SSR Journal of Engineering and Technology (SSRJET)

Homepage: https://ssrpublisher.com/ssrjet/

Volume 3, Issue 8, 2026 ISSN: 3049-0383

A Zero Trust Architecture for Protecting Cyber-Physical Systems in the Aviation Sector

Odey Peter Wonah, Ohagwam Chidinma Maureen and Onoja Sunday Onoja

Centre for Cyberspace Studies, Department of Cybersecurity, Nasarawa State University, Keffi, Nigeria

Received: *11.06.2026* | ***Accepted:*** *24.07.2026* | ***Published:*** *07.08.2026*

***Corresponding author:** Ohagwam Chidinma Maureen

DOI: *10.5281/zenodo.21837972*

Abstract **Original Research Article**

Contemporary aviation depends on tightly coupled information-technology, operational-technology and cyber-physical assets spanning aircraft, airports, air-traffic services, maintenance, ground operations and cloud platforms. Conventional perimeter security provides insufficient protection once credentials, suppliers, mobile devices or trusted network segments are compromised. This study proposes a Safety-Aware Adaptive Zero Trust Architecture (SA-AZTA) that continuously evaluates human, device, service and workload requests using identity confidence, device integrity, behavioural conformity, mission context, threat intelligence, privilege risk, asset criticality and prospective safety impact. A safety gate constrains automated response so that suspicious activity can be restricted, isolated or escalated without indiscriminately interrupting safety-critical services. The architecture was evaluated in a reproducible synthetic major-airport case study. A detailed main run comprised 50,000 access events, while 30 Monte Carlo runs of 20,000 events each compared perimeter security, role-based access control, static attribute-based access control, non-adaptive Zero Trust and SA-AZTA. Across the Monte Carlo runs, SA-AZTA achieved mean accuracy of 96.99%, precision of 92.78%, recall of 92.10%, F1-score of 92.43%, ROC-AUC of 99.25% and a false-positive rate of 1.79%. Mean attack containment was 93.09%, service availability was 99.40% and policy-decision latency was 7.58 ms. The improvement over every baseline was statistically significant in paired Wilcoxon tests ($p < 1.9 \times 10^{-9}$), although the proposed model introduced approximately 2.17 ms additional latency relative to non-adaptive Zero Trust. Ablation analysis showed that behavioural analytics contributed most to detection performance, whereas microsegmentation contributed most to limiting lateral movement. The results provide evidence that context-rich, safety-constrained Zero Trust can improve cyber resilience in aviation CPS; however, findings remain simulation-based and require digital-twin, hardware-in-the-loop and regulated operational validation.

Keywords: zero trust architecture; aviation cybersecurity; cyber-physical systems; adaptive access control; safety-security co-engineering; microsegmentation; trust modelling; synthetic simulation.

1. Introduction

Civil aviation is increasingly organised as a system of interconnected cyber-physical systems rather than a collection of isolated aircraft and airport technologies. Digital flight operations, airport operational databases, passenger processing, baggage handling, maintenance applications, surveillance sensors, air-traffic interfaces, mobile devices, cloud services and

third-party platforms exchange data across organisational and technical boundaries. Sampigethaya and Poovendran described this evolution as an aviation cyber-physical ecosystem in which computation and communication increasingly shape physical operations [11]. The resulting connectivity creates operational efficiency, but also permits cyber events to propagate into service disruption, erroneous situational awareness or unsafe operating conditions.

Aviation security practice has traditionally relied on certification boundaries, network zoning, trusted organisational roles and perimeter controls. These controls remain necessary, yet they are insufficient when an attacker operates with valid credentials, compromises an authorised maintenance device, exploits a supplier connection or moves laterally after entering a trusted segment. Reviews of aviation cybersecurity consistently identify expanding attack surfaces, heterogeneous legacy systems, unencrypted or weakly authenticated communications, remote maintenance, insider threats and supply-chain dependence as persistent challenges [12-14]. The wider CPS literature likewise shows that security mechanisms must account for physical consequences, recovery and safety rather than treating confidentiality as the only objective [15].

Zero Trust changes the decision premise. NIST defines Zero Trust as an architecture that removes implicit trust based on network location and requires explicit, continuously informed decisions about access to resources [1]. NIST SP 800-207A extends this principle to granular application-level policy in hybrid and multi-location environments [2], while the CISA maturity model organises implementation around identity, devices, networks, applications/workloads and data [3]. These principles align well with aviation's distributed architecture, but enterprise Zero Trust cannot be transferred uncritically to safety-critical CPS. A policy engine that immediately disconnects a suspicious component may protect data while creating an operational hazard. The problem is therefore not simply how to deny more requests; it is how to make risk-adaptive decisions while preserving essential safety functions.

This article addresses that problem through SA-AZTA, an aviation-specific architecture that integrates continuous verification, behaviour-aware trust, microsegmentation, threat intelligence, asset criticality and a safety-constrained response gate. The work is evaluated through a transparent synthetic case study rather than confidential operational data. The simulation compares the proposed model with four progressively stronger baselines under compromised-credential, insider, lateral-movement, device-compromise, false-data-injection, ransomware and supply-chain scenarios.

The research contributions are:

- An aviation-specific Zero Trust reference architecture that spans human, device, service and workload identities across airport IT, OT, cloud and third-party domains.
- A composite trust-and-risk model that explicitly includes asset criticality and potential safety impact rather than relying only on identity and network posture.
- A bounded safety gate that maps cyber decisions to allow, step-up, restrict, quarantine or supervised degradation actions according to operational criticality.
- A reproducible synthetic benchmark and Monte Carlo evaluation against perimeter, RBAC, static ABAC and non-adaptive Zero Trust baselines.
- A component-level ablation, sensitivity and scalability analysis that identifies the security value and operational cost of behavioural analytics, device posture, microsegmentation, threat intelligence and safety gating.

2. Background and Theoretical Foundations

2.1 Aviation cyber-physical systems

An aviation CPS integrates sensing, computation, communication, decision support and physical processes. At airport level, examples include baggage conveyors, access-control devices, surveillance cameras, passenger-processing kiosks, fuel and ground-support systems, environmental controls and operational databases. Aircraft-related domains

may include aircraft control, airline information services, passenger information and entertainment, maintenance interfaces and wireless sensor networks. Air-traffic-management and communication, navigation and surveillance infrastructure add further dependencies. The defining security property is that a cyber decision may influence availability, integrity or timing of a physical or operational function.

The security objectives consequently extend beyond the conventional confidentiality-integrity-availability triad. Deterministic timing, graceful degradation, recoverability, traceability, maintenance integrity, human oversight and safety assurance are also relevant. NIST SP 800-82 Rev. 3 emphasises that OT controls must respect performance, reliability and safety requirements [4]. EASA Part-IS similarly requires the management of information-security risks that may affect aviation safety [7]. These sources support a safety-security co-engineering approach in which cyber risk treatment is evaluated against operational consequences.

2.2 Zero Trust principles

The architecture adopts seven operational principles derived from NIST guidance [1,2]: protect resources rather than network segments; authenticate and authorise every session; minimise privilege; use dynamic policy; continuously evaluate identity and device evidence; assume compromise; and collect telemetry for policy improvement. These principles are implemented through a policy engine, policy administrator and policy-enforcement points. They are supplemented by microsegmentation, workload identity, device-health assessment, threat intelligence and continuous monitoring.

Zero Trust is not synonymous with distrust of all information. Policy decisions inevitably depend on evidence whose confidence varies. The useful interpretation is therefore evidence-calibrated trust: no entity receives enduring access merely because it is inside a boundary, but access may be granted when independent signals provide sufficient confidence for the requested action. Fernandez and Brazhuk caution that Zero Trust can be reduced to slogans if trust assumptions, policy correctness and system dependencies remain implicit [21]. SA-AZTA responds by exposing the trust variables, weights, thresholds, missing-evidence treatment and safety constraints.

2.3 Safety-security co-engineering and resilience

Safety engineering seeks to prevent unacceptable harm arising from failures and hazardous interactions, whereas cybersecurity addresses intentional and accidental compromise of information and control. In aviation CPS the two are coupled. A cyberattack can generate a safety hazard, but an aggressive cyber response can also create one. SA-AZTA therefore treats response selection as a constrained optimisation problem: reduce cyber risk while maintaining an acceptable safety state. This is consistent with ICAO's objective that global civil aviation remain resilient, safe and trusted while continuing to innovate [5,6] and with airworthiness-security processes represented by RTCA DO-326A and EUROCAE ED-202B [8,9].

3. Structured Evidence Review

A targeted structured review was undertaken to establish the architecture and identify the research gap. This was not represented as a PRISMA systematic review because the search was purposefully focused on authoritative standards and directly relevant peer-reviewed studies. Sources were selected from NIST, CISA, ICAO, EASA, RTCA, EUROCAE and MITRE portals, together with peer-reviewed work located through IEEE Xplore, ScienceDirect, SpringerLink and major scholarly indexes. Search concepts combined "zero trust", "aviation", "cyber-physical systems", "industrial IoT", "airborne wireless sensor networks", "adaptive access control", "microsegmentation", "anomaly detection" and "safety". Priority was given to 2020-2026 work, with older seminal sources retained where necessary.

Inclusion required a direct contribution to Zero Trust architecture, aviation cybersecurity, CPS security, trust evaluation, assurance or safety-

related security governance. Purely descriptive commentary without a traceable method was excluded from the analytical comparison. The resulting corpus combined standards and peer-reviewed studies. Table 1 summarises the most directly relevant research.

Table 1. Critical comparison of directly relevant studies.

Study	Domain and method	Principal contribution	Gap retained
Sampigethaya and Poovendran [11]	Aviation CPS conceptual foundation	Defines aviation as a coupled cyber-physical ecosystem	Predates operational Zero Trust architecture
Elmarady and Rahouma [12]	Civil-aviation risk assessment	Structures aviation cyber-risk identification and assessment	Does not implement continuous adaptive access control
Ukwandu et al. [13]	Aviation cybersecurity review	Synthesises threat actors, attack surfaces and future trends	Primarily descriptive; no evaluated architecture
Ain et al. [14]	Aviation CPS anomaly-detection review	Maps ML opportunities and limitations for aviation anomaly detection	Detection is not integrated with policy enforcement or safety response
Feng and Hu [16]	Industrial CPS simulation	Cross-layer trust scoring and policy optimisation	Not aviation-specific; limited treatment of safety-preserving response
Hasan et al. [17]	CPS assurance patterns and UAV case	Provides Zero Trust design and assurance patterns	Focuses on design assurance rather than operational adaptive evaluation
Zanasi et al. [18]	Industrial IoT prototype	Flexible SDN microsegmentation and resilient policy distribution	No aviation mission context or safety-impact term
Wang et al. [19]	Airborne wireless sensor networks	Distributed Zero Trust and dynamic identity authentication	Narrow airborne sensor scope; no airport/enterprise/OT integration

4. Research Gap, Aim and Hypotheses

4.1 Research gap and novelty

The evidence reveals four connected gaps. First, an empirical gap exists because many aviation cybersecurity publications catalogue threats without evaluating a complete adaptive access architecture. Second, a methodological gap exists because Zero Trust evaluations commonly report detection or authentication performance without modelling how policy responses interact with safety-critical availability. Third, a contextual gap exists between enterprise or industrial Zero Trust and the multi-

organisational aviation ecosystem, where aircraft, airports, airlines, air-traffic services, contractors and cloud systems have different trust and certification boundaries. Fourth, an integration gap remains between anomaly detection, access control, microsegmentation, threat intelligence and resilience-aware response.

The novelty of SA-AZTA lies in combining these elements within one aviation-specific decision cycle. Unlike static ABAC, the model updates evidence continuously. Unlike conventional Zero Trust, it incorporates asset criticality and potential safety impact directly into risk evaluation. Unlike purely automated containment, the safety gate selects bounded responses, including step-up authentication, privilege reduction, zone isolation, supervised degradation and human escalation. The architecture is evaluated across heterogeneous airport assets rather than a single protocol or sensor network.

4.2 Aim, objectives, research questions and hypotheses

The research aim is to design and evaluate a safety-aware adaptive Zero Trust architecture that reduces unauthorised access and lateral movement in aviation cyber-physical systems while maintaining acceptable latency and availability.

Table 2. Alignment of objectives, research questions and hypotheses.

Objective	Research question	Hypothesis
O1: Design an aviation-specific Zero Trust architecture	RQ1: Which evidence and control components are required for cross-domain aviation CPS access decisions?	Design question; no statistical hypothesis
O2: Develop a dynamic trust and safety-risk model	RQ2: Can context-rich trust evaluation distinguish malicious from legitimate requests more effectively than baseline controls?	H1: SA-AZTA produces a higher F1-score than each baseline
O3: Evaluate containment and operational performance	RQ3: Does the architecture reduce lateral movement while preserving availability and latency?	H2: SA-AZTA lowers false positives and compromised-node propagation; H3: mean decision latency remains below 20 ms
O4: Determine component contribution and robustness	RQ4: Which architecture components contribute most to detection and containment?	H4: Removing behaviour analytics or microsegmentation materially reduces performance

5. Proposed Safety-Aware Adaptive Zero Trust Architecture

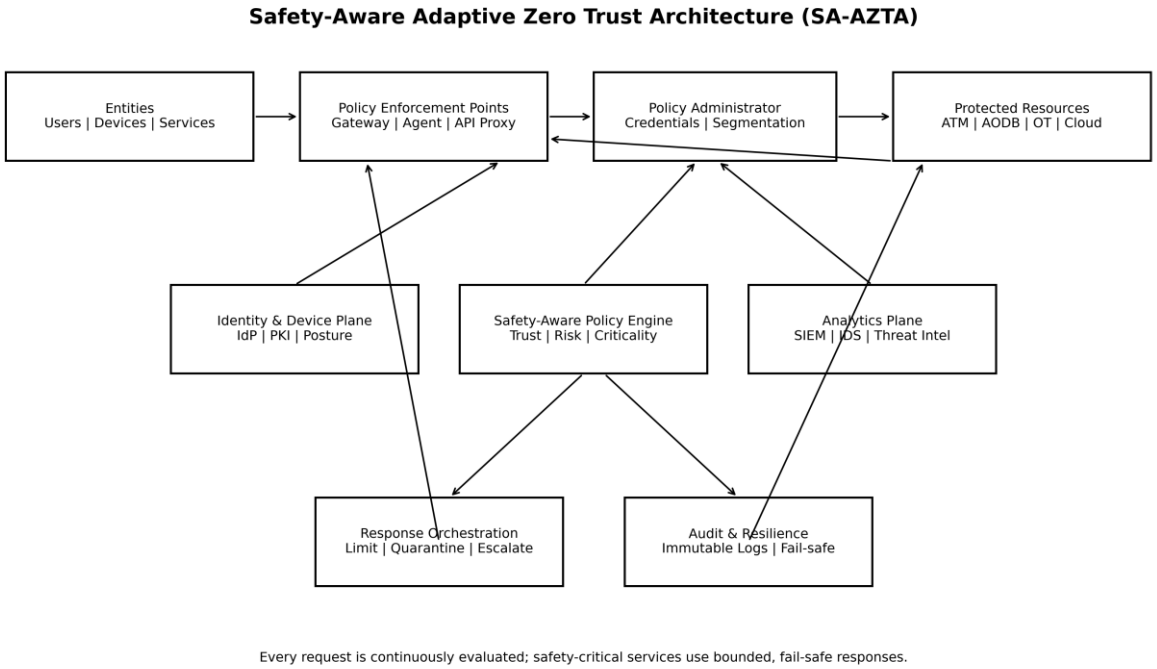

Figure 1. Logical architecture of the proposed Safety-Aware Adaptive Zero Trust Architecture (SA-AZTA).

Figure 1 places enforcement close to the requesting entity and protected resource while retaining logically coordinated policy. Human users, devices, services and workloads present identity and session evidence to policy-enforcement points (PEPs), such as identity-aware gateways, host agents, API proxies and industrial gateways. PEPs do not decide access independently; they submit the request and evidence to the policy decision plane.

The policy engine combines the trust-evaluation engine, risk-scoring module and safety-aware rule base. The policy administrator translates decisions into short-lived credentials, network segmentation rules, API permissions and endpoint actions. Identity and device services provide multifactor authentication, public-key infrastructure, workload identity, certificate status, software inventory, vulnerability state and device-health evidence. The analytics plane contributes behavioural baselines, anomaly scores, threat intelligence and security-event correlation. Response orchestration coordinates isolation, privilege reduction, alerting and recovery. Immutable or tamper-evident logging supports investigation and compliance.

Table 3. Architecture components and aviation-specific functions.

Component	Inputs	Function	Aviation-specific consideration
Policy-enforcement point	Request, identity token, device and session metadata	Intercepts and enforces every resource request	Deployment must not create a single unsafe failure mode
Identity and PKI services	Human, device, service and workload identity	Strong authentication and short-lived	Supports contractors, maintenance devices and

		credentials	machine identities
Trust-evaluation engine	Identity, posture, behaviour, context and history	Computes composite confidence	Accounts for intermittent links and missing evidence
Risk and safety engine	Threat, privilege, vulnerability, criticality and safety impact	Estimates cyber risk and response constraint	High criticality reduces tolerance but limits abrupt disconnection
Policy administrator	Decision and required controls	Issues credentials and segmentation rules	Uses fail-safe, redundant distribution
Analytics plane	SIEM, IDS, anomaly and threat intelligence	Detects abnormal and known malicious patterns	Correlates IT and OT evidence
Microsegmentation fabric	Identity-aware network and workload policies	Limits lateral movement and blast radius	Separates passenger, airline, airport and safety-related zones
Response orchestrator	Decision, asset state and runbook	Restricts, isolates, escalates or restores	Human confirmation for potentially safety-significant actions

5.1 Policy-decision workflow

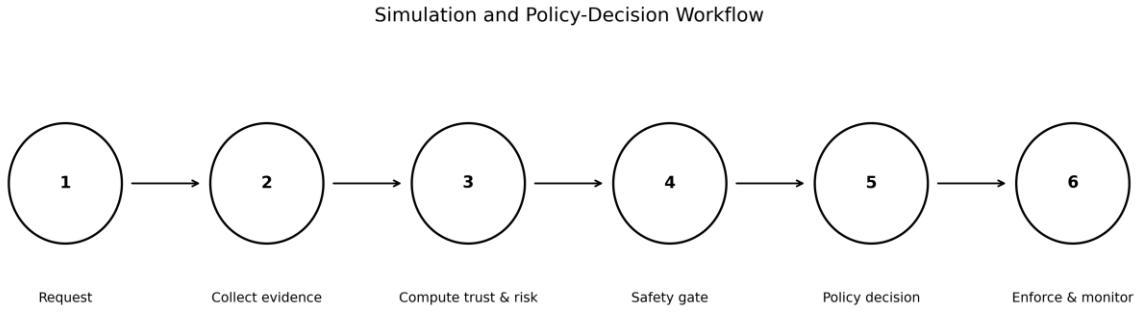

Figure 2. Continuous evidence collection, safety gating and policy-enforcement workflow.

The workflow begins with a request to access a resource or execute an operation. The PEP gathers evidence from identity, device, network and analytics services. The trust engine computes a positive-confidence score, while the risk engine calculates threat pressure and prospective safety impact. The safety gate then determines which response classes are permissible for the target resource. The policy engine selects the least-privileged action and the PEP enforces it. Telemetry from the resulting session updates subsequent decisions, enabling continuous verification rather than one-time login approval.

5.2 Trust and risk model

All evidence variables are normalised to [0,1]. Higher values are favourable for identity confidence (I), device integrity (D), behavioural conformity (B), location consistency (L), mission-context consistency (M), network health (N) and authentication strength (A). Higher values are adverse for threat-intelligence risk (X), privilege risk (P), vulnerability risk (V), anomaly score (Y), asset criticality (C) and safety impact (S). Missing device evidence is conservatively imputed to 0.50 and increases uncertainty; production deployment should use source-specific confidence and freshness terms.

The positive trust score is:

T = 0.18I + 0.16D + 0.17B + 0.08L + 0.10M + 0.08N + 0.08A + 0.05(1-P) + 0.05(1-V) + 0.05(1-X).

The adverse risk pressure is:

R = 0.43Y + 0.18X + 0.14P + 0.10V + 0.08S + 0.07C.

The probability that a request is malicious is calculated by a logistic decision function:

p_attack = sigma[2.05(1-T) + 1.55R + 0.55(Y x S) - 1.43],

where sigma(z)=1/(1+e^-z). The interaction term Y x S increases scrutiny when abnormal behaviour affects a safety-significant asset. In the main simulation, p_attack >= 0.508 triggered an attack classification. Operational access actions use a separate trust policy: T >= 0.72 permits least-privileged access; 0.58 <= T < 0.72 requires step-up authentication or reduced privilege; 0.42 <= T < 0.58 restricts the session to an isolated or read-only mode; and T < 0.42 denies or quarantines the entity. The safety gate can replace immediate disconnection with supervised degradation, redundant-channel switching or human escalation when a hard denial could create a greater hazard.

Table 4. Trust variables, simulation weights and rationale.

Variable	Weight	Interpretation	Rationale
Identity confidence (I)	0.18	Strength and consistency of asserted identity	Fundamental, but insufficient against valid credential compromise
Device integrity (D)	0.16	Patch, attestation, configuration and malware posture	Important for maintenance devices, sensors and gateways
Behavioural conformity (B)	0.17	Deviation from historical and peer behaviour	Detects insiders and compromised valid accounts
Location consistency (L)	0.08	Expected physical/network location	Useful contextual signal but vulnerable to spoofing
Mission context (M)	0.10	Consistency with role, shift, flight or maintenance task	Reduces legitimate-but-inappropriate access
Network health (N)	0.08	Connection integrity and segment condition	Supports detection of propagation and abnormal channels
Authentication strength (A)	0.08	MFA, certificate and credential assurance	Raises identity assurance
Anomaly score (Y)	0.43 in R	Behavioural and traffic anomaly severity	Largest risk term because attacks often retain valid identity

Threat risk (X)	0.18 in R	Known malicious indicator or campaign association	Provides external and cross-system context
Privilege and vulnerability risk (P,V)	0.24 in R	Requested power and exposed weakness	Limits high-impact access from weak devices
Safety impact and criticality (S,C)	0.15 in R	Potential physical/operational consequence	Makes policy aviation-specific and constrains response

Algorithm 1. Safety-aware adaptive access decision

Input: request q, entity e, resource r, evidence vector E, safety state H
Output: action in {ALLOW, STEP_UP, RESTRICT, QUARANTINE, DENY, ESCALATE}
1: validate evidence signatures, freshness and provenance
2: impute missing evidence conservatively and mark uncertainty
3: compute positive trust T and adverse risk pressure R
4: compute p_attack = sigmoid(2.05(1-T)+1.55R+0.55(Y*S)-1.43)
5: determine permitted response set G from resource criticality and safety state H
6: if p_attack >= theta_attack or T < theta_deny then propose DENY/QUARANTINE
7: else if T < theta_restrict then propose RESTRICT
8: else if T < theta_stepup then propose STEP_UP
9: else propose least-privileged ALLOW
10: map proposed action to the safest action in G; require human approval where specified
11: issue short-lived authorisation, enforce segmentation and begin continuous monitoring
12: update evidence and re-evaluate on material change or policy interval
Complexity: O(k) for k evidence features, excluding external analytics and policy lookup.

6. Threat Model

Threat modelling combined STRIDE-style security properties with ATT&CK for ICS behaviours [10]. The model assumes adversaries may obtain credentials, exploit remote services, compromise supplier software, alter device state, inject false data or use a trusted host for lateral movement. It does not include detailed exploit instructions or target a real airport. The protected assets are an airport operational database (AODB), baggage systems, passenger processing, maintenance applications, an air-traffic-management interface, surveillance and IoT devices, ground-support systems and cloud APIs.

Table 5. Threat scenarios, consequences and SA-AZTA controls.

Scenario	Primary weakness	Potential consequence	SA-AZTA treatment	Residual risk
Compromised credentials	Valid account and authentication artefacts	Unauthorised data or operational access	Behaviour, mission context, privilege limits and step-up authentication	Novel behaviour may initially resemble legitimate work
Malicious insider	Legitimate identity and device	Misuse of authorised privilege or data	Continuous behaviour scoring, just-in-time privilege and immutable audit	Collusion and low-and-slow activity
Lateral movement	Flat or overly permissive network trust	Propagation from IT to OT or sensitive service	Identity-aware microsegmentation and per-resource policy	Misconfigured policy or unmanaged legacy path
Device compromise	Weak posture, malware or vulnerable firmware	Manipulated sensing, commands or credential theft	Posture attestation, anomaly detection, restricted zone and quarantine	Incomplete attestation on legacy equipment
False-data injection	Weak message integrity or compromised source	Erroneous operational picture or unsafe decision support	Source identity, behaviour correlation, mission plausibility and safety gate	Sophisticated coordinated manipulation
Ransomware	Endpoint compromise and privilege escalation	Loss of availability and operational disruption	Segmentation, least privilege, rapid containment and recovery runbooks	Dependence on shared services and backups
Supply-chain compromise	Trusted update or contractor channel	Cross-domain entry with apparently valid software	Workload identity, software provenance, staged rollout and behavioural monitoring	Unknown upstream compromise

7. Research Methodology

7.1 Case-study boundary

The case study represents a synthetic major international airport ecosystem. It includes eight asset classes: AODB, baggage handling, passenger processing, maintenance, an ATM interface, IoT surveillance, ground-support equipment and cloud APIs. The topology includes users, contractor devices, service accounts, sensors, gateways, enterprise networks, OT zones and cloud services. No confidential airline, airport, aircraft or government information was used. The configuration is an analytical abstraction intended to test security mechanisms under controlled conditions.

Normal requests represent routine passenger, operations, maintenance, monitoring and application interactions. Malicious events are distributed across credential compromise,

malicious insiders, lateral movement, device compromise, false-data injection, ransomware and supply-chain compromise. Each event contains identity, posture, behaviour, location, mission, network, authentication, threat, privilege, vulnerability, anomaly, criticality and safety variables. Two per cent of device-posture observations are missing to test conservative evidence handling.

7.2 Synthetic data generation

Table 6. Synthetic dataset characteristics.

Event class	Main-run observations	Share
Normal	40,006	80.01%
Credential Compromise	2,002	4.00%
Lateral Movement	1,772	3.54%
Device Compromise	1,533	3.07%
False Data Injection	1,274	2.55%
Malicious Insider	1,242	2.48%
Supply Chain	1,161	2.32%
Ransomware	1,010	2.02%

Feature distributions were generated with bounded beta distributions because they allow realistic concentration near healthy or unhealthy states while remaining within [0,1]. Attack-specific transformations preserve valid identity evidence in credential and insider scenarios while degrading behaviour, privilege, device or network signals. This prevents the experiment from becoming a trivial classification task based on one feature. Random noise, missing posture evidence and overlapping distributions introduce uncertainty. The main random seed was 20260716; the Monte Carlo seeds ranged from 20260 to 20289.

7.3 Baselines and experimental design

Table 7. Compared security models.

Model	Decision evidence	Adaptive?	Segmentation/containment capability
Perimeter	Network health, location, authentication and limited threat/anomaly evidence	No	Low; assumes trusted internal network

Model	Decision evidence	Adaptive?	Segmentation/containment capability
RBAC	Identity, role, authentication and privilege	No	Role-scoped but limited session context
Static ABAC	Identity, device, location, authentication and privilege attributes	No	Moderate; fixed attribute policies
Non-adaptive Zero Trust	Identity, device, authentication, network and vulnerability evidence	Limited	Resource-oriented policy without behavioural adaptation
Proposed SA-AZTA	Full trust, anomaly, threat, criticality and safety evidence	Yes	Identity-aware microsegmentation and bounded response

The main run used 50,000 events for detailed confusion matrices and curves. Thirty independent Monte Carlo runs used 20,000 events per run. Each model was evaluated on the same generated events within a run. Decision thresholds were selected from ROC operating points using an objective that favoured F1-score and recall while penalising false positives; the proposed model was additionally constrained to a false-positive rate below 2.5% during calibration. Security metrics were accuracy, precision, recall, F1-score, false-positive rate, false-negative rate, ROC-AUC, precision-recall AUC and containment rate. Operational metrics were policy-decision latency and service availability.

Paired Wilcoxon signed-rank tests compared the proposed model with each baseline across the 30 seeds because the sample of run-level metrics was moderate and normality was not assumed. A two-sided alpha of 0.05 was used. Effect sizes were calculated as the mean paired difference divided by its standard deviation; these simulation effect sizes are descriptive and should not be interpreted as estimates from an operational population.

7.4 Reproducibility and implementation

The simulation was implemented in Python using NumPy, pandas, scikit-learn, SciPy and Matplotlib. The supplied code generates the dataset, calibrates thresholds, computes all metrics, runs the Monte Carlo experiment, performs paired tests and exports every figure. Metadata records the random seeds, event counts, thresholds and software versions. MATLAB/Simulink can reproduce the logic through a request-event generator, Stateflow policy states and SimEvents queues; Appendix B provides implementation-oriented pseudocode. Python was used because it was available in the execution environment, and no claim is made that MATLAB itself was run.

8. Results

8.1 Comparative detection performance

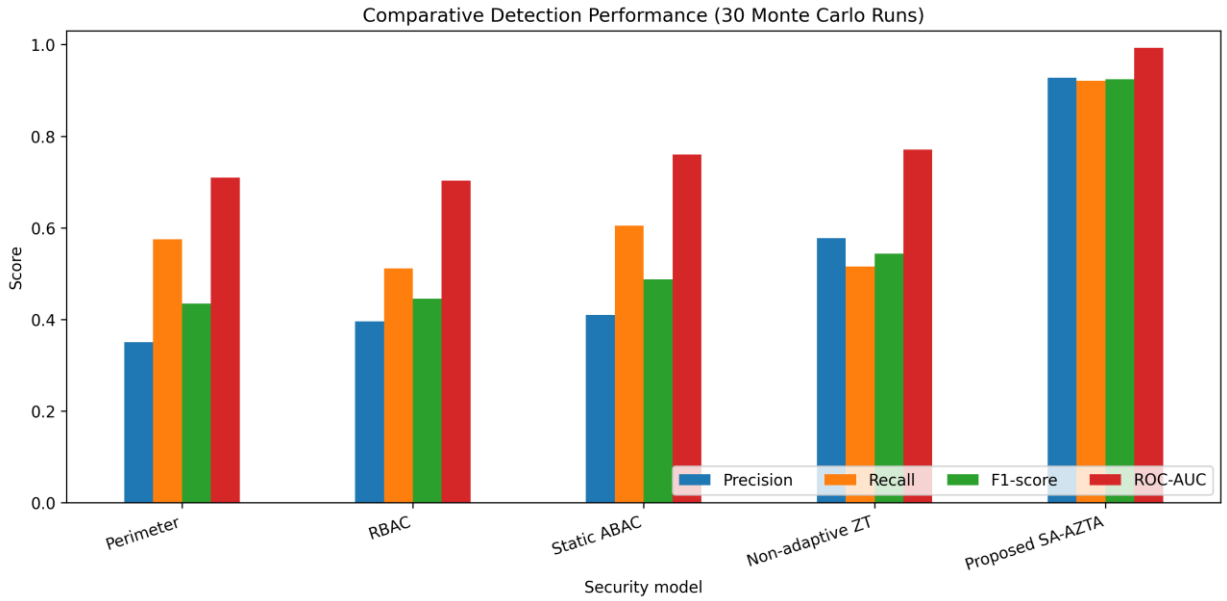

Figure 3. Mean precision, recall, F1-score and ROC-AUC over 30 Monte Carlo runs.

Table 8. Mean detection results over 30 Monte Carlo runs.

Model	Accuracy	Precision	Recall	F1	FPR	ROC-AUC	PR-AUC
Perimeter	70.11%	35.08%	57.49%	43.47%	26.74%	0.7092	0.4328
RBAC	74.54%	39.60%	51.13%	44.52%	19.62%	0.7031	0.4603
Static ABAC	74.58%	40.98%	60.45%	48.73%	21.90%	0.7599	0.4897
Non-adaptive ZT	82.71%	57.72%	51.58%	54.39%	9.52%	0.7710	0.6110
Proposed SA-AZTA	96.99%	92.78%	92.10%	92.43%	1.79%	0.9925	0.9776

SA-AZTA achieved the strongest overall performance: mean accuracy 96.99%, precision 92.78%, recall 92.10% and F1-score 92.43%. The closest baseline, non-adaptive Zero Trust, produced an F1-score of 54.39%. This difference reflects the proposed model's use of behaviour, threat intelligence and mission/safety context, particularly for attacks that retain valid credentials or originate from trusted segments.

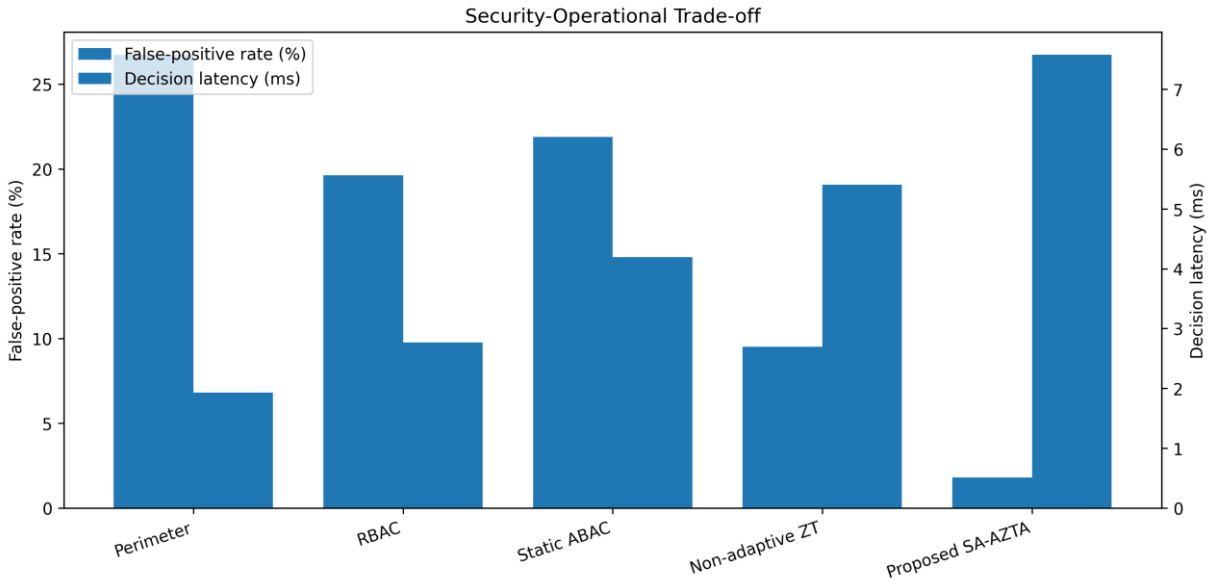

Figure 4. False-positive rate and policy-decision latency trade-off.

The improved security performance imposed a measurable but bounded latency cost. Mean decision latency increased from 5.40 ms for non-adaptive Zero Trust to 7.58 ms for SA-AZTA. The proposed false-positive rate was 1.79%, substantially below the 9.52% recorded for the closest baseline. The latency remains below the 20 ms experimental operational budget; this budget is a simulation requirement and not a universal certification threshold.

8.2 Main-run classification behavior

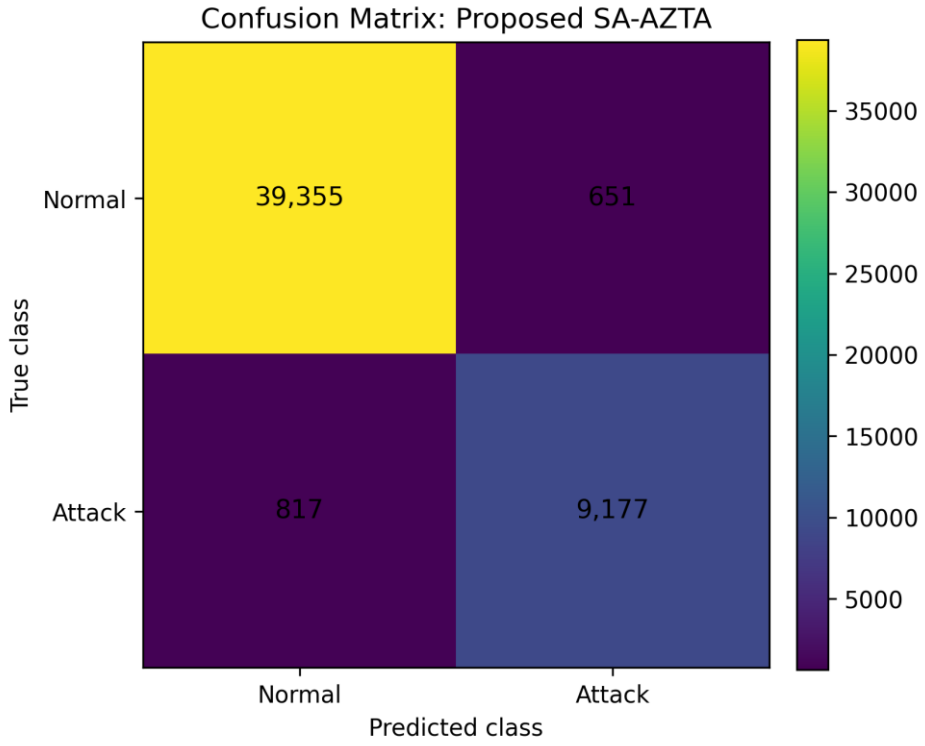

Figure 5. Confusion matrix for SA-AZTA in the 50,000-event main run.

In the main run, the proposed model correctly classified 39,355 normal events and 9,177 attacks, with 651 false positives and 817 false negatives. The resulting performance was 97.06% accuracy, 93.34% precision, 91.83% recall and 92.59% F1-score. False negatives were concentrated in low-and-slow insider or credential-compromise events where valid identity, authentication and device signals remained strong.

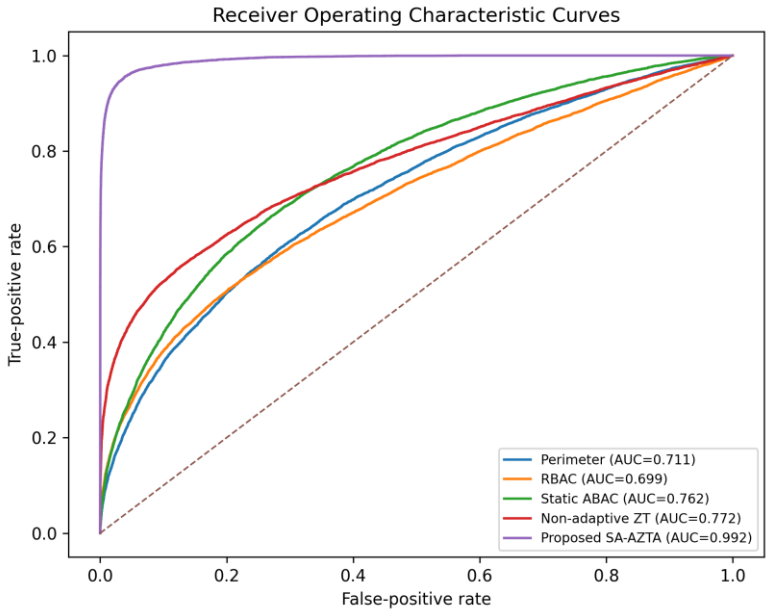

Figure 6. ROC curves for the five compared security models.

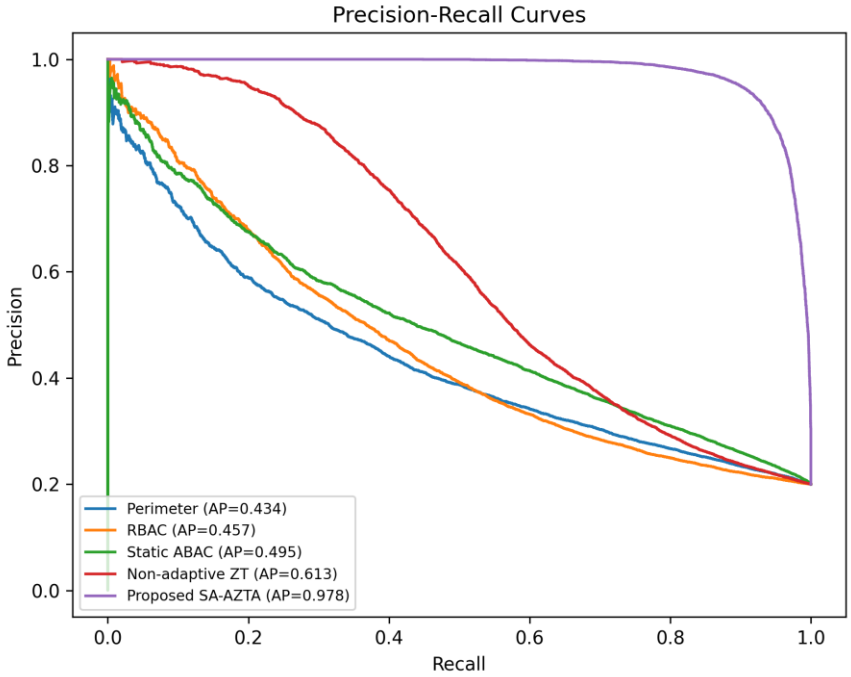

Figure 7. Precision-recall curves under the 20% attack prevalence used in the main simulation.

The ROC and precision-recall curves confirm that the improvement was not an artefact of one threshold. SA-AZTA achieved main-run ROC-AUC of 0.9923 and average precision of 0.9777. The large separation from the baselines follows from richer evidence and the interaction between anomaly and safety impact. Operational validation must examine whether comparable separation persists when real evidence sources are noisier, correlated or adversarially manipulated.

8.3 Dynamic trust, containment and blast radius

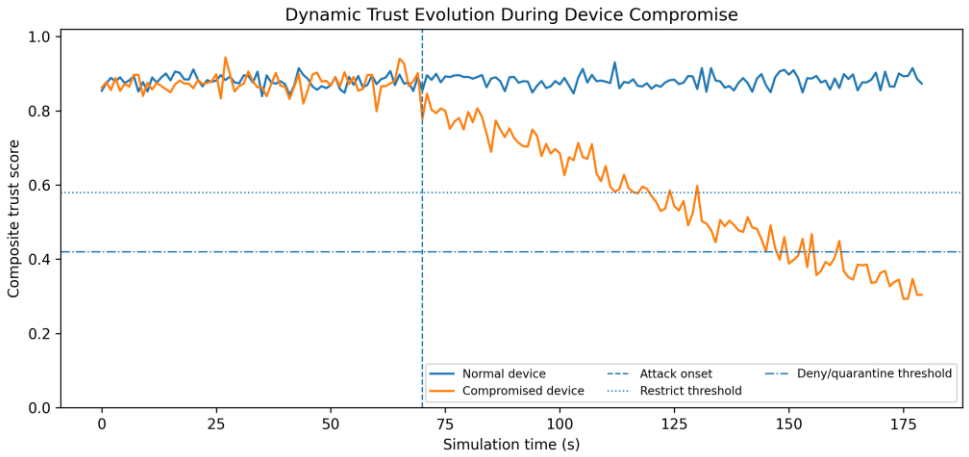

Figure 8. Illustrative dynamic trust evolution following device compromise at 70 s.

Figure 8 illustrates the intended continuous-verification behaviour. A normal device remains near its established trust range, whereas a compromised device decays after attack onset as behaviour and posture evidence deteriorate. The entity crosses the restrict threshold before reaching the deny/quarantine threshold, providing a bounded interval for step-up verification, privilege reduction and safe transition.

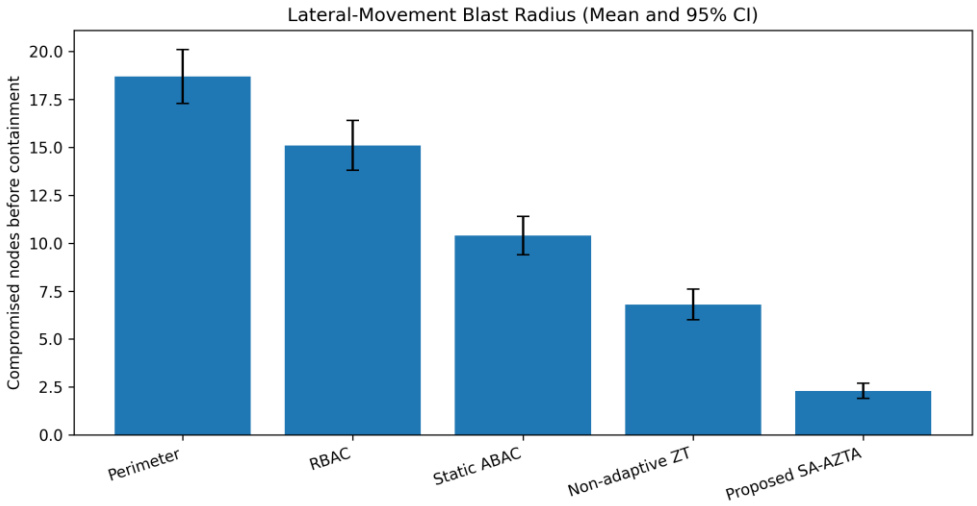

Figure 9. Simulated lateral-movement blast radius before containment.

The proposed architecture reduced the mean blast radius to 2.3 compromised nodes, compared with 6.8 under non-adaptive Zero Trust and 18.7 under perimeter security. Mean containment across Monte Carlo runs was 93.09%, while mean availability remained 99.40%. These results support the use of identity-aware microsegmentation as a resilience control rather than merely a network-management feature.

8.4 Statistical evaluation

Table 9. Paired Wilcoxon comparisons over 30 runs.

Baseline	Baseline F1	SA-AZTA F1	Two-sided p	F1 paired effect dz	FPR paired effect dz
Perimeter	0.4347	0.9243	<1.9 x 10^-9	83.82	-7.64
RBAC	0.4452	0.9243	<1.9 x 10^-9	78.80	-6.37
Static ABAC	0.4873	0.9243	<1.9 x 10^-9	73.77	-6.93
Non-adaptive ZT	0.5439	0.9243	<1.9 x 10^-9	48.21	-5.17

All paired F1, recall and false-positive comparisons rejected the null hypothesis of equal run-level distributions at $p < 1.9 \times 10^{-9}$. The very large simulation effect sizes arise because the same generative rules were used consistently across seeds; they should not be reported as evidence of equivalent effect magnitude in real airports. H1 and H2 were supported within the synthetic environment. H3 was also supported because the highest simulated mean latency was 14.82 ms at 20,000 requests per second, below the stated 20 ms budget.

8.5 Ablation and sensitivity analysis

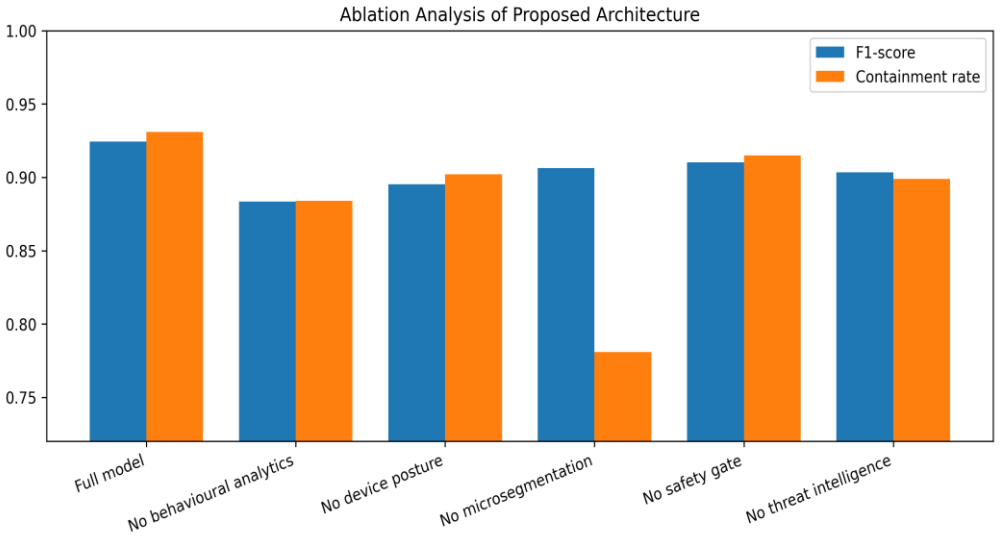

Figure 10. Ablation analysis of major SA-AZTA components.

Table 10. Component-ablation results.

Configuration	F1-score	Containment rate
Full model	0.9243	0.931
No behavioural analytics	0.8833	0.884
No device posture	0.8953	0.902
No microsegmentation	0.9063	0.781
No safety gate	0.9103	0.915
No threat intelligence	0.9033	0.899

Removing behavioural analytics reduced F1-score by 4.10 percentage points, the largest detection loss. Removing device posture reduced F1 by 2.90 points. Microsegmentation had a smaller direct effect on F1 but reduced containment from 93.1% to 78.1%, demonstrating that classification and resilience are distinct. Removing the safety gate improved neither detection nor containment in the synthetic model; its contribution is principally assurance that response does not create unsafe disruption.

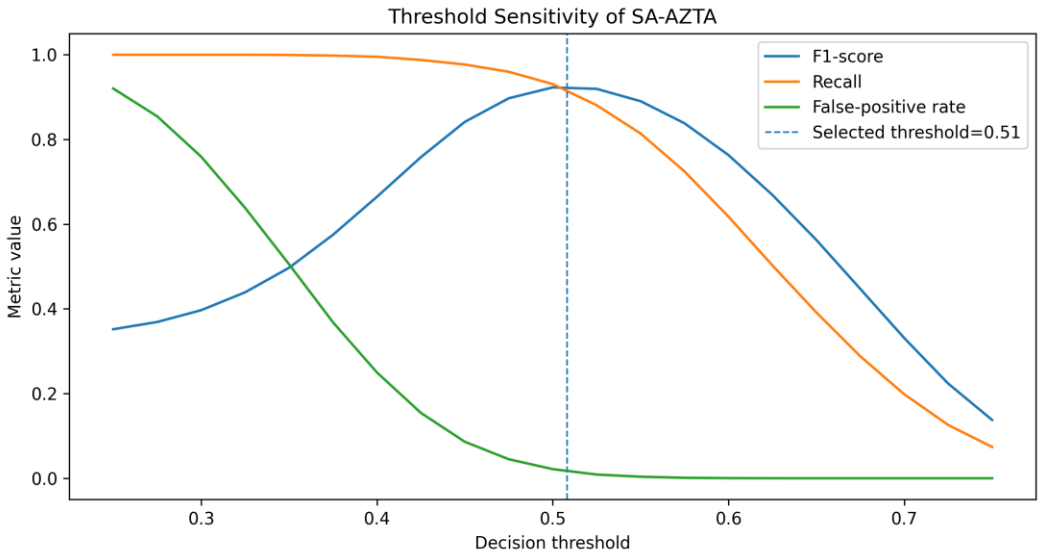

Figure 11. Sensitivity of F1-score, recall and false-positive rate to the attack-probability threshold.

The sensitivity curve shows the expected trade-off. Low thresholds maximise recall but generate unacceptable false positives; high thresholds suppress false positives while missing attacks. The selected threshold of approximately 0.51 lies near the maximum F1 region under the conservative false-positive constraint. In production, thresholds should vary by resource criticality, mission phase, uncertainty and available fallback controls rather than remain globally fixed.

8.6 Scalability

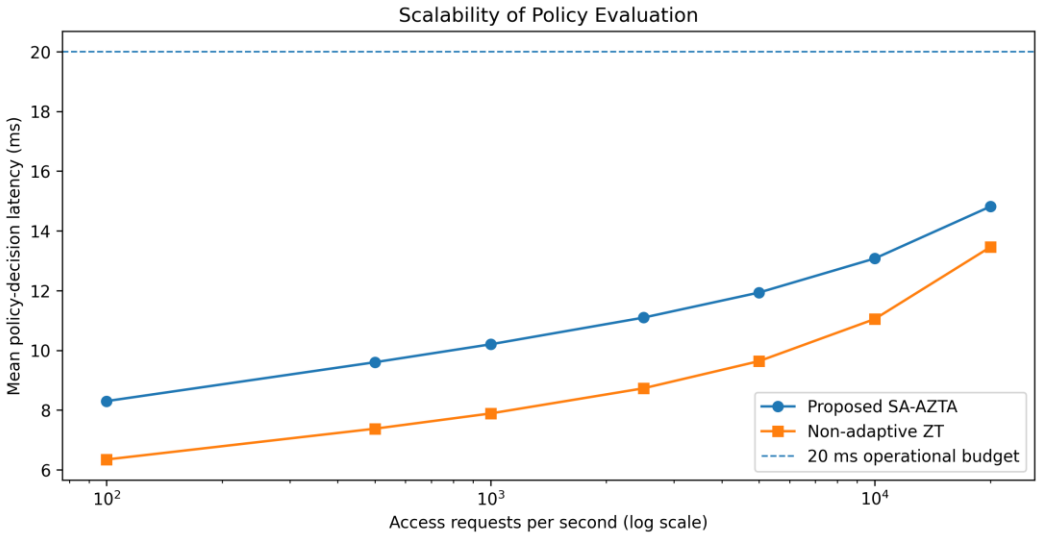

Figure 12. Mean policy-decision latency as simulated request load increases.

Table 11. Scalability results.

Requested events/s	SA-AZTA latency (ms)	Non-adaptive ZT latency (ms)	SA-AZTA processed events/s
100	8.30	6.34	100
500	9.60	7.38	500
1,000	10.20	7.89	1,000
2,500	11.10	8.73	2,500
5,000	11.94	9.64	5,000
10,000	13.08	11.05	10,000
20,000	14.82	13.46	16,833

SA-AZTA latency increased from 8.30 ms at 100 requests/s to 14.82 ms at 20,000 requests/s. The simulated processing capacity began to saturate at approximately 16,833 events/s under the highest offered load, indicating that a central policy engine would become a bottleneck. The operational architecture should therefore use horizontally scaled regional policy engines, local PEP caching for low-risk short-lived decisions, replicated policy stores and deterministic fallback rules. Safety-related fallback must be explicitly designed rather than relying on an uncontrolled fail-open or fail-closed default.

9. Discussion

9.1 Answers to the research questions

RQ1 asked which evidence and controls are required. The architecture indicates that identity and device posture are necessary but not sufficient. Aviation-specific policy also requires behavioural conformity, mission context, asset criticality, safety impact, network condition and threat intelligence. Enforcement requires workload identity, microsegmentation, short-lived authorisation, logging and safety-aware response orchestration.

RQ2 asked whether context-rich evaluation outperforms baseline controls. Within the synthetic benchmark, SA-AZTA produced substantially higher F1, ROC-AUC and precision-recall performance. The main mechanism was improved discrimination of attacks that retained valid identity or location evidence. This finding is consistent with the broader Zero Trust literature's emphasis on continuous verification [20,23] and with CPS-specific work showing the importance of cross-layer evidence [16].

RQ3 concerned containment, latency and availability. Microsegmentation reduced the lateral-movement blast radius, while the safety gate allowed bounded response. The additional latency was small relative to the experimental budget, but it remains operationally important. Authentication and authorisation latency can accumulate with network delay, certificate validation, analytics and logging. Real deployment therefore requires edge enforcement, precomputed policy, asynchronous telemetry and fail-safe redundancy.

RQ4 asked which components matter most. Behavioural analytics contributed most to F1-score because compromised credentials and insiders often preserve static attributes. Microsegmentation contributed most to containment. Device posture and threat intelligence provided intermediate value. The safety gate's benefit cannot be captured fully by classification metrics; it is a risk-control mechanism that must be evaluated through safety cases and hazardous-scenario testing.

9.2 Comparison with existing studies

SA-AZTA extends prior work in several ways. Feng and Hu model cross-layer penetration in industrial CPS [16], whereas the present architecture adds aviation mission and safety terms and tests a broader set of airport assets. Hasan et al. provide reusable Zero Trust assurance patterns for CPS and demonstrate them on a UAV application [17]; SA-AZTA complements design assurance with operational trust scoring, attack simulation and component ablation. Zanasi et al. demonstrate resilient microsegmentation for heterogeneous industrial IoT [18]; the present results similarly identify segmentation as central to containment, while adding identity, behavioural and safety-aware policy. Wang et al. address distributed Zero Trust and dynamic identity authentication for airborne wireless sensor networks [19]; SA-AZTA expands the scope to an airport system-of-systems and treats human, service, cloud and legacy OT identities together.

The proposed architecture should not be interpreted as replacing aviation airworthiness standards or information-security management. Instead, it provides a technical control architecture that can supply evidence to those processes. RTCA/EUROCAE airworthiness-security guidance addresses intentional unauthorised electronic interaction and certification activities [8,9], while EASA Part-IS addresses organisational management of information-security risks with potential safety impact [7]. SA-AZTA supports these objectives through explicit trust boundaries, auditable decisions and safety-constrained response, but formal compliance requires organisation-specific assurance.

9.3 Security, safety, privacy and ethical implications

Continuous verification creates privacy and governance risks. Behavioural monitoring can reveal work patterns, location, relationships and performance indicators. Aviation organisations should apply purpose limitation, data minimisation, role separation, retention limits and transparent workforce governance. Behavioural models must be tested for bias across roles and shifts. Security analysts should receive interpretable evidence rather than unexplained scores, especially when a decision affects employment, contractor access or operational responsibility.

Automated response is a dual-use capability. Isolation can limit attack propagation, but poorly designed automation can disrupt legitimate operations. Safety-significant decisions should use pre-approved runbooks, redundant communication paths, reversible actions and human confirmation where time permits. The architecture deliberately avoids publishing real

airport configurations, exploit sequences or vulnerable endpoints. Synthetic scenarios are sufficient for methodological evaluation without increasing operational risk.

10. Standards and Governance Mapping

Table 12. Mapping of SA-AZTA capabilities to selected guidance.

Guidance	Relevant expectation	SA-AZTA capability	Evidence produced
NIST SP 800-207 [1]	Explicit verification, least privilege, resource protection	PEPs, dynamic policy engine, short-lived authorisation	Decision logs, policy and session evidence
NIST SP 800-207A [2]	Granular application/workload access in distributed environments	API proxies, workload identity and multi-location policy	Application policy and workload credentials
CISA ZTMM 2.0 [3]	Identity, devices, networks, applications/workloads and data pillars	Integrated evidence and phased control planes	Maturity and coverage metrics
NIST SP 800-82r3 [4]	OT security respecting performance, reliability and safety	Legacy gateways, segmentation and bounded response	Latency, availability and fallback tests
ICAO strategy/action plan [5,6]	Resilience, collaboration, governance and capacity	Cross-domain telemetry, response and audit	Incident, resilience and stakeholder reports
EASA Part-IS [7]	Manage information-security risks with potential safety impact	Safety-impact term and controlled response	Risk assessments and safety-security decision records
RTCA DO-326A / EUROCAE ED-202B [8,9]	Airworthiness-security process and intentional unauthorised interaction	Traceable security requirements and enforcement controls	Assurance artefacts and test results
MITRE ATT&CK for ICS [10]	Behaviour-based adversary modelling	Threat scenarios and detection coverage	Technique-to-control mapping

11. Contributions to Knowledge

11.1 Theoretical contribution

The study extends Zero Trust reasoning from resource-centric verification to safety-constrained cyber-physical policy. It treats trust as evidence-calibrated confidence and separates attack probability from permissible response. The safety gate formalises the proposition that the most cyber-secure action may not be the safest immediate action in a CPS.

11.2 Methodological contribution

The reproducible benchmark combines detailed event generation, identical-baseline comparison, Monte Carlo uncertainty, paired non-parametric testing, threshold sensitivity, component

ablation, blast-radius analysis and scalability. The method distinguishes detection quality, containment and operational cost rather than collapsing them into one accuracy measure.

11.3 Technical and practical contribution

The technical contribution is a layered architecture integrating identity, posture, behavioural analytics, threat intelligence, microsegmentation, safety-aware policy and bounded response. Practically, the framework gives airports, airlines, service providers and regulators a migration path: inventory resources and identities; deploy PEPs at high-value interfaces; establish workload identity and device posture; segment cross-domain pathways; add behavioural analytics; and introduce safety-reviewed automated response incrementally.

12. Limitations and Future Research

The principal limitation is synthetic evidence. The generative distributions and attack transformations were designed to be plausible, not to replicate a specific airport. Results may overestimate separability because real telemetry contains correlated errors, unknown attack behaviour, inconsistent clocks, legacy blind spots and adversarial manipulation. The 20% attack prevalence is useful for comparative evaluation but much higher than normal operational prevalence; precision would fall in a rarer-event environment unless thresholds and triage were adjusted.

The architecture also assumes reliable evidence provenance and policy distribution. Compromise of the identity provider, policy engine, telemetry pipeline or time source could undermine all decisions. The simulation models latency analytically rather than through a production network and does not include certification artefacts, human-in-the-loop timing, electromagnetic interference, intermittent aircraft links or hardware resource limits. The attack-propagation and availability models are simplified.

Future research should validate the framework with an aviation digital twin, hardware-in-the-loop testbed and representative airport/airline data under appropriate agreements. Formal methods should verify safety-gate invariants and policy conflicts. Further work should examine federated behavioural learning, privacy-preserving telemetry, adversarial robustness, post-quantum device identity, cross-airport trust federation, human factors, economic feasibility and integration with security operations. NIST's digital-twin trust guidance provides a useful basis for this transition [24].

13. Conclusion

This study developed and evaluated a Safety-Aware Adaptive Zero Trust Architecture for aviation cyber-physical systems. The architecture removes implicit network trust, continuously evaluates human and machine identities, incorporates behaviour and mission context, limits lateral movement through microsegmentation and constrains automated response according to asset criticality and potential safety impact. In a reproducible synthetic airport case study, SA-AZTA materially outperformed perimeter, RBAC, static ABAC and non-adaptive Zero Trust baselines. Across 30 runs it achieved a mean F1-score of 92.43%, a false-positive rate of 1.79%, containment of 93.09% and availability of 99.40%, while maintaining mean decision latency of 7.58 ms.

The central implication is that Zero Trust for aviation should not be implemented as a generic enterprise access-control product. It must be engineered as a safety-security system with explicit trust evidence, bounded response, resilient policy distribution, auditable decisions and sector-specific assurance. The findings provide a strong proof of concept, but they remain synthetic. Publication and operational use should therefore present the framework as a reproducible research prototype requiring regulated validation rather than a certified aviation solution.

Declarations

Funding: [Insert funding information or state "This research received no external funding."]

Conflicts of interest: The author(s) declare no

known competing financial interests or personal relationships that could have appeared to influence the work.

Ethics statement: The study used synthetic data and publicly available secondary literature. It involved no human participants, personal data, live airport systems or confidential aviation infrastructure.

Data availability: The synthetic event dataset, Monte Carlo outputs, statistical tests, figures and metadata are included in the accompanying reproducibility package.

Code availability: Executable Python source code is included. MATLAB/Simulink-compatible pseudocode is provided, but MATLAB was not executed in this study.

Author contributions (CRediT): [Insert Conceptualisation; Methodology; Software; Validation; Formal analysis; Investigation; Writing - original draft; Writing - review and editing; Visualisation; Supervision.]

References

[1] Rose S, Borchert O, Mitchell S, Connelly S. Zero Trust Architecture. NIST Special Publication 800-207. National Institute of Standards and Technology; 2020. doi:10.6028/NIST.SP.800-207.

[2] Chandramouli R, Butcher Z. A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Location Environments. NIST Special Publication 800-207A. National Institute of Standards and Technology; 2023. doi:10.6028/NIST.SP.800-207A.

[3] Cybersecurity and Infrastructure Security Agency. Zero Trust Maturity Model, Version 2.0. CISA; 2023.

[4] Stouffer K, Pease M, Tang C, et al. Guide to Operational Technology (OT) Security. NIST Special Publication 800-82 Revision 3. National Institute of Standards and Technology; 2023. doi:10.6028/NIST.SP.800-82r3.

[5] International Civil Aviation Organization. Aviation Cybersecurity Strategy. ICAO; current edition.

[6] International Civil Aviation Organization. Cybersecurity Action Plan. 2nd ed. ICAO; 2022.

[7] European Union Aviation Safety Agency. Easy Access Rules for Information Security (Regulations (EU) 2023/203 and 2022/1645). Revision December 2025. EASA; 2025.

[8] RTCA. DO-326A: Airworthiness Security Process Specification. RTCA; 2014.

[9] EUROCAE. ED-202B: Airworthiness Security Process Specification. EUROCAE; 2024.

[10] MITRE. ATT&CK for Industrial Control Systems: Matrix, tactics, techniques and assets. MITRE; accessed July 2026.

[11] Sampigethaya K, Poovendran R. Aviation cyber-physical systems: foundations for future aircraft and air transport. Proceedings of the IEEE. 2013;101(8):1834-1855. doi:10.1109/JPROC.2012.2235131.

[12] Elmarady AA, Rahouma K. Studying cybersecurity in civil aviation, including developing and applying aviation cybersecurity risk assessment. IEEE Access. 2021;9:143997-144016. doi:10.1109/ACCESS.2021.3121230.

[13] Ukwandu E, Ben-Farah MA, Hindy H, Bures M, Atkinson R, Tachtatzis C, Andonovic I, Bellekens X. Cyber-security challenges in aviation industry: a review of current and future trends. Information. 2022;13(3):146. doi:10.3390/info13030146.

[14] Ain QU, Jilani AAA, Butt NA, Rehman SU, Alhulayyil HA. Anomaly detection for aviation cyber-physical system: opportunities and challenges. IEEE Access. 2024;12:175905-175925. doi:10.1109/ACCESS.2024.3495519.

[15] Harkat H, Camarinha-Matos LM, Goes J, Ahmed HFT. Cyber-physical systems security: a systematic review. Computers & Industrial Engineering. 2024;188:109891. doi:10.1016/j.cie.2024.109891.

[16] Feng X, Hu S. Cyber-physical zero trust architecture for industrial cyber-physical systems. IEEE Transactions on Industrial Cyber-Physical Systems. 2023;1:394-405. doi:10.1109/TICPS.2023.3333850.

[17] Hasan S, Amundson I, Hardin D. Zero-trust design and assurance patterns for cyber-physical

systems. Journal of Systems Architecture. 2024;155:103261. doi:10.1016/j.sysarc.2024.103261.

[18] Zanasi C, Russo S, Colajanni M. Flexible zero trust architecture for the cybersecurity of industrial IoT infrastructures. Ad Hoc Networks. 2024;156:103414. doi:10.1016/j.adhoc.2024.103414.

[19] Wang K, Hong Y, Li Y, et al. A distributed zero-trust scheme for airborne wireless sensor networks using dynamic identity authentication. Scientific Reports. 2025;15:8036. doi:10.1038/s41598-025-91957-2.

[20] Kang H, Liu G, Wang Q, Meng L, Liu J. Theory and application of zero trust security: a brief survey. Entropy. 2023;25(12):1595. doi:10.3390/e25121595.

[21] Fernandez EB, Brazhuk A. A critical analysis of Zero Trust Architecture (ZTA). Computer Standards & Interfaces. 2024;89:103832. doi:10.1016/j.csi.2024.103832.

[22] Lee S, Huh J-H, Woo H. Security system design and verification for Zero Trust Architecture. Electronics. 2025;14(4):643. doi:10.3390/electronics14040643.

[23] Mushtaq S, Mohsin M, Mushtaq MM. A systematic literature review on the implementation and challenges of Zero Trust Architecture across domains. Sensors. 2025;25(19):6118. doi:10.3390/s25196118.

[24] Voas J, et al. Security and Trust Considerations for Digital Twin Technology. NIST Interagency Report 8356. National Institute of Standards and Technology; 2025. doi:10.6028/NIST.IR.8356.

[25] Phiayura P, Teerakanok S. A comprehensive framework for migrating to Zero Trust Architecture. IEEE Access. 2023;11:19487-19511. doi:10.1109/ACCESS.2023.3248622.

[26] Cybersecurity and Infrastructure Security Agency. Microsegmentation in Zero Trust, Part One. CISA; 2025.

Appendix A. Reproducibility Parameters

Parameter	Value
Main simulation seed	20260716
Main detailed events	50,000
Monte Carlo runs	30
Events per Monte Carlo run	20,000
Attack prevalence	Approximately 20%
Missing device-posture evidence	2%
Main SA-AZTA attack threshold	0.5082
Trust action thresholds	Allow >=0.72; step-up 0.58-0.72; restrict 0.42-0.58; deny/quarantine <0.42
Operational latency budget	20 ms (experimental assumption, not certification threshold)
Software	Python 3.x; NumPy; pandas; scikit-learn; SciPy; Matplotlib

Appendix B. MATLAB/Simulink-Compatible Implementation Outline

Algorithm B1. MATLAB-style simulation outline
rng(20260716);
N = 50000;
[events, labels] = generateAviationEvents(N);
for model = [PERIMETER, RBAC, ABAC, NONADAPTIVE_ZT, SA_AZTA]
scores = evaluateModel(model, events);
theta = calibrateThreshold(scores, labels);
predictions = scores >= theta;
metrics(model) = computeMetrics(predictions, labels);
end
for seed = 20260:20289
rng(seed); runResults(seed) = runMonteCarlo(20000);
end
performPairedWilcoxon(runResults);
exportFiguresAndTables();

A Simulink implementation can represent entities and resources as subsystems, the evidence vector as a bus, the policy engine as a MATLAB Function block, and access states as a Stateflow chart. SimEvents can model request queues and processing latency. The safety gate should be a separate Stateflow layer so that cyber decisions cannot bypass operational constraints. Hardware-in-the-loop validation should replace analytical latency functions before any operational claim.

Appendix C. Reference and Quality-Control Statement

Reference consistency: Every numbered in-text citation corresponds to an entry in the reference list. Bibliographic details and DOI strings for the principal peer-reviewed sources were checked against publisher, DOI, institutional or standards-organisation records. Online standards and regulatory materials should be rechecked for revised editions immediately before submission.

Quality-control status: The manuscript distinguishes synthetic data, simulated outputs, public secondary sources and unexecuted MATLAB guidance. It does not claim access to a real airport or certified aviation system. Objectives, questions, hypotheses, methods, results and conclusions were cross-checked for alignment. Figures and tables are original outputs of the accompanying simulation. Author information, funding, CRediT roles and the exact Elsevier template remain placeholders requiring completion.