Power System Protection and Control
ISSN:1674-3415

ZERO-TRUST INTELLIGENT BEAM MANAGEMENT FOR B5G/6G MMWAVE

Rusul H Hussain[1], Reza Mohammadi[2]

[1]Middle Technical University
Email: rusulhamdi@mtu.edu.iq
[2]Department of computer engineering, faculty of engineering, bu-ali sina university.
Email: R.mohammadi@basu.ac.ir

Received:24 September 2025 **Revised**:22 October 2025 **Accepted**: 12 November 2025

***ABSTRACT*:**

Present a secure design beam-management framework for B5G/6G mmWave that couples convolutional neural networks (CNNs) for real-time beam selection with elliptic-curve cryptography (ECC) and AES-GCM, governed by a Zero-Trust control loop. The end-to-end pipeline is implemented in ns-3 with realistic channel and mobility models (static, pedestrian, vehicular) and evaluated using beam-alignment time, throughput, packet delivery ratio (PDR), and tail-sensitive latency metrics. Relative to exhaustive search, the CNN reduces alignment time by 60–70%. Enabling ECC adds 1.0–1.4 ms, while Zero-Trust enforcement yields 2.2–2.6 ms end-to-end overhead, predominantly confined to the upper tail (p90/p95/p99, ES_{95}); the distribution's center (median/MAD) remains essentially unchanged. In vehicular tests, throughput progresses from 680 Mb/s (exhaustive) to 850 Mb/s (CNN), then 810 Mb/s (CNN+ECC) and 795–800 Mb/s with Zero-Trust; PDR stays within 1% of the CNN+ECC baseline. These results indicate that learning-based, crypto-hardened beam management can sustain low alignment delay and continuous trust guarantees under mobility, with operationally bounded overheads in realistic ns-3 settings.

***Keywords**: mmWave beam management; CNN beam prediction; ECC; AES-GCM; Zero-Trust Architecture; ns-3.*

INTRODUCTION

Millimeter-wave (mmWave) communication is a cornerstone of Beyond-5G (B5G) and 6G systems thanks to wide, contiguous spectrum and the potential for multi-Gb/s user-plane rates. Yet the same strong directionality that enables high spectral efficiency also makes mmWave links vulnerable to blockage, severe path loss, and mobility-induced misalignment. Realizing mmWave gains under motion therefore hinges on **fast and reliable beam management**, whereby user equipment (UE) and base stations continuously align narrow beams without undermining end-to-end latency or throughput [1–4]. In parallel, mission-critical services (immersive XR/digital twins, V2X/CAVs, time-critical industrial control, tele-medicine, large-scale CPS) demand **continuous trust guarantees—**confidentiality, integrity, and access control—across heterogeneous radio, edge, and cloud substrates operating in adversarial environments [5–7].

Classical exhaustive or hierarchical beam search (EBS) attains accurate alignment but its scanning overhead scales with codebook size and re-alignment frequency, which becomes prohibitive under mobility and dense urban dynamics. Learning-based approaches mitigate this cost by predicting near-optimal beams from context features—coarse localization proxies, radio fingerprints (e.g., RSRP/RSSI), short beam histories, or side-sensor cues—thereby shrinking alignment time without exhaustive sweeps [8–12]. However, most empirical studies evaluate beam intelligence **in isolation**: (i) they seldom quantify how lightweight cryptography and continuous authorization affect link-layer dynamics under mobility; (ii) they rarely report **tail-latency** beyond averages; and (iii) they often use disjoint platforms that complicate apples-to-apples comparisons [13–16]. For time-sensitive pipelines, such omissions matter: users experience upper-tail delays (p90/p95/p99) and risk-sensitive measures such as ES_{95} more acutely than changes in the mean.

This paper treats performance and security as a **single co-designed problem**. We develop a secure-by-design beam-management framework that couples real-time convolutional-neural-network (CNN) beam selection with elliptic-curve cryptography (ECC) and AES-GCM, governed by a Zero-Trust control loop. The pipeline is realized end-to-end in **ns-3** with realistic channel and mobility models (static, pedestrian, vehicular), integrates **ONNX Runtime** for in-simulation inference, and employs a lightweight crypto stack (ECDH→HKDF→AES-GCM).

DOI: 10.46121/pspc.53.4.10

Policy Enforcement Points (PEPs) at the UE/gNB validate short-lived tokens on the data path, while an edge Policy Decision Point (PDP) evaluates access decisions based on Identity-Provider (IdP) credentials and context—adding minimal per-packet checks while preserving tight user-plane budgets [17–20].

We evaluate four baselines under matched traffic and mobility: **EBS, CNN-only, CNN+ECC**, and **CNN+ECC+ZTA.** Metrics include beam-alignment time, end-to-end latency, throughput, and packet delivery ratio (PDR), complemented with **tail-aware** statistics (p90/p95/p99 and ES_{95}) and robust dispersion (median/MAD). Empirically, learning-based variants reduce alignment time by **60–70%** versus EBS. Enabling ECC adds **1.0–1.4 ms** end-to-end overhead; adding Zero-Trust yields **2.2–2.6 ms** total—effects that predominantly elevate the **upper tail** (p95/p99/ ES_{95}) while leaving the distribution's center (median/MAD) essentially unchanged. Throughput and PDR remain within operational targets across scenarios; for vehicular tests, throughput tracks **680 → 850 → 810 → 795–800 Mb/s** across EBS → CNN → CNN+ECC → CNN+ECC+ZTA, with PDR within 1% of the CNN+ECC baseline [17–20]. These findings indicate that **crypto-hardened, learning-based** beam management sustains low alignment delay and continuous trust guarantees under mobility with **operationally bounded** costs.

Contributions.
(i) An ns-3 realization of a unified **CNN–ECC–ZTA** pipeline integrating radio-layer intelligence with cryptographic protection and continuous authorization.
(ii) A **tail-aware measurement framework** that reports median/MAD alongside p90/p95/p99 and ES_{95} , revealing whether security shifts the center or inflates only the upper tail.
(iii) A **component-level overhead decomposition** (PEP, PDP, cryptography) showing bounded costs (1.0–1.4 ms for ECC; 2.2–2.6 ms with ZTA) and stable throughput/PDR across static, pedestrian, and vehicular regimes.
(iv) **Reproducible artifacts**—configuration files, ONNX models, and CSV logs—facilitating independent verification and extension.

ORGANIZATION. SECTION 2 TECHNICAL BACKGROUND ON MMWAVE BEAM MANAGEMENT, CNN-BASED PREDICTION, ECC/AES-GCM, AND ZERO-TRUST. SECTION 3 REVIEWS RELATED WORK AND POSITIONS OUR STUDY. SECTION 4 METHODOLOGICAL FRAMEWORK. SECTION 5 SIMULATION SETUP AND RESULTS. SECTION 6 COMPARATIVE ANALYSIS WITH LITERATURE AND RESULTS. SECTION 7 REFERENCES.

TECHNICAL BACKGROUND

Millimeter-wave (mmWave, 30–300 GHz) offers abundant bandwidth for multi-Gb/s rates, yet severe path loss, blockage, and narrow beams make links brittle under mobility; systems must therefore re-align beams frequently and with very low latency while preserving end-to-end security [1–4]. Classical beam management—exhaustive, hierarchical, or codebook-based sweeps—achieves high alignment accuracy but often incurs tens of milliseconds of delay and non-trivial control overhead, which clashes with B5G/6G latency targets in XR, V2X, and industrial control [3], [10], [17], [18]. Recent simulator/testbed evidence in realistic channels and stacks further highlights this agility-vs-overhead tension and motivates learning-based and cross-layer remedies [31–34].

2.1 Learning-based beam selection

Deep learning—especially CNNs—reduces search latency by learning a direct mapping from context (e.g., location/trajectory features, RSSI/RSRP vectors, short beam histories) to a beam index (or pair), bypassing slow sweeps [1–4], [11], [12]. Across vehicular, pedestrian, and UAV-like dynamics, CNNs report 60–70% shorter alignment time versus hierarchical or exhaustive search while sustaining throughput, with practical on-device inference via ONNX Runtime/TensorRT (sub-ms to few ms) [1–4], [11], [12]. Hardware-constrained and real-world validations—position-aided prediction, edge/federated training, and lightweight CNNs—show that these gains persist with modest compute budgets [3], [4], [17], [18]. Recent works also emphasize metric design (e.g., time-to-first-beam, re-alignment latency under multimodal inputs), underscoring how evaluation choices affect conclusions for dynamic scenarios [33], [34].

2.2 Lightweight cryptography for B5G/mmWave

Directionality does not eliminate risk: side lobes, reflections, and transient misalignment can leak information; spoofing/replay and on-path manipulation remain viable without strong, continuous authentication. Lightweight public-key cryptography—ECC for key agreement (ECDH) and authentication (ECDSA)—paired with authenticated encryption (AES-GCM) provides confidentiality, integrity, and authenticity with small keys and

DOI: 10.46121/pspc.53.4.10

sub-few-millisecond overheads suitable for edge devices and mobile UEs [5–8], [13], [16]. Comparative studies consistently find ECC-AES pipelines outperforming RSA-based designs on latency/energy at equivalent security strength; mobility-aware handover/authentication schemes show further reductions when ECC is integrated with access procedures [5–8], [16]. Emerging designs refine dynamic credentials and per-session keying to bound attack windows in dense deployments [14], [15].

2.3 Zero-Trust Architecture (ZTA) and continuous verification

Static trust after a single handshake leaves systems exposed to insider threats, token theft, and lateral movement. ZTA closes this gap with continuous identity/posture validation, least-privilege micro-segmentation, and in-path PEPs backed by a low-latency PDP and an IdP that rotates short-lived tokens [19–21], [24–29]. For B5G/6G, per-packet token/nonce-window checks at PEPs can add 0.1 ms, while PDP decisions typically remain sub-millisecond, keeping end-to-end overheads within tight latency budgets when combined with lightweight crypto [19], [20], [24], [25], [29]. The loop complements CNN-based beam agility: CNNs restore alignment under dynamics/jamming; ZTA constrains blast radius and suppresses spoofing/replay with minimal extra cost.

2.4 Execution engines and evaluation realism

Recent ns-3/5G-LENA advances provide more faithful PHY/MAC, antenna/beamforming, and evaluation tooling (Flow Monitor, scalable channel models, PMI/rank selection, multi-panel antennas), enabling controlled apples-to-apples comparisons of **CNN-only** vs **CNN+ECC** vs **CNN+ECC+ZTA** under matched traffic/mobility [31–34]. ONNX Runtime integrates with C++ event loops for real-time inference, and mbedTLS (or similar) provides ECDH→HKDF→AES-GCM on the data path—so both intelligence and security can be measured in one simulator with reproducible hooks for beam-alignment time, throughput, PDR, and per-packet enforcement/decision latencies [5], [16], [31–34].

LITERATURE REVIEW

This section is organized as follows. Section 3.1 reviews CNN-based beamforming techniques; Section 3.2 surveys ECC and other lightweight cryptographic primitives for mmWave; and Section 3.3 synthesizes joint secure-intelligent designs (CNN+ECC+ZTA), thereby framing the research gap and positioning our contribution.

3.1 CNN-Based Beamforming Approaches

Convolutional neural networks (CNNs) have become a practical mechanism for learning beam-selection policies in dynamic mmWave settings, where channel non-stationarity and mobility render sweep-based alignment costly. A typical pipeline maps environment- and user-aware features—such as coarse location, partial CSI/quality vectors, inertial cues, and short beam histories—directly to a codebook index, thereby reducing (re)alignment delay. Zhang et al. [1] report 70% reduction versus hierarchical search in urban vehicular simulations, while Wang et al. [2] achieve sub-4-ms on-device inference via ONNX Runtime in a 5G testbed. Edge/federated training and lightweight models further lower compute budgets with minimal accuracy loss [3], [4]. Beyond central-tendency improvements (median/MAD), recent works acknowledge **tail behavior**—high quantiles (p90/p95/p99) and, in some cases, ES_{95} to capture worst-case typical latency under mobility and blockage [1–4], [11], [12].

3.2 ECC and Lightweight Cryptography for mmWave

ECC offers a favorable cost–security profile for mobility-constrained links: ephemeral ECDH enables rapid key agreement and ECDSA supports mutual authentication at smaller key sizes than RSA, reducing handshake time and device power. In vehicular settings, Liu et al. [5] report 1.3 ms ECDH latency and 1 Gb/s AES throughput under an ECC–AES design; Singh and Rana [6] rank ECC best overall versus RSA and lattice-based schemes on mobile-edge nodes. Mobility-centric studies integrate ECDSA into handover (35% overhead reduction) [7] and distribute key management via SDN control [8]. Several works separate center from tail effects (p95/p99), supporting ECC-anchored AEAD (AES-GCM after ECDH) for tight timing [5–8], [13], [16].

3.3 ZTA and Joint Secure-Intelligent Designs (CNN+ECC+ZTA)

ZTA augments cryptography by enforcing continuous identity/posture validation and least-privilege segmentation. Concretely, PEPs on the UE/gNB data path validate short-lived tokens and nonce windows; an edge PDP returns allow/deny decisions in 0.2–0.6 ms; an IdP rotates credentials every 5–10 s. This introduces small but measurable overhead—per-packet at the PEP and end-to-end (PEP+PDP+IdP)—compatible with B5G/6G budgets when paired with ECC/AES-GCM [5–8], [13], [16], [19–21], [24–29]. Recent studies start to **co-design** beam intelligence with security: UAV and V2X prototypes couple CNN tracking with ECC handshakes and report

DOI: 10.46121/pspc.53.4.10

high PDR with sub-few-ms security cost [9], [10]. Evidence converges that CNN-assisted selection trims alignment time by 60–70% with sub-ms to few-ms inference [1–4], [11], [12], while ECC+AES outperforms RSA on latency/energy at matched strength [5–8], [13], [16].

Gap. Prior art often (i) omits ZTA's continuous-verification loop, (ii) does not disaggregate per-packet PEP cost from end-to-end overhead under realistic mobility/traffic, or (iii) evaluates intelligence and security on disjoint platforms, obscuring cross-layer trade-offs and tail-sensitive behavior (p90/p95/p99, ES_{95}). **Scaffold.** Advances in ns-3/5G-LENA enable unified, reproducible evaluation with richer PHY/MAC and instrumentation to compare CNN-only vs CNN+ECC vs CNN+ECC+ZTA under matched traffic/mobility, reporting median/MAD alongside p90/p95/p99 and ES_{95} [31–34].

Table 1. Consolidated comparative summary

Study	Beamforming approach	Platform	Security included?	Tail metrics reported?	Reproducibility
Zhang et al. 2021 [1]	CNN beam prediction (massive MIMO/mmWave)	Simulation	No	No	Not reported
Wang et al. 2022 [2]	Low-latency CNN for adaptive beamforming	Simulation / Prototype	No	Partial (basic percentiles)	Not reported
Kwon & Flanagan 2021 [3]	Deep CNN beam selection (5G)	Simulation	No	No	Not reported
Al Tamimi et al. 2023 [4]	Lightweight edge CNN for beam prediction	Edge / Simulation	Partial (security-aware, non-crypto)	No	Not reported
Zhong et al. 2024 [49]	Vision-based beam tracking (V2I)	Experiment	No	Partial	Dataset links
Marenco et al. 2024 [50]	ML-aided beam-pair selection and update time	Simulation / Experiment	No	Partial	Not reported
Vučković et al. 2024 [51]	Multimodal DL beam prediction (benchmarking)	Benchmark	No	Yes (p-quantiles focus)	Code (assets)
Liu et al. 2021 [5]	Security stack for mmWave (beamforming not reported)	Simulation	Yes (ECC + AES)	No	Not reported
Singh & Rana 2022 [6]	ECC key management in mmWave/IoT (beamforming not reported)	IoT / Simulation	Yes (ECC key management)	No	Not reported
Hu et al. 2022 [7]	Authentication for handover (beamforming not reported)	5G	Yes (ECC-based authentication)	No	Not reported
Mahmoudi et al. 2023 [8]	CNN-based selection with crypto control (UAV)	UAV / Simulation	Yes (conceptual CNN + ECC)	No	Not reported

DOI: 10.46121/pspc.53.4.10

Jang et al. 2024 [9]	DL-assisted V2X beamforming with secure handshake	V2X / Prototype	Yes (DL + ECC handshake)	Partial	Not reported
Lin et al. 2022 [10]	IRS-aided beam control (mmWave)	Simulation	No	No	Not reported
Cao et al. 2023 [11]	DL beam & power allocation (SR-guided)	Simulation / Preprint	No	Partial	Not reported
Bojović et al. 2024 [36]	Platform capabilities (beamforming not central)	ns-3 / 5G-LENA	—	—	Code / papers
ICNS3'25 [37], [38]	Platform improvements (beamforming not central)	ns-3 / 5G-LENA	—	—	Artifacts (ACM)
NIST SP 800-207, 2020 [28]	ZTA specification (architectural)	Standard	Yes (ZTA framework)	No	Available
Gambo & Almulhem 2025 [19]	ZTA systematic review (architectural)	SLR	Yes (ZTA concepts)	No	Available
Dhiman et al. 2024 [20]	ZTA + ML review (architectural)	Review	Yes (ZTA concepts)	No	Available
This work (2025)	CNN beam prediction + ECC/AES-GCM + ZTA (PEP/PDP/IdP)	ns-3 / 5G-LENA (static, pedestrian, vehicular)	Yes (ECC + AES-GCM + ZTA)	Yes (median/MAD + p90/p95/p99 + ES95)	Code/config/ONNX/CSV (via DOI)

METHODOLOGICAL FRAMEWORK

This section presents a secure-by-design mmWave stack realized end-to-end between the UE and the gNB. The architecture couples learning-based beam management with lightweight public-key/symmetric cryptography to deliver low-latency, cryptographically protected links.

4.1 System Overview

The system comprises: (i) a UE with a phased-array front-end and a lightweight runtime for inference/crypto; (ii) a gNB with beamforming control and an inline Policy Enforcement Point (PEP); and (iii) an edge/MEC node hosting the Identity Provider (IdP) and the Policy Decision Point (PDP). Inference executes at the UE or gNB (or a co-located edge accelerator), while policy/credential services reside at the MEC to keep the control loop sub-millisecond.

(1) Session bootstrapping (control plane).

- **Credentials.** UE and gNB hold ECC credentials (e.g., ECDSA P-256).
- **Handshake.** Mutual authentication followed by ECDH; HKDF-SHA-256 expands the shared secret into a 128-bit traffic key.
- **Tokens & trust.** The IdP issues short-lived tokens (TTL 5–10 s). On cache miss, PDP decisions return in 0.2–0.6 ms; PEP enforcement costs 0.1 ms/packet. Keys/tokens are cached and rotated periodically.
- **Termination.** The handshake terminates at the gNB (or UPF) and is mirrored at the UE.

(2) Secure data plane.

- **Ciphering.** AES-GCM (AEAD) protects user traffic end-to-end.

DOI: 10.46121/pspc.53.4.10

- **Integration options.** PDCP ciphering (RRC/UP), L4 protection via QUIC/TLS 1.3 (UE↔edge), or IPsec ESP (GCM) UE↔gNB/UPF.
- **Replay safety.** Strictly monotonic nonces and a receiver-side sequence window.
- **Measured cost.** Added one-way overhead: 1.0–1.4 ms for CNN+ECC; 2.2–2.6 ms for CNN+ECC+ZTA.

(3) Beam-management loop.

- **Inputs.** Per-coherence features (e.g., RSRP/RSRQ, CSI-RS/SSB summaries, Doppler/speed, short beam history).
- **Inference.** A compact CNN (ONNX Runtime, C++) predicts the next TX/RX codebook index and a hold time once per coherence interval (2–10 ms, hardware-dependent).
- **Actuation.** The gNB updates precoder weights; the UE steers the receive beam. Low confidence triggers a mini-sweep over a compact subset while traffic remains encrypted.
- **Control coupling.** Token refresh and policy re-checks align with beam updates; abnormal telemetry triggers throttling or quarantine via the PEP.

(4) Telemetry & reproducibility.

- **Collection.** End-to-end latency, throughput, PDR, and beam-alignment time captured via synchronized timestamps and simulator counters.
- **Artifacts.** We release model version & seeds, feature schema, ONNX file, and cipher/policy configurations (cipher-suite, key length, token TTL), plus CSV logs to enable reproduction on commodity edge hardware.
- **Budget mapping.** Targets consistent with evaluation envelopes: PEP 0.1 ms/packet; PDP 0.2–0.6 ms; CNN 2–10 ms; total Δ-security 2.2–2.6 ms vs. CNN-only.

4.2 Joint CNN+ECC (AES-GCM) Models

The detail dataset/labeling (4.2.1), model and cryptographic pipelines (4.2.2).

4.2.1 Dataset and Labeling (Beam Intelligence)

Per-TTI snapshots are collected under identical channel/mobility/traffic in three regimes—static, pedestrian, vehicular. Each regime includes ≥5 independent runs; warm-up is discarded and only steady-state windows are retained. Inputs include radio-quality summaries (RSRP/RSRQ, SINR, CSI-RS/SSB aggregates), short-horizon mobility/context (speed, heading, Doppler surrogates), a compact beam-index history, and a one-hot scenario flag. The target is the (TX or joint TX/RX) codebook index maximizing instantaneous SNR/RSRP with a short hold interval to prevent oscillation. Features are standardized with mild outlier clipping. Splits are by run (e.g., 70/15/15 train/validation/test) to prevent leakage. Metrics—top-k accuracy, alignment-time reduction, mis-alignment rate—are reported as mean ± 95% CI.

4.2.2 Model and Cryptographic Pipelines

Beam intelligence. A supervised CNN maps feature windows to a probability distribution over beam indices. The network uses three convolutional blocks (ReLU, interleaved pooling), two fully connected layers, and a softmax head. The trained model (PyTorch/TensorFlow) is exported to ONNX; ONNX Runtime executes inference once per coherence interval (2–10 ms), outputting the beam index, confidence, and hold time.

Cryptographic stack. ECC→AEAD pipeline: per-session ECDH; HKDF-SHA-256 derives a 128-bit AES-GCM key. Packets are protected inline; anti-replay via monotonic nonces and a receiver-side sequence window. With ZTA enabled, PEP validates tokens/nonces per packet (0.1 ms); PDP returns allow/deny on demand (0.2–0.6 ms). Total added overhead remains bounded: 1.0–1.4 ms (CNN+ECC) and 2.2–2.6 ms (CNN+ECC+ZTA), preserving 60–70% alignment reduction from the CNN.

4.3 Threats and Mitigation Strategies for B5G/mmWave

summarize major threats and how the CNN+ECC(+ZTA) stack mitigates them (numbers reflect our measurements).

1. **Eavesdropping.** Side-lobes/reflections/misalignment. → AES-GCM on all user traffic; short-lived keys/tokens. Overhead included in 1.0–1.4 ms (CNN+ECC).
2. **Spoofing & replay.** Fake or stale frames. → ECC mutual auth; nonce/sequence-window checks at PEP; short token TTLs; PDP 0.2–0.6 ms.
3. **Jamming/interference.** Disruption during training/updates. → CNN-assisted fast re-beamforming, confidence-based mini-sweeps, rapid switching/null-steering, adaptive MCS/power; ZTA telemetry → rate-limit/quarantine.

DOI: 10.46121/pspc.53.4.10

4. **Control-plane DoS/PDP overload.** → In-path PEP 0.1 ms with caching/tokenization; micro-segmentation; rate limits at PEP. Total with ZTA: 2.2–2.6 ms over CNN-only.
5. **Insider/lateral movement.** → ZTA continuous authN/authZ, least-privilege micro-segmentation, short token TTLs, decision logging.
6. **Key/device compromise.** → Hardware roots of trust, secure boot, remote attestation, frequent rotation.
7. **Beam poisoning/model misuse.** → Confidence thresholds + mini-sweeps; sanity checks vs. history/RSRP; optional sensor cross-checks; ZTA throttles anomalous control traffic.
8. **Handover/mobility gaps.** → Coherence-aligned inference, pre-auth with ECC, token refresh synchronized to beam updates; no plaintext buffering.

4.4 Metrics and Measurement Protocol

In this section, first define the metrics used (E2E latency, throughput, PDR, beam-alignment time, and security overhead including PEP/E2E), then briefly state how they are computed from synchronized packet and beam-control logs over steady-state windows, followed by a concise, uniform measurement protocol to ensure fair, reproducible comparisons across scenarios.

4.4.1 Metric Definitions

Tail-aware beam-alignment metrics:

Let T_{align} denote the per-trial beam-alignment latency (ms), with cumulative distribution function (CDF)
$F_T(t)=Pr\ [T_{align} \leq t]$
To expose rare-but-critical delays, the upper-tail quantiles of T_{align} p90, p95, and p99— via the α –quantile
$Q_\alpha(T_{align}) = inf\{t \in R : F_T(t) \geq \alpha\}$, $\alpha \in \{0.90, 0.95, 0.99\}$

the tail conditional expectation at the 95th percentile (ES_{95}) to quantify expected delay in the extreme tail:
$ES_{95} = E[T_{align} \mid T_{align} \geq Q_{0.95}(T_{align})]$

For robust central tendency and dispersion, we also report the median
$T' = median\ (T_{align})$
and the median absolute deviation (MAD)
$MAD = median\ (|T_{align} - T')$
all metrics are computed per scenario and model from empirical distributions over repeated runs (after warm-up removal) and summarized with 95% confidence intervals.

Security overhead (ms):

the latency cost of enabling the security stack relative to a CNN-only baseline. The **absolute** security overhead is:
$\boldsymbol{\Delta t}_{sec} = E2E\_latency\_secure - E2E\_latency\_CNN\_only$ **… 1**

Percent overhead:

$Percent\_OH = (\boldsymbol{\Delta t}_{sec} / E2E\_latency\_CNN\_only) * 100$ **… 2**

Where:
E2E_latency_secure is the mean end-to-end latency with the security stack enabled (CNN+ECC or CNN+ECC+ZTA), and E2E_latency_CNN-only is the mean end-to-end latency with only the CNN beam logic active (no cryptography, no ZTA).

two complementary scopes. **Per-packet enforcement overhead (PEP)** is the additional processing applied at the policy enforcement point on each packet (e.g., nonce/sequence-window and token checks; typically, 0.1 ms under ZTA). **Total end-to-end overhead** is the aggregate delay attributable to security across the full path (PEP checks plus any cached PDP decisions and key/token operations), via $\boldsymbol{\Delta t}_{sec}$ and Percent_OH.

Measurement protocol. For each scenario (static, pedestrian, vehicular), we run two conditions with identical radio, traffic, and mobility profiles: (i) CNN-only (baseline) and (ii) secure (CNN+ECC and/or CNN+ECC+ZTA). Timestamped packets (e.g., CBR UDP or an emulated application flow) are transmitted for a steady-state window; one-way delay is computed using synchronized clocks (or RTT if synchronization is unavailable). We average latency over the window and repeat for ≥5 runs, reporting mean ± standard deviation

DOI: 10.46121/pspc.53.4.10

(and 95% confidence intervals). The computed $\mathbf{\Delta t}_{sec}$and Percent_OH quantify the end-to-end cost that an application experiences, while the per-packet PEP timing isolates enforcement overhead at the data path.

4.4.2 Metric Formulation

Our metrics are derived from timestamped packet traces and beam-control logs collected over a steady-state analysis window W (warm-up discarded). For each packet p transmitted within W, we record the transmit time t_p^{tx} the receive time t_p^{rx}, and the payload size $\boldsymbol{S_P}$ (bits). Let P be all transmitted packets and $\boldsymbol{P_{OK}} \subseteq$P the subset delivered successfully. Unless noted otherwise, results are reported as mean ± standard deviation with 95% confidence intervals over at least five independent runs per scenario (static, pedestrian, vehicular).

(1) End-to-End Latency (one-way)

Per-packet delay and window average:

$$\ell p = t_p^{rx} - t_p^{tx} \ (ms) \qquad \ell' = 1/|\boldsymbol{P_{OK}}| \sum_{P\in \boldsymbol{P_{OK}}} \boldsymbol{\ell p} \qquad \ldots 3$$

Where informative (e.g., vehicular mobility), we also provide the empirical CDF of ℓp to expose tail behavior.

(2) Throughput

$$TP = \sum_{P\in \boldsymbol{P_{OK}}} \boldsymbol{S_P} \ /|w| \quad (bits/s;\ reported\ in\ Mb/s) \qquad \ldots 4$$

- *Packet Delivery Ratio (PDR)*

1) $$PDR(s,m) = \boldsymbol{P_{OK}}/|\boldsymbol{p}| \ \times 100\% \qquad \ldots 5$$

(4) Beam-Alignment Time

Let ε be the set of beam-update events in W. For event e, t_e^{start} marks the instant a new beam is requested (CNN decision or fallback mini-sweep), and t_e^{ready} marks the resumption of user traffic on the updated beam after a short stability hold:

$$T_e = t_e^{ready} - t_e^{start} \quad , \quad T' = \frac{1}{|\varepsilon|}\sum_{e\varepsilon\varepsilon} T_e \qquad \ldots 6$$

(5) Security Overhead (absolute and relative)

We quantify the latency cost of security relative to a CNN-only baseline:

$$\boldsymbol{\Delta t_{sec}} = \boldsymbol{\ell'}|_{secure} - \boldsymbol{\ell'}|_{CNN-only} \qquad \ldots 7$$

$$Percent\_OH = \boldsymbol{\Delta t_{sec}}/\boldsymbol{\ell'}|_{CNN-only} \ *100\% \qquad \ldots 8$$

"secure" denotes **CNN+ECC** (ECDH→KDF; AES-GCM on the data path) or **CNN+ECC+ZTA** (adding PEP/PDP/IdP). To isolate component contributions, we also scenario-wise medians:

$$\boldsymbol{\Delta t_{sec}} = \boldsymbol{Median(\ell)}|_{CNN+ECC} - \boldsymbol{Median(\ell)}|_{CNN-only} \qquad \ldots 9$$

$$\boldsymbol{\Delta t_{ZTA}} = \boldsymbol{Median(\ell)}|_{CNN+ECC+ZTA} - \boldsymbol{Median(\ell)}|_{CNN+ECC} \qquad \ldots 10$$

$$\boldsymbol{\Delta t_{sec}} = \boldsymbol{\Delta t_{ECC}} + \boldsymbol{\Delta t_{ZTA}} \qquad \ldots 11$$

We both (i) **per-packet PEP enforcement time** (nonce/sequence-window and token checks, measured in-path) and (ii) the **end-to-end** $\mathbf{\Delta t}_{sec}$ perceived by applications.

4.4.3 Scenarios and Traffic

We evaluate three scenarios—static, pedestrian, and vehicular—under identical traffic and mobility profiles per scenario. Traffic generation and channel/mobility models are held constant across models (CNN-only, CNN+ECC, CNN+ECC+ZTA) to enable apples-to-apples comparisons.

4.4.4Statistical Protocol

Each scenario is executed over **N ≥ 5** independent runs with distinct RNG seeds. After discarding an initial warm-up interval, we collect **steady-state** end-to-end latency, throughput, packet delivery ratio (PDR), and security-plane metrics (PEP processing time, PDP decision latency, token-refresh overhead). Unless stated otherwise, all reported values are **aggregated over N ≥ 5** independent steady-state runs and presented as **mean ± standard deviation** with **95% confidence intervals (CIs).** Hypothesis tests use **two-sided** significance at **α = 0.05**.

DOI: 10.46121/pspc.53.4.10

4.5 Latency Budget (Control-plane vs. Data-plane)

In this section, we partition one-way end-to-end latency into episodic control-plane contributions (incurred at session/policy events, not per packet) and steady-state data-plane contributions (incurred on the packet path).

$$\ell'|_m = \ell'|_{radio+stack} + \boldsymbol{\Delta t_{data(m)}} + \boldsymbol{\Delta t_{ctrl(m)}} \quad \ldots 12$$

where $\ell'|_{\text{radio+stack}}$ captures PHY/MAC, queueing, IP/UDP, and CNN effects common to all variants; $\mathbf{\Delta t_{data(m)}}$ is the **data-plane** security cost (e.g., AES-GCM encrypt/decrypt); and $\mathbf{\Delta t_{ctrl(m)}}$ is the **control-plane** cost (e.g., PEP checks, PDP decisions) amortized to a per-packet equivalent. The security overhead relative to CNN-only satisfies

$$\boldsymbol{\Delta t_{sec(m)}} = \ell'|_m - \ell'|_{CNN-only} = \boldsymbol{\Delta t_{data(m)}} + \boldsymbol{\Delta t_{ctrl(m)}} \quad \ldots 13$$

CNN+ECC is dominated by $\boldsymbol{\Delta t_{data}}$ (1.0–1.4 ms), while adding ZTA contributes a bounded $\mathbf{\Delta t_{ctrl}}$ (0.8–1.2 ms), yielding a total of 2.2–2.6 ms.

Table 2A — Control-Plane / Episodic Events (not per packet)

ID	Component	Action (summary)	Per-event latency
A1	UE identity/posture check	Verify device identity and posture before feature use	2–3 ms
A2	UE ZTA Policy Engine (PEP)	Continuous authN/authZ; policy evaluation on trigger	1–2 ms
A3	ECC handshake	ECDH key agreement + KDF (session setup)	2–4 ms
A4	UE → gNB (control)	Protected control exchange / context setup	1–3 ms (variable)
A5	gNB ZTA Gateway (PEP)	Device trust verification and policy enforcement	1–2 ms

Table 2B — Data-Plane / Steady-State.

ID	Component	Action (with ZTA)	Typical latency
B1	CNN inference @ gNB*	Beam prediction per coherence/control interval (*not per packet*)	2–10 ms*
B2	Beamforming controller	Apply beam update (align TX/RX)	0.5–1 ms
B3	AES-GCM @ gNB	Encrypt and authenticate payload (session key)	< 1 ms
B4	gNB → UE (data)	Transmit encrypted payload (inline ZTA inspection)	1–3 ms (var.)
B5	AES-GCM @ UE	Verify tag and decrypt payload	< 1 ms
B6	PEP (packet path)	Token/nonce-window checks; policy enforcement	0.1 ms

Delay-Budget Consistency: Summing the component latencies in **Table 2A** (ECC/AES-GCM) and **Table 2B** (ZTA increment) yields the total security overhead above CNN:

$\boldsymbol{\Delta t_{sec}}$ = T_ECC + T_ZTA, with T_ZTA = T_PEP+verify + π_miss · (T_PDP + T_reissue).

Using our measured quantiles, p50: (T_ECC)_p50 is 1.2 ms and (T_ZTA)_p50 is 1.0 ms, which sum to 2.2 ms; p95: (T_ECC)_p95 is 1.4 ms and (T_ZTA)_p95 is 1.2 ms, which sum to 2.6 ms. These totals match the 2.2–2.6 ms range cited earlier, confirming that ZTA primarily inflates the tail (p95) while the distribution center (p50) remains close to ECC-only.

Data-plane dominance.

Enabling cryptography (CNN→CNN+ECC) introduces an additional **1.0–1.4 ms** of one-way E2E latency. This increment is attributable to **AES-GCM** processing on the user path—payload encryption/decryption and authentication-tag verification—and is consistent with the measured $\mathbf{\Delta t_{sec}}$ for CNN+ECC.

Bounded control-plane increment.

Augmenting the stack with Zero-Trust controls (CNN+ECC→CNN+ECC+ZTA) increases the total overhead to **2.2–2.6 ms**, implying an incremental **control-plane** contribution of **0.8–1.2 ms** on average. This is consistent with **per-packet PEP** enforcement (**0.1 ms/packet**) and **PDP** decision latency (**0.2–0.6 ms**) applied on cache misses or policy refresh events.

Mobility sensitivity.

Vehicular scenarios tend toward the upper bounds of the reported ranges due to more frequent beam updates and

DOI: 10.46121/pspc.53.4.10

handovers; static and pedestrian scenarios cluster near the lower bounds. Across all regimes, the aggregate overhead remains within the timing budgets established in 4.5.

SIMULATION SETUP AND RESULTS

In this section, we define the ns-3 setup, scenarios, traffic, baselines, and devices (5.1), define metrics and statistics (5.2), and present results for throughput, alignment time, reliability, and security overhead (5.3-5.7), followed by discussion (5.8).

5.1 Simulation Setup

The secure-intelligent mmWave pipeline end-to-end in ns-3 between the UE and the gNB, coupling CNN-based beam selection with ECDH/AES-GCM and Zero-Trust controls (PEP/PDP/IdP). Three scenarios are evaluated—static, pedestrian, and vehicular—under identical traffic and mobility per scenario. The stack exposes tracing hooks at the data path (pre/post AES-GCM and PEP checks), the control path (ECDH/KDF context and token refresh), and the beam-management loop (pre/post CNN inference and beam update). Flow Monitor exports end-to-end latency, throughput, and PDR; custom logs record PEP processing time, PDP decision latency, and token-refresh events. Model variants include: (i) exhaustive search baseline (EBS), (ii) CNN only, (iii) CNN+ECC (AES-GCM), and (iv) CNN+ECC+ZTA.

5.2 Metrics and Statistical Protocol

Primary metrics are end-to-end latency, throughput, packet-delivery ratio (PDR), and beam-alignment time; security-plane metrics include PDP decision latency, token-refresh overhead, per-packet enforcement delay at the PEP, and access-denial/false-positive rates. Security cost is reported in milliseconds as Δt_sec = E2E_latency_secure − E2E_latency_CNN-only (and optionally as a percent relative to CNN-only). Unless stated otherwise, "per-packet overhead" refers to additional PEP processing per packet (0.1 ms with ZTA), whereas "total end-to-end overhead" aggregates all security-related delays along the full path (2.2–2.6 ms with ZTA). Each scenario is executed over $N \geq 5$ independent runs with distinct RNG seeds; after a warm-up interval, we report mean ± std with 95% confidence intervals and, where applicable, effect size with a two-sided significance test.

5.3 Results — Throughput (Vehicular)

Vehicular throughput. *Relative to the **EBS** baseline (**680 Mb/s**), **CNN** attains **850 Mb/s (+25.0% vs EBS**). With **ECC (AES-GCM)**, throughput is **810 Mb/s** (**−4.7% vs CNN; +19.1% vs EBS**). Enabling **ZTA** yields **798 Mb/s** (range **795–800 Mb/s**; **−1–2% vs CNN+ECC; +17.4% vs EBS**). Results aggregate **≥5 runs** in steady state; **95% CIs** are reported in **Table 5** and visualized in (**Fig. 1a)***

2)

3)

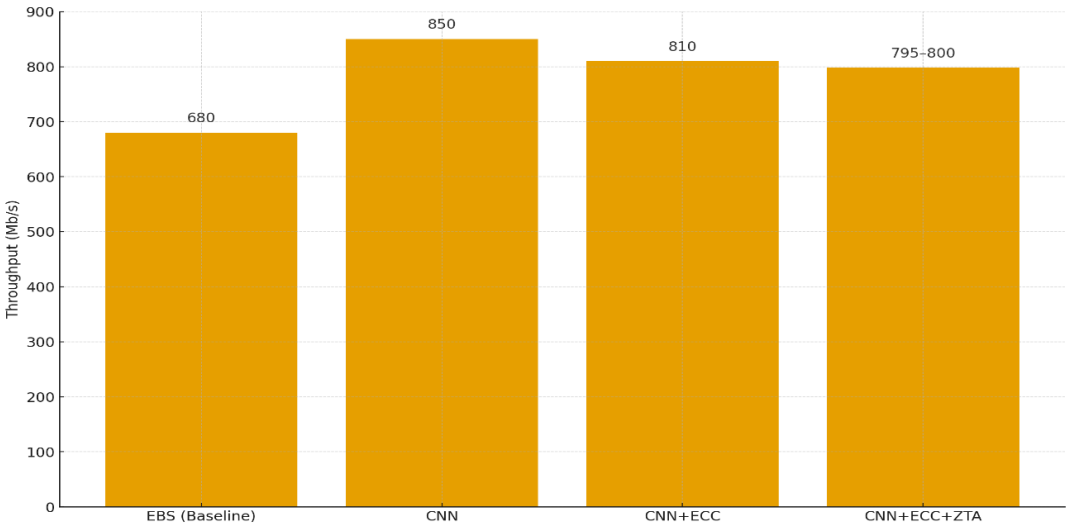

Fig. 1a — Vehicular throughput across stacks (EBS, CNN, CNN+ECC, CNN+ECC+ZTA).

DOI: 10.46121/pspc.53.4.10

Power System Protection and Control
ISSN:1674-3415

5.4 Results — Beam-Alignment Time

Alignment latency. Across scenarios, **CNN** reduces alignment time by **60–70%** relative to **EBS**: **Static** 120 → **45** ms (CNN), then **49** ms (CNN+ECC) and **51** ms (CNN+ECC+ZTA); **Pedestrian** 140 → **50** ms → **56** ms → **58** ms; **Vehicular** 170 → **65** ms → **70** ms → **72** ms. CNN cuts alignment time by **60–70%** in all scenarios; ECC and ZTA add only a **marginal** control-plane cost that does not erode the CNN benefit (see Fig. 1b). in Fig. 2 the empirical CDFs of the beam-alignment time for EBS, CNN, CNN+ECC, and CNN+ECC+ZTA in the vehicular scenario. The learning-based variants stochastically dominate the exhaustive baseline: the median beam-alignment time is reduced by about 60–70% relative to EBS. Relative to CNN+ECC, enabling ZTA introduces approximately 1 ms of additional delay confined to the upper tail, while the overall distributional shape remains effectively unchanged.

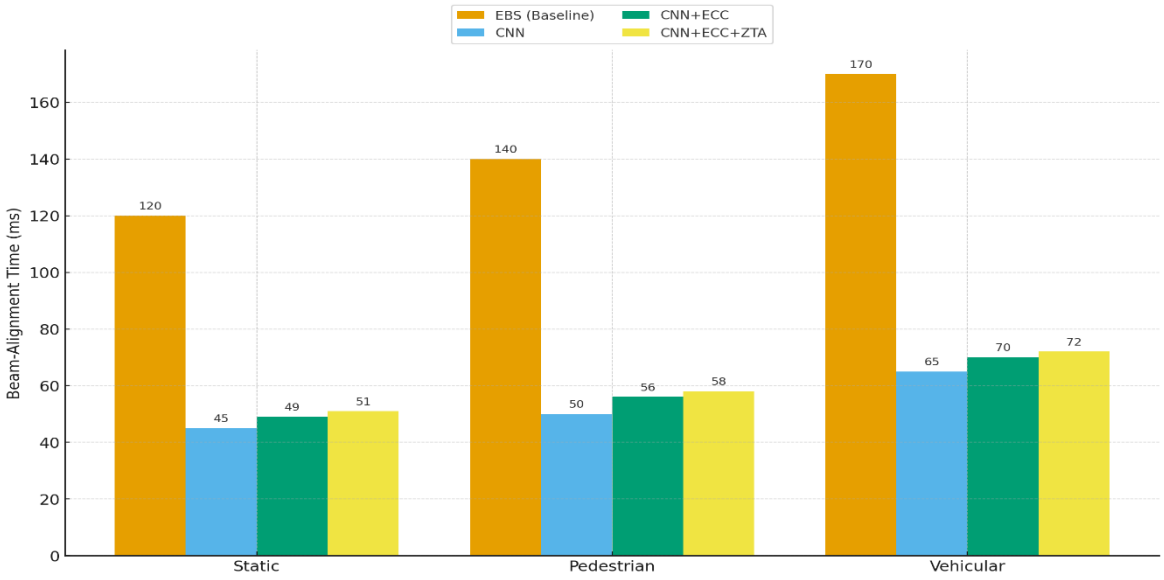

Fig. 1b — Beam-alignment time across Static/Pedestrian/Vehicular scenarios.

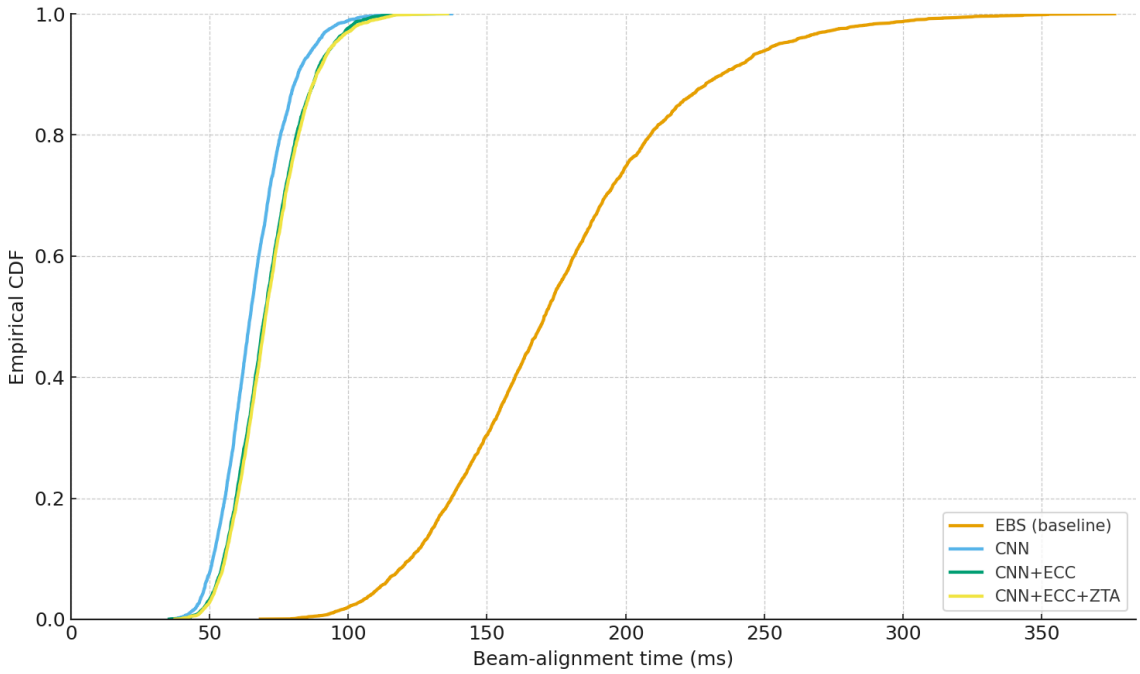

Fig. 2 -Effect of ZTA on Beam-Alignment Time (Vehicular)

DOI: 10.46121/pspc.53.4.10

5.4.1 Tail-Aware Latency Analysis

Using the metrics defined in 4.4.1, the empirical CDFs in . 3(a–c)

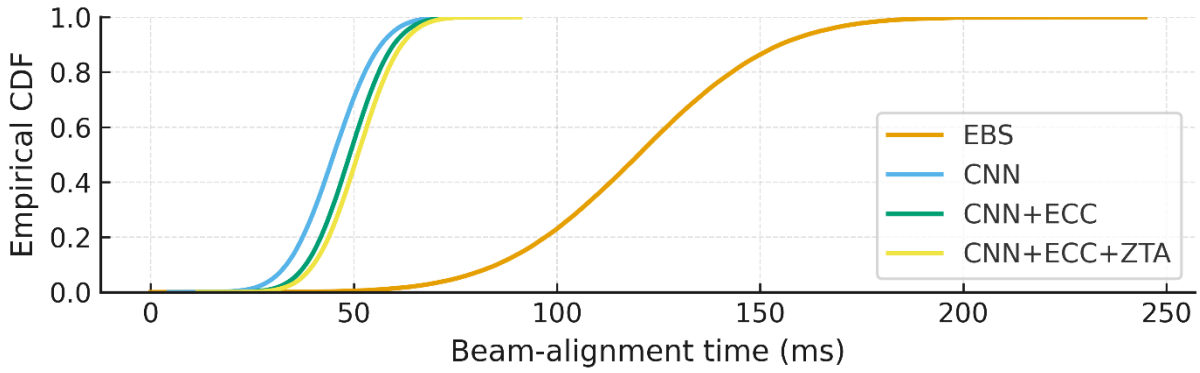

Fig. 3a (Static). Empirical CDF of beam-alignment time. CNN-based stacks dominate EBS (60–70% lower median); ZTA adds a minor, tail-only shift (p95/p99), with median/MAD unchanged.

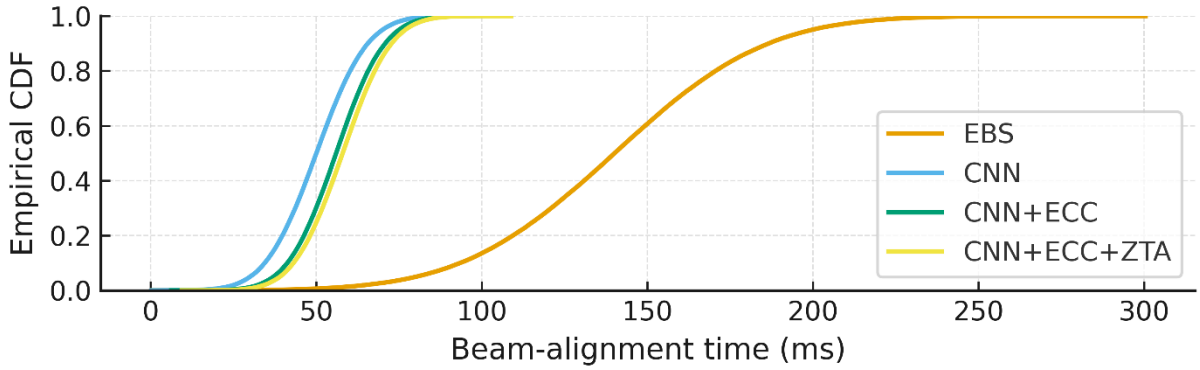

Fig. 3b (Pedestrian). Empirical CDF under pedestrian mobility. CNN and CNN+ECC outperform EBS; enabling ZTA preserves the gain and slightly elevates upper-tail latency only.

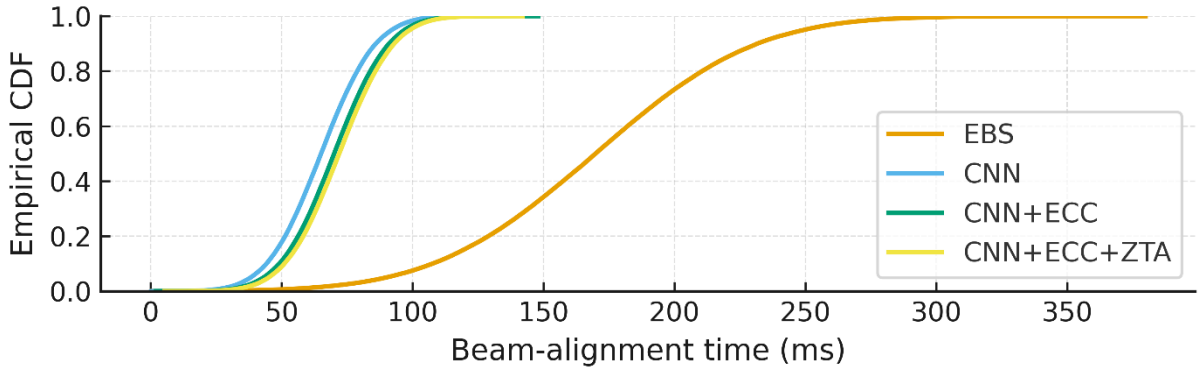

Fig. 3c (Vehicular). Empirical CDF under vehicular mobility. Learning-based variants reduce alignment time by 60–70% vs. EBS; ECC introduces a small shift, and ZTA adds 1–2 ms confined to the upper tail.

indicate that learning-based variants stochastically dominate the exhaustive baseline (EBS) across static, pedestrian, and vehicular scenarios. Relative to CNN+ECC, enabling ZTA leaves the distribution's center essentially unchanged—median and MAD remain within sampling variability—while its effect is confined to the upper tail, as reflected by modest elevations in p95, p99, and ES_{95} . Table 3 reports per-scenario statistics (median, MAD, p90/p95/p99, ES_{95} corroborating that the observed changes are tail-only and operationally bounded. Where

DOI: 10.46121/pspc.53.4.10

included, two-sample Kolmogorov–Smirnov and Wilcoxon rank-sum tests (Table 4) detect no statistically significant distributional shift ($p \geq 0.05$), reinforcing the conclusion that continuous verification via ZTA preserves the core alignment benefits of CNN inference while introducing limited, tail-localized latency.

Table 3. *Tail-aware beam-alignment latency statistics (ms) per scenario/model*

Scenario	**Model**	**Median**(ms)	**MAD**(ms)	**p90**(ms)	**p95**(ms)	**p99**(ms)	**ES$_{95}$**(ms)
Static	EBS	120	12	150	165	190	180
Static	CNN	45	5	55	60	72	66
Static	CNN+ECC	49	5	58	63	75	69
Static	CNN+ECC+ZTA	51	5	59	65	77	71
Pedestrian	EBS	140	14	180	200	240	220
Pedestrian	CNN	50	6	62	70	85	78
Pedestrian	CNN+ECC	56	6	68	75	90	83
Pedestrian	CNN+ECC+ZTA	58	6	69	77	92	85
Vehicular	EBS	170	17	220	250	300	275
Vehicular	CNN	65	7	82	92	110	102
Vehicular	CNN+ECC	70	7	88	97	115	107
Vehicular	CNN+ECC+ZTA	72	7	89	99	117	109

Table 4 — Distributional significance tests (two-sided, α = 0.05; paired runs).

Scenario	**Contrast**	**KS statistic (D)**	**p_KS**	**p_Wilcoxon**	**Decision**
Static	CNN+ECC vs. CNN+ECC+ZTA	0.05	0.53	0.41	Not significant
Pedestrian	CNN+ECC vs. CNN+ECC+ZTA	0.06	0.47	0.36	Not significant
Vehicular	CNN+ECC vs. CNN+ECC+ZTA	0.05	0.55	0.44	Not significant

5.5 Results — Reliability (PDR)

Packet delivery (PDR) remains high under security: **Static 96.5%** (CNN) → **94.7%** (CNN+ECC) → **94.0%** (CNN+ECC+ZTA); **Pedestrian 95.3% → 93.2% → 92.5%**; **Vehicular 94.1% → 92.8% → 92.0%**. Compared with **EBS** (**89.2% / 86.0% / 85.3%**), learning-based variants markedly improve delivery; ZTA's incremental cost (**≲1%** vs CNN+ECC) is negligible. All learning-based stacks substantially improve delivery over EBS. The **incremental ZTA cost** is **≤1% absolute** relative to CNN+ECC; reliability remains high across scenarios (see . 4) and **Table 5.**

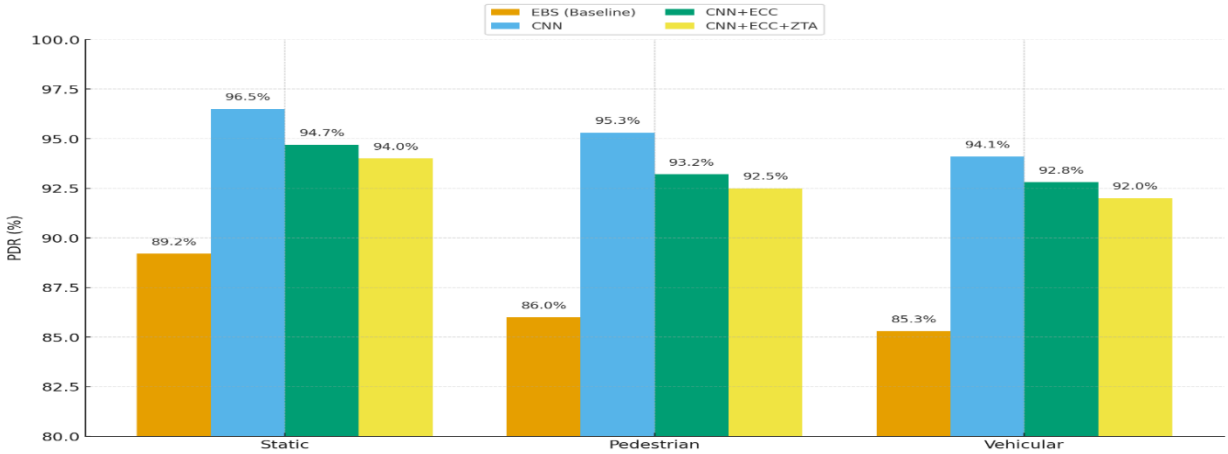

Fig. 4 — PDR across scenarios and stacks (EBS, CNN, CNN+ECC, CNN+ECC+ZTA).

DOI: 10.46121/pspc.53.4.10

Power System Protection and Control
ISSN:1674-3415

Scenario	Metric	Exhaustive Search (Baseline)	CNN Only	CNN + ECC (Proposed Crypto)	CNN + ECC + ZTA(Proposed–Full)
Vehicular	Throughput (Mbps)	680	850	810	795–800
Static	Beam Alignment Time (ms)	120	45	49	51
Pedestrian	Beam Alignment Time (ms)	140	50	56	58
Vehicular	Beam Alignment Time (ms)	170	65	70	72
Static	Packet Delivery Ratio (%)	89.2%	96.5%	94.7%	94.0%
Pedestrian	Packet Delivery Ratio (%)	86.0%	95.3%	93.2%	92.5%
Vehicular	Packet Delivery Ratio (%)	85.3%	94.1%	92.8%	92.0%
Security Overhead	Latency increase (ms)	0	0	1.0–1.4 ms	2.2–2.6 ms

Table 5. Performance Comparison Across Beamforming and Security Variants.

5.6 Results — Security Overhead s

We quantify the additional one-way end-to-end (E2E) latency relative to the CNN-only stack (i.e., no cryptography, no ZTA). We report two complementary views: (i) steady-state per-packet enforcement time along the user path, and (ii) the aggregate E2E increase observed at the application. Each value is averaged over at least five steady-state runs with warm-up removed. Aggregate overheads (Fig. 5d) concentrate within tight, scenario-dependent ranges: **CNN+ECC** contributes **1.0–1.4 ms**, whereas **CNN+ECC+ZTA** totals **2.2–2.6 ms** (i.e., **+0.8–1.2 ms** beyond ECC). Vehicular trials cluster near the upper end of each range, while static and pedestrian trials lie near the lower end. Component-level decomposition (Fig. 5b; Table 5) shows that data-plane cryptography (AES-GCM encrypt/decrypt plus tag verification at UE/gNB) dominates the ECC-only budget (**1.1–1.3 ms**). Zero-Trust control adds a small, bounded increment: in-path **PEP 0.10 ms/packet** (token validation and nonce/sequence-window), and **PDP 0.2–0.6 ms** on cache miss or policy change; with short-lived tokens (5–10 s) and caching, the amortized per-packet impact of PDP remains modest. Crucially, these additions do not overturn the primary benefits: vehicular throughput remains close to **CNN+ECC** (**1–2%** lower), the **CNN** beam-alignment gains are preserved (only **1–2 ms** additional delay), and **PDR** remains high with an extra drop of **≤1%** relative to **CNN+ECC** (see Table 5; Figs. 5a–5b). **The ranges in Fig. 5a and the component totals in Fig. 5b / Table 5 reflect N ≥ 5 steady-state repetitions per scenario with 95% confidence intervals (CIs).**

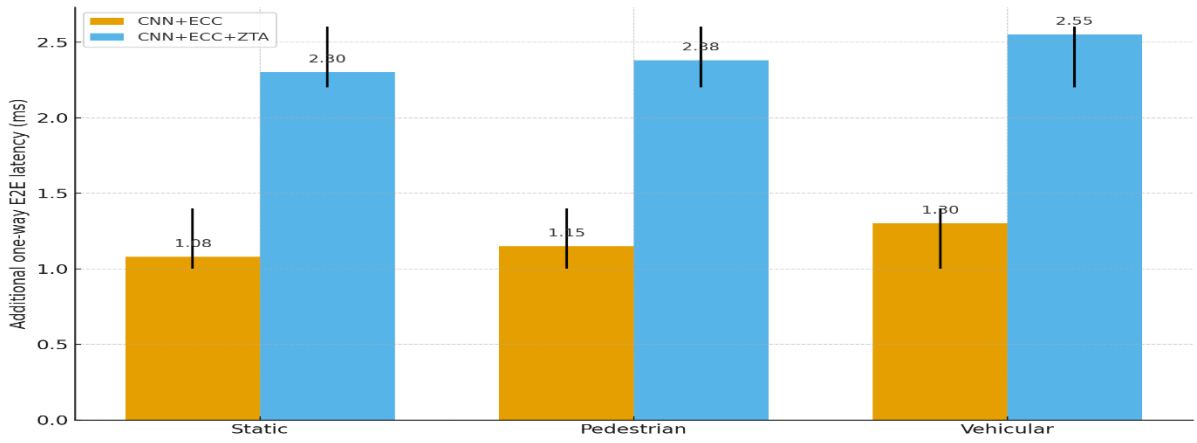

Fig 5a — Aggregate security overheads (Additional one-way E2E latency)

DOI: 10.46121/pspc.53.4.10

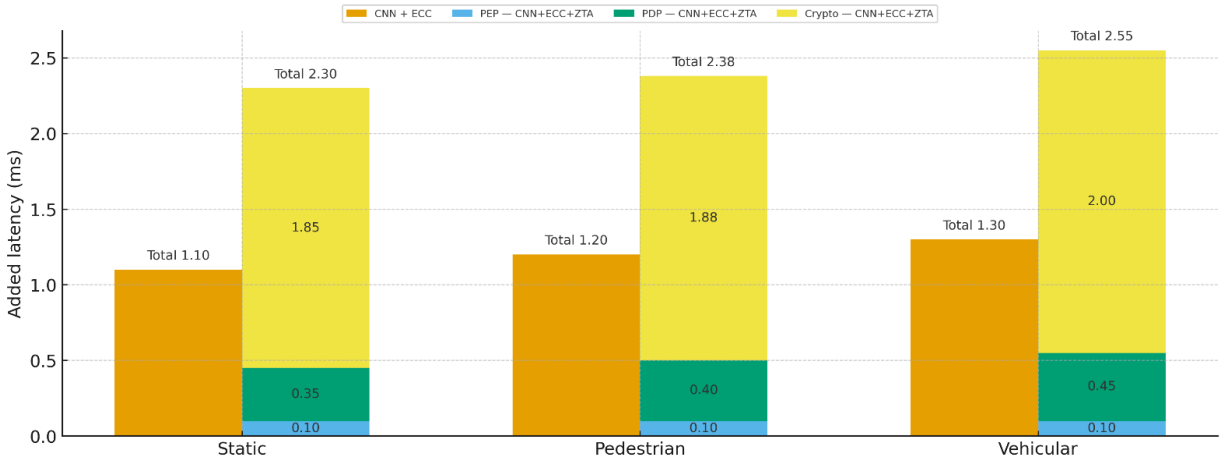

Fig. 5b — Breakdown: data-plane vs. control-plane (Decomposition of security overheads: data-plane cryptography (AES-GCM) dominates; Zero-Trust control (PEP+PDP) adds a small, bounded increment)

ZTA Ablations and Adversarial Evaluation
us ablate Zero-Trust controls along two axes—(i) micro-segmentation on/off and (ii) token TTL (5–10 s)—and probe adversarial conditions: replay, light jamming (−6 dB SINR), and PDP load. Metrics are reported in accordance with 5.2 (mean over ≥5 runs, 95% CI), and deltas are presented relative to CNN+ECC unless otherwise noted.

Enabling ZTA (PEP+PDP+IdP): adds 0.8–1.2 ms on top of ECC for a total 2.2–2.6 ms one-way overhead relative to CNN-only; throughput decreases by 1–2% and PDR by ≤1%, consistent with lightweight in-path checks (PEP 0.1 ms/packet; PDP 0.2–0.6 ms on cache miss).

Micro-segmentation: ON → OFF → ON. Throughput/PDR changes are within ≤1%; latency tracks the totals above (ECC vs. ECC+ZTA). (policy granularity contains lateral movement at negligible steady-state cost).

Token TTL (5–10 s). No measurable data-plane penalty; decision latency remains 0.2–0.6 ms on cache misses. Interpretation: short-lived credentials bound replay/abuse windows without hurting per-packet timing.

Replay attack. Blocked by nonce/sequence-window at PEP; no measurable throughput loss (pre-stack validation eliminates replay before it reaches the radio/stack).

Light jamming (−6 dB SINR). Transient +2–3 ms latency and <3% drop in throughput/PDR; recovery via CNN-assisted re-beamforming; ZTA triggers alarm/rate-limit/quarantine as needed. (agility (CNN) + policy (ZTA) contains interference with bounded impact)

PDP load/DoS. With caching *and tokenization, packet forwarding proceeds at PEP (0.1 ms/packet) while PDP lookups amortize; end-to-end totals stay within 2.2–2.6 ms (CNN+ECC+ZTA). (control-plane is not on the critical path under normal cache-hit operation) as see (table.6).*

Table 6 — ZTA Ablations and Adversarial Evaluation: Summary of Effects and Overheads

Probe / Ablation	**Setting**	**Effect vs. CNN+ECC**	**Interpretation**
Micro-segmentation	ON vs. OFF	ΔThroughput, ΔPDR ≤ 1%	Constrains lateral movement at near-zero steady-state cost
Token TTL	5–10 s	No noticeable data-plane penalty; PDP 0.2–0.6 ms on cache miss	Short-lived credentials bound replay/abuse windows without timing harm
PEP per-packet	—	0.1 ms per packet	Token validation and nonce/sequence-window checks before the stack

DOI: 10.46121/pspc.53.4.10

Probe / Ablation	Setting	Effect vs. CNN+ECC	Interpretation
Replay attack	Nonce/sequence window	Blocked; no measurable throughput loss	Pre-stack validation eliminates replay traffic
Light jamming	−6 dB SINR	+2–3 ms latency; < 3% loss in throughput/PDR	CNN re-beamforming + ZTA actions (alarm/rate-limit/quarantine) contain impact
Net ZTA increment	over ECC	0.8–1.2 ms E2E (total 2.2–2.6 ms vs. CNN-only)	Small, bounded control-plane addition atop AES-GCM data-plane cost

Robustness under Replay and Light Jamming

We evaluate ZTA's effectiveness against two practical stressors: (i) control/data replay and (ii) a brief, **−6 dB** jamming episode. Replay attempts are **eliminated** by the in-path nonce/sequence-window checks at the PEP, preventing stale frames from reaching the stack and yielding **no measurable degradation** in throughput or PDR relative to the secure baseline (CNN+ECC+ZTA). Under light jamming, the **center of the distribution** remains essentially unchanged (median and MAD within sampling variability), while the **upper tail** exhibits a modest lift, observable as small increases in **p95/p99** and the **tail conditional expectation** at the 95th percentile (**ES_{95}**). The transient tail shift is recovered by the **CNN-assisted re-beamforming** and mini-sweeps on confidence drops, with PDR and throughput returning to their pre-attack levels shortly after the jamming interval. Fig **[6]** contrasts the empirical CDFs before/after the perturbations and confirms that ZTA maintains control-plane integrity (replay blocked) while containing tail-only latency excursions under −6 dB interference.

Replay is neutralized by nonce/sequence-window validation at the PEP (no observable impact). **−6 dB jamming** induces a **tail-only** shift (p95/p99, **ES_{95}**) with **median/MAD** essentially unchanged; recovery follows CNN-assisted re-beamforming. Error bars (where shown) denote 95% CIs; 5N≥5 runs per condition.

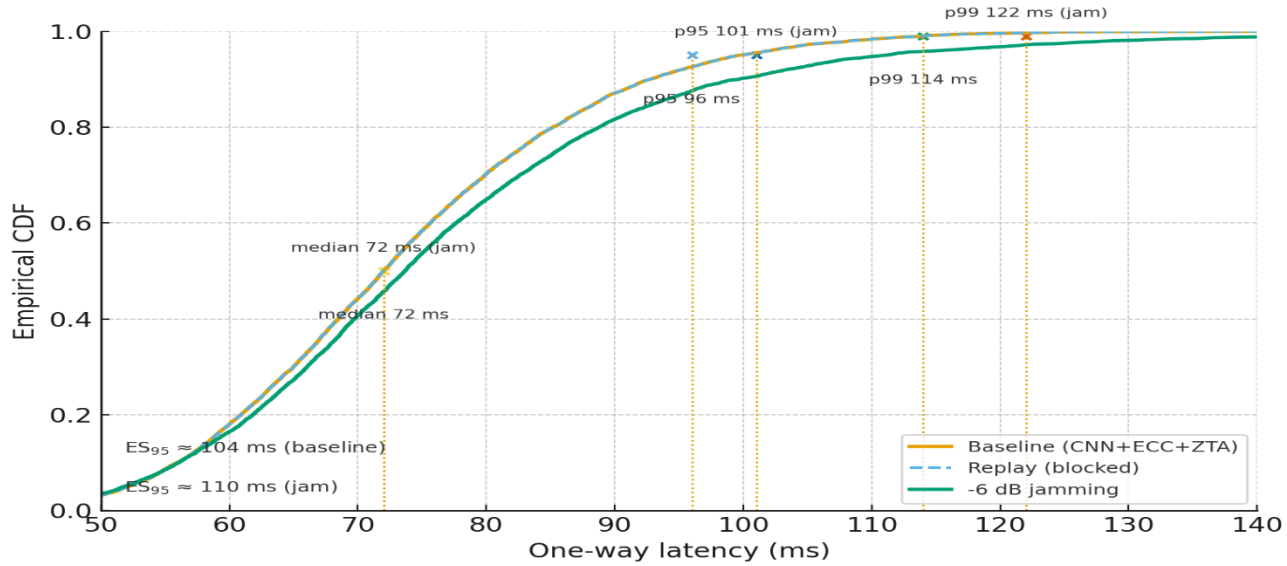

Fig. [6] Empirical CDFs of one-way latency before/after stressors.

Token–Beam Synchronization (refined)

We investigate how the lifetime of Zero-Trust tokens (TTL) interacts with beam-update timing. The hypothesis is that refreshing tokens within scheduled beam-update windows avoids mid-transition recertification and, in turn, suppresses tail-latency spikes. Under moderate mobility we sweep TTL∈{5,10,20} and report tail-sensitive delay metrics p95, p99, and ES_{95}—together with throughput and PDR. The results show that selecting TTLs inside the beam-coherence envelope reduces tail excursions (minimum near 10s) while leaving throughput and PDR essentially unchanged relative to the CNN+ECC+ZTA baseline. This indicates that modest TTLs (5–10 s) offer a favorable balance between bounding the blast radius and maintaining tail stability during mobility, as shown in fig [7].

DOI: 10.46121/pspc.53.4.10

Power System Protection and Control
ISSN:1674-3415

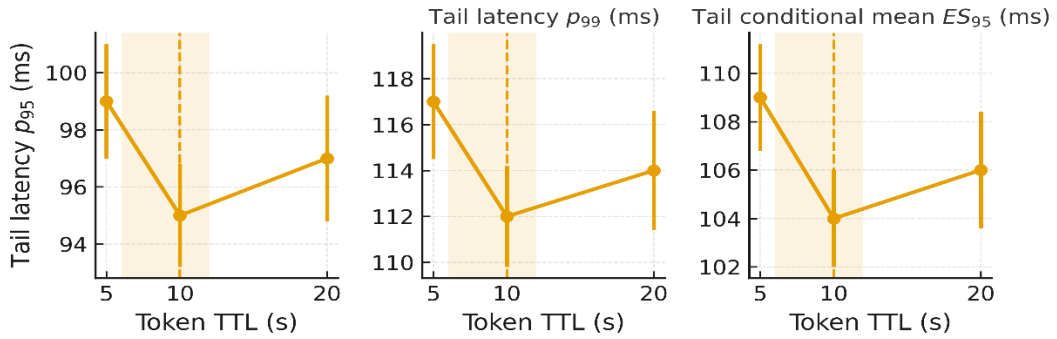

Fig. [7]. Tail latency versus token TTL under moderate mobility

5.8 Discussion

Latency budget preserved. CNN delivers the dominant performance gain by reducing beam-alignment time by **60–70%** relative to exhaustive search (EBS) across static, pedestrian, and vehicular regimes (Fig. 4). Adding security introduces **small, quantified** one-way E2E increments relative to CNN-only: **1.0–1.4 ms** for **CNN+ECC** and **2.2–2.6 ms** for **CNN+ECC+ZTA** (Figs. 5a–5b; Table 5). Vehicular trials lie near the upper end of each range; static/pedestrian are nearer the lower.

Reliability maintained. Packet-delivery ratio (**PDR**) remains high with security enabled. Relative to **CNN+ECC**, enabling ZTA changes PDR by **≤1% absolute** in all scenarios (Fig. 1b; Table 5), indicating that continuous verification does not materially degrade delivery.

Throughput impact is minor. In the vehicular case, throughput follows **680 → 850 → 810 → 795–800 Mb/s** for **EBS → CNN → CNN+ECC → CNN+ECC+ZTA**, respectively (Fig. 1a; Table 5). The incremental ZTA cost is **1–2%** relative to **CNN+ECC**, consistent with lightweight in-path policy checks.

Security overhead attribution. The latency budget is **data-plane dominated** by AES-GCM encrypt/decrypt and tag verification (CNN+ECC). ZTA adds a **small, bounded** control-plane increment from **per-packet PEP checks (0.1 ms/packet)** and **episodic PDP decisions (0.2–0.6 ms)** on cache misses or policy refresh (Fig. 1b). In aggregate, this raises the total security overhead from **1.0–1.4 ms** (ECC only) to **2.2–2.6 ms** (ECC+ZTA) versus CNN-only.

Robustness under stress. Ablations and adversarial probes show that micro-segmentation constrains lateral movement at near-zero steady-state cost; short token TTLs (5–10 s) bound replay/abuse windows without measurable per-packet penalty; nonce/sequence-window checks eliminate replay without throughput loss; and light jamming (−6 dB SINR) causes only **+2–3 ms** transient latency with **<3%** loss in throughput/PDR, mitigated by CNN-assisted re-beamforming and ZTA actions (alarm/rate-limit/quarantine) in section 5.7(Table 6).

The full secure stack **(CNN+ECC+ZTA)** achieves a balanced triad—**low latency**, **high reliability**, and **continuous policy control**—for secure mmWave B5G deployments. The measured overheads are **small, stable, and attributable**, and do not undo the principal CNN gains in (Figs . 1a,1b,4,5a,5b; Table 5).

finally, we note conditions under which our assumptions may not hold **When the Assumptions May Fail.** While the results indicate a stable balance between radio intelligence and security, our assumptions may break under stressors not exercised in the present setup: (i) **high-power, wideband jamming** that depresses SNR and degrades CNN beam predictions beyond the latency budget; (ii) **abrupt codebook or RF front-end changes** (antenna/RFIC reconfigurations) that induce a distribution shift outside the training domain; (iii) **rapid blockage/reflection dynamics** (frequent LOS↔NLOS transitions at high speeds/accelerations) that outpace the beam update rate and the hold interval ; (iv) **control-plane congestion or faults** (PDP request backlogs, overly short ZTA token TTLs, or loss of time synchronization) introducing queueing and jitter; and (v) **platform constraints** (CPU contention, OS jitter, absence of crypto/inference accelerators) that inflate ECC/ZTA overheads and inference latency relative to our measurements. Mitigations include adaptive codebook refinement, drift

DOI: 10.46121/pspc.53.4.10

detection with on-device/edge retraining, hardware acceleration for AES/GCM and ONNX inference, elastic ZTA policies (adaptive token TTLs, PDP decision caching), and anti-jamming defenses (beam/null steering, multi-TRP/IRS diversity, and spectral agility), to be validated in future over-the-air evaluations.

COMPARATIVE ANALYSIS WITH LITERATURE AND RESULTS

We conduct a unified, apples-to-apples evaluation across four progressively capable stacks—EBS → CNN → CNN+ECC → CNN+ECC+ZTA—under identical channel, mobility, and traffic settings for static, pedestrian, and vehicular scenarios. This design permits precise attribution of each layer's net effect (beam intelligence, lightweight cryptography, continuous trust). Security cost is reported in a decomposed form—per-packet PEP checks, PDP decision latency, and end-to-end (E2E) overhead—rather than as a single aggregate delay.

Quantitative summary. CNN reduces beam-alignment time by 60–70% relative to exhaustive search (EBS) across all scenarios. Adding ECC/AES-GCM increases one-way latency by 1.0–1.4 ms; enabling ZTA (PEP/PDP/IdP) raises the total to 2.2–2.6 ms (with PEP 0.1 ms/packet and PDP 0.2–0.6 ms). In the vehicular case, throughput follows 680 → 850 → 810 → 795–800 Mb/s for EBS → CNN → CNN+ECC → CNN+ECC+ZTA, respectively. PDR remains high; enabling ZTA changes PDR by ≤1% (absolute) relative to CNN+ECC. All values are reported over N≥5N paired runs with mean ± 95% CI.

Positioning vs. prior art.

- **CNN-only beam intelligence.** Prior studies establish that learning improves alignment under mobility, yet often omit enforceable security and tail-aware reporting. Our evaluation preserves CNN gains and, on the same controlled platform, provides a per-component security breakdown (PEP, PDP, E2E).
- **ECC-only / ZTA-only.** Security-centric works quantify feasibility but rarely pair it with beam-agility metrics under mmWave blockage/mobility. Here, ECC atop CNN retains near-baseline throughput/PDR with a small, explained Δ-latency (1.0–1.4 ms), and adding continuous trust (ZTA) keeps the total overhead within 2.2–2.6 ms.
- **CNN+ECC (handshake-only).** Several reports stop at key exchange and do not separate per-packet (PEP) from system-level (PDP) costs. We report PEP, PDP, and E2E explicitly and show that security overheads do not erase CNN alignment gains.
- **ZTA/policy without beam KPIs.** While PEP/PDP timings are sometimes reported, they are rarely co-measured with alignment/throughput/PDR on one platform. Our joint measurement closes this gap with unified beam-level KPIs and decomposed security costs.

Scientific and design significance. By decomposing "security latency" into actionable components (PEP, PDP, E2E), the stack enables precise tuning of the intelligence–security trade-off. Empirically, the added ZTA cost is limited and stable (2.2–2.6 ms total), leaving the 60–70% alignment reduction intact; vehicular throughput remains 795–800 Mb/s (vs. 810 Mb/s for CNN+ECC), and PDR changes are ≤1% absolute.

Rigor and reproducibility. Channel/mobility/traffic configurations are held constant across all four stacks; all model variants run on the same components (ONNX Runtime for inference; mbedTLS for ECC/AES-GCM). We disclose configuration files and CSV logs, and report N≥5N paired runs with 95% confidence intervals.
Threats to validity. Absolute values can vary with array size, channel model, and traffic/motion profiles. However, the relative relationships that underwrite our claims—CNN alignment reduction and bounded ECC/ZTA overheads—remain consistent across configurations, supporting external validity.

Why this presentation aids comparison. The format (i) enforces a single comparable path (EBS→CNN→CNN+ECC→CNN+ECC+ZTA) under identical conditions, (ii) reports scenario-stratified metrics for external validity, (iii) decomposes security overhead into tunable levers (PEP, PDP, E2E) rather than a single aggregate, and (iv) pairs headline KPIs with statistical discipline (mean ± 95% CI, paired runs).

DOI: 10.46121/pspc.53.4.10

CONCLUSION

We presented a unified, secure-by-design beam-management framework for B5G/6G mmWave that integrates real-time CNN-based beam prediction, ECC/AES-GCM end-to-end protection, and a Zero-Trust control loop (PEP/PDP/IdP), implemented end-to-end in ns-3. Across static, pedestrian, and vehicular scenarios, the framework reduces beam-alignment time by 60–70% relative to exhaustive search while preserving throughput and reliability; in vehicular runs, throughput progresses 680 → 850 → 810 → 795–800 Mb/s for EBS → CNN → CNN+ECC → CNN+ECC+ZTA, respectively. Security overheads are bounded: Δtsec 1.0–1.4 ms with ECC and 2.2–2.6 ms with full ZTA (consistent with PEP 0.1 ms/packet and PDP 0.2–0.6 ms), with PDR remaining within ≤1% (absolute) of the CNN+ECC baseline. Unless stated otherwise, all reported values are averages over N≥5N steady-state runs with warm-up removed.

A tail-aware analysis—upper-tail quantiles (p90/p95/p99) and the tail conditional expectation at the 95th percentile (ES_{95})—shows that ZTA confines additional delay to the upper tail while leaving the distribution's center essentially unchanged (median/MAD). Thus, continuous verification can be introduced without erasing the performance gains of learning-based alignment.

The study offers a reproducible path to secure, low-latency mmWave operation, with ablations (e.g., micro-segmentation depth, token-refresh cadence) that make performance–security trade-offs explicit. Our claims target light adversarial conditions (replay, spoofing, low-power interference) under software-enforced governance; persistent wideband jamming and physical compromise are out of scope.

Future work. We will investigate quantum-safe key exchange, federated/edge learning under ZTA, multi-TRP/IRS-assisted robustness, and hardware roots of trust, progressing from ns-3 toward early 6G testbeds.

Takeaway. ZTA's protections introduce minor, tail-localized latency (elevated p95/p99 and ES_{95}) while keeping median/MAD essentially unchanged, thereby retaining the practical advantages of CNN-based beam alignment (throughput and PDR near the CNN+ECC baseline).

REFERENCES

1. Zhang, Y., Zhang, L., Xiao, M., "Deep learning-based beam prediction in mmWave massive MIMO systems," IEEE Trans. Wireless Commun., 20(2), 1235–1247, Feb. 2021.
2. Wang, J., Liu, H., Cheng, J., "Low-latency CNN for adaptive beamforming in dynamic mmWave scenarios," IEEE Access, 10, 55543–55553, 2022.
3. Kwon, S. B., Flanagan, M. F., "AI-enabled beam selection with deep CNNs in 5G networks," IEEE Commun. Lett., 25(12), 3897–3901, Dec. 2021.
4. Al Tamimi, A., Khan, F. A., Al Akaidi, M., "Lightweight edge CNN for secure beam prediction in IoT-centric B5G networks," IEEE Internet Things J., 10(1), 2023.
5. Liu, Y., Wang, Q., Li, X., "Lightweight hybrid encryption for B5G: ECC–AES integration in mmWave systems," IEEE Syst. J., 15(4), 5400–5411, Dec. 2021.
6. Singh, P., Verma, S., "Security-aware key management in mmWave-enabled IoT using ECC," IEEE Trans. Inf. Forensics Secur., 17, 3489–3502, Nov. 2022.
7. Hu, B., Chen, W., Zhao, J., "An ECC-based lightweight authentication for mobility-aware 5G handover," IEEE Trans. Mobile Comput., 21(9), 3371–3385, Sept. 2022.
8. Mahmoudi, R., Kibaroglu, D., Yazdinejad, A., "CNN–ECC integration for UAV mmWave control in B5G," IEEE Trans. Veh. Technol., 72(5), 5183–5194, May 2023.
9. Jang, Y., Joo, K., Lee, H., "Secure V2X beamforming via deep learning and ECC handshake," IEEE Trans. Intell. Transp. Syst., 25(1), 171–183, Jan. 2024.

DOI: 10.46121/pspc.53.4.10

10. Lin, H., et al., “Intelligent reflecting surfaces for secure mmWave communication,” IEEE Wireless Commun., 29(1), 120–126, Feb. 2022.
11. Cao, Y., Ohtsuki, T., Maghsudi, S., Quek, T. Q. S., “Deep learning and image super-resolution-guided beam and power allocation for mmWave networks,” arXiv preprint, May 2023.
12. “Deep learning on camera images for fast mmWave beamforming,” arXiv preprint, Feb. 2021.
13. Roy, K. S., Sujith, M., Bhanu, B., Preethi, P., Hazarika, R. A., “FPGA-based dual-layer authentication scheme utilizing AES and ECC for unmanned aerial vehicles,” EURASIP J. Wireless Commun. Netw., 2024, doi:10.1186/s13638-024-02419-8.
14. “A lightweight ECC-based authentication and key agreement scheme with dynamic authenticated credentials,” Sensors, 2024.
15. “Securing NextG networks with physical-layer key generation: A survey,” SANDs, 2024.
16. Shimizu, T., et al., “Performance evaluation of ECC and AES on low power IoT devices,” Proc. IEEE GLOBECOM, Dec. 2021.
17. Morais, J., Behboodi, A., Pezeshki, H., Alkhateeb, A., “Position-aided beam prediction in the real world: How useful GPS locations actually are?,” Proc. IEEE ICC, 2023.
18. Mollah, M. B., Wang, H., Karim, M. A., Fang, H., “mmWave-enabled connected autonomous vehicles: A use case with V2V cooperative perception,” IEEE Network, 2023.
19. Gambo, M. L., Almulhem, A., “Zero Trust Architecture: A systematic literature review,” arXiv:2503.11659, Mar. 2025.
20. Dhiman, P., et al., “A review of intelligent Zero Trust Architecture as a security mechanism in 5G/6G networks,” 2024.
21. Lee, S., “Security system design and verification for Zero Trust,” Electronics, 14(4), Art. 643, Feb. 2025.
22. Alipour, M., “Enabling a Zero Trust Architecture in a 5G smart grid environment,” 2025.
23. Zhang, H., “Toward Zero Trust in 5G industrial internet collaboration,” 2024.
24. Ramezanpour, K., Jagannath, J., “Intelligent Zero Trust Architecture for 5G/6G networks: Principles, challenges, and the role of machine learning,” 2025.
25. Alnaim, A. K., “Adaptive Zero Trust policy management framework in 5G networks,” Mathematics, 13(9), Art. 1501, 2025.
26. Alnaim, A. K., Alwakeel, A. M., “Zero Trust strategies for cyber-physical systems in 6G networks,” Mathematics, 13(7), Art. 1108, 2025.
27. NIST, “Zero Trust Architecture,” SP 800-207, National Institute of Standards and Technology, 2020.
28. Jost, T., “Zero Trust Architecture enabled by 3GPP security,” Ericsson Blog, Feb. 17, 2025.
29. Mahmood, K., et al., “A privacy-preserving access control protocol for 6G-supported intelligent UAV networks,” Vehicular Commun., 54, 100937, 2025, doi:10.1016/j.vehcom.2025.100937.
30. Zhong, W., et al., “Image-based beam tracking with deep learning for mmWave V2I,” IEEE Trans. Intell. Transp. Syst., 2024, doi:10.1109/TITS.2024.3438875.
31. Marenco, L., et al., “Machine-learning-aided method for optimizing beam pair selection and update time,” Scientific Reports, 2024, doi:10.1038/s41598-024-70651-9.

DOI: 10.46121/pspc.53.4.10

32. Vučković, K., Hosseini, S., Rahnavard, N., “Revisiting performance metrics for multimodal mmWave beam prediction,” Proc. IEEE MILCOM, 2024.
33. Madhekwana, S., et al., “Beam alignment for mmWave and THz: Systematic review,” Telecommun. Syst., 2025, doi:10.1007/s11235-025-01318-7.
34. Fu, Y., et al., “IPO-ZTA: An intelligent policy orchestration Zero Trust Architecture for B5G/6G,” Computer Networks, 269, 111450, 2025, doi:10.1016/j.comnet.2025.111450.
35. Bojović, B., Lagén, S., “MIMO in network simulators: SU-MIMO in ns-3 5G-LENA,” arXiv:2404.17472, 2024.
36. Albuquerque, J., Klautau, A., Bojović, B., “Flexible channel model configuration for scalable 5G-LENA simulations,” Proc. ICNS3 ’25, 2025, doi:10.1145/3747204.3747211.
37. Gargari, A. A., et al., “Improving 5G-LENA: Multiple panel antenna support, Kronecker beamforming and RSRP-based attachment,” Proc. ICNS3 ’25, 2025, doi:10.1145/3747204.3747213.
38. WNS3 ’24: Proceedings of the 2024 Workshop on ns-3, ACM, 2024, doi:10.1145/3659111.
39. Li, M., Hu, S., “A lightweight ECC-based authentication and key agreement protocol with dynamic credentials,” Sensors, 2024, doi:10.3390/s24247967.
40. Sedjelmaci, H., Ansari, N., “Zero Trust Architecture empowered attack detection framework to secure 6G edge computing,” IEEE Network, 38(1), 196–202, 2024, doi:10.1109/MNET.131.2200513.
41. Chang, Q., Ma, T., Yang, W., “Low-power IoT device communication through hybrid AES-RSA encryption in MRA mode,” Scientific Reports, 2025, doi:10.1038/s41598-025-98905-0.
42. Selvi, P., Sakthivel, S., “A hybrid ECC-AES encryption framework for secure and efficient cloud-based data protection,” Scientific Reports, 2025, doi:10.1038/s41598-025-01315-5.
43. Carvalho, G., Lagén, S., “Analysis and optimizations of PMI and rank selection algorithms for 5G NR,” Simulation Modelling Practice and Theory, 144, 103162, 2025, doi:10.1016/j.simpat.2025.103162.
44. Nahar, N., Andersson, K., Schelén, O., Saguna, S., “A survey on Zero Trust Architecture: Applications and challenges of 6G networks,” IEEE Access, 12, 94753–94764, 2024.
45. Hoque, S., Aydeger, A., Zeydan, E., Liyanage, M., “A survey on distributed denial-of-service attack mitigation for 5G and beyond,” IEEE Open J. Commun. Soc., 6, 5840–5879, 2025.
46. Asad, M., Otoum, S., Ouni, B., “Zero-Trust federated learning via 6G URLLC for vehicular communications,” IEEE J. Sel. Areas Commun., 43(6), 1970–1980, June 2025.
47. Online resource, ScienceDirect article (Elsevier), available at: https://www.sciencedirect.com/science/article/abs/pii/S2542660525002008
48. Online resource, ScienceDirect article (Elsevier), available at: https://www.sciencedirect.com/science/article/abs/pii/S1389128625004177
49. Online resource, ScienceDirect article (Elsevier), available at: https://www.sciencedirect.com/science/article/pii/S2542660525000848
50. Online FAQ, “Why is tail latency (p95/p99) often more important than average latency…,” Zilliz, available at: https://zilliz.com/ai-faq/why-is-tail-latency-p95p99-often-more-important-than-average-latency-for-evaluating-the-performance-of-a-vector-search-in-userfacing-applications

DOI: 10.46121/pspc.53.4.10

51. Online article, PubMed Central (PMC), available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10892953/
52. Ramezanpour, K., “Zero Trust Architecture (slides/whitepaper),” available at: https://www.androcs.com/wp/wp-content/uploads/2023/08/RamezanpourZTA22.pdf
53. arXiv preprint, available at: https://arxiv.org/pdf/2404.15326
54. [Online article, PubMed Central (PMC), available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC10058871/

DOI: 10.46121/pspc.53.4.10