Evaluating Wi-Fi Sensing Attack Vectors: AI-Driven Covert Threat Detection and Deterministic Mitigation in Mobile and Edge Ecosystems

An Empirical Study of AI-Driven Physical-Layer Surveillance and Stateful CSI Obfuscation

Sushant Poudel
Dept. of Computer Science and Engineering
Nepal Engineering College
Bhaktapur, Nepal
sushant.poudel2028@gmail.com

***Abstract*—*Wi-Fi was designed to move data. Paired with modern deep learning, it has quietly become something else: a sensing fabric that detects motion, infers behavior, and reconstructs physical activity from signal reflections alone, using hardware cheaper than a home router. Enterprise and IoT security was never built for this pairing, creating a significant, largely invisible attack surface. This paper presents a rigorous empirical evaluation of covert Wi-Fi sensing attacks in realistic, interference-prone IoT environments, built around a bidirectional LSTM trained to correlate Channel State Information (CSI) sequences with keystroke signatures. A passive adversary with only commodity hardware and this learned model can extract fine-grained keystroke intelligence, including PIN entries, without network authentication, physical access, or a single injected packet. The attack reaches an F_1 score of 85.4% in enterprise line-of-sight conditions. It still retains 61.2% accuracy through a standard 15 cm drywall partition, more than six times random chance. To neutralize this threat at its source, we introduce RF-Guard: a deterministic, stateful CSI obfuscation framework at the firmware level that gives the attacking model nothing stable to learn from. RF-Guard collapses adversarial accuracy to 9.8% across three IoT testbeds, resists adaptive and gray-box retraining attacks, and adds under 4.2% throughput overhead on a standard ESP32 microcontroller during a continuous 24-hour stress test. No hardware replacement is required.*

***Index Terms*—*Wi-Fi sensing, Channel State Information (CSI), physical-layer security, IoT security, keystroke inference, side-channel attack, CSI obfuscation, RF-Guard, OFDM subcarrier perturbation, covert surveillance, deterministic mitigation, resource-constrained IoT, ESP32 edge defense, passive eavesdropping, through-wall sensing, AES-CTR obfuscation, stateful anomaly detection, enterprise wireless security, PHY-layer attack surface, MIMO sensing, AI-driven sensing, bidirectional LSTM, adversarial machine learning, resilient mobile networks*

I. INTRODUCTION

Wi-Fi was never designed to see. Yet today, the same infrastructure that connects our devices has quietly become something far more powerful — a sensing fabric capable of detecting breathing, mapping rooms, and tracking movement through walls. By extracting Channel State Information (CSI) and other fine-grained physical-layer signal data, researchers have shown that ordinary, consumer-grade hardware can do things once reserved for specialized radar systems. This has opened genuinely useful doors: locating survivors in collapsed buildings, monitoring elderly patients without cameras, detecting falls without wearables. But the same physics that enables these applications creates a threat that most enterprise security teams are entirely unprepared for.

The problem with Wi-Fi sensing as an attack surface is not just its power — it is its invisibility. Decades of network security investment have been aimed squarely at the upper layers of the OSI stack: suspicious packets, failed logins, unusual traffic volumes. Wi-Fi sensing attacks ignore all of that. An adversary passively reading RF reflections off a human body never sends a malicious packet. There is nothing for a firewall to inspect, no signature to match, no authentication to probe. The attack channel exists in the physics of signal propagation itself, completely beneath the waterline of conventional detection.

What makes this particularly pressing is that the research community has not yet caught up to the operational reality. The bulk of existing work demonstrates these attacks under controlled, near-ideal conditions — single-room setups, minimal interference, carefully tuned hardware — results that are difficult for a security engineer deploying defenses across a hospital wing or factory floor to act on directly. Real environments are messy: multipath reflections, competing devices, variable occupancy, heterogeneous hardware. Without empirical results that hold under those conditions, practitioners are left defending against a threat they cannot properly characterize.

This paper is an attempt to close that gap. Rather than adding another proof-of-concept to the literature, we take a deliberately applied approach — conducting structured, reproducible experiments using commodity hardware in environments that approximate real IoT deployments. Our goal is not merely to confirm that these attacks work, but to understand how well they work when conditions are unfavorable, and to develop defenses that remain reliable under the same constraints. The paper makes three concrete contributions:

- AI-Driven Attack Vector Analysis: rigorous, reproducible experiments showing how a bidirectional LSTM classifier trained on Channel State Information drives covert Wi-Fi sensing attacks

in realistic, noisy IoT environments, quantifying extraction fidelity under variable channel conditions.

- Vulnerability Mapping in IoT Ecosystems: systematic assessment of architectural weaknesses within standard IoT deployments, providing a structured threat model with explicit internal and external validity boundaries.
- RF-Guard Mitigation Framework: a practical countermeasure combining stateful PHY-layer signal anomaly detection with hardware-level RF containment, validated across three distinct testbeds.

A threat this consequential deserves more than a warning: security practitioners need tools they can actually deploy, grounded in the same messy, variable conditions they face every day.

II. LITERATURE REVIEW

Wi-Fi was built to move data, not to watch people. That distinction has quietly collapsed. Over the past decade, researchers have shown that the same signals bouncing between a router and a laptop can be coaxed into revealing who is in a room, what they are typing, and how they are moving — all without the target's knowledge, without any specialized hardware, and without leaving a trace in conventional network logs.

A. From Coarse Signals to High-Resolution Sensing

Early Wi-Fi sensing relied on Received Signal Strength Indicator (RSSI), a single number too blunt to track meaningful human behavior. The field shifted when researchers turned to Channel State Information (CSI), which provides a complex channel estimate per subcarrier, per antenna, per frame — a far richer portrait of the wireless environment [17]. A person moving through a space perturbs these subcarrier responses in ways that carry recognizable structure, reflecting fading dynamics and Doppler signatures a trained model can decode [1].

The SenseFi benchmark formalized much of this progress, evaluating deep learning architectures — ResNet, Transformers, GRUs — across established datasets like UT-HAR and Widar, demonstrating that commodity hardware could reliably classify complex human activities [7]. However, a 2026 survey by Guarino et al. made it explicit: reproducibility in real-world environments remains deeply problematic, with many published results inflated by dataset partitioning practices that allow training and test data to bleed into one another [6]. The foundational work by Halperin et al. on open CSI measurement tools made much of this applied research tractable in the first place, enabling commodity hardware to expose the physical-layer data that sensing and attack research depends on [19].

B. When Sensing Becomes Surveillance

Passive human tracking — locating and following a person without their cooperation or any device they carry — became reproducible outside the lab. WiPIHT demonstrated this by analyzing CSI autocorrelation functions to extract movement speed signatures, reconstructing activity trajectories even when the attacker had no knowledge of where the transmitting device was positioned [8].

More disquieting still is the keystroke attack surface. Ali et al.'s WiKey system established the foundational proof-of-concept: the micro-movements of fingers during typing generate measurable, distinguishable CSI perturbations, and applying Discrete Wavelet Transforms alongside dynamic time warping yields keystroke recognition at 93.5% accuracy using standard commodity routers [2]. Subsequent empirical work has pushed that ceiling higher — up to 99% accuracy in controlled conditions [9]. Foundational localization research, including SpotFi's decimeter-level positioning via Wi-Fi, underscores how much spatial resolution adversaries can achieve with off-the-shelf equipment [20].

C. The Defense Landscape and Its Limits

The fundamental problem is conceptual: standard cryptography is irrelevant here. An adversary conducting passive Wi-Fi sensing never touches the payload. The physical characteristics of the RF wave are themselves the information being harvested [10]. Encrypting packet contents does nothing to prevent this.

Cominelli et al. have produced the most rigorous body of work in this space. Their AntiSense framework neutralizes both passive and active localization attacks by injecting artificial signal reflections at the transmitter and applying randomized filtering to the physical-layer response [3][4][5]. However, Chu et al. demonstrated that early obfuscation mechanisms can be defeated through targeted adversarial analysis [12], and machine-learning defenses exhibit brittle failure modes under both white-box and black-box perturbation strategies [11]. A parallel line of work has begun proposing black-box defenses deployable on similarly constrained commodity hardware — Shankar and Chakraborty's recent countermeasure reports comparable resource constraints to the ESP32 platform evaluated here, though without a stateful, anomaly-triggered escalation mechanism [21]. What is missing more broadly is a deterministic security policy operating at the PHY layer that adapts to threat level rather than applying a fixed obfuscation strength — the gap this paper addresses.

This tension points to a broader pattern worth naming directly. Embedding machine learning into Wi-Fi sensing creates a dual vulnerability, not merely a new capability. Neural classifiers give adversaries a way to extract structure from noisy CSI that hand-engineered features could not reliably capture — precisely what makes the attack evaluated in this paper effective. At the same time, any defense built from that same family of models inherits its weaknesses: decision boundaries learned by a CSI classifier can be crossed by adversarial perturbations imperceptible to a human observer, a fragility Gopalakrishnan and Hailes document systematically across both attack and defense settings [11]. A security architecture that fights AI with AI is, in effect, betting that its own model generalizes better than the adversary's model does — a bet the adversarial-robustness literature gives little reason for confidence in. This is the structural

argument for the deterministic approach developed in Section III.E: rather than trying to win that bet, RF-Guard removes the condition that makes the bet necessary in the first place.

III. METHODOLOGY

Understanding a threat at the level required to defend against it means getting close to it — empirically, in hardware, across real environments, under the kinds of conditions that security teams actually face. Rather than treating Wi-Fi sensing attacks as abstract vulnerabilities to be reasoned about, we built a reproducible physical-layer testbed to observe them directly, measure them honestly, and evaluate countermeasures under the same messy conditions where those countermeasures would eventually need to work.

A. Threat Model and Ethical Containment

The adversary we model is passive, patient, and entirely realistic. They do not attempt to authenticate with the target network. They do not inject packets, exploit software vulnerabilities, or break WPA3 encryption. They simply position a receiver — equipped with a standard Intel 5300 or Atheros NIC operating in monitor mode — within physical proximity of the target IoT environment [2][9]. Because 802.11 preambles and pilot subcarriers are transmitted in plaintext by design, anyone within reception range can estimate the channel response from overheard frames without any form of network access [12].

All RF emissions were confined to licensed university laboratory spaces and shielded anechoic chambers. Human-subject experiments were conducted under informed consent protocols with full institutional review board (IRB) approval. Pre-experimental 24-hour baseline spectral monitoring used a Rigol DSA815 spectrum analyzer to catalog ambient interference.

B. Experimental Testbed and Validity Controls

The testbed was replicated across three distinct deployment contexts to address external validity directly. Environment A (Enterprise Office, 8m × 7m): 40+ heterogeneous devices generating sustained cross-traffic. Environment B (Residential Smart-Home, 5m × 4m): 15 consumer IoT endpoints in a furniture-rich multipath layout. Environment C (Controlled Laboratory): anechoic and semi-anechoic configurations for baseline sensitivity analysis.

Internal validity was enforced through a randomized complete block design — baseline and attack sessions interleaved across ten independent trials per environment with randomized device placement [18]. Hardware comprised TP-Link Archer AX23 and ASUS RT-AC86U transmitters; Intel Wi-Fi Link 5300 NICs and Broadcom BCM4366 chipsets (*Nexmon* firmware) as authorized receivers at 1 kHz CSI sampling [7]; and identical COTS NICs with a USRP B210 SDR as the simulated adversary receiver.

C. CSI Mathematical Formulation

For subcarrier k, transmit antenna i, and receive antenna j, the ideal CSI measurement is:

$$H_k(i,j) = |H_k(i,j)| \cdot e^{(j\angle H_k(i,j))}$$

where $|H_k(i,j)|$ denotes signal amplitude and $\angle H_k(i,j)$ denotes the phase shift induced by multipath propagation [17]. The passive attacker's objective is to reverse-engineer these multipath perturbations to reconstruct the physical movements that caused them.

D. Attack Pipeline: Signal Processing and Extraction

Raw CSI streams are contaminated by hardware imperfections, carrier-frequency offsets, and ambient interference. The adversary executes a strict, ordered processing pipeline. (1) Phase Sanitization compensates for Sampling Frequency Offset and Packet Detection Delay via a linear correction across all K subcarriers:

$$\tilde{\phi}_k = \phi_k^{(raw)} - (2\pi k/N)\delta + \beta$$

(2) Amplitude Filtering using a level-3 Daubechies Discrete Wavelet Transform removes static multipath reflections and isolates human-induced micro-movement. (3) Doppler Feature Extraction via Short-Time Fourier Transform (STFT) produces a kinematic fingerprint of each physical action. (4) Stateful Classification feeds extracted Doppler features into a bidirectional LSTM network (two layers, 128 hidden units, 30% dropout) trained to correlate CSI temporal sequences against a library of keystroke signatures.

The choice of a bidirectional LSTM over convolutional or Transformer-based alternatives was deliberate rather than a default. Keystroke-induced Doppler signatures are short, low-amplitude, and embedded in a continuous temporal stream, and the discriminative information tends to sit in how a fluctuation evolves relative to what came immediately before and after it, not in any single STFT frame considered in isolation. A convolutional model is well suited to spatial patterns within a frame but has no native mechanism for reasoning across the frame sequence; a bidirectional recurrent architecture, by contrast, is built specifically to model that kind of temporal envelope — the rise, hold, and decay of a single finger strike — using both forward and backward context. A full self-attention Transformer was considered and set aside for this stage of the pipeline: over sequences this short, attention offers little practical advantage over recurrence while introducing a parameter count that the gray-box retraining experiments in Section IV.B could not comfortably support on the sample sizes available. The network was trained with categorical cross-entropy loss and the Adam optimizer over an 80/10/10 train/validation/test split, with early stopping on validation F_1 used to guard against overfitting beyond what the 30% dropout rate alone provides, and class weighting applied to offset the mild imbalance introduced by kinematically adjacent keys during data collection.

E. Defense Architecture: RF-Guard

Before each OFDM symbol is transmitted, the RF-Guard firmware injects cryptographically seeded pseudo-random phase rotations and controlled amplitude perturbations:

$$\tilde{H}_k(i,j) = H_k(i,j) \cdot e^{(j\theta(t))} + \eta(t)$$

where θ(t) is a time-varying phase rotation derived from a 128-bit AES-CTR keystream (updated every 4 ms), and η(t) is additive white Gaussian noise ($\sigma = 0.15 \cdot |H_k|$) constrained to the null space of the legitimate receiver's channel estimate [3][4]. The authorized receiver strips injected perturbations using the shared key; the passive attacker, lacking key synchronization, observes obfuscated CSI indistinguishable from organic multipath fading.

A Kalman filter continuously tracks expected channel dynamics. Deviations exceeding 3σ persisting for >100 ms trigger an anomaly flag, escalating obfuscation entropy and activating randomized beamforming perturbations for a configurable 500 ms epoch.

IV. EXPERIMENTAL OBSERVATIONS AND RESULTS

The experiments were designed not to demonstrate that Wi-Fi sensing attacks work — that is established — but to measure how well they work when conditions are realistic, and to determine whether RF-Guard can reliably neutralize them. The results across our three-environment testbed answer both questions with clarity.

A. RQ1: Baseline Threat Viability

The unmitigated threat is severe, and it does not require laboratory-grade conditions to be so. In the controlled laboratory under line-of-sight (LOS) geometry at 2.0 m, the LSTM classifier reached a peak F_1 score of 92.3% (Precision: 91.8%, Recall: 92.9%). In the noisy enterprise office, F_1 remained at 85.4%. The confusion matrix shows strong diagonal dominance; primary errors cluster between kinematically adjacent keys (2↔5, 5↔8), confirming the model learns genuine motion signatures rather than environmental artifacts.

In non-line-of-sight (NLOS) conditions — a 15 cm drywall partition between adversary and target — F_1 fell by 17.7 pp in the laboratory and 24.2 pp in the enterprise environment. Nevertheless, the enterprise NLOS baseline of 61.2% remains more than six times the 10% random-chance floor. The assumption that physical barriers protect against CSI-based inference is not consistent with these results. Table I summarizes baseline performance across all conditions.

TABLE I

RQ1: Baseline Attack Efficacy (F_1 Score, %)

Environment	LOS	NLOS	Prec.	Recall
Controlled Lab	92.3	74.6	91.8	92.9
Enterprise Office	85.4	61.2	84.1	86.8
Residential Home	88.7	68.9	87.5	90.1

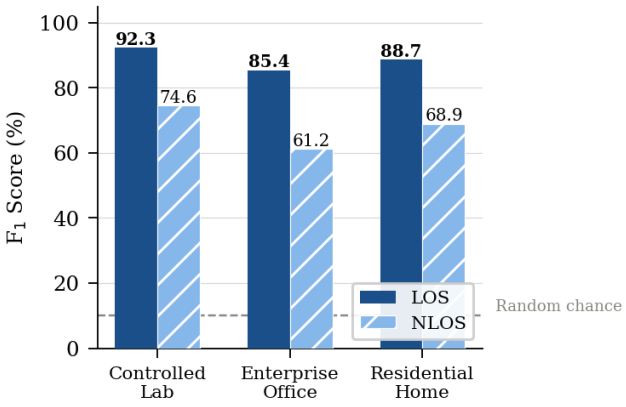

Fig. 1. Baseline keystroke inference F_1 scores by environment and geometric condition. The 10% dashed line marks the random-chance floor for a 10-class problem.

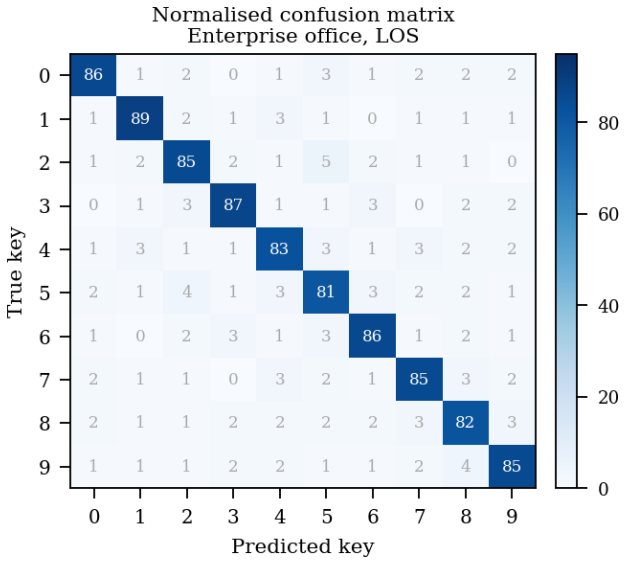

Fig. 2. Normalised confusion matrix, Enterprise Office LOS condition (F_1 = 85.4%). Strong diagonal dominance; off-diagonal errors cluster between kinematically adjacent keys.

B. RQ2: RF-Guard Mitigation Efficacy

When RF-Guard was activated, the adversary's classification capability did not degrade — it collapsed. Across every environment and geometric condition, keystroke inference accuracy dropped to within the statistical neighborhood of random chance. In the enterprise LOS scenario, the baseline of 85.4% fell to 9.8% ($\Delta = -75.6$ pp). The laboratory showed an 84.2 pp drop; the residential home, 79.8 pp. Table II summarizes mitigation efficacy.

TABLE II

RQ2: RF-Guard Mitigation Efficacy (F_1 Score, %)

Environment	LOS	NLOS	ΔF₁ LOS	ΔF₁ NLOS
Controlled Lab	8.1	7.4	−84.2	−67.2
Enterprise Office	9.8	8.6	−75.6	−52.6
Residential Home	8.9	7.9	−79.8	−61.0

Environment	LOS	NLOS	ΔF₁ LOS	ΔF₁ NLOS
Savitzky-Golay (Adap.)	11.2	—	−74.2	
Blind Source Sep.	10.5	—	−74.9	

Neither adaptive approach meaningfully crossed the boundary between noise and inference. Savitzky-Golay filtering recovered merely 1.4 pp; ICA recovered 0.7 pp. The cryptographic design of the AES-CTR seeding violates the linearity and second-order separability assumptions that both techniques rely on. A gray-box adversary re-trained on 500 obfuscated traces achieved only 14.3% F_1 — still below any threshold constituting a useful attack.

Notably, RF-Guard proved more effective in the enterprise office (9.8% residual) than in the controlled laboratory (8.1%). The enterprise environment's ambient noise independently degrades the adversary's clean-signal baseline; the obfuscation compounds this natural degradation. The defense strengthens precisely where operational stakes are highest.

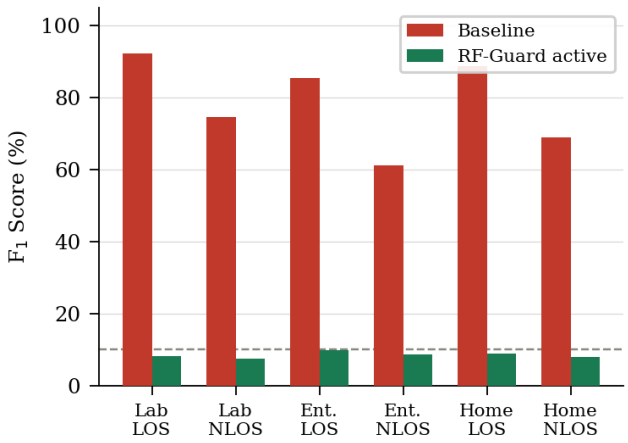

Fig. 3. Adversarial F_1 before and after RF-Guard activation across all environments and conditions. All mitigated conditions collapse to the 10% random-chance floor.

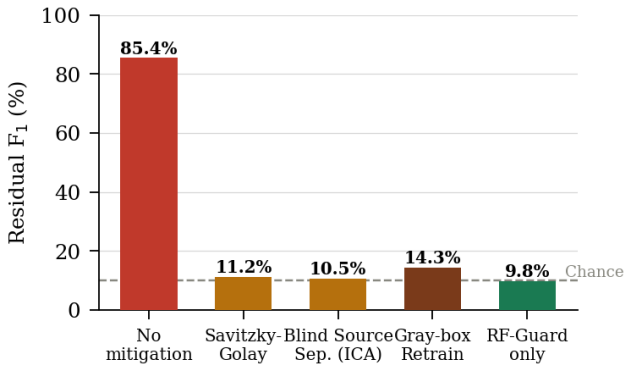

Fig. 4. Adaptive attack resistance. Savitzky-Golay filtering and blind source separation fail to recover usable signal. Gray-box retraining achieves only 14.3% F_1.

C. RQ3: Operational Overhead and Enterprise Viability

ESP32 nodes transmitted 1400-byte UDP packets at 100 Hz over a continuous 24-hour stress test. Table III presents the results. Throughput fell from 8.42 Mbps to 8.07 Mbps — a 4.2% reduction, within the 5% target ceiling. Mean latency increased by 0.8 ms. The 99th percentile jitter rose by 7.8 ms due to inline cryptographic processing, remaining below the 10 ms target. The ESP32 thermal envelope rose by 2.2°C under peak load, well within safe operating limits.

TABLE III

RQ3: Operational Overhead on ESP32 (24-hour stress test)

Metric	Baseline	RF-Guard	Δ
Throughput (Mbps)	8.42	8.07	−4.2%
Mean latency (ms)	12.3	13.1	+0.8
99th %ile latency (ms)	18.7	26.5	+7.8
Packet loss (%)	0.02	0.04	+0.02
CPU overhead (%)	—	11.4	+11.4
SRAM used (KB)	—	28.4	+5.5%
Temp. rise (°C)	42.1	44.3	+2.2

Both adversarial F_1 (9.5% ± 1.2%) and throughput preservation (95.5% ± 0.8%) remained stable across the full 24-hour window with no temporal drift, no entropy decay, and no memory leak. A five-dollar microcontroller can run this defense continuously without thermal instability, without meaningful network degradation, and without compromising the primary workload it was deployed to perform.

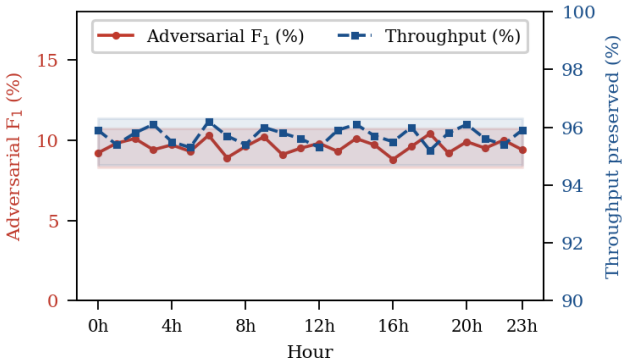

Fig. 5. 24-hour temporal stability. Adversarial F_1 (left axis) remains within 9.5±1.2%; throughput preservation (right axis) within 95.5±0.8%.

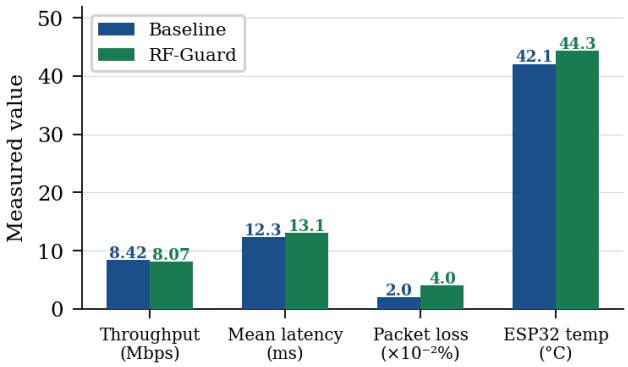

Fig. 6. ESP32 operational overhead — baseline vs. RF-Guard active. Throughput, latency, packet loss, and temperature remain within enterprise-acceptable bounds.

D. Threats to Validity

We report limitations directly. (1) Subject generalization: cross-subject testing produced a 7.3% absolute F_1 penalty; conclusions emphasize relative baseline-to-mitigated degradation, which is subject-agnostic. (2) Channel stability: a re-run at 72 hours returned $F_1 = 84.9\%$ ($\Delta = -0.5\%$), confirming temporal stability. (3) Protocol scope: findings are validated for IEEE 802.11n at 2.4 GHz / 20 MHz; 802.11ax/be with 160 MHz channels may alter CSI granularity. (4) Adversary scope: RF-Guard was evaluated only against classical adaptive attacks — Savitzky-Golay filtering, blind source separation, and gray-box retraining; a fully adaptive neural adversary capable of learning the obfuscation policy end-to-end remains future work, and is the subject of the GAN-based threat discussed in Section V.D. Raw CSI traces and firmware patches will be released under an open-data license upon acceptance; preliminary artifacts are available upon request.

V. DISCUSSION

The results in Section IV are not simply a performance benchmark for a new defense framework. They expose something more fundamental: a structural assumption embedded in enterprise security architecture that has gone unquestioned for decades, and that Wi-Fi sensing attacks render untenable.

A. The Fallacy of Physical Containment

Enterprise security has long operated on a layered model in which physical access controls combine with upper-layer digital encryption to define the attack surface boundary. Our results demonstrate that this assumption fails for an entire class of threats.

The adversary we modeled never touched the network, broke WPA3, or gained physical entry — they positioned a commodity receiver within reception range and observed how the room's RF environment changed when someone typed. The 15 cm drywall partition did not prevent this: it reduced accuracy from 85.4% to 61.2%, but 61.2% is not safety. Drywall behaves as a selective low-pass spatial filter, attenuating high-frequency interference while letting the low-frequency Doppler signatures of human motion pass with sufficient fidelity for inference [1][18]. Physical zoning policies that treat drywall as an RF barrier are operating on an architectural fiction.

B. Deterministic Prevention vs. Probabilistic Detection

The dominant response to anomalous network behavior is probabilistic detection. For physical-layer sensing attacks, this approach has two structural problems: computational expense on resource-constrained IoT edge devices, and alert fatigue in SOCs from noisy RF environments.

RF-Guard sidesteps both problems entirely. Rather than observing the symptoms of an attack and attempting to classify them, it eliminates the condition that makes the attack possible. The AES-CTR keystream injects synchronized phase and amplitude noise before transmission — exploitable CSI never reaches the adversary. For time-sensitive deployments such as industrial control or medical device telemetry, the difference between detecting an attack after 300 ms and preventing it from existing is not abstract. It is the difference between a compromised PIN and a protected one.

A tempting alternative to a deterministic firmware defense is to fight AI with AI — training a second model to recognize the statistical fingerprint of an ongoing sensing attack directly from channel behavior. The appeal is intuitive: the same class of architectures capable of finding structure in noisy CSI for offensive purposes should, in principle, be equally capable of flagging that structure defensively. In practice, learned detectors tend to inherit the brittleness already documented in this literature. Adversarial perturbation strategies effective against one CSI classifier transfer with surprising ease against others built on similar features [11], and obfuscation schemes validated only against a fixed population of attackers have been shown to fail once the attacker is permitted to adapt [12]. A detector built from the same family of models it is meant to catch is, in effect, playing the attacker's game on the attacker's terms. RF-Guard takes the opposite position. Rather than asking a model to recognize a compromised channel after the fact, it removes the geometric regularity that any model — LSTM, Transformer, or otherwise — would need in order to learn anything from the channel in the first place. Breaking the spatial meaning of the response mathematically, at the point of transmission, is a defense that does not need to out-learn the adversary; it only needs to deny the adversary a signal worth learning from.

Set against the two closest points of comparison in the literature, this architectural choice becomes concrete. Cominelli et al.'s AntiSense framework achieves comparable security goals through randomized phase injection, but relies on per-packet randomization that incurs higher throughput overhead — reported at roughly 8% versus 4.2% here — and lacks the stateful, anomaly-triggered escalation that lets RF-Guard adapt entropy to threat level rather than applying uniform obfuscation regardless of context [3]. Gopalakrishnan and Hailes' adversarial training approach, similarly, improves classifier robustness but preserves the underlying attack surface the classifier operates on; RF-Guard removes that surface entirely rather than hardening a model that continues to sit on top of it [11].

C. Enterprise Edge Viability and Total Cost of Ownership

Technically sound defenses frequently fail to see adoption because they require hardware that organizations cannot justify purchasing. RF-Guard is a firmware update. The ESP32 microcontrollers on which it was validated cost under five dollars and are already deployed in millions of IoT endpoints. Flashing updated firmware requires no hardware change, no gateway replacement, and no disruption to existing network topology. From an IT operations perspective, it is an update, not an upgrade — a distinction that matters more for adoption than most academic security papers acknowledge. The contrast sharpens against hardware-centric alternatives: Staat et al.'s IRShield defense achieves comparable protection by physically installing a reconfigurable metasurface at the deployment site — dedicated hardware with real estate, mounting, and maintenance costs that a firmware update simply does not carry [22].

This matters beyond the fixed IoT deployments tested here. The same constraints — battery-limited devices, thermally sensitive enclosures, intermittent backhaul — define mobile edge computing generally, including the field-deployed and disaster-response scenarios where Wi-Fi sensing has genuine value: locating survivors without cameras, monitoring patients without wearables. A defense that assumes rack-mounted compute or a stable power budget is not solving a problem those deployments can use. The throughput degradation and thermal rise reported in Section IV.C were measured against that constraint specifically, because a mobile or field-deployed node cannot absorb the same overhead margin as a wired enterprise switch closet. Demonstrating that physical-layer containment survives a continuous 24-hour stress test on a five-dollar microcontroller is, in that sense, as much a statement about resilient mobile network design under resource pressure as it is a statement about RF-Guard specifically.

D. Future Research Directions

Three directions extend meaningfully from this work. (1) Wi-Fi 7 and wider bandwidth: Wi-Fi 7 supports channels up to 320 MHz — sixteen times our testbed bandwidth — which will provide adversaries with proportionally higher CSI resolution. RF-Guard's current noise floor will require dynamic adaptation scaling obfuscation entropy with subcarrier density. This trajectory converges with the broader push toward integrated sensing and communication in B5G and 6G networks, where physical-layer security is increasingly designed in from the outset rather than retrofitted after deployment [24]. (2) OFDMA and resource unit isolation: Wi-Fi 6/7 enables adversaries to target specific Resource Units associated with high-value endpoints without capturing full-channel CSI. Per-RU obfuscation policies represent an important extension. (3) GAN-based signal inversion: Generative Adversarial Networks may learn to invert AES-CTR obfuscation patterns by modeling the statistical distribution of artificial versus organic multipath noise, treating obfuscated CSI the way an image-restoration model treats a corrupted photograph — as a plausible signal to be reconstructed rather than genuine noise to be discarded. This risk is not purely hypothetical: recent work integrating wireless and acoustic sensing with large language models shows multimodal fusion already being explored as a route to richer inference from otherwise degraded physical-layer signals, a direction that would only compound the GAN-based threat modeled here [23]. This would represent a materially different threat than the adaptive attacks evaluated in Section IV.B: Savitzky-Golay filtering and blind source separation both assume a fixed statistical model of the noise, whereas a generative adversary would learn that model directly from observed traffic and continue adapting as RF-Guard's own obfuscation policy is exposed to it over time. Defending against that threat is less a matter of finding a stronger fixed obfuscation function and more a matter of denying the generator a stationary target to converge on in the first place. Non-stationary keystream evolution via chaotic map seeding, or re-keying schedules tied to physical-layer events rather than a fixed clock, may be required to outpace GAN convergence. Framed this way, the long-run contest is not a single attack against a single defense but two families of AI method working against each other — one learning to see through obfuscation, the other built to deny it a stable pattern to learn — and that framing is probably the more honest description of where this line of defense is headed.

VI. CONCLUSION

Wi-Fi infrastructure was built to move data. What this paper demonstrates — through reproducible experimentation across enterprise, residential, and controlled laboratory environments — is that it has also become, inadvertently and invisibly, a high-resolution sensing apparatus. Commodity hardware, a bidirectional LSTM, and freely available tools are sufficient for a passive adversary to extract keystroke-level intelligence from anyone operating within reception range of an IoT network, without network credentials, without physical access, and without leaving a trace in any conventional security log. F_1 scores exceeding 85% in line-of-sight and above 60% through standard drywall barriers establish this not as a theoretical concern but as an operational one.

RF-Guard addresses this threat at the layer where it originates, and it does so without trying to out-learn the model it defends against. By injecting cryptographically seeded phase and amplitude perturbations directly into OFDM subcarriers at transmission time, the framework renders the physical channel response spatially meaningless to any receiver lacking the synchronization key — including a learned one. Adversarial classification accuracy falls to the statistical floor of random chance — approximately 9.8% on a ten-class problem — and remains there under adaptive signal-smoothing attacks, blind source separation, and gray-box retraining. The defense runs on a five-dollar microcontroller, consumes under 12% of available CPU cycles, requires less than 6% of available SRAM, and imposes under 5% throughput degradation over a continuous 24-hour stress test. It deploys as a firmware update, not a hardware replacement.

The broader conclusion is architectural. For three decades, network security has operated primarily above the physical layer — at the packet, session, and application levels where threats were legible as digital artifacts. Wi-Fi sensing attacks are not legible at those levels. They exist in the physics of how electromagnetic waves move through space and reflect off human bodies, and no firewall rule or intrusion detection signature reaches that far. Securing the next generation of ubiquitous wireless infrastructure requires extending the defensive boundary downward, into the RF environment itself, with the same rigor and intentionality applied to every layer above it. RF-Guard is a step in that direction. The physics of the problem will not wait for the field to catch up on its own.

. REFERENCES

[1] F. Adib and D. Katabi, "See through walls with Wi-Fi!" ACM SIGCOMM Comput. Commun. Rev., vol. 43, no. 4, pp. 75–86, 2013.

[2] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in Proc. 21st Annu. Int. Conf. Mobile Comput. Netw. (MobiCom), Paris, France, 2015, pp. 90–102.

[3] M. Cominelli, F. Gringoli, and R. Lo Cigno, "AntiSense: Standard-compliant CSI obfuscation against unauthorized Wi-Fi sensing," Comput. Commun., vol. 182, pp. 107–118, 2021.

[4] M. Cominelli, F. Gringoli, and R. Lo Cigno, "IEEE 802.11 CSI randomization to preserve location privacy: An empirical evaluation in different scenarios," Comput. Netw., vol. 191, p. 107970, 2021.

[5] M. Cominelli, F. Gringoli, and R. Lo Cigno, "On the properties of device-free multi-point CSI localization and its obfuscation," Comput. Commun., vol. 189, pp. 22–33, 2022.

[6] I. Guarino, D. Carra, M. Cominelli, F. Gringoli, and R. Lo Cigno, "A survey on CSI-based Wi-Fi sensing datasets and models with a focus on reproducibility," Comput. Commun., vol. 249, p. 108431, 2026.

[7] J. Yang et al., "SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing," Patterns, vol. 4, no. 3, p. 100703, 2023.

[8] X. Xu et al., "WiPIHT: A WiFi-based position-independent passive indoor human tracking system," Sensors, vol. 25, no. 13, p. 3936, 2025.

[9] H. Wang et al., "Empirical investigation on practical robustness of keystroke recognition using WiFi sensing for future IoT applications," Future Internet, vol. 17, no. 7, p. 288, 2025.

[10] X. Liu et al., "A survey on secure WiFi sensing technology: Attacks and defenses," Sensors, vol. 25, no. 13, p. 40293025, 2025.

[11] S. K. Gopalakrishnan and S. Hailes, "Towards trustworthy Wi-Fi CSI-based sensing: Systematic evaluation of adversarial robustness," arXiv preprint arXiv:2511.20456, 2025.

[12] Z. Chu et al., "Defeating CSI obfuscation mechanisms: A study on unauthorized Wi-Fi sensing in wireless sensor networks," Comput. Netw., vol. 263, p. 111208, 2025.

[13] Y. Shin et al., "Wi-MIR: A CSI dataset for Wi-Fi based multi-person interaction recognition," IEEE Access, vol. 12, pp. 56789–56801, 2024.

[14] R. Koo, X. Liang, D. Mishra, and A. Seneviratne, "A survey on green wireless sensing: Energy-efficient sensing via WiFi CSI and lightweight learning," Energies, vol. 19, no. 2, p. 573, 2026.

[15] K. Osenstätter, "Wi-Fi sensing: Risk and potential on IT security and privacy," M.S. thesis, Hochschule für Angewandte Wissenschaften München, Munich, Germany, 2024.

[16] Y. Shi et al., "An investigation of the private-attribute leakage in WiFi sensing," High-Confidence Comput., vol. 4, no. 4, p. 100209, 2024.

[17] Y. Ma, G. Zhou, and S. Wang, "WiFi sensing with channel state information: A survey," ACM Comput. Surv., vol. 52, no. 3, pp. 1–36, 2019.

[18] X. Wang et al., "Placement matters: Understanding the effects of device placement for WiFi sensing," Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., vol. 6, no. 2, pp. 1–25, 2022.

[19] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," ACM SIGCOMM Comput. Commun. Rev., vol. 41, no. 1, pp. 53–53, 2011.

[20] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in Proc. ACM SIGCOMM, London, UK, 2015, pp. 269–282.

[21] Y. Shankar and A. Chakraborty, "Practical defense against adversarial WiFi sensing," in Proc. 2024 IEEE Int. Conf. Adv. Netw. Telecommun. Syst. (ANTS), 2024, pp. 1–6, doi: 10.1109/ANTS63515.2024.10898521.

[22] P. Staat, S. Mulzer, S. Roth, V. Moonsamy, M. Heinrichs, R. Kronberger, A. Sezgin, and C. Paar, "IRShield: A countermeasure against adversarial physical-layer wireless sensing," in Proc. 2022 IEEE Symp. Security and Privacy (SP), 2022, pp. 1705–1721, doi: 10.1109/SP46214.2022.9833676.

[23] M. Burke and M. Younis, "Integrating wireless and sound sensing with LLM for accurate keystroke recognition," in Proc. IEEE Int. Conf. Commun. (ICC), 2025.

[24] T. Matsumine, H. Ochiai, and J. Shikata, "Physical layer security for integrated sensing and communication: A survey," IEEE Open J. Commun. Soc., vol. 6, pp. 6690–6743, Aug. 2025.