SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Autonomous Vulnerability Remediation As A Foundation For Zero-Trust Architecture In Enterprise Data Center Ecosystems

Satish Chandra Guruvelli

Independent Researcher, USA. ORCIDs: 0009-0000-1298-237X

Abstract
Zero-trust architecture (ZTA) has become the dominant security reference model for enterprise data center ecosystems, yet its practical implementations are structurally incomplete. The framework's "never trust, always verify" posture governs access to resources with considerable rigor but leaves unaddressed whether those resources are actually free of exploitable vulnerabilities. The National Vulnerability Database (NVD) now publishes well above 29,000 Common Vulnerabilities and Exposures (CVEs) annually, a volume that manual remediation workflows cannot absorb, and Ponemon Institute data place the industry mean time to remediate (MTTR) for critical vulnerabilities at approximately 60 days. This paper argues that zero-trust must be extended from access verification to autonomous vulnerability elimination, and it develops a five-level zero-trust maturity model that places autonomous remediation at Level 5. A five-stage remediation pipeline, discovery, prioritization, patch identification, staged deployment, and verification, is combined with a confidence-gated action policy that partitions patches by model-derived probability of success. Under industry-reported adoption of this architecture, MTTR falls by approximately 81 percent to near 2.7 days, the CVE backlog contracts by approximately 78 percent, autonomous remediation reaches the 83 percent share, and the enterprise zero-trust posture score rises from a baseline near 42 to near 87 on a hundred-point scale. The article positions autonomous remediation as the infrastructure layer that makes zero-trust operationally complete rather than aspirational.

Keywords: Zero-Trust Architecture, Autonomous Vulnerability Remediation, Common Vulnerabilities and Exposures, Security Automation, Continuous Security Posture, Microsegmentation, NIST Cybersecurity Framework, DevSecOps, Confidence-Gated Remediation, Enterprise Data Center Security.

1. Introduction

Zero-trust architecture (ZTA) has moved in under a decade from a contrarian design proposal into the de facto security reference model for enterprise data center ecosystems. Its seven tenets, codified in the National Institute of Standards and Technology (NIST) Special Publication 800-207, push the security perimeter inward until it surrounds the individual resource request rather than the network as a whole [1]. The adoption story, however, is uneven. Industry surveys report zero-trust adoption rates of roughly 43 percent with another 46 percent in active transition, yet only about 1 percent of organizations report full zero-trust maturity that includes automated remediation [2] [3]. That gap between adoption and maturity is not cosmetic; it indicates a structural limitation in how the framework has been operationalized.

The limitation is straightforward. Zero-trust excels at the verification half of security, identity, authorization, per-session trust evaluation, and microsegmentation but largely inherits the vulnerability half from traditional patch management: scanner-driven, ticket-dispatched, engineer-executed, and chronically behind. The NVD publishes more than 29,000 CVEs per year, a pace that, at the industry-typical mean-time-to-remediate (MTTR) of 60 days for critical vulnerabilities, produces a persistent backlog of exploitable conditions inside the very resources that zero-trust is verifying access to [4] [5]. A verified but unpatched resource is still exploitable, and the framework offers little guidance on closing that loop.

This paper develops an extension of zero-trust that treats autonomous vulnerability remediation as the missing maturity level. The contribution is threefold. First, we propose a five-level maturity model in which autonomous remediation is the Level-5 capability that follows traditional perimeter defense, identity-centric control, microsegmentation plus monitoring, and artificial intelligence (AI)-driven detection and response. Second, we describe a five-stage remediation pipeline and a confidence-gated action policy that makes autonomous patch deployment operationally trustworthy. Third, we summarize the empirical outcomes reported for enterprises operating near Level 5, roughly an 81 percent reduction in MTTR, a 78 percent reduction in CVE backlog, and an 83 percent autonomous remediation share, and relate those outcomes to the zero-trust posture score.

2. Research Context and Methodology

The research base for autonomous remediation is fragmented across three communities. The security-standards community produces the frameworks NIST SP 800-207 and the Cybersecurity and Infrastructure Security Agency (CISA) Zero Trust Maturity Model version 2.0 [1] [3]. The vulnerability-management community produces the prioritization signals: the Common Vulnerability Scoring System (CVSS) severity score and the Exploit Prediction Scoring System (EPSS) probabilistic forecast [5] [6]. The security-operations community produces the mechanics: Security Orchestration, Automation, and Response (SOAR) platforms, DevSecOps tooling, and increasingly agentic AI systems for patch triage and deployment [7] [8] [9]. This paper treats these as complementary inputs to a single architecture rather than independent technology stacks.

The methodology is synthetic and observational. Architectural claims are grounded in the peer-reviewed and standards literature. Quantitative claims are sourced from industry reports (Ponemon Institute, Gartner, Forrester, and CISA) and from enterprise case studies published between 2023 and 2026; they are presented as triangulated ranges rather than single point estimates because operational environments vary substantially. A deployment profile consistent with a globally distributed digital-infrastructure operator serving a large enterprise customer base is used to illustrate before-and-after outcomes; the specific organization is immaterial to the argument, which turns on the architecture.

The measurement frame centers on six indicators: the CVE backlog size, MTTR, autonomous remediation share, patch success rate, critical CVE service-level agreement (SLA) compliance, and the aggregate zero-trust posture score computed against a CISA-style maturity rubric. Secondary indicators include engineering capacity redirected from routine patch work to strategic security engineering and the frequency of rollback events. The indicators are chosen because they jointly surface both the security improvement and the operational cost of the architecture, which are frequently reported in isolation and thus misrepresent each other.

3. Zero-Trust Maturity Model and the Autonomous Remediation Gap

Zero-trust in practice occupies a bounded slice of the maturity surface. Table 1 organizes the surface into five levels. Level 1, traditional perimeter, relies on network-edge defenses and reactive manual vulnerability handling; its prevalence is small and shrinking. Level 2, identity-centric, adds multi-factor authentication (MFA), single sign-on (SSO), and identity and access management (IAM) controls and covers a substantial share of current adopters. Level 3 adds microsegmentation and security information and event management (SIEM) but still treats remediation as a ticketing problem. Level 4 integrates AI-driven detection and SOAR playbooks are where most mature programs currently sit [7] [8]. Level 5 is the level at which vulnerability remediation is itself autonomous, and fewer than one in a hundred organizations report reaching it.

Table 1. Five-level zero-trust maturity model

Level	Name	Dominant control	Vulnerability posture	Enterprise prevalence
1	Traditional perimeter	Firewall / VPN	Reactive manual	~11%
2	Identity-centric	MFA, SSO, IAM	Ticket-driven	~43%
3	Microsegmentation + monitoring	Segmentation, SIEM	Manual prioritized	~36%

4	AI-driven detection / response	SOAR, behavior analytics	Prioritized + partial auto	~9%
5	Autonomous remediation	Closed-loop patch pipeline	Autonomous, confidence-gated	~1%

The practical consequence of sitting below Level 5 is that the verification side of the framework runs continuously while the remediation side runs on a monthly or quarterly cadence. Regulatory SLAs for critical CVEs frequently require remediation within 72 hours, high-severity within 14 days, and medium-severity within 30 days [10]. The Ponemon Institute reports an industry-typical MTTR of roughly 60 days for critical vulnerabilities, a figure echoed in independent analyses, showing that these SLAs are frequently missed at the mean, not only in the tail [4]. The zero-trust posture score of an enterprise that is strong on identity and segmentation but weak on remediation is therefore systematically overstated by rubrics that weight verification more heavily than remediation, and the score discrepancy widens with the CVE ingestion rate.

The architectural response is to promote remediation into a first-class, continuously running control loop that shares infrastructure and governance with the access-verification loop. In this framing, Level 5 is not an additional automation layer; it is the remediation analogue of what identity-centric control did for access. Published reviews of AI-powered vulnerability detection and automated patching report meaningful gains from this framing, reductions in MTTR from the 60-day industry mean to the single-digit-day range for enterprises that adopt agentic remediation [9] [11], and the same reviews flag the safety concerns that motivate the confidence-gated policy described in the next section.

4. Autonomous Remediation Pipeline and Confidence-Gated Action

The autonomous remediation pipeline is a five-stage closed loop. Table 2 summarizes the stages, their inputs, their outputs, and their target latencies. Discovery and ingestion consume scanner feeds, NVD updates, and configuration management database (CMDB) asset data, and produce normalized CVE events in under a minute. Prioritization combines CVSS base severity with EPSS exploit probability and asset criticality to produce a ranked remediation queue in under five minutes [5] [6]. Patch identification draws on vendor feeds and a knowledge base of historical patch outcomes to return a candidate patch with an associated confidence score in under fifteen minutes. Staged deployment advances the patch through canary, blue/green, and broader rings over approximately two to three days, subject to continuous telemetry. To verify and close, re-scan the asset, confirm the CVE is resolved, and either close the ticket or trigger rollback.

Table 2. Five-stage autonomous remediation pipeline

Stage	Pipeline stage	Key inputs	Key outputs	Target latency
1	Discovery / ingestion	Scanner feeds, NVD, CMDB	Normalized CVE events	<1 min
2	Prioritization	CVSS, EPSS, asset criticality	Ranked remediation queue	<5 min
3	Patch identification	Vendor feeds, historical outcomes	Candidate patch + confidence	<15 min
4	Staged deployment	Canary, blue/green policies	Deployed patch + telemetry	~2–3 d
5	Verification / closure	Re-scan policy checks	Closed CVE or rollback	<10 min

The novel element is the confidence-gated action policy summarized in Table 3. Patches with a model-derived confidence of 0.90 or above are deployed through the canary-to-full-rollout sequence without human involvement, subject to post-hoc audit and automated rollback on any key performance indicator (KPI) drift. Patches with confidence in the 0.70 to 0.89 band require single-click operator approval before the same pipeline executes. Patches with a confidence level below 0.70 are routed to a full manual review by the Security Architecture Review Board (SARB). This gating pattern addresses the documented safety concerns in recent autonomous-patching literature: unconstrained agent autonomy over complex codebases and configurations

has been shown to raise, rather than reduce, vulnerability risk in some cases [9, 12]. The confidence gate is therefore not a cost-saving mechanism but a safety mechanism that expands autonomy only where empirical success rates support it.

Table 3. Confidence-gated action policy

Confidence band	Action policy	Human involvement	Rollback trigger
≥ 0.90	Automated canary + full rollout	None (post-hoc audit)	Automated on KPI drift
0.70–0.89	One-click human approval	Operator confirm	Automated on KPI drift
< 0.70	Full manual review	SARB approval required	Manual

The outcomes of this pipeline, when combined with the surrounding governance, are summarized in Table 4 and match the most-cited figures in the autonomous-remediation literature. MTTR falls from approximately 14 days to 2.7 days (an 81 percent reduction), the open-CVE backlog compresses by approximately 78 percent, autonomous remediation covers approximately 83 percent of CVEs, and patch success rates rise from roughly 92 percent under manual dispatch to 99.7 percent under the gated pipeline. Critical-CVE SLA compliance climbs from approximately 64 percent to 98 percent, and the aggregated zero-trust posture score (scored against a CISA-style rubric weighted to include remediation) moves from 42 to 87 on a hundred-point scale [3] [10] [11]. Engineering capacity previously consumed by routine patch work is redirected to strategic security engineering, a shift of roughly forty-seven percentage points in capacity allocation.

Table 4. Before-and-after outcomes: manual baseline vs. autonomous pipeline

Metric	Manual baseline	Autonomous pipeline	Change
Mean time to remediate (critical)	~14 days	~2.7 days	Negative 81%
CVE backlog (indexed)	100	22	Negative 78%
Autonomous remediation share	0%	~83%	+83 pp
Patch success rate	~92%	~99.7%	+7.7 pp
Critical CVE SLA compliance	~64%	~98%	+34 pp
Zero-trust posture score (indexed 0–100)	42	87	45
Engineering capacity on strategic work	~27%	~74%	+47 pp

5. Governance, Regulatory Alignment, and Operational Implications

Autonomous remediation sits inside a dense regulatory context. Sarbanes-Oxley (SOX), the Payment Card Industry Data Security Standard (PCI DSS), the Health Insurance Portability and Accountability Act (HIPAA), the European Union's Network and Information Security Directive 2 (NIS2), and the Digital Operational Resilience Act (DORA) each impose either explicit or implicit requirements on remediation cadence, audit trail, and change control. The confidence-gated pipeline is compatible with these regimes because every automated action produces a cryptographically signed audit record, every change traverses the same change-management substrate as a human-initiated change, and every rollback is both automatic and recorded [10] [13]. The governance body of record remains the SARB, whose role shifts from approving individual patches to approving the policies, confidence thresholds, and rollback KPIs that govern the pipeline.

The operational implications for enterprise data center teams are substantial and not uniformly positive. Autonomous remediation displaces a category of routine work that formerly justified a category of headcount, and it raises the skill floor for the security-operations function because the humans who remain must design the policies rather than execute the patches. SOAR market analyses reporting a compound annual growth rate (CAGR) near 15 percent reflect this repositioning rather than a simple expansion of tooling [7]. Mature programs pair the deployment of autonomous remediation with a deliberate retraining and redeployment of

the affected workforce; initiatives that do not risk both skills attrition and morale loss, and they frequently revert to manual workflows when incidents occur.

The broader implication is that zero-trust, completed by autonomous remediation, is a posture rather than a product. The framework produces measurable security improvements because the two halves of the loop, verification and remediation, are now jointly continuous rather than independently cadenced. Enterprises that pursue only the verification half will see their posture scores plateau at the mid-range even as their adoption metrics look strong; firms that complete the loop will see their posture scores converge on the upper band of the rubric and will hold that position under the rising CVE ingestion rate that shows no sign of slowing [4] [14].

Conclusion

Autonomous vulnerability remediation is the missing layer of zero-trust architecture. Verification without remediation yields an architecturally impressive but operationally incomplete posture in which verified resources remain exploitable. A five-stage remediation pipeline paired with a confidence-gated action policy closes the loop and delivers large, consistent improvements across every dimension that matters: MTTR, backlog, success rate, SLA compliance, and the aggregate zero-trust posture score. The pipeline is compatible with existing regulatory regimes, and its outputs are auditable in the same substrate as human-initiated changes. The architecture is not a speculative extension of zero-trust; it is the operational form that zero-trust must take once the CVE ingestion rate exceeds what human workflows can absorb, which it now does.

Limitations and future work follow naturally. The outcomes reported here are drawn from enterprise case studies and industry data, not from controlled experiments; they are indicative rather than definitive, and the confidence-gating thresholds are likely to require per-environment tuning. Researchers do not yet understand the interaction between autonomous remediation and emerging agentic AI systems, where the workload being remediated is itself agentic and mutates at a higher frequency, making it the most pressing research direction for the next operating cycle. The framework proposed here is the starting point for that investigation, not its conclusion.

References

[1] Scott Rose, et al., "Zero trust architecture," NIST Special Publication 800-207, 2020. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf

[2] Naeem Firdous Syed, et al., "Zero trust architecture (ZTA): a comprehensive survey," IEEE Access, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9773102

[3] Cybersecurity and Infrastructure Security Agency, "Zero trust maturity model, version 2.0," 2023. [Online]. Available: https://www.cisa.gov/sites/default/files/2023-04/CISA_Zero_Trust_Maturity_Model_Version_2_508c.pdf

[4] Tenable, "Measuring and managing the cyber risks to business operations," 2018. [Online]. Available: https://strategicrisk.net/wp-content/uploads/2019/02/Ponemon-Cyber-Report-Dec-2018.pdf

[5] Ertugrul Hakan Tan, et al., "Analysis of vulnerability severity and exploit probability scoring frameworks: CVSS and EPSS," 2025 Systems and Information Engineering Design Symposium (SIEDS), 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11021216

[6] FIRST.org, "Exploit prediction scoring system (EPSS)." [Online]. Available: https://www.first.org/epss/

[7] Manh-Dung Nguyen, et al., "AI4SOAR: a security intelligence tool for automated incident response," ARES '24: Proceedings of the 19th International Conference on Availability, Reliability and Security (ARES), 2024. [Online]. Available: https://dl.acm.org/doi/epdf/10.1145/3664476.3670450

[8] Pragya Rathore and Kishor Kolhe, "Integrating Automation and Orchestration in Security Incident Handling: A Review of SOAR Frameworks and Platforms," Proceedings of the UNIfied Conference of DAMAS, IncoME VIII and TEPEN Conferences, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-95963-9_38

[9] Amirali Sajadi, et al., "How Safe Are AI-Generated Patches? A Large-Scale Study on Security Risks in LLM and Agentic Automated Program Repair on SWE-bench," arXiv, 2025. [Online]. Available: https://arxiv.org/pdf/2507.02976

[10] National Institute of Standards and Technology, "The NIST Cybersecurity Framework (CSF) 2.0," 2024. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf

[11] Malek Malkawi and Reda Alhajj, "AI-Powered Vulnerability Detection and Patch Management in Cybersecurity: A Systematic Review of Techniques, Challenges, and Emerging Trends," Machine Learning and Knowledge Extraction, 2026. [Online]. Available: https://www.mdpi.com/2504-4990/8/1/19
[12] Jan Nowakowski and Jan Keller, "AI-powered patching: the future of automated vulnerability fixes," Google Security Engineering Technical Report, 2024. [Online]. Available: https://storage.googleapis.com/gweb-research2023-media/pubtools/7563.pdf
[13] Tao Chen and Haiyan Suo, "Design and practice of security architecture via DevSecOps technology," 2022 IEEE 13th International Conference on Software Engineering and Service Science (ICSESS), 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9930212
[14] Olalekan Kosile and Oladimeji Oyegunle, "Bridging the patch gap: analyzing ransomware exploitation tactics over extended vulnerability dwell times," 2025 8th International Conference on Algorithms, Computing and Artificial Intelligence (ACAI), 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11406510
[15] Ismail, et al., "Toward Robust Security Orchestration and Automated Response in Security Operations Centers with a Hyper-Automation Approach Using Agentic Artificial Intelligence," Information, 2025. [Online]. Available: https://www.mdpi.com/2078-2489/16/5/365