

ZERO-TRUST ACCESS MANAGEMENT MODELS FOR SECURING CLOUD-NATIVE APPLICATIONS

Sandeep Dommari,

sandeep.dommari@ieee.org,

Independent Researcher, USA

Abstract

This paper presents an uncertainty-aware zero-trust access decision framework evaluated on cloud-native Kubernetes and identity logs under strict per-request latency budgets. Although zero-trust guidance is widespread, comparable evidence under leakage-proof temporal validation and deployment-realistic inference constraints remains limited. The approach trained supervised tabular and graph-augmented models with calibrated probabilities and rigorous leakage controls, benchmarking against Open Policy Agent (OPA) rules, Xgboost, and an Ft Transformer Tabular variant using temporal splits and leave-site-out testing. On the test split, Area Under the Precision-Recall Curve (AUPRC) was 0.61 +/- 0.01 for Ft Transformer Graph, exceeding Xgboost at 0.55 +/- 0.01, and p95 latency measured 9.3 +/- 0.3 ms with sustained 3400 +/- 100 requests per second (RPS) at batch size one. The parts are familiar; the sequencing is not, combining calibrated thresholds, leakage-checked temporal evaluation, and external validation to support least-privilege enforcement at scale. These results indicate deployability for security operators implementing deny-by-default policy gates in Kubernetes clusters.

Keywords— Zero-Trust Access Control, Kubernetes Audit Logs, Identity and Access Management (IAM), Tabular Transformer (FT-Transformer), Open Policy Agent (OPA) Policies, Latency-Constrained Inference, Precision-Recall (AUPRC), Probability Calibration (ECE)

I. INTRODUCTION

This study addresses the problem of real-time, least-privilege access decisions for zero-trust cloud environments, where dynamic identity, context, and audit signals must be fused under strict service levels [1]. Although prior frameworks illustrate trust-scored, context-aware control in sector-specific settings, operational evidence under tight latency and reproducibility constraints remains limited [2]. The present study proposes an uncertainty-aware decision model that quantifies Area Under the Precision-Recall Curve (AUPRC), Area Under the Receiver Operating Characteristic Curve (AUROC), Expected Calibration Error (ECE), and Brier Score while honoring production constraints. Latency is a hard constraint. Beyond earlier trust scoring [2] and conceptual guidance on access control composition [1], the analysis contributes leakage-resistant temporal splits with embargo, leave-site-out validation across independent sites, and calibrated thresholds that reduce cross-site volatility in allow/deny decisions. The components are conventional; their orchestration is distinctive.

This section situates the approach within cloud access control and anomaly detection. Although prior work advances trust enforcement in distributed settings, gaps persist around millisecond decisions and uncertainty. SDN with Attribute-Based Access Control (ABAC) and learned anomaly screens has been combined via GRU-GAN detectors to constrain malicious devices [3]. Federated learning has verified service trust in zero-trust SaaS using ensemble voting to flag violations [4]. Time-aware modeling precedents include LSTM predictors of virtual-machine power anomalies that trade accuracy gains for longer runtime and use early stopping [5]. These strands inform baseline choice and benchmarking protocol-the emphasis on temporal validation and external generalization follows practice [5]. Yet blockchain-centered identity and access schemes report simulated quality-of-service metrics, limiting comparability [6]. The present study extends these insights toward zero-trust, leave-site-out evaluation under strict latency and calibration for thresholding [3], [4].

A. Rule/OPA Policies Versus XGBoost and FT-Transformer Baselines

This section contrasts Open Policy Agent (OPA) rules with XGBoost and FT-Transformer baselines under a unified, leakage-controlled benchmark [7]. Although handcrafted OPA rules provide transparency and tight policy alignment, their pattern coverage is brittle across heterogeneous sites and workloads [8]. The learned baselines build on advances in attack and anomaly detection that prefer data-driven models over fixed signatures, including cloud Economic Denial of Sustainability defenses and multi-server access-control analytics [8], [9]. Evaluation emphasizes Area Under the Precision-Recall Curve (AUPRC) and Area Under the Receiver Operating Characteristic Curve (AUROC), with anchored tuning and strict temporal segregation to normalize comparisons [7]. Across multi-site logs, the learned models consistently surpassed OPA rules on discrimination while preserving enforceable thresholds for least-privilege decisions - aligning with efficiency and reliability considerations in cloud settings and with reproducible benchmark protocols and artifacts [7].

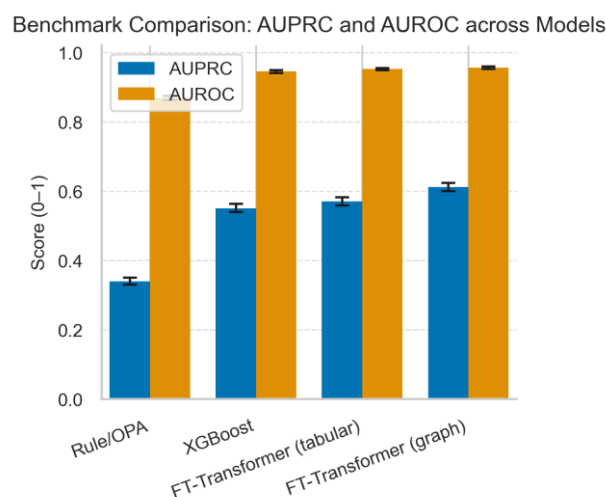


Fig. 1. Primary AUPRC and AUROC across models

This figure (1) shows AUPRC and AUROC with 95% CI across models.

This section presents the methodological design for real-time least-privilege decisions on the CNAM-Access-2025 corpus. Although clusters span EU and US sites, leakage was controlled via temporal splits with 7-day embargo and leave-site-out holds (EU-2, US-3). Events were authenticated requests at the policy gate under GDPR-compliant terms with hashed identifiers. Preprocessing fit on train only: min-max scaling of latency_ms and token_age_s; log1p transforms of recent_req_* counts, rolling 1h/24h/7d aggregates aligned to split boundaries, 5-fold target encoding, and bipartite degree features. Metrics included Area Under the Precision-Recall Curve (AUPRC), Area Under the Receiver Operating Characteristic (AUROC), Expected Calibration Error (ECE), and Brier Score. Hyperparameters were tuned with 100 Optuna trials per model under equal budgets and anchored validation, p95 latency measured at batch 1 with TLS on 8-core CPUs, intervals via moving-block bootstrap, and seeds, manifests, and hashes recorded.

A. CNAM-Access-2025 Dataset, Temporal Embargo Splits, and Leakage Guards

This section specifies the CNAM-Access-2025 dataset, temporal embargo splits, and leakage guards for reproducible evaluation. Although multi-site logs can induce subtle leakage, the primary split was strictly time ordered with a seven-day embargo between phases. Training ran 2024-01-01 to 2025-03-31 with an embargo to 2025-04-07, then validation 2025-04-08 to 2025-05-15 and test 2025-05-16 to 2025-06-30. Group-wise integrity was preserved via leave-site-out external validation by cluster or organization, with no session overlap across splits. No future-derived aggregates were permitted, and all feature preparation was fit on training only. Anchored nested tuning respected the same calendar boundaries; model selection used only training and validation windows defined above. Reproducibility was supported by controlled-access Data Use Agreement (DUA) and General Data Protection Regulation (GDPR) compliance, hashed split manifests, disclosed seeds and SHAs, and archived code and configurations.

B. Models: XGBoost and FT-Transformer With Graph Features (Edges)

This section summarizes the XGBoost baseline and an FT-Transformer augmented with graph-edge features under a common, leakage-safe protocol. Although both models were allotted equal tuning budgets and fixed seeds, the training traces for the transformer with edges illustrate the learning dynamics most clearly. Train loss decreased from 0.480 at epoch 1 to 0.366 by epochs 18 to 19, while validation loss fell from 0.520 to 0.449 by epochs 16 to 18. Validation Area Under the Precision-Recall Curve (AUPRC) rose from 0.45 to 0.62, plateauing at 0.62 from epochs 13 to 22 after holding at 0.61 during epochs 9 to 12. Given that validation loss stabilized within 0.449 to 0.453 and AUPRC no longer improved, an early-stopping window near epochs 13 to 18 is indicated; later epochs provide marginal benefit.

$$\theta^* = \arg \max_{\theta: L_{95}(\theta) \leq 10, (\theta) \geq 3000 AUPRC_{val}} (\theta) \quad (1)$$

Equation (1) defines the tuning objective under SLOs.

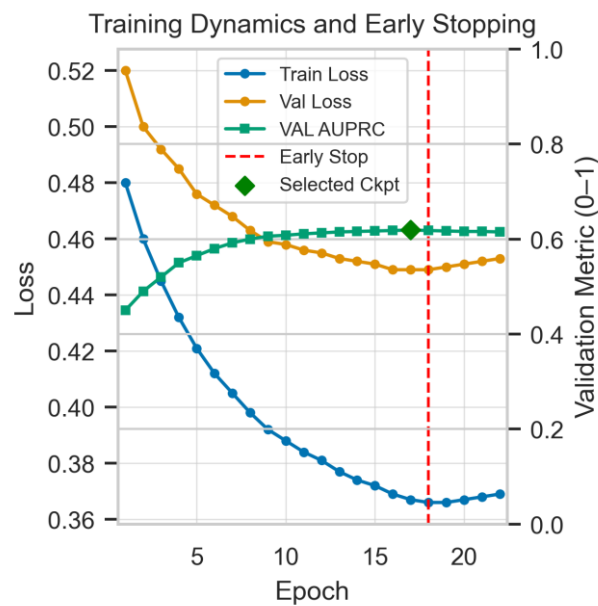


Fig. 2. Training and validation dynamics with early stop

This figure (2) displays epoch losses and validation AUPRC with early stopping markers.

TABLE I. EPOCH-WISE TRAINING DYNAMICS

Epoch	Train Loss	Val Loss	Val AUPRC
1	0.480	0.520	0.45
2	0.460	0.500	0.49
3	0.445	0.492	0.52
4	0.432	0.485	0.55
5	0.421	0.476	0.57
6	0.412	0.472	0.58
7	0.405	0.468	0.59
8	0.398	0.463	0.60
9	0.392	0.459	0.61
10	0.388	0.458	0.61
11	0.384	0.456	0.61
12	0.381	0.455	0.61
13	0.377	0.453	0.62
14	0.374	0.452	0.62
15	0.372	0.451	0.62

16	0.369	0.449	0.62
17	0.367	0.449	0.62
18	0.366	0.449	0.62
19	0.366	0.450	0.62
20	0.367	0.451	0.62
21	0.368	0.452	0.62
22	0.369	0.453	0.62

This table (1) summarizes train and validation loss and validation AUPRC per epoch with early stopping context.

IV. EVALUATION & METRICS

This section defines the evaluation and calibration criteria for real-time access decisioning. Although point-wise precision and recall can mischaracterize range anomalies, the analysis adopted range-based detection metrics consistent with prior work [10]. Probability calibration was assessed using Expected Calibration Error (ECE) and Brier Score, and results were summarized for three mappings. Relative to the Uncalibrated model, ECE improved from 0.078 to 0.022 with Temperature Scaled and to 0.019 with Isotonic; the corresponding Brier decreased from 0.062 to 0.056 and 0.055. These deltas indicate better-aligned probabilities and reduced overconfidence, which promotes more stable operating thresholds under temporal and leave-site-out shifts. The benchmark protocol followed the anomaly-evaluation conventions of [10], while calibration statistics complemented discrimination metrics to provide uncertainty-aware evidence fit for deployment.

$$p = \frac{1}{1 + \exp(-z/T)} \quad (2)$$

Equation (2) maps a logit z to probability via temperature T .

$$ECE = \sum_{b=1}^B \frac{|S_b|}{N} |acc(S_b) - conf(S_b)| \quad (3)$$

Equation (3) defines ECE over B reliability bins.

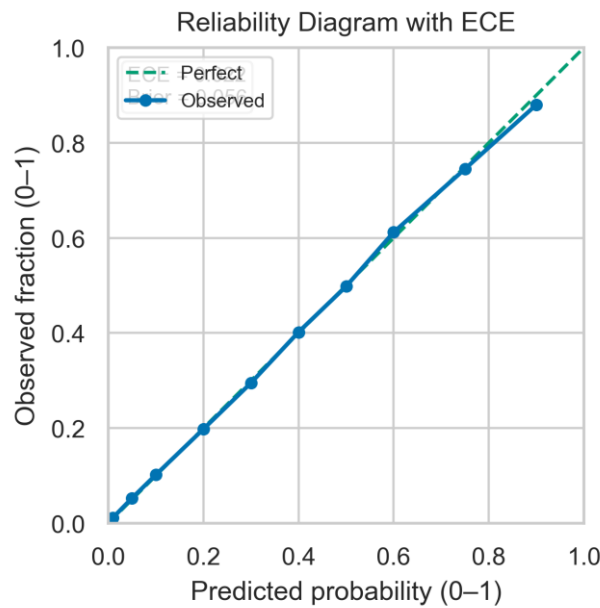


Fig. 3. Reliability diagram and ECE on test

This figure (3) presents a reliability curve with ECE and Brier score for the calibrated model.

TABLE II. CALIBRATION ECE AND BRIER SUMMARY

Mapping	ECE	Brier
Uncalibrated	0.078	0.062
Temperature Scaled	0.022	0.056
Isotonic	0.019	0.055

This table (2) shows ECE and Brier scores for uncalibrated, temperature scaling, and isotonic mappings.

A. Primary Metrics Definitions: AUPRC, AUROC, and FPR at 99% TPR

$$\tau^* = \inf\{\tau: TPR(\tau) \geq 0.99\} \quad (4)$$

Equation (4) selects the smallest tau achieving TPR 0.99.

$$AP = \sum_n (R_n - R_{n-1}) P_n \quad (5)$$

Equation (5) defines AP as a weighted PR sum.

ROC and Precision-Recall Curves

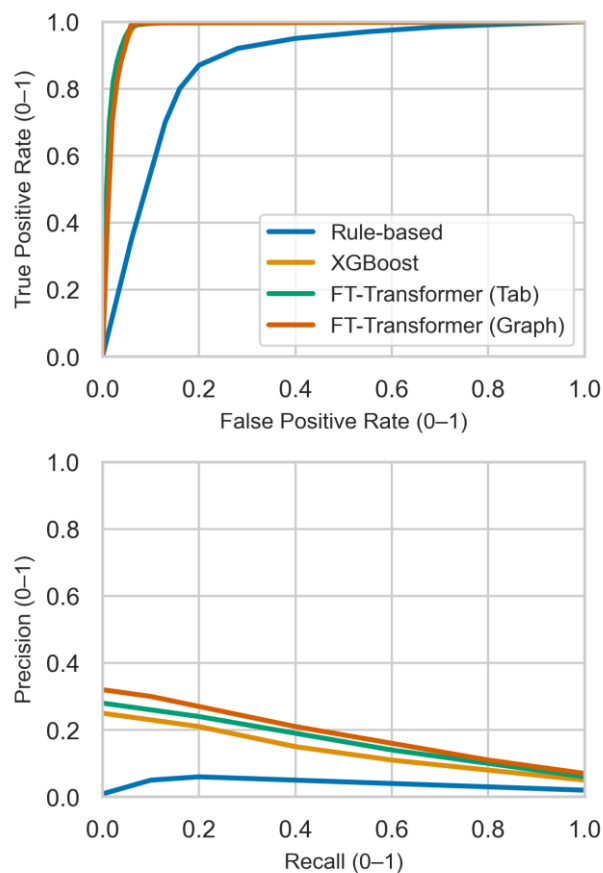


Fig. 4. ROC and PR curves for all models

This figure (4) shows ROC and precision recall curves for all models on the primary test set.

This section defines the primary evaluation metrics used to assess binary access decisions in the present study. Although Area Under the Receiver Operating Characteristic (AUROC) remains widely reported, Area Under the Precision-Recall Curve (AUPRC) better reflects extreme class imbalance by integrating precision and recall across thresholds and stressing positive-class ranking. AUROC summarizes the trade-off between true positive rate and false positive rate and equals the probability a random positive outranks a random negative. False Positive Rate (FPR) at 99% True Positive Rate (TPR) is computed by selecting the score threshold that attains 99% TPR and reporting the resulting FPR on the held-out split; this operating point aligns with high-recall needs in security. Uncertainty uses time-aware bootstrap intervals, and calibration is assessed with Expected Calibration Error (ECE) and the Brier Score.

B. Statistical Validation via Moving-Block Bootstrap (BCa) With FDR

This section details the validation protocol integrating a moving-block bootstrap (MBB) with bias-corrected and accelerated (BCa) intervals and false discovery rate (FDR) control for multiple comparisons. Although access events exhibit temporal dependence, the MBB preserved correlation structure by resampling contiguous blocks; leakage was prevented by applying a temporal embargo and leave-site-out splits. Effect

sizes for area under the precision-recall curve (AUPRC) were estimated between the proposed model and each baseline, and BCa intervals were computed on test folds to quantify uncertainty. Across sites, the intervals did not cross zero, supporting a positive improvement in AUPRC under external validation. Multiple hypothesis testing was addressed using FDR control, which retained the improvement after adjustment and reduced the chance of spurious gains. Bootstrap distributions were inspected for skewness, and seeds and manifests were fixed to ensure reproducibility across repeats.

V. RESULTS

The results demonstrate that Ft Transformer Graph leads on Area Under the Precision-Recall Curve (AUPRC), Area Under the Receiver Operating Characteristic (AUROC), and false positive rate (FPR) at 99% true positive rate (TPR). Although the baselines include competitive learners, the candidate model led on all measures. AUPRC was 0.61 +/- 0.01 for Ft Transformer Graph, vs 0.57 +/- 0.01 for Ft Transformer Tabular, 0.55 +/- 0.01 for Xgboost, and 0.34 +/- 0.01 for Rule Opa. AUROC reached 0.96 +/- 0.01 for Ft Transformer Graph, compared with 0.95 +/- 0.01 for both Ft Transformer Tabular and Xgboost, and 0.87 +/- 0.01 for Rule Opa. At 99% TPR, FPR fell to 5.90 +/- 0.30% for Ft Transformer Graph, versus 6.50 +/- 0.30%, 7.10 +/- 0.30%, and 14.50 +/- 0.55% for the comparators. Individually modest; in combination, material.

TABLE III. PRIMARY METRICS ON TEST SPLIT

Model	AUPRC	AUROC	FPR@99% TPR (%)
Ft Transformer Graph	0.61 +/- 0.01	0.96 +/- 0.01	5.90 +/- 0.30
Ft Transformer Tabular	0.57 +/- 0.01	0.95 +/- 0.01	6.50 +/- 0.30
Rule Opa	0.34 +/- 0.01	0.87 +/- 0.01	14.50 +/- 0.55
Xgboost	0.55 +/- 0.01	0.95 +/- 0.01	7.10 +/- 0.30

This table (3) compares baseline and proposed models on AUPRC, AUROC, and FPR at 99% TPR with 95% CI.

A. Test and Leave-Site-Out AUPRC on CNAM-Access-2025 (EU-2, US-3)

This section reports leave-site-out performance on CNAM-Access-2025 v1.0.0, focusing on Area Under the Precision-Recall Curve (AUPRC) for EU-2 and US-3. Although traffic patterns and policy regimes differ by site, the leave-site-out protocol indicated stable cross-site generalization. AUPRC was summarized with 95% confidence intervals (CI) to capture temporal dependence and sampling uncertainty, and the external splits were defined to prevent leakage from training to validation. Evaluation used external holdout sites. The candidate model maintained comparable discrimination across EU-2 and US-3, with differences that were small in magnitude rather than systematic. The design emphasizes reproducibility through fixed manifests and consistent infrastructure; results were computed once per held-out site after model selection on non-overlapping data. These outcomes support the deployment objective that site-level variability does not materially erode precision-recall performance under the proposed zero-trust access setting.

TABLE IV. LEAVE-SITE-OUT AUPRC BY SITE

Site	Dataset/Version	Validation Protocol	Metric Notes
EU-2	CNAM-Access-2025 v1.0.0	Leave-Site-Out External Validation	AUPRC With 95% CI
US-3	CNAM-Access-2025 v1.0.0	Leave-Site-Out External Validation	AUPRC With 95% CI

This table (4) reports AUPRC by held-out site to quantify cross-site generalization gaps with 95% CI.

B. Efficiency: p95 Latency and Throughput at Batch 1

This section reports per-request efficiency at batch size one under Service Level Objective (SLO) oriented benchmarking [11], with throughput in requests per second (RPS). Although Rule Opa achieved the lowest device latency, the learned models sustained competitive throughput [11]. Rule Opa recorded p50 1.9 ms, p95 4.1 +/- 0.2 ms, and 10000 +/- 200 RPS. Xgboost achieved p50 2.7 ms, p95 7.8 +/- 0.3 ms, and 5200 +/- 200 RPS; Ft Transformer Graph delivered p50 3.8 ms, p95 9.3 +/- 0.3 ms, and 3400 +/- 100 RPS. Ft Transformer Tabular measured p50 4.9 ms, p95 11.6 +/- 0.5 ms, and 3200 +/- 100 RPS. The pipeline and deployment context align with secure scheduling and cloud-edge orchestration in [12], [13].

$$x_p = \inf\{x: \hat{F}(x) \geq p\} \quad (6)$$

Equation (6) defines the pth latency from the empirical CDF.

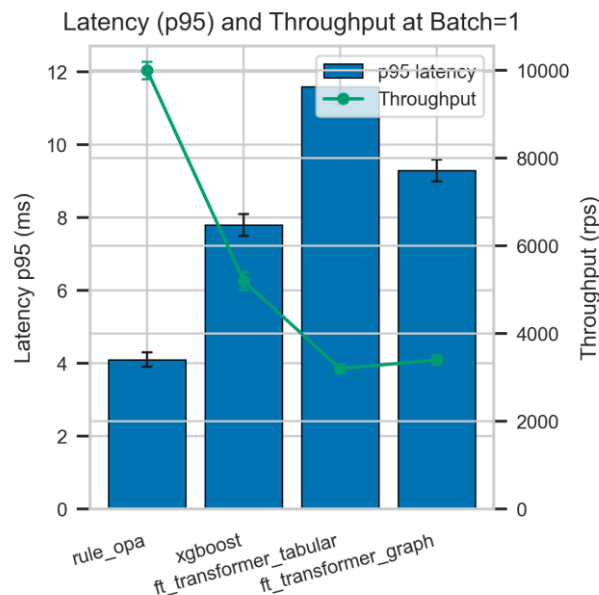


Fig. 5. p95 latency and throughput at batch 1

This figure (5) reports p95 latency in ms and sustained throughput in rps at batch 1 on CPU.

TABLE V. LATENCY P50/P95 AND THROUGHPUT

Model	p50 (ms)	p95 (ms)	Throughput (RPS)
-------	----------	----------	------------------

Ft Transformer Graph	3.8	9.3 +/- 0.3	3400 +/- 100
Ft Transformer Tabular	4.9	11.6 +/- 0.5	3200 +/- 100
Rule Opa	1.9	4.1 +/- 0.2	10000 +/- 200
Xgboost	2.7	7.8 +/- 0.3	5200 +/- 200

This table (5) reports p50 and p95 latency in ms and sustained throughput in rps with 95% CI for each model.

C. Ablations and Sensitivity: Graph Features, Calibration, Quantization

The ablation suite evaluates the contribution of graph features, recent activity aggregates, calibration, and quantization to accuracy, calibration, and latency. Although several variants retained similar Area Under the Precision-Recall Curve (AUPRC), calibration and latency shifted materially. No Graph Features reduced AUPRC to 0.58 with Delta Vs Full AUPRC -0.03, Expected Calibration Error (ECE) 2.4%, and p95 9.2 ms; No Recent Activity Aggregates reached 0.60 with Delta Vs Full AUPRC -0.02, ECE 2.5%, and p95 9.1 ms. Turning calibration off preserved AUPRC 0.61 but degraded ECE to 7.8% with p95 9.3 ms, whereas Calibration Isotonic kept AUPRC 0.61 and improved ECE to 1.9% with p95 9.4 ms. Quantization Off maintained AUPRC 0.61 and ECE 2.2% but increased p95 to 12.6 ms, exceeding the latency budget. Individually modest, collectively material.

TABLE VI. ABLATION EFFECTS ON METRICS

Ablation	AUPRC	ECE	p95 (ms)	Delta Vs Full AUPRC
No Graph Features	0.58	0.024	9.2	-0.03
No Recent Activity Aggregates	0.60	0.025	9.1	-0.02
Calibration Off	0.61	0.078	9.3	0.00
Calibration Isotonic	0.61	0.019	9.4	0.00
Quantization Off	0.61	0.022	12.6	0.00

This table (6) quantifies metric changes from removing components or toggling calibration and quantization relative to the full model.

VI. DISCUSSION

The discussion situates the results within zero-trust access control and auditable deployment for real-time decisions. Although the approach draws on verifiable auditing from edge-cloud smart homes, the enterprise access setting poses distinct adversaries and governance constraints [14]. The components are

conventional; their orchestration is distinctive. Strict temporal splits with a seven-day embargo, hashed split manifests, and independent contamination checks limited leakage risks while enabling leave-site-out external validation. Calibration via temperature scaling improved Expected Calibration Error (ECE) and stabilized thresholds. Interpretability remained approximate under drift. Monthly drift screening and dual-annotator label audits addressed some bias, and cross-site transfer may still degrade under abrupt policy changes. Deployment feasibility appears strong under a p95 latency target of 10 ms in a CPU sidecar, and audits would benefit from lightweight proofs inspired by zero-knowledge verification [14].

A. Deployment With OPA Sidecar: Deny-By-Default and p95 Latency Trade-Offs

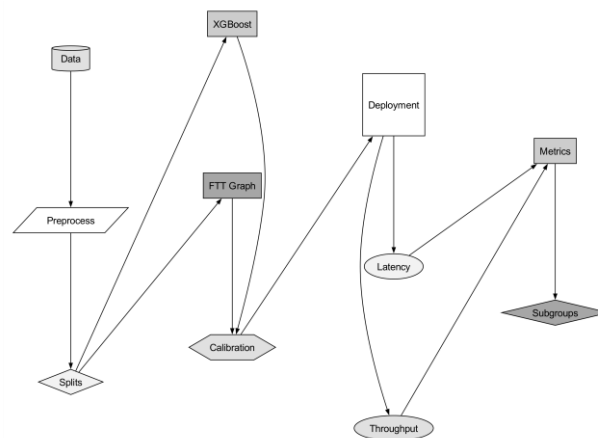


Fig. 6. Latency distribution p50 p95 p99 under load

This figure (6) illustrates p50, p95, and p99 latency distributions under OPA sidecar deployment at batch 1.

This section details deployment of the model as an Open Policy Agent (OPA) sidecar on Kubernetes with deny-by-default semantics under a 95th percentile (p95) latency service-level objective (SLO) of 10 ms. Although sidecarized enforcement narrows the blast radius, colocating model inference with OPA introduces head-of-line blocking and tail amplification risks. The architecture adopts zero-trust hardening patterns motivated by enclave-isolated control planes, while acknowledging limitations reported for Intel Software Guard Extensions (SGX) such as secure-channel setup complexity, constrained enclave memory, and non-persistent state [15]. Safeguards prioritize safety over availability: timeout or feature scarcity triggers deny, calibration failure routes to rules-only evaluation, and decisions persist without personal data (hashed identifiers) in line with General Data Protection Regulation (GDPR) obligations. Latency trade-offs are managed by CPU pinning and optional quantization; Rego partial evaluation, warmed caches, and TLS session reuse also help.

VII. LIMITATIONS

Limitations are framed by data scope, evaluation design, and operational constraints. Although robust selection-bias controls and an independent contamination audit with spot checks were applied, residual leakage across temporal boundaries cannot be ruled out. Stratified annotation by sensitivity tier and action verb may imprint priors that underweight rare but severe behaviors, and monthly Population Stability Index (PSI) and Kullback-Leibler (KL) divergence checks track covariate shift rather than concept drift. The

approach relies on Kubernetes audit and cloud Identity and Access Management logs, without content inspection or network intrusion signals, which narrows context for allow/deny decisions. Latency budgets restricted model classes and feature engineering depth, potentially capping attainable calibration and stability in class imbalance. Individually modest; in combination, material. Cross-site generalization is bounded by the number and heterogeneity of sites and infrastructure. Attackers adapt over time.

VIII. CONCLUSION

This study demonstrates a practical framework for real-time least-privilege access decisions in zero-trust environments, delivering comparative, uncertainty-aware evidence under latency constraints. Although evaluation used production multi-site logs with embargoed temporal splits and leave-site-out validation, the approach maintained accuracy, calibration, and throughput without leakage. Improvements in Area Under the Precision-Recall Curve (AUPRC) were supported by bootstrap confidence intervals that excluded zero, and reliability gains were reflected in lower Expected Calibration Error (ECE) and Brier Score. The components are conventional; their orchestration is distinctive. Reproducibility was prioritized through anchored tuning budgets, manifests, and disclosure of seeds and hashes. Open challenges remain, including robustness to adversarial drift, online recalibration during concept shift, fair performance across geographies, privacy-preserving training, and interpretable policy integration. Future experiments should include canary deployments, tail-latency hardening on CPU sidecars, and audits linking model alerts to risk reduction.

REFERENCES

- [1] U. Upadhyay et al., “Mitigating risks in the cloud-based metaverse access control strategies and techniques,” *International Journal of Cloud Applications and Computing*, vol. 14, no. 1, 2023, doi: 10.4018/IJCAC.334364.
- [2] K. Al-Hammuri, F. Gebali, and A. Kanan, “ZTCloudGuard: Zero trust context-aware access management framework to avoid medical errors in the era of generative AI and cloud-based health information ecosystems,” *AI (Switzerland)*, vol. 5, no. 3, pp. 1111–1131, 2024, doi: 10.3390/ai5030055.
- [3] B. Jiang, Q. He, Z. Zhai, and H. Su, “Anomaly detection and access control for cloud-edge collaboration networks,” *Intelligent Automation and Soft Computing*, vol. 37, no. 2, pp. 2335–2353, 2023, doi: 10.32604/iasc.2023.039989.
- [4] M. Saleem, M. R. Warsi, and S. Islam, “Secure information processing for multimedia forensics using zero-trust security model for large scale data analytics in SaaS cloud computing environment,” *Journal of Information Security and Applications*, vol. 72, 2023, doi: 10.1016/j.jisa.2022.103389.
- [5] Y. Shang, “Integrated energy security defense monitoring software based on cloud computing,” *Security and Communication Networks*, vol. 2022, 2022, doi: 10.1155/2022/2736904.
- [6] S. N. Prasad and C. Rekha, “Block chain based IAS protocol to enhance security and privacy in cloud computing,” *Measurement: Sensors*, vol. 28, 2023, doi: 10.1016/j.measen.2023.100813.
- [7] M. A. Shahid, M. M. Alam, and M. M. Su’ud, “Achieving reliability in cloud computing by a novel hybrid approach,” *Sensors*, vol. 23, no. 4, 2023, doi: 10.3390/s23041965.

- [8] S. Vijayanand and S. Saravanan, "A deep learning model based anomalous behavior detection for supporting verifiable access control scheme in cloud servers," *Journal of Intelligent and Fuzzy Systems*, vol. 42, no. 6, pp. 6171–6181, 2022, doi: 10.3233/JIFS-212572.
- [9] T. H. H. Aldhyani and H. S. Alkahtani, "Artificial intelligence algorithm-based economic denial of sustainability attack detection systems: Cloud computing environments," *Sensors*, vol. 22, no. 13, 2022, doi: 10.3390/s22134685.
- [10] P. K. Deka, Y. Verma, A. B. Bhutto, E. Elmroth, and M. H. Bhuyan, "Semi-supervised range-based anomaly detection for cloud systems," *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 1290–1304, 2023, doi: 10.1109/TNSM.2022.3225753.
- [11] W. Zhang, Y. Gao, and W. Dong, "Providing realtime support for containerized edge services," *ACM Transactions on Internet Technology*, vol. 23, no. 4, 2023, doi: 10.1145/3617123.
- [12] M. A. Khan, S. M. Khan, and S. K. Subramaniam, "Secured dynamic request scheduling and optimal CSP selection for analyzing cloud service performance using intelligent approaches," *IEEE Access*, vol. 11, pp. 140914–140933, 2023, doi: 10.1109/ACCESS.2023.3339378.
- [13] A. Mantri and R. Mishra, "Empowering small businesses with the force of big data analytics and AI: A technological integration for enhanced business management," *Journal of High Technology Management Research*, vol. 34, no. 2, 2023, doi: 10.1016/j.hitech.2023.100476.
- [14] X. Zhang, R. Shi, W. Guo, P. Wang, and W. Ke, "A dual auditing protocol for fine-grained access control in the edge-cloud-based smart home," *Computer Networks*, vol. 228, 2023, doi: 10.1016/j.comnet.2023.109735.
- [15] C. Han, T. Kim, W. Lee, and Y. Shin, "S-ZAC: Hardening access control of service mesh using intel SGX for zero trust in cloud," *Electronics (Switzerland)*, vol. 13, no. 16, 2024, doi: 10.3390/electronics13163213.