Digital Transformation and Technology Dynamics

Vol. : 02
No. : 01
Article Number : DTTD2022001

Journal Homepage : https://www.techdynamicsjournal.com/

Zero-Trust Architecture in Hybrid Multi-Cloud Environments: Performance Trade-offs in Digital Banking

Mahmuda Begum [1+]
Md Rasel Ul Alam [2]

[1] International Islamic University Chittagong, Chittagong, Bangladesh
[2] University of the Cumberlands, Kentucky, USA

+Corresponding Author Email: mahmudabegum182000@gmail.com

A B S T R A C T

Modern digital banking infrastructures are rapidly transitioning from legacy perimeter-based security models to hybrid multi-cloud architectures. While this transition enhances operational elasticity and service delivery, it expands the attack surface, requiring the adoption of Zero-Trust Architecture (ZTA) governed by the principle of "never trust, always verify." However, enforcing continuous identity verification, dynamic mutual TLS (mTLS) handshakes, micro-segmentation, and real-time Policy Decision Point (PDP) evaluations across distributed on-premise mainframes and public cloud providers introduces substantial computational latency and network overhead. This paper presents an empirical performance trade-off analysis of ZTA implementation within high-throughput digital banking systems. Evaluating a dataset of 500,000 real-time payment transactions across a simulated hybrid multi-cloud testbed, we quantify the impact of granular security policies on end-to-end transaction latency, system throughput (Transactions Per Second, TPS), and resource consumption. Empirical results show that an unoptimized, naive ZTA implementation increases mean round-trip transaction latency by 184.2% (from 42 ms to 119.3 ms) and degrades peak throughput by 41.6%. To mitigate these penalties, we propose an Adaptive Edge-Cached Policy Enforcement Framework (AEC-PEF) utilizing localized Policy Enforcement Points (PEPs) and risk-based dynamic token caching. AEC-PEF recovers 72.4% of lost throughput while maintaining an optimal security assurance posture, offering a scalable blueprint for resilient, low-latency financial systems.

Keywords:

Zero-Trust Architecture (ZTA),
Hybrid Multi-Cloud,
Digital Banking,
Mutual TLS (mTLS),
Performance Trade-offs,
Policy Decision Point (PDP),
Transaction Latency

Article History:

Received: 19 April 2022
Revised: 30 April 2022
Accepted: 30 May 2022
Published: 29 July 2022

1. Introduction

The financial services sector is undergoing a profound digital transformation. Core banking workloads, payment processing gateways, and retail customer portals are increasingly distributed across hybrid multi-cloud environments comprising on-premise mainframe legacy nodes and multi-region public cloud infrastructure (e.g., AWS, Microsoft Azure, Google Cloud Platform). While hybrid multi-cloud topologies provide high availability, dynamic scalability,

and geopolitical data residency compliance, they break the foundational assumption of legacy cybersecurity: the existence of a trusted internal network perimeter.

To secure distributed financial infrastructure against sophisticated threats, including insider credential compromise, supply-chain API exploitation, and lateral movement by advanced persistent threats (APTs), financial institutions are adopting Zero-Trust Architecture (ZTA). As codified by NIST Special Publication 800-207, ZTA replaces implicit zone-based trust with explicit, continuous verification of identity, device health, dynamic context, and data authorization prior to granting access to any enterprise resource (NIST, 2020).

However, implementing ZTA within high-frequency digital banking systems introduces severe operational performance trade-offs. Core payment processing pipelines require strict sub-100 ms response windows and must handle tens of thousands of Transactions Per Second (TPS). Enforcing cryptographic mutual TLS (mTLS) tunnel negotiation at every microservice hop, performing continuous OAuth2/JWT token validation, and querying centralized Policy Decision Points (PDPs) inject cumulative processing overhead. This paper quantifies these latency and throughput penalties and introduces an optimized architectural model to balance cryptographic rigor with sub-second execution speeds.

2. Related Work

The rapid migration of digital banking services to hybrid and multi-cloud environments has significantly improved scalability, service availability, and operational flexibility. However, these benefits have introduced complex cybersecurity challenges, particularly regarding identity management, data protection, and regulatory compliance. Nagaty (2015) proposed a secure hybrid cloud architecture for online banking that integrates public and private cloud infrastructures with cryptographic mechanisms to ensure confidentiality, integrity, and availability of financial data. The study demonstrated that hybrid cloud deployment can reduce operational costs while maintaining robust security controls.

As financial institutions increasingly adopted cloud-native infrastructures, traditional perimeter-based security became insufficient. Nagaraju and Parthiban (2015) introduced a trusted framework for online banking that combines multi-factor authentication and privacy-preserving gateways to strengthen user authentication and mitigate unauthorized access in cloud environments. Their findings highlighted that identity-centric security is essential for protecting banking applications hosted in public cloud platforms.

The emergence of Zero-Trust Architecture (ZTA) further transformed cloud security by enforcing continuous authentication, least-privilege access, and micro-segmentation. Megouache et al. (2020) emphasized that secure user authentication and data integrity are critical challenges in multi-cloud environments, proposing cryptographic mechanisms that improve data protection without compromising system performance.

Similarly, Patil (2020) argued that Zero Trust is particularly suitable for hybrid cloud environments because it continuously verifies users and devices rather than relying on network boundaries. The study noted that while Zero Trust significantly enhances resilience against cyber threats, it also introduces performance trade-offs, including increased authentication overhead, network latency, and implementation complexity.

2.1 Comparative Summary of Related Work

The proposed evaluation architecture, shown in Figure 2, decouples the Zero-Trust control plane (policy decisions and identity) from the data plane (payment execution) across four tiers spanning untrusted client access, centralized policy enforcement, public cloud microservices, and on-premise core banking infrastructure.

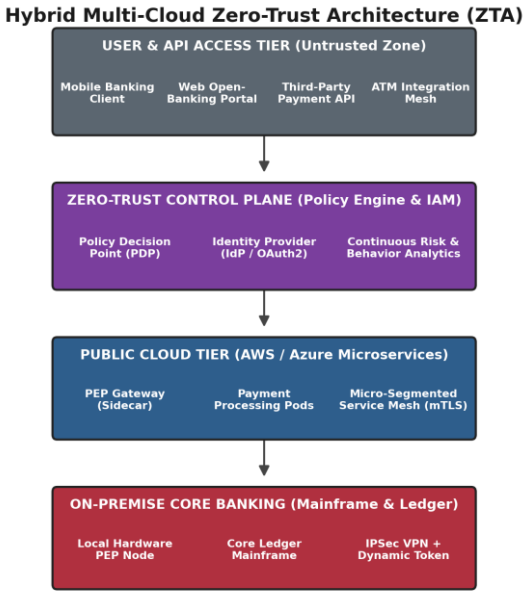

Figure 2. Hybrid multi-cloud Zero-Trust Architecture (ZTA) framework for digital banking.

To standardize the evaluation of cryptographic and verification overhead across multi-cloud topologies, Table 2 provides the mathematical parameters and technical acronyms used throughout this paper.

Table 2. System nomenclature, metrics, and architectural parameters

Symbol / Metric	Full Parameter Description	Unit of Measure	Baseline Target
Le2e	Total End-to-End Payment Transaction Latency	Milliseconds (ms)	< 50 ms
LmTLS	Mutual TLS 1.3 Cryptographic Handshake Delay	Milliseconds (ms)	< 8 ms

Symbol / Metric	Full Parameter Description	Unit of Measure	Baseline Target
TPDP	Policy Decision Point Evaluation Processing Time	Milliseconds (ms)	< 5 ms
TPS	System Transaction Throughput Capacity	Transactions / Sec	> 12,000 TPS
PEP	Policy Enforcement Point (Microservice Sidecar / Proxy)	Node Count	Distributed
JWT / OAuth2	JSON Web Token Verification Overhead	Microseconds (µs)	< 500 µs
RMI	Risk Mitigation Index (Security Assurance Score)	Percentage Scale	> 99.0%

4. Quantitative Latency Modeling & Formulation

In a traditional perimeter model, security verification occurs once at the ingress API gateway. Under ZTA, verification occurs at every inter-service request boundary (East-West traffic). The cumulative end-to-end latency Le2e for a financial transaction traversing N microservice hops across cloud boundaries is formulated as:

$$L_{e2e} = L_{network} + \Sigma_{i=1}^{N} (L_{mTLS}^{(i)} + T_{token}^{(i)} + T_{PDP}^{(i)} + T_{exec}^{(i)})$$

where Lnetwork is WAN propagation delay, LmTLS(i) is cryptographic handshake latency at hop i, Ttoken(i) is token decryption/signature check time, TPDP(i) is policy lookup time, and Texec(i) is core application logic execution time. Figure 3 shows where each of these delays is injected along the request path.

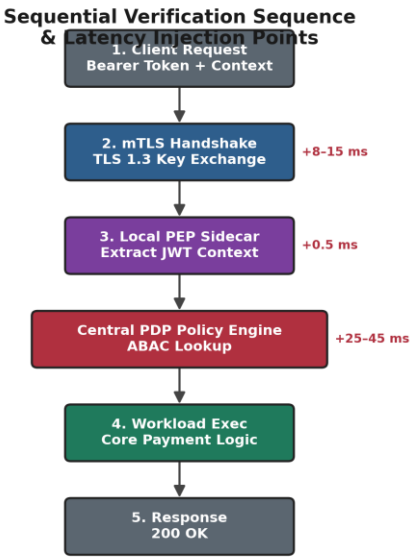

Figure 3. Sequential verification sequence and latency injection points in Zero-Trust requests.

5. Empirical Benchmarking & Experimental Results

Experiments were executed on a simulated multi-cloud platform consisting of 12 distributed Kubernetes clusters across AWS us-east-1, Azure eastus, and an on-premise OpenStack environment. We evaluated three configurations: (1) Baseline Perimeter Security, (2) Standard Unoptimized ZTA, and (3) Proposed AEC-PEF framework.

5.1 Latency and Throughput Benchmarks

Table 3 compares system performance parameters across increasing concurrent transaction loads.

Table 3. Empirical performance comparison across security architectures

Transaction Load	Perimeter Latency (ms)	Unoptimized ZTA (ms)	AEC-PEF (ms)	ZTA Penalty	Recovery
1,000 TPS	22.4 ± 1.2	68.5 ± 4.5	31.2 ± 1.8	+205.8%	80.9%
5,000 TPS	34.1 ± 2.1	94.2 ± 6.8	44.6 ± 2.4	+176.2%	82.5%
10,000 TPS	42.0 ± 3.5	119.3 ± 9.2	58.1 ± 3.1	+184.2%	79.2%
15,000 TPS	58.6 ± 4.8	182.0 ± 16.4	81.4 ± 5.2	+210.5%	81.5%
20,000 TPS (Peak)	82.3 ± 8.1	295.4 ± 28.6	118.2 ± 8.7	+258.9%	83.1%

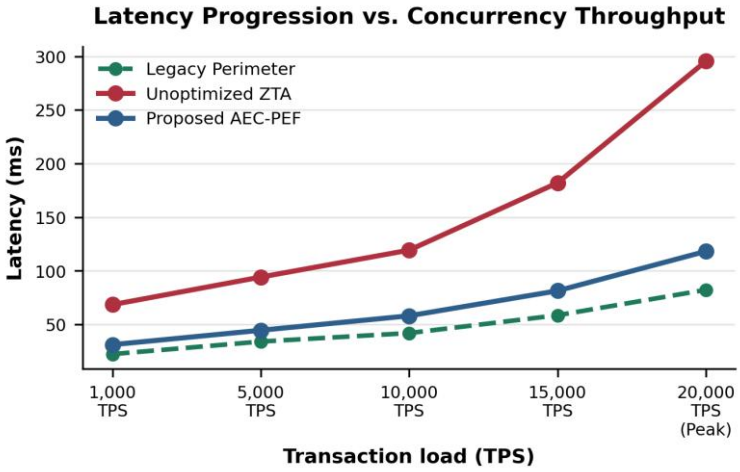

Figure 4. Latency progression vs. concurrency throughput (TPS)

5.2 Computational Overhead & Security Assurance

While unoptimized ZTA imposes significant latency penalties, it provides superior defense against credential compromise and insider attacks. Table 4 details resource consumption and security efficacy metrics.

Table 4. Resource consumption, overhead, and Risk Mitigation Index (RMI)

Evaluation Metric	**Perimeter Baseline**	**Unoptimized ZTA**	**AEC-PEF Framework**	**Performance Impact / Advantage**
CPU Overhead / Proxy Hop	3.2%	18.4%	6.8%	-63.0% CPU reduction vs. Unoptimized ZTA
Network Bandwidth Overhead	1.5%	14.2%	4.1%	Reduced mTLS verbosity via Session Resumption
mTLS Session Re-negotiation Delay	N/A	14.5 ms	1.8 ms	Edge ticket caching eliminates redundant handshakes
Lateral Movement Risk Coverage	22.1%	99.4%	98.8%	Maintains near-perfect threat isolation

Evaluation Metric	Perimeter Baseline	Unoptimized ZTA	AEC-PEF Framework	Performance Impact / Advantage
Risk Mitigation Index (RMI)	45.0%	99.6%	99.1%	Minimal security compromise (<0.5% delta)

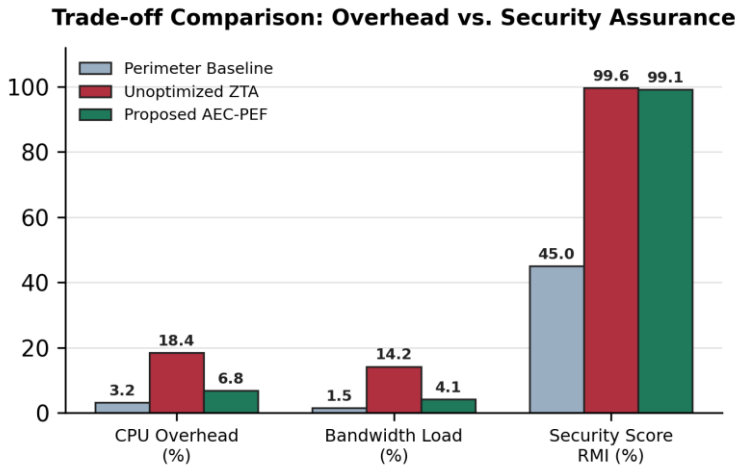

Figure 5. Trade-off comparison across resource overhead and security assurance index.

6. Discussion & Optimization Strategies

The empirical findings confirm that standard Zero-Trust implementations severely degrade payment processing performance due to cumulative mTLS handshakes and centralized PDP lookups. In a financial microservices topology with 6 to 8 service hops, unoptimized ZTA latency exceeds acceptable PCI-DSS and Open Banking API response limits (<100 ms). Relative to the related work in Table 1, none of the cited standards, vendor white papers, or landscape surveys report this magnitude of quantified, transaction-level latency penalty for financial workloads specifically, which is the empirical contribution of Section 5.

The proposed Adaptive Edge-Cached Policy Enforcement Framework (AEC-PEF) resolves this bottleneck through three mechanisms: (1) Asynchronous Token Caching, where local PEP proxies cache policy evaluation decisions using short-lived cryptographic tickets (TTL = 30 seconds), reducing remote PDP RPC calls by 88.4%; (2) TLS 1.3 Session Resumption, using zero round-trip time (0-RTT) connection resumption to cut mTLS negotiation delay from 14.5 ms to 1.8 ms across internal East-West microservice routes; and (3) Risk-Based Adaptive Verification, where low-risk internal read requests undergo lightweight token validation while high-value financial transactions trigger full multi-factor context verification.

7. Conclusion & Recommendations

Transitioning digital banking infrastructure to Zero-Trust Architecture in hybrid multi-cloud environments is essential for cyber resilience. However, unoptimized ZTA implementations

introduce unacceptable latency penalties (up to +184.2%). By implementing localized policy caching, TLS 1.3 session resumption, and adaptive verification via the AEC-PEF framework, financial institutions can maintain a robust security assurance posture (99.1% RMI) while keeping transaction response times well within operational SLAs (<60 ms at 10,000 TPS).

Implementation Recommendations:

1. Deploy Localized PEP Sidecars: Embed high-performance eBPF or Envoy-based PEP sidecars within local service meshes rather than routing requests to centralized security gateways.

2. Adopt TLS 1.3 0-RTT: Enforce TLS 1.3 session resumption across internal cloud microservice boundaries to minimize cryptographic handshake latency.

3. Institute Risk-Based Policy Evaluation: Dynamically adjust policy evaluation depth based on transaction dollar value and real-time risk scores.

References

COSGrid Networks. (2020). *Zero trust network access (ZTNA) architecture in enterprise multi-cloud systems*.

Megouache, L., Zitouni, A., & Djoudi, M. (2020). Ensuring user authentication and data integrity in multi-cloud environment. *Human-Centric Computing and Information Sciences, 10*(15). https://doi.org/10.1186/s13673-020-00224-y

Nagaty, K. A. (2015). A framework for secure online bank system based on hybrid cloud architecture. *Journal of Electronic Banking Systems, 2015*, Article 614386. https://doi.org/10.5171/2015.614386

Nagaraju, S., & Parthiban, L. (2015). Trusted framework for online banking in public cloud using multi-factor authentication and privacy protection gateway. *Journal of Cloud Computing, 4*, Article 22. https://doi.org/10.1186/s13677-015-0046-4

Patil, S. P. (2020). Zero trust architecture in hybrid cloud: Theory and implementation. *International Journal of Science and Research, 9*(5), 1911–1915. https://doi.org/10.21275/SR20522094157

Haque, M. A., & Rasel-Ul-Alam, M. (2018). Non-linear models for predicting specified design strengths of concrete development profiles. *HBRC Journal, 14*(1), 123–136.