International Journal of AI, Big Data, Computational and Management Studies
Noble Scholar Research Group | Volume 7, Issue 2, 70-75, 2026
ISSN: 3050-9416 | https://doi.org/10.63282/3050-9416.IJAIBDCMS-V7I2P112

Original Article

Architecting Large-Scale Identity Governance Frameworks for Zero Trust Enterprises

Thrinaath Vajragowni
Independent Researcher / Identity Security Architect.

Received On: 19/02/2026 ***Revised On: 28/03/2026*** ***Accepted On: 06/04/2026*** ***Published On: 13/04/2026***

Abstract: *Identity governance has emerged as the foundational control plane of modern enterprise security. As organizations increasingly operate across hybrid and multi-cloud environments, traditional perimeter-based security models have proven inadequate. This paper examines the architectural principles, design patterns, and operational frameworks required to implement large-scale Identity Governance and Administration (IGA) within a Zero Trust paradigm. We analyze the convergence of identity lifecycle management, least-privilege enforcement, continuous verification, and AI-augmented risk intelligence. Drawing on current industry frameworks including NIST SP 800-207 and NIST SP 800-63 [8], we propose a five-layer reference architecture for scalable, policy-driven identity governance that spans human and non-human identities across enterprise-scale deployments. Empirical data from industry research underscores the urgency of this transition and informs our proposed design guidance.*

Keywords: *Identity Governance and Administration, Zero Trust Architecture, Identity Lifecycle Management, Least Privilege, Privileged Access Management, NIST SP 800-207, Role-Based Access Control, Continuous Verification, Identity Threat Detection and Response, Policy-as-Code.*

1. Introduction

The dissolution of the traditional network perimeter - accelerated by widespread cloud adoption, distributed remote work, and the proliferation of SaaS applications - has fundamentally redefined the enterprise security landscape. What was once a defensible border, enforced through firewalls and VPNs, has given way to an environment in which resources are accessed from any device, any location, and through an expanding portfolio of identity types. In this context, identity has become the new security perimeter. Every user account, service account, API token, and device certificate represents a potential attack vector; every access grant is, in effect, a trust decision. Identity Governance and Administration (IGA) encompasses the policies, workflows, and technology platforms that determine who has access to what resources, how that access is granted and reviewed, and how compliance with regulatory mandates is maintained over time. At enterprise scale - where organizations may manage hundreds of thousands of identities across dozens of cloud platforms and legacy systems - IGA must be architected with the same engineering rigor applied to network and application infrastructure.

The Identity Defined Security Alliance (IDSA) reported that in 2024, 90% of organizations experienced an identity-related breach, with 84% indicating it was directly tied to inadequate access governance [1]. IBM's Cost of a Data Breach Report 2024 documented that compromised credentials remain the most common initial attack vector, contributing to breaches averaging $4.88 million in cost [2]. These figures substantiate what security architects have long understood: identity governance is not a peripheral compliance function but a core operational security discipline. Zero Trust Architecture (ZTA), as formally defined by NIST SP 800-207, establishes that no user, device, or network segment should be implicitly trusted. Every access request must be continuously authenticated, authorized, and encrypted regardless of its origin [3]. This paradigm elevates identity governance from a periodic review exercise to a real-time security function, demanding that IGA platforms support contextual, dynamic access decisions rather than static role assignments reviewed once per quarter. The intersection of IGA and ZTA creates what this paper terms the governance-verification gap: a temporal window during which entitlements exist in the IGA system that are no longer appropriate, but have not yet been identified through the next certification cycle. Research by Gartner indicates the average enterprise takes 23 days to deprovision a departing employee's access, and over 40% of organizations have experienced unauthorized access via residual entitlements [4]. In a Zero Trust environment, this gap is not merely a compliance risk; it is a persistent open door. This paper makes the following contributions: (1) a five-layer reference architecture for enterprise IGA-ZT convergence; (2) a formal access control policy model for continuous evaluation; (3) pseudocode for the core continuous access evaluation algorithm; (4) design patterns for JIT privileged access and Policy-as-Code; and (5) a regulatory alignment matrix mapping IGA-ZT to key compliance frameworks.

2. Background and Related Work

2.1. Evolution of Identity Governance

Early IAM implementations focused on directory synchronization and single sign-on (SSO). Role-Based Access Control (RBAC), formalized by Sandhu et al. [5], enabled coarse-grained entitlement management organized around organizational roles. However, RBAC proved insufficient as application portfolios diversified and access requirements grew contextually complex. Attribute-Based Access Control (ABAC) introduced contextual dimensions - role, department, location, device health - into access decisions. Modern IGA platforms synthesize RBAC, ABAC, and Policy-Based Access Control (PBAC) into unified entitlement engines capable of enforcing complex, contextual policies at scale.

Cloud computing further fragmented the identity control plane across AWS IAM, Azure Entra ID, GCP IAM, and dozens of SaaS-specific provisioning systems. Cloud Infrastructure Entitlement Management (CIEM) emerged as a discipline focused specifically on discovering and right-sizing the excessive permissions endemic to cloud identity configurations, where overly permissive IAM policies create vast, often unrecognized attack surfaces.

2.2. Zero Trust Architecture Principles

The Zero Trust model, originally articulated by John Kindervag at Forrester Research in 2010 and formalized by NIST in SP 800-207, rests on three foundational tenets: (1) all resources must be accessed securely regardless of network location; (2) access is granted through dynamic policy incorporating as many data sources as feasible; and (3) the enterprise continuously monitors and measures the integrity of all assets and infrastructure [3]. NIST further identifies seven operational tenets of ZTA, including requirements that all communication be secured, that access to individual resources be granted on a per-session basis, and that access decisions incorporate real-time asset state and context signals. Operating a Zero Trust architecture demands that identity infrastructure consume rich contextual signals - device compliance status, network path, geographic location, behavioral baselines, threat intelligence - and render access decisions against this context in real time. This is categorically different from static role assignment, and requires IGA platforms to operate as dynamic policy engines rather than administrative record systems.

2.3. The Governance-Verification Gap

Despite widespread adoption of IGA platforms and Zero Trust frameworks as separate initiatives, organizations frequently fail to integrate them into a coherent whole. IGA platforms were designed for periodic governance cycles: access certifications conducted quarterly, provisioning workflows triggered by HR events, role reviews conducted annually. Zero Trust demands continuous, per-request verification. This temporal mismatch - periodic governance applied to a continuously operating environment - is the governance-verification gap. Adversaries targeting organizations with this gap have a reliable exploitation window between governance cycles during which compromised or excessive entitlements remain active.

3. Proposed Architecture: Iga-Zt Reference Model

3.1. Five-Layer Architecture

We propose a five-layer Identity Governance Architecture for Zero Trust (IGA-ZT). The architecture provides logical partitioning of governance functions enabling independent scaling, technology substitution within layers, and clear integration contracts between components.

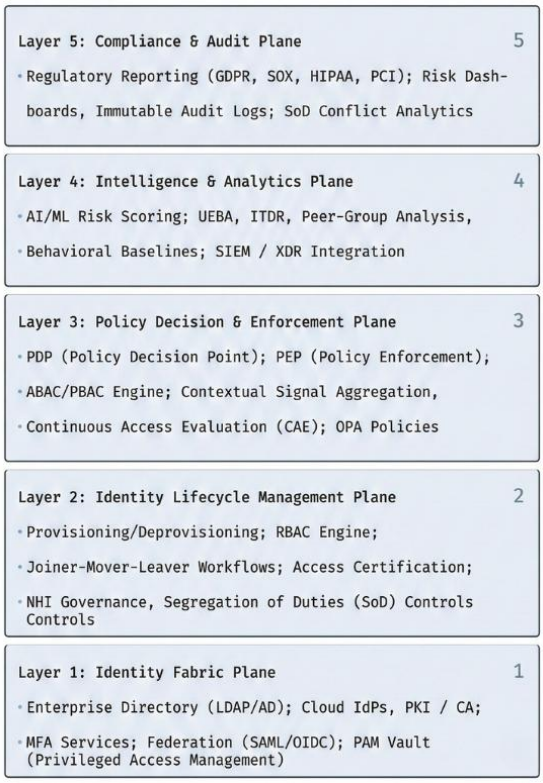

Fig 1: IGA-ZT Reference Architecture

- **Layer 1 - Identity Fabric** provides the authoritative identity store and authentication services. This includes enterprise directories (Active Directory, LDAP), cloud-native Identity Providers (Azure Entra ID, Okta, Ping Identity), PKI infrastructure for certificate-based identities, MFA services, and PAM vaults for secrets storage. This layer answers the fundamental question: *who is this entity, and can we verify that claim cryptographically?*
- **Layer 2 - Identity Lifecycle Management** governs the complete span of an identity's existence. It automates the Joiner-Mover-Leaver (JML) process: provisioning access at onboarding based on role, updating entitlements when employees change positions, and revoking access reliably upon separation. This layer also governs non-human identities - service accounts, API keys, machine certificates, RPA bot credentials - applying the same lifecycle rigor to NHIs as to human users.

Segregation of Duties (SoD) controls are enforced at this layer to prevent conflicting entitlement combinations that could enable fraud.

- **Layer 3 - Policy Decision & Enforcement** implements the real-time access evaluation engine that is the operational heart of Zero Trust. The Policy Decision Point (PDP) evaluates incoming access requests against policies defined in the IGA system, consuming contextual signals from Layer 4 including device posture attestation, geolocation, and behavioral risk scores. The Policy Enforcement Point (PEP) enforces the PDP's decision at each resource boundary. This directly maps to the ZTA logical components defined in NIST SP 800-207, where the Policy Engine and Policy Administrator form the control plane of the Zero Trust system [3].
- **Layer 4 - Intelligence & Analytics** provides continuous monitoring and AI-augmented risk scoring. User and Entity Behavior Analytics (UEBA) establish behavioral baselines for each identity and detects deviations indicative of compromise or insider threat. Identity Threat Detection and Response (ITDR) capabilities correlate identity signals with broader threat intelligence to identify attack patterns such as credential stuffing, pass-the-hash, and token replay. SIEM and XDR integrations ensure identity signals are available to the broader security operations center.
- **Layer 5 - Compliance & Audit** provides the management and reporting capabilities necessary to demonstrate regulatory compliance. Immutable audit logs capture all access events, governance decisions, and policy changes. Automated compliance reporting generates evidence packages for SOX, HIPAA, GDPR, and PCI DSS. Risk dashboards provide real-time visibility into IGA health metrics including orphaned accounts, over-privileged identities, and pending certification actions.

3.2. Joiner-Mover-Leaver with Continuous Verification

The operational backbone of Layer 2 is the JML process, augmented with continuous verification loops that ensure access appropriateness throughout the identity lifecycle - not just at provisioning time.

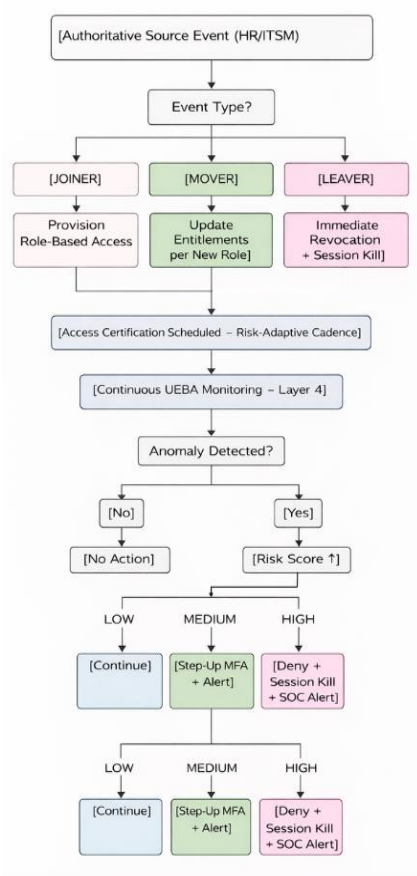

Fig 2: JML with Continuous Verification Loop

3.3. Access Control Policy Model

The policy engine within Layer 3 implements a hybrid RBAC-ABAC-PBAC model. Access decisions are computed as a function of the requesting identity's verified attributes, their assigned roles, environmental context at the moment of access, and the AI-computed risk score for the current session:

```
Access Decision = f(Identity, Role, Attributes, Context, Risk, Policy)
Variables:
  Identity   = Cryptographically verified principal (human or NHI)
  Role       = Role(s) assigned from authoritative HR/ITSM source
  Attributes = {job_function, department, clearance_level,
                employment_type, project_membership}
  Context    = {device_health_score, geo_location, network_path,
                time_of_day, session_history, ip_reputation}
  Risk Score = AI-computed session risk index ∈ [0.0, 1.0]
  Policy     = Resource-specific access policy expressions
Decision Logic:
  IF session_token is expired → DENY
  IF Risk ≥ 0.75 (HIGH)  → DENY + SOC alert
  IF Risk ≥ 0.45 (MEDIUM)  → STEP_UP (require additional factor)
  IF Role ∩ required_roles ≠ ∅
    AND Attributes ⊆ policy.constraints
      AND Context satisfies policy.conditions → PERMIT
```

```
  ELSE                              → DENY
```

3.4. Pseudocode: Continuous Access Evaluation (CAE)

The following algorithm implements the real-time access evaluation loop forming the operational core of Layer 3. It is invoked on every access request and periodically during active sessions to reflect changing conditions - device posture, risk score escalation, policy updates.

Algorithm 1: Continuous Access Evaluation (CAE)

```
INPUT:   identity_id, resource_id, session_token,
context_vector
OUTPUT: access_decision ∈ {PERMIT, DENY, STEP_UP}
CONSTANTS:
  HIGH_RISK   = 0.75
  MEDIUM_RISK = 0.45
  REVIEW_SEC   =  300     //  Re-evaluate  every  5
minutes
BEGIN
  // Phase 1: Token and Identity Validation
  IF session_token.is_expired():
      audit_log(identity_id,           resource_id,
"SESSION_EXPIRED")
      RETURN DENY
  identity ← resolve_identity(identity_id)
  IF identity.status IN {SUSPENDED, TERMINATED}:
      revoke_all_sessions(identity_id)
      RETURN DENY
  // Phase 2: Role and Attribute Resolution
  roles ← get_current_roles(identity_id)
  attrs ← get_identity_attributes(identity_id)

  // Phase 3: Risk Computation
  risk ← compute_risk_score(
      identity        = identity,
      context         = context_vector,
      behavioral_model                          =
load_baseline(identity_id),
      threat_intel                              =
fetch_threat_context(context_vector.ip)
  )
  // Phase 4: Risk-Adaptive Decision
  IF risk ≥ HIGH_RISK:
      alert_soc(identity_id, resource_id, risk)
      audit_log(identity_id,           resource_id,
"HIGH_RISK_DENIED")
      RETURN DENY
  IF risk ≥ MEDIUM_RISK:
      RETURN STEP_UP
  // Phase 5: Policy and SoD Evaluation
  policy ← load_resource_policy(resource_id)
  IF   NOT   evaluate_entitlements(roles,   attrs,
policy).grants_access():
      RETURN DENY
  IF                          check_sod(identity_id,
resource_id).has_conflicts():
      alert_compliance_team(identity_id)
      RETURN DENY
  // Phase 6: Issue Permit with Monitored Session
  session ← create_monitored_session(
      identity_id  = identity_id,
      resource_id  = resource_id,
      risk_score   = risk,
      review_timer = REVIEW_SEC
  )
  audit_log(identity_id, resource_id, "PERMITTED",
session.id)
  RETURN PERMIT
END
```

4. Key Design Patterns

4.1. Just-In-Time (JIT) Privileged Access

The JIT pattern eliminates standing privileged access, replacing persistent elevated credentials with dynamically issued, time-bounded, purpose-scoped access grants. A privileged user requiring administrative access must explicitly request it, justify the need, obtain approval through a defined workflow, and receive a temporary credential expiring at session end. Following the session, the PAM vault rotates the underlying secret and archives the session recording for audit.

Table 1: Standing Privileged Access vs. Just-In-Time Access

Metric	Standing Access	JIT Access
Attack Surface	Persistent credential always active	Credential exists only during session window
Breach Blast Radius	Large - attacker inherits standing privileges	Constrained to time-bounded session
Auditability	Requires manual event correlation	Automated per-request logging with justification chain
SoD Enforcement	Difficult - static grants bypass SoD checks	Enforced at each JIT request evaluation
Lateral Movement Risk	High - credential usable indefinitely	Low - credential expires before movement completes
Compliance Posture	Difficult to demonstrate least-privilege	Provable least-privilege by design

4.2. Policy-as-Code (PaC)

Policy-as-Code transforms identity governance policies from prose documents and GUI configurations into machine-readable, version-controlled, testable artifacts. Open Policy Agent (OPA), using the Rego policy language [9], enables governance teams to express access rules as code tested in CI/CD pipelines, peer-reviewed through pull requests, and deployed consistently across cloud and on-premises enforcement points. This approach eliminates the policy drift that occurs when policies are managed through multiple disconnected administrative interfaces, ensuring that governance intent is universally and uniformly enforced.

A sample Rego policy for a financial reporting API enforces role, time, device compliance, and SoD constraints simultaneously:

```
package finance.reports.read
default allow = false
allow {
    input.identity.roles[_] == "finance_analyst"
    time.clock(time.now_ns())[0]  >=  8           #
Business hours UTC
    time.clock(time.now_ns())[0] <= 18
    input.context.device_managed == true
    input.context.device_compliance_score >= 80
    count(input.identity.sod_conflicts) == 0
}
allow {
```

```
    input.identity.roles[_] == "finance_manager"
    input.context.device_managed == true
    count(input.identity.sod_conflicts) == 0
}
```

4.3. Identity Threat Detection and Response (ITDR)

ITDR converges identity-centric behavioral analytics with security incident response. Where traditional SIEM approaches analyze logs from network and endpoint sources, ITDR focuses specifically on identity signals: authentication events, entitlement changes, access pattern deviations, and privileged session activities. Key detection scenarios include:

- Impossible travel: Authentication from geographically distant locations within a timeframe physically impossible for travel, indicating credential compromise
- Privilege escalation anomalies: Requests for elevated privileges outside the identity's normal operational pattern or approved JIT workflows
- Bulk data access: Unusually large volume of records accessed in a compressed timeframe, consistent with exfiltration
- Dormant account activation: Sudden activity from historically inactive accounts, a common indicator of orphaned credential exploitation
- Token replay attacks: Reuse of authentication tokens beyond their intended scope or after revocation.

5. AI Augmentation in Enterprise Iga

5.1. AI-Driven Risk Scoring

Modern IGA platforms incorporate ML models trained on historical identity and access telemetry to compute continuous risk scores for identity sessions. Feature vectors include time-of-day access patterns, resource access frequency, peer group behavioral baselines, device and network attributes, and threat intelligence feeds. Industry research indicates organizations leveraging AI-augmented IAM reduce mean time to detect identity incidents by approximately 62% and reduce false positive alert rates by up to 45%, compared to rule-based approaches [6]. The precision improvement stems from the models' capacity to distinguish genuine anomalies from ordinary behavioral variation.

5.2. Peer Group Access Analytics and Right-Sizing

AI-driven peer analysis compares each identity's entitlements against a dynamically computed peer group - identities with comparable roles, departments, seniority, and location. Entitlements significantly exceeding peer baselines are flagged for review and potential right-sizing. The model may also identify access patterns common to the peer group that the subject identity lacks, generating positive access recommendations. This approach accelerates role engineering and enables continuous least-privilege optimization [10] without requiring manual entitlement analysis at scale.

```
FOR EACH identity IN enterprise_directory:
  peer_group ← compute_peer_group(role, dept, level, location)
  peer_common ← entitlements present in ≥ 80% of peer_group
  current ← get_entitlements(identity)
  // Detect excess entitlements
  excess ← current - union(peer_group entitlements)
  FOR EACH e IN excess:
    IF last_used(e) > DORMANT_THRESHOLD OR access_count(e) == 0:
      flag_for_revocation(identity, e, "Excess + dormant")
    ELSE:
      flag_for_review(identity, e, "Excess but active")
  // Positive recommendations
  missing ← peer_common - current
  IF |missing| > 0:
    recommend_access(identity, missing, "Peer pattern match")
```

Algorithm 2: Peer-Group Access Right-Sizing

5.3. Autonomous Access Certification

Traditional access certifications present reviewers with exhaustive entitlement lists, producing reviewer fatigue and rubber-stamp approval rates exceeding 85% regardless of actual appropriateness [7]. AI-augmented certification pre-screens entitlements: automatically certifying those meeting defined criteria (active usage, within peer norms, low risk) and surfacing only genuinely anomalous or high-risk items for human review. This shifts reviewer focus from volume to quality, transforming certification from a documentation exercise into a substantive governance control.

6. Implementation Roadmap and Regulatory Alignment

6.1. Phased Deployment Model

Large-scale IGA-ZT deployments should follow a phased model to manage complexity and organizational change:

- Phase 1 - Discovery and Normalization (Months 1-3): Conduct comprehensive identity discovery across all target systems, producing a normalized inventory of all human and non-human identities, their owning systems, current entitlements, and last-used dates.
- Phase 2 - Role Engineering and RBAC Baseline (Months 3-6): Define a clean, right-sized role model from first principles - derived from job function and organizational structure, not inherited from accumulated historical entitlements.
- Phase 3 - Lifecycle Automation and JML Integration (Months 6-9): Connect IGA to authoritative HR and ITSM sources. Automate JML workflows. Instrument SoD controls across all in-scope applications.
- Phase 4 - Continuous Monitoring Integration (Months 9-12): Integrate IGA with SIEM and UEBA platforms. Activate ITDR capabilities. Establish SOC runbooks for identity incident response.
- Phase 5 - AI Enablement and ZT Policy Integration (Months 12-18): Activate AI-driven risk scoring, peer analysis, and autonomous certification. Connect the IGA policy engine to ZT PDP/PEP infrastructure for real-time, contextual access evaluation.

6.2. Regulatory Alignment Matrix

Table 2: Regulatory Alignment of IGA-ZT Components

Regulation	Primary Layer(s)	Key Requirement	IGA-ZT Control
SOX	Layers 2, 5	Segregation of duties; access audit trails	SoD engine; immutable audit logs
HIPAA	Layers 2-5	Minimum necessary access; workforce controls; audit	RBAC least-privilege; UEBA; compliance reporting
GDPR	Layers 2, 5	Data minimization; right of erasure; accountability	Lifecycle automation; identity deletion workflows
PCI DSS	Layers 2-4	Quarterly access reviews; MFA; session logging	Access certification; MFA enforcement
NIST CSF	All Layers	Identify, Protect, Detect, Respond, Recover	Full five-layer coverage
FedRAMP	Layers 1-3	Continuous monitoring; least privilege; MFA	CAE; JIT access; PIV/MFA enforcement

7. Limitations and Open Challenges

The effectiveness of AI risk scoring models is contingent on the quality and volume of historical identity telemetry. Organizations with immature logging infrastructure or significant identity estate blind spots will produce less accurate risk scores, potentially generating both false positives - unnecessary friction for legitimate users - and false negatives - missed anomalies that adversaries exploit. Addressing this requires investment in logging completeness before AI-driven scoring is activated. Integration of IGA with Zero Trust policy infrastructure remains technically complex in heterogeneous environments. Most legacy applications do not expose policy enforcement points compatible with modern ZTA architectures, requiring custom integration work or the deployment of proxy-based enforcement solutions that add latency and operational complexity. The governance of non-human identities at scale - particularly short-lived cloud workload identities, CI/CD pipeline credentials, and AI agent sessions - continues to outpace the maturity of available tooling. Vendors and standards bodies continue to work toward interoperable NHI governance solutions, but organizations implementing IGA-ZT today should expect significant bespoke engineering requirements in this area.

8. Conclusion

The convergence of Identity Governance and Administration with Zero Trust Architecture represents a fundamental transformation in enterprise security posture. As organizations manage identity populations spanning employees, contractors, service accounts, cloud workload identities, and increasingly autonomous AI agents, the governance frameworks of the perimeter era are structurally inadequate for the demands of the modern distributed environment. The IGA-ZT reference architecture proposed in this paper provides a principled foundation for this transformation. By grounding identity governance in a five-layer model connecting authoritative identity fabric through lifecycle management, dynamic policy enforcement, AI-augmented intelligence, and compliance automation, enterprises can build a governance posture that is simultaneously rigorous, adaptive, and auditable. The design patterns discussed - JIT privileged access, Policy-as-Code, ITDR, AI-driven right-sizing, and autonomous certification - are not theoretical constructs but operational capabilities being implemented by security-mature organizations today. Closing the governance-verification gap - by integrating the IGA policy engine with real-time PDP/PEP infrastructure and enabling continuous AI-driven certification - transforms identity governance from a quarterly compliance exercise into a dynamically operating security control. As identity attack surfaces continue to expand and sophisticated adversaries increasingly target identity infrastructure as their primary means of enterprise penetration, this architectural transformation will prove one of the most consequential security investments of the decade.

References

1. Identity Defined Security Alliance, "2024 Trends in Securing Digital Identities," IDSA, Denver, CO, USA, 2024. [Online]. Available: https://www.idsalliance.org
2. IBM Security, "Cost of a Data Breach Report 2024," IBM Corporation, Armonk, NY, USA, 2024. [Online]. Available: https://www.ibm.com/reports/data-breach
3. National Institute of Standards and Technology, "Zero Trust Architecture," NIST Special Publication 800-207, Aug. 2020. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-207
4. Gartner, "Predicts 2025: Identity and Access Management," Gartner Research Note G00800432, Stamford, CT, USA, 2024.
5. R. Sandhu, E. Coyne, H. Feinstein, and C. Youman, "Role-based access control models," *IEEE Computer*, vol. 29, no. 2, pp. 38-47, Feb. 1996.
6. Avatier Corporation, "The ROI of AI-Augmented Identity Management," Avatier, Pleasanton, CA, USA, 2024.
7. SailPoint Technologies, "The Convergence of Identity Governance and Zero Trust," SailPoint White Paper, Austin, TX, USA, 2024.
8. National Institute of Standards and Technology, "Digital Identity Guidelines," NIST Special Publication 800-63-3, Jun. 2017. [Online]. Available: https://doi.org/10.6028/NIST.SP.800-63-3
9. Open Policy Agent Project, "OPA Rego Language Reference," CNCF, 2024. [Online]. Available: https://www.openpolicyagent.org/docs/latest/policy-language/
10. M. Saltzer and M. Schroeder, "The protection of information in computer systems," Proceedings of the IEEE, vol. 63, no. 9, pp. 1278-1308, Sep. 1975.