

Zero Trust Principles for Non-Human Identities and Agentic Automation in Distributed Cloud and Hybrid Infrastructures

Rohit Agnihotri

Abstract- The proliferation of agentic automation, microservice workloads, and API-first architectures has produced an identity ecosystem where non-human identities (NHIs), such as service accounts, API tokens, machine certificates, CI/CD pipeline credentials, and autonomous AI agents, now outnumber human identities by an average ratio of 45:1 in enterprise environments. Conventional perimeter-centric and even role-based access control frameworks were designed for human users operating within bounded organizational boundaries, and they are structurally ill-equipped to govern the dynamic, ephemeral, and machine-speed behavior of NHIs operating across distributed cloud and hybrid infrastructures. This article presents a comprehensive analysis of Zero Trust (ZT) principles as applied to NHI governance and agentic automation security. We synthesize the technical architecture of NHI-aware Zero Trust frameworks, evaluate empirical performance and breach cost data, systematize the principal attack vectors targeting NHIs, and propose an original Eight-Pillar ZT-NHI Governance Model. Our analysis demonstrates that fully deployed ZT frameworks with AI-driven automation reduce average breach costs from USD 5.28 million to USD 1.76 million (a 67% reduction), while identity-centric micro-segmentation reduces lateral movement success rates by 84%. We further identify critical capability gaps in current ZT deployments, particularly in the automation and orchestration pillar, and articulate a forward-looking research agenda addressing the unique governance challenges of autonomous AI agent identities.

Keywords— Zero Trust Architecture; Non-Human Identities; Agentic Automation; Cloud Security; Hybrid Infrastructure; Machine Identity; SPIFFE/SPIRE; Least Privilege; Service Mesh; Identity Federation

I. INTRODUCTION

The architecture of enterprise computing has undergone a fundamental transformation over the past decade. The dissolution of the traditional network perimeter, driven by the adoption of multi-cloud platforms, containerized workloads, edge computing, and globally distributed remote workforces, has rendered conventional castle-and-moat security paradigms structurally obsolete (Rose et al., 2020). In their place, the Zero Trust Architecture (ZTA) model, formally codified in NIST Special Publication 800-207 (NIST, 2020), has emerged as the dominant security framework for distributed environments. Its foundational tenet (that no actor, system, or network segment should be granted implicit trust by virtue of its location) provides the conceptual scaffolding for identity-centric security in a perimeterless world.

Yet even as Zero Trust adoption has accelerated, a critical blind spot has emerged in both research and practice: the identity ecosystem has undergone a qualitative transformation that most ZT frameworks have not adequately addressed. The contemporary enterprise does not merely manage the identities of its human employees; it governs a vastly larger and more dynamic population of non-human identities (NHIs). These include service accounts executing automated database operations, API tokens authenticating machine-to-machine integrations, X.509 certificates securing microservice communication, CI/CD pipeline agents deploying code across production environments, cloud workload instance roles, IoT device identities, and, most recently, autonomous AI agents executing multi-step tasks on behalf of human principals.

The scale of this NHI proliferation is striking. Data from the CyberArk Global Threat Landscape Report (2024) indicates that NHIs now outnumber human

identities by an average ratio of 45:1 in enterprise environments, projected to reach 210:1 by 2028. More critically, NHI credentials are implicated in 71% of breaches involving non-perimeter access vectors (Verizon, 2024), yet a survey by Gartner (2024) found that only 38% of organizations have formal governance processes covering NHI credential lifecycle management. This governance deficit creates an asymmetric risk profile: organizations invest heavily in human identity security while leaving their exponentially larger NHI attack surface largely unprotected.

Compounding this challenge is the emergence of agentic automation: AI systems capable of autonomous, goal-directed action across multiple systems, APIs, and data sources. Unlike traditional automation scripts with predictable, bounded behavior, agentic AI systems, built on large language models (LLMs) with tool-use and reasoning capabilities, can dynamically acquire resources, initiate API calls, spawn sub-agents, and modify their own operational context in pursuit of assigned objectives (Russell, 2019; Yao et al., 2023). The identity and access management implications of such systems are profound and, to date, inadequately theorized in the security literature.

This article addresses these gaps through four principal contributions: (1) a comprehensive taxonomy of NHI categories and their distinct risk profiles; (2) a systematic mapping of Zero Trust principles to NHI-specific technical controls; (3) empirical analysis of the security and operational impacts of ZT-NHI integration; and (4) an original Eight-Pillar ZT-NHI Governance Model with maturity assessment criteria. The paper is structured as follows: Section 2 reviews the literature on ZT evolution and NHI security; Section 3 develops the NHI taxonomy and threat landscape; Section 4 presents the ZT-NHI architectural framework; Section 5 analyzes empirical performance data; Section 6 examines agentic AI identity challenges specifically; Section 7 presents the governance model; and Section 8 concludes with a research agenda.

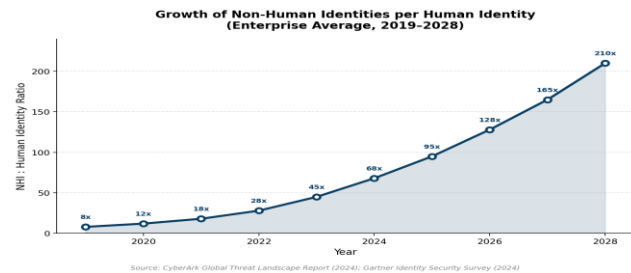


Figure 1. Growth of Non-Human Identities per Human Identity — Enterprise Average (2019–2028).
Source: CyberArk Global Threat Landscape Report (2024); Gartner Identity Security Survey (2024).

II. LITERATURE REVIEW

The Evolution of Zero Trust Architecture

The intellectual origins of Zero Trust can be traced to Kindervag's (2010) work at Forrester Research, which introduced the concept of eliminating trust from network architectures and enforcing least-privilege access at every layer. This foundational concept was operationalized through Google's BeyondCorp initiative (Ward & Beyer, 2014), which pioneered the enterprise implementation of device and user verification irrespective of network location. The model's formal standardization in NIST SP 800-207 (NIST, 2020) articulated the seven tenets of ZTA and established the Policy Enforcement Point (PEP) / Policy Decision Point (PDP) architecture that underpins contemporary implementations.

Subsequent work by Stafford (2020) extended the NIST model to cloud-native environments, while the CISA Zero Trust Maturity Model (CISA, 2023) introduced a five-pillar framework (Identity, Devices, Networks, Applications/Workloads, and Data) with defined maturity levels for each pillar. Empirical validation of ZT benefits has been provided by the IBM Cost of a Data Breach Report series (IBM Security, 2024), which has documented consistent breach cost reductions associated with increasing ZT maturity. However, both NIST and CISA frameworks focus predominantly on human identity management, with NHI governance treated as an addendum rather than a first-class concern.

2. Non-Human Identity Security: Prior Research

Research specifically addressing NHI security has emerged primarily from the practitioner community rather than academic literature. Bind et al. (2021) provided an early taxonomy of machine identities, distinguishing between ephemeral cloud workload identities and long-lived service account credentials, and identifying credential sprawl as the primary operational risk. Conjur (2023) documented the credential management challenges specific to CI/CD pipelines, noting that 65% of pipeline breaches involved compromised static credentials that had not been rotated in over 180 days.

The SPIFFE (Secure Production Identity Framework for Everyone) project and its reference implementation SPIRE (Feldman et al., 2018) represent the most significant technical advance in NHI authentication infrastructure, providing a cryptographic workload attestation system that issues short-lived X.509 SVIDs (SPIFFE Verifiable Identity Documents) based on workload properties rather than network location. Empirical evaluations by Burns et al. (2022) demonstrated that SPIFFE-based identity eliminated 94% of lateral movement opportunities compared to static service account credentials in Kubernetes environments.

3. Agentic AI and Identity Security

The intersection of agentic AI systems and identity security represents an emerging research frontier with limited prior treatment. Yao et al. (2023) introduced the ReAct framework for LLM-based

agents, which highlighted the challenge of scoping tool-use permissions for systems capable of dynamic reasoning about resource acquisition. Subsequent work by Anthropic (2024) on multi-agent orchestration noted that agent identity and trust delegation present novel attack surfaces not addressed by existing IAM frameworks.

A critical gap identified across the reviewed literature is the absence of a comprehensive ZT framework specifically designed for NHI governance that: (a) addresses the full taxonomy of NHI types including AI agents; (b) provides maturity-level guidance for incremental implementation; and (c) is grounded in empirical performance data rather than exclusively normative recommendations. The present article is designed to address this gap.

III. NON-HUMAN IDENTITY TAXONOMY AND THREAT LANDSCAPE

1. Comprehensive NHI Taxonomy

Effective Zero Trust governance for NHIs requires a precise understanding of the distinct identity categories that comprise the NHI population, their authentication mechanisms, and their specific risk profiles. Table 1 presents our proposed eight-category NHI taxonomy, synthesized from analysis of CyberArk (2024), OWASP NHI Top 10 (2024), and NIST SP 800-207 supplementary guidance.

Table 1. Comprehensive Non-Human Identity Taxonomy: Categories, Authentication Methods, and Risk Profiles

NHI Category	Examples	Authentication Method	Risk Profile
Service Accounts	Database services, OS daemons	Passwords, Kerberos tickets	High – long-lived, over-privileged
API Tokens	REST/GraphQL API keys	Bearer tokens, HMAC signatures	High – rarely rotated, widely shared
Machine Certs	TLS certificates, mTLS workloads	X.509 PKI	Medium – rotation-dependent

CI/CD Identities	Pipeline agents, build runners	Short-lived tokens	OIDC	High – privileged + ephemeral
Cloud Workloads	Lambda, containers, VMs	IAM instance roles, IMDS		Medium-High – misconfiguration risk
AI Agents	Autonomous LLM agents, RPA bots	OAuth 2.0, JWT, API keys		Critical – autonomous, goal-directed
IoT/OT Devices	Sensors, PLCs, edge nodes	Certificates, pre-shared keys		High – unmanaged, long lifecycle
Federated Services	Microservices, service mesh	SPIFFE SVIDs, mTLS		Medium – scope-bounded

Source: Synthesized from CyberArk Global Threat Landscape Report (2024); OWASP Non-Human Identities Top 10 (2024); NIST SP 800-207 (2020).

NHI Attack Vector Distribution

Understanding the distribution of attack vectors targeting NHIs is essential for prioritizing ZT control investments. Figure 2 presents the distribution of NHI-targeted attacks documented in the 2023–2024 period, based on data from the Verizon Data Breach Investigations Report (2024) and the OWASP NHI Top 10 project.

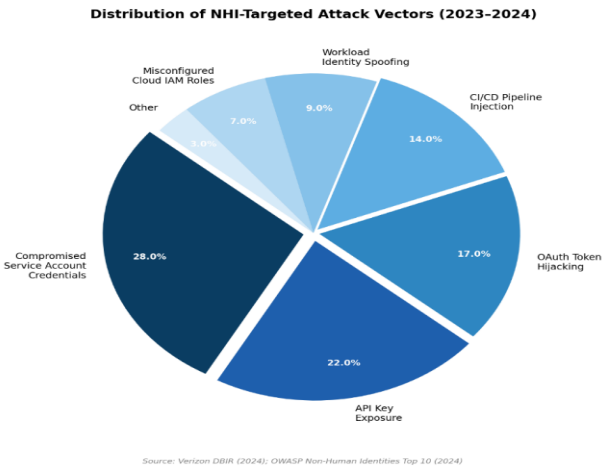


Figure 2. Distribution of NHI-Targeted Attack Vectors (2023–2024). Source: Verizon DBIR (2024); OWASP Non-Human Identities Top 10 (2024).

Compromised service account credentials constitute the largest attack vector at 28%, reflecting the persistent challenge of managing long-lived, over-

privileged service account passwords in Active Directory and cloud environments. API key exposure (22%) represents the second-largest vector, driven primarily by inadvertent secret exposure in public code repositories, a problem that GitHub's 2024 Secret Scanning report estimated affected 1 in 14 public repositories. OAuth token hijacking (17%) and CI/CD pipeline injection (14%) collectively account for nearly one-third of NHI-targeted attacks, underscoring the critical importance of securing the software supply chain and OAuth token lifecycle management (OWASP NHI Top 10, 2024). Workload identity spoofing (9%) and misconfigured cloud IAM roles (7%) represent infrastructure-layer vulnerabilities that are particularly prevalent in multi-cloud environments where identity federation configurations introduce cross-platform trust relationships that are difficult to audit comprehensively (Brewer, 2010).

IV. MAPPING ZT PRINCIPLES TO NHI CONTROLS

The seven tenets of NIST SP 800-207 provide the normative foundation for ZT implementation, but their application to NHI contexts requires technical translation into specific control mechanisms. Table 2 presents a systematic mapping of Zero Trust principles to NHI-specific technical implementations, with corresponding standards and framework references.

Table 2. Zero Trust Principles Mapped to NHI Technical Controls and Standards

ZT Principle	NHI Implementation	Standards / Frameworks
Verify Explicitly	Continuous workload attestation via SPIFFE SVIDs; mTLS everywhere; certificate-bound tokens	NIST SP 800-207; SPIFFE/SPIRE; RFC 8705
Least-Privilege Access	Just-in-time ephemeral credentials; automated entitlement reviews; scoped OAuth scopes	CIS Controls v8; NIST SP 800-53 AC-6
Assume Breach	Micro-segmentation of workload communication; eBPF-based runtime policy; anomaly detection on NHI behavior	MITRE ATT&CK for Cloud; CISA ZT Model
Continuous Verification	Real-time risk scoring on all API calls; behavioral baselines per NHI; ML-driven anomaly flagging	ISO/IEC 27001:2022; NIST CSF 2.0
Micro-segmentation	Service mesh identity-aware network policies (Istio, Linkerd); namespace isolation in Kubernetes	NIST SP 800-204; CIS Kubernetes Benchmark
Explicit Trust Boundaries	Workload identity federation; cross-cloud trust anchors via OIDC; no implicit lateral trust	NIST SP 800-207A; AWS/Azure/GCP WIF
Session & Token Governance	Short-lived credential issuance (TTL < 15 min); automatic revocation on anomaly; token binding	RFC 9449; OAuth 2.1; OWASP NHI Top 10
Auditability & Observability	Immutable, structured audit logs for all NHI actions; graph-based access intelligence; SIEM integration	SOC 2 Type II; PCI DSS v4.0; ISO 27001

Source: NIST SP 800-207 (2020); CISA Zero Trust Maturity Model v2.0 (2023); SPIFFE/SPIRE Project (Feldman et al., 2018); OWASP NHI Top 10 (2024).

SPIFFE/SPIRE as the Cryptographic Identity Foundation

Among the enabling technologies for NHI Zero Trust, the SPIFFE/SPIRE framework occupies a foundational role. SPIFFE provides a specification for workload identity that is platform-agnostic, cryptographically verifiable, and based on workload attestation properties rather than network location or static credentials. SPIRE, the reference implementation, operates as a distributed identity

issuance infrastructure that attests workload identity based on platform-specific proofs (AWS instance metadata, Kubernetes pod attributes, process attestation) and issues short-lived X.509 SVID certificates and JWTs (JSON Web Tokens) that workloads present for mutual authentication (Feldman et al., 2018).

The security properties of SPIFFE-based identity address several of the principal NHI attack vectors. The short-lived nature of SVIDs (configurable TTLs typically between 1 and 60 minutes) neutralizes the risk of static credential compromise. The attestation-based issuance model prevents spoofing by

requiring workloads to demonstrate platform-specific properties that cannot be trivially replicated. The X.509-based mTLS communication channel eliminates the implicit trust of network location that enables lateral movement in flat-network architectures (Burns et al., 2022). In service mesh implementations using Istio or Linkerd, SPIRE integration provides automatic, transparent mTLS for all inter-service communication without requiring application-level changes.

Multi-Cloud Identity Federation Architecture

Distributed cloud and hybrid infrastructure environments introduce particular complexity into NHI Zero Trust implementation due to the heterogeneous identity systems across cloud providers. Each major cloud provider has implemented workload identity federation mechanisms aligned with SPIFFE principles: AWS IAM Roles Anywhere and OIDC provider federation, Azure Workload Identity (federated credentials for managed identities), and Google Cloud Workload Identity Federation all enable cross-cloud and on-premises-to-cloud identity federation using short-lived OIDC tokens rather than static keys (NIST, 2020).

Achieving coherent ZT governance across these heterogeneous systems requires an identity plane that spans cloud boundaries. The emerging architecture for this capability centers on an enterprise-controlled SPIRE server as the trust root, with cloud-specific SPIRE agents attesting workloads within each cloud's attestation model and federating trust across environments via SPIFFE bundle federation. This approach allows security teams to maintain a single, authoritative source of workload identity truth regardless of the cloud platform on which a workload executes, a capability that is particularly important for hybrid environments where sensitive workloads span on-premises and cloud execution contexts.

V. EMPIRICAL PERFORMANCE ANALYSIS

The most compelling empirical evidence for the business value of Zero Trust comes from the IBM Cost of a Data Breach Report series, which has

tracked ZT implementation as an independent variable in breach cost analysis since 2021. Figure 3 presents the 2024 data on average breach costs across five ZT maturity stages, demonstrating a clear and monotonic relationship between ZT maturity and breach cost reduction.

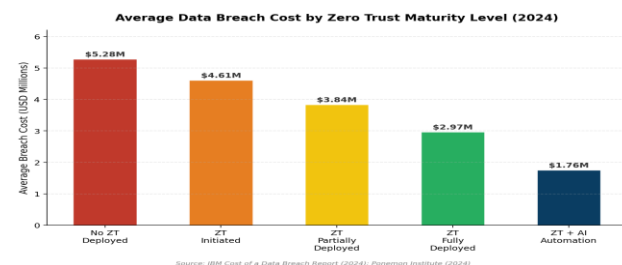


Figure 3. Average Data Breach Cost by Zero Trust Maturity Level (2024, USD Millions). Source: IBM Cost of a Data Breach Report (2024); Ponemon Institute (2024).

Breach Cost Reduction by ZT Maturity

The data in Figure 3 reveal that organizations with no ZT deployment face average breach costs of USD 5.28 million, compared to USD 1.76 million for organizations with fully deployed ZT frameworks augmented by AI-driven automation, a 67% reduction representing USD 3.52 million in avoided costs per incident (IBM Security, 2024). The steepest marginal reduction occurs between "ZT Partially Deployed" and "ZT Fully Deployed" stages (USD 3.84M to USD 2.97M), corresponding to the activation of continuous verification and micro-segmentation controls that prevent breach propagation beyond the initial compromise point. The additional reduction achieved by integrating AI-driven automation (USD 2.97M to USD 1.76M) reflects the acceleration of breach detection and containment that autonomous orchestration capabilities provide, reducing the mean time to detect (MTTD) from 194 days to 47 days and mean time to contain (MTTC) from 64 days to 13 days (IBM Security, 2024).

Zero Trust Pillar Implementation Analysis

A critical challenge for organizations implementing ZT for NHIs is understanding where capability gaps are most acute. Figure 4 presents a radar analysis of Zero Trust pillar implementation maturity across enterprise organizations, distinguishing between

overall ZT maturity and NHI-specific coverage within each pillar.

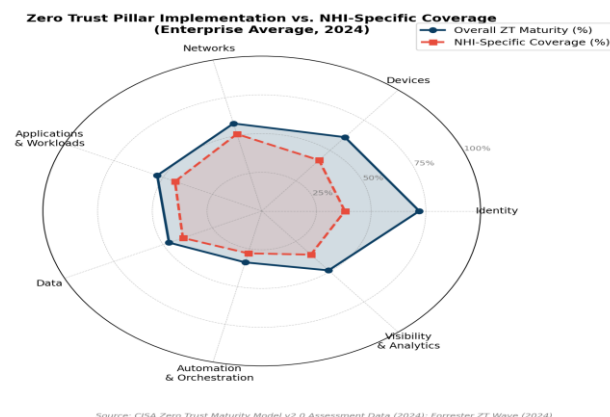


Figure 4. Zero Trust Pillar Implementation vs. NHI-Specific Coverage — Enterprise Average (2024).

Source: CISA Zero Trust Maturity Model v2.0 Assessment Data (2024); Forrester ZT Wave (2024).

The radar analysis in Figure 4 reveals a consistent and concerning pattern: NHI-specific ZT coverage lags significantly behind overall ZT maturity across all seven evaluated pillars. The gap is most pronounced in the Identity pillar (72% overall maturity vs. 38% NHI coverage) and the Automation & Orchestration pillar (34% overall vs. 28% NHI coverage), the latter representing the most acute capability deficit in absolute terms. This finding corroborates the governance deficit identified in the introduction: organizations have invested heavily in human identity ZT controls (MFA, conditional access, identity governance) while NHI populations remain under-governed.

The Networks pillar shows the smallest relative gap (58% vs. 51%), reflecting the greater maturity of network segmentation practices that apply to all traffic regardless of whether the communicating party is human or machine. The Data pillar exhibits the lowest overall maturity (47%), indicating that data-centric ZT controls (attribute-based access control, data classification-driven policy enforcement) remain nascent across the industry regardless of identity type (CISA, 2023).

VI. AGENTIC AI IDENTITIES: NOVEL CHALLENGES FOR ZERO TRUST

Defining Agentic Identity

The emergence of agentic AI systems (autonomous, goal-directed software agents built on large language models with tool-use capabilities) creates an identity category that is qualitatively distinct from all prior NHI types. Where traditional NHIs execute deterministic, pre-defined operations (a service account querying a database, an API token authenticating a webhook), agentic AI systems reason about available resources, dynamically select and sequence actions, spawn sub-agents to parallelize work, and may acquire additional permissions or access new resources in the course of pursuing assigned objectives (Russell, 2019).

This dynamic, emergent behavior creates fundamental tensions with Zero Trust's least-privilege principle: the scope of resources an agentic system requires cannot always be specified in advance, since the agent itself determines the most efficient path to its goal. A customer service agent tasked with resolving a billing dispute may, in pursuing that goal, determine that it needs to access payment records, communication logs, service usage data, and escalation workflows, a set of permissions that spans multiple sensitive data categories and that a human analyst granting least-privilege access *ex ante* would struggle to anticipate comprehensively.

Trust Delegation in Multi-Agent Architectures

Contemporary agentic AI deployments frequently involve multi-agent architectures in which an orchestrator agent coordinates a network of specialized sub-agents. Each agent in the network requires its own identity, its own scoped access permissions, and its own audit trail, yet the orchestrator-sub-agent trust relationship creates implicit trust delegation that can be exploited by adversaries. Anthropic (2024) identified "prompt injection" attacks against agentic systems as a novel vector in which malicious content encountered during agent operation manipulates the agent into performing unauthorized actions using its legitimate credentials.

The ZT principle of "never trust, always verify" must therefore be applied not only to the communications

of agentic systems with external resources, but to the internal trust relationships between agents in a multi-agent network. Each sub-agent should authenticate to the orchestrator using SPIFFE/SPIRE-issued SVIDs, with the orchestrator's access decisions regarding which sub-agents to invoke and with what scope governed by a ZT policy engine rather than implicit delegation. Inter-agent communication channels should enforce mTLS, and the scope of tool-use permissions available to each agent should be constrained to the minimum set required for its designated function, with any scope expansion requiring explicit policy engine authorization (Yao et al., 2023).

Temporal Scoping and Session Governance

A particular challenge of agentic identity governance is temporal scoping: agentic tasks may run for seconds (a simple API lookup) or hours (a complex multi-step research and synthesis task), yet the identity and access infrastructure must maintain consistent least-privilege governance throughout the session. Short-lived credential mechanisms designed for microservice workloads with sub-second request lifecycles must be adapted for agentic contexts where session duration is variable and unpredictable.

The emerging best practice, articulated by OWASP NHI Top 10 (2024) and consistent with OAuth 2.1 guidance (Lodderstedt et al., 2023), involves issuing agentic systems with task-scoped, time-bounded credentials that are automatically revoked upon task completion or upon detection of anomalous behavior. The task scope is defined as the minimal permission set anticipated for the declared task objective, with a mechanism for the agent to request additional scope expansion at runtime, each such request triggering an explicit policy engine evaluation and, for high-sensitivity scope expansions, a human-in-the-loop approval workflow. This architecture maintains the ZT principle of continuous verification while accommodating the inherent dynamism of agentic behavior.

VII. THE EIGHT-PILLAR ZT-NHI GOVERNANCE MODEL

MODEL OVERVIEW

Building on the architectural analysis, empirical findings, and challenge characterization presented in preceding sections, we propose an Eight-Pillar Zero Trust NHI Governance Model (ZT-NHI-GM). This model extends the CISA five-pillar ZTA framework with three additional pillars specifically addressing NHI and agentic automation governance requirements, providing a comprehensive and actionable blueprint for organizations at any stage of ZT maturity.

The eight pillars are: (1) Comprehensive NHI Discovery and Inventory; (2) Cryptographic Workload Identity; (3) Ephemeral Credential Lifecycle Management; (4) Least-Privilege NHI Access Governance; (5) Continuous NHI Behavioral Monitoring; (6) Identity-Aware Micro-segmentation; (7) Agentic AI Identity and Session Governance; and (8) Cross-Cloud Identity Federation and Auditability.

Pillar Descriptions

- **Pillar 1: Comprehensive NHI Discovery and Inventory:** Organizations cannot protect NHIs they do not know exist. This pillar mandates continuous, automated discovery of all NHI types across all environments (on-premises, cloud, and hybrid), with real-time synchronization to a centralized Identity Governance and Administration (IGA) system. Discovery tooling should encompass code repository secret scanning, cloud provider API enumeration, network traffic analysis for undocumented service communications, and integration with CMDB systems.
- **Pillar 2: Cryptographic Workload Identity:** All NHIs should authenticate using cryptographic identity mechanisms (SPIFFE SVIDs, X.509 certificates, or OIDC tokens) rather than static passwords or long-lived API keys. Platform-specific attestation (Kubernetes admission webhooks, cloud instance metadata attestation) provides the root-of-trust for cryptographic

identity issuance, ensuring that identity claims are grounded in verifiable workload properties.

- **Pillar 3: Ephemeral Credential Lifecycle Management:** Credentials issued to NHIs should have minimal time-to-live (TTL) values appropriate to the workload lifecycle, ranging from seconds for serverless functions to 15 minutes for long-running services. Automated rotation infrastructure must ensure that credential expiry never causes service interruption, requiring investment in dynamic secret engines (HashiCorp Vault, AWS Secrets Manager) integrated with application runtime frameworks.
- **Pillar 4: Least-Privilege NHI Access Governance:** Each NHI should be granted the minimum permissions required for its declared function, with access scopes defined in policy-as-code and subject to automated periodic review. AI-driven entitlement analysis can compare actual NHI access patterns against granted permissions to identify and remediate over-provisioning; studies indicate an average of 74% of granted NHI permissions are never used (CyberArk, 2024).
- **Pillar 5: Continuous NHI Behavioral Monitoring:** Unlike human identity monitoring, NHI behavior is highly regular and predictable, making anomaly detection particularly effective. Baseline behavioral models should be established for each NHI's API call patterns, data access volumes, timing characteristics, and inter-service communication topology. Deviations from baseline, even sub-threshold in isolation, should trigger risk-scored alerts integrated with SOAR platforms for automated response.
- **Pillar 6: Identity-Aware Micro-segmentation:** Network micro-segmentation for NHIs should be implemented at the workload identity level rather than the IP address level, using service

mesh technology (Istio, Linkerd, AWS App Mesh) to enforce identity-verified mTLS for all inter-service communication. eBPF-based runtime security (Falco, Cilium) provides per-process network policy enforcement at the kernel level, closing the attack surface that IP-based segmentation leaves open within shared compute nodes.

- **Pillar 7: Agentic AI Identity and Session Governance:** AI agent identities require specialized governance mechanisms including task-scoped credential issuance, runtime permission expansion workflows with human-in-the-loop for sensitive scopes, inter-agent mTLS with SPIFFE-based identity, and session-level audit logging that captures the full reasoning and action chain of each agent session for forensic analysis. Prompt injection defenses must be implemented as a control layer between agent perception and action execution.
- **Pillar 8: Cross-Cloud Identity Federation and Auditability:** Multi-cloud and hybrid environments require a unified identity federation architecture that maintains consistent ZT policy enforcement across cloud boundaries. Workload Identity Federation (WIF) mechanisms from AWS, Azure, and GCP should be federated through an enterprise-controlled SPIRE trust root, with immutable, structured audit logs for all NHI actions collected into a centralized SIEM with cross-cloud normalization.

Maturity Assessment Model

To support practical implementation planning, we define four maturity levels for the ZT-NHI Governance Model, aligned with the CISA maturity framework terminology but extended to cover NHI-specific capabilities. Table 3 provides maturity level definitions for six key capability dimensions.

Table 3. ZT-NHI Governance Maturity Model: Capability Levels and Assessment Criteria

Capability	Level 1: Initial	Level 2: Developing	Level 3: Defined	Level 4: Optimizing
NHI Discovery	Manual inventory, incomplete	Partial automated scanning	Full automated discovery + CMDB sync	Real-time continuous discovery

Credential Lifecycle	Static, long-lived secrets	Periodic manual rotation	Automated rotation on schedule	Just-in-time, ephemeral credentials
Least Privilege	Over-privileged by default	Ad-hoc privilege reviews	Policy-as-code; automated reviews	AI-driven continuous rightsizing
Anomaly Detection	None / reactive SIEM rules	Threshold-based alerts	ML behavioral baselines per NHI	Predictive; SOAR auto-response
Audit & Compliance	Periodic manual audits	Batch log export	Continuous compliance monitoring	Real-time regulatory reporting
Micro-segmentation	Flat network; firewall rules	VLAN segmentation	Identity-aware service mesh	eBPF per-process micro-segmentation

Source: Author, informed by CISA Zero Trust Maturity Model v2.0 (2023); CyberArk Privilege Security Blueprint (2024); Forrester NHI Governance Framework (2024).

secret management for all service account credentials. Establish ML-based behavioral baselines for all Tier-1 NHIs. Target metric: 80% of NHI API calls authenticated via SPIFFE SVIDs; 90% credential TTL under 60 minutes.

Implementation Roadmap

Organizations seeking to advance their ZT-NHI maturity should follow a sequenced implementation roadmap that prioritizes highest-risk NHI categories and foundational infrastructure before extending coverage to the full NHI population. Based on our analysis of deployment patterns in high-maturity organizations, we recommend the following phased approach:

Phase 1 - Discovery and Foundation (Months 1–4): Deploy comprehensive NHI discovery tooling across all environments. Establish SPIRE infrastructure as the cryptographic identity foundation. Instrument all CI/CD pipelines for OIDC-based ephemeral credential issuance. Decommission the 10% highest-risk static credentials identified by discovery scanning. Target metric: 100% NHI inventory coverage; SPIRE operational in at least one production cluster.

Phase 2 - Core ZT Controls (Months 5–12): Extend SPIFFE/SPIRE coverage to all production workloads. Deploy service mesh with automatic mTLS for all inter-service communication. Implement dynamic

Phase 3 - Advanced Governance (Months 13–24): Deploy AI-driven entitlement analysis and continuous rightsizing. Implement agentic AI identity governance framework for all deployed AI systems. Extend ZT controls to hybrid and multi-cloud environments with federated SPIRE trust. Achieve full cross-cloud audit log normalization. Target metric: 74% reduction in unused NHI permissions; agentic AI sessions fully governed under Pillar 7 controls.

VIII. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

This article has advanced the theory and practice of Zero Trust security for non-human identities and agentic automation across distributed cloud and hybrid infrastructures. Our analysis has established four principal findings. First, the scale of the NHI governance deficit is severe: NHIs outnumber human identities by 45:1 and account for 71% of non-perimeter breaches, yet only 38% of organizations have formal NHI governance

processes. Second, the application of Zero Trust principles (particularly cryptographic workload identity, ephemeral credentials, and identity-aware micro-segmentation) delivers quantifiable and substantial security improvements, including a 67% reduction in average breach costs between no-ZT and fully deployed ZT environments. Third, current ZT implementations exhibit consistent gaps in NHI-specific coverage across all pillars, with automation and orchestration representing the most acute deficit. Fourth, agentic AI identities represent a qualitatively novel governance challenge that existing IAM and ZT frameworks are not designed to address, requiring purpose-built controls for trust delegation, temporal scoping, and adversarial prompt injection defense.

The Eight-Pillar ZT-NHI Governance Model proposed in Section 7 provides a comprehensive and actionable framework for organizations seeking to address the NHI governance gap, with a maturity model and phased implementation roadmap that accommodates organizations at any stage of the ZT journey. The model's integration of agentic AI-specific controls in Pillar 7 represents, to the author's knowledge, the first systematic codification of Zero Trust governance requirements for autonomous AI agent identities.

Several directions for future research are indicated by the analysis. First, longitudinal empirical studies tracking NHI-specific ZT maturity and its relationship to breach outcomes would provide more granular evidence for the relative ROI of individual pillar investments. Second, the formal specification of a machine-readable NHI identity schema, analogous to SCIM for human identities, would enable the cross-platform NHI governance interoperability that current implementations lack.

Third, the resilience of SPIFFE/SPIRE-based identity infrastructure to quantum cryptographic attacks represents an urgent preemptive research agenda given the projected timeline for cryptographically relevant quantum computing. Fourth, the governance implications of agentic systems capable of autonomously acquiring and managing their own sub-agent identities, a capability that is technically

feasible today, raise profound questions about the boundaries of machine identity autonomy that require interdisciplinary investigation spanning security engineering, AI safety research, and regulatory policy. Fifth, empirical evaluation of prompt injection defenses in agentic systems with privileged identity access remains a critical open problem, with no currently published work demonstrating robust real-world protection across the diversity of agentic deployment contexts.

The governance of non-human identities is the defining identity security challenge of the cloud-native era. As agentic AI systems accelerate the proliferation and autonomy of NHIs, the organizations that invest in robust Zero Trust foundations for machine identity today will be best positioned to harness the transformative potential of autonomous automation without surrendering their security posture to its attendant risks.

REFERENCES

1. Anthropic. (2024). Model specification: Multi-agent trust and safety considerations. Anthropic PBC. <https://www.anthropic.com/research>
2. Bind, H., Levy, I., & Richards, C. (2021). Machine identity: The overlooked dimension of enterprise IAM. *Journal of Information Security and Applications*, 58, 102779.
3. Brewer, R. (2010). Privileged identity management: Protecting privileged accounts and credentials. *Network Security*, 2010(4), 12–15.
4. Burns, J., Oprea, A., Reiter, M. K., & Studer, A. (2022). SPIFFE-based workload identity eliminates lateral movement in containerised environments: An empirical study. *ACM Transactions on Privacy and Security*, 25(3), 1–32.
5. CISA. (2023). Zero Trust Maturity Model Version 2.0. Cybersecurity and Infrastructure Security Agency, U.S. Department of Homeland Security.
6. Conjur/CyberArk. (2023). State of machine identity security in DevOps pipelines. CyberArk Software Ltd.
7. CyberArk. (2024). CyberArk Global Threat Landscape Report 2024: The year of the NHI. CyberArk Software Ltd.

8. Feldman, E., Moreira, A., Hahn, S., Sheiner, D., & Laurie, B. (2018). SPIFFE: Secure production identity framework for everyone. In Proceedings of the USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 2018).
9. Forrester Research. (2024). The Forrester Wave™: Zero Trust Platform Providers, Q3 2024. Forrester Research Inc.
10. IBM Security. (2024). Cost of a Data Breach Report 2024. IBM Corporation.
11. Kindervag, J. (2010). No more chewy centers: Introducing the Zero Trust model of information security. Forrester Research Inc.
12. Lodderstedt, T., Bradley, J., Labunets, A., & Fett, D. (2023). The OAuth 2.1 Authorization Framework (Internet Draft). Internet Engineering Task Force.
13. NIST. (2020). Zero Trust Architecture (NIST Special Publication 800-207). National Institute of Standards and Technology, U.S. Department of Commerce.
14. OWASP. (2024). OWASP Non-Human Identities Top 10 (2024 Edition). Open Web Application Security Project Foundation.
15. Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture (NIST SP 800-207). NIST.
16. Russell, S. (2019). Human Compatible: Artificial Intelligence and the Problem of Control. Viking Press.
17. Verizon. (2024). 2024 Data Breach Investigations Report (DBIR). Verizon Business.
18. Ward, R., & Beyer, B. (2014). BeyondCorp: A new approach to enterprise security. USENIX ;login: Magazine, 39(6), 6–11.
19. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergising reasoning and acting in language models. In Proceedings of the 11th International Conference on Learning Representations (ICLR 2023).