Front-code.com

SURVEY PAPER

Zero Trust Architecture

Identity-Centric Security: A Comprehensive Survey of **NIST SP 800-207**, Federal Maturity Models, Phishing-Resistant Authentication, and Migration Roadmaps

NIST SP 800-207 | CISA ZTMM 2.0 | NIST SP 800-63-4



OCTOBER 2026

NEVER TRUST · ALWAYS VERIFY
IDENTITY · DEVICE · WORKLOAD

SURVEY PAPER · OCTOBER 2026

Zero Trust Architecture (ZTA) and Identity-Centric Security: A Comprehensive Survey of NIST SP 800-207, Federal Maturity Models, Phishing-Resistant Authentication, and Enterprise Migration Roadmaps

Front-code.com

***Abstract*—**The perimeter-based security model—in which a hardened network boundary separates a trusted interior from an untrusted exterior—has been eroded by cloud computing, software-as-a-service (SaaS), remote work, and dense supply-chain interconnection. Adversaries rarely need to breach the wall any longer: they log in. Stolen passwords, session cookies, and OAuth tokens feature in a large share of breaches, and once inside, implicitly trusted networks permit largely unchecked lateral movement. Zero Trust Architecture (ZTA) responds by removing implicit trust based on network location and by evaluating every access request, per session, against dynamic policy informed by the identity of the subject, the posture of the device, and the sensitivity of the resource. This paper presents a comprehensive survey of zero trust with an emphasis on identity as the primary control plane. We trace the evolution from the Jericho Forum's de-perimeterization and Google's BeyondCorp to NIST SP 800-207 and its companion publications, SP 800-207A and SP 1800-35; formalize the trust algorithm and the policy decision and enforcement model; analyze phishing-resistant authentication under the revised NIST SP 800-63-4 guidelines, including synced and device-bound passkeys; examine token binding, continuous access evaluation through the OpenID Shared Signals Framework, and attestation-based workload identity with SPIFFE; compare the CISA Zero Trust Maturity Model 2.0 with the U.S. Department of Defense zero trust strategy and its 2026 implementation guidelines; map prevalent identity-centric attack techniques to zero trust countermeasures; and construct a phased, metrics-driven enterprise migration roadmap. We close with open problems, including the authorization of autonomous AI agents and the migration of identity infrastructure to post-quantum signatures.

***Index Terms*—**Zero trust architecture (ZTA), NIST SP 800-207, identity and access management (IAM), phishing-resistant MFA, FIDO2/WebAuthn, passkeys, continuous access evaluation, Shared Signals Framework (SSF), CAEP, micro-segmentation, workload identity, SPIFFE, CISA Zero Trust Maturity Model, least privilege.

I. INTRODUCTION

For three decades, enterprise security was organized around the castle and its moat. Firewalls, demilitarized zones, and virtual private networks (VPNs) defined a perimeter, and whatever sat inside it was trusted by default. That assumption has collapsed. Workloads run across several public clouds, business data lives in dozens of SaaS tenants, staff connect from networks the enterprise does not control, and every organization is stitched to its suppliers through APIs and delegated tokens. The "inside" of the network no longer contains the assets worth protecting, and the people who need those assets are rarely inside it either.

Attackers adapted first: rather than breaching a boundary, they acquire valid identities. In the 2026 Verizon Data Breach Investigations Report (DBIR), vulnerability exploitation (31%) overtook credential

abuse as the leading initial access vector for the first time, yet stolen or misused credentials still appeared somewhere along the attack chain in 39% of breaches—more than any other vector [1]. The 2021 Colonial Pipeline ransomware attack began with a legacy VPN account that lacked multi-factor authentication (MFA) [2]; in 2024, infostealer-harvested passwords opened customer tenants of the Snowflake data platform, and about 165 potentially exposed organizations were notified [3]; and in August 2025, OAuth tokens stolen from the Salesloft Drift integration were used to exfiltrate data from the Salesforce instances of more than 700 potentially affected organizations [4]. In none of these cases did the adversary need to defeat a firewall.

Zero trust is the architectural answer. NIST Special Publication (SP) 800-207 describes it as an evolving set of paradigms that move defenses from static, network-based perimeters to users, assets, and resources [5]. No implicit trust is granted on the basis of physical or network location or of ownership: every request is authenticated, authorized, and encrypted; access is granted per session and with least privilege; and the posture of every participant is monitored continuously. Because location no longer confers trust, *identity*—of people, devices, and workloads—becomes the primary control plane, the substrate on which every other pillar depends.

> **The core principle: never trust, always verify**
> Zero trust assumes that the network is already hostile and that a breach has already occurred. Every access request is explicitly verified against identity, device posture, and context, granted with least privilege for a single session, and re-evaluated as conditions change.

Section II traces the lineage of zero trust; Section III formalizes the NIST reference architecture and its trust algorithm; Sections IV and V analyze identity as the control plane and extend it to devices, networks, workloads, and AI agents; Section VI compares maturity models; Section VII maps identity-centric attacks to countermeasures; Sections VIII and IX cover implementation challenges and a migration roadmap; and Sections X and XI give research directions and conclusions.

II. HISTORICAL EVOLUTION: FROM PERIMETER DEFENSE TO ZERO TRUST

A. De-Perimeterization and the Zero Trust Model

The roots of zero trust lie in the Jericho Forum, convened under The Open Group in 2004. The Forum argued that mobility and business-to-business integration were dissolving the perimeter, and its *Commandments* (2007) called for protection that travels with the data and for authenticated communication irrespective of the network in between [6]. In 2010, John Kindervag of Forrester Research coined the term “zero trust” for a model in which all traffic is inspected and logged, access follows strict need-to-know, and networks are segmented around the most valuable assets—the *protect surface*—instead of trusting a soft interior [7].

B. BeyondCorp and the Software-Defined Perimeter

The first large-scale production implementation came from Google. After the Operation Aurora intrusions of 2009, Google moved its corporate applications off the privileged intranet and behind an internet-facing access proxy, so that access depended on the user and the device rather than on the network. BeyondCorp, published from 2014 onward, rested on an inventory of managed devices with per-device certificates, a user and group database, and an access proxy that consulted a trust-inference engine before forwarding each request [8], [9]. In parallel, the Cloud Security Alliance specified the Software-Defined Perimeter

(SDP), in which a controller authenticates and authorizes a client before any connectivity to the protected service exists, so that services stay invisible to unauthorized hosts [10].

C. From Concept to Mandate

Zero trust became public policy between 2020 and 2023. NIST published SP 800-207 in August 2020 [5]. After the SolarWinds compromise, Executive Order 14028 of May 2021 directed federal agencies to plan for zero trust [11], and OMB Memorandum M-22-09 (January 2022) set goals for the end of fiscal year 2024: enterprise-wide identity, phishing-resistant MFA, a full device inventory with endpoint detection and response (EDR), encrypted DNS and HTTP traffic, and internal applications treated as internet-accessible [12]. The UK National Cyber Security Centre (NCSC) published eight design principles in 2021 [13], the U.S. Department of Defense (DoD) released its Zero Trust Strategy in November 2022 [14], and the Cybersecurity and Infrastructure Security Agency (CISA) issued version 2.0 of its Zero Trust Maturity Model in April 2023 [15].

The emphasis has since shifted to implementation. SP 800-207A (2023) extended the model to cloud-native, multi-cloud applications [16]; the NIST National Cybersecurity Center of Excellence (NCCoE) finalized SP 1800-35 in June 2025, with 19 example implementations built with 24 industry collaborators [17]; NIST finalized SP 800-63-4 in July 2025 [18]; the OpenID Foundation approved the Shared Signals Framework, CAEP, and RISC as final in September 2025 [19], [20]; the National Security Agency (NSA) published the DoD Zero Trust Implementation Guidelines in January 2026 [21]; and Web Authentication Level 3 became a W3C Recommendation in August 2026 [22]. Fig. 1 summarizes this progression.

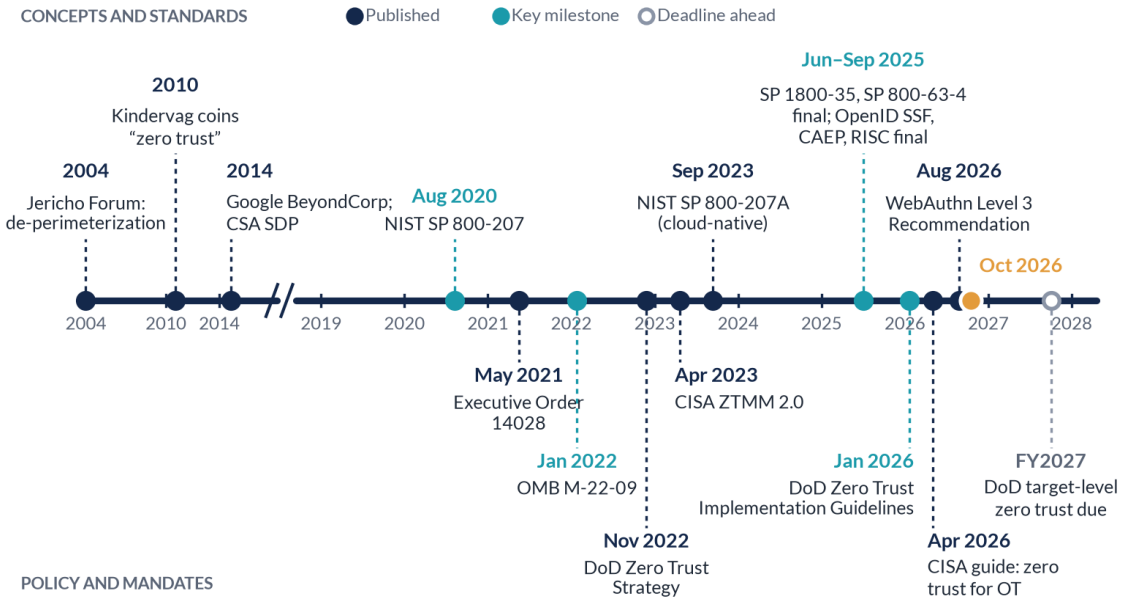

Fig. 1. Evolution of zero trust from de-perimeterization to federal implementation guidance, 2004–2027. Concepts and standards are shown above the axis and policy milestones below it; the time axis is compressed before 2019.

III. THE NIST SP 800-207 REFERENCE ARCHITECTURE

A. Tenets of Zero Trust

SP 800-207 sets out seven tenets [5]. They are design goals rather than a checklist; few enterprises satisfy all of them for every resource.

1. **Everything is a resource**, including personal devices that reach enterprise data and SaaS platforms.

2. **Location confers no trust:** all communication is secured, and a request from inside the network meets the same requirements as one from the internet.
3. **Per-session access:** authenticating to one resource does not grant access to another.
4. **Dynamic policy** weighs the observable state of the client identity, the application, and the requesting asset, plus behavioral and environmental attributes.
5. **Continuous posture monitoring** of all owned and associated assets; no asset is inherently trusted.
6. **Strict, dynamic enforcement** before every access, within a constant cycle of assessing threats, adapting, and re-evaluating trust.
7. **Telemetry-driven improvement** from data on assets, infrastructure, and communications.

B. Logical Components: Policy Decision and Enforcement

SP 800-207 decomposes a deployment into a few logical components (Fig. 2). The *policy engine* (PE) decides whether to grant, deny, or revoke access, feeding enterprise policy and external inputs into a *trust algorithm*. The *policy administrator* (PA) executes the decision by establishing or tearing down the communication path, for example by issuing a session-specific token. Together they form the *policy decision point* (PDP). The *policy enforcement point* (PEP) sits in the data path between subject and resource and enables, monitors, and terminates connections on the PA's instruction. Separating the *control plane* from the *data plane* is fundamental: the subject reaches the resource only through a PEP that the PDP has configured for that session.

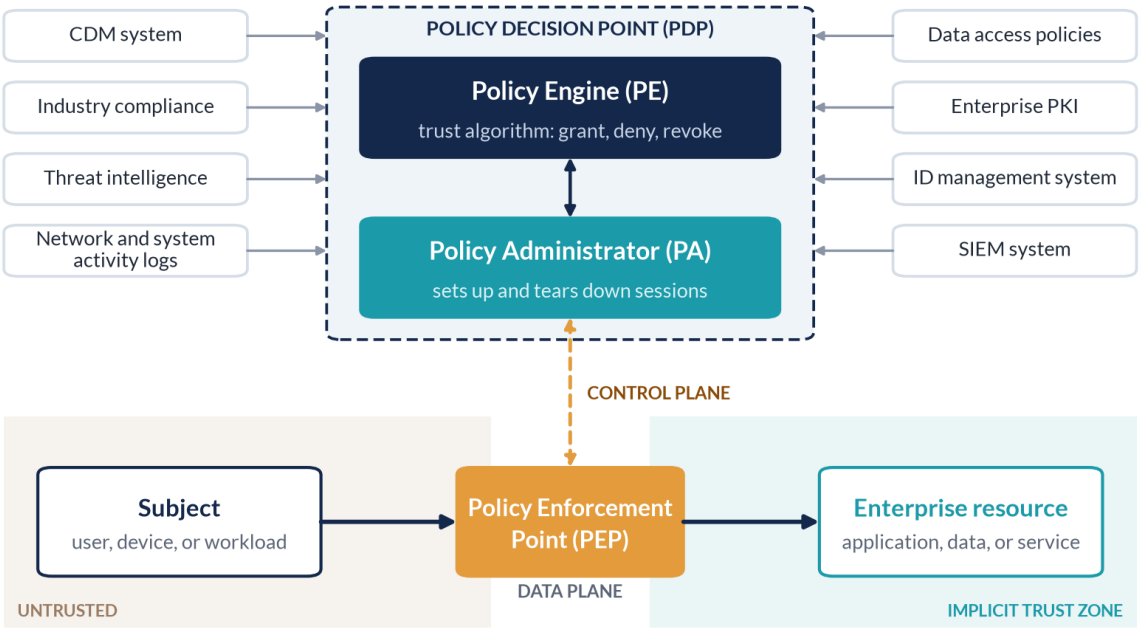

Fig. 2. Core logical components of a zero trust architecture after NIST SP 800-207 [5]. The policy decision point (PE and PA) configures the policy enforcement point over the control plane; the subject reaches the resource only through the PEP.

The PE draws context from a continuous diagnostics and mitigation (CDM) system, compliance requirements, threat intelligence, activity logs, data access policies, the enterprise public-key infrastructure (PKI), the identity management system, and a security information and event management (SIEM) platform. Each decision is only as good as these inputs, which is why visibility and analytics are cross-cutting capabilities in every maturity model in Section VI.

C. The Trust Algorithm

SP 800-207 distinguishes *criteria-based* trust algorithms, which grant access only when a fixed set of attributes is satisfied, from *score-based* ones, which compare a weighted confidence level with a threshold set by the resource owner; and *singular* algorithms, which judge each request in isolation, from *contextual* ones, which use the subject's recent history to spot anomalies such as a burst of downloads [5]. A score-based, contextual algorithm can be written as (1), where each signal $\varphi_k \in [0, 1]$ scores one factor—authentication strength, device compliance, network reputation, behavioral deviation, or threat intelligence—for subject s on device d in context c, and the weights w_k sum to one.

$$T(s,d,c) = \sum_{k=1}^{m} w_k \varphi_k(s,d,c)\,, \quad \sum_{k=1}^{m} w_k = 1, \quad 0 \le \varphi_k \le 1 \tag{1}$$

Two resource-specific thresholds, $\theta_{lo}(r) < \theta_{hi}(r)$, map the confidence to a decision as in (2). The intermediate *step-up* outcome is essential to usability: instead of denying a legitimate user whose context has drifted, the PDP demands more assurance, such as a fresh phishing-resistant authentication or a device health attestation. High-value resources set both thresholds higher.

$$\text{decision}(s,r) = \begin{cases} \text{permit}, & \text{if} \quad T \ge \theta_{\text{hi}}(r) \\ \text{step-up}, & \text{if} \quad \theta_{\text{lo}}(r) \le T < \theta_{\text{hi}}(r) \\ \text{deny}, & \text{if} \quad T < \theta_{\text{lo}}(r) \end{cases} \tag{2}$$

Weights and thresholds are policy, and policy that cannot be explained cannot be audited, so transparent, testable rules are preferable to opaque risk scores for high-impact decisions. Some gates should also stay criteria-based: no score should compensate for a missing phishing-resistant authenticator on an administrative session.

D. Deployment Approaches and Models

SP 800-207 identifies three complementary approaches. *Enhanced identity governance* (EIG) makes subject identity and attributes the main policy input and suits cloud and SaaS, where the enterprise does not control the network. *Micro-segmentation* places resources behind gateway PEPs such as next-generation firewalls or host agents to contain lateral movement. *Software-defined perimeters* use an overlay network, such as an SDP or a service mesh, with the PA acting as network controller. Mature deployments combine all three [5], [17].

PEPs are placed through four models: device agent/gateway, enclave gateway, resource portal, and application sandboxing. SP 800-207A adds cloud-native guidance: policy is expressed at both the network tier and the *identity tier*, with service-mesh proxies acting as PEPs for every service-to-service call [16].

IV. IDENTITY AS THE CONTROL PLANE

Once location confers no trust, "who is asking?" and "how sure are we?" carry the full weight of every decision. Identity-centric security spans proofing, authentication, session and token security, and authorization.

A. Assurance Levels and Phishing-Resistant Authentication

NIST SP 800-63-4 (July 2025) structures digital identity around identity assurance levels (IAL) for proofing, authentication assurance levels (AAL) for authenticators, and federation assurance levels (FAL) for assertions [18], [23]. The revision strengthens risk management, recognizes subscriber-controlled

digital wallets, and formally incorporates *syncable authenticators* (passkeys), whose private keys may be copied among a user's devices.

The decisive property is *phishing resistance*: the protocol binds the authenticator's output to the verifier's identity, so a look-alike site or an adversary-in-the-middle (AiTM) proxy cannot relay it. FIDO2/WebAuthn scopes each credential to the relying party and signs a challenge bound to the requesting origin; PIV and CAC smart cards rely on certificate-based mutual TLS. M-22-09 requires phishing-resistant MFA for federal staff and names PIV and WebAuthn as the two widely used approaches [12], [24]. Table I compares common authenticators.

Simple approve-or-deny push prompts invite *MFA fatigue*, in which an attacker who holds a password triggers prompts until the user accepts one, as in the September 2022 compromise of Uber [25]. Number matching defeats blind approval but not phishing, because an AiTM proxy simply relays the number [24]; only origin-bound authenticators close that gap. Synced passkeys, which relying parties can recognize through the backup-eligibility and backup-state flags standardized in WebAuthn Level 3 [22], have largely removed the usability argument against phishing-resistant authentication for the workforce, while device-bound security keys remain the choice for administrators and AAL3.

B. Session and Token Security

Strong authentication protects only the moment of sign-in; afterward, access rests on bearer artifacts—session cookies, OAuth tokens, SAML assertions—that anyone holding them can use. AiTM phishing kits harvest session cookies after the victim completes MFA, and infostealers copy them from browser profiles, so the session itself must be protected.

Table I. Comparison of common authenticators against identity-centric attacks

Authenticator	Example	Phishing-resistant	Resists prompt fatigue	Key or secret	Highest AAL
Password only	Memorized secret	No	—	Reusable, phishable	AAL1
SMS or voice OTP	Code over the phone network	No	Yes	SIM-swap exposure	AAL2 (restricted)
Software OTP (TOTP)	Authenticator-app code	No	Yes	Shared seed	AAL2
Push with number matching	Mobile-app prompt	No	Yes	Device-bound	AAL2
Synced passkey	WebAuthn, platform-synced	Yes	Yes	Exportable via sync	AAL2
Device-bound passkey	FIDO2 security key	Yes	Yes	Non-exportable	AAL3
PIV / CAC smart card	X.509 certificate + PIN	Yes	Yes	Non-exportable	AAL3

Highest assurance level achievable in a conforming multi-factor configuration under NIST SP 800-63B-4 [23]; authenticators that use the public telephone network are designated *restricted*. Synced passkeys are limited to AAL2 because their keys can be exported.

Sender-constrained tokens bind a token to a key held by the client: OAuth 2.0 mutual TLS binds it to a client certificate [26], and Demonstrating Proof of Possession (DPoP) to a key that signs every request [27]. The OAuth 2.0 Security Best Current Practice, RFC 9700, recommends sender-constraining and requires PKCE and exact redirect-URI matching [28]. The complementary lever is *short exposure*. With token lifetime τ_{TTL}, detection time τ_{det}, and revocation propagation time τ_{prop}, the exposure window of a stolen token is bounded by (3).

$$W_{\text{exp}} = \min\left(\tau_{\text{TTL}},\ \tau_{\text{det}} + \tau_{\text{prop}}\right) \tag{3}$$

Without a revocation channel, τ_{prop} is unbounded and the window equals the full token lifetime—commonly an hour or more for access tokens and far longer for refresh tokens. Continuous access evaluation attacks the second term.

C. Continuous Access Evaluation and Shared Signals

Classic federation evaluates policy only at sign-in: a relying party has no standard way to learn that an account was disabled, a device fell out of compliance, or a session was revoked at the identity provider. The OpenID Shared Signals Framework (SSF) closes this gap by defining how a *transmitter* and a *receiver* exchange signed Security Event Tokens over an event stream [19]. The Continuous Access Evaluation Profile (CAEP) carries session-relevant events such as *session-revoked*, *credential-change*, *device-compliance-change*, *assurance-level-change*, and *token-claims-change* [20], while RISC carries account-level risk such as credential compromise. With all three approved as final in September 2025, near-real-time revocation across vendors is now a standards-based capability.

In Fig. 3, a session that begins with a passkey carries high confidence; an impossible-travel signal drops it below θ_{hi} and triggers step-up authentication; a later *device-compliance-change* event drops it below θ_{lo}, and the session ends within seconds. Under a static model it would have stayed valid until the token expired.

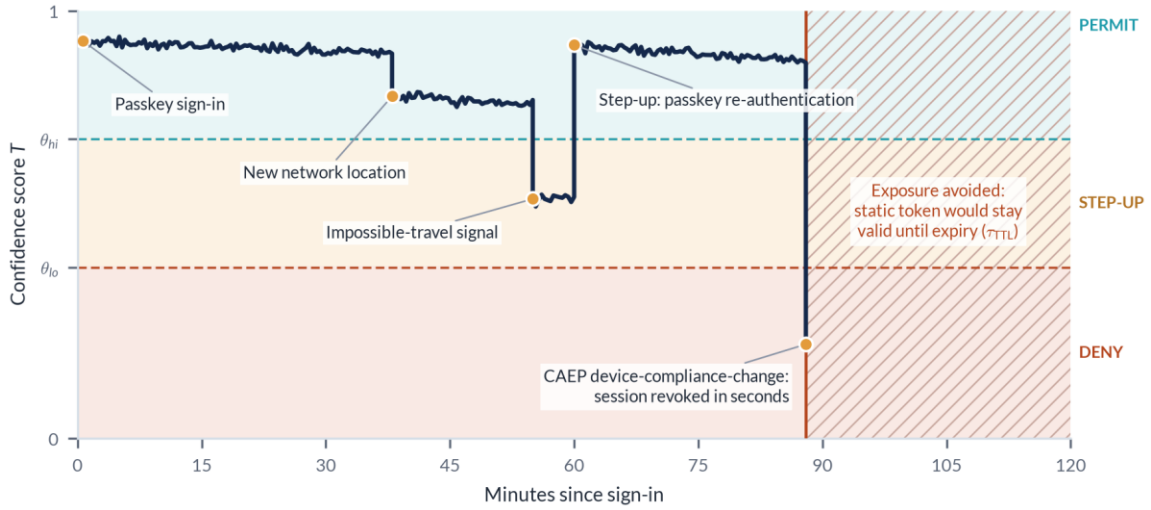

Fig. 3. Continuous access evaluation of a single session (illustrative). Identity-provider signals and CAEP events move the confidence score of (1) across the step-up and deny thresholds of (2); the hatched band is the exposure that a static token lifetime would have allowed.

D. Authorization: From Roles to Policy as Code

Authorization in a zero trust architecture must be fine-grained, contextual, and auditable. Role-based access control (RBAC) remains common, but roles accumulate privileges and cannot express context. Attribute-based access control (ABAC), formalized in NIST SP 800-162, evaluates attributes of subject, object, action, and environment and maps naturally onto the trust algorithm [29]. Relationship-based access control (ReBAC), popularized by Google's Zanzibar, derives permissions from a relationship graph at global scale [30]. Increasingly, such policies are written as *code*—in Rego for the Open Policy Agent, or in Cedar, designed for formal analysis [31]—and versioned, tested, and verified like software, while just-in-time elevation applies least privilege in time as well as scope.

V. DEVICE, NETWORK, WORKLOAD, AND AGENT IDENTITY

A. Device Trust and Posture

A legitimate user on a compromised device is not a trustworthy subject. Zero trust needs an authoritative device inventory and continuous posture assessment—patch level, disk encryption, EDR status, configuration compliance—and M-22-09 required both inventory and EDR [12]. The strongest signals are *attested*: a Trusted Platform Module (TPM) or secure enclave signs statements about device identity and boot state, so the PDP relies on cryptographic evidence rather than an agent's self-report. Device identity should sit on non-exportable keys so a certificate cannot be copied to another machine, as with BeyondCorp's per-device certificates [9].

B. Network: Micro-Segmentation and ZTNA

Zero trust demotes the network from an authority to an enforcement fabric. Micro-segmentation divides the environment into small default-deny zones, ideally down to individual workloads. Zero trust network access (ZTNA) replaces broad VPN tunnels with per-application connections brokered by an identity-aware proxy, so a user authorized for one application cannot even discover the others; the secure access service edge (SASE) model packages ZTNA with web gateways and cloud access security brokers as a cloud service [32]. Both techniques minimize the *blast radius* of a compromised identity i, defined in (4) as the share of enterprise value it can reach, where R is the set of resources, $R(i) \subseteq R$ the subset reachable by i, and $v(r)$ the business value of resource r.

$$\beta(i) = \frac{\sum_{r \in R(i)} v(r)}{\sum_{r \in R} v(r)} \tag{4}$$

In a flat network with standing privileges, $\beta(i)$ approaches one; least privilege, micro-segmentation, and per-session access drive it toward what each identity needs, turning "containing lateral movement" into a measurable quantity.

C. Workload and Non-Human Identities

Machine identities—service accounts, API keys, OAuth clients, CI/CD pipelines, containers, and serverless functions—greatly outnumber human users, yet often authenticate with long-lived static secrets kept in code or configuration. These non-human identities (NHIs) are seldom reviewed and often over-privileged. In the 2025 Salesloft Drift campaign, one vendor's compromise yielded OAuth tokens with programmatic access to hundreds of customer Salesforce tenants, bypassing every human-facing control [4].

The remedy is cryptographic, short-lived, automatically rotated workload identity issued on the basis of *attestation* rather than a shared secret. SPIFFE, a graduated Cloud Native Computing Foundation project, defines a workload identifier of the form spiffe://trust-domain/path and two identity documents: X.509-SVIDs for mutual TLS and JWT-SVIDs for tokens [33]. Its reference implementation, SPIRE, attests the node and the workload (for example, from cloud instance identity and Kubernetes pod metadata) before issuing and rotating an SVID. Service meshes use these identities for mutual TLS and identity-tier policy on every call [16], and federation exchanges an SVID or a CI/CD platform's OIDC token for short-lived cloud credentials through OAuth 2.0 token exchange [34] (Fig. 4).

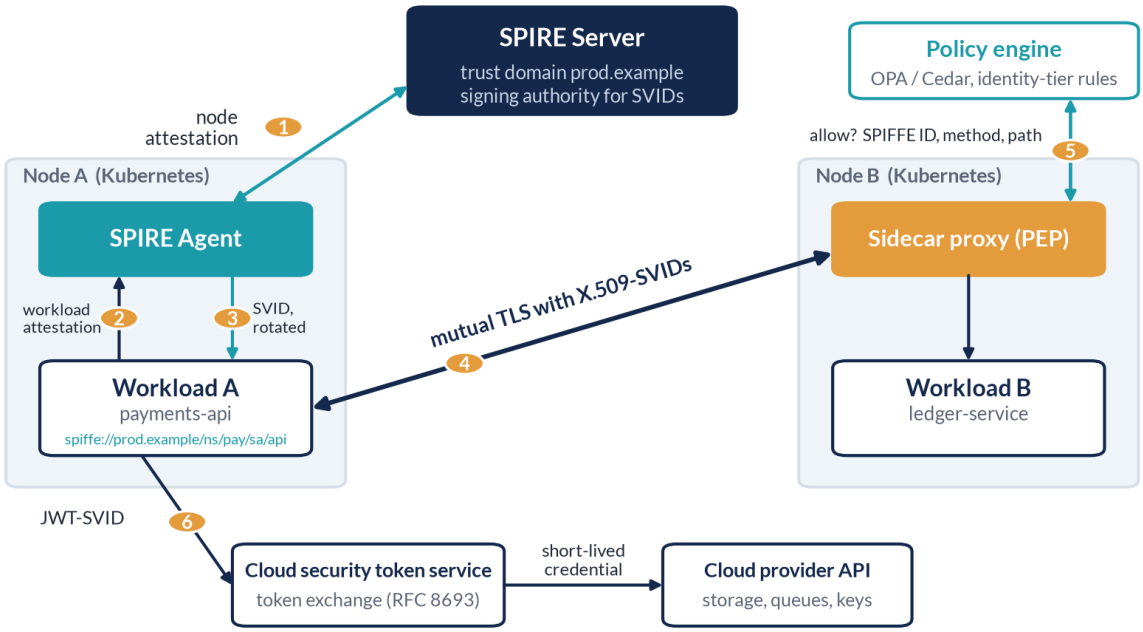

Fig. 4. Attestation-based workload identity with SPIFFE/SPIRE. Workloads receive short-lived SVIDs after node and workload attestation, authenticate to each other with mutual TLS, and are authorized by an identity-tier policy decision at the receiving proxy; federation exchanges a JWT-SVID for short-lived cloud credentials.

D. AI Agents as First-Class Identities

AI agents, which invoke tools and APIs on behalf of users, are the fastest-growing class of non-human identity: they act with delegated human authority yet can be steered by untrusted content through prompt injection. The Model Context Protocol (MCP) classified MCP servers as OAuth 2.1 resource servers in its June 2025 revision, requiring protected-resource metadata discovery (RFC 9728) and audience-restricted tokens through resource indicators (RFC 8707) [35]. In February 2026, the NCCoE proposed applying OAuth 2.0/2.1, OpenID Connect, SPIFFE/SPIRE, SCIM, and Next Generation Access Control to the identification, authorization, auditing, and non-repudiation of agents [36]. Each agent therefore needs a distinct, attestable identity, task-scoped and short-lived delegated tokens, logged delegation hops, and fresh human approval for high-impact actions—a step-up decision in the sense of (2).

E. Data-Centric Controls

Data is what zero trust ultimately protects. The CISA data pillar calls for inventory and classification, encryption at rest and in transit, data-loss prevention, and sensitivity-aware access decisions [15]. Classification labels are what let the PDP set the thresholds of (2) per resource.

VI. FEDERAL AND INDUSTRY MATURITY MODELS

Architecture describes the destination; maturity models describe the journey. Table II compares the most influential frameworks.

A. CISA Zero Trust Maturity Model 2.0

CISA organizes zero trust into five pillars—Identity, Devices, Networks, Applications and Workloads, and Data—supported by three cross-cutting capabilities: Visibility and Analytics, Automation and Orchestration, and Governance [15]. Version 2.0 added an *Initial* stage, giving four: *Traditional* (manual, static, siloed), *Initial* (first automation and cross-pillar integration), *Advanced* (centralized visibility, risk-

based least privilege), and *Optimal* (fully automated, just-in-time, dynamic policy). Being technology-neutral, the model applies well beyond federal agencies.

Table II. Comparison of zero trust frameworks and maturity models

Framework	Issuer, year	Structure	Maturity scale	Primary scope
NIST SP 800-207	NIST, 2020	7 tenets; PE, PA, and PEP logical model; 3 approaches	None (architecture)	All sectors
NIST SP 800-207A	NIST, 2023	Identity-tier and network-tier policy for cloud-native apps	None	Multi-cloud, service mesh
NIST SP 1800-35	NIST NCCoE, 2025	19 example builds (EIG, micro-segmentation, SDP)	None (practice guide)	All sectors
CISA ZTMM 2.0	CISA, 2023	5 pillars and 3 cross-cutting capabilities	Traditional, Initial, Advanced, Optimal	U.S. federal civilian agencies
DoD ZT Strategy and ZIGs	DoD, 2022; NSA, 2026	7 pillars, 45 capabilities, 152 activities	Target (91 activities), Advanced (61)	DoD and defense industrial base
NCSC design principles	UK NCSC, 2021	8 design principles	None	UK public and private sectors
Google BeyondCorp	Google, 2014–2017	Device inventory, access proxy, trust inference	None (implementation)	Industry reference model

ZIG = Zero Trust Implementation Guideline; EIG = enhanced identity governance; SDP = software-defined perimeter.

B. DoD Zero Trust Strategy and Implementation Guidelines

The DoD strategy defines seven pillars—User; Device; Applications and Workloads; Data; Network and Environment; Automation and Orchestration; and Visibility and Analytics—decomposed into 45 capabilities and 152 activities [14]. Ninety-one activities define *target-level* zero trust, due by the end of fiscal year 2027; the other 61 define the *advanced* level. The implementation guidelines of January 2026 sequence the target-level work into a Discovery phase (14 activities), Phase One (36), and Phase Two (41), with later phases to cover advanced capabilities [21]—ordered, dependency-aware tasks of the kind most enterprises lack.

C. Other Frameworks

The NCSC principles stress knowing one's architecture and identities, assessing user, device, and service health, authenticating and authorizing everywhere, and distrusting every network, including one's own [13]. SP 1800-35 maps its builds to the NIST Cybersecurity Framework and SP 800-53 controls [17]. In April 2026, CISA and four federal partners issued guidance on zero trust for operational technology (OT), where availability, legacy protocols, and safety limit the use of agents, encryption, and active scanning [37].

VII. THREAT LANDSCAPE: IDENTITY-CENTRIC ATTACKS

Table III maps prevalent identity-centric techniques from MITRE ATT&CK [38] to representative incidents and to the zero trust controls that would have prevented or contained them. Three lessons recur.

First, *MFA coverage gaps are exploited systematically*. Colonial Pipeline, the Snowflake intrusions, and the February 2024 Change Healthcare ransomware attack—which began with stolen credentials on a remote-access portal without MFA [39]—all hinged on one account or portal outside the MFA boundary. The Midnight Blizzard intrusion disclosed by Microsoft in January 2024 began with a password spray against a legacy test tenant account without MFA [40]. Policy must cover every identity, including test tenants, break-glass accounts, and service principals.

Second, *the help desk is part of the authentication system*. Scattered Spider, behind the 2023 MGM Resorts intrusion among others, impersonated employees to get passwords reset or attacker-controlled MFA devices enrolled [41]. Recovery must meet the assurance level of the credential it restores.

Third, *the identity provider is a crown jewel*. In 2023, Storm-0558 used an acquired Microsoft consumer signing key to forge tokens accepted by Exchange Online and read senior officials' mail; the Cyber Safety Review Board attributed it to a cascade of avoidable failures [42]. In the SolarWinds campaign, attackers forged SAML assertions with the "Golden SAML" technique described in 2017 [43]. Signing keys belong in hardware security modules, rotated and monitored, and identity threat detection and response (ITDR) should treat anomalous token issuance as a first-order alert.

Table III. Identity-centric attack techniques, representative incidents, and zero trust countermeasures

Technique	ATT&CK ID	Representative incident	Zero trust countermeasure
Credential stuffing with stolen passwords	T1078, T1110.004	Snowflake customer tenants, 2024 [3]	Phishing-resistant MFA on every account; leaked-credential detection; network allow-lists for data platforms
Password spraying	T1110.003	Midnight Blizzard vs. Microsoft, 2023–2024 [40]	MFA on test and legacy tenants; tenant and account inventory; anomalous sign-in analytics
MFA request flooding (fatigue)	T1621	Uber, 2022 [25]	Number matching at minimum; phishing-resistant WebAuthn or PIV; prompt rate limits
Adversary-in-the-middle phishing and cookie theft	T1557, T1539	Phishing-as-a-service AiTM kits	Origin-bound passkeys; sender-constrained tokens (DPoP, mTLS); CAEP session revocation
Help-desk impersonation and MFA reset	T1656	Scattered Spider, MGM Resorts, 2023 [41]	Re-proofing at the original IAL for recovery; out-of-band call-back; JIT administrative access
External remote services without MFA	T1133	Colonial Pipeline, 2021 [2]; Change Healthcare, 2024 [39]	Replace VPNs and remote portals with ZTNA; per-application access; device posture checks
Forged tokens from stolen signing keys	T1606	Storm-0558, 2023 [42]; SolarWinds, 2020 [43]	HSM-protected, rotated signing keys; strict issuer and audience validation; ITDR on token issuance
Theft of OAuth tokens held by third-party apps	T1528, T1550.001	Salesloft Drift, 2025 [4]	NHI inventory; scoped, short-lived, sender-constrained tokens; SSF/RISC signals between vendors
Unsecured workload secrets	T1552	Secrets leaked from code and CI/CD pipelines	Attested workload identity (SPIFFE); workload identity federation; secret scanning

Technique identifiers from the MITRE ATT&CK Enterprise matrix [38].

VIII. IMPLEMENTATION CHALLENGES

Legacy and OT systems. Mainframes, controllers, thick clients, and appliances often cannot federate or host agents. Enclave gateways and resource portals can wrap them at the cost of coarser policy; they should be tracked as explicit exceptions with compensating controls and retirement dates [37].

The PDP as a critical dependency. Centralizing decisions creates a high-value target and a potential single point of failure. The PDP and identity provider need high availability, defined fail-open or fail-closed behavior per resource class, monitored break-glass procedures, and the strongest protection for their signing keys [42].

Policy sprawl, signal quality, and privacy. Conditional-access rules multiply until nobody can say why a request was allowed; policy as code with testing and formal analysis helps [31]. False positives invite workarounds, and behavioral monitoring calls for data minimization and transparent governance. Zero trust is also an operating model, not a product: offerings should be judged against the functions of SP 800-207, not against labels.

Cryptographic agility of identity. Device and workload certificates, SAML assertions, JSON Web Tokens, and WebAuthn attestations all rest on signatures that must migrate to post-quantum schemes such as ML-DSA, whose larger signatures affect token sizes, certificate chains, and constrained devices. The identity tier belongs in post-quantum inventories now [44].

IX. ENTERPRISE MIGRATION ROADMAP

Zero trust is adopted one protect surface at a time, not in a single cut-over. The roadmap below synthesizes SP 800-207, the CISA and DoD models, and the NCCoE builds [5], [15], [17], [21]; phases overlap and durations are indicative.

1. **Discover and prioritize (months 0–6).** Inventory human and non-human identities, devices, applications, data, and flows; identify protect surfaces, including the identity infrastructure itself.
2. **Consolidate identity and deploy phishing-resistant MFA (months 3–12).** Converge on an authoritative identity provider; enforce passkeys, security keys, or PIV, administrators first; retire legacy protocols; harden recovery.
3. **Bind device posture to access (months 6–18).** Require managed, compliant, preferably attested devices for sensitive resources, and feed EDR signals into policy.
4. **Replace implicit network trust (months 9–24).** Move remote access from VPN to ZTNA; micro-segment around protect surfaces; encrypt internal traffic.
5. **Secure workloads and NHIs (months 12–30).** Replace static secrets with attested identities and federation; scope third-party OAuth grants; enforce mutual TLS between services.
6. **Automate continuous evaluation (months 18–36).** Consume SSF/CAEP events, adopt sender-constrained tokens, automate response, and measure against a maturity model.

Progress is measured by outcomes: the share of users and of administrators on phishing-resistant authenticators, of applications behind a PEP, and of workloads without static secrets; the median $\tau_{det} + \tau_{prop}$ of (3); and the blast radius β of (4) for the riskiest identities.

> **Executive recommendation: immediate action items**
>
> 1. Enforce phishing-resistant MFA for every administrator now, and for the whole workforce on a published timeline.
> 2. Close MFA coverage gaps: legacy protocols, test tenants, remote-access portals, break-glass and service accounts.
> 3. Treat the help desk as part of the authentication system and require re-proofing for credential resets.
> 4. Protect identity-provider signing keys in HSMs and monitor token issuance.
> 5. Inventory non-human identities and third-party OAuth grants; replace static secrets with short-lived, attested credentials.

X. FUTURE RESEARCH DIRECTIONS

Agentic authorization needs delegation models that preserve the originating user's intent across multi-hop chains of agents and tools; OAuth profiles express scope and audience but not intent [36]. **Adaptive policy with machine learning** must remain explainable, robust to manipulation, and fair. **Digital wallets**, recognized in SP 800-63-4, may let subjects present selectively disclosed attributes directly to a

PDP [18]. **OT zero trust** must reconcile verification with determinism and safety [37], **post-quantum identity** must fit existing size and latency budgets [44], and **formal verification** of composed identity-, network-, and application-tier policies remains open.

XI. CONCLUSION

Zero trust has matured from a slogan into a standards-based architecture with federal mandates, reference implementations, and maturity models. Its insight is simple: trust is earned for each request from verifiable evidence about identity, device, and context, never inherited from network location. The attacks that dominate breach statistics target exactly the identity layer that zero trust makes central, and the defenses now exist as open standards: phishing-resistant authenticators, sender-constrained tokens, shared signals, attested workload identity, and policy as code. What remains is execution—closing coverage gaps, guarding identity infrastructure, extending the model to non-human identities and AI agents, and measuring outcomes—so that one stolen credential no longer opens the whole enterprise.

REFERENCES

[1] Verizon Business, "2026 Data Breach Investigations Report," Verizon, May 2026.

[2] J. Blount, "Threats to critical infrastructure: Examining the Colonial Pipeline cyber attack," testimony before the U.S. Senate Committee on Homeland Security and Governmental Affairs, Washington, DC, USA, Jun. 8, 2021.

[3] Mandiant (Google Cloud), "UNC5537 targets Snowflake customer instances for data theft and extortion," Google Cloud Threat Intelligence Blog, Jun. 2024.

[4] Google Threat Intelligence Group, "Widespread data theft targets Salesforce instances via Salesloft Drift," Google Cloud Threat Intelligence Blog, Aug. 2025.

[5] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero trust architecture," National Institute of Standards and Technology, Gaithersburg, MD, USA, NIST SP 800-207, Aug. 2020.

[6] Jericho Forum, "Jericho Forum commandments," The Open Group, version 1.2, May 2007.

[7] J. Kindervag, "No more chewy centers: Introducing the zero trust model of information security," Forrester Research, Cambridge, MA, USA, Sep. 2010.

[8] R. Ward and B. Beyer, "BeyondCorp: A new approach to enterprise security," *;login:*, vol. 39, no. 6, pp. 6–11, Dec. 2014.

[9] B. Osborn, J. McWilliams, B. Beyer, and M. Saltonstall, "BeyondCorp: Design to deployment at Google," *;login:*, vol. 41, no. 1, pp. 28–34, 2016.

[10] Cloud Security Alliance, "Software-defined perimeter (SDP) specification v2.0," CSA Zero Trust Working Group, Mar. 2022.

[11] Executive Office of the President, "Improving the nation's cybersecurity," Executive Order 14028, 86 *Fed. Reg.* 26633, May 2021.

[12] S. D. Young, "Moving the U.S. government toward zero trust cybersecurity principles," Office of Management and Budget, Memorandum M-22-09, Jan. 26, 2022.

[13] National Cyber Security Centre (NCSC), "Zero trust architecture design principles," London, U.K., Jul. 2021.

[14] U.S. Department of Defense, Office of the Chief Information Officer, "DoD zero trust strategy," Washington, DC, USA, Nov. 2022.

[15] Cybersecurity and Infrastructure Security Agency (CISA), "Zero trust maturity model," version 2.0, Apr. 2023.

[16] R. Chandramouli and Z. Butcher, "A zero trust architecture model for access control in cloud-native applications in multi-cloud environments," NIST SP 800-207A, Sep. 2023.

[17] National Cybersecurity Center of Excellence (NCCoE), "Implementing a zero trust architecture," NIST SP 1800-35, Jun. 2025.

[18] D. Temoshok *et al.*, "Digital identity guidelines," NIST SP 800-63-4, Jul. 2025.

[19] OpenID Foundation, "OpenID Shared Signals Framework specification 1.0," Final Specification, 2025.

[20] OpenID Foundation, "OpenID Continuous Access Evaluation Profile 1.0," Final Specification, 2025.

[21] National Security Agency (NSA), "Zero trust implementation guideline (ZIG): Primer, Discovery Phase, Phase One, and Phase Two," Cybersecurity Technical Reports, Jan. 2026.

[22] World Wide Web Consortium (W3C), “Web Authentication: An API for accessing public key credentials, Level 3,” W3C Recommendation, Aug. 2026.

[23] D. Temoshok *et al.*, “Digital identity guidelines: Authentication and authenticator management,” NIST SP 800-63B-4, Jul. 2025.

[24] Cybersecurity and Infrastructure Security Agency (CISA), “Implementing phishing-resistant MFA,” fact sheet, Oct. 2022.

[25] Uber Technologies, “Security update,” Uber Newsroom, Sep. 2022.

[26] B. Campbell, J. Bradley, N. Sakimura, and T. Lodderstedt, “OAuth 2.0 mutual-TLS client authentication and certificate-bound access tokens,” IETF RFC 8705, Feb. 2020.

[27] D. Fett, B. Campbell, J. Bradley, T. Lodderstedt, M. Jones, and D. Waite, “OAuth 2.0 demonstrating proof of possession (DPoP),” IETF RFC 9449, Sep. 2023.

[28] T. Lodderstedt, J. Bradley, A. Labunets, and D. Fett, “Best current practice for OAuth 2.0 security,” IETF RFC 9700 (BCP 240), Jan. 2025.

[29] V. C. Hu *et al.*, “Guide to attribute based access control (ABAC) definition and considerations,” NIST SP 800-162, Jan. 2014.

[30] R. Pang *et al.*, “Zanzibar: Google’s consistent, global authorization system,” in *Proc. USENIX Annu. Tech. Conf. (ATC)*, Renton, WA, USA, 2019, pp. 33–46.

[31] J. W. Cutler *et al.*, “Cedar: A new language for expressive, fast, safe, and analyzable authorization,” *Proc. ACM Program. Lang.*, vol. 8, no. OOPSLA1, 2024.

[32] N. MacDonald, L. Orans, and J. Skorupa, “The future of network security is in the cloud,” Gartner Research, Aug. 2019.

[33] D. Feldman *et al.*, *Solving the Bottom Turtle: A SPIFFE Way to Establishing Trust in Your Infrastructure via Universal Identity*. SPIFFE Project, Cloud Native Computing Foundation, 2020.

[34] M. Jones, A. Nadalin, B. Campbell, J. Bradley, and C. Mortimore, “OAuth 2.0 token exchange,” IETF RFC 8693, Jan. 2020.

[35] Model Context Protocol, “Authorization,” MCP Specification, revision 2025-06-18, Jun. 2025.

[36] National Cybersecurity Center of Excellence (NCCoE), “Accelerating the adoption of software and artificial intelligence agent identity and authorization,” concept paper (initial public draft), Feb. 2026.

[37] CISA, U.S. Department of War, U.S. Department of Energy, FBI, and U.S. Department of State, “Adapting zero trust principles to operational technology,” joint guide, Apr. 2026.

[38] The MITRE Corporation, “MITRE ATT&CK: Enterprise matrix,” 2026. [Online]. Available: https://attack.mitre.org

[39] A. Witty, “Hacking America’s health care: Assessing the Change Healthcare cyber attack and what’s next,” testimony before the U.S. Senate Committee on Finance, Washington, DC, USA, May 1, 2024.

[40] Microsoft Security Response Center, “Microsoft actions following attack by nation state actor Midnight Blizzard,” MSRC Blog, Jan. 2024.

[41] CISA and FBI, “Scattered Spider,” Joint Cybersecurity Advisory AA23-320A, Nov. 2023.

[42] Cyber Safety Review Board, “Review of the Summer 2023 Microsoft Exchange Online intrusion,” U.S. Department of Homeland Security, Mar. 2024.

[43] S. Reiner, “Golden SAML: Newly discovered attack technique forges authentication to cloud apps,” CyberArk Threat Research Blog, Nov. 2017.