IEEE Open Journal of Signal Processing

Received 9 April 2024; revised 26 July 2024; accepted 14 August 2024. Date of publication 2 September 2024; date of current version 24 September 2024. The review of this article was arranged by Associate Editor Mauro Barni.

Digital Object Identifier 10.1109/OJSP.2024.3453176

Adversarial Training for Jamming-Robust Channel Estimation in OFDM Systems

MARCELE O. K. MENDONÇA [1,2] (Member, IEEE), PAULO S. R. DINIZ [2] (Life Fellow, IEEE), JAVIER MAROTO MORALES [3] (Member, IEEE), AND PASCAL FROSSARD [3] (Fellow, IEEE)

[1]Interdisciplinary Centre for Security, Reliability, and Trust (SnT), 1855 Kirchberg, Luxembourg
[2]Department of Electrical and Computer Engineering, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-972, Brazil
[3]École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

CORRESPONDING AUTHOR: MARCELE O. K. MENDONCA (e-mail: marcele.kuhfuss@smp.ufrj.br).

The work of Javier Maroto was supported by Armasuisse Science and Technology project TRACIE under Grant AR-CYD-C-025. This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)–Finance Code 001 and in part by the Swiss Government Excellence Scholarships for Foreign Students.

ABSTRACT Orthogonal frequency-division multiplexing (OFDM) is widely used to mitigate inter-symbol interference (ISI) from multipath fading. However, the open nature of wireless OFDM systems makes them vulnerable to jamming attacks. In this context, pilot jamming is critical as it focuses on corrupting the symbols used for channel estimation and equalization, degrading the system performance. Although neural networks (NNs) can improve channel estimation and mitigate pilot jamming penalty, they are also themselves susceptible to malicious perturbations known as adversarial examples. If the jamming attack is crafted in order to fool the NN, it represents an adversarial example that impairs the proper behavior of OFDM systems. In this work, we explore two machine learning (ML)-based jamming strategies that are especially intended to degrade the performance of ML-based channel estimators, in addition to a traditional Additive White Gaussian Noise (AWGN) jamming attack. These ML-based attacks create noise patterns designed to reduce the precision of the channel estimation process, thereby compromising the reliability and robustness of the communication system. We highlight the vulnerabilities of wireless communication systems to ML-based pilot jamming attacks that corrupts symbols used for channel estimation, leading to system performance degradation. To mitigate these threats, this paper proposes an adversarial training defense mechanism desined to counter jamming attacks. The effectiveness of this defense is validated through simulation results, demonstrating improved channel estimation performance in the presence of jamming attacks. The proposed defense methods aim to enhance the resilience of OFDM systems against pilot jamming attacks, ensuring more robust communication in wireless environments.

INDEX TERMS Pilot jamming attacks, machine-learning, channel-estimation, OFDM, adversarial training.

I. INTRODUCTION

Within the context of broadband wireless communications, the management of multipath fading has long been a crucial concern [1]. A popular and effective technique employed to address this challenge is Orthogonal Frequency Division Multiplexing (OFDM). Pilot symbols are used in OFDM to obtain channel state information (CSI), which is often estimated using techniques like least squares (LS) and minimum mean square error (MMSE). Accurate CSI is vital to ensure that the receiver can perform channel equalization and properly detect the data symbols.

Deep learning (DL) has recently been proposed to build effective methods that increase the accuracy of channel estimation [2]. For example, two neural networks were designed to refine the CSI accuracy and improve data detection in [3]. In [4], a simple NN is sufficient to improve the LS-based channel estimates when insufficient redundancy is employed and nonlinear clipping distortion is present. Convolutional neural networks (CNNs) are also commonly used to obtain CSI [5].

Yet, the OFDM systems find application in many over-the-air systems [6], [7] where in all cases there are vulnerable to

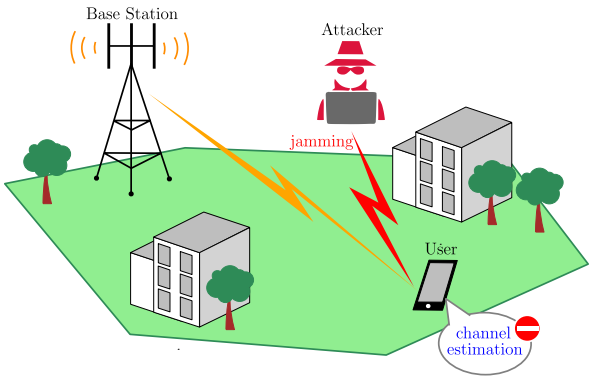


FIGURE 1. Illustration of a jamming attack during the channel estimation of a wireless OFDM system.

adversarial attacks. As the wireless communication channel is open and exposed, the OFDM systems are however prone to jamming attacks [8]. In wireless communications, jamming is defined as an intentional interfering signal hindering the transmission or distorting the legitimate signal. These harmful attacks have the potential to seriously impair communication. This risk possibly increases when utilizing machine learning-based methods in wireless communications systems, as neural networks are known to be particularly vulnerable to adversarial examples. Adversarial examples are small and intended perturbations designed to fool neural networks, and it is crucial to derive methods that can be robust to such attacks.

In this work, we explore several pilot jamming attacks in wireless OFDM systems, as illustrated in Fig. 1. Although we focus on pilot jamming attacks, similar concepts can be extended to various other physical layer functions. In fact, whenever ML is used to improve the physical layer of wireless communication systems, it also poses new possibilities of threats. These threats could manifest in the form of adversarial attacks aimed at disrupting synchronization, channel estimation, signal demodulation, and other critical functionalities. Thus, our exploration of pilot jamming attacks acts as a stepping stone towards understanding and mitigating broader security challenges in ML-enhanced wireless communication systems. In pilot jamming attacks, the pilot tones are corrupted by an intentional interfering signal that aims to impair the channel estimation task. We consider that when a neural network is used to estimate the channel, the jamming attacks can be divided into two categories: the conventional jamming attack that disregards the NN vulnerability to adversarial examples, and the NN-based jamming attack that rather exploits this vulnerability. In addition, our research explores physical layer security strategies specifically designed to counter ML-based jamming attacks. As such, our focus remains primarily on the physical layer, and we do not consider the security aspects of higher layers in the communication protocol stack. As such, investigating cybersecurity strategies is outside the scope of this work. In this context, we propose the following contributions:

- We extend the concept of adversarial attacks, commonly used in other applications, into the context of wireless communication, with a specific focus on their implications for the critical task of modern channel estimation based on machine learning techniques. By identifying the potential vulnerabilities, we raise awareness about the security implications of incorporating machine learning in wireless communications.
- We particularly investigate pilot jamming attacks within wireless OFDM systems, showcasing how adversarial attacks are related to jamming attacks and can deliberately target pilot tones to disrupt channel estimation. Our exploration of these attacks sheds light on the underexplored area of the physical layer of wireless communications, providing crucial insights into potential threats to wireless communication systems.
- We propose a novel defense mechanism based on adversarial training, specifically customized for protecting CSI acquisition in OFDM systems. Our approach offers a robust shield against pilot jamming attacks, effectively preserving the accuracy of channel estimation.

We start by describing the OFDM system in Section II. We then formalize three different attacks, including conventional and NN-based jamming attacks in Section III. In Section IV, we propose an adversarial training framework adapted to CSI acquisition in OFDM systems as a defense to pilot jamming attacks. Section V shows some simulation results to evaluate the proposed robust OFDM receiver that is shown to enhance CSI estimation when the system is being attacked. Section VI includes concluding remarks.

II. ML-BASED OFDM CHANNEL ESTIMATION

A. GENERAL FRAMEWORK

Consider the OFDM transmitter illustrated in Fig. 2.

The incoming data streams are modulated by using the M-ary quadrature amplitude modulation (M-QAM), resulting in the symbols $\boldsymbol{x}_P^{\mathrm{T}} \in \mathbb{C}^{1\times N}$, in which N is the number of subcarriers. We consider a block-type pilot arrangement where all subcarriers either contain pilots or data, as in [3], [9]. A comb-type pilot arrangement can also be considered as shown in [10]. The symbols are then converted to a parallel data stream $\boldsymbol{x}_P \in \mathbb{C}^{N\times 1}$. An N-point inverse fast Fourier transform (IFFT) converts the signal to the time domain, $\mathbf{x}_P = \mathbf{W}^{\mathrm{H}} N \boldsymbol{x}_P$, where $\mathbf{W}_N$ is the $N \times N$ discrete Fourier transform (DFT) matrix.

The OFDM signal is transmitted after adding redundancy and performing a clipping operation to control nonlinear distortion, defined by

$$\mathbf{u}_c = \begin{cases} \mathbf{A}\mathbf{x}_P, & \text{if } |\mathbf{A}\mathbf{x}_{\mathrm{P}}| < (\mathrm{C_R}\sigma_{\mathrm{u}}). \\ \frac{\mathbf{A}\mathbf{x}_P}{|\mathbf{A}\mathbf{x}_P|}(C_R\sigma_u), & \text{otherwise}, \end{cases} \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{S\times N}$ adds redundancy as in [11] and σ_u is the root mean square (RMS) value of the OFDM signal.

The channel model is described by the impulse response $\mathbf{h} = [\mathrm{h}(0)\ \mathrm{h}(1)\ \cdots \mathrm{h}(\mathrm{L})]^{\mathrm{T}}$, with the pseudo-circulant channel

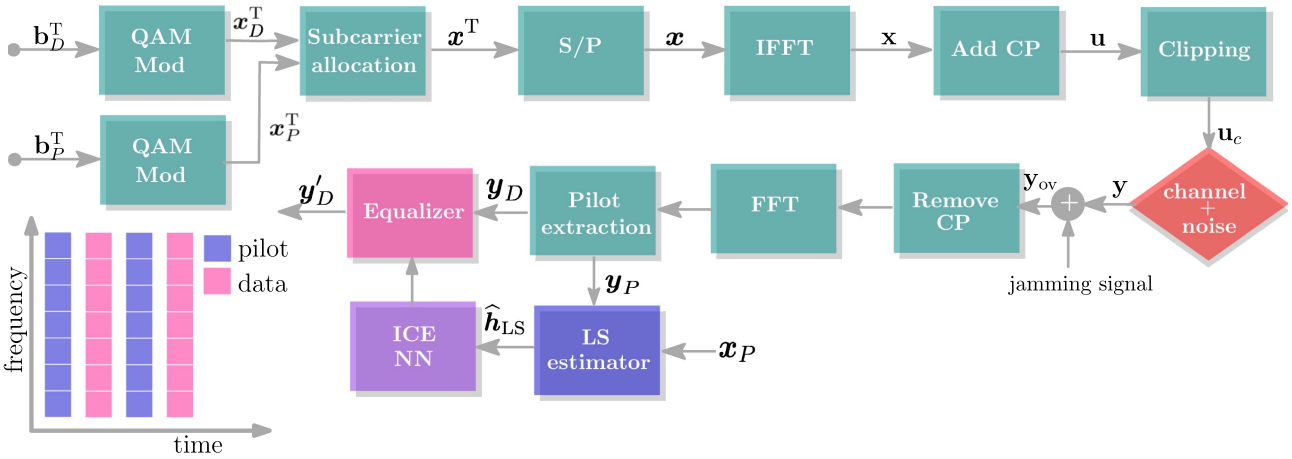


FIGURE 2. The system employs OFDM with a neural network to estimate the wireless channel. Pilot x_P and data x_D signals are utilized, organized into a resource grid block x. The time-domain representation of the block is x, and u includes the cyclic prefix (CP). Then, the clipping is applied generating u_c that, in turn, traverses the channel generating y. The received signals, including both intentional y signal and a jamming signal, are processed at the receiver. The pilot phase involves obtaining a channel estimate using LS and refining it with the ICE neural network. This estimate is then used to equalize the data symbols.

matrix given by $\mathbf{H}(z) = \mathbf{H}_{\text{ISI}} + z^{-1}\mathbf{H}_{\text{IBI}}$ as shown in [11]. We assume the channel remains constant over an OFDM frame [3], [9].

The received signal in the time domain, without noise, is

$$\mathbf{y}(k) = \mathbf{R}\mathbf{H}_{\text{ISI}}\mathbf{A}\mathbf{x}(k) + \mathbf{R}\mathbf{H}_{\text{IBI}}\mathbf{A}\mathbf{x}(k-1), \tag{2}$$

with redundancy removed by $\mathbf{R}$ [11].

Using the estimated CSI, the receiver achieves effective channel equalization, ensuring precise reconstruction of the original data symbols. Accurate CSI estimation is crucial, impacting system performance, interference mitigation, and signal recovery.

B. NN-BASED CHANNEL ESTIMATION

The amount of redundancy plays a significant role in the spectrum efficiency of OFDM systems. Therefore, operating with reduced or no redundancy is highly desired [9]. When redundancy is insufficient, the channel matrix is no longer circulant, affecting the LS-based channel estimate [12]:

$$\widehat{h}_{\text{LS}}(n) = \frac{y_P(n)}{x_P(n)} \quad \text{for } n = 1, \cdots N, \tag{3}$$

where $\boldsymbol{x}_P \in \mathbb{C}^{N\times 1}$ and $\boldsymbol{y}_P \in \mathbb{C}^{N\times 1}$ are the transmitted and received pilot signals in the frequency domain.

NNs can enhance channel estimates under these conditions. They excel at pattern recognition, making them suitable for mitigating issues like insufficient redundancy and nonlinear clipping distortion. NNs, after training, are also less computationally intensive compared to methods like LMMSE [13]. The linear minimum mean-squared error (LMMSE) estimate is given by

$$\widehat{\boldsymbol{h}}_{\text{LMMSE}} = \mathbf{W}_{\text{LMMSE}}\widehat{\boldsymbol{h}}_{\text{LS}}, \tag{4}$$

where $\mathbf{W}_{\text{LMMSE}}$ is the LMMSE weight matrix defines as follows

$$\mathbf{W}_{\text{LMMSE}} = \mathbf{R}_{\boldsymbol{h}_N\widehat{\boldsymbol{h}}_{\text{LS}}}\left(\mathbf{R}_{\boldsymbol{h}_N\boldsymbol{h}_N} + \frac{\sigma_v^2}{E[||\mathbf{x}||^2]}\mathbf{I}_N\right)^{-1}, \tag{5}$$

in which $\mathbf{R}_{\boldsymbol{h}_N\widehat{\boldsymbol{h}}_{\text{LS}}}$ is the cross-correlation matrix between the true channel vector and channel estimate vector in the frequency domain. The auto-correlation matrix of $\boldsymbol{h}_N$ is $\mathbf{R}_{\boldsymbol{h}_N\boldsymbol{h}_N}$. The energy of the transmitted symbol is $E[||\mathbf{x}||^2]$, and $\boldsymbol{h}_N$ is $\mathbf{h}_N$ in the frequency domain. This solution has some practical forms of implementation, see [13].

In this paper, we use the improved channel estimator (ICE) NN model proposed in [14]. ICE is designed to enhance LS channel estimation by minimizing the MSE between the LS estimate and the true channel response. This neural network model not only improves CSI but also offers computational efficiency, making it a valuable asset in the context of channel estimation. The ICE subnet is an improved version of the CE NN proposed in [3] including one hidden layer with hyperbolic tangent as activation function.

III. PILOT JAMMING ATTACKS IN OFDM SYSTEMS

A. JAMMING ATTACKS

There are many types of jamming threats in wireless communications, many of those occurring in the physical layer implemented with OFDM-type transceivers. A jamming attack is defined as the intentional use of radio noise or waveform signals to obstruct communication [15]. The simplest type of jamming attack is called barrage jamming, in which the attacker, with no prior knowledge of the target, attempts to jam a full band of OFDM waveform with noise-like signal [16]. On the other hand, correlated jamming attacks exploit the knowledge about the OFDM waveform to craft the attacks [17]. Among correlated jamming attacks, pilot jamming attacks target the pilot symbols used for channel

estimation leading to noisy CSI and hence performance degradation [18].

The risk posed by pilot jamming is exacerbated when neural networks (NNs) are used for channel estimation. NNs, while powerful, are vulnerable to adversarial examples—deliberately crafted inputs designed to deceive the network [19]. This vulnerability is not limited to channel estimation alone. Any physical layer task enhanced by ML techniques is potentially at risk of being targeted by jamming attacks.

In this work, we focus on the channel estimation task enhanced by ML, exploring how jamming attacks can increase the MSE between the estimated and actual channels. However, the concepts and vulnerabilities discussed here can be extended to other physical layer tasks that leverage ML. This broader implication underscores the importance of developing robust ML models that can withstand adversarial conditions across various applications in the physical layer of wireless communication systems.

This section introduces three jamming attack scenarios. The first is a conventional attack that disregards the NN vulnerability to adversarial examples. The remaining attacks explore the vulnerability of NNs to adversarial examples, and for this reason we call them NN-based attacks.

In the inner-receiver invasion scenario described in Section III-C, we assume the attacker has the capability to compromise the receiver chain, allowing for direct manipulation of the network input. While this may seem improbable in conventional scenarios, it's worth considering the potential ramifications of such an attack, especially in the context of increasingly sophisticated cyber threats. On the other hand, the eavesdropping-assisted jamming attack in Section III-D represents a more practical variant of the inner-receiver invasion scenario discussed earlier. In this attack, the jamming entity strategically situates itself in close proximity to the transmitter. By eavesdropping on the communication signals, the attacker gains information about the transmission parameters, which can then be used to craft the jamming signals.

Our threat model assumes that adversaries have the capability to inject jamming signals and possess sufficient knowledge to interfere with the OFDM system effectively. This involves the adversaries acquiring a LS channel estimate that exhibits a significant correlation with the LS channel estimate obtained by the victim. This capability allows attackers to tailor their jamming signals in a manner that specifically targets and interferes with the channel estimation process used by the legitimate system. We also assume that while channel conditions can vary, they remain relatively stable over the duration of an OFDM frame. This stability allows the attacker to predict and exploit the channel characteristics effectively during the attack period.

B. CONVENTIONAL RANDOM JAMMING ATTACK

Similar to barrage jamming described earlier, no prior information of the target is necessary for the so-called conventional random jamming attack against NN-based channel estimation algorithms. In this case, the overall received signal is

$$\mathbf{y}_{\text{ov}}(k) = \mathbf{y}_P(k) + \mathbf{a}(k), \tag{6}$$

in which $\mathbf{y}_P(k)$ corresponds to the original signal and $\mathbf{a}(k)$ is the spurious signal modeled as an AWGN with variance σ_n^2 [16]. The jammer can be located anywhere between the transmitter and the receiver. The random jamming attack does not intend to fool the NN specifically. Instead, it just adds random noise that results in increasing the MSE between the channel estimate and the true channel response for both LS and ML-based channel estimators.

C. NN-BASED INNER-RECEIVER INVASION JAMMING ATTACK

Many types of adversarial attacks happen on the receiver side after information leaves the physical layer. For instance, gradient-based attacks might occur by disturbing the gradient of the cost function internally at the receiver. The work [20] describes several classes of possible adversarial attacks in open radio access networks that could fall in the class of inner-receiver attacks discussed here, as well as in many IoT deployments.

We can characterize a NN-based inner-receiver invasion[1] jamming attack by assuming that the signal processing chain has been compromised [21]. In this case, the attacker has the capability to compromise the receiver chain, and it can directly perturb the NN input which is the LS-based channel estimate defined in (3)

$$\widehat{h}_{\text{wcj}}(n) = \frac{y_P(n)}{x_P(n)} + \eta(n) \text{ for } n = 1, \cdots N, \tag{7}$$

by adding a perturbation $\eta(n)$. The inner-receiver invasion scenario jamming attack is assumed to know the NN architecture, allowing for the training and acquisition of the NN weights $\boldsymbol{\theta}$ to craft a minimum perturbation $\boldsymbol{\eta} \in \mathbb{C}^{N\times 1}$ so that the MSE between the channel estimate obtained by the NN and the actual channel, is maximally increased. The perturbation $\boldsymbol{\eta}$ can be computed using for instance the Fast Gradient Sign Method (FGSM) attack [22] as

$$\boldsymbol{\eta} = \epsilon \text{sign}(\nabla_{\mathbf{h}} \mathcal{L}(\boldsymbol{\theta}, \mathbf{h}, \boldsymbol{h}_{\text{N}})), \tag{8}$$

with sign$(\cdot)$ the sign function, $\nabla_{\mathbf{h}} \mathcal{L}(\boldsymbol{\theta}, \mathbf{h}, \boldsymbol{h}_{\text{N}})$ the loss gradient with respect to $\mathbf{h}$, $\boldsymbol{h}_{\text{N}}$ is the training label, and ϵ a scalar. In this way, when adding the perturbation to the NN input, the attacker modifies the input towards the direction where the loss $\mathcal{L}$ (MSE) increases. As the added perturbation is computed using an adversarial attack, the output of the inner-receiver invasion jamming attack in (7) can be interpreted as an adversarial example. The FGSM is among the simplest and most effective adversarial attacks possible, being the most widely used.

[1] In this work, the concept of inner-receiver invasion represents that the attacker possesses the capability to compromise the receiver chain, thereby enabling direct manipulation of the network input.

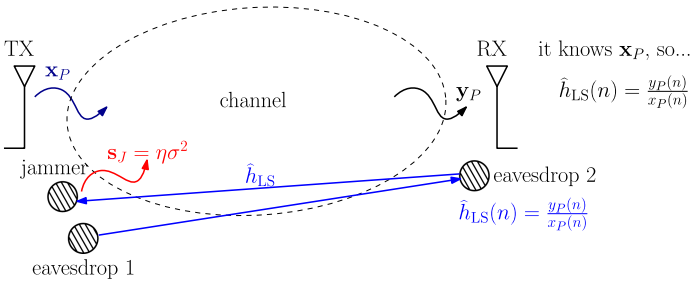


FIGURE 3. Illustration of jamming attack based on adversarial samples.

D. NN-BASED ATTACK: EAVESDROPPING-ASSISTED JAMMING

We introduce an attack strategy termed the eavesdropping-assisted jamming attack, which represents a more practical variant of the inner-receiver invasion scenario discussed earlier. In this attack, the jamming entity strategically positions itself in close proximity to the transmitter, allowing for a more effective interference strategy. The jamming signal is crafted based on the inner-receiver invasion jamming attack defined in (7). This approach reflects a threat scenario where adversaries leverage proximity to exploit vulnerabilities in communication systems.

In this type of attack, the attacker does not need to have precise knowledge of the legitimate communication channel. Instead, it is sufficient to create a channel model that exhibits a significant correlation with the actual channel. In many cases, success can be achieved by deploying the eavesdropping devices close enough to the infrastructure installations [23]. In the GHz deployment range, this type of attack is more challenging despite its very short connection range. An alternative approach in this attack scenario involves the eavesdropper initiating an attack by transmitting a fake pilot symbol in environments with massive MIMO deployments. By doing so, the attacker can disrupt the channel estimation process and degrade the quality of the CSI.

Regardless of the wireless system, predictable synchronization signaling usually does not include encryption. This is a window of opportunity for low-cost attacks.

In future works, we plan to address the issue of the effect of partial knowledge, or the correlation, of the legitimate channel model with respect to the attacker model. Also, a subject of future work is the use of ML solutions at the receiver to deal with the situations where the assumption made about channel estimate is not so realistic.

In this type of attack, the jammer needs a LS-based channel estimate denoted as $\widehat{\boldsymbol{h}}_{\mathrm{LS}}$ correlated with the LS channel estimate obtained by the victim. This channel estimate serves as the foundation upon which the perturbation for the jamming signal is generated. To fulfill this requirement, a pair of eavesdropping sensors are utilized to estimate the channel, with the obtained information transmitted to the jammer. Fig. 3 illustrates this process. Initially, eavesdropper 1 initiates communication by transmitting a predetermined message to eavesdropper 2. Upon receiving this message, eavesdropper 2 leverages the information to estimate channel characteristics. Subsequently, eavesdropper 2 relays this estimated channel information to the jammer, facilitating the crafting of jamming signals. While this elucidates the coordination between the eavesdroppers and the jammer, issues related to synchronization and delays are not within the current scope of this work. Additionally, the potential mismatch between the channel estimates obtained by the victim and the pair of eavesdroppers warrants further analysisis an aspect that needs further analysis to understand the implications of such discrepancies in practical scenarios.

From the perspective of the attack, the fundamental approach remains similar to Section III-C, but with the perturbation applied indirectly via the jamming signal. Importantly, this setup mirrors scenarios within the context of Internet of Things (IoT) environments, where distributed sensors and devices collaborate to optimize wireless communication functions.

The goal of this attack is to transmit a jamming signal $\mathbf{s}_J$ that leads to a channel estimate corrupted by $\eta(n)$

$$\widehat{h}'_{\mathrm{LS}}(n) = \frac{y_{\mathrm{ov}}(n)}{x_P(n)} = \frac{y_P(n)}{x_P(n)} + \frac{h_N(n)s_J(n)}{x_P(n)} = \frac{y_P(n)}{x_P(n)} + \eta(n), \tag{9}$$

which is similar to the inner-receiver invasion scenario in (7). In this case, $\widehat{\boldsymbol{h}}'_{\mathrm{LS}}$ is the estimated channel by the eavesdropper sensors.

The jammer transmits a signal $\mathbf{s}_J$ so that the overall received signal, in time domain, is

$$\mathbf{y}_{\mathrm{ov}}(k) = \mathbf{y}_P(k) + \mathbf{y}_J(k), \tag{10}$$

which is composed of the received original message $\mathbf{y}_P$ and the received jamming signal $\mathbf{y}_J$ during the pilot phase. The received original signal can be described as

$$\mathbf{y}_P(k) = \mathbf{H}_{\mathrm{ISI}}\mathbf{u}(k) + \mathbf{H}_{\mathrm{IBI}}\mathbf{u}(k-1) + \mathbf{v}(k), \tag{11}$$

whereas the received jamming signal is

$$\mathbf{y}_J(k) = \mathbf{H}_{\mathrm{ISI}}\mathbf{u}_J(k) + \mathbf{H}_{\mathrm{IBI}}\mathbf{u}_J(k-1) + \mathbf{v}'(k), \tag{12}$$

where $\mathbf{u}_J = \mathbf{A}\mathbf{s}_J$. Here, $\mathbf{v}(k)$ and $\mathbf{v}'(k)$ represent additive Gaussian noise vectors with zero mean. The matrices $\mathbf{H}_{\mathrm{ISI}}$ and $\mathbf{H}_{\mathrm{IBI}}$ represent intersymbol interference and interblock interference, respectively. In the absence of noise and after removing the redundancy, we obtain

$$\mathbf{y}_{\mathrm{ov}}(k) = \mathbf{R}\mathbf{H}_{\mathrm{ISI}}\mathbf{A}\mathbf{x}(k) + \mathbf{R}\mathbf{H}_{\mathrm{IBI}}\mathbf{A}\mathbf{x}(k-1) + \mathbf{R}\mathbf{H}_{\mathrm{ISI}}\mathbf{A}\mathbf{s}_J(k) + \mathbf{R}\mathbf{H}_{\mathrm{IBI}}\mathbf{A}\mathbf{s}_J(k-1), \tag{13}$$

where $\mathbf{A}$ is the matrix that includes the redundancy [11].

If the redundancy length is adequate, the IBI is eliminated, $\mathbf{R}\mathbf{H}_{\mathrm{IBI}}\mathbf{A} = \mathbf{0}$ and $\mathbf{R}\mathbf{H}_{\mathrm{ISI}}\mathbf{A} = \mathbf{H}_{\mathrm{c}}$ is a circulant matrix. Therefore, the matrix multiplication is equivalent to a circular convolution,

$$\mathbf{y}_{\mathrm{ov}}(k) = \mathbf{H}_{\mathrm{c}}\mathbf{x}(k) + \mathbf{H}_{\mathrm{c}}\mathbf{s}_J(k) = \mathbf{h}_N \circledast [\mathbf{x}(k) + \mathbf{s}_J(k)]. \tag{14}$$

In the frequency domain, we can then use the LS method to estimate the channel at each subcarrier $n = 1, \cdots N$, as described in (9). From (9), the jamming signal can then be expressed as

$$s_J(n) = \eta(n)\frac{x_P(n)}{h_N(n)} \text{ for } n = 1, \cdots N. \tag{15}$$

We proceed to derive an approximate expression for the jamming signal

$$s'_J(n) = \frac{\eta(n)}{||\eta(n)||} E\left[\left(\eta(n)\frac{x_P(n)}{h_N(n)}\right)^2\right] = \eta'(n)\sigma^2 \tag{16}$$

for $n = 1, \cdots N$. The objective is to craft a signal that incorporates the adversarial perturbation denoted as $\eta'(n)$ into its waveform, while ensuring that the power of this signal surpasses a predefined threshold of σ^2.

The perturbation $\boldsymbol{\eta} = [\eta(1), \cdots \eta(N)]^{\mathrm{T}}$ is then computed using the FGSM attack in (8). The FGSM attack is employed to strategically perturb the signal at each subcarrier, ensuring that the resultant jamming signal possesses the desired adversarial characteristics. By computing the perturbation vector $\boldsymbol{\eta}$ through the application of the FGSM attack, we ensure that the jamming signal is not only adversarial in nature but also strong enough to fool the NN in the receiver.

IV. ADVERSARIAL TRAINING-BASED DEFENSE AGAINST PILOT JAMMING ATTACKS

A. ADVERSARIAL TRAINING

In the previous Section, we presented several possible jamming attacks to impair the channel estimation at the OFDM receiver. We now propose a defense to be used against these attacks.

A pivotal connection emerges between the crafted pilot jamming attacks and the broader domain of adversarial attacks. These attacks share a common thread - the capability to degrade the quality of the task at hand. Pilot jamming attacks aim at distorting the essential channel state information within OFDM systems by contaminating the received pilot signals. In stark contrast, adversarial attacks focus on leading neural networks into making highly confident yet erroneous predictions through the use of adversarial examples. Adversarial examples are meticulously crafted perturbations of the network's input data. While these perturbations are very small in scale, their impact is far from trivial. When introduced into the input of the network, adversarial examples induce a significant shift in the its behavior, causing the production of predictions that confidently deviate from the actual or expected outcomes.

Therefore, both pilot jamming attacks and adversarial attacks share the capacity to perform precision-driven subversion. Whether it be the contamination of pilot signals or the manipulation of neural networks, both categories of attacks impair the task. Such a connection serves as a foundation upon which we develop our defense mechanisms, as discussed in subsequent sections, to bolster the resilience of OFDM systems in the face of such formidable threats.

So far, adversarial training is the most effective approach to mitigate the effect of adversarial attacks by training the NN with perturbed versions of the original samples to improve the accuracy on unseen adversarial examples. Adversarial training has emerged as the state-of-the-art defense mechanism in various domains, including computer vision and natural language processing [22]. This technique has proven its efficacy in bolstering the robustness of NNs against adversarial attacks by exposing them to diverse perturbations during the training phase.

B. GENERATION OF ADVERSARIAL EXAMPLES

In a common classification setting, adversarial training continually creates and incorporates adversarial examples into the training process of a deep neural network classifier

$$f_{\boldsymbol{\theta}}(\mathbf{h}) : \mathbb{R}^D \rightarrow \{1 \cdots C\}, \tag{17}$$

with $\boldsymbol{\theta}$ weights, which maps an input $\mathbf{h}$ to a label r from a dataset with C possible classes. Adversarial training attempts to solve the min-max optimization problem

$$\min_{\boldsymbol{\theta}} \frac{1}{|\mathcal{D}|} \sum_{\mathbf{h}, r \in \mathcal{D}} \max_{\boldsymbol{\eta}} \mathcal{L}(f_{\boldsymbol{\theta}}(\mathbf{h}+\boldsymbol{\eta}), \boldsymbol{h}_{\mathrm{N}}) \quad \text{s.t} \, ||\boldsymbol{\eta}||_p \leq \epsilon, \tag{18}$$

where $\mathcal{L}(f_{\boldsymbol{\theta}}(\mathbf{h}+\boldsymbol{\eta}), \boldsymbol{h}_{\mathrm{N}})$ is the loss function on the adversarial sample and $\boldsymbol{\eta}$ is a small perturbation constrained by ϵ.

Creating adversarial samples involves solving the inner maximization problem in (18), in which the loss function $\mathcal{L}$ is maximized in an effort to change the prediction, that is, $f_{\boldsymbol{\theta}}(\mathbf{h}+\boldsymbol{\eta}) \neq f_{\boldsymbol{\theta}}(\mathbf{h})$. The optimization constraints ensure that the distance between the adversarial and original example should be less than ϵ under a particular norm, $||\boldsymbol{\eta}||_p \leq \epsilon$.

The outer minimization problem in (20) is then solved to find the model parameters that minimize the loss on the generated adversarial examples. The original dataset $\mathcal{D}$ is split into small batches $\mathcal{B}$ and stochastic gradient descent (SGD) is employed to update the model parameters

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \mu \frac{1}{|\mathcal{B}|} \sum_{\mathbf{h}, r \in \mathcal{B}} \nabla_{\boldsymbol{\theta}} \mathcal{L}(f_{\boldsymbol{\theta}}(\mathbf{h}+\boldsymbol{\eta}^*), \boldsymbol{h}_{\mathrm{N}}), \tag{19}$$

where the gradient is evaluated at the maximum point $\boldsymbol{\eta}^*$ found in the inner maximization problem, thanks to the Danskin's theorem [24].

C. PROPOSED DEFENSE

Inspired by adversarial training approach, we propose a defense based on adversarial training to cope with jamming attacks. The OFDM receiver is under attack by a jamming signal that aims at perturbing the channel estimation performed by a NN at the receiver. Since a NN is used to estimate the channel, the jamming attack exploits its vulnerability to adversarial examples when crafting the attack. However, as channel estimation is a regression task in this context, we need to adjust the objective function in (18) accordingly.

The mean square error (MSE) is a suitable metric to access the quality of the channel estimation and it is used as loss function. We then aim at finding the perturbation $\boldsymbol{\eta}$ that increases the MSE between the channel estimate obtained by the ICE model [14] and the target channel, as expressed by $\mathcal{L}(f_{\boldsymbol{\theta}}(\widehat{\boldsymbol{h}}_{\mathrm{LS}}+\boldsymbol{\eta}),\boldsymbol{h}_{\mathrm{N}})$, but which also keeps the MSE between the channel estimate obtained with the LS method and the target channel, denoted as $\mathcal{L}(\widehat{\boldsymbol{h}}_{\mathrm{LS}}+\boldsymbol{\eta},\boldsymbol{h}_{\mathrm{N}})$, barely affected. To meet these requirements, we propose the following modified optimization problem

$$\begin{aligned}\min_{\boldsymbol{\theta}}\frac{1}{|\mathcal{D}|}\sum_{\widehat{\boldsymbol{h}}_{\mathrm{LS}},\boldsymbol{h}_{\mathrm{N}}\in\mathcal{D}}&\max_{\boldsymbol{\eta}}\left(\mathcal{L}(f_{\boldsymbol{\theta}}(\widehat{\boldsymbol{h}}_{\mathrm{LS}}+\boldsymbol{\eta}),\boldsymbol{h}_{\mathrm{N}})\right.\\&\left.-\lambda\mathcal{L}(\widehat{\boldsymbol{h}}_{\mathrm{LS}}+\boldsymbol{\eta},\boldsymbol{h}_{\mathrm{N}})\right)\\&\text{s.t. }||\boldsymbol{\eta}||_p\leq\epsilon_0(1-r)^{\frac{SNR}{5}-1},\end{aligned}\tag{20}$$

in which λ is a scaling factor that balances the importance of the two requirements. The initial perturbation constraint is ϵ_0, and r is the decay factor that controls how the perturbation decreases. Since less perturbation is required to increase the overall loss function for high SNR values, the constraint in the perturbation vector $\boldsymbol{\eta}$ is now dependent on the SNR, decaying with a rate r. We remark that by setting $\lambda=0$ and $r=0$ in (20), we end up with the original optimization problem for adversarial training in classification problems in (18).

The adversarial perturbation should also guarantee the similarity between the adversarial sample and the original, unaltered sample. It emphasizes the need for adversarial perturbations to deceive the neural network while preserving the original sample's likeness. In this regard, the optimization constraints ensure that the distance between the adversarial and original examples is less than ϵ under a particular norm, $||\boldsymbol{\eta}||_p\leq\epsilon$. The norms aim to quantify how imperceptible an adversarial example is to humans, particularly in the context of computer vision tasks. Some examples of norms are the l_0 norm, l_2 norm, and l_∞.

In our specific scenario, we draw an analogy between the imperceptibility of adversarial examples to humans in a computer vision problem and the impact on the LS estimator in the jamming problem, treating the LS estimator as a surrogate for human perception. Consequently, our primary objective is to craft adversarial examples that deceive or fool solely the NN, while exerting a relatively lesser impact on the LS estimate. This ensures that the perturbations introduced for adversarial purposes do not unduly distort the estimated channel in a way that diverges significantly from the actual, unaltered channel. In this context, deceiving or fooling the neural network translates to increasing the MSE between the channel estimate and the target channel (label). This approach allows us to effectively measure the similarity between adversarial and clean samples.

The imperceptibility constraint in adversarial attacks ensures that perturbations to input data are subtle yet effective in causing misclassification or model degradation. This principle extends to LS estimation, where perturbations should minimally affect the estimation process while significantly distorting the estimated channel. Future work could explore leveraging conventional anti-jamming techniques to counteract or minimize the impact of adversarial attacks.

The inner maximization problem in (20) is then solved by considering the Fast Gradient Sign Method (FGSM) attack. FGSM generates adversarial examples by modifying the input towards the direction where the overall loss

$$\mathcal{L}_{\mathrm{o}}(\boldsymbol{\theta},\widehat{\boldsymbol{h}}_{\mathrm{LS}},\boldsymbol{h}_{\mathrm{N}})=\mathcal{L}(f_{\boldsymbol{\theta}}(\widehat{\boldsymbol{h}}_{\mathrm{LS}}),\boldsymbol{h}_{\mathrm{N}})-\lambda\mathcal{L}(\widehat{\boldsymbol{h}}_{\mathrm{LS}},\boldsymbol{h}_{\mathrm{N}}))\tag{21}$$

increases. For the original optimization problem, $\lambda=0$ in (21). The adversarial example is then computed

$$\boldsymbol{h}_{\mathrm{adv}}=\widehat{\boldsymbol{h}}_{\mathrm{LS}}+\epsilon\,\mathrm{sign}(\nabla_{\widehat{\boldsymbol{h}}_{\mathrm{LS}}}\mathcal{L}_{\mathrm{o}}(\boldsymbol{\theta},\widehat{\boldsymbol{h}}_{\mathrm{LS}},\boldsymbol{h}_{\mathrm{N}})),\tag{22}$$

where

$$\epsilon=\epsilon_0(1-r)^{\frac{SNR}{5}-1}\tag{23}$$

for the modified optimization problem, $\mathrm{sign}(\cdot)$ is the sign function, and $\nabla_{\widehat{\boldsymbol{h}}_{\mathrm{LS}}}\mathcal{L}_{\mathrm{o}}(\boldsymbol{\theta},\widehat{\boldsymbol{h}}_{\mathrm{LS}},\boldsymbol{h}_{\mathrm{N}})$ is the loss gradient with respect to the input $\widehat{\boldsymbol{h}}_{\mathrm{LS}}$.

Then, using SGD, we can solve the outer minimization problem in (20) and obtain the model parameters that reduce the loss on the created adversarial examples.

V. SIMULATION RESULTS

In this section, we evaluate the proposed defense based on adversarial training via some simulation results, following the setup described in Section V-A. We first consider the original optimization problem when performing the defense and adversarial attacks in Section V-B. Then, we show how the results are improved when using the proposed optimization problem in Section V-C. Table 1 summarizes the type of attacks and defenses utilized.

A. SIMULATION SETUP

1) OFDM SYSTEM

We consider a CP-OFDM system with $N=64$ subcarriers. The input symbols are 16-QAM samples, and the block-type pilot arrangement is considered. The ITU Pedestrian A channel [25] is generated with Matlab's `stdchan` using the channel model `itur3GPAx` with a carrier frequency $f_c=2$ GHz, 4 km/h as velocity, $T_s=200$ ns and order $L=10$. The redundancy length is $K=L/2=5$. As the consider NN model was originally conceived to cope with reduced redundancy, we consider $K=L/2$ in the simulations. Nevertheless, the results also apply for the case we have sufficient redundant elements $K=L$.

2) NN STANDARD TRAINING (NN STA)

In the context of standard training, the ICE model [14] is trained with clean samples for each Signal-to-Noise Ratio (SNR). This means that the NN is trained without any prior knowledge or exposure to adversarial attacks.

TABLE 1. The Table Describes the Attack Classifications Along With Their Respective Defenses Utilized in the Simulations

Type of system		$\widehat{h}(n)$	Channel	estimation	
No attack (clean)	-	$\frac{y_P(n)}{x_P(n)}$	LS	no defense	
Attack (adv)	Conventional random	$\frac{y_P(n)+a(n)}{x_P(n)}$	NN Sta	no defense	
	NN-based inner-receiver invasion	$\frac{y_P(n)}{x_P(n)}+\eta(n)$	NN Rob	defense scenario 1	equation (20) with $\lambda=0$, $r=0$ and $\epsilon_o=10^{-2}$
	NN-based eavesdropping	$\frac{y_P(n)+s_J(n)}{x_P(n)}$		defense scenario 2	equation (20) with $\lambda=0.2$, $r=0.2$ and $\epsilon_o=10^{-2}$

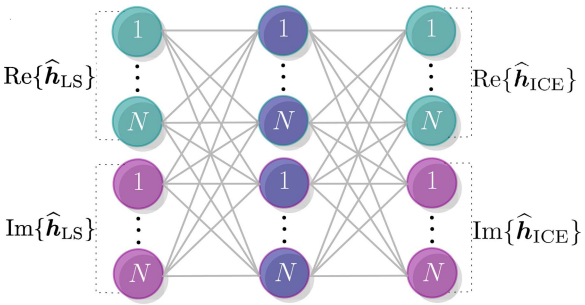


FIGURE 4. ICE model used for ML-based channel estimation [14].

The loss function is minimized by using the adaptive moment estimator (Adam) optimizer with learning rate $\mu = 0.001$.

Since obtaining the true channel response is difficult in practice, the noisy channel training labels are more suitable than the true channel response labels used in [3], [9]. The NN target or label is then defined as the real-valued block version of the noisy version of the ground truth channel gains, $\overline{\boldsymbol{h}}_N = \boldsymbol{h}_N + \boldsymbol{v}$, where $\boldsymbol{v}$ is an additive Gaussian noise vector with zero mean and covariance matrix $\sigma_v^2 \mathbf{I}_N$. The NN input is the real-valued block version of the complex LS-based channel estimate $\widehat{\boldsymbol{h}}_{LS} \in \mathbb{C}^{N\times 1}$, as illustrated in Fig. 4. The real-valued block conversion is required since the channel gains are complex values, and the NN expects real values.

The ML-based channel estimation using the ICE model is described as

$$\begin{bmatrix} \text{Re}\left\{\widehat{\boldsymbol{h}}_{\text{ICE}}\right\} \\ \text{Im}\left\{\widehat{\boldsymbol{h}}_{\text{ICE}}\right\} \end{bmatrix} = \mathcal{F}\left\{\begin{bmatrix} \text{Re}\left\{\widehat{\boldsymbol{h}}_{\text{LS}}\right\} \\ \text{Im}\left\{\widehat{\boldsymbol{h}}_{\text{LS}}\right\} \end{bmatrix}\right\} = \mathcal{F}\left\{\mathbf{g}^{(0)}\right\} \tag{24}$$

where $\mathcal{F}\{\cdot\}$ represents the ICE mapping.

The ICE feedforward process entails the repetition of two steps in each hidden layer. The first step consists of the sum of the weighted outputs of the previous layer,

$$\mathbf{a}^{(l)} = \mathbf{W}^{(l)^T}\mathbf{g}^{(l-1)}, \tag{25}$$

where $\mathbf{a}^{(l)}$ is the input of layer l. The second step consists of applying an activation function at layer l to obtain the output vector,

$$\mathbf{g}^{(l)} = \begin{bmatrix} 1 \\ f(\mathbf{a}^{(l)}) \end{bmatrix}, \text{ for } l = 1, 2, \tag{26}$$

where the entries of $f(\mathbf{a}^{(l)})$ are $f(a_j^{(l)})$ with $a_j^{(l)} = \sum_{i=0}^{d^{(l-1)}} w_{ij}^{(l)} g_i^{(l-1)}$, for $j = 1, \ldots, d^{(l)}$.

The LS-based channel estimate is used as input instead of the received signal because it represents better features for learning the problem. The training set contains 3000 samples. The mini-batch size comprises 50 samples, and 200 epochs are required to train the ICE model. At each epoch, 60 iterations are then required to explore the dataset.

3) NN ROBUST TRAINING (NN ROB)

Conversely, in the context of robust training, the ICE model is trained with the inclusion of adversarial examples for each SNR. In this scenario, the NN is aware of the existence of adversarial attacks and utilizes this knowledge to fortify its defenses and improve its resilience against such attacks. The NN is trained by solving the optimization problem in (20). The loss function $\mathcal{L}$ is minimized by using the adaptive moment estimator (Adam) optimizer with learning rate $\mu = 0.001$.

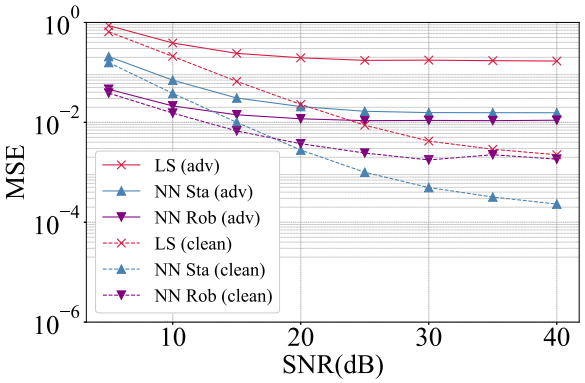
B. SCENARIO 1: ORIGINAL OPTIMIZATION PROBLEM

We first consider the optimization problem in (20) with $\lambda = 0$, $r = 0$, and $\epsilon_0 = 10^{-2}$ to perform the defense. To generate the conventional random jamming attack, we use (6) in which $\mathbf{a}(k)$ is obtained with variance $\sigma_n^2 = \epsilon_0 = 10^{-2}$. Then, for both inner-receiver invasion jamming attack and eavesdropping-assisted jamming attack, the attack is created via FGSM attack in (8) with $\epsilon = \epsilon_0 = 10^{-2}$.

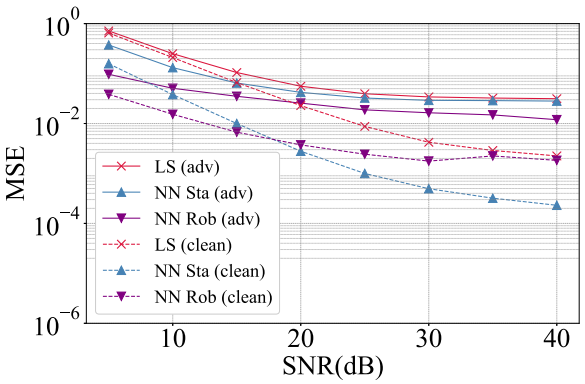
The MSE between the channel estimate and the actual channel is presented as a function of the SNR for random, inner-receiver invasion and eavesdropping-assisted jamming attacks in Fig. 5(a)–(c), respectively. The dashed lines represent a system free of attacks, and they remain consistent across all plots in Fig. 5. Conversely, the solid lines depict the system's performance under attack. Three distinct configurations are presented for comparison:

1) NN Sta (Standard): The ICE model trained exclusively with clean samples.
2) NN Rob (Robust): The ICE model trained using adversarial examples to enhance robustness against attacks.
3) LS (Least Squares): The basic method for estimating the channel without the use of a neural network.

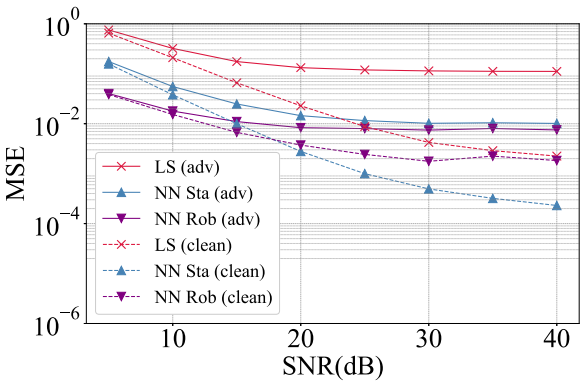
When the original optimization problem is considered, the impairment caused by the inner-receiver invasion jamming attack in the standard trained network is the most severe among



(a) random jamming attack



(b) inner-receiver invasion jamming attack



(c) eavesdropping-assisted jamming attack

FIGURE 5. MSE results when the OFDM receiver is under jamming attacks. The defense is based on the optimization problem in (20) with $\lambda = 0$ and $r = 0$.

the considered attacks. This can be noticed by the gap between the solid and dashed blue curves in Fig. 5(b). Although the inner-receiver invasion jamming attack impairs the LS method less than the other attacks, there is still a significant gap between the solid and dashed red curves in Fig. 5(b). This big gap is expected as the loss function in the original optimization problem does not include any similarity constraints between the adversarial example and the original sample. It is like adding noise to an image without worrying about it being visible to humans. Therefore, the LS performance is clearly impaired in this case. As shown in Fig. 5(a) and (c), both the random and the eavesdropping-assisted jamming attacks yield similar results. This unveils that we are possibly adding an overpowered adversarial perturbation; that is, the adversarial perturbation is so high that it does not only damage the NN results but also the LS results significantly as in the random case.

C. SCENARIO 2: PROPOSED OPTIMIZATION PROBLEM

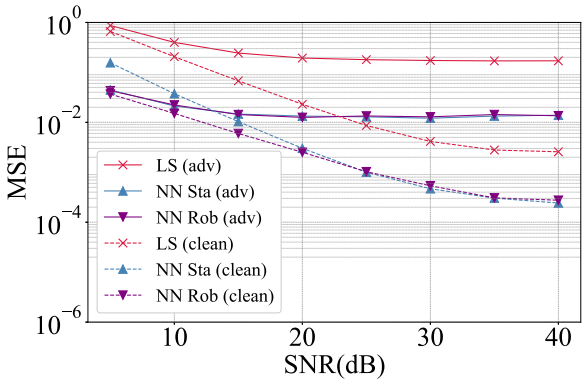
We now consider the optimization problem proposed in (20) with $\lambda = 0.2$ and $r = 0.2$ to perform the defense. To generate the conventional random jamming attack, we use (6) in which $\mathbf{a}(k)$ is obtained with variance $\sigma_n^2 = \epsilon$. For both inner-receiver invasion jamming attack and eavesdropping-assisted jamming attack, the attack is created via FGSM attack in (8) with ϵ in (23) and $\epsilon_0 = 10^{-2}$.

The MSE between the channel estimate and the actual channel is presented as a function of the SNR for random, inner-receiver invasion and eavesdropping-assisted jamming attacks in Fig. 6(a)–(c), respectively. The dashed lines represent a system free of attacks, and they remain consistent across all plots in Fig. 5. The solid lines depict the system's performance under attack.

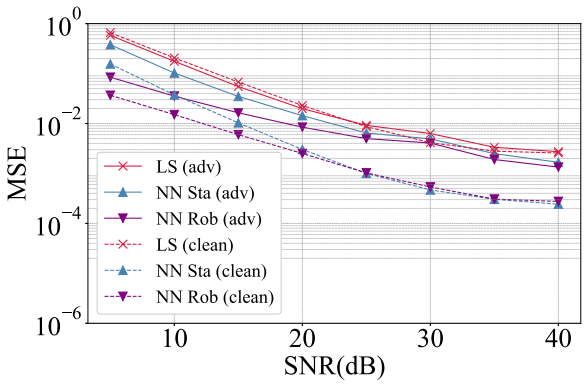
With the proposed optimization problem, we are not only interested in increasing the MSE when the NN is used for channel estimation, but also in minimally affecting the MSE when using the LS channel estimation method. To perform a fair comparison, we varied the power of the jamming signal as in (20) when performing the conventional random jamming attack in Fig. 6(a).

In this case, the NN robust outperforms or performs quite similarly to the NN standard for clean samples (dashed lines). This shows that we can benefit from adversarial training even when the system is free of jamming attacks. As shown in Fig. 6(b), the impairment caused by the inner-receiver invasion jamming attack in the standard trained network is no longer the most severe among the attacks. This shift in behavior is attributed to the fact that the inner-receiver invasion attack is now subject to constraints that limit its perceptibly. The constraints imposed on the inner-receiver invasion attack compel it to be less discernible, resulting in a notable alteration of its impact on the system's performance.

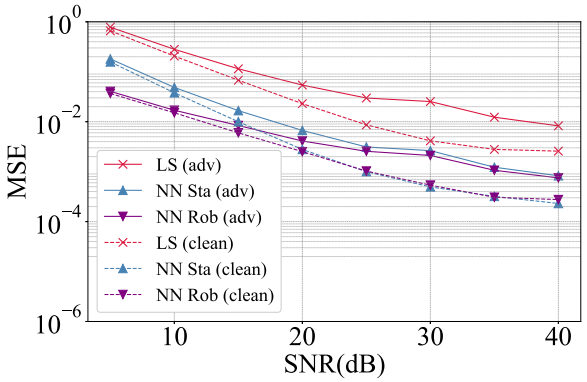
Even though the impact on the MSE is more severe for the conventional random jamming attack, the gap between the solid and dashed red curves is larger if we compare it with the inner-receiver invasion and the eavesdropping-assisted jamming attacks. Unlike the previous scenario, the eavesdropping-assisted jamming attack in Fig. 6(c) is more similar to the inner-receiver invasion jamming attack. This encourages using the eavesdropping-assisted scheme to craft jamming attacks that try to fool the NN without impacting the LS method too much in practice, contrasting to what the random attack does.



(a) random jamming attack



(b) inner-receiver invasion jamming attack



(c) eavesdropping-assisted jamming attack

FIGURE 6. MSE results when the OFDM receiver is under jamming attacks. The defense is based on the optimization problem in (20) with $\lambda = 0.2$ and $r = 0.2$.

VI. CONCLUSION

In this work, we have introduced and analyzed two NN-based pilot jamming attacks and presented a defense mechanism based on adversarial training to address jamming attacks during the channel estimation process in a wireless CP-OFDM system. Our findings confirm that NNs in the context of wireless communications are indeed susceptible to adversarial attacks, underscoring the emergence of new threats posed by ML at the physical layer of wireless communication systems. Moreover, the simulations results demonstrate the efficacy of adversarial training in significantly enhancing the performance of ML-based channel estimators when the CP-OFDM system is subjected to ML-based jamming attacks. This highlights the importance of proactive defense strategies, such as adversarial training, in safeguarding wireless communication systems against adversarial interference. Furthermore, we have shown that implementing the proposed defense does not compromise system performance in the absence of jamming attacks. Through these findings, we emphasize the importance of proactive defense strategies, such as adversarial training, in safeguarding wireless communication systems against adversarial attacks. This highlights the potential of adversarial training as a viable approach to bolstering the resilience and robustness of ML-based channel estimation techniques in the presence of sophisticated jamming attacks.

REFERENCES

[1] R. Prasad, *OFDM for Wireless Communications Systems*. Boston, MA, USA: Artech House Publishers, 2004.
[2] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep learning based channel estimation for massive MIMO with mixed-resolution ADCs," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1989–1993, Nov. 2019.
[3] X. Gao, S. Jin, C. K. Wen, and G. Y. Li, "ComNet: Combination of deep learning and expert knowledge in OFDM receivers," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2627–2630, Dec. 2018.
[4] M. O. K. Mendonça and P. S. R. Diniz, "OFDM receiver using deep learning: Redundancy issues," in *Proc. Eur. Signal Process. Conf.*, 2020, pp. 1687–1691.
[5] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 652–655, Apr. 2019.
[6] S. Sen and A. Nehorai, "Sparsity-based multi-target tracking using OFDM radar," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1902–1906, Apr. 2011.
[7] S. Sen and A. Nehorai, "Target detection in clutter using adaptive OFDM radar," *IEEE Signal Process. Lett.*, vol. 16, no. 7, pp. 592–595, Jul. 2009.
[8] K. Grover, A. Lim, and Q. Yang, "Jamming and anti-jamming techniques in wireless networks: A survey," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 17, no. 4, pp. 197–215, 2014.
[9] J. Zhang, H. He, C. K. Wen, S. Jin, and G. Y. Li, "Deep learning based on orthogonal approximate message passing for CP-free OFDM," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brighton, 2019, pp. 8414–8418.
[10] M. O. K. Mendonça, P. S. R. Diniz, and T. N. Ferreira, "Machine learning-based channel estimation for insufficient redundancy OFDM receivers using comb-type pilot arrangement," in *2022 IEEE Latin-Amer. Conf. Commun.*. IEEE, 2022, pp. 1–6.
[11] P. S. R. Diniz, W. A. Martins, and M. V. S. Lima, *Block Transceivers: OFDM and Beyond*. San Rafael, CA, USA: Morgan & Claypool Publishers, 2012.
[12] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002.
[13] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," *IEEE Trans. Signal Process.*, vol. 68, pp. 1702–1715, 2020.
[14] P. S. R. Diniz and M. O. K. Mendonça, "Zero-padding OFDM receiver using machine learning," in *2021 IEEE Stat. Signal Process. Workshop*, IEEE, 2021, pp. 26–30.
[15] R. Poisel, *Modern Communications Jamming Principles and Techniques*. Norwood, MA, USA: Artech House, 2011.
[16] T. Basar, "The Gaussian test channel with an intelligent jammer," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 1, pp. 152–157, Jan. 1983.

[17] C. Shahriar et al., "PHY-layer resiliency in OFDM communications: A tutorial," *IEEE Commun. Surv. Tut.*, vol. 17, no. 1, pp. 292–314, Firstquarter 2014.
[18] T. C. Clancy, "Efficient OFDM denial: Pilot jamming and pilot nulling," in *2011 IEEE Int. Conf. Commun.*, 2011, pp. 1–5.
[19] M. A. Durmaz, H. Alakoca, G. K. Kurt, and C. Ayyildiz, "Chirp subcarrier jamming attacks: An OFDM based smart jammer design," in *2017 16th Annu. Mediterranean Ad Hoc Netw. Workshop*, 2017, pp. 1–5.
[20] E. Habler et al., "Adversarial machine learning threat analysis and remediation in open radio access network (o-ran)," Mar. 4, 2023, *arXiv:2201.06093v2*.
[21] B. Flowers, R. M. Buehrer, and W. C. Headley, "Evaluating adversarial evasion attacks in the context of wireless communications," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 1102–1113, 2020.
[22] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," 2014, *arXiv:1412.6572*.
[23] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 21–27, Jun. 2015.
[24] J. M. Danskin, "The theory of max-min, with applications," *SIAM J. Appl. Math.*, vol. 14, no. 4, pp. 641–664, 1966.
[25] A. G. Marques, "Guidelines for evaluation of radio transmission technologies for IMT-2000," Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA: Recommendation ITU-R, M. 1225, 1997.

MARCELE O. K. MENDONÇA (Member, IEEE) was born in Mesquita, Brazil. She received the Telecommunications Engineering degree from the Fluminense Federal University (UFF), Brazil, in 2016, and the M.Sc. and Ph.D. degrees from the Signal Multimedia and Telecommunications Laboratory, Federal University of Rio de Janeiro (UFRJ), Brazil, in 2018 and 2022, respectively. Her research interests are mainly signal processing, wireless communications and machine learning. She was the 2nd place in the UFF Vasconcellos Torres Award of Scientific Initiation with the project OFDM Systems in Software Gnuradio using USRP in 2014. She was the recipient of the Best Demo Award at the Brazilian Telecom Symposium with the project FM transmission and reception using software-defined radios in 2019. She was a Ph.D. Visiting Scholar by the Swiss Government Excellence Scholarships Program with LTS4 Research Group, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, during 2021–2022 supervised by Prof. Pascal Frossard. In 2022, she joined the Interdisciplinary Centre for Security, Reliability, and Trust, University of Luxembourg, Luxembourg, as a Research Associate.

PAULO S. R. DINIZ (Life Fellow, IEEE) was born in Niterói, Brazil. He received the Electronics Engineering degree (cum laude) in electrical engineering from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 1978, the M.Sc. degree in electrical engineering from COPPE/UFRJ, Rio de Janeiro, in 1981, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, QC, Canada, in 1984. He is currently with the Program of Electrical Engineering and the Department of Electronics and Computer Engineering, COPPE/Poli, Universidade Federal do Rio de Janeiro. He has authored or coauthored more than 300 refereed papers in journals and conference papers and the following textbooks: *Online Learning and Adaptive Filters* (Cambridge, U.K.: Cambridge University Press, 2021, with M. L. Campos, W. A. Martins, M. V.S. Lima, and J. A. Apolinrio Jr.), *Adaptive Filtering: Algorithms and Practical Implementation* (NY: Springer, Fifth Edition, 2020), and *Digital Signal Processing: System Analysis and Design* (Cambridge, U.K.: Cambridge University Press, Second Edition, 2010, with E. A. B. da Silva and S. L. Netto), and the monograph *Block Transceivers: OFDM and Beyond* (NY: Springer, 2012, with W. A. Martins and M. V.S. Lima). His research interests include analog and digital signal processing, adaptive signal processing, digital communications, wireless communications, multirate systems, stochastic processes, and machine learning. Dr. Diniz is a Fellow of EURASIP. He is a National Academy of Engineering and the Brazilian Academy of Science member. He was the recipient of the 2014 Charles Desoer Technical Achievement Award of the IEEE Circuits and Systems Society. He was the recipient of some of the best paper awards from the conferences and an IEEE journal. He was an Associate Editor for IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: ANALOG AND DIGITAL SIGNAL PROCESSING from 1996 to 1999, *Circuits, Systems, and Signal Processing* journal from 1998 to 2002, and IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2002. From 2000 to 2001, he was a Distinguished IEEE Circuits and Systems Society Lecturer. In 2004, he was a Distinguished IEEE Signal Processing Society Lecturer.

JAVIER MAROTO MORALES (Member, IEEE) received the Ph.D. degree in electrical engineering from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, where he focused on enhancing the robustness of neural networks against adversarial perturbations with applications in computer vision and wireless communications. His research interests include adversarial training, knowledge distillation, generative AI and recommender systems, which has led to publications in leading machine learning conferences.

PASCAL FROSSARD photograph and biography not available at the time of publication.