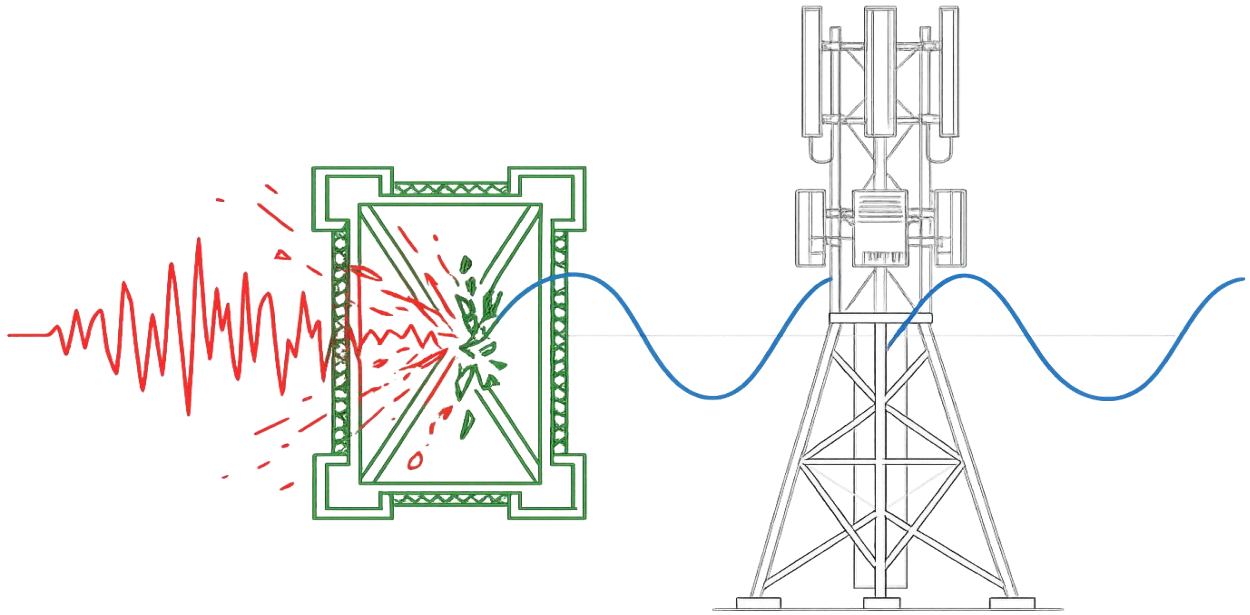
JCCI2026
The 36th Joint Conference on Communications and Information
April 15-17, 2026, Yeosu, Republic of Korea
https://jcci.kr/

2026.04.15

RAN-Aware Adversarial Training for Robust Near-RT RIC Handover Control

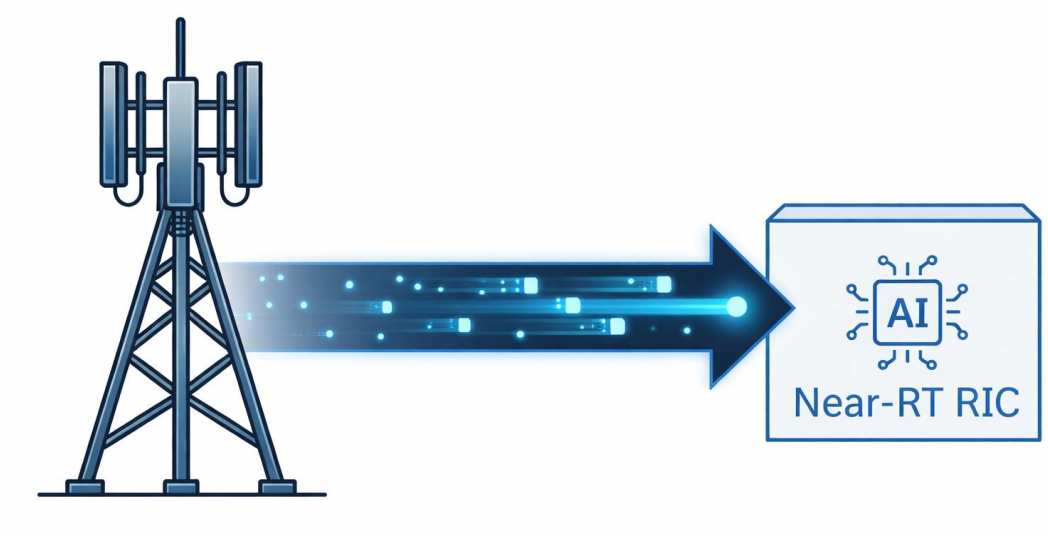
Securing AI-driven Handover Control Against Physically Plausible Attacks in O-RAN Architectures



Vaskar Chakma, and Wooyeol Choi

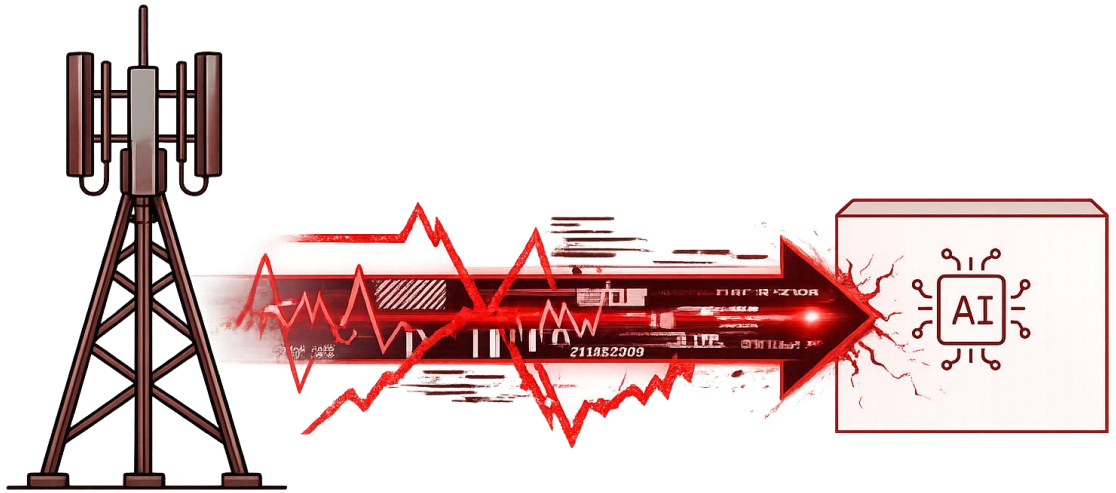
Chung-Ang University, Seoul, Republic of Korea

The AI-RAN Promise



AI-driven control applications (xApps) utilize Bi-directional LSTMs to evaluate RSRP, CQI, and mobility attributes, drastically optimizing cell selection and handover management.

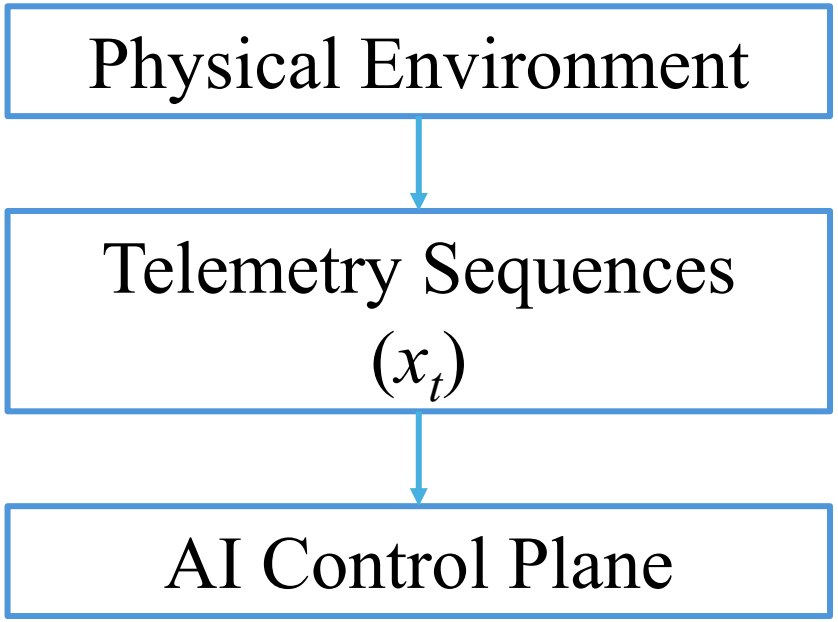
The Hidden Vulnerability



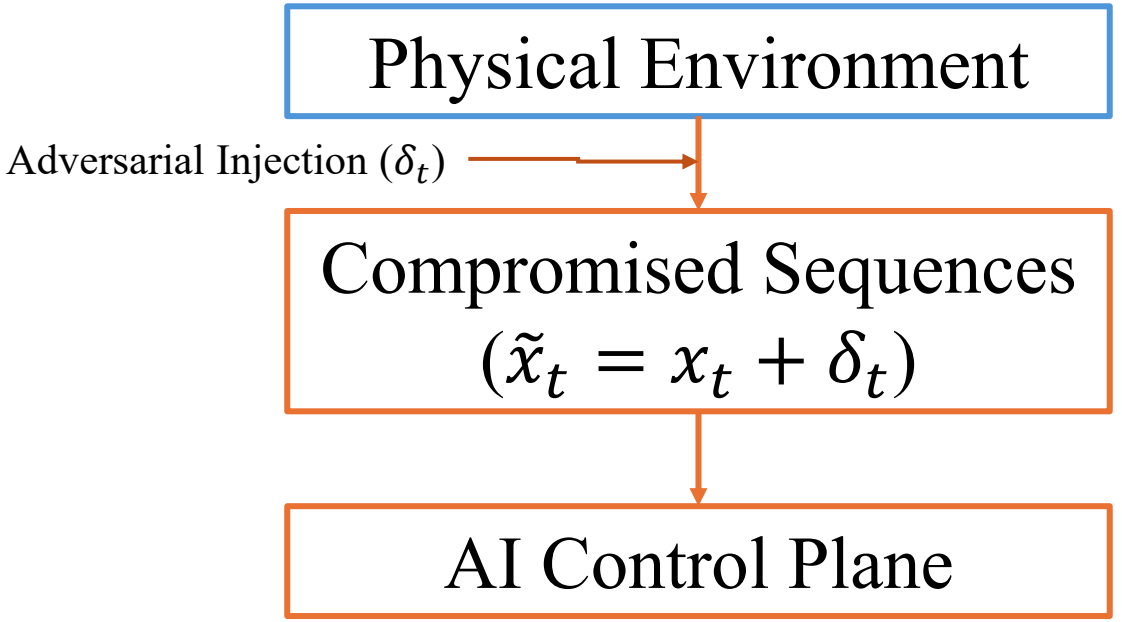
Adversarial Machine Learning (AML) attacks can manipulate resource allocation, traffic steering, and critically, handover control, by injecting stealthy perturbations into the observation sequence.

The Adversarial Blind Spot in O-RAN

The Hidden Vulnerability

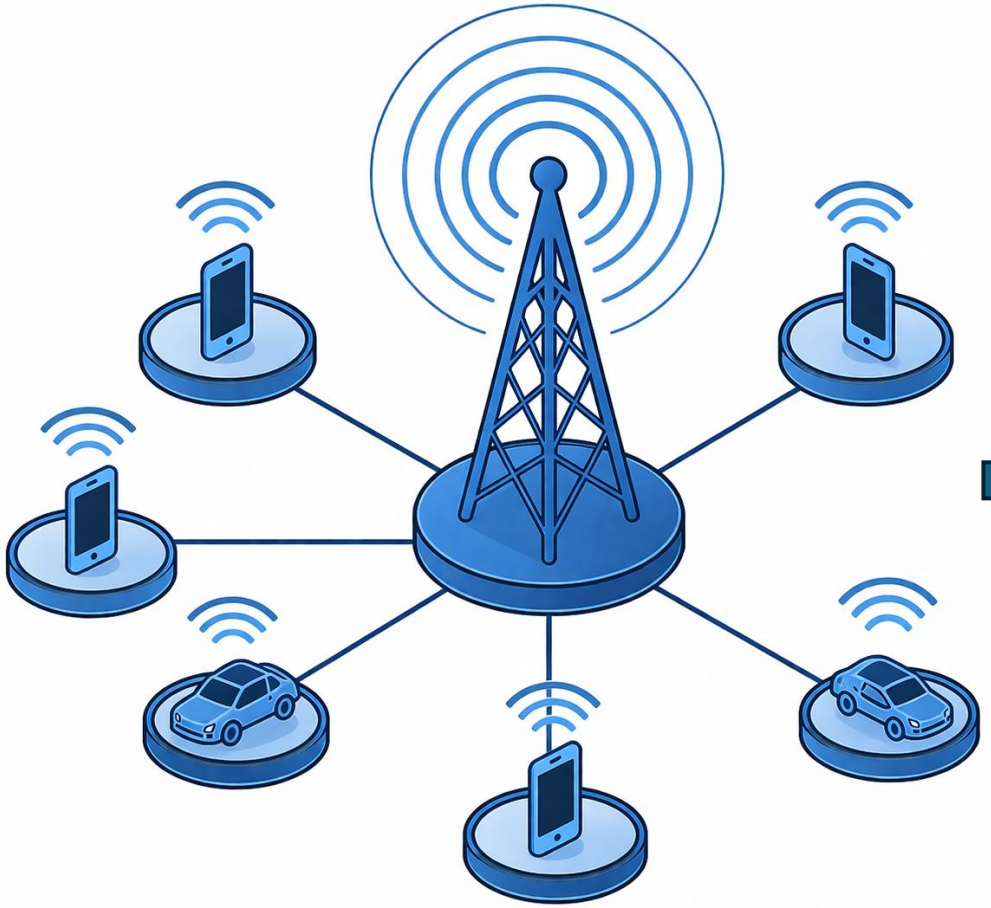


Compared Traffic (Attacked)



AI-driven control applications blindly trust RSRP and CQI telemetry. If an adversary injects micro-perturbations into these sequences, they can fundamentally alter the Bi-directional LSTM's handover policy without triggering conventional network alarms.

Architecture Phase I: The Clean Observation



Observed RAN State at time t:

$$x_t = [RSRP_t(s), RSRP_t(n_1), \ldots, RSRP_t(n_k), CQI, p_t, v_t, L_t]$$

The Near-RT RIC continuously ingests this complex observation sequence – serving cells, neighboring cells, UE velocity, and cell load- to predict the optimal handover moment.

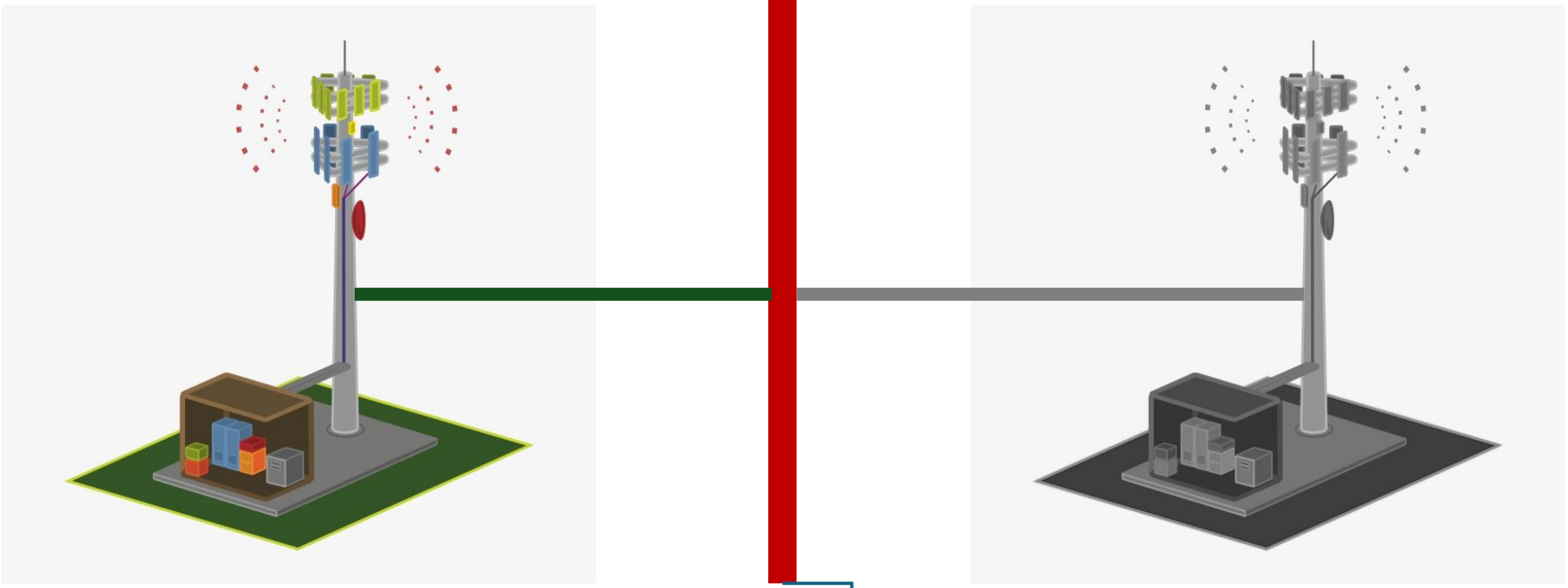
Architecture Phase II: The Physics Filter

Expanded Constraint Definition (C_{RAN})

Only physically plausible perturbations successfully bypass the physical filter and reach the AI-Based Handover Policy.

The Danger Zone: Surgical Stealth

$$RSRP_t(s) - max_k RSRP_{n_k}(t) \leq \Delta HO + \varepsilon$$



Hysteresis Border

Sparse Activation

The attack does not operate constantly. It remains dormant to avoid detection, injecting perturbations ONLY when the serving-cell margin enters this vulnerable hysteresis window.

The Robust Training Objective

The Attacker
(Inner Maximization)

Utilizes a targeted sign-gradient PGD method to search the CRAN constraint space, actively generating the absolute worst-case perturbation (δ).

$$\min_{\boldsymbol{\theta}} \mathbf{E}_{(\boldsymbol{X},\boldsymbol{y})\sim \boldsymbol{D}}[\max_{\boldsymbol{\delta}\in C_{RAN}} \mathrm{L}(\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{X}+\boldsymbol{\delta}),\boldsymbol{y})]$$

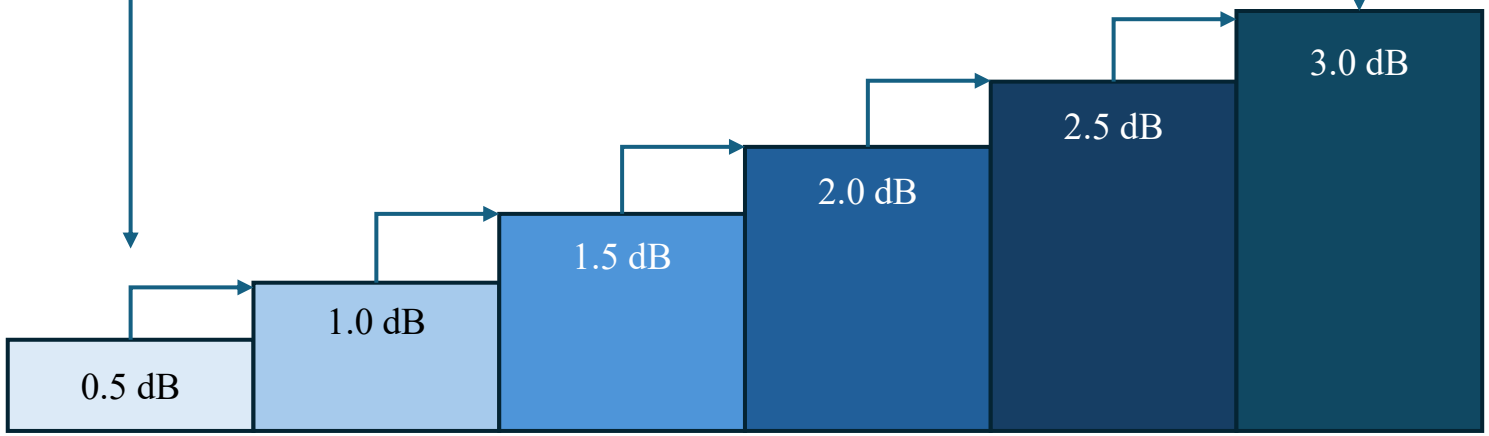
The Defender
(Outer Minimization)

The Near- RT RIC algorithm continuously optimizes its neural network weights (θ) to minimize the handover loss despite the worse-case attack.

Resulting Robust Parameter: θ*

Building Immunity: The Multi- ε Curriculum

Projected Gradient Descent (PGD):
20 ascent steps finding the absolute worst-case perturbation.



Result: A hardened model trained over 35 epochs that retains real-time inference speed for latency-sensitive operations.

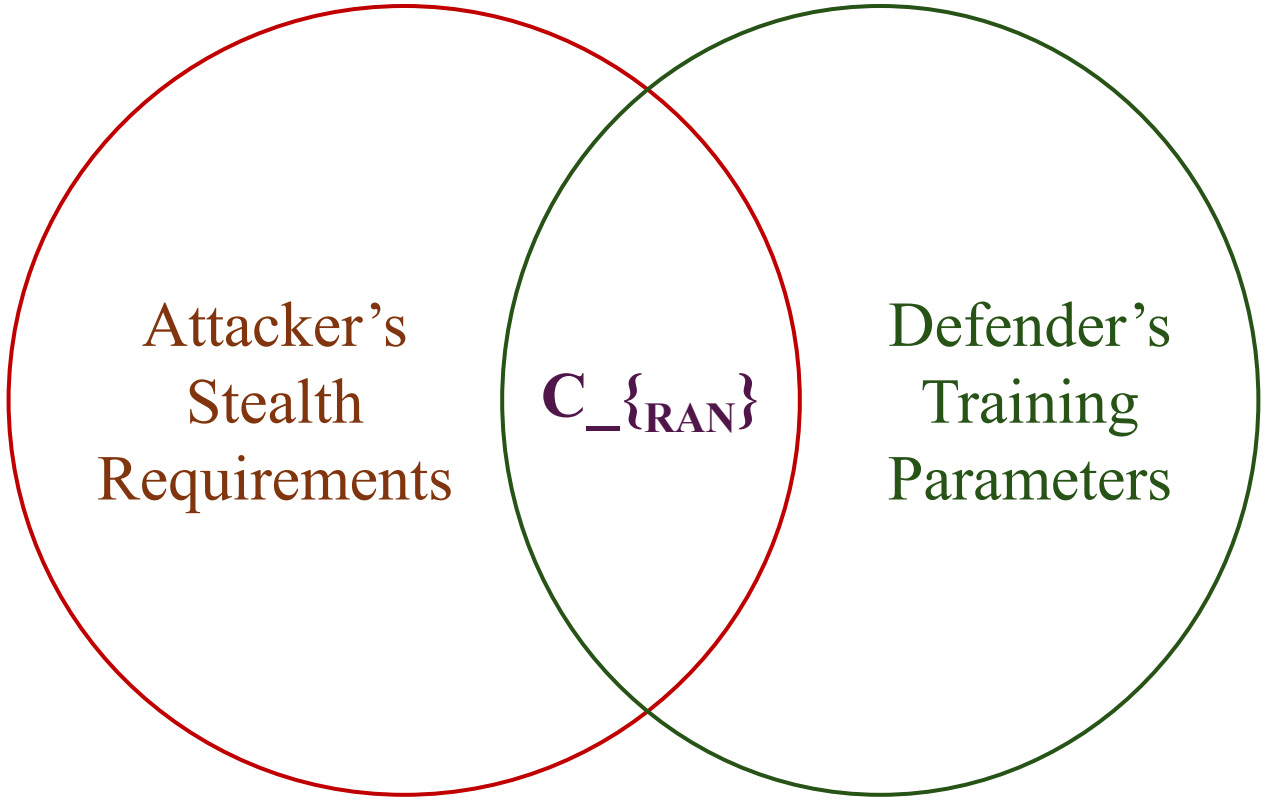
1. Historical Traces
Start with clean 3GPP Urban Macro Data (6 neighbors, UEs at 2-30 m/s)

2. PGD Injection
Generate worst-case perturbations using a 20-step targeted sign-gradient method (0.4 dB steeps).

3. Curriculum Scaling
Expose the model to progressively stronger attacks along the multi-ε staircase.

4. Projection
At every step, mathematically force the attacks to remain strictly within the physical RAN constraints.

Physics as Both Weapon and Shield

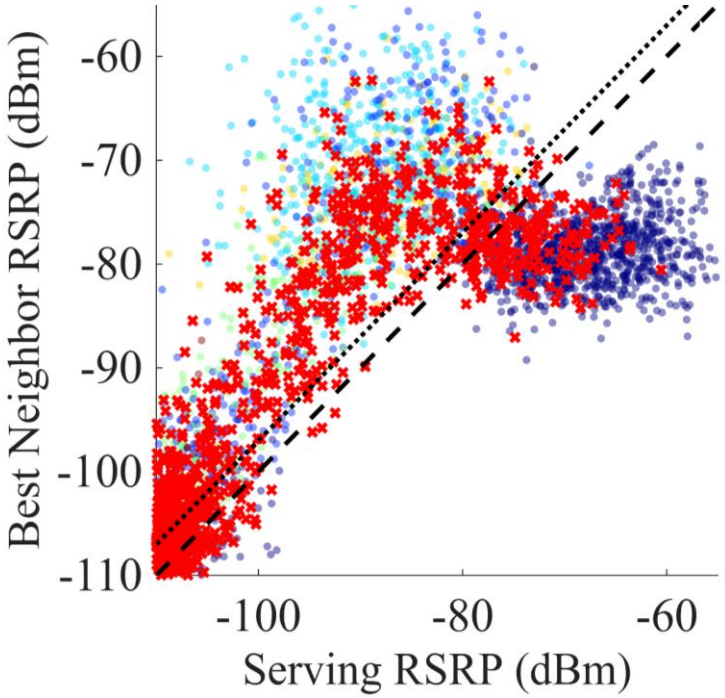


The defining insight of this study: Because the attacker must obey the physical constraints to remain undetected, the defense can map the precise, finite geometry of the attack surface. The physical limitations that allow the attack to be stealthy are the exact parameters we use to perfectly define the training curriculum.

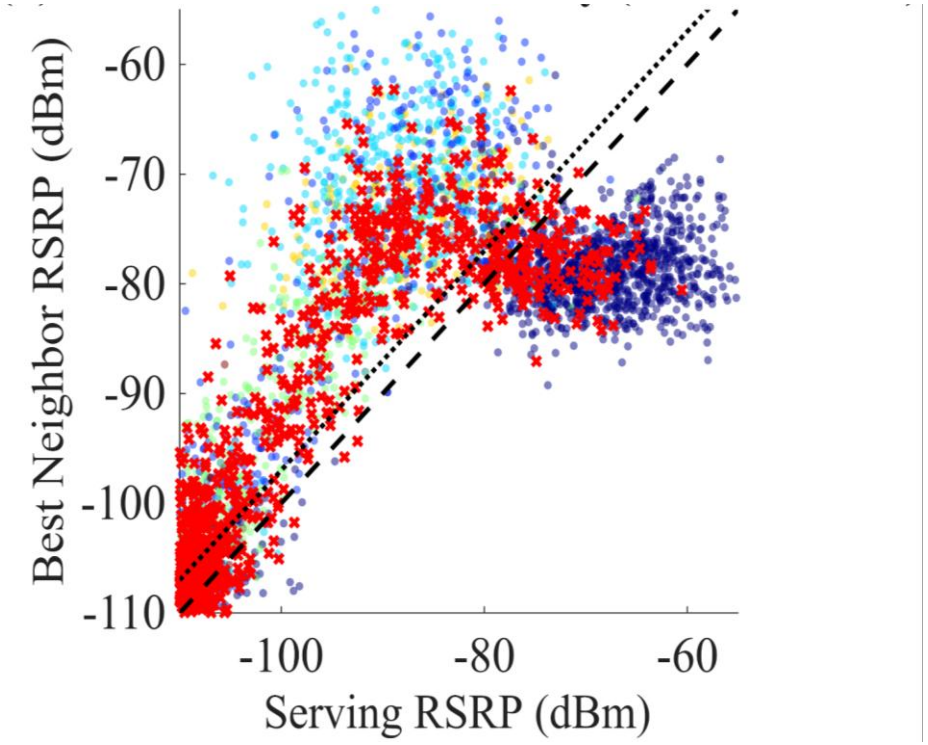
The Synthesis: Closing the Vulnerability Gap

Unprotected models collapse exactly at the hysteresis border under a 2.5 dB attack. By training the AI exclusively on physically plausible attacks, the model learns to reshape its decision boundary, creating an impenetrable shield exactly where attackers try to exploit.

Baseline Model (Under Attack)

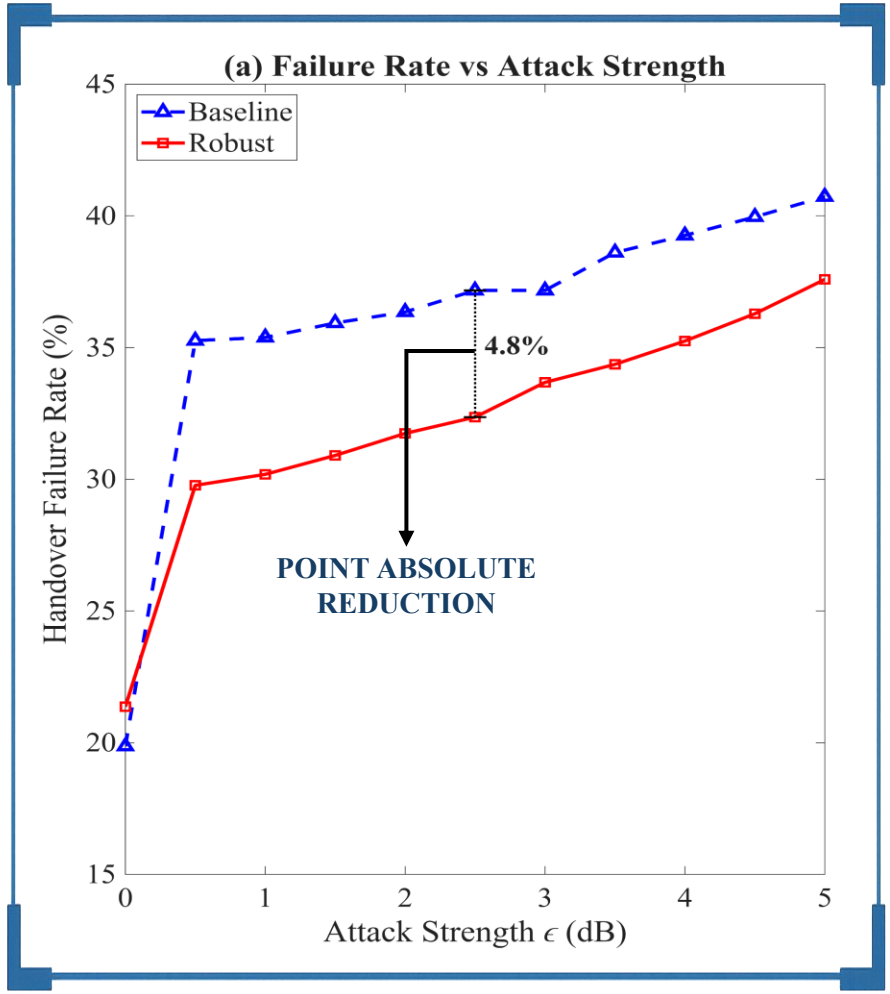


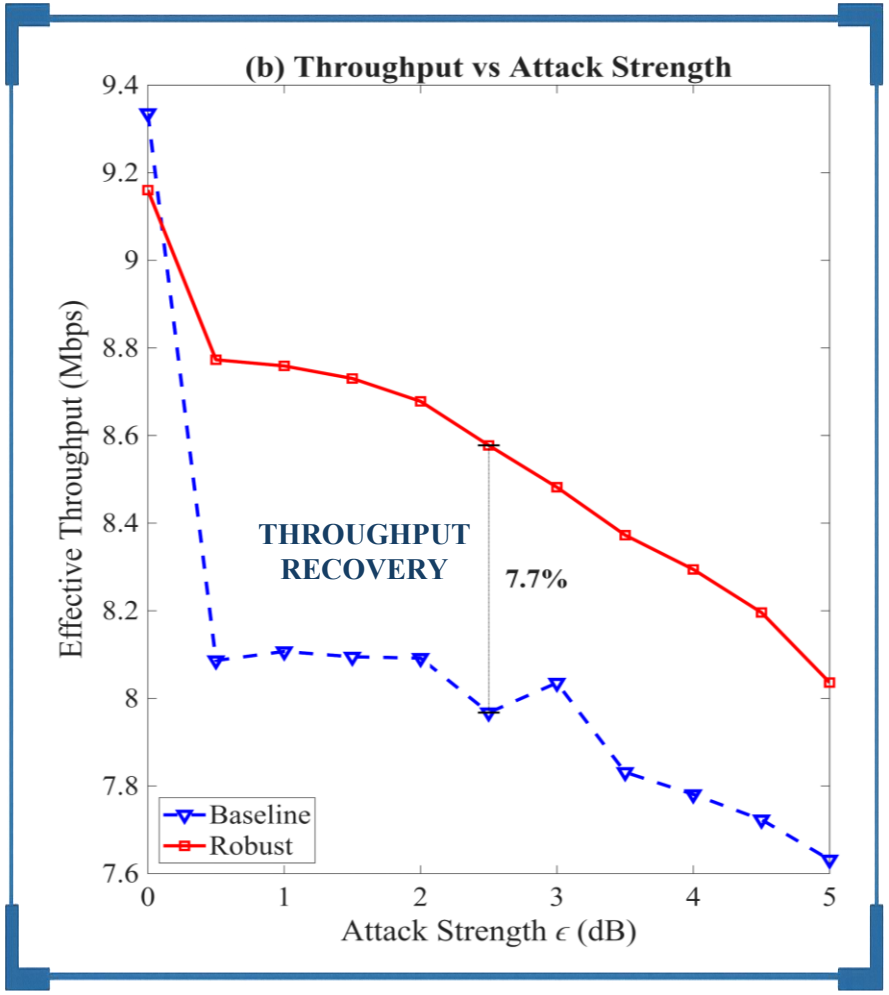
Robust Model (Under Attack)



Key Takeaway: The red × distribution is identical in both panels — same attack, same poisoned data. The difference is entirely in how each model's decision boundary responds: the baseline is misled, while the robust model maintains a more appropriate boundary that resists the attack's influence.

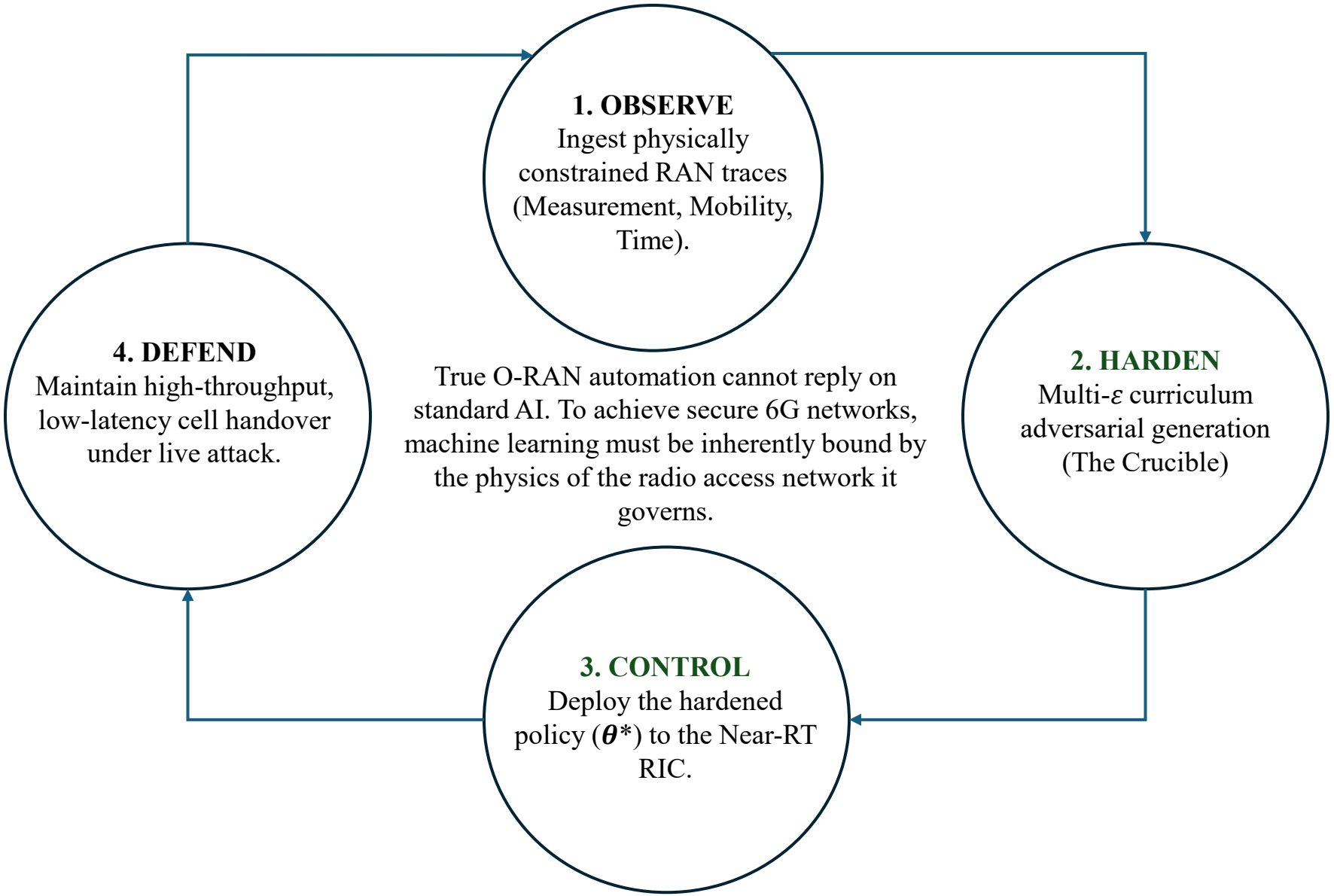
System Impact Under 2.5 dB Attack Conditions





Context: Theoretical Shannon capacity limit at -1.2 dB mean SINR is ~16.3 Mbps. Preserving 8.7 Mbps under severe targeted attack represents critical operational resilience.

The Foundation for Reliable 6G Automation



Thank You!

Intelligent Networking Lab (INL)
Department of Computer Science and Engineering
Chung-Ang University
Seoul, Republic of Korea

vaskar@cau.ac.kr