Practical Defense Against Adversarial WiFi Sensing

Yamini Shankar, Ayon Chakraborty
Department of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai, India
Email: {cs22d002, ayon}@cse.iitm.ac.in

***Abstract*—WiFi-based sensing, a non-intrusive technology that leverages existing communication infrastructure, has become widely used for environmental monitoring by extracting Channel State Information (CSI). However, the vulnerability of these systems to adversarial attacks, a core challenge within Integrated Sensing and Communication (ISAC), highlights the need for practical and robust defenses. In this work, we introduce a practical black-box defense strategy designed to protect CSI data from adversarial manipulation, significantly reducing an attacker's classification accuracy from 98% to 17% while preserving communication quality. Our approach achieves a minimal median Signal-to-Noise Ratio (SNR) difference of 1 dB, ensuring stable throughput and reliable system performance. This defense represents a crucial step forward in securing WiFi-based sensing systems, offering a resilient, low-impact solution that safeguards both sensing integrity and communication efficacy.**

I. Introduction

Recently, radio frequency (RF)-based sensing has gained significant attention from the research community due to its promising potential for joint communication and sensing tasks, as well as its ability to serve as a standalone sensing modality. This sensing leverages information related to multipath fading, which is typically estimated at a wireless receiver. For instance, the Channel State Information (CSI) available from a WiFi receiver captures fading statistics at the granularity of individual OFDM subcarriers. These fading characteristics are superimposed on the signal itself, providing valuable environmental data. By processing the CSI collected over time, one can extract physical signatures or dynamic patterns for a variety of sensing tasks, including human sensing, elderly fall detection, activity recognition, location tracking, and imaging applications [1].

While much of the scientific focus has been on enabling innovative applications through wireless sensing [1]–[3], relatively little attention has been devoted to the security risks and privacy concerns associated with such sensing modalities [4]–[6]. Although wireless sensing might appear to be beneficial and innocuous, this only presents a partial view of the picture. A malicious receiver eavesdropping on the wireless medium can observe and learn from the relevant RF signatures, potentially launching an adversarial sensing attack [7], [8]. These attacks exploit the wireless channel to infer sensitive contextual information about the surrounding environment, which is otherwise considered private and vulnerable to misuse. Our primary focus is to develop effective countermeasures against such attacks, while minimizing disruption to the underlying data communication in the network.

Existing Defence Mechanisms. Current defense strategies generally involve perturbing the transmitted signal in a way that the CSI observed at the receiver (from the attacker's perspective) serves as an *adversarial sample* for the attacker's sensing model (details are discussed in Section 2). A related strategy involves using Intelligent Reflecting Surfaces (IRS) [9] to artificially alter the channel state. However, we focus here on perturbations at the transmitter side. It is important to note that such perturbations must be carefully crafted, as they can unintentionally affect channel estimation, thereby degrading communication performance for legitimate receivers within the network. For example, depending on the extent of the perturbation, the receiver may switch to a lower MCS (modulation-and-coding-scheme) such as 16-QAM, where it would have otherwise used 64-QAM. The throughput of the link is also influenced by the Signal-to-Noise Ratio (SNR), which can be estimated using the CSI, as shown in Equation4. Most of the existing defense methods ([8], [10], [11]) focus on minimizing the sensing accuracy for the attacker, often without considering the communication penalties incurred by the perturbation. These methods primarily manipulate the CSI estimated at the receiver, creating adversarial samples by perturbing it. Some examples include adding zero-mean Gaussian noise to each OFDM subcarrier amplitude [10], or zeroing out specific subcarrier frequencies deemed important for sensing based on PCA [11], among others. Recent work has demonstrated the effectiveness of adversarial machine learning techniques, such as the *Fast Gradient Sign Method* (FGSM) and *Projected Gradient Descent Method* (PGDM), for generating adversarial perturbations in CSI [8] and for deep learning-based attacks like those in [12]. In particular, [13] introduces perturbations that are applied selectively to priority pixels, determined using Class Activation Mapping (CAM), to minimize the impact on the original signal while maximizing the attack's effectiveness. Similarly, [14] takes an optimization approach to perturbation generation but requires access to the model, though it presents transfer learning-based black-box results minimally. While above mentioned techniques effectively create adversarial samples with *minimal* perturbations to the original data, their practical applicability in adversarial settings is limited. These methods, often referred to as *white-box* attacks, assume full access to the internals of the attacker's sensing model (e.g., neural network architecture, trained weights, and gradients), which is unrealistic in real-world scenarios.

We argue that such assumptions are not practical, as they overlook the challenges posed by unknown or black-box sensing models used by attackers. Therefore, developing practical countermeasures against adversarial sensing attacks requires strategies that do not rely on privileged access to the attacker's model, ensuring robustness against a wide range of adversarial conditions. Based on the aforementioned background, in this work, we propose a *blackbox* solution towards defending against adversarial WiFi sensing attacks. We formulate the CSI perturbation as an optimization problem so as to decrease the inference accuracy of the attacker's model while making minimal changes to the CSI (which also roughly preserves the SNR). Unlike methods based on FGSM or PGDM, we do not assume access to the model's internal parameters. We validate our approach on a real testbed where we defend against an adversarial receiver that attempts to localize itself in the region of interest using estimated CSI instances. Additionally, using simulation studies, we show that our method is robust against throughput drops caused by the SNR fluctuations due to perturbation of CSI. To our knowledge, this is the first work which addresses the SNR aspect of integrated communication and sensing problem.

II. System Model

We consider a system schematic as shown in Figure 1. The system represents a generic WiFi communication network consisting of an access-point (AP) and legitimate receiver clients (CL). The CLs have access to the downlink CSI that are leveraged to perform sensing tasks or train models. In addition, there are eavesdropper nodes (ED) that have access to the wireless medium and can use their respective CSI estimates to perform adversarial sensing.

Consider $\mathbf{R}$ to be the OFDM signal at the receiver location for a transmitted signal $\mathbf{S}$, where the estimated channel state due to the physical environment is denoted by $\mathbf{H_{env}}$ and η denotes noise, i.e., $\mathbf{R} = \mathbf{H_{env}}\mathbf{S} + \eta$. The adversarial agent ED can estimate $\mathbf{H_{env}}$ at its location and either has access to or can train a model $\mathbf{M_{adv}}$ to infer the physical context surrounding the channel. To protect the network from adversarial sensing, we propose to distort/perturb the estimated $\mathbf{H_{env}}$ such that, its inference model, $\mathbf{M_{adv}}$ has substantially reduced accuracy and communication at CL sites (e.g., channel equalization, MCS choice) are not reasonably degraded.

A. CSI Perturbation - Defense Against Adversarial Sensing

The perturbations are generally modeled using a finite impulse response filter, $\mathbf{FIR_{spoof}}$. In time-domain, the transmitted signal can be estimated by $\mathbf{FIR_{spoof}}$ convoluted over the original. In frequency domain, this manifests as an artificial channel response $\mathbf{H_{spoof}}$ bundled with the transmitted signal as $\mathbf{H_{spoof}S}$. Hence, the received signal at receiver $\mathbf{i}$,

$$\mathbf{R^i} = \mathbf{H^i_{env}}(\mathbf{H_{spoof}S}) + \eta \tag{1}$$

Effectively, the receiver node estimates the CSI to be $\mathbf{H^i_{per}} = \mathbf{H^i_{env}}\mathbf{H_{spoof}}$, where $\mathbf{H^i_{env}}$ is the original channel state at the receiver $\mathbf{i}$. In this work, we focus on understanding how $\mathbf{H_{per}}$ can be used to attack the adversarial model $\mathbf{M_{adv}}$ while approximately maintaining the signal's SNR, i.e., $|\mathbf{H_{per}}| \sim |\mathbf{H_{env}}| \leq \varepsilon$, where ε is the magnitude of a small perturbation.

Our experimental evaluations (§ IV) are based on a SISO case, however, the principles apply to MIMO channels as well. Further, although the CSI is a complex-valued vector, for $\mathbf{H}$ we only consider its magnitude, i.e., $\mathbf{H} \in \mathbb{R}^{1\times N}$ for N WiFi (OFDM) sub carrier frequencies. Second, we restrict ourselves *only* to the methodology of perturbation and its effect on the adversarial sensing model – implementation details (eqn. 1) on real WiFi hardware is beyond the scope of this letter.

III. Proposed Method

For distorting the CSI, we introduce perturbations at the transmitter itself, directly into the digital signal at the final stage of the OFDM block (precisely after the IFFT operation) before passing on to the analog processing chain.
Given a CSI sample $\mathbf{H} \in \mathbb{R}^{1\times N}$, our task is to construct another sample $\mathbf{H_{per}} = \mathbf{H} + \delta$, where $\delta \in \mathbb{R}^{1\times N}$ is amount of perturbation we introduce with the following constraints. First, the accuracy of the inference model $\mathbf{M_{adv}}$ must be diminished. For instance, if $\mathbf{M_{adv}}(\mathbf{H})$ be the predicted inference class for input $\mathbf{H}$, on an average $\mathbf{M_{adv}}(\mathbf{H})$ and $\mathbf{M_{adv}}(\mathbf{H_{per}})$ should create different outputs. Second, $\mathbf{H}$ and $\mathbf{H_{per}}$ must result in similar communication performance. For instance, the MCS chosen for communication must not change drastically with a perturbed CSI.

A. Exisitng Defense with Whitebox Approach

Before moving on to our proposed method, to calibrate the reader, we present a sketch of the existing *whitebox* technique (FGSM) that is primarily used in majority of the adversarial WiFi sensing literature. For simplicity, consider a linear classifier model $\mathbf{M_{adv}} : \mathbf{w}^\top\mathbf{H}$, where $\mathbf{w}$ is the weight matrix and $\mathbf{H}$ the CSI. On passing $\mathbf{H_{per}}$ to the model, the logits take the form:

$$\mathbf{w}^\top\mathbf{H_{per}} = \mathbf{w}^\top\mathbf{H} + \mathbf{w}^\top\delta \tag{2}$$

The crucial idea here is that, even if δ is small enough such that $\mathbf{H} \approx \mathbf{H_{per}}$, the factor $\mathbf{w}^\top\delta$ is what affects the model's logits and hence the accuracy. As we want the adversarial sensing model $\mathbf{M_{adv}}$ to be unsuccessful, as a defensive mechanism $\mathbf{w}^\top\delta$ should drive the model towards an incorrect prediction. In theory, it is possible to create a strong adversarial sample even with a negligible perturbation δ, dictated by the average magnitude of the weights in $\mathbf{w}$ as well as its dimension [15]. Now, to estimate δ, FGSM adopts an innovative strategy - it computes the gradient over the model's loss function with respect to the input ($\mathbf{H}$), however steps in the direction *opposite* to the gradient, thereby *maximizing* the model's loss (in contrast, recall that the gradient descent used while training a model *minimizes* such loss moving *in the direction* of the gradient). The perturbed sample, as estimated by FGSM, can be expressed as,

$$\mathbf{H_{per}} = \mathbf{H} + \varepsilon.\mathbf{sign}(\nabla_{\mathbf{H}}(\mathbf{L}(\mathbf{w},\mathbf{H},\mathbf{Y}))) \tag{3}$$

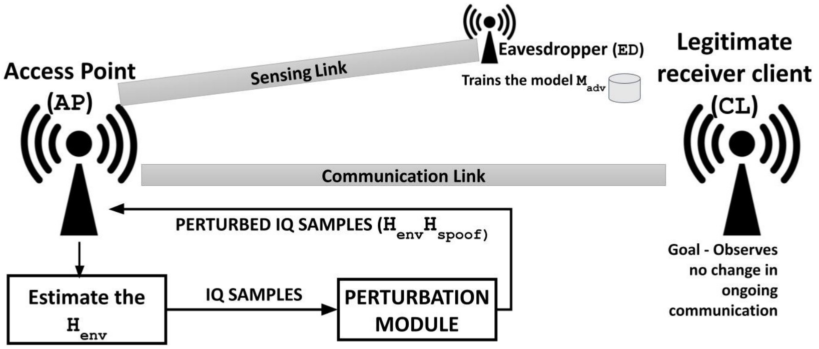

Fig. 1. Schematic of our system framework. As the transmitter infuses the original $\mathbf{H_{env}}$ with $\mathbf{H_{spoof}}$ precisely after the IFFT operation to transmit the spoofed time domain signals.

$\mathbf{L}$ is the loss function, ε is a control parameter determining the magnitude of perturbation and $\mathbf{Y}$ denotes the corresponding true inference for input $\mathbf{H}$.

B. Practical Defense with Blackbox Approach

Being a *whitebox* approach, FGSM assumes access to the internals of the adversarial model including its weights. In our setting, such assumption is impractical. Also the technique does not explicitly attempt to restrict the modulation and coding scheme (MCS) switching by preserving the signal's Signal-to-Noise Ratio (SNR).

Why is SNR Important? SNR is a critical factor in wireless communication because it directly influences the Modulation and Coding Scheme (MCS) used by the transmitter. Higher SNR values allow the use of higher MCS levels, which increases the number of bits sent per symbol, effectively raising the total data rate. Conversely, lower SNR values require lower MCS levels to maintain reliable communication, reducing the number of bits per symbol and thus decreasing the achievable data rate. Our method stabilizes the SNR while introducing controlled perturbations. This ensures that the adversary's inference model is disrupted, without degrading the communication quality or link reliability between the Access Point (AP) and legitimate users.

We illustrate how different SNR ranges affect the modulation type, coding rate, and resulting data rate in Table I [16]. This table highlights how maintaining a high SNR affects the MCS levels to be used for ongoing communication. High MCS levels imply maximizing data throughput, while lower SNR levels imply a reduction in data rate to preserve communication reliability.

In our proposed method, we do not assume any knowledge on the internal model parameters. The only information that we assume access to are the inferences drawn by the model, typical of a blackbox technique. In a nutshell, we formulate the CSI perturbation as an optimization problem, where we attempt to minimize the SNR difference between the original and the perturbed version. We also introduce a loss term to keep track of the model's failure rate (without accessing the model's internal weights).

SNR Estimation. To address the SNR constraint, i.e., making minimal effective changes to the SNR, we estimate a metric, shown below, based on the CSI ($\mathbf{H}$) that correlates highly with the signal's true SNR. Let $\mathbf{f_{snr}}(\mathbf{H})$ be such an estimator. We use this to estimate the change in SNR as a result of the perturbation, i.e., $\mathbf{f_{snr}}(\mathbf{H}+\delta)$. Specifically, we use,

$$\mathbf{f_{snr}}(\mathbf{H}) = -10\log_{10}(\sum_{n=1}^{N}|h_n|^2) + C \tag{4}$$

In the above equation, $|h_n|$ represents the gain magnitude at the n^{th} OFDM subcarrier, where $\mathbf{H}$ = $[h_1, h_2, \cdots h_N]$. The offset C is learned from the multitude of packets historically received and automatically takes care of artifacts like automatic gain control, noise floor etc. Additionally, on using neural network models for $\mathbf{f_{snr}}$ we observe marginal improvement in the estimation accuracy, however the inference delay (few 10s of ms) was prohibitive to adopt in a real-time system.

Adversarial Sample Generation. In this section, we formalize the process of generating an adversarial sample as an optimization problem. Given an original CSI sample, represented by $\mathbf{H}$, and a desired target output, $\mathbf{Y_t}$ (distinct from the original model output, $\mathbf{Y}$), we aim to find a perturbation δ that modifies $\mathbf{H}$ to approximate $\mathbf{Y_t}$, while minimally affecting signal properties critical for transmission quality, such as SNR. We pose this as the following optimization problem:

$$\arg\min_{\delta} |\mathbf{f_{snr}}(\mathbf{H}) - \mathbf{f_{snr}}(\mathbf{H}+\delta)| + \lambda \cdot \mathbf{L_{adv}}(\mathbf{H}+\delta, \mathbf{Y_t}) \tag{5}$$

In Equation 5, $\mathbf{L_{adv}}(.)$ represents the adversarial loss function, designed to guide $\mathbf{H}+\delta$ towards being misclassified as the target class $\mathbf{Y_t}$. The hyperparameter λ balances the preservation of the signal-to-noise ratio (SNR) with the strength of the adversarial perturbation. The first term in the minimization problem ensures that the SNR difference between the original and perturbed sample is minimal, preserving the transmission quality and reducing the likelihood of detection. For the adversarial loss $\mathbf{L_{adv}}(.)$, we use the *Cross-Entropy Loss*, which operates solely on the model's output without requiring access

TABLE I
SNR VS. MCS INDEX

SNR (dB) Range	MCS Index	Modulation Type	Coding Rate	Data Rate (Mbps)
< 5	0	BPSK	1/2	Low
5 – 10	1	QPSK	1/2	Moderate
10 – 15	2	QPSK	3/4	Moderate
15 – 20	3	16-QAM	1/2	High
20 – 25	4	16-QAM	3/4	High
25 – 30	5	64-QAM	2/3	Very High
30 – 35	6	64-QAM	3/4	Very High
> 35	7	64-QAM	5/6	Maximum

to its internal weights or parameters. This loss function is given by:

$$\mathbf{L_{adv}}(H+\delta, Y_t) = -\sum_c \mathbf{Y_t}^{(c)} \log\left(\mathbf{f}(H+\delta)^{(c)}\right) \tag{6}$$

where $\mathbf{f}(H+\delta)^{(c)}$ represents the output probability of class c for the perturbed sample $\mathbf{H}+\delta$. This formulation allows the perturbation δ to be optimized, yielding an adversarial sample that can induce misclassification with minimal impact on the underlying transmission characteristics, as shown in Equation 4.

IV. Testbed and Empirical Evaluation

This section describes our experimental setup and data collection methodology for evaluating the proposed black-box defense mechanism against an adversarial eavesdropper, referred to as ED. We assume ED is equipped with a pretrained localization model, $\mathbf{M_{adv}}$, boasting an accuracy of 98% when tested on unperturbed data. Our defense is compared with two baseline perturbation techniques: FGSM [8], a common adversarial attack method, and Gaussian noise as a naive noise addition approach.

A. Arena Description

Our testbed is a controlled indoor localization environment measuring 13.5 × 11 meters, specifically structured to facilitate reliable CSI data collection. The arena layout includes 100 distinct positions, chosen to provide spatial diversity for consistent evaluation of our defense method. The controlled environment allows for reproducibility in experiments and enables precise localization by both the transmitter and receiver devices.

B. Dataset Description

For dataset creation, we collect Channel State Information (CSI) data from 100 designated locations within the arena. The data collection setup employs two ESP32 devices [17], which are low-cost, power-efficient microcontrollers with built-in Wi-Fi and Bluetooth capabilities, commonly used in IoT and wireless sensing applications. One ESP32 device acts as a static transmitter, fixed in position, while the other serves as a dynamic receiver, systematically moving across specified positions within the arena to capture CSI data signatures.

Each CSI entry consists of 64 complex IQ values (In-phase and Quadrature components), representing the signal properties across 64 subcarriers, as defined by the WiFi 802.11 standard. This high-resolution CSI data provides detailed insights into the multipath environment, essential for accurate localization and sensing.

Instead of using a traditional train-test split, we collect two temporally distinct datasets—one for training and another for testing the eavesdropper's model (ED)—to ensure temporal diversity and reduce potential data leakage. Each dataset (train/test) includes approximately 3000 samples per location, providing a robust basis for accurate and unbiased evaluation of the adversarial model $\mathbf{M_{adv}}$ under different perturbation strategies. This setup enables reliable testing of the proposed defense mechanism in realistic conditions while maintaining control over environmental variables.

C. Model Training

To assess the effectiveness of the proposed perturbation technique, we simulate ED's localization capabilities by training a multilayer perceptron (MLP) model on the training dataset, using PyTorch [18]. The model architecture consists of four fully connected layers with ReLU activations [19]. Each sample from the training dataset represents one of the 100 locations, and the model is trained exclusively on this dataset, achieving high accuracy to mimic ED's optimal localization performance in an unperturbed setting.

D. Defense Evaluation

The evaluation involves applying the proposed perturbation technique to the CSI samples in the test dataset and measuring its impact on the adversarial model $\mathbf{M_{adv}}$'s localization accuracy. We compare our method's effectiveness with FGSM-based perturbations and Gaussian noise addition, assessing how each technique degrades the localization accuracy of $\mathbf{M_{adv}}$. For each perturbation method, we report key performance metrics, including the adversary's post-perturbation localization error, signal quality (SNR), and trade-off in communication reliability, i.e., the throughput.

E. Results

1) SNR Variations Across Perturbation Methods: We first evaluate the Signal-to-Noise Ratio (SNR) variations introduced by different perturbation methods. The cumulative distribution function (CDF) of SNR differences for each perturbation method is shown in Figure 2. As observed, the FGSM technique causes minimal SNR alteration, making it effective for preserving transmission quality, while our proposed defense also achieves a median SNR change of less than 1 dB. This balance of maintaining low SNR impact is essential, as large deviations can disrupt primary communication objectives.

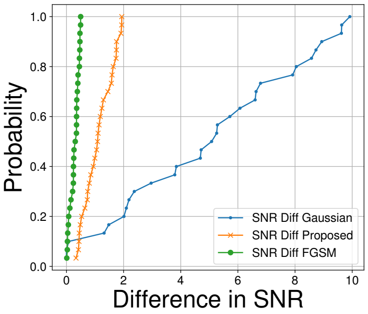

Fig. 2. CDF of difference in SNR by different perturbation techniques.

By contrast, Gaussian noise introduces significant SNR variations, which, while effective in disrupting eavesdropping, undermines the primary communication link quality. This result demonstrates that FGSM performs well under a full knowledge assumption of the adversarial model, which may be impractical. Our defense, by limiting SNR variation without assuming access to the adversary's model, is therefore well-suited for real-world scenarios where stealth is critical.

2) Accuracy and Throughput Impact of Perturbations: To assess the effectiveness of each perturbation in reducing the accuracy of $\mathbf{M_{adv}}$ and its impact on link throughput, we conducted a Monte Carlo analysis with 1000 randomly selected samples. Figure 3 shows the trade-off between model accuracy and effective throughput under each perturbation method. FGSM is shown to best preserve throughput while applying perturbations, an advantageous feature for network performance. However, our method achieves a more pronounced drop in adversarial model accuracy (median accuracy of 0.17) with an acceptable SNR difference (median of 1.8 dB), positioning it as a favorable choice for adversarial defense in real-world blackbox scenarios. The figure indicates that our approach achieves an effective balance, reducing adversarial model accuracy significantly while keeping SNR impacts manageable. This supports the idea that our blackbox defense is more practical for real-time applications where full adversarial model knowledge is unavailable.

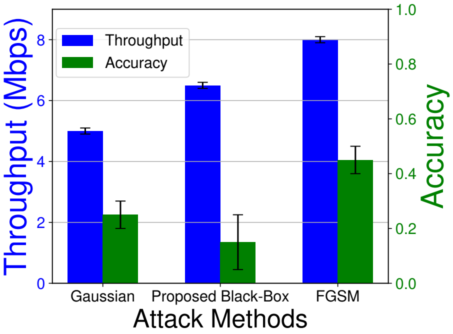

Fig. 3. The figure above shows how accuracy and effective throughput are affected with the perturbation techniques.

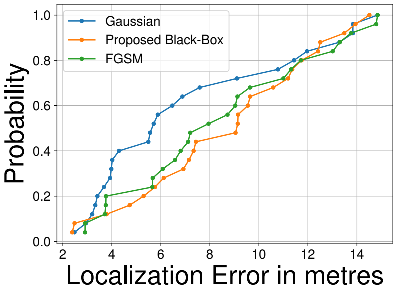

Fig. 4. The figure above shows the localization error by the different perturbation techniques.

3) Localization Error Analysis: In addition to accuracy, we measure the localization error introduced by each perturbation technique. Figure 4 illustrates the localization error across methods, showing that both FGSM and our proposed defense yield a median localization error of approximately 7 meters. This result is noteworthy because, despite the different approaches, both FGSM and our method achieve similar localization error rates. This suggests that our defense maintains spatial fidelity within an acceptable range, even without model knowledge. These findings underscore that our proposed defense introduces sufficient localization error to mislead the eavesdropper while preserving communication quality. It demonstrates that our technique provides a practical balance, safeguarding against eavesdropping without relying on model-specific information.

V. Conclusion

This study presents a robust blackbox defense strategy for perturbing CSI samples, designed to counter adversarial attacks on WiFi-based sensing systems without compromising communication quality. Our experimental results demonstrate the effectiveness of our approach, achieving a median SNR difference of only 1 dB (Figure 2), a dramatic reduction in the adversarial model's classification accuracy from 98% to 17% (Figure 3), and a median localization error of 7 meters (Figure 4), comparable to the best-performing techniques. These outcomes highlight our method's capacity to protect against eavesdropping while preserving the essential characteristics of WiFi signals for reliable sensing.

By enhancing security in integrated sensing and communication systems, our approach addresses critical challenges posed by adversarial entities in modern wireless environments. This work provides a valuable contribution to the advancement of secure, resilient, and efficient wireless sensing systems, paving the way for future developments in safeguarding sensing applications against sophisticated threats.

References

[1] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–36, 2019.

[2] E. Cianca, M. De Sanctis, and S. Di Domenico, "Radios as sensors," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 363–373, 2016.

[3] "Taskgroupbf webpage :," https://www.ieee802.org/11/Reports/tgbf_update.htm.

[4] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using wifi signals," in *Proceedings of the 21st annual international conference on mobile computing and networking*, 2015, pp. 90–102.

[5] Y. Zhu, Z. Xiao, Y. Chen, Z. Li, M. Liu, B. Y. Zhao, and H. Zheng, "Et tu alexa? when commodity wifi devices turn into adversarial motion sensors," *arXiv preprint arXiv:1810.10109*, 2018.

[6] J. Li, D. Mishra, D. Krishnaswamy, A. Chakraborty, J. G. Davis, and A. Seneviratne, "Wifi interference-based adversarial attacks on ntc using csi sensing," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 4354–4359.

[7] Y. Qiao, O. Zhang, W. Zhou, K. Srinivasan, and A. Arora, "{PhyCloak}: Obfuscating sensing from communication signals," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, 2016, pp. 685–699.

[8] H. Ambalkar, X. Wang, and S. Mao, "Adversarial human activity recognition using wi-fi csi," in *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, 2021, pp. 1–5.

[9] 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 *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2022, pp. 1705–1721.

[10] J. Yang, H. Zou, and L. Xie, "Securesense: defending adversarial attack for secure device-free human activity recognition," *IEEE Transactions on Mobile Computing*, 2022.

[11] M. Cominelli, F. Kosterhon, F. Gringoli, R. L. Cigno, and A. Asadi, "Ieee 802.11 csi randomization to preserve location privacy: An empirical evaluation in different scenarios," *Computer Networks*, vol. 191, p. 107970, 2021.

[12] Y. Zhou, H. Chen, C. Huang, and Q. Zhang, "Wiadv: Practical and robust adversarial attack against wifi-based gesture recognition system," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1–25, 2022.

[13] L. Xu, X. Zheng, X. Li, Y. Zhang, L. Liu, and H. Ma, "Wicam: Imperceptible adversarial attack on deep learning based wifi sensing," in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2022, pp. 10–18.

[14] C. Li, M. Xu, Y. Du, L. Liu, C. Shi, Y. Wang, H. Liu, and Y. Chen, "Practical adversarial attack on wifi sensing through unnoticeable communication packet perturbation," in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, 2024, pp. 373–387.

[15] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in *Proceedings of the 2017 ACM on Asia conference on computer and communications security*, 2017, pp. 506–519.

[16] "wirelesslan webpage :," https://wlanprofessionals.com/mcs-table-and-how-to-use-it/.

[17] "Esp32 webpage :," https://www.espressif.com/en/products/socs/esp32, accessed: 2023-08-06.

[18] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[19] A. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.