. The relationship between the firm’s professional activities and the subject of this study is disclosed for institutional transparency. Such a relationship does not, by itself, establish a conflict of interest or invalidate the method. The study’s methodological independence should be assessed from its analytical rules, source traceability, separation of facts from inferences, and disclosure of favorable and adverse results.

CCA-01 may be relevant to architecture, risk, and resilience advisory services, including those provided by ZTC. This potential convergence of professional interests is disclosed expressly, without equating it with improper influence or an externally independent assessment. The study publishes assumptions, criteria, and synthetic scenarios producing favorable, failed, and non-estimable outcomes. Inferences are limited to documentary evidence and the constructed model, and therefore do not demonstrate operational control effectiveness or adoption by third parties.

References and Sources

References for the principal research [01–32] and framework crosswalk sources [33–45]. Citations to organizations document their existing frameworks and do not imply acceptance of CCA-01. Documentary update: 2 November 2026.

[01] CVETodo — CVE Statistics 2026; 2025–2026 monthly series. Accessed 31 October 2026. | Link

[02] CVE Tools — Monthly CVE trends; snapshots and additions to CISA KEV. Accessed 31 October 2026. | Link

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

[04] FIRST — 2026 Vulnerability Forecast, 11 February 2026. | Link
[05] FIRST — Mid-Year Vulnerability Forecast, 15 June 2026. | Link
[06] IBM X-Force — AI can find serious vulnerabilities at unprecedented speed. Can defenders keep up?, October 2026. | Link
[07] Palo Alto Networks Unit 42 — The Frontier AI Vulnerability Burst: Industrializing Autonomous Zero-Day Discovery, 4 August 2026. | Link
[08] Securin — AI-Discovered CVE Tracker; attribution database with evidence grades, 27 October 2026. | Link
[09] PHP Group — PHP2Shell security advisory and patched versions, 21 and 23 October 2026. | Link
[10] OpenAI — Patch the Planet: global response coordination as agents uncover critical vulnerabilities, 21 October 2026. | Link
[11] CISA — BOD 26-04: Prioritizing Security Updates Based on Risk, June 2026. | Link
[12] FIRST — EPSS Frequently Asked Questions: scope, use cases, and limitations. | Link
[13] Google Project Zero — From Naptime to Big Sleep, 1 November 2024. | Link
[14] Jon Hernández — La IA está rompiendo la seguridad de Internet [AI Is Breaking Internet Security], 26 October 2026. | Link
[15] Jason D. Clinton — X post, 22 October 2026; content corroborated through third-party reproduction. | Link
[16] Mantic — Black October and the future of cybersecurity, 27 October 2026. | Link
[17] Google — Big Sleep and the discovery of CVE-2025-6965, 2025. | Link
[18] CISA — Vulnerability Review: baseline controls, prioritization, and exploitation, 2026. | Link
[19] PCI Security Standards Council — FAQ 1597, distinguishing requirements 6.3.3 and 11.3; updated May 2025. | Link
[20] CISA — Official BOD 26-04 Directive, 10 June 2026; conditional deadlines. | Link
[21] Center for Internet Security — CIS Controls v8.1, Control 7, process and remediation frequencies. | Link
[22] European Commission — Delegated Regulation (EU) 2024/1774, Article 10, DORA. | Link
[23] NIST — SP 800-207, Zero Trust Architecture (2020). | Link
[24] NIST — SP 800-160 Vol. 2 Rev. 1, cyber resiliency engineering (December 2021). | Link
[25] RSAC Conference — Jason Clinton, professional biography and verified career history. | Link
[26] Clarín / DEF CON 34 — Jason Clinton statement on an 18-to-24-month period (August 2026). | Link
[27] Microsoft Learn — Zero Trust, trust boundaries, and defense in depth; architecture guidance. | Link
[28] NIST — SP 800-40 Rev. 4, patch management planning (April 2022). | Link
[29] Verizon — 2026 Data Breach Investigations Report; KEV coverage and median closure time (2025 data). | Link
[30] Google Cloud / Mandiant — M-Trends 2026; estimated time to exploitation of −7 days in the sample. | Link
[31] NIST — SP 800-207A, service-level identity and access controls in cloud-native environments. | Link
[32] FedRAMP — Institutional response to CISA BOD 26-04 and implementation timeline. | Link
[33] NIST. The NIST Cybersecurity Framework (CSF) 2.0. CSWP 29 (2024). | Link
[34] NIST. SP 800-61 Revision 3, Incident Response Recommendations (2025). | Link
[35] NIST. SP 800-207, Zero Trust Architecture (2020). | Link
[36] ISO/IEC. 27001:2022, Information Security Management Systems — Requirements. | Link
[37] ISO/IEC. 27002:2022, Information security controls. | Link
[38] ISO/IEC. 27005:2022, Guidance on managing information security risks. | Link
[39] PCI SSC. Document Library: PCI DSS v4.0.1 (June 2024). | Link
[40] PCI SSC. Guidance for Compensating Controls and the Customized Approach (10 June 2026). | Link
[41] CPMI-IOSCO / BIS. Guidance on cyber resilience for financial market infrastructures (2016). | Link
[42] CPMI-IOSCO / BIS. Cyber resilience toolkit: practical considerations for FMIs, consultation (8 September 2026). | Link
[43] CPMI-IOSCO / BIS. FMIs reliance on third-party service providers: discussion paper (8 September 2026). | Link
[44] G7 Cyber Expert Group. Statement on AI and Cybersecurity (September 2025; published October 2025). | Link
[45] ISO/IEC. 42001:2023, Artificial Intelligence Management System. | Link

Suggested Citation

ZTC Research / Zero Trust Consulting (2026). The New Asymmetry of Cybersecurity: Containment Architecture as a Verifiable Bound on Harm in the Age of Artificial Intelligence. ZTCR-AI-VM-2026-01, Version 1.0, 2 November 2026. Mexico.

© 2026 Zero Trust Consulting. Quotation and citation are permitted with attribution to the study and preservation of its methodological scope. Synthetic scenarios must not be presented as evidence of operational protection, certification, or third-party endorsement.