

Research Article

A Zero-Trust Framework for Automated Protection of Non-Human Identities in Cloud IAM

Minh Nguyen ¹

¹Faculty of Information Technology, Vietnam Institute of Advanced Computing, Vietnam

Abstract

The rapid adoption of cloud-native applications, automation pipelines, artificial intelligence services, containers, serverless workloads, and machine-to-machine communication has substantially increased the number and operational importance of non-human identities (NHIs). Unlike human identities, NHIs frequently operate continuously, interact across distributed environments, and may possess permissions that are difficult to observe, evaluate, and revoke dynamically. Consequently, conventional identity and access management approaches based primarily on static credentials and manually administered permissions are insufficient for highly automated cloud environments. This research proposes a Zero-Trust Framework for Automated Protection of Non-Human Identities in Cloud IAM. The study follows a conceptual research-and-review methodology based exclusively on the supplied literature and synthesizes principles from automated governance, distributed computing, machine learning, optimization, DataOps, IoT, and cyber-physical systems. The proposed framework combines continuous identity discovery, contextual authentication, least-privilege authorization, behavioral monitoring, risk scoring, automated credential governance, policy optimization, and continuous revocation. The analysis indicates that NHI protection should be treated as a continuously optimized governance process rather than a one-time access-control configuration. The framework further establishes a feedback mechanism through which observed identity behavior can influence future authorization decisions. The research contributes a structured conceptual architecture for integrating Zero Trust principles with automated NHI lifecycle management and identifies implementation challenges involving policy complexity, false positives, interoperability, explainability, and operational scalability.

Keywords: Zero Trust, Non-Human Identities, Cloud IAM, Identity Governance, Automated Security, Least Privilege, Behavioral Analytics, Risk-Based Access Control, Cloud Security.



Received: 12 June 2026
Revised: 2 July 2026
Accepted: 20 August 2026
Published: 02 September 2026

Doi:10.55640/ijiot-06-09-01

Page No: 01-09

Copyright: © 2026 Authors retain the copyright of their manuscripts, and all Open Access articles are disseminated under the terms of the Creative Commons Attribution License 4.0 (CC-BY), which licenses unrestricted use, distribution, and reproduction in any medium, provided that the original work is appropriately cited.

INTRODUCTION

Cloud computing has transformed enterprise information systems from relatively static infrastructure into highly distributed and automated environments. Modern workloads can communicate through application programming interfaces, service accounts, workload identities, automation agents, containers, data-processing systems, and machine-learning pipelines. These entities require identities and permissions to perform computational operations even when no human user is directly involved. Such identities constitute a growing operational layer within cloud Identity and Access Management (IAM).

The security challenge is not simply the existence of these identities but the difficulty of maintaining appropriate authorization throughout their lifecycle. A non-human identity

may be created automatically, receive permissions from several systems, communicate with multiple services, and remain active after its original purpose has changed. Static governance therefore creates a risk that excessive or obsolete privileges remain available. Ganapathy (2025) specifically positions automated governance and protection of non-human identities as an important dimension of cloud IAM, establishing the relevance of treating NHI security as an automated governance problem rather than solely as credential management.

The problem becomes more complex in environments where distributed computation and automated decision-making are fundamental operating characteristics. Research on distributed neural networks demonstrates the importance of coordinating computational processes across distributed clusters (Hamdi et al., 2024), while DataOps research emphasizes lifecycle-oriented management of data operations (Bahaa et al., 2023). These perspectives suggest that identity protection must similarly account for lifecycle transitions, distributed execution, and automated operational dependencies.

The objective of this research is to develop a conceptual Zero-Trust framework capable of continuously discovering, evaluating, authorizing, monitoring, and governing NHIs in cloud environments. The study focuses on four questions: how NHIs can be continuously governed; how Zero Trust principles can be translated into automated authorization decisions; how behavioral and contextual information can support adaptive risk assessment; and how automated remediation can reduce the persistence of excessive privileges.

The significance of the research lies in integrating identity governance with continuous security decision-making. Rather than assuming that an identity is trustworthy after initial authentication, the proposed approach evaluates each access request according to identity attributes, workload context, resource sensitivity, behavioral evidence, and current policy requirements.

LITERATURE REVIEW

The supplied literature does not constitute a conventional cloud-IAM corpus. Instead, it provides several complementary methodological perspectives that can inform an NHI protection architecture.

Ganapathy (2025) provides the most direct conceptual foundation by addressing automated governance and protection of NHIs in cloud IAM. The central implication is that machine identities require dedicated governance mechanisms because their operational characteristics differ from those of conventional human accounts. The present research extends this premise by positioning automated NHI governance within a Zero-Trust control model.

The literature on distributed and intelligent computing contributes a second foundation. Hamdi et al. (2024) investigate distributed recurrent neural networks across Hadoop/Spark infrastructure, illustrating how computational workloads may operate across multiple distributed components. In an analogous cloud environment, a workload identity may interact with several distributed services, meaning that authorization cannot be considered solely at an individual application boundary.

Machine-learning research also demonstrates the potential of adaptive models for complex operational decisions. Liu et al. (2023) examine deep learning for intelligent production scheduling optimization, while Raska et al. (2024) investigate adaptive neural networks for discrete-event simulation. These studies are relevant conceptually because identity governance can similarly be formulated as an optimization problem involving competing objectives such as security, availability, operational efficiency, and policy compliance.

Wen (2024) investigates multi-feature fusion involving long short-term memory networks and an improved artificial bee colony algorithm. The methodological relevance is that security decisions can potentially incorporate multiple behavioral and contextual features rather than relying on a single indicator. Similarly, Zhang et al. (2024) explore data fusion through Q-learning and cascade classification, suggesting the usefulness of combining heterogeneous signals in automated decision systems.

The IoT literature provides another useful analogy. Montoya et al. (2025) address prioritized services in energy-constrained IoT networks, where limited resources require differentiated prioritization. In cloud IAM, resources and privileges likewise possess different levels of sensitivity. An automated policy engine therefore needs mechanisms for prioritizing high-risk resources and applying stricter authorization requirements.

Almeida et al. (2024) examine knowledge transfer in multi-objective multi-agent reinforcement learning. This work supports the theoretical possibility of treating complex governance as a multi-objective decision problem, where security and operational objectives must coexist. The proposed framework uses this insight conceptually rather than claiming that reinforcement learning is necessary for every IAM deployment.

Other studies reinforce the importance of structured modeling and lifecycle-aware decision-making. El Faydy and El Abbadi (2023) apply interpretive structural modeling to critical success factors in Industry 4.0, illustrating the usefulness of identifying relationships among interdependent organizational and technological factors. Li and Yao (2024) examine formal activity diagrams for safety-oriented system testing, providing methodological support for modeling control flows and validating security-relevant activities.

Research on cyber-physical and smart-grid systems further demonstrates the importance of real-time coordination. Gligor et al. (2025) present an augmented cyber-physical model for real-time smart-grid co-simulation, while Li et al. (2024) investigate lifecycle aging modes in energy and transportation systems. Although these studies are not IAM studies, they reinforce the broader proposition that security and operational decisions in interconnected systems require continuous state awareness.

A major gap emerges from the synthesis. The supplied studies separately address automation, optimization, distributed computation, lifecycle management, and intelligent decision-making, while Ganapathy (2025) directly addresses NHI governance. However, there is limited integration among these perspectives into a unified Zero-Trust architecture specifically focused on automated NHI protection. This research addresses that gap conceptually.

METHODOLOGY

3.1 Research Design

This study adopts a conceptual research-and-review methodology. The objective is not to evaluate a particular cloud vendor or produce empirical benchmark measurements but to construct a technically coherent framework from the supplied literature. The methodology combines thematic synthesis, architectural abstraction, control-flow modeling, and conceptual risk analysis.

The literature was organized into five analytical themes: identity governance, distributed workload management, intelligent decision-making, lifecycle management, and automated optimization. These themes were subsequently mapped to the security requirements of NHIs.

The resulting framework follows a continuous control cycle:

Discover → Authenticate → Evaluate Context → Authorize → Monitor → Recalculate Risk → Remediate → Revalidate

This cycle is intended to prevent the assumption that authorization is permanent after initial authentication.

3.2 NHI Discovery and Inventory Layer

The first layer establishes an authoritative inventory of machine identities. It includes service accounts, application identities, workload identities, automation agents, API identities, pipeline identities, and other machine-operated credentials.

Each identity should possess a machine-readable identity profile containing ownership, purpose, associated workload, resource scope, credential type, creation date, last observed activity, privilege level, and lifecycle state.

The principle is comparable to lifecycle-oriented operational management emphasized in DataOps research (Bahaa et al., 2023). An identity should not exist as an isolated credential object; it should be associated with an operational lifecycle.

A hypothetical example is a deployment pipeline that creates a temporary workload identity to deploy an application. Once deployment is complete, the identity should automatically transition to an inactive or revoked state rather than remaining indefinitely available.

3.3 Zero-Trust Authentication Layer

The second layer applies continuous verification. Authentication establishes that an identity is genuine, but Zero Trust requires authorization decisions to incorporate contextual information.

A conceptual trust score can be represented as:

$$T_i = f(I, C, B, R, P) \quad T_i = f(I, C, B, R, P)$$

where II represents identity confidence, CC represents contextual attributes, BB represents behavioral evidence, RR represents resource sensitivity, and PP represents applicable policy conditions.

The purpose of this formulation is not to prescribe a specific algorithm but to establish that authorization should depend on multiple dimensions.

For example, a workload identity normally accessing a database from its expected service environment may receive a low-risk classification. The same identity suddenly requesting administrative access to an unrelated production resource should trigger additional evaluation.

3.4 Least-Privilege Authorization Engine

The authorization layer translates identity and context into permissions. Instead of assigning broad permissions to an identity, the framework evaluates the minimum permissions required for a specific operation.

A conceptual authorization decision is:

$$A = P(I, R, O, C) - EA = P(I, R, O, C) - E$$

where AA represents the permitted access set, II is identity context, RR is resource sensitivity, OO is requested operation, CC is contextual information, PP is the applicable

policy set, and EE represents permissions excluded because they exceed the current operational requirement.

The critical distinction is between authentication and authorization. A valid credential does not automatically justify unrestricted access.

This principle is particularly important for machine identities because automation can execute actions at machine speed and scale. Excessive privileges can therefore propagate rapidly across interconnected resources.

3.5 Behavioral Monitoring and Risk Scoring

The proposed framework continuously evaluates identity behavior. Relevant features may include access frequency, resource diversity, request timing, geographic or network context, privilege changes, failed operations, and deviations from historical behavior.

The conceptual risk function can be expressed as:

$$Risk_i = w_1D + w_2P + w_3B + w_4S + w_5L$$

where DD represents deviation from normal behavior, PP represents privilege exposure, BB represents behavioral anomalies, SS represents resource sensitivity, and LL represents lifecycle-related risk.

Multi-feature decision-making is consistent with the methodological direction found in intelligent classification and optimization studies (Wen, 2024; Zhang et al., 2024). However, the framework does not assume that machine learning must replace deterministic policy. Instead, analytics should supplement explicit security policies.

3.6 Adaptive Policy Optimization

Cloud environments change continuously. Therefore, static policies can become inaccurate as workloads evolve. The framework introduces an adaptive policy optimization layer capable of identifying redundant, excessive, or obsolete permissions.

Optimization can be expressed as:

$$\min F = \alpha R + \beta O + \gamma C$$

where RR is security risk, OO is operational overhead, and CC represents policy complexity. The coefficients reflect organizational priorities.

The multi-objective perspective is consistent with research involving multi-agent reinforcement learning and optimization (Almeida et al., 2024). Similarly, adaptive computational methods examined by Liu et al. (2023) and Raska et al. (2024) provide conceptual support for continuously improving decision processes.

3.7 Automated Remediation

Detection without remediation limits the practical value of continuous monitoring. When an NHI exceeds a predefined risk threshold, the system should automatically initiate an appropriate response.

Possible actions include temporary privilege reduction, credential rotation, session termination, policy restriction, identity quarantine, or complete revocation.

The remediation decision can follow graduated controls rather than an immediate binary response. A low-confidence anomaly may trigger additional verification, whereas a high-confidence compromise indicator may result in immediate credential revocation.

This lifecycle approach is consistent with the broader emphasis on adaptive and automated operational processes evident across the supplied literature.

3.8 Continuous Validation and Audit

The final layer validates whether the implemented controls remain effective. Every authorization decision should generate an auditable record containing the identity, resource, operation, contextual evidence, policy decision, risk classification, and resulting action.

Formalized activity modeling is relevant to this requirement because security workflows benefit from explicit representations of expected transitions (Li and Yao, 2024). Similarly, structural modeling can assist organizations in understanding relationships among critical governance factors (El Faydy and El Abbadi, 2023).

The complete framework therefore operates as a feedback system rather than a linear sequence. Monitoring results influence risk assessment, risk influences authorization, and authorization outcomes influence subsequent governance actions.

RESULTS

The conceptual analysis produced five principal findings.

First, NHI security is fundamentally a lifecycle governance problem. An identity may be secure at creation but become risky when its workload changes, privileges accumulate, ownership becomes unclear, or its operational purpose ends. Consequently, automated discovery and lifecycle transitions should be core IAM capabilities rather than supplementary functions. This conclusion is consistent with the lifecycle orientation found in DataOps research and the direct focus on automated NHI governance presented by Ganapathy (2025).

Second, Zero Trust is most effective for NHIs when authentication is separated from continuous authorization. A valid credential should establish identity evidence, not permanent trust. Authorization must remain conditional on current context, behavior, resource sensitivity, and policy.

Third, the literature supports treating automated authorization as a multi-objective decision problem. Security, availability, performance, administrative effort, and policy complexity can conflict. Research on multi-objective learning and optimization provides methodological support for this perspective (Almeida et al., 2024; Liu et al., 2023).

Fourth, behavioral intelligence can improve NHI governance when used as an additional signal rather than as an unrestricted replacement for deterministic policy. Multi-feature and adaptive models provide a conceptual foundation for detecting deviations in complex operational environments (Wen, 2024; Zhang et al., 2024).

Fifth, automated remediation is necessary to close the governance loop. Detection without response allows risky identities to remain active. A continuous model should therefore connect monitoring with privilege reduction, credential rotation, quarantine, or revocation.

The findings collectively indicate that effective NHI protection requires integration among identity inventory, policy enforcement, behavioral analytics, adaptive optimization, and automated remediation. The proposed architecture consequently moves beyond static IAM toward a continuously evaluated control system.

DISCUSSION

The proposed framework has both theoretical and practical implications. Theoretically, it positions NHI protection as an adaptive control problem rather than a conventional access-control problem. Traditional IAM thinking can emphasize who receives access, whereas the proposed Zero-Trust perspective additionally asks why access is required, under what context it is requested, whether the behavior is consistent with the identity's role, and whether the permission should continue to exist.

This perspective extends the direct NHI governance focus of Ganapathy (2025) by introducing a continuous feedback architecture. The distinction is important because machine identities frequently operate without direct human supervision. Automated governance is therefore not merely an efficiency mechanism; it is a security control.

A second implication concerns distributed cloud architectures. Research on distributed processing illustrates that workloads may span multiple computational components (Hamdi et al., 2024). Consequently, identity governance cannot be restricted to a single application boundary. A workload may legitimately require access to several services, but the authorization system must distinguish legitimate dependency chains from anomalous privilege expansion.

The framework also introduces a significant trade-off between security sensitivity and operational continuity. Excessively aggressive anomaly detection can interrupt legitimate automated workloads. Conversely, overly permissive thresholds may allow compromised identities to continue operating. Adaptive approaches discussed in the supplied AI literature provide a potential direction for balancing these competing requirements (Raska et al., 2024; Wen, 2024).

Another challenge concerns explainability. If an automated system denies a machine identity's request because of a behavioral risk score, administrators must understand the factors responsible for the decision. This requirement limits the usefulness of opaque models in security-critical IAM environments. Therefore, deterministic policy rules and interpretable risk features should remain visible even when advanced analytics are introduced.

Policy complexity represents another limitation. As organizations create thousands of workloads and identities, manually defining relationships among permissions can become difficult. Optimization research suggests that computational approaches can help manage complex decision spaces (Liu et al., 2023; Montoya et al., 2025), but optimization itself introduces additional operational complexity.

The proposed framework is also limited by the conceptual nature of the study. No empirical cloud deployment, attack dataset, quantitative benchmark, or comparative experiment was performed. The framework should therefore be viewed as an architectural proposition rather than validated evidence of security improvement. Future empirical work should test the framework using realistic cloud workloads, simulated compromised service accounts, privilege-escalation scenarios, and measurable indicators such as mean time to revoke, excessive-permission reduction, false-positive rate, and policy evaluation latency.

Nevertheless, the convergence of lifecycle management, distributed computing, adaptive analytics, and automated governance provides a strong theoretical basis for continuous NHI protection. The broader systems literature demonstrates that interconnected and automated environments require continuous state awareness rather than isolated control points (Gligor et al., 2025). This principle is directly applicable to cloud identity governance.

CONCLUSION

This research proposed a Zero-Trust Framework for Automated Protection of Non-

Human Identities in Cloud IAM. The framework addresses a growing architectural problem in cloud environments: machine identities increasingly perform critical operations while traditional identity controls often remain static, manually managed, and insufficiently contextual.

The research synthesized the supplied literature across automated governance, distributed systems, intelligent optimization, lifecycle management, IoT, and cyber-physical computing. The resulting architecture consists of continuous NHI discovery, contextual authentication, least-privilege authorization, behavioral risk assessment, adaptive policy optimization, automated remediation, and continuous auditing.

The primary contribution is the conceptual integration of these mechanisms into a feedback-driven Zero-Trust lifecycle. Rather than treating identity authorization as a permanent decision, the framework treats it as a continuously reassessed state. This approach is particularly appropriate for NHIs because their actions can occur automatically, rapidly, and across multiple interconnected services.

The study also identifies important implementation limitations, particularly policy complexity, explainability, false positives, interoperability, and the absence of empirical validation. Future research should therefore implement the framework in representative multi-cloud or cloud-native environments and evaluate its effectiveness against measurable security and operational metrics.

Ultimately, automated NHI protection should evolve from static credential administration toward continuous identity governance. The central principle is that a machine identity should receive only the access necessary for its current task, under its current context, for only as long as that access remains justified. Such an architecture provides a foundation for making Zero Trust operationally applicable to the increasingly autonomous identity landscape of modern cloud computing (Ganapathy, 2025).

REFERENCES

1. Almeida, V. N. D., et al. (2024). "Knowledge Transfer in Multi-objective Multi-agent Reinforcement Learning via Generalized Policy Improvement," *Computer Science and Information Systems*, vol. 21, no. 1, pp. 335–362.
2. Bahaa, S., et al. (2023). "DataOps Lifecycle with a Case Study in Healthcare," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, pp. 136–144.
3. El Faydy, N., and El Abbadi, L. (2023). "Interpretive Structural Modelling of Critical Success Factor for Lean Product Lifecycle Management in Industry 4.0," *International Journal of Production Management and Engineering*, vol. 11, no. 1, pp. 65–72.
4. Gligor, A., et al. (2025). "Augmented Cyber-Physical Model for Real-Time Smart-Grid Co-Simulation," *International Journal of Computers Communications & Control*, vol. 20, no. 1, p. 6914.
5. Hamdi, M., et al. (2024). "Deep Recurrent Neural Networks Distributed on a Hadoop/Spark Cluster for Fall Detection," *International Journal of Computers Communications & Control*, vol. 19, no. 3, p. 6428.
6. Han, L. N., et al. (2023). "Balancing Material Supply-Demand with ARIMA and Neural Networks," *International Journal of Simulation Modelling*, vol. 22, no. 4, pp. 712–722.
7. Li, Z., and Yao, M. (2024). "Optimized Test Case Design for Automotive Safety Driving Virtual Training System Using Formal Activity Diagrams," *Tehnički vjesnik – Technical Gazette*, vol. 31, no. 6, pp. 2014–2021.
8. Li, S., et al. (2024). "Factoring Electrochemical and Full-Lifecycle Aging Modes of Battery Participating in Energy and Transportation Systems," *IEEE Transactions on Smart Grid*, vol. 15, no. 5, pp. 4932–4945.

9. Liu, Y., et al. (2023). "Deep Learning for Intelligent Production Scheduling Optimization," *International Journal of Simulation Modelling*, vol. 23, no. 1, pp. 172–183.
10. Liu, Z., et al. (2024). "A Hybrid CNN and LSTM Based Model for Financial Crisis Prediction," *Tehnički vjesnik – Technical Gazette*, vol. 31, no. 1, pp. 185–192.
11. Montoya, G. A., et al. (2025). "A Mathematical Optimization Approach for Prioritized Services in IoT Networks for Energy-Constrained Smart Cities," *International Journal of Computers Communications & Control*, vol. 20, no. 1, p. 6912.
12. Motiejauskas, M., and Dzemyda, G. (2024). "EfficientNet Convolutional Neural Network with Gram-Matrices Modules for Predicting Sadness Emotion," *International Journal of Computers Communications & Control*, vol. 19, no. 5, p. 6697.
13. Raska, P., et al. (2024). "Using Adaptive Neural Networks for Optimising Discrete Event Simulation," *International Journal of Simulation Modelling*, vol. 23, no. 2, pp. 227–238.
14. Sun, H. (2023). "Optimizing Manufacturing Scheduling with Genetic Algorithm and LSTM Neural Networks," *International Journal of Simulation Modelling*, vol. 22, no. 3, pp. 508–519.
15. Wen, T. (2024). "A Multi-feature Fusion Model Based on Long and Short Term Memory Network and Improved Artificial Bee Colony Algorithm for English Text Classification," *Computer Science and Information Systems*, vol. 21, no. 4, pp. 1607–1627.
16. Yang, X., et al. (2024). "Vehicle Driving Behavior Recognition and Optimization Strategies Based on Cloud Computing and SSA-BP Algorithm," *Studies in Informatics and Control*, vol. 33, no. 3, pp. 17–28.
17. Yu, J., et al. (2024). "News Recommendation Model Based on Encoder Graph Neural Network and BAT Optimization in Online Social Multimedia Art Education," *Computer Science and Information Systems*, vol. 21, no. 3, pp. 989–1012.
18. Zhang, X., et al. (2024). "A Novel Industrial Big Data Fusion Method Based on Q-learning and Cascade Classifier," *Computer Science and Information Systems*, vol. 21, no. 4, pp. 1629–1649.
19. Ganapathy, S. K. (2025). "Automated Governance and Protection of Non-Human Identities in Cloud IAM," *Frontiers in Emerging Computer Science and Information Technology*, vol. 2, no. 2, pp. 08–20. Available at: <https://irjernet.com/index.php/fecsit/article/view/cloud-iam-non-human-identities>.