

Collusion-Driven Impersonation Attack on Channel-Resistant RF Fingerprinting

Zhou Xu

*School of Cyber Science
and Engineering
Southeast University
Nanjing, China
xu_zhou@seu.edu.cn*

Guyue Li

*School of Cyber Science
and Engineering
Southeast University
Nanjing, China
guyuelee@seu.edu.cn
(Corresponding author)*

Zhe Peng

*Dept. of Industrial
and Systems Engineering
The Hong Kong
Polytechnic University
Hong Kong, China
jeffrey-zhe.peng@polyu.edu.hk*

Aiqun Hu

*National Mobile Communications
Research Laboratory
Southeast University
Nanjing, China
aqhu@seu.edu.cn*

Abstract—Radio frequency fingerprint (RFF) is a promising device identification technology, with recent research shifting from robustness to security due to growing concerns over vulnerabilities. To date, while the security of RFF against basic spoofing such as MAC address tampering has been validated, its resilience to advanced mimicry remains unknown. To address this gap, we propose a collusion-driven impersonation attack that achieves RF-level mimicry, successfully breaking RFF identification systems across diverse environments. Specifically, the attacker synchronizes with a colluding receiver to match the centralized logarithmic power spectrum (CLPS) of the legitimate transmitter; once the colluder deems the CLPS identical, the victim receiver will also accept the forged fingerprint, completing RF-level spoofing. Given that the distribution of CLPS features is relatively concentrated and has a clear underlying structure, we design a spoofed signal generation network that integrates a variational autoencoder (VAE) with a multi-objective loss function to enhance the similarity and deceptive capability of the generated samples. We carry out extensive simulations, validating cross-channel attacks in environments that incorporate standard channel variations including additive white Gaussian noise (AWGN), multipath fading, and Doppler shift. The results indicate that the proposed attack scheme essentially maintains a success rate of over 95% under different channel conditions, revealing the effectiveness of this attack.

Index Terms—Radio frequency fingerprint, physical-layer security, impersonation attack, channel-resistant, variational autoencoder.

I. INTRODUCTION

With the rapid proliferation of Internet of Things (IoT) and wireless communication technologies, device authentication becomes essential to ensure network security [1]. Radio frequency fingerprinting (RFF), which exploits the inherent hardware impairments of radio transmitters for device identification, has emerged as a promising non-cryptographic physical-layer solution [2]. Compared with traditional cryptographic methods, RFF-based techniques avoid key management and require minimal protocol overhead, making them suitable for resource-constrained IoT environments. Over the past decade, the research agenda around RFF has undergone a notable pivot: early efforts concentrated almost exclusively on improving robustness under benign channel variations like multipath fading,

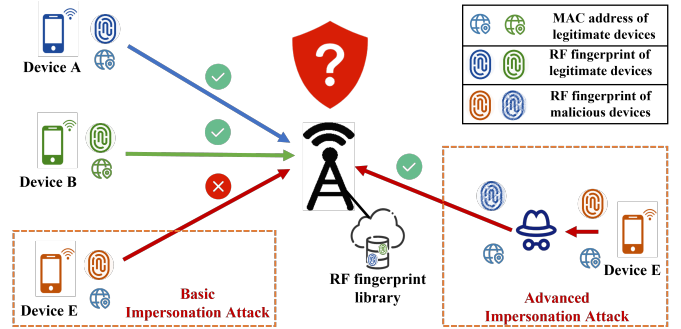


Fig. 1. Beyond Basic Impersonation: the missed threat of Advanced RF mimicry.

Doppler spread, and thermal noise. These studies validated that, once trained on sufficiently diverse channel conditions, RFF classifiers could maintain high accuracy without requiring frequent re-calibration [3].

Recently, the community's focus has decisively pivoted towards RFF vulnerabilities which stem from two sources: the neural network-based recognition system and the RF features it relies on. Driven by deep learning's ability to extract effective features directly from I/Q samples, researchers have widely adopted neural networks for RFF classification, yet this shift simultaneously exposes the system to evasion, backdoor, and impersonation attacks akin to those plaguing computer vision and NLP [4]–[11]. Yet these threats somewhat remain loosely coupled to RFF: they generally rely on strong assumptions that limit their real-world impact. Fig. 1 spotlights another blind spot in the security of RF features: while basic impersonation attacks are easily caught, the more sophisticated advanced impersonation attack may slip through unnoticed. The prevailing belief is that RF fingerprint offers solid security: an attacker who merely forges a MAC address still exposes a distinct RF fingerprint and is therefore detectable; we call this the Basic Impersonation Attack. Yet attackers may go further and fabricate the RF fingerprint itself to deceive the authentication system, an Advanced Impersonation Attack. The feasibility of

such an attack is critical for the credibility of RF fingerprint security, yet it remains largely uninvestigated and its potential success is uncertain.

RF-level impersonation is challenging primarily because the attacker's receiver is distinct from the legitimate receiver, rendering precise RFF information inaccessible. However, this very limitation can be overcome through a collusion attack. The attacker places an RFF module at the colluder matching the legitimate receiver. It captures RF from both sources, extracts fingerprints, infers the target's features, and crafts spoofed signals indistinguishable to the receiver. This colluding attack framework can reduce the attacker's reliance on prior knowledge. Inspired by this insight, Xu et al. [12] inserted a colluder and, under additive white Gaussian noise (AWGN) channels, successfully breached the RFF system to achieve RF-level forgery. However, the method cannot be extended to richer multipath environments. It relies on the premise that if the colluder cannot distinguish the attacker's RF signature from the legitimate one, the legitimate receiver will likewise fail. Because the colluder and the legitimate receiver occupy different physical locations, this premise seldom holds. Fortunately, we show this limitation is surmountable. To work in multipath channels, legitimate receivers must rely on channel-robust RF features that yield identical fingerprints at different locations. For example, methods such as the amplitude-only channel-independent spectrogram [13], amplitude-phase channel-independent spectrogram [14], and centralized logarithmic power spectrum (CLPS) [15] have demonstrated the ability to suppress channel-induced distortions while preserving unique device characteristics. By equipping the colluder with the same feature extractor, we can produce spoofed signals that remain indistinguishable across varying propagation paths, rendering the attack resilient to multipath effects.

Motivated by these considerations, this paper investigates the feasibility and effectiveness of a collusion-driven impersonation attack towards RFF identification in multipath environments. We analyze the limitations of current attack models under realistic channel assumptions and investigate how attackers can exploit channel-resistant feature extraction methods to improve impersonation success rates. By introducing CLPS into the attack model and leveraging a variational autoencoder (VAE) for signal generation, we demonstrate that the collusion-driven impersonation attack can achieve high attack success rates across a wide range of challenging wireless environments.

The main contributions of this paper can be summarized as follows:

- We propose a collusion-driven impersonation attack strategy. To address the issue with existing impersonation attacks, we construct a new attack model. This model introduces CLPS as the channel-resistant RFF, and with the assistance of a colluder, enables the attacker to generate signals that can effectively mimic the target device's channel-resistant features under complex channel

conditions.

- We design a VAE-based spoofed signal generation network with a multi-objective optimization strategy. This strategy directly guides spoofing performance at both the feature and classifier decision levels, thereby addressing the problem of high-concentration feature distributions and improving the overall attack effectiveness.
- We build a simulation-based evaluation system incorporating typical channel variations (e.g., additive white Gaussian noise (AWGN), multipath fading, Doppler shift). Experimental results demonstrate that the proposed method maintains high spoofing performance in cross-channel scenarios and achieves over 95% success rate under various channel parameters (e.g., SNR, K-factor).

The rest of this paper is organized as follows: Section II examines current attack strategies and identifies their issues; Section III introduces the relevant theoretical foundations; Section IV details the proposed attack model and strategy; Section V details the framework of the proposed attack; Section VI presents experimental results and analysis; finally, Section VII concludes the paper.

II. RELATED WORKS

This section reviews existing attacks on RFF systems and discusses their limitations.

A. Overview of existing attack methods

Vulnerabilities in RFF systems are targeted within two main components: the neural networks used for identification and the extracted RFF features.

Neural network attacks can be executed in the following manner. Evasion attacks, for instance, create perturbation signals that overlay genuine transmissions, leading to misclassification. The algorithms for creating these perturbations often rely on gradient information or the neural network's decision boundaries. Ma et al. [4] reported success rates exceeding 90% with the Fast Gradient Sign Method and Projected Gradient Descent on channel-independent spectrogram features. Liu et al. [5] designed algorithms effective in both ideal and non-ideal channel scenarios, achieving similarly high success rates. Similarly, backdoor attacks incorporate corrupted samples into the training phase and employ trigger signals during validation to induce misclassification. Creating both these tainted samples and triggers necessitates comprehension of the neural network's decision mechanics. Huang et al. [6] attained nearly 100% success in attacks, while preserving system accuracy on untainted samples, effectively ensuring stealth.

Impersonation attacks aim directly at RFF features by replicating or altering them, necessitating partial insight into the neural network's identifier. Danev et al. [7] investigated methods like signal and feature replay. Merchant and Noursain [9] utilized Generative Adversarial Networks (GANs) for creating spoofed signals and boosting defense. Karunaratne et al. [10] initiated black-box attacks with reinforcement learning using solely binary feedback. Xu et al. [12] employed colluders

TABLE I
SUMMARY OF ATTACKS ON RF FINGERPRINT

Ref	Attack Type	Target Feature	Attack assumptions								
		Channel-resilient Feature	Known Legal Entity Information				Consider attacker's RFF	Channel conditions			
			Model structure	Model parameters	RFF method	Fixed packet		AWGN	Multipath	Doppler	Non-Identical
[4]	Evasion	●	●	●	●	●	○	○	○	○	○
[5]	Evasion	○	●	●	●	○	○	●	●	●	●
[6]	Backdoor	○	●	●	●	●	○	○	○	○	○
[7]	Impersonation	○	●	●	●	●	○	○	○	○	○
[9]	Impersonation	○	●	●	●	○	○	●	○	○	○
[10]	Impersonation	○	○	○	●	●	○	●	●	○	◐
[12]	Colluding Impersonation	○	●	○	●	○	●	●	◐	○	○
This Work	Colluding Impersonation	●	●	●	●	●	●	●	●	●	●

In this table, solid circles (●) denote fully satisfied attack assumptions, hollow circles (○) denote unsatisfied attack assumptions, and semi-solid circles (◐) denote partially satisfied attack assumptions, e.g., the similar channel condition required in [10], the single-tap Rayleigh channel condition in [12].

to overcome hardware limitations, achieving success rates exceeding 90%.

B. Limitations of current approaches

While these works have made significant contributions, several limitations merit consideration:

1) *Information assumptions:* Many studies assume white-box scenarios with comprehensive system knowledge. Although [12] reduced these requirements and [10] introduced black-box approaches, strong assumptions remain prevalent.

2) *Channel modeling:* Several works employ simplified channel models, such as AWGN-only conditions or direct signal superposition without channel effects. While recent studies incorporate multipath fading, they often assume attackers experience similar channel conditions as legitimate users—an assumption rarely valid in practice.

3) *Attack targets:* Current approaches primarily focus on I/Q-level mimicry, which becomes challenging under complex channels. The similarity in time-domain I/Q data may not translate to similarity in channel-robust features, particularly under significant channel variations.

Table I provides an overview of attack types, target characteristics, and underlying assumptions. In summary, current RFF impersonation attack research often relies on simplified channel assumptions, limiting practical efficacy. This work addresses these limitations by introducing channel-resistant RFF into attack scenarios and proposing difference measurement at the feature and classifier decision levels to effectively target these robust features.

III. PRELIMINARY OF CHANNEL-RESISTANT RFF

This section covers the system fundamentals and derives the channel-resistant RFF that remains reliable across multipath environments.

A. System Model

In an RFF-based authentication system, a transmitter modulates a baseband signal $s(t)$ and transmits it through its RF front-end. During this process, unique distortions arise due to hardware imperfections specific to the device, including power amplifier nonlinearity, local oscillator frequency offset, and I/Q imbalance, et al. These distortions resulting from hardware are expressed as a function $f(\cdot)$ applied to the baseband signal. The transmitted signal then propagates through a wireless channel with impulse response $h(t)$ and is corrupted by additive white Gaussian noise $n(t)$. The signal $r(t)$ received at the receiver is represented by

$$r(t) = f(s(t)) * h(t) + n(t) \quad (1)$$

where $*$ denotes convolution. Both hardware and channel distortions are present in $r(t)$. Since the fingerprint at the receiver's front-end is identical for a given receiver, this model omits modeling the receiver's fingerprint. The goal of RFF-based identification is to extract device-specific features that reflect the impact of $f(s(t))$, while being robust to variations in $h(t)$ and $n(t)$.

B. Channel-Resistant RFF

To isolate hardware-specific distortions from channel variations, channel-resistant RFF methods aim to extract features that depend only on hardware imperfections, regardless of the wireless channel conditions. Among these, CLPS features exhibit outstanding RFF-identification performance over unknown channel statistics [15].

CLPS begins by computing the power spectrum (PS) of the received signal, which is estimated as the squared magnitude of the Fast Fourier Transform (FFT) of the signal's autocorrelation function. Subsequently, the logarithmic PS is obtained by

taking the logarithm of the PS $P_r(\omega)$ of $r(t)$. This logarithmic scale representation can be expressed as:

$$P_r(\omega)_{\log} = \log\left(\frac{1}{T}\right) + 2 * \log |\text{FFT}(f(s(t)))| + 2 * \log |\text{FFT}(h(t))| \quad (2)$$

where T denotes the duration of the captured signal.

Given that the wireless channel experiences minimal variation over short durations, it can be reasonably assumed that the channel remains constant for each preamble. Based on this assumption, the CLPS is obtained by subtracting its mean from $P_{r,\log}(\omega)$ to mitigate channel variations:

$$\begin{aligned} \text{CLPS}(\omega) &= P_r(\omega)_{\log} - \text{mean}(P_r(\omega)_{\log}) \\ &= 2 * (\log |\text{FFT}(f(s(t)))| - \text{mean}(\log |\text{FFT}(f(s(t)))|)) \end{aligned} \quad (3)$$

This formulation removes the channel-dependent term and preserves the spectral features introduced by hardware imperfections. Consequently, with a fixed baseband preamble, CLPS operates as a channel-resistant RFF that maintains stability, regardless of changing channel conditions [16].

IV. IMPERSONATION ATTACK STRATEGY UNDER COMPLEX CHANNEL CONDITIONS

In this section, we will introduce the attack model and attack strategy, in order to clarify the attack scenario, assumptions, and the roles of each member within this scenario.

A. Attack Model

The attack model is shown in Fig. 2. We consider a collusion-driven impersonation attack targeting channel-resistant RFF systems. The scenario consists of the following four primary entities: a set of legitimate transmitters, a legitimate receiver, an attacker, and a colluder. Each transmitter periodically sends packets that include a fixed baseband preamble. Each transmitter introduces distinct distortions due to hardware imperfections, like I/Q imbalance, affecting the signal and acting as the device's specific RFF.

During propagation, the signal experiences channel variations including multipath fading, Doppler shift, and AWGN. The legitimate receiver extracts a channel-resistant RFF, from the received signal's preamble segment and performs device identification via a deep neural network classifier. Here we use CLPS as an example, and the approach can later be extended to other channel-independent RFF features.

The attacker aims to impersonate a specific legitimate transmitter by generating spoofed signals that, when passed through its own transmitter hardware and an unknown wireless channel, yield CLPS features at the receiver that resemble those of the target device. In this model, the attacker does not possess knowledge of the receiver's channel conditions or the exact effect of its own hardware on the transmitted signal.

To overcome these limitations, the attacker collaborates with a colluder. The colluder is located at a different position and is capable of receiving transmissions from both the attacker

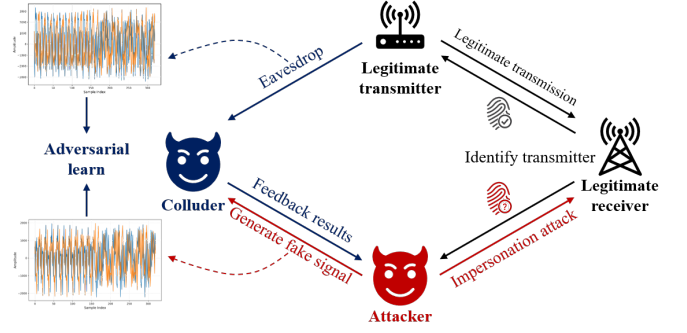


Fig. 2. The Proposed Collusion-Driven Impersonation Attack Model Illustration

and the target transmitter. It is assumed that the colluder has full knowledge of the CLPS feature extraction method and the classifier used by the legitimate receiver. Furthermore, the colluder employs the same RFF identification method as the legitimate receiver. This setup allows the attacker to train a spoofed signal generation network with limited prior information about the target environment. The use of CLPS ensures that even under different channels between the colluder and the receiver, consistent device features can be extracted.

The overall attack assumption settings are summarized in Table I. The attacker is unaware of the impact of its own hardware impairments and relies on a colluding impersonation strategy to circumvent and optimize for this. To address the core challenge that attackers cannot obtain the channel conditions between devices, we resolve this issue by introducing channel-resistant RFF, aiming to enhance the feasibility of impersonation attacks in realistic channel environments. This study addresses an attack scenario operating under a white-box assumption, wherein the attacker has partial knowledge of the legitimate system, including classifier structures, parameters, known baseband signals, and fingerprint extraction techniques. The public access to various RFF methodologies provides a basis for this assumption. In practice, attackers might acquire details about the target RFF system, including feature extraction methods and protocol specifications, by consulting relevant technical documents or employing social engineering tactics. Furthermore, considering the widespread use of deep learning models, the attacker may also leverage existing model stealing techniques [17], in conjunction with eavesdropped legitimate signals, to train a surrogate model that approximates the behavior of the legitimate classifier. This surrogate model can then be used to guide the generation of spoofed signals during the attack process, without requiring knowledge of the exact classifier.

B. Attack Strategy

Fig. 3 illustrates the procedure of the suggested impersonation attack method. The attacker seeks to execute an impersonation attack where their transmitted signals are recognized as originating from the intended legitimate transmitter. Un-

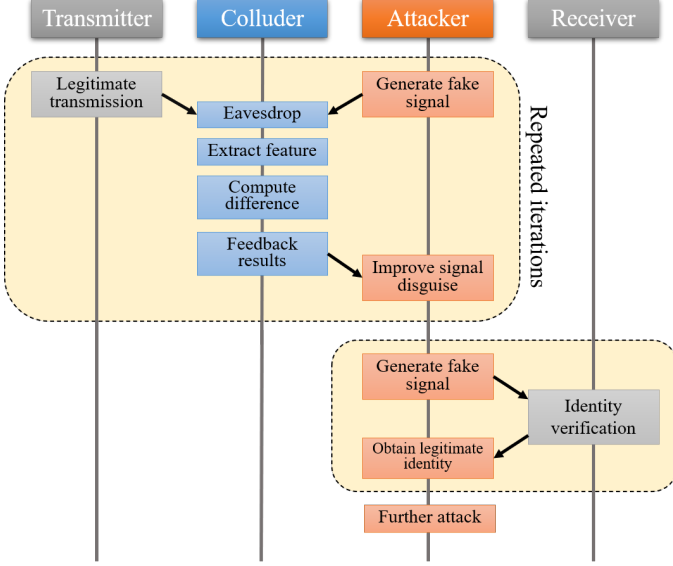


Fig. 3. The workflow of the proposed collusion-driven impersonation attack.

like traditional RF-level spoofing, which replicates hardware-induced distortions, the attacker aims to modify its signal to ensure that the spoofed signal exhibits a similar feature representation to that of the target signal. Accordingly, the attacker must learn the latent feature distribution associated with the target transmitter's transmissions and then iteratively adapt its own signal to conform to this representation, thereby realizing the impersonation attack.

Considering the different channel conditions, the attacker cannot confirm the effect of its own hardware imperfections and channel variations; a colluder works with the attacker to help improve the way the attacker spoofs its RFF features. The colluder can be distant from the legitimate receiver. It captures the altered signal from the adversary and monitors the transmission from the target. The colluder then extracts features akin to the legitimate receiver and examines the differences. Given that the colluder utilizes the same classifier as the legitimate receiver, it can discern the receiver's classification outcome for the spoofed signal and subsequently send critical details regarding the inconsistencies to the attacker. With the feedback, the attacker can improve its signal disguise to minimize the feature distance between its own and the target's RFF representation. Through an iterative process, the attacker can gradually achieve a high RFF feature similarity to the target transmitter.

After finalizing the preparations for the attack, the attacker transmits optimized spoofed signals to the receiver. If the impersonation attack is successful, the receiver will misidentify the attacker's transmission as originating from the target transmitter and subsequently return the classification result to the attacker. Thereafter, the attacker can exploit the misclassifications to launch further attacks as a legitimate user.

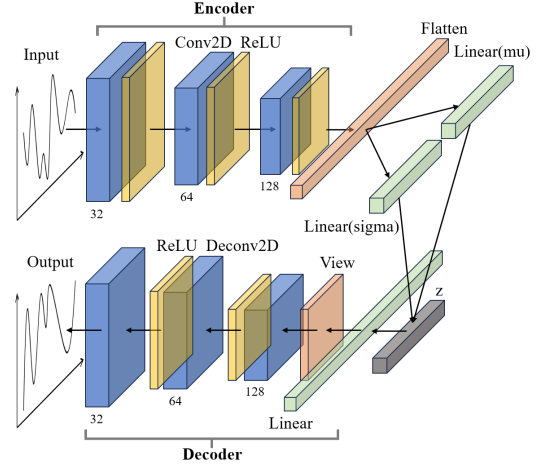


Fig. 4. Proposed spoofed signal generation network based on VAE model.

V. VAE-BASED IMPERSONATION ATTACK FRAMEWORK AGAINST CHANNEL-RESISTANT RFF

This section first presents the overall framework of the proposed impersonation attack and then derives the multi-objective loss function that drives the attack.

A. Attack Implementation Framework

1) *Spoofed Signal Generation Network*: The two mainstream generative models for wireless signal spoofing are GAN and VAE. The previous work of [12] adopted GAN, yet we opt for VAE. In a GAN, the generator depends on the discriminator's feedback. However, in RF fingerprints, channel robust features are so unique that the discriminator soon gains overconfidence, ceasing to offer valuable guidance and leading to stalled training. The VAE circumvents this by utilizing a defined loss function, combining reconstruction error with Kullback-Leibler (KL) regularization, which ensures a consistent gradient during training. Consequently, the VAE continuously aligns the spoofed signal with the target device's RF fingerprint, avoiding the premature convergence or training failure typical with GANs.

Therefore, as depicted in Fig. 4, we develop a spoofed signal generation network utilizing the VAE model. The proposed network is composed of an encoder, decoder, and a latent sampling module. The encoder includes three convolutional layers with (1×5) kernels, stride 2, and ReLU activations, which progressively compress the input into a compact latent space. Two fully connected layers produce the mean and log-variance of a 128-dimensional latent vector. The decoder operates as the inverse of the encoder, beginning with a fully connected layer and followed by three transposed convolutional layers to reconstruct the input's original shape.

2) *CLPS Feature Extraction and Classifier*: Algorithm 1 outlines the method for CLPS extraction, used by both the legitimate receiver and the colluder, with the goal of isolating channel-resistant RFF. Section II discusses how these CLPS

Algorithm 1 Extract CLPS feature**Require:** Received signal \mathbf{R} **Ensure:** Channel-resistant RFF CLPS

- 1: $\mathbf{ACF} \leftarrow \text{Autocorrelation}(\mathbf{R})$
- 2: $\mathbf{P}_{\text{spec}} \leftarrow |\mathcal{FFT}(\mathbf{ACF})|^2$
- 3: $\mathbf{P}_{\text{log}} \leftarrow \log_{10}(\mathbf{P}_{\text{spec}})$
- 4: $\mathbf{CLPS} \leftarrow \mathbf{P}_{\text{log}} - \text{Mean}(\mathbf{P}_{\text{log}})$
- 5: **return** CLPS

features, derived from the signal's preamble, mainly differ due to the device's hardware imperfections rather than data content or channel variability. These robust features allow the colluder to evaluate feature discrepancies based on impersonated and genuine signals consistently with the legitimate receiver.

We utilize a conventional ResNet-based classifier for both the receiver and colluder, adhering to the architecture detailed in [15].

B. Multi-objective Loss functions

To facilitate successful impersonation attacks via the VAE-based framework, we construct a multi-objective loss function that balances waveform naturalness, feature alignment, and adversarial deception. This aims to generate signals that preserve legitimate waveform structures, closely mimic the target device's CLPS features, and effectively trick the classifier. The detailed loss function is as follows:

$$\text{Loss} = \lambda_{\text{recon}} * \text{loss}_{\text{recon}} + \lambda_{KL} * \text{loss}_{KL} + \lambda_{CLPS} * \text{loss}_{CLPS} + \lambda_{Cls} * \text{loss}_{Cls} \quad (4)$$

where λ are hyperparameters for balancing each loss term. In our experiment, they are set to 2.0, 0.1, 1.0, and 0.5, respectively.

1) *Reconstruction Loss* ($\text{loss}_{\text{recon}}$): Measures the mean squared error (MSE) between the altered baseband signals and the initial input signals. This component guarantees that the spoofed signals preserve appropriate temporal characteristics, thereby avoiding the generation of aberrant samples that could arise when optimizing solely based on other loss functions. The loss is calculated as:

$$\text{loss}_{\text{recon}} = \|r(t)_{\text{recon}} - r(t)_{\text{input}}\|_2^2 \quad (4a)$$

2) *KL Divergence Loss* (loss_{KL}): Regularizes the latent space distribution by minimizing the KL divergence between the learned distribution and a standard normal distribution. This term primarily serves to maintain training stability rather than strictly enforcing output samples to follow a particular distribution. The loss is calculated as:

$$\text{loss}_{KL} = -0.5 \sum (1 + \log(\sigma^2) - \mu^2 - \sigma^2) \quad (4b)$$

3) *CLPS Loss* (loss_{CLPS}): Computes the mean squared error between the CLPS features of reconstructed signals and target CLPS features. This term ensures the spoofed signals'

TABLE II
SUMMARY OF SIMULATION SETUP

Component	Setting / Value
Devices	10 legitimate + 1 attacker
I/Q Imbalance	Fixed per device, gain $\in [-0.3, 0.3]$, phase $\in [-15^\circ, 15^\circ]$
Channel Models	AWGN, Rician, Rayleigh
Sample Length	5120 complex I/Q samples
Dataset Size	55000 training samples + 5500 test samples

TABLE III
DEFAULT TRAINING PARAMETERS

Parameter	Setting / Value
Channel Type	Rician + AWGN
Delay Vector (ns)	[0, 50, 110, 170, 290, 310]
Gain Vector (dB)	[0, -3, -10, -18, -26, -32]
K Factor	5
Max Doppler Shift (Hz)	10
E_b/N_0 (dB)	10

CLPS closely match those of the target device. The loss is calculated as:

$$\text{loss}_{CLPS} = \|\text{CLPS}_{\text{recon}} - \text{CLPS}_{\text{target}}\|_2^2 \quad (4c)$$

4) *Classification Loss* (loss_{Cls}): Uses cross-entropy loss to measure how effectively the spoofed signals deceive the target classifier into predicting the desired device label. This adversarial term drives the VAE to produce signals that successfully fool the final classification decision. The loss is calculated as:

$$\text{loss}_{Cls} = \text{CrossEntropy}(y_{\text{recon}}, y_{\text{target}}) \quad (4d)$$

With the multi-objective loss fully specified, the design of the impersonation attack is now complete.

VI. PERFORMANCE EVALUATION

This section details the experimental setup and presents the results alongside comparative evaluations.

A. Simulation Setup

The simulation settings are summarized in Table II. The specific explanation of the simulation setup is as follows.

1) *Device Configuration*: The simulation involves 10 legitimate transmitters and a single attacker, each characterized by distinct I/Q imbalance values to simulate hardware-specific distortions. Every device employs a fixed baseband preamble per IEEE 802.15.4 standards and transmits 5120-sample signal segments. The gain and phase imbalance parameters, based on realistic hardware constraints, align with previous research [12].

2) *Channel Configuration*: We examine robustness across various propagation scenarios by simulating four typical environments: AWGN, Indoor Office Channel A/B, and Vehicular Channel A, utilizing ITU-R M.1225 models. Channels are synthesized using MATLAB's `comm.AWGNChannel` and `comm.RicianChannel` functions, featuring independent

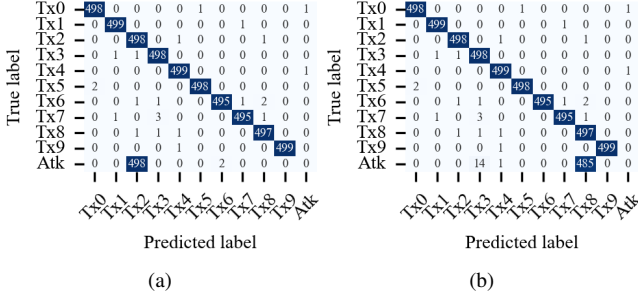


Fig. 5. Classification result under the proposed collusion-driven impersonation attack. (a) Target on legitimate transmitter 2. (b) Target on legitimate transmitter 8.

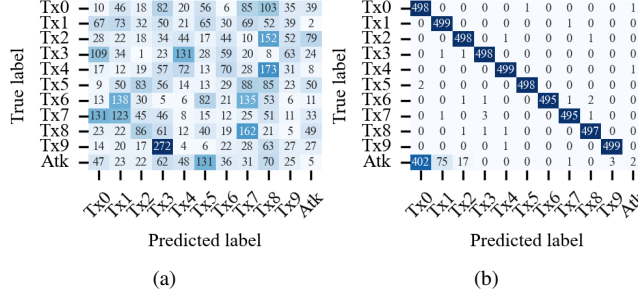


Fig. 6. Classification result of the existing attack strategy [12], target on Tx2. (a) Classification result when targeting I/Q data. (b) Classification result when using GAN.

pathways for each entity. The default training configuration employs Rician fading with a K factor of 5, a six-tap multipath profile, and $E_b/N_0 = 10$ dB, as outlined in Table III.

3) *Dataset Description*: The dataset is composed of simulated generated signals. Each device produces 5000 training segments and 500 testing segments, each with 5120 I/Q samples. CLPS features are extracted from each segment to train a ResNet-based classifier. To generate spoofed signals, the attacker feeds perturbed baseband preambles into its VAE model. The colluder supplies the target CLPS features and classifier outputs required for loss calculation during training.

B. Effectiveness of the Proposed Impersonation Attack

We first assess the proposed collusion-driven attack's efficacy by analyzing classification results from the legitimate classifier, as illustrated in Fig. 5. For visualization purposes, Tx 2 and 8 were chosen at random. Fig. 5 shows that the legitimate classifier accurately identifies legitimate devices, while spoofed signals aimed at specific devices are classified as the intended target. This implies that the attacker's spoofed signals possess the RFF characteristics of the target device, leading to a high likelihood of misclassification across various devices, indicating a probable attack success.

C. Comparison with Existing Attack Strategies

We evaluate our approach in relation to current attack methods, specifically where attackers alter raw I/Q data via a GAN-based network [12]. Experiments involve varying attack targets

True label \ Predicted label	AWGN	Indoor A	Indoor B	Vehicular
AWGN	99.94	99.99	99.96	99.99
Indoor A	84.33	98.14	99.96	99.98
Indoor B	53.66	99.68	99.92	98.98
Vehicular	42.19	94.08	94.89	99.38

Fig. 7. Cross-channel mean attack success rate.

(I/Q data and CLPS) and generation networks (GAN and VAE), validating the effectiveness of our attack framework.

1) *I/Q-Domain Attacks*: To evaluate the effectiveness of channel-resistant RFF modeling, we conduct a baseline experiment using raw I/Q samples as the input feature, without the CLPS extraction stage. Both the receiver and the colluder directly operate on I/Q data, and the spoofed signal is generated by a VAE trained to mimic the I/Q characteristics of the target transmitter. As shown in Fig. 6(a), this approach fails to generate effective spoofed signals, resulting in almost random classification at the receiver.

The failure arises because I/Q data is extremely susceptible to channel variations. Lacking channel-resistant RFF, colluders obtain unstable features, which hinders them from offering reliable feedback for spoofed signal generation. This instability obstructs model convergence, making I/Q-level impersonation impractical in realistic multipath environments. It underscores the need for channel-resistant features such as CLPS for effective attack implementation.

2) *GAN-Based Signal Generation*: To further assess the attack design, we compare the proposed VAE-based generation method with the GAN-based approach. Both models aim to generate spoofed signals whose CLPS features match the target. As shown in Fig. 6(b), the GAN model fails to achieve effective impersonation, with spoofed signals consistently misclassified as a different device (e.g., Tx0 instead of the intended Tx2).

The GAN's performance declines due to high similarity in CLPS samples, causing fast discriminator convergence and resulting in vanishing gradients for the generator, thereby obstructing training. Conversely, VAE training uses explicit reconstruction and feature-aligned losses for consistent network updates, allowing for a smooth latent space and improved approximation of the target device's CLPS distribution. Thus, the proposed VAE framework improves spoofing stability and effectiveness, proving superior in generating high-fidelity RFF features for impersonation in channel-resistant modeling.

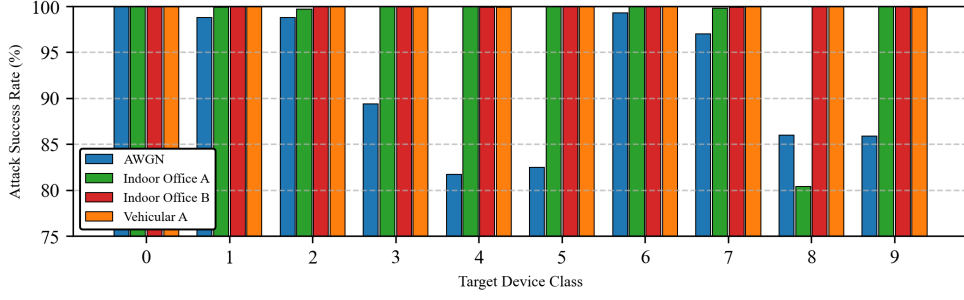


Fig. 8. Attack success rates across target devices under models trained under different channels.

D. Attack Success Rate in Complex Channel Conditions

The proposed attack's effectiveness is assessed under intricate channel conditions. The attack success rate (ASR), defined as the proportion of signals mistakenly identified as legitimate targets to those transmitted by the attacker, is calculated. Evaluation begins with varied channel types and parameters.

1) *Performance under Different Channel Types:* We trained attack models for each of the four channel conditions and conducted comprehensive cross-channel evaluations. The results in Fig. 7 indicate that, except for the model trained under the AWGN-Only channel, the other attack models demonstrate high resilience across different channel conditions. Notably, the model trained under the Vehicular channel condition achieves the best overall performance. In contrast, models trained under the Indoor channel condition show a slight drop in ASR when tested on the Vehicular channel. This is because the Vehicular channel presents a more complex multipath environment than the Indoor channel. As a result, a model trained under Vehicular conditions can better adapt to more challenging channel scenarios, thereby achieving superior performance. These results also reflect the impact of mismatched channel conditions between the colluder and the receiver, indicating that the impersonation attack performance degrades as their channel divergence increases.

To further understand the impact of channel mismatch and device-specific factors on spoofing effectiveness, Fig. 8 presents a detailed breakdown of attack success rates across different devices, where the test channel is fixed to Indoor Office Channel A. The results show a consistent trend with Fig. 7, in that models trained under all channel conditions except AWGN maintain high ASR across most target devices—often nearing 100%. Particularly, models trained under Vehicular and Indoor Office Channel B conditions continue to achieve strong performance, consistent with their ability to generalize better in the presence of multipath fading and other channel variations.

However, a notable performance drop is observed for attack models trained under AWGN and Indoor Office Channel A when evaluated across different devices, with the ASR decreasing to around 80% in certain cases. This drop is primarily due to increased signal distortion under mismatched channel conditions, which reduces the similarity between the crafted spoofed signal and the target signal. Additionally, in the sim-

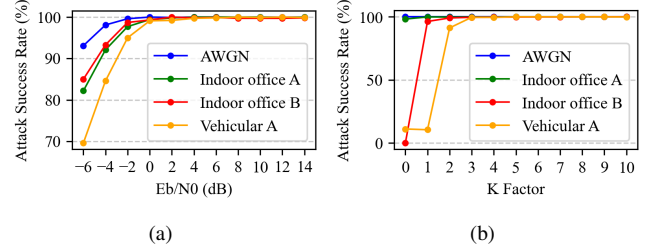


Fig. 9. ASR performance under different channel parameters. (a) ASR vs E_b/N_0 under different channels. (b) ASR vs K-factor under different channels.

ulation setup, hardware impairments are limited to modeling IQ imbalance only. As a result, devices with similar IQ imbalance parameters may become less distinguishable, making them harder to spoof or more susceptible to misclassification, thereby reducing the attack success rate for those specific model-device pairs. This highlights the need to consider a wider range of hardware features in modeling device-specific characteristics to further improve spoofing robustness.

2) *Impact of E_b/N_0 :* We evaluate the attack performance in different SNR, the result is shown in Fig. 9(a). From the figure, it can be observed that the success rate of impersonation attacks increases with the rise of E_b/N_0 . This is in line with expectations, as a higher E_b/N_0 indicates that the legitimate receiver experiences less distortion in the received spoofed signals.

Furthermore, it can be observed that channels A and B maintain a generally consistent attack rate under different E_b/N_0 conditions. Although the vehicular channel performs relatively poorly at lower E_b/N_0 levels, the attack rate under this channel condition quickly approaches 100% as E_b/N_0 increases. This indicates that the proposed attack has relatively weak channel robustness at low E_b/N_0 levels, but exhibits good channel robustness as E_b/N_0 rises (for $E_b/N_0 \geq 0$).

3) *Impact of K-factor:* To validate the robustness of the proposed attack, we assess its performance over different K-Factors (e.g., $K = [0, 1, \dots, 10]$). The results are presented in Fig. 9(b). This figure shows that the attack success rate (ASR) in the Indoor Office Channel A is stable and comparable to that in an AWGN channel across varying K-factors. Notably, the ASR is approximately 97.40% at K

= 0 and rises to nearly 100.0% as K grows to 10. Using MATLAB's comm.RicianChannel function, a K-factor of zero simulates a Rayleigh fading model. The findings highlight the attack's robustness to variations in the K-factor, ensuring steady performance under line-of-sight (LoS, $K > 0$) and non-line-of-sight (NLoS, $K = 0$) conditions, assuming consistent channel characteristics.

Conversely, in Indoor Office Channel B and Vehicular Channel A, attack effectiveness declines markedly at low K-factors. For example, in Indoor Office Channel B, the ASR is merely 0.20% at $K=0$, but climbs to 100.0% with improved channel conditions. Similarly, in Vehicular Channel A, the ASR reduces to 12.40% at $K=0$, escalating to 99.80% at $K=10$. This is due to the lack of a direct LoS path in NLoS scenarios, where signal propagation depends fully on multipath reflections, causing significant fading and distortion. The steep degradation in NLoS conditions underscores the sensitivity of these channels to specific multipath variables. However, under LoS conditions, the proposed attack remains resilient to changes in the K-factor.

VII. CONCLUSION

In this paper, we present a collusion-driven impersonation attack targeting channel-resistant RFF systems. By incorporating CLPS as a channel-invariant feature and introducing a VAE-based spoofed signal generation network, the attacker is able to mimic the target device's RF fingerprint under complex wireless channel conditions. Leveraging a colluder that shares the same feature extractor and classifier as the legitimate receiver, the attacker iteratively optimizes signal generation to minimize discrepancies in both feature space and classification outcome. Extensive simulations across diverse scenarios—including AWGN, multipath fading, and Doppler shift—demonstrate that the proposed attack achieves stable success rates ($>95\%$) across fading channels and SNR conditions. These results reveal the practical feasibility and significant threat posed by collusion-driven impersonation attack in realistic RFF systems with channel-resilient features.

Future work may explore adversarial training techniques to improve the classifier's robustness against synthesized spoofed samples. Additionally, active defense strategies at the legitimate transmitter, such as embedding recoverable perturbations known only to the intended receiver, may offer effective protection by limiting the attacker's ability to accurately observe and replicate legitimate transmissions.

ACKNOWLEDGMENT

This work was supported in part by the Frontier Technologies Research and Development Program of Jiangsu under Grant BF2024065; in part by the National Natural Science Foundation of China under Grant 62171121 and Grant U22A2001; in part by the Natural Science Foundation on Frontier Leading Technology Basic Research Project of Jiangsu under Grant BK20222001.

REFERENCES

- [1] Chaimae Saadouni, Saad El Jaouhari, Nouredine Tamani, Soumia Ziti, Lina Mroueh, and Karim El Bouchti. Identification techniques in the internet of things: Survey, taxonomy and research frontier. *IEEE Communications Surveys & Tutorials*, pages 1–1, 2025.
- [2] Naeimeh Soltanieh, Yaser Norouzi, Yang Yang, and Nema Chandra Karmakar. A review of radio frequency fingerprinting techniques. *IEEE Journal of Radio Frequency Identification*, 4(3):222–233, 2020.
- [3] Jiabao Yu, Aiqun Hu, Guyue Li, and Linning Peng. A robust RF fingerprinting approach using multisampling convolutional neural network. *IEEE Internet of Things Journal*, 6(4):6786–6799, 2019.
- [4] Jie Ma, Junqing Zhang, Guanxiong Shen, Alan Marshall, and Chip-Hong Chang. White-box adversarial attacks on deep learning-based radio frequency fingerprint identification. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pages 3714–3719, Rome, Italy, May 2023.
- [5] Boyang Liu, Haoran Zhang, Yiyao Wan, Fuhui Zhou, Qihui Wu, and Derrick Wing Kwan Ng. Robust adversarial attacks on deep learning-based RF fingerprint identification. *IEEE Wireless Communications Letters*, 12(6):1037–1041, 2023.
- [6] Yunsong Huang, Weicheng Liu, and Hui-Ming Wang. Hidden backdoor attack: A new threat to learning-aided physical layer authentication. In *Proceedings of the International Conference on Ubiquitous Communications (UCOM)*, pages 310–314, Xi'an, China, July 2023.
- [7] Boris Danev, Heinrich Lueken, Srdjan Capkun, and Karim El Defrawy. Attacks on physical-layer identification. In *Proceedings of the Third ACM Conference on Wireless Network Security (WiSec)*, pages 89–98, Hoboken, NJ, USA, March 2010. ACM.
- [8] Kaisheng Liang and Bin Xiao. Styles: Boosting the transferability of adversarial examples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8163–8172, 2023.
- [9] Kevin Merchant and Bryan Nossain. Securing IoT RF fingerprinting systems with generative adversarial networks. In *Proceedings of the IEEE Military Communications Conference (MILCOM)*, pages 584–589, Norfolk, VA, USA, November 2019.
- [10] Samudhi Karunaratne, Enes Krijestorac, and Danijela Cabric. Penetrating RF fingerprinting-based authentication with a generative adversarial attack. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pages 1–6, Montreal, QC, Canada, June 2021.
- [11] Xuelong Dai, Kaisheng Liang, and Bin Xiao. Advdiff: Generating unrestricted adversarial examples using diffusion models. In *European Conference on Computer Vision (ECCV)*, pages 93–109, Milan, Italy, October 2024.
- [12] Yuxuan Xu, Ming Liu, Linning Peng, Junqing Zhang, and Yawen Zheng. Colluding RF fingerprint impersonation attack based on generative adversarial network. In *Proceedings of the IEEE International Conference on Communications (ICC)*, pages 3220–3225, Seoul, South Korea, May 2022.
- [13] Guanxiong Shen, Junqing Zhang, Alan Marshall, and Joseph R Cavallaro. Towards scalable and channel-robust radio frequency fingerprint identification for LoRa. *IEEE Transactions on Information Forensics and Security*, 17:774–787, 2022.
- [14] Abdullahi Mohammad, Mateen Ashraf, Mikko Valkama, and Bo Tan. Learning-based RF fingerprinting for device identification using amplitude-phase spectrograms. In *Proceedings of the 98th Vehicular Technology Conference (VTC)*, pages 1–6, Hong Kong, China, October 2023.
- [15] Peng Tang, Guoru Ding, Yitao Xu, Yutao Jiao, Yehui Song, and Guofeng Wei. Causal learning for robust specific emitter identification over unknown channel statistics. *IEEE Transactions on Information Forensics and Security*, 19:5316–5329, 2024.
- [16] Ning Yang, Bangning Zhang, Guoru Ding, Yimin Wei, Guofeng Wei, Jian Wang, and Daoxing Guo. Specific emitter identification with limited samples: A model-agnostic meta-learning approach. *IEEE Communications Letters*, 26(2):345–349, 2021.
- [17] Daryna Oliynyk, Rudolf Mayer, and Andreas Rauber. I know what you trained last summer: A survey on stealing machine learning models and defences. *ACM Computing Surveys*, 55(14s):1–41, 2023.