Stay Connected, Leave no Trace: Enhancing Security and Privacy in WiFi via Obfuscating Radiometric Fingerprints

LUIS F. ABANTO-LEON*, TU Darmstadt, Secure Mobile Networking Lab (SEEMOO), Germany
ANDREAS BÄUML*, TU Darmstadt, Secure Mobile Networking Lab (SEEMOO), Germany
GEK HONG (ALLYSON) SIM, TU Darmstadt, Secure Mobile Networking Lab (SEEMOO), Germany
MATTHIAS HOLLICK, TU Darmstadt, Secure Mobile Networking Lab (SEEMOO), Germany
ARASH ASADI, TU Darmstadt, Wireless Communication and Sensing Lab (WiSe) & SEEMOO, Germany

The intrinsic hardware imperfection of WiFi chipsets manifests itself in the transmitted signal, leading to a unique radiometric fingerprint. This fingerprint can be used as an additional means of authentication to enhance security. In fact, recent works propose practical fingerprinting solutions that can be readily implemented in commercial-off-the-shelf devices. In this paper, we prove analytically and experimentally that these solutions are highly vulnerable to impersonation attacks. We also demonstrate that such a unique device-based signature can be abused to violate privacy by tracking the user device, and, as of today, users do not have any means to prevent such privacy attacks other than turning off the device.

We propose `RF-Veil`, ***a radiometric fingerprinting solution that not only is robust against impersonation attacks but also protects user privacy by obfuscating the radiometric fingerprint of the transmitter for non-legitimate receivers***. Specifically, we introduce a ***randomized pattern of phase errors*** to the transmitted signal such that only the intended receiver can extract the original fingerprint of the transmitter. In a series of experiments and analyses, we expose the vulnerability of adopting naive randomization to statistical attacks and introduce countermeasures. Finally, we show the efficacy of `RF-Veil` experimentally in protecting user privacy and enhancing security. More importantly, our proposed solution allows communicating with other devices, which do not employ `RF-Veil`.

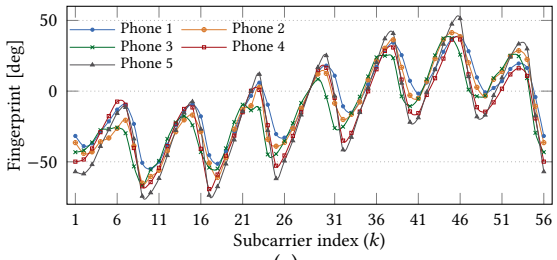
ACM Reference Format:
Luis F. Abanto-Leon, Andreas Bäuml, Gek Hong (Allyson) Sim, Matthias Hollick, and Arash Asadi. 2020. Stay Connected, Leave no Trace: Enhancing Security and Privacy in WiFi via Obfuscating Radiometric Fingerprints. In , Vol. 1, 1 (November 2020). ACM, New York, NY. 30 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The omnipresence of WiFi devices in our daily lives demands *strong and quantifiable security and privacy* mechanisms to protect us from attackers. WiFi security mechanisms traditionally reside above the physical layer. This can be augmented by using physical layer characteristics (e.g., channel fading, interference, hardware impairments), which further enhance the security of WiFi. In fact, physical layer security gained momentum after a chain of acute vulnerabilities rendered these *high-layer* security mechanisms unsecure. This includes the disastrous RC4 vulnerability in WEP [12] as well as the more recent attacks on WPA2 (e.g., KRACK [37] and Kr00k [7]). We have also witnessed a variety of masquerading attacks in which the adversary mounts a machine-in-the-middle (MitM) attack by creating a rogue access point (AP), mimicking the identity (i.e., SSID) of

*Both authors contributed equally to this research.

XXXX-XXXX/2020/11-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn



(a)

Actual \ Predicted	Phone 1	Phone 2	Phone 3	Phone 4	Phone 5
Phone 1	0.985	0.0	0.0	0.0	0.0
Phone 2	0.0	0.985	0.085	0.0	0.0
Phone 3	0.0	0.115	0.965	0.0	0.0
Phone 4	0.0	0.0	0.025	1.0	0.0
Phone 5	0.010	0.0	0.0	0.0	0.995

(b)

Fig. 1. WiFi radiometric fingerprints of 5 identical phone (Samsung Galaxy S6). Fig. 1a shows that the *fingerprints differ from one another even though the chipsets belong to the same series and manufacturer*, thus allowing to distinguish among multifarious devices. Fig. 1b shows that *the devices can be distinguished with* 96.5% *accuracy using a simple mean absolute error (MAE)-based classifier* (MAE threshold = 4.5°).

a legitimate AP. It has been shown that physical layer security, in particular, radiometric (radio frequency) fingerprinting can thwart such attacks [5, 21, 23].

Radiometric fingerprinting techniques rely on measuring and extracting device-specific imperfections of the transmitter RF circuitry embedded in the emitted signal, which manifest in form of negligible but distinguishable errors, e.g., in phase (e.g., [23]) or frequency (e.g., [18]). These imperfections are so individualized that even chipsets from the same manufacturer have different fingerprints [5, 23]. In Fig. 1a, we demonstrate that the radiometric fingerprint[1] of five identical phones with the same WiFi chipset are visually distinguishable. Thus, it is not surprising that these devices can be easily differentiated from one another with high success ratio (i.e., 96.5%). Such degree of accuracy, on the one hand, reveals the *potential of radiometric fingerprinting for achieving accurate authentication, thus enhancing security.* On the other hand, it raises *major privacy concerns since adversaries can locate/track devices* using these unique fingerprints. Our work is motivated by the potential of radiometric fingerprinting in coping with security and privacy challenges.

Challenge I: Privacy. Any unique identifier which can be easily measured/accessed by an adversary poses a significant privacy threat. Indeed, this is the motivation behind MAC address randomization in WiFi or temporary identifiers in cellular networks to prevent potential adversaries from tracking users. Radiometric fingerprints expose users to the same privacy vulnerability, and as of today, users do not have any means to prevent such privacy attacks other than turning off the device. While randomizing the physical layer characteristics of the signal is a plausible solution to enhance privacy, such procedure may degrade the communication link and disrupt or prevent legitimate radiometric fingerprinting, which brings us to the next challenges.

Challenge II: Security. Radiometric fingerprints are typically considered a secure anchor for device authentication. Still, they are *collectible by anyone in the vicinity* of the transmitter who is capable of "overhearing" the packets, e.g., 50-100 meters for WiFi. This *exposes the fingerprinting methods to impersonation attacks.* Initial proposals argued that the cost of mimicking the fingerprints is too high [31]. To date, a wide range of software-defined radios (SDRs) costing from a few hundred (up to a few thousand) euros can collect and forge the fingerprints of other devices, e.g., through modifying the phases of emitted signals, as shown in Section 2.2. This issue is further exacerbated by the emergence of WiFi firmware patching tools [34], which enables commercial WiFi chipsets to shape their signals and impersonate other devices.

Challenge III: Allowing for legitimate radiometric fingerprinting. There are several solutions to hide ones' fingerprint: *(i) Jamming*, which defeats the primary purpose of WiFi, i.e., communication; *(ii) Constructive interference.* The seminal work of Oh *et al.* on location privacy [25]

[1]These fingerprints are extracted from non-linear phase errors derived from device-specific hardware imperfections [23]

and recent literature on privacy against WiFi sensing [26, 40] use coordinated transmissions or a secondary signal repeater to obfuscate the physical layer information, which are not scalable and can be costly due to reliance on secondary devices; *(iii) Fingerprint randomization at the transmitter* has the advantage of scalability, but it can disrupt the communication link by distorting the channel estimation at the receiver. In [8], the authors randomize the transmitted signal to obfuscate device-free localization but their approach introduces marginal impact on the quality of the communication. Furthermore, we must ensure that the randomization is reversible to allow legitimate fingerprinting.

1.1 Our approach

In this paper, we propose `RF-Veil`, a scalable approach that enhances the user privacy by obfuscating the radiometric fingerprints of the device from adversaries while allowing the use of channel state information (CSI)-based fingerprinting at legitimate receivers to strengthen the security of the network.

In essence, `RF-Veil` *adds a crafted randomized phase noise to the signal at the transmitter such that the radiometric fingerprints are obfuscated, but the quality of communication remains intact. Furthermore, we facilitate fingerprint extraction through a low-overhead synchronized random noise generation process between legitimate transmitters and receivers.* The properties of `RF-Veil` are:

Privacy-preserving. The latest radiometric fingerprinting solutions extract device-specific phase errors from the CSI [18, 23]. `RF-Veil` introduces *deliberate phase noise to the subcarriers in the OFDM symbols* on a per-frame basis such that the adversary can no longer estimate the actual radiometric fingerprint by analyzing the CSI, thus preventing the device identification/tracking via radiometric fingerprint.

Secure against impersonation. We strive to maintain the possibility of legitimate fingerprinting without exposing the user to impersonation attacks. To this aim, we first devise a technique (synchronized phase noise generation), which enables only the legitimate receivers to denoise the transmitted signals and extract the original fingerprint. Secondly, we apply the obfuscation on a per-frame basis to eliminate the possibility of impersonation or reply attack via over-the-air packet sniffing. The effectiveness of this method is proven both theoretically and experimentally, even in presence of sophisticated adversaries with the capability of realizing statistical attacks.

Dual mode. `RF-Veil` is designed to allow the legitimate use of wireless fingerprinting techniques (e.g., for authentication as in [23]) in presence of our obfuscation method. Furthermore, a reduced form of `RF-Veil` can be used to obfuscate the fingerprint of the device in order to only protect the device's privacy when fingerprinting is not used as an additional security feature, i.e., reversing the phase noise is not required. *We refer to this second operational mode as* `RF-Veil-Standalone`. In this mode, we can hide the fingerprint of the transmitter by executing the obfuscation blocks without any handshake or coordination with other receivers. As a result, we can ensure privacy protection in a much broader scenario, e.g., communicating with non-`RF-Veil`-enabled devices, in absence of any active connections, or in connection establishment phase.

Low-overhead and scalable. `RF-Veil` has low overhead from both computational and signaling/control message perspective. Our simple yet effective obfuscation technique enabled extraction of CSI-based radiometric fingerprints at the legitimate receiver without any additional complex signal processing. Furthermore, `RF-Veil` is highly scalable since it is implemented directly at the transmitter and does not rely on any secondary device [26, 40]. Therefore, any WiFi device can obfuscate its fingerprint easily and independently.

Fig. 2. IEEE 802.11ac PHY frame format.

1.2 Our contributions

To the best of our knowledge, *this is the first work exposing privacy and security vulnerabilities of radiometric fingerprints as well as devising practical methods to resolve them.* Note that prior works such as [27, 28] propose techniques for preventing the exposure of unencrypted fields (e.g., headers and payload information) to counter, for instance, reactive jamming attacks that can adapt to the rate of transmission. While the approach prevents data analysis and acquisition of transmission cues, it does not protect the radiometric fingerprint of the device. The following summarizes our main contributions: *(i)* Showing vulnerabilities of recent CSI-based radiometric fingerprinting solutions [5, 23, 38] to impersonation attacks (Section 3); *(ii)* Proposing a method for injecting artificial noise to the fingerprint without impacting communication quality (Section 4). *(iii)* Designing `RF-Scope`, a benchmarking tool to assess the effectiveness of radiometric fingerprint obfuscation against statistical attacks (Section 5). Specifically, `RF-Scope` is a maximum-likelihood-based estimator of the CSI, which we prove to be near-optimal through derivation of Cramer-Rao bounds (Appendix B); *(iv)* Devising `RF-Veil`, a fingerprint obfuscation framework that circumvents the privacy issues without impacting the communication quality (Section 6). *(v)* We prove the efficacy of our proposals, both analytically and experimentally.

2 FINGERPRINTING PRIMER

Radio signal analysis to identify devices and distinguish between friends and foes dates back to the time of the Vietnam war. In the same line, *radiometric fingerprinting* has gained momentum in recent years with the surge of attacks that leverage hardware impairments to breach privacy and security in wireless networks. Recently, CSI-based radiometric fingerprinting gained popularity due to the availability of CSI extraction tools [14, 17, 38]. These tools allow per-frame CSI collection from commercial WiFi chipsets (e.g., Intel, Qualcomm, Broadcom), making CSI-based fingerprinting practical and feasible for all devices. In the following, a short overview of CSI estimation and CSI-based fingerprinting is provided.

2.1 Channel estimation in WiFi

As a prelude to CSI-based fingerprinting, we describe channel estimation in WiFi throughout this section. Fig. 2 shows the IEEE 802.11ac PHY frame structure, wherein we recognize four distinct fields: short training field (STF), long training field (LTF), SIG, and DATA (cf. 17.3 in [35]). The receiver uses the STF field for signal detection, automatic gain control, time synchronization, and coarse carrier frequency offset (CFO). The LTF field is employed for fine CFO estimation and channel estimation. Channel estimation is performed by sending BPSK pilots over the LTF subcarriers of two consecutive OFDM symbols. The SIG and DATA fields convey the MCS level and the payload, respectively. The OFDM symbols in the SIG and DATA fields are equalized using the channel estimated by the preceding LTF field. The prefix "L-" denotes the legacy fields, which are included for compatibility with IEEE 802.11a.

Let K denote the number of subcarriers and $\mathbf{s} = [s_1, \cdots, s_K]^T \in \mathbb{C}^{K\times 1}$ the BPSK pilot symbols (defined in Equation 19-23 of [35]). Also, let $\mathbf{F} \in \mathbb{C}^{K\times K}$ and $\mathbf{F}^H \in \mathbb{C}^{K\times K}$ denote the discrete Fourier transform (DFT) matrix and the inverse DFT (IDFT) matrix, respectively. Moreover, $\mathbf{F}\mathbf{F}^H = \mathbf{I}$, with $(\cdot)^H$ representing the Hermitian transpose. The discrete-time OFDM symbol is given by $\mathbf{x} = \mathbf{F}^H\mathbf{s}$ [1]. In order to improve the signal robustness against multi-path interference, the periodic OFDM

symbol $\tilde{\mathbf{x}}$ is produced by appending the cyclic prefix (CP) to $\mathbf{x}$. The CP consists of the last L samples of $\mathbf{x}$, thus $\tilde{\mathbf{x}} = \left[\mathbf{x}^T_{[K-L+1:K]}, \mathbf{x}^T\right]^T \in \mathbb{C}^{(K+L)\times 1}$. Appending the CP to $\mathbf{x}$ transforms the linear convolution between $\tilde{\mathbf{x}}$ and the channel $\mathbf{c} = [c_1, \cdots, c_J]^T \in \mathbb{C}^{J\times 1}$ (with $J < L$ paths) into a circular convolution. This has the advantage of simplifying OFDM demodulation and equalization at the receiver. Upon transmitting $\tilde{\mathbf{x}}$ over the channel $\mathbf{c}$, the receiver obtains $\tilde{\mathbf{r}} = \tilde{\mathbf{x}} * \mathbf{c}$, where $*$ denotes the convolution operator. To remove the impact of inter-block interference between adjacent LTF frames (caused by multi-path propagation), we discard the first L elements of $\tilde{\mathbf{r}}$, thus yielding $\mathbf{r} = \tilde{\mathbf{r}}_{[L+1:L+K]} \in \mathbb{C}^{K\times 1}$. The received signal $\mathbf{r}$ can be expressed as:

$$\underbrace{\begin{pmatrix} r_1 \\ \vdots \\ r_K \end{pmatrix}}_{\mathbf{r}} = \underbrace{\begin{pmatrix} c_1 & 0 & \cdots & 0 & c_J & \cdots & c_2 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ c_J & c_{J-1} & & 0 & 0 & & 0 \\ 0 & c_J & & 0 & 0 & & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & c_J & c_{J-1} & \cdots & c_1 \end{pmatrix}}_{\mathbf{C}} \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix}}_{\mathbf{x}} + \underbrace{\begin{pmatrix} w_1 \\ \vdots \\ w_K \end{pmatrix}}_{\mathbf{w}}, \tag{1}$$

where $\mathbf{w} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ denotes circularly-symmetric complex Gaussian noise. The receiver demodulates the received signal $\mathbf{r}$ by multiplying it with $\mathbf{F}$ to obtain $\mathbf{y} = \mathbf{Fr} = \mathbf{FCx} + \mathbf{Fw}$. The convolution matrix $\mathbf{C} \in \mathbb{C}^{K\times K}$ is circulant, which is a consequence of adding the CP to the transmitted signal. Circulant matrices can be expressed via eigen-decomposition as $\mathbf{C} = \mathbf{F}^H\mathbf{HF}$, where $\mathbf{H} = \text{diag}([h_1, \cdots, h_K])$ represents a diagonal matrix containing the eigenvalues of $\mathbf{C}$ [13]. As a result, the demodulated signal $\mathbf{y}$ collapses to $\mathbf{y} = \mathbf{F}(\mathbf{F}^H\mathbf{HF})(\mathbf{F}^H\mathbf{s}) + \mathbf{F}^H\mathbf{w} = \mathbf{Hs} + \mathbf{w}$. More specifically,

$$\underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_K \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} h_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_K \end{pmatrix}}_{\mathbf{H}} \underbrace{\begin{pmatrix} s_1 \\ \vdots \\ s_K \end{pmatrix}}_{\mathbf{s}} + \underbrace{\begin{pmatrix} w_1 \\ \vdots \\ w_K \end{pmatrix}}_{\mathbf{w}}.$$

Thus, for any subcarrier $k \in \mathcal{K} = \{1, \cdots, K\}$, the received symbol is expressed as

$$y_k = h_k s_k + w_k = |h_k|\, e^{j\phi_k} s_k + w_k, \tag{2}$$

which shows that the channel affects each pilot symbol s_k by a complex-valued factor $h_k = |h_k|\, e^{j\phi_k}$ and additive noise w_k. Since the pilot symbols $\mathbf{s}$ are known by the receiver, the CSI vector $\mathbf{h} = [h_1, \cdots, h_K]^T$ can be obtained upon equalizing each received symbol y_k with the compensation factor $\frac{s_k^*}{|s_k|^2}$. Thus, the estimated channel in subcarrier k is given by:

$$\tilde{h_k} = h_k s_k \frac{s_k^*}{|s_k|^2} + w_k \frac{s_k^*}{|s_k|^2} = |h_k|\, e^{j\phi_k} + w_k. \tag{3}$$

2.2 CSI-based fingerprinting

CSI-based radiometric fingerprinting techniques consist of analyzing the CSI to extract features that are unique to the transmitting device. Specifically, Zhuo *et al.* [41] found that WiFi chipsets exhibit non-linear phase errors that change across subcarriers and are analogous to a sinusoidal function, as shown in Fig. 1a. These phase errors are caused by I/Q imbalance as a result of hardware imperfections. It was shown in [41] that these errors are latent signatures that can be extracted upon removing the linear phase errors from the CSI. Building on this finding, Liu *et al.* [23] harness these non-linear errors as CSI-based *radiometric fingerprints* for identification, thus preventing impersonation by unauthorized WiFi devices. Following the same notation in (2) and (3), we denote the CSI phases by $\mathbf{\Phi} = [\phi_1, \cdots, \phi_K]^T$, which can be further decomposed into

$$\mathbf{\Phi} = \underbrace{\boldsymbol{\varphi} + \boldsymbol{\omega} + \boldsymbol{\theta} + \boldsymbol{\psi}}_{\text{linear errors}} + \underbrace{\boldsymbol{\epsilon}}_{\text{non-linear error (fingerprint)}}, \tag{4}$$

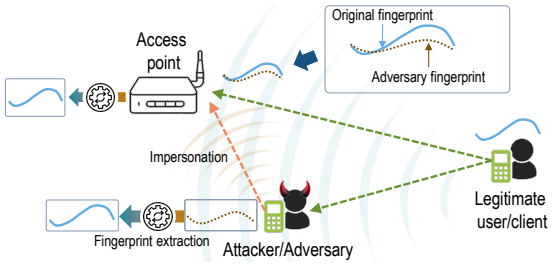


Fig. 3. Adversary model. *We consider an impersonation scenario, where the adversary has captured the fingerprint of the victim (legitimate user). The adversary forges the fingerprint of the victim by introducing additional phase rotations to its own fingerprint.*

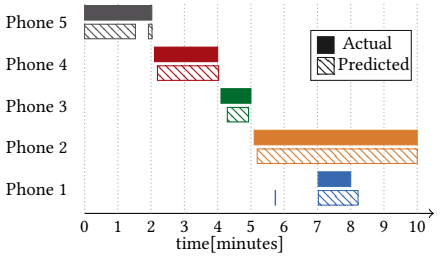


Fig. 4. Adversary prediction on the presence time of 5 different phones turned on and off at different time intervals. *An adversary can accurately determine the presence of a specific user in a network by tracking the radiometric fingerprints.*

where $\boldsymbol{\varphi} \in \mathbb{R}^{K\times 1}$ represents the phase of the signal at the transmitter while $\boldsymbol{\omega} \in \mathbb{R}^{K\times 1}$, $\boldsymbol{\theta} \in \mathbb{R}^{K\times 1}$ and $\boldsymbol{\psi} \in \mathbb{R}^{K\times 1}$ denote the phase errors due to sampling frequency offset, frame detection delay and time of flight, respectively. By using the mirror subcarriers, the linear part of the phase errors can be canceled [23]. Hence, the non-linear phase errors $\boldsymbol{\epsilon} \in \mathbb{R}^{K\times 1}$ are obtained by the following equation

$$\boldsymbol{\epsilon} = \boldsymbol{\Phi} - \left(2\pi\lambda \cdot \mathbf{v} + \mathbf{1}Q^{\star}\right), \tag{5}$$

where $\mathbf{v} = \left[-K/2, \cdots, -1, 1, \cdots, K/2\right]^T$ and λ is a constant used for nullifying the linear phase rotation in a specific frame whereas $Q^{\star}$ is used for phase error normalization [23].

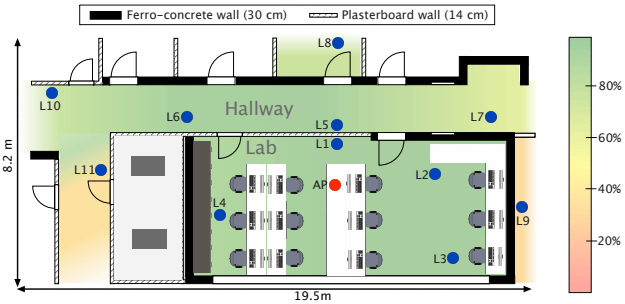
The authors show that the non-linear phase errors exhibit both time and location invariance and change significantly even across devices of the same manufacturer. As a result, these non-linear phase errors can be used as highly distinctive radiometric fingerprints for device identification by leveraging the above described approach in [23]. Even though the difference is very small, the phones are distinguishable from one another, as illustrated in Fig. 1b. As a result, the authors conclude that these fingerprints can be used as countermeasures against impersonation attacks. However, we show in the next section that impersonation is indeed possible.

3 ADVERSARY MODEL AND ATTACK SCENARIO

In this section, we introduce the adversary model and devise two attack scenarios, which aim at breaching privacy and security.

3.1 Adversary model

We consider a scenario where the legitimate device (i.e., client) communicates with an AP and vice versa, see Fig. 3. We further consider an adversary in transmission range of the legitimate communication with the following capabilities: *(i)* sniffing packets sent by the legitimate device; *(ii)* extracting the fingerprint of the legitimate device from the CSI; *(iii)* knowing its own fingerprint and the ability to change it arbitrarily. Hence, the adversary can breach the user privacy upon extracting the fingerprint from the sniffed packets. Even if the client employs MAC address randomization to remain anonymous, the adversary can identify and track the client via the radiometric fingerprint (Attack scenario I). Further, having the ability to change its fingerprint arbitrarily, the adversary can subsequently modify its own fingerprint to impersonate the client, thus compromising the security of the system (Attack scenario II). Note that we do not consider an adversary launching a denial-of-service attack by jamming the WiFi signals as this will disrupt the communication in WiFi channels as a whole.



Location	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
2.4 GHz	99.8%	99.6%	98.3%	88.7%	98.3%	99.4%	63.9%	76.5%	31.9%	78.3%	34.9%
5 GHz	95.8%	99.7%	99.5%	97.6%	99.6%	72.1%	74.9%	79.4%	0%	0%	0%

Fig. 5. The success rate of impersonation attacks is shown as a heatmap throughout the experimental site when using 2.4 GHz. *Impersonation is possible even in challenging scenarios, i.e., through a 30-centimeter thick ferro-concrete wall.* The table shows the success rate at 2.4 and 5 GHz bands. *The success rate at* 5 *GHz is lower due to higher propagation and penetration loss compared to* 2.4 *GHz.*

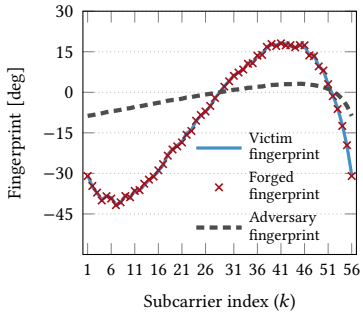


Fig. 6. Impersonation attack on CSI-based radiometric fingerprinting. *Impersonation is feasible when the adversary is capable of introducing additional crafted phase rotations per subcarrier to match the fingerprint of the victim.*

3.2 Attack scenario I: Violating user privacy by tracking the radiometric fingerprints

This attack focuses on tracking the presence of specific devices in the network using their radiometric fingerprints. In this scenario, the privacy-invading adversary silently sniffs the encrypted traffic over a WiFi network, extracts the fingerprints from the CSI, and creates a database recording time and duration in which a device was present in the vicinity. In order to show the efficacy of this attack experimentally, we setup an adversary and 5 phones that are entering and leaving the network at different times over the course of 10 minutes. We depict the results of this experiment in Fig. 4, in which the ground truth is presented in solid bars, whereas the hatched bars indicate the adversary's prediction. *We observe that the adversary is able to determine the presence time of the different phones with fairly high accuracy.* Note that MAC layer anonymization techniques cannot stop our adversary from tracking the presence of users across networks since such techniques do not conceal the inherent physical cues of the device, i.e., radiometric fingerprint. In an era where smartphones, smartwatches, and other WiFi-enabled wearables are omnipresent, these simple attacks expose us to significant privacy risks at workplace and at home.

3.3 Attack scenario II: Compromising security via impersonation attacks

Radiometric fingerprinting can enhance the security of networks by enabling means of additional authentication based on physical layer properties of devices [5, 21, 23]. However, we found that *an adversary can easily impersonate other devices* exploiting the fingerprinting scheme proposed by Liu *et al.* [23]. We mount such an *impersonation attack* using an SDR as follows: we first compute the fingerprint of the SDR by connecting it to another receiver (e.g., another SDR, signal analyzer). This needs to be done only once. Next, we measure the fingerprint of the target device, which only requires the adversary to sniff one (encrypted or unencrypted) packet sent by the target device. Knowing the proprietary fingerprint and that of a target, we can compute the phase offset on each subcarrier. These phase offsets are added to the LTF subcarriers and all subcarriers in the succeeding OFDM symbols (cf. Fig. 2) at the SDR. Consequently, the fingerprint extracted by the receiver matches that of the target device. The SDR provides flexible processing capabilities that allow us to introduce phase rotations to the transmission chain easily. In Fig. 6, we demonstrate how accurately the SDR (i.e., adversary) can replicate the fingerprint of another device (i.e., victim).

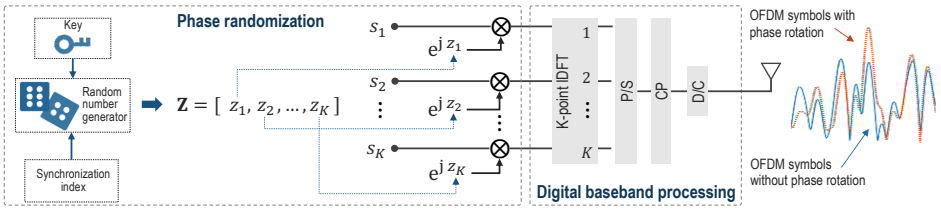


Fig. 7. Diagram illustrating the phase randomization method at the transmitter. *Additional random phase rotation at every subcarrier protects the fingerprint of the transmitter and therefore prevents impersonation.*

We now analyze the efficacy of the *impersonation attack* in a real-world scenario (i.e., an office building). We set up an AP which employs the CSI-based fingerprinting mechanism in [23] for authentication. To show the severity of the attack, we mount the attack in different locations in the vicinity of the victims, as depicted in Fig. 5. The adversary is transmitting 1000 packets in each location from which the access point calculates the fingerprints and compares them against a reference fingerprint of the legitimate user. We conduct this experiment for each of the WiFi bands (i.e., 2.4 and 5 GHz) and show the results in the table in Fig. 5. In the 2.4 GHz band, we observe that the adversary can successfully impersonate the victim in all locations. While the attack is very successful in line-of-sight scenarios (i.e., inside the lab), we observe that it still yields very high success rates in non-line-of-sight (NLOS) scenarios, e.g., the hallway or even in the office across the hallway. We kept all doors closed throughout the experiments. We also observe that the impersonation attack is possible even in highly challenging scenarios, i.e., behind a 30-centimeter thick ferro-concrete wall. However, the success rate is lower due to high signal attenuation. Due to higher propagation and penetration loss at 5 GHz band, the success rate in the NLOS locations (i.e., L5 to L11) is lower. In particular, in locations L9 to L11, no signal was received by the AP. However, locations L5 to L8 yield similar results in 5 GHz and 2.4 GHz bands. We conclude that, as long as the adversary is in range of the access point, they can successfully effectuate an impersonation attack regardless of the frequency band used by the access point.

3.4 Takeaway

In this section, we emphasize the need for a secure and privacy-preserving fingerprinting solution. Existing fingerprinting solutions based on CSI are capable of *distinguishing between different devices, even of the same model, allowing adversaries to track the presence of users in a network*. Further, we show that *an adversary can successfully impersonate the victim's device even through thick composite steel–concrete walls, which are among the most disruptive construction materials for wireless signals.* Hence, there are two main takeaway messages from this section: (i) *device fingerprints can be used to invade privacy of users and, as of the writing of this paper, there is no protection for users; and* (ii) *active deployments of CSI-based fingerprinting schemes can be attacked.*

4 HOW TO INJECT ARTIFICIAL NOISE TO FINGERPRINTS WITHOUT IMPACTING COMMUNICATION

Here we discuss our method for injecting artificial noise (i.e., randomized phase rotation) to the radiometric fingerprints. We further *prove why it does not impact the quality of communication.*

Randomizing the radiometric fingerprints is the first logical step towards maintaining user privacy. However, if not carefully designed, the randomization can potentially break/degrade the communication link. As described in Section 2.1, the WiFi receiver relies on the LTF field for channel estimation. To ensure that *obfuscation via randomization does not disrupt communication, we maintain the introduced phase rotations on each subcarrier constant for the duration of the whole*

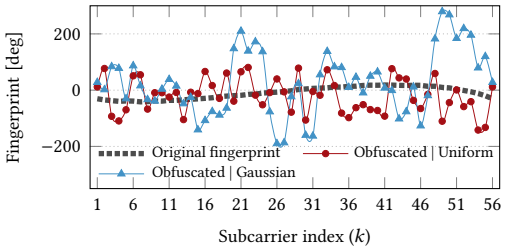


Fig. 8. Fingerprints before and after obfuscation. *Upon including random phase rotations in the subcarriers, the recovered fingerprint differs from the original.*

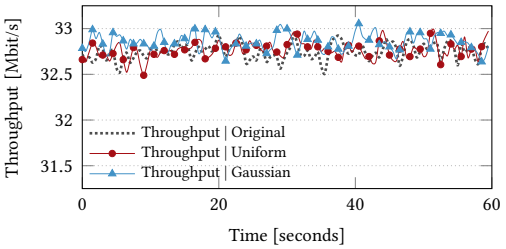


Fig. 9. Throughput before and after obfuscation. *The throughput is not affected by the phase rotations, as these can be reverted at the receiver.*

frame. As a result, the estimated CSI from the preambles remains valid for the succeeding VHT-DATA frame, as shown in Fig. 2, thus allowing successful decoding of information. In the following, we describe the process. Fig. 7 shows the transmitter chain of our proposed fingerprint obfuscation method, in which we include deliberate phase randomization across all OFDM subcarriers. Having the *pre-shared key* and *randomization index*, the receiver can decode the message and extract the fingerprint without impacting the communication.

Let z_k denote the phase rotation in subcarrier k intentionally included by the transmitter. From (2), the signal received in the k-th subcarrier is given by $y_k = h_k s_k e^{jz_k} + w_k$. Using (3), the CSI at the receiver is expressed as

$$\tilde{h}_k = h_k e^{jz_k} + w_k = |h_k|\, e^{j(\phi_k + z_k)} + w_k. \tag{6}$$

Compared to (3), the factor e^{jz_k} in (6) obfuscates the legitimate CSI by shifting its phase information. As a result, the phase z_k will appear in the radiometric fingerprint extracted by an adversary, thus safeguarding the device original fingerprint. The effects of this phase randomization mechanism can only be reverted by a trusted receiver that is aware of z_k. In particular, we assume that the phase z_k is a realization of a random variable Z_k, that can be generated locally at the receiver since the pre-shared key to the random generator is known. As a result, the receiver generates z_k and multiplies the perturbed $\tilde{h}_k$ in (6) by e^{-jz_k} yielding $e^{-jz_k}\left(|h_k|\, e^{j(\phi_k+z_k)} + w_k\right) = |h_k|\, e^{j\phi_k} + w_k$, which is equivalent to (3), and therefore showing that the CSI remains unaffected as the phase randomization can be removed. In addition, we denote the capacity of the channel in (3) by $C' = \log_2\left(1 + \left||h_k|\, e^{j\phi_k}\right|^2/\sigma^2\right) = \log_2\left(1 + |h_k|^2/\sigma^2\right)$. Similarly, the channel capacity of (6) is denoted by $C'' = \log_2\left(1 + \left||h_k|\, e^{j(\phi_k+z_k)}\right|^2/\sigma^2\right) = \log_2\left(1 + |h_k|^2/\sigma^2\right)$, thus revealing the equivalence $C' \equiv C''$. This shows that the channel capacity before and after randomization does not change. Therefore, for a given MCS level, the throughput is not altered by phase randomization as long as the phase z_k is generated correctly at each receiver. We generalize this idea for every subcarrier $k \in \mathcal{K}$. In Fig. 8, we illustrate the original fingerprint of a device as well as the obfuscated versions, in which the random phase rotations are obtained from uniform and Gaussian distributions, i.e., $Z_k \sim \mathcal{U}(\mu_k, \xi_k^2)$ and $Z_k \sim \mathcal{N}(\mu_k, \xi_k^2)$ with $\mu_k = 0$ deg and $\xi_k^2 = 60$ deg^2 (deg $\equiv$ °). Since the additional randomized phase z_k differs for each subcarrier, the adversary cannot leverage the linear phase error difference among subcarriers to identify the users. *In particular, if the same phase rotation z_k is used for all K subcarriers, the original fingerprint can be easily extracted via the method proposed in [23] as such method exploits the phase difference among adjacent subcarriers, which in this case would be constant and easy to remove.* Moreover, we corroborate experimentally that the throughput is not affected by our proposed obfuscation method. In particular, for the obfuscated signals depicted in Fig. 8, we show the throughput in Fig. 9.

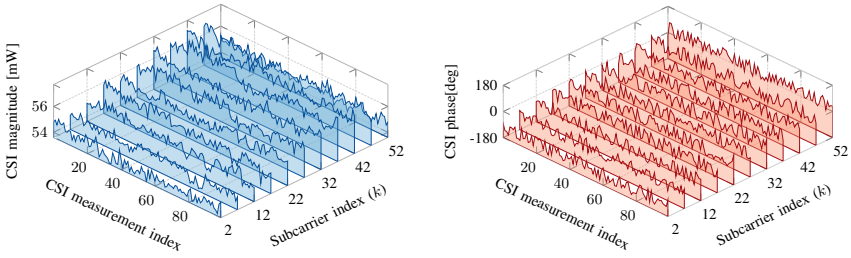


Fig. 10. Collected CSI measurements with additional synchronization phase rotations obtained from a zero-mean unit-variance Gaussian probability density function.

In the next section, we discuss the robustness of this approach against statistical attacks.

5 RF-SCOPE: A BENCHMARKING TOOL FOR ASSESSING VULNERABILITY TO STATISTICAL ATTACKS

Statistical attacks are common in cryptography where the adversary exploits statistical weaknesses of the underlying random number generators or hashing algorithms to discover the secrets, e.g., birthday attacks [3]. In the course of our experiments, we discovered that an adversary can mount similar attacks on phase randomization to restore the original fingerprint. *Viewing this as an estimation problem, we devise* RF-Scope*, which is a maximum likelihood-based approach design to restore the legitimate (unimpaired) CSI from a set of captured CSI measurements with obfuscated fingerprints.* Thus, if the legitimate CSI is restored accurately, the radiometric fingerprint can be extracted by the method described in Section 2.2 and used for malicious purposes. *In essence, we designed* RF-Scope *as a tool to evaluate the efficiency of RF-fingerprint obfuscation against statistical attacks.* Specifically, we designed an experiment in which the adversary captures 10000 CSI samples (within ~ 10 seconds) and uses RF-Scope to estimate the legitimate CSI. This experiment showed that *an adversary can denoise the fingerprint even without the knowledge of the probability density function used for phase randomization.* We will elaborate on RF-Scope and the experimental results in Section 5.1. We prove that this vulnerability stems from the zero-mean nature of the selected distributions, see Section 5.2.

Fig. 10 shows the magnitude and phases of CSI measurements. We assume that the channel impulse response is invariant for a short interval τ compliant with the channel coherence time T_c. Thus, small-scale oscillations in the CSI magnitude are attributed to noise. On the other hand, the CSI phase changes abruptly between contiguous measurements due to phase randomization.

5.1 A maximum-likelihood-based estimator for evaluating statistical attacks

RF-Scope minimizes the overall approximation error between the unknown CSI and the collected measurements. The premise is that adversaries do not have information on the probability density function used for CSI phase randomization. Let $\mathbf{M} = [\mathbf{m}_1, \cdots, \mathbf{m}_N] \in \mathbb{C}^{K\times N}$ denote a matrix that collects N measurements in all K subcarriers, where vector $\mathbf{m}_n \in \mathbb{C}^{K\times 1}$ represents the CSI (contaminated with phase randomization and noise) in the n-th captured LTF frame. Also, let $\mathbf{u} = \left[|h_1|\, e^{j\phi_1}, \cdots, |h_K|\, e^{j\phi_K}\right]^T \in \mathbb{C}^{K\times 1}$ denote the unknown unrandomized CSI vector. Further, $\Delta = [\mathbf{m}_1 - \mathbf{u}, \cdots, \mathbf{m}_N - \mathbf{u}] = \mathbf{M} - (\mathbf{1}^T \otimes \mathbf{u})$ represents the error matrix between the unknown CSI $\mathbf{u}$ and the measurements $\mathbf{M}$, where $\otimes$ is the Kronecker product. We define the following problem:

$$\mathcal{B} : \mathbf{u}^\star = \underset{\mathbf{u}\in\mathbb{C}^{K\times 1}}{\operatorname{argmin}} \underbrace{\left\|\mathbf{M} - \mathbf{1}^T \otimes \mathbf{u}\right\|_{\mathrm{F}}^2}_{J}, \tag{7}$$

where $\|\cdot\|_{\mathrm{F}}^2$ denotes the Frobenius norm. To solve problem $\mathcal{B}$, we have used several Kronecker product properties specified in *Appendix A*. Recalling that $\|\Delta\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(\Delta^T\Delta\right)$, the objective function can be recast

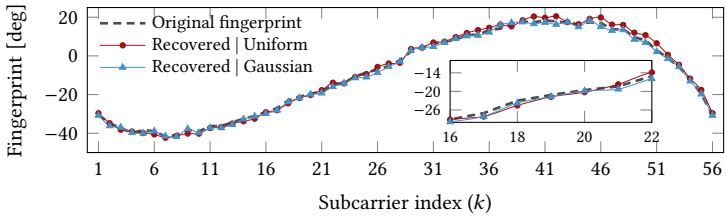


Fig. 11. Restored fingerprint upon CSI denoising. *We observe that an adversary capable of mounting a statistical attack can obtain the original fingerprint even after randomization.*

as $J = \mathrm{Tr}\Big((\mathbf{M} - \mathbf{1}^T \otimes \mathbf{u})^T (\mathbf{M} - \mathbf{1}^T \otimes \mathbf{u})\Big)$. By employing *Property 1* and *Property 2*, the objective collapses to $J = \mathrm{Tr}(\mathbf{M}^T\mathbf{M} - (\mathbf{1} \otimes \mathbf{u}^T)\,\mathbf{M} - \mathbf{M}^T\,(\mathbf{1}^T \otimes \mathbf{u}) + \mathbf{L} \otimes (\mathbf{u}^T\mathbf{u}))$, where $\mathbf{L} = \mathbf{1}\mathbf{1}^T$. To find a critical point $\mathbf{u}^\star$ that minimizes J, we compute the gradient of J with respect to $\mathbf{u}$ and equate it to zero, i.e., $\nabla_{\mathbf{u}} J = 0$. To this purpose, we resort to the use of differentials. Thus, $dJ = \mathrm{Tr}\Big(-(\mathbf{1} \otimes d\mathbf{u}^T)\mathbf{M} - \mathbf{M}^T(\mathbf{1}^T \otimes d\mathbf{u}) + \mathbf{L} \otimes (d\mathbf{u}^T\mathbf{u}) + \mathbf{L} \otimes (\mathbf{u}^T d\mathbf{u})\Big)$, where d denotes the differential operator and $d\mathbf{M} = 0$, $d\mathbf{L} = 0$. Using *Property 2*, *Property 3* and *Property 4*, the differential of J is expressed as $dJ = \mathrm{Tr}\Big(-(\mathbf{M1}) \otimes d\mathbf{u}^T - \mathbf{1}^T \otimes (\mathbf{M}^T d\mathbf{u}) + \mathbf{L} \otimes ((d\mathbf{u}^T)\mathbf{u}) + \mathbf{L} \otimes (\mathbf{u}^T d\mathbf{u})\Big)$. Now, by means of *Property 4* and *Property 5* we obtain $dJ = 2\mathrm{Tr}\Big(\big(-(\mathbf{M1})^T + N\mathbf{u}^T\big)d\mathbf{u}\Big)$. The Frobenius inner product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as $\langle \mathbf{A}, \mathbf{B}\rangle_{\mathrm{F}} \equiv \mathrm{Tr}(\mathbf{A}^T\mathbf{B})$. Therefore, $dJ = 2\,\langle -\mathbf{M1} + N\mathbf{u}, d\mathbf{u}\rangle_{\mathrm{F}}$, from where we obtain $\nabla_{\mathbf{u}} J = 2(-\mathbf{M1} + N\mathbf{u})$. Upon equating $\nabla_{\mathbf{u}} J$ to zero, we obtain $\mathbf{u}^\star = \frac{1}{N}\mathbf{M1} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{m}_n$. The denoised CSI phase $\mathbf{\Phi}$ for all subcarriers is computed as

$$\mathbf{\Phi}^\star = \arctan\left(\Im\mathfrak{m}\left\{\frac{1}{N}\sum_{n=1}^{N}\mathbf{m}_n\right\} \oslash \Re\mathfrak{e}\left\{\frac{1}{N}\sum_{n=1}^{N}\mathbf{m}_n\right\}\right). \tag{8}$$

Since denoised CSI is available through (8), the radiometric fingerprint $\boldsymbol{\epsilon}^\star$ can be extracted using (5). [23] Fig. 11 shows the restored fingerprints for uniform and Gaussian distributions with mean $\mu_k = 0°$ and variance $\xi_k^2 = 60$ deg^2, for all subcarriers $k \in \mathcal{K}$. In both cases, we have collected $N = 10000$ measurements. We observe that the obtained fingerprints exhibit a small deviation with respect to the original one. When uniform distribution is used, the mean absolute error (MAE) is $0.7489°$, whereas that of Gaussian distribution is $1.2252°$. Although both distributions have a variance of $\xi_k^2 = 60$ deg^2, for the uniform case, this signifies that the range of phase rotations is bounded to $[-99.2°; 99.2°]$. However, for the Gaussian case, the range of rotation phases spans $[-180°; 180°]$.

In *Appendix B*, we analyze the RF-Scope estimator under the Cramer-Rao bound (CRB) framework. We show that RF-Scope attains near-optimality in estimating the CSI.

5.2 Statistical rationale for CSI denoising feasibility via RF-Scope

If an efficient estimator does not exist for an unknown variable, the maximum-likelihood estimation often yields an asymptotically efficient estimator for sufficiently large number of samples. Based on this premise, we expect the effect of randomization to be averaged out. Thus, *motivated by the outcome of* RF-Scope*, we justify why the effect of randomization, introduced in Section 4, can be removed.* By assuming that an adversary is capable of collecting an infinite number of measurements, we RF-Scope within the law of large numbers; which states that the average of outcomes obtained from a large number of experiments approximates the expected value.

Assumption: *Let $f_{Z_k}(z_k)$ be a symmetric zero-mean probability density function governing the random phase rotation Z_k, spanning an interval with upper and lower bounds $z_k^U = R_k$ and $z_k^L = -R_k$, respectively.*

Invoking the assumption above, the expected value of the corrupted CSI information in subcarrier k according to (6) is defined as $\mathbb{E}\left[\tilde{h}_k\right] = \mathbb{E}\left[h_k e^{jZ_k}\right] + \mathbb{E}\left[w_k\right]$, where $\mathbb{E}\left[h_k e^{jZ_k}\right] = h_k \mathbb{E}\left[e^{jZ_k}\right] =$

$h_k \int_{-R_k}^{R_k} e^{jz_k} f_{Z_k}(z_k) dz_k$. Using integration by parts, $\mathbb{E}\left[e^{jZ_k}\right]$ can be recast as,

$$\mathbb{E}\left[e^{jZ_k}\right] = 2\sin(R_k) f_{Z_k}(R_k) - \int_{-R_k}^{R_k} \sin(z_k) f'_{Z_k}(z_k) dz_k + j \int_{-R_k}^{R_k} \cos(z_k) f'_{Z_k}(z_k) dz_k = \beta_k^{\text{real}} + j\beta_k^{\text{imag}}, \tag{9}$$

where the equivalence $\int udv = uv - \int vdu$ is used assuming that $u = f_{Z_k}(z_k)$, $dv = e^{jz_k} dz_k$ and $f_{Z_k}(-R_k) = f_{Z_k}(R_k)$ due to symmetry. In the following, we instantiate three fundamental corollaries that allow us to gain insights on the characteristics of (9).

Corollary 1: *If $g(x)$ is an even function, then its derivative $g'(x)$ is an odd function.*

Corollary 2: *If $g(x)$ is even and $h(x)$ is odd, then $q(x) = g(x)h(x)$ is odd.*

Corollary 3: *If $g(x)$ is odd, then $\int_{-a}^{a} g(x)dx = 0$ for $a > 0$.*

By means of *Corollary 1*, we assert that $f'_{Z_k}(z_k)$ is an odd function. Also, via *Corollary 2*, the function $\cos(z_k) f'_{Z_k}(z_k)$ is odd. Finally, by means of *Corollary 3* the value of $\beta_k^{\text{imag}} = \int_{-R_k}^{R_k} \cos(z_k) f'_{Z_k}(z_k) = 0$. As a result, $\mathbb{E}\left[h_k e^{jZ_k}\right] = \beta_k^{\text{real}} h_k$, which shows that (on average) the CSI in every subcarrier k is affected only by a real-value attenuation factor β_k^{real} without altering the phase.

Claim: *When we obfuscate the fingerprints through phase randomization using symmetric zero-mean distributions,* `RF-Scope` *produces an unbiased estimator for the CSI phase.*

Harnessing this outcome, we compute the expected value of the proposed `RF-Scope` estimator, i.e., $\mathbb{E}\left[\mathbf{u}^\star\right] = \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}\left[\mathbf{m}_n\right] = \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}\left[\text{diag}\left(e^{j\mathbf{z}_n}\right)\mathbf{h} + \mathbf{w}_n\right]$, where $\mathbf{z}_n = \left[Z_{n,1}, \cdots, Z_{n,K}\right]^T$ and $\mathbf{w}_n = \left[w_{n,1}, \cdots, w_{n,K}\right]^T$. Thus, $\mathbb{E}\left[\mathbf{u}^\star\right]$ reduces to

$$\mathbb{E}\left\{\mathbf{u}^\star\right\} = \begin{pmatrix} \beta_1^{\text{real}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_K^{\text{real}} \end{pmatrix} \begin{pmatrix} |h_1| e^{j\phi_1} \\ \vdots \\ |h_K| e^{j\phi_K} \end{pmatrix}. \tag{10}$$

From (10), we note that when the randomization scheme in Section 4 is used for CSI obfuscation, its effect can be removed via `RF-Scope`. Essentially, the restored CSI magnitudes $|h_k|$ are scaled by β_k^{real} but the phases ϕ_k remain unaffected. As a result, an adversary can extract the radiometric fingerprint $\boldsymbol{\epsilon}$ (defined in (4)) from the restored CSI phase Φ. In order to prevent this outcome that infringes secrecy, a specific type of probability density function is required that prevents CSI denoising from collected measurements. This aspect is elaborated thoroughly in Section 6.2.

5.3 Takeaway

Any system relying on randomization for improving security/privacy should prove robust against statistical attacks. Here we propose `RF-Scope` to *assess the vulnerability of fingerprint randomization against these attacks.* This tool will be later used to demonstrate the robustness of our proposed fingerprint obfuscation method (i.e., `RF-Veil`) against statistical attacks. Furthermore, we analyze the statistical rationale behind the aforementioned vulnerability. *This analysis is then leveraged to devise suitable countermeasures in the next section.*

6 RF-VEIL: A PRIVACY- AND SECURITY-PRESERVING SOLUTION FOR RADIOMETRIC FINGERPRINTING

In this section, we introduce our proposed technique `RF-Veil`, which injects crafted artificial noise to fingerprints in order to improve the robustness of WiFi transmissions against statistical attacks aiming at fingerprint acquisition. In Fig. 12, we illustrate the building blocks of `RF-Veil`. Note that, in `RF-Veil-Standalone` mode, we only need a subset of the blocks at the transmitter since the receiver does not perform any radiometric fingerprinting. To avoid repetition, we highlight the `RF-Veil-Standalone`-specific blocks and the algorithm workflow in this mode in Section 6.4.

A short overview of `RF-Veil`. As shown in Fig. 12, *the transmitter uses a random number generator to generate a pattern that obfuscates its radiometric fingerprint on a per-frame basis. The*

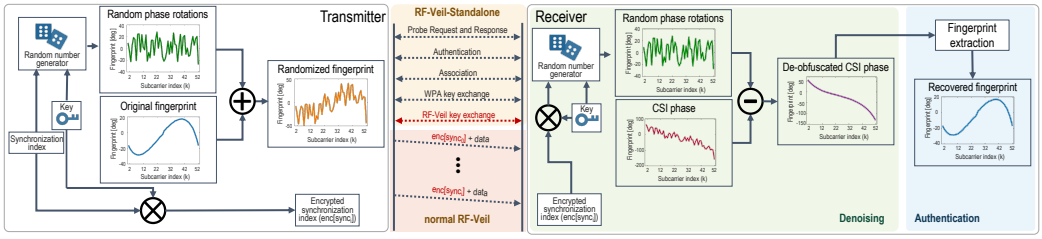


Fig. 12. Schematic overview of an RF-Veil transmitter and receiver. The flow-diagram in the center depicts the WiFi connection establishment and data exchange procedure (changes due to RF-Veil marked in red color). The ACK messages are not shown in the figure for readability. They are not modified in RF-Veil.

random number generator follows a specific distribution that is robust against statistical attacks. The receiver requires the seed to the random generator in order to generate the same pattern, which is used for CSI denoising and fingerprint extraction. The details about standard compliancy, random sequence generation, and key exchange are elaborated below.

6.1 Association

In WiFi, every new device first associates to the AP upon arrival to the network. This includes exchanging the probe, authentication, and association request and response messages. It is within this stage that the AP and the device establish a secure connection. In RF-Veil, we require the access point and the client to exchange one more key, which is used as one of the inputs to the random number generator, as shown in Fig. 7. We choose to use a pre-shared key due to ease of implementation. However, one can leverage alternative secret key extraction methods that rely on channel response [22]. As a result, the receiver and transmitter do not require a security handshake in advance but use physical layer information to generate the secret keys.

At this stage, the AP can extract the real fingerprint of the client after obtaining the shared key. We elaborate further on this in Section 6.3.

6.2 Obfuscation at the transmitter

The main task of the transmitter consists in obfuscation, as depicted on the left-hand side in Fig. 12. For every frame, a random sequence is generated using the pre-shared key and the synchronization index.

Pre-shared key. In our implementation, we used a 128-bit key, which is refreshed every time the device re-associates with the AP. As a privacy protection measure, we obfuscate the fingerprint even before the association with an AP takes place. Hence, any frame transmitted from the devices (e.g., beacon, discovery) has an obfuscated fingerprint. In this case, it is advised to generate a new key periodically in order to protect against statistical attacks (see Section 4). We leave the frequency of key renewal as a design choice. Since renewing the pre-shared key does not impose considerable overhead, we suggest to lean towards higher security.

Synchronization index. Attaching a synchronization index to each frame has two purposes: *(i)* synchronize the random generator between the receiver and the transmitter and *(ii)* protect the receiver from replay attacks. The synchronization is important because the pre-shared key only ensures that the random generators at both ends produce the same string of random numbers. However, if a frame is lost, then the receiver may try to de-obfuscate the frame with the wrong pattern. To prevent this, we attach an index for each frame, so that the receiver can use this index in combination with the pre-shared key to generate a synchronized and secure randomization pattern. We intentionally refrained from using the existing 12-bit MAC frame sequence number due to its

vulnerability to replay attacks. Even at low data rates, the 12-bit sequence number resets within seconds, whereas our 32-bit sequence number takes 24 days to reset at the rate of 1000 frames per second. We expect the WiFi connection to be re-initiated within such an interval. Even though the exposure of this synchronization index does not expose legitimate users to security threats, it can still be abused for tracking. Therefore, we encrypt this index with the pre-shared key via XOR operations. We further discuss this approach in Section 8.

Once the obfuscation pattern is generated for all subcarriers, the symbols of the regular WiFi transmitter are rotated accordingly. The frame sequence number is then updated for the next frame and stored in a lookup-table (LUT). Finally, the symbols with phase rotations can be sent out over the air. However, one question still remains: *how do we ensure robustness against statistical attacks?*

Robustness against statistical attacks. In Section 5, we showed experimentally and analytically that obfuscation with symmetric zero-mean distributions is susceptible to statistical attacks. Recalling the analysis therein, a robust distribution against such attacks should have the following properties.

$$(\mathbf{P}_1): f_Z(z) \geq 0 \qquad (\mathbf{P}_2): \int_{-\infty}^{\infty} f_Z(z)dz = 1 \qquad (\mathbf{P}_3): f_Z(z) \neq f_Z(-z) \qquad (\mathbf{P}_4): \mathbb{E}\left[f_Z(z)\right] \neq 0$$

Essentially, $(\mathbf{P}_1)$ and $(\mathbf{P}_2)$ are inherent properties of all probability density functions, i.e., they are non-negative, and the total area under the graph $f_Z(z)$ is equal to unity. On the one hand, $(\mathbf{P}_3)$ requires the probability density function to be non-symmetric while $(\mathbf{P}_4)$ states that it must not be centered around zero. These properties ensure that the effect of the random phase rotations will prevail even if a statistical attack is perpetrated. In Section 7, we corroborate experimentally that probability density functions complying with $(\mathbf{P}_1)$, $(\mathbf{P}_2)$, $(\mathbf{P}_3)$ and $(\mathbf{P}_4)$ can conceal the radiometric fingerprint effectively.

In the following, we justify the necessity for $(\mathbf{P}_3)$ and $(\mathbf{P}_4)$. From (9), we note that $\mathbb{E}\left[e^{jZ_k}\right] = \beta_k^{\text{real}} + j\beta_k^{\text{imag}}$ must be complex-valued in order to prevent the phase randomization effect from being removed. This is attained when the term $\beta_k^{\text{imag}} = \int_{-R_k}^{R_k} \cos(z_k) f'_{Z_k}(z_k)dz_k \neq 0$, which produces a non-zero phase shift that is absorbed by the CSI phase thus concealing the fingerprint. In order for this to hold, $\cos(z_k) f'_{Z_k}(z_k)$ must not be an odd function according to *Corollary 3*. Since $\cos(z_k)$ is an even function, this also signifies that $f'_{Z_k}(z_k)$ must not be an odd function according to *Corollary 2*. Via *Corollary 1*, this requirement is satisfied when $f_{Z_k}(z_k)$ is not an even function. Therefore, it is revealed that we can design arbitrary probability density functions $f_{Z_k}(z_k)$ that are not even with non-zero mean, thus yielding the desired effect that prevents phase randomization removal.

A simple yet effective manner to meet the above criteria is using a shifted even probability density function (e.g., shifted Gaussian or uniform distribution). We will experimentally prove that in Section 7.2.

6.3 De-obfuscation and authentication at the receiver

The right-hand side of Fig. 12 shows the two main tasks of the receiver: de-obfuscation and authentication.

De-obfuscation. Having the synchronization index and pre-shared key, the receiver can regenerate the obfuscation pattern (i.e., randomized phase rotations) of the transmitted frame. This allows the receiver to extract the original fingerprint. This is done easily by subtracting the obfuscation pattern from the phase of the received signal.

Authentication. The receiver verifies the restored fingerprint against the original fingerprint of the transmitter to authenticate the received frame. In addition, the receiver verifies that the synchronization index is larger than that in the last received frame. A frame whose synchronization index is less than or equal to the last frame is probably sent from an adversary attempting a replay attack. *We highlight that with* `RF-Veil`*, WiFi devices can always obfuscate their fingerprint.* We mentioned

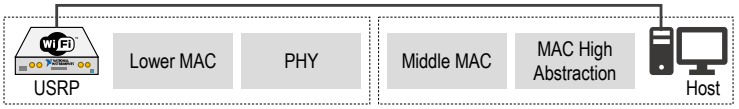


Fig. 13. Schematic overview of the hardware setup.

in Section 4 that RF-Veil is designed such that obfuscation does not impact the communication performance. Hence, user privacy is always ensured through fingerprint concealment.

6.4 RF-Veil-Standalone mode

In this mode, we allow the transmitting device to hide its fingerprint by executing the obfuscation blocks without any handshake or coordination with other receivers. Specifically, the device generates locally a synchronization index and the key, which are used as inputs for fingerprint obfuscation, as depicted in the transmitter side of Fig. 12. As a result, we can ensure privacy protection in a much broader scenario, e.g., communicating with non-RF-Veil-enabled devices, in absence of any active connections, or in connection establishment phase.

6.5 SDR implementation

We have implemented RF-Veil using the USRP 2954R SDR platform. A simplified overview of the hardware used in our setup is depicted in Fig. 13. Each USRP is connected via PCI-e interface to a host machine running NI-Linux RT (kernel version 4.1.13-rt15-nilrt). We build RF-Veil using NI 802.11 application framework (AFW) [2], which provides the physical layer and lower MAC layer functions in the FPGA, while the rest of the MAC procedures run at the host (Linux RT in our setup). We provide a detailed overview of the existing implementation in Appendix D. Due to space constraints, we do not delve into the SDR implementation details. Our implementation and data is available online[3]. The following briefly describes the setup.

Fingerprint extraction at the receiver. The physical layer implementation of 802.11 AFW already includes CSI estimation in the FPGA. For our implementation, we have transferred the CSI from the FPGA to the host via a Target-to-Host (T2H) FIFO on a per-frame basis. This enables fast prototyping while maintaining real-time operation of the testbed. Having the CSI, we implemented the radiometric fingerprinting using non-linear phase errors, as described in Section 2.2.

Fingerprint modification at the transmitter. These are required modifications at both the FPGA and the host. At the host, we compute the obfuscation pattern, which is sent to the FPGA on a per-frame basis. We made use of the interprocess communication protocol by NI to send packets containing the additional phase rotations. Then, we modified the transmitter chain at the FPGA to read the obfuscation pattern and multiply each outgoing symbol with the corresponding phase rotations. This increases the latency of the transmission chain by 5 clock cycles (12.5 ns).

Secure fingerprinting. We implement RF-Veil on top of the *Fingerprint Extraction* and *Fingerprint Modification* modules on the host. We extend the packet headers so as to also carry the 32-bit synchronization index chosen at the transmitter. When a new packet is being prepared for transmission, the MAC header is used to obtain the key and synchronization index from the LUT. Then, the obfuscation pattern is generated using the key and synchronization index. This pattern serves as input for the *Fingerprint Modification* module, which then pushes the values to the PHY.

At the receiver side, the CSI is written into the *T2H Channel Estimation* FIFO at the PHY. The frame reception continues on the FPGA while the implementation of RF-Veil runs on the output of the FIFO at the host. After the information for random pattern generation is obtained, and the

[2]http://www.ni.com/pdf/manuals/376779f.pdf
[3]https://github.com/seemoo-lab/RF-Veil

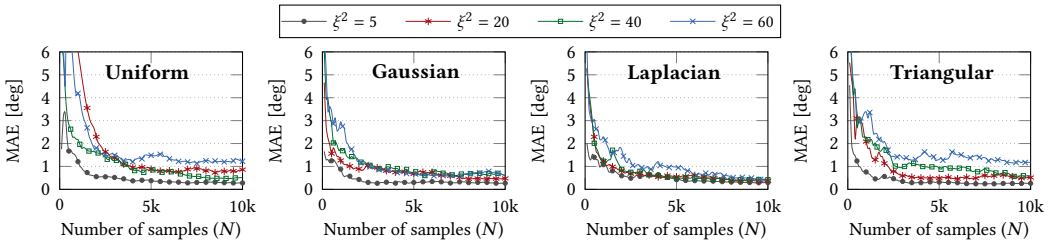


Fig. 14. Mean absolute error (MAE) for different symmetric zero-mean distributions using RF-Scope in 802.11ac. *These results show that zero-mean randomization is not robust against statistical attacks since the fingerprint can be obtained with high accuracy and negligible error (below 2%).*

randomization is reverted, the fingerprint is calculated by the *Fingerprint Extraction* module. The obtained fingerprint is then passed to the *Matcher* to be compared with the original fingerprint for authentication.

6.6 Takeaway

In this section, we elaborated on the workings behind RF-Veil *and its standalone-mode. We devised the idea of synchronized obfuscation with special probability density functions to counter the statistical attacks introduced in Section 5, as well as the tracking and impersonation attacks introduced in Section 3. The prototype implementation of* RF-Veil *on a USRP SDR platform enables us to experimentally evaluate the performance of our approach. The takeaway message is that* RF-Veil *introduces low overhead to the existing WiFi message flow while providing enhanced privacy for users and a secure way of physical layer device identification.*

7 EVALUATION

In this section, we first evaluate the efficacy of the impersonation attack introduced in Section 3. We then leverage RF-Scope to provide a broader assessment of the performance of naive randomization (i.e., obfuscation via zero-mean distributions) and RF-Veil against statistical attacks.

7.1 Performance of naive randomization

In Section 5, we demonstrated the vulnerability of obfuscation, with zero-mean distributions, to statistical attacks experimentally and analytically. In particular, we showed that an adversary can easily restore the original fingerprint from 10000 frames. However, we have neither studied the impact of number of samples, nor considered the effect of the distributions variance on the accuracy of the restored fingerprint by the adversary. To this aim, in Fig. 14, we show the mean absolute error (MAE) of the adversary's estimate of the original fingerprint when using RF-Scope in 802.11ac. The figure demonstrates the results under four distributions, namely, uniform, Gaussian, Laplacian, and triangular. For each distribution, we compute the MAE with four variances. Here we make two key observations: *(i)* the adversary can *restore the original fingerprint with very high accuracy by just processing the CSI of ~2000 frames* (a couple of seconds[4]), and *(ii)* the CSI-recovery error increases with the variance of randomization *since larger variance leads to higher entropy of the obfuscated fingerprints*. This behavior is mainly observed when the number of samples is low. As more samples are processed, the estimation error converges to nearly the same value (this is also supported by equation (B7) in Appendix B). Nonetheless, an adversary can still obtain accurate estimates of the original fingerprint with negligible error even when distributions with large variances are

[4]In estimating the time for collecting a given number of frames, we assume that the user transmits at ~ 8Mbps. This number is referential and intendeds to provide an estimate of how fast an adversary can mount an statistical attack.

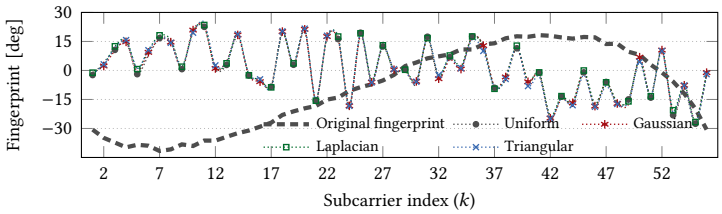


Fig. 15. Restored fingerprint after obfuscation with RF-Veil. *Note that* RF-Veil *prevents potential adversaries from infringing privacy and security since the original fingerprint cannot be recovered. In this experiment, the* RF-Veil *transmitters use the same values for shifting the means across the subcarriers. Hence, it is the expected behavior that the restored fingerprint for the different random distributions are similar.*

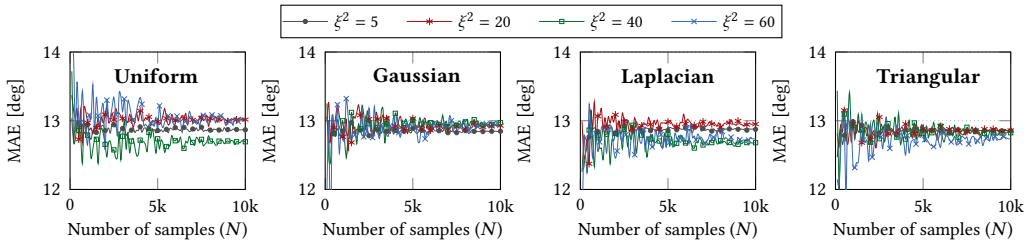


Fig. 16. Mean absolute error (MAE) using RF-Scope in 802.11ac after fingerprint obfuscation via RF-Veil. Since the error is high, RF-Veil *prevents adversaries from obtaining the fingerprint of the targeted victim.*

employed. For instance, with only 1000 CSI samples, the MAE is below 3° for distributions with a variance of 60 deg^2. For small variances such as $\xi^2 = 5$ deg^2, with only 500 samples (roughly 0.5 seconds), the estimation error is consistently below 1° for all distributions. We observe a similar trend with 802.11a, whose results are available in Appendix C.

Remarks: We have shown experimentally that the effect of naive randomization can be removed if an attacker is capable of collecting a few thousand samples to mount an statistical attack. Thus, naive randomization does not protect the fingerprint of devices.

7.2 RF-Veil performance

In this experiment, we evaluate the security and privacy enhancement achieved by RF-Veil. Following the conditions for randomization patterns that are robust to statistical attacks (see Section 6.2), we obfuscate the original signature of the device using the same four distributions whose mean values are now shifted according to a random pattern, unlike the previous experiment. The results of this experiment in legacy mode 802.11a are provided in Appendix C. As depicted in Fig. 15, the recovered fingerprints using RF-Scope deviate from the original fingerprint substantially and follow the course of a random pattern. Essentially, when an adversary uses statistical analysis to identify devices, the extracted fingerprint will not match with the original fingerprint. To shed light on this aspect, Fig. 16 depicts the MAE of fingerprints restored by RF-Scope with increasing sample sizes and under different variances. As compared to the low MAE in Fig. 14 (3°) where zero-mean distributions are used, the MAE in Fig. 16 increases approximately by 4-fold (13°) regardless of the variances and sample sizes used. While all the variances lead to nearly the same error when N is large, we observe that a large variance produces more variability in the MAE, specially with a small number of samples N. On the other hand, small variances produce a more condensed range of MAE values throughout all N. *This result demonstrates two promising properties of* RF-Veil*: (i)*

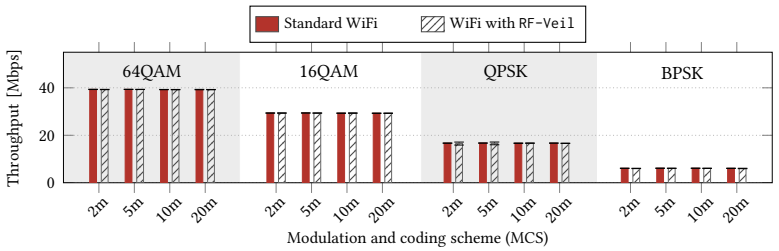


Fig. 17. Average throughput of regular 802.11ac and RF-Veil in different distances and with different MCS over the course of 60 seconds. *Here, we prove experimentally that* RF-Veil *does not impact the throughput.*

the adversary's estimate does not improve even with large number of samples, and (ii) the impact of the variance on estimation is almost negligible as all the errors converge to a similar value (also supported by equation (B7) in Appendix B). Note that the random pattern in Fig. 15 is generated approximately within the same phase-error range as the original fingerprint (i.e., between $-25°$ and $40°$). Therefore, the attained MAE is not excessively large. However, the MAE can be arbitrarily larger if we construct the pattern spanning a wider range.

Remarks: RF-Veil *protects users' privacy by preventing adversaries from estimating the original fingerprint for tracking/locating the user. Furthermore, the security is also enhanced, since the adversary cannot successfully forge the original fingerprints of other devices.*

Effect of RF-Veil on throughput. In Section 4, we analytically showed that RF-Veil does not impact the throughput of WiFi communication. Here, we confirm our analysis with experiments.

We design an experiment in which we measure the throughput of two WiFi devices at different distances (up to 20m) and under distinct modulation and coding schemes (MCS) (up to 64 QAM) with/without RF-Veil. Fig. 17 demonstrates that RF-Veil does not impact the throughput of the system, thus confirming our analysis. This is because the fingerprint obfuscation of RF-Veil is based only on phase rotations of the I/Q symbols within a frame. In particular, such rotations do not affect the WiFi channel estimation since their effect is removed at the legitimate receivers. In the figure, we only show the result of obfuscation with uniform random distribution with $\xi^2 = 60°$. However, we report that the other random distributions (Gaussian, Laplacian, and triangular) do neither impact the throughput. In this experiment, we also measured the computational overhead of RF-Veil. Our measurements show that RF-Veil has an average execution time of 49.495 microseconds, even though we implemented most parts of RF-Veil on the host (i.e., a windows machine). We expect the execution time to drop by at least an order of magnitude in real-time kernel or FPGA implementation.

Remark: RF-Veil *has low computational overhead and does not impact the communication quality.*

8 DISCUSSION

In this section, we discuss some of the practical aspects of RF-Veil.

To share or not to share? For the secure CSI-denoising when using RF-Veil, we use a symmetric shared key as it provides the easiest way of synchronizing the random number generators at the transmitter and receiver. Furthermore, the synchronization index can be easily encrypted by XOR-ing the index with the shared key. An alternative to the key exchange we use in Section 6.1 is a key extraction mechanism based on physical layer properties, such as [22]. This key extraction method leverages channel response information at the transmitter and receiver to generate symmetric keys. *Note that* RF-Veil *is compatible with both methods.* Regardless of the method, the key should be renewed at certain intervals, which brings us to the next point in our discussion.

How often should we renew the key? The monotonically increasing 32-bit synchronization index ensures that, even if the transmitter keeps the symmetric key static for a certain time, they

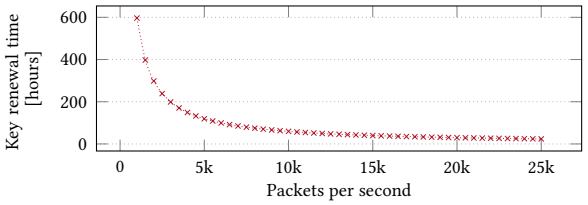


Fig. 18. Time to renew the key for different sending rates.

do not repeat the pattern of phase rotations. If the synchronization index wraps around at its maximum of $2^{32} = 4\,294\,967\,296$, the pattern of random phase rotations is repeated, and an adversary can potentially launch a replay attack. In order to thwart such an attack, the key has to be renewed before the transmitter starts to re-use their synchronization indices for this key. Furthermore, it is important to point out that the transmitter initializes the synchronization index with a number between 0 and 2 147 483 648, which ensures that a part of the key cannot be guessed from the encrypted synchronization index in the frame. We plot the minimum time for key refreshment under different transmission rates in Fig. 18 (for the worst case in which the transmitter chose to start at 2 147 483 648). We observe that the key has to be refreshed every 596 hours (24.8 days), assuming an average of 1000 packets per second. Even if we assume that the transmitter sends on average 25 000 packets per second, which would imply a rate of 462.4 Mbps, it will exhaust the number of available synchronization indices in 23.8 hours. Thus, even at very high rates (5 TByte of traffic per day), the key exchange is not too frequent. Note that increasing the frequency of key exchange does not decrease the security level but increase the overhead since the keys are either transmitted encrypted by WPA2 or extracted by both transmitter and receiver using key extraction methods [22]. Subsequently, an exposure of this key would affect the privacy and security of the connection until a new key is exchanged.

What if a frame is rejected? A frame can be rejected for two different events: (*i*) *the calculated frame check sequence (FCS) of the frame does not match the actual FCS in the frame*, and (*ii*) *the fingerprinting algorithm rejects the frame (e.g., the extracted fingerprint differs from the expected one)*. In both cases, we let the MAC layer handle the re-transmission. In the case of a rejected frame, an `RF-Veil` transmitter does not re-use the synchronization index of the frame that is to be re-transmitted; instead, it increases the count as if a new frame was transmitted. This is crucial to guarantee the security of the system as the encryption of the same synchronization index would lead to the same cipher-text.

What if a device is not yet connected? If a device is not connected to an AP, it can still obfuscate its fingerprint by using `RF-Veil-Standalone` mode in order to lead a potential privacy-intruding adversary astray. Once the device is connected to an AP and the pre-shared key has been established, it can switch into `RF-Veil` mode, allowing the AP to securely extract the unrandomized fingerprint. This same mechanism applies to the probe requests and acknowledgments in response to probe responses from APs during active scanning. Note that, when the device is not associated to an AP, it can simply use a random key.

How does an `RF-Veil` transmitter communicate with a non-`RF-Veil` receiver? Recalling Section 4, the obfuscation of fingerprints does not degrade the channel quality as the channel estimation and equalization at the receiver can handle the arbitrary phase shifts introduced by the transmitter. Specifically, the additional phase rotations are absorbed by the CSI, and as long as the same phase rotation pattern is used for all the subcarriers within the frame, the receiver will assume that such CSI is legitimate. Hence, a receiver that is not aware of `RF-Veil` will simply revert the phase shifts together with the channel effects. In other words, a transmitter using the

RF-Veil-Standalone mode can still communicate with a legacy receiver. This receiver, however, will not be able to extract the correct fingerprint of the transmitter.

Can we implement RF-Veil on commercial off-the-shelf (COTS) devices? In recent years, a number of research groups have developed firmware modification/patching tools which allow manipulating MAC/PHY layer operations of the WiFi chipset. Although out of scope of this work, we believe that RF-Veil can be implemented on COTS devices using such tools. In particular, Schulz *et al.* [33] demonstrate the feasibility of modifying IQ symbols in commercial APs equipped Broadcom chipsets using their firmware patching framework, i.e., nexmon[5].

9 RELATED WORK

To date, we have not found any prior work on radiometric fingerprint obfuscation. Prior works only focused on thwarting identification techniques that used packet metadata (frame size, data rate, inter-packet time, etc.) and friendly jamming [29], upper-layer characteristics such as jitter of beacon timestamps [2], rate switching mechanisms [9], and under-specification of the MAC layer protocols and procedures [4, 6]. The proposed countermeasures for these upper-layer fingerprinting techniques consist of pattern randomization [15, 19, 28, 36], similar to ours. *However, unlike* RF-Veil*, their approach eliminates the possibility of legitimate fingerprinting. Furthermore, the solutions therein are not tested against statistical attacks.*

In the following, we provide a broader overview of the radiometric fingerprinting solutions, which can be categorized into *transient-based* and *modulation-based* approaches.

9.1 Transient-based approaches

The transient refers to the part of the signal in which the amplitude rises from background noise to full power [30]. Given its dependence on the hardware characteristics, a transient is a reliable feature for device identification by tracking the small but measurable differences in the turn-on transients. This can, for example, include the duration of turn-on transient [30] or standard deviation of normalized amplitude, phase, and frequency [16]. These approaches are cumbersome since they rely on the exact extraction of the transient portion of signals, which further depends on the channel noise. To ensure accurate and timely detection of the transient despite the channel noise, a very high sampling rate is required, which is typically achievable by high-end oscilloscopes (e.g., 4 Giga samples per second in [11]).

9.2 Modulation-based approaches

Modulation- or steady-state approaches, as the name suggests, make use of errors in the modulated signal. The seminal work of Brik *et al.* [5] proposes to collect the fingerprints from five features of the modulated signal, that is, magnitude, phase and frequency error, I/Q origin offset, and SYNC correlation. They show experimentally that their solutions, called PARADIS, can differentiate among 130 identical IEEE 802.11b devices with an accuracy above 99% even under mobility and varying noise conditions. Similar to the transient-based approaches, their approach requires additional equipment since they rely on high-end vector analyzers for channel sampling. Motivated by their work, recent approaches [18, 23] propose to use the CSI obtained from the pilot symbols which are readily available on WiFi chipsets, such as the Intel 5300 or Atheros AR9380. Specifically, Hua *et al.* [18] propose to compute the fingerprint using a combination of CFO extracted from the CSI and time difference of arrival (TDoA) computed from capturing 5000 adjacent frames. Furthermore, they require the device to remain stationary for at least 10 seconds for authenticating a device based on the previously collected fingerprint. The most recent work on

[5]https://github.com/seemoo-lab/nexmon

radiometric fingerprinting [23] makes use of the non-linear phase errors extracted from CSI. Their work takes advantage of non-linear phase error extraction methods proposed by Zhuo *et al.* in [41]. In this paper, we work toward obfuscating the radiometric fingerprints caused by non-linear phase errors [23, 41] since it neither relies on RF equipment with very high sampling rates nor requires large number of frames or stationary user behavior for fingerprinting. Nonetheless, `RF-Veil`'s approach can be extended to other features of the signal, which is controllable at the chipset, such as CFO and amplitude.

10 CONCLUSIONS

Radiometric fingerprinting is typically considered a secure method for device identification [23, 31, 39]. In this paper, we first *demonstrate the vulnerability of the latest CSI-based radiometric identification schemes to impersonation attacks*, which emphasizes the need for fingerprinting solutions that are robust against adversarial attacks on user security and privacy. We also illustrate that a *naive fingerprint-randomization approach does not withhold adversaries capable of mounting statistical attacks (i.e.,* `RF-Scope` *in this paper).* Consequently, we devise `RF-Veil`, a framework that enhances user privacy against fingerprint-based tracking/localization attacks, and is robust to statistical, impersonation, and replay attacks.

To the best of our knowledge, this is the first article that addresses the vulnerabilities of radiometric fingerprints. Hence, we foresee a few avenues of research as future work. Leveraging the randomization patterns to create a side-channel between the receiver and transmitter is an interesting method for exchanging the synchronization index. Furthermore, extending `RF-Veil` *to support MIMO transmissions or other signal characteristics such as CFO is another direction to further enhance user privacy. Randomizing the STFs and its impact on the communication and radiometric fingerprints is also an interesting research avenue. Further, investigating new distributions functions for the phase rotations that not only preserve security and communication but also reduce the peak-to-average power ratio is an interest research direction, especially for achieving high energy efficiency in low-power IoT devices.*

ACKNOWLEDGMENTS

This research is conducted in the context of the DFG-funded project SenShield (447586980). This work is in part supported by the B5G-Cell project in SFB 1053 MAKI and by the LOEWE initiative (Hesse, Germany) within the emergenCITY center. We would like to thank Clemens Felber and Dr. Walter P. Nitzold from NI, Dresden, for their valuable guidance in modification of WiFi LabVIEW AFW.

REFERENCES

[1] Luis F. Abanto-Leon, Gek Hong (Allyson) Sim, Matthias Hollick, Amnart Boonkajay, and Fumiyuki Adachi. 2020. SWAN: Swarm-Based Low-Complexity Scheme for PAPR Reduction. In *IEEE GLOBECOM*. 1–7.

[2] Chrisil Arackaparambil, Sergey Bratus, Anna Shubina, and David Kotz. 2010. On the Reliability of Wireless Fingerprinting using Clock Skews. In *ACM WiSec*. 169–174.

[3] Mihir Bellare and Tadayoshi Kohno. 2004. Hash function balance and its impact on birthday attacks. In *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 401–418.

[4] Sergey Bratus, Cory Cornelius, David Kotz, and Daniel Peebles. 2008. Active Behavioral Fingerprinting of Wireless Devices. In *ACM WiSec*. 56–61.

[5] Vladimir Brik, Suman Banerjee, Marco Gruteser, and Sangho Oh. 2008. Wireless Device Identification with Radiometric Signatures. In *ACM MobiCom*. 116–127.

[6] Johnny Cache. 2006. *Fingerprinting 802.11 Devices*. Ph.D. Dissertation. Naval Postgraduate School.

[7] Milos Cermak, Stefan Svorencik, Robert Lipovsky, and Ondrej Kubovic. 2020. *KR00K - CVE-2019-15126*. Technical Report. ESET.

[8] Marco Cominelli, Felix Kosterhon, Francesco Gringoli, Renato Lo Cigno, and Arash Asadi. 2020. An Experimental Study of CSI Management to Preserve Location Privacy. In *ACM WiNTECH*. 64–71.

[9] Cherita L Corbett, Raheem A Beyah, and John A Copeland. 2008. Passive Classification of Wireless NICs during Active Scanning. *International Journal of Information Security* 7, 5 (2008), 335–348.

[10] A. N. D'Andrea, U. Mengali, and R. Reggiannini. 1994. The Modified Cramer-Rao Bound and its Application to Synchronization Problems. *IEEE Transactions on Communications* 42, 234 (Feb 1994), 1391–1399.

[11] Boris Danev and Srdjan Capkun. 2009. Transient-based Identification of Wireless Sensor Nodes. In *ACM IPSN*. 25–36.

[12] Scott Fluhrer, Itsik Mantin, and Adi Shamir. 2001. Weaknesses in the Key Scheduling Algorithm of RC4. In *SAC*. Springer Berlin Heidelberg, 1–24.

[13] Robert M. Gray. 2006. Toeplitz and Circulant Matrices: A Review. *Foundations and Trends® in Communications and Information Theory* 2, 3 (2006), 155–239. https://doi.org/10.1561/0100000006

[14] Francesco Gringoli, Matthias Schulz, Jakob Link, and Matthias Hollick. 2019. Free Your CSI: A Channel State Information Extraction Platform For Modern Wi-Fi Chipsets. In *ACM WiNTECH*. 21–28.

[15] Marco Gruteser and Dirk Grunwald. 2005. Enhancing Location Privacy in Wireless LAN through Disposable Interface Identifiers: A Quantitative Analysis. *Mobile Networks and Applications* 10, 3 (2005), 315–325.

[16] Jeyanthi Hall, Michel Barbeau, and Evangelos Kranakis. 2004. Enhancing intrusion detection in wireless networks using radio frequency fingerprinting. *ICCIIT*, 1–6.

[17] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool Release: Gathering 802.11 n Traces with Channel State Information. *ACM SIGCOMM* 41, 1 (Jan 2011), 53–53.

[18] Jingyu Hua, Hongyi Sun, Zhenyu Shen, Zhiyun Qian, and Sheng Zhong. 2018. Accurate and Efficient Wireless Device Fingerprinting Using Channel State Information. In *IEEE INFOCOM*. 1700–1708.

[19] Jafar Haadi Jafarian, Amirreza Niakanlahiji, Ehab Al-Shaer, and Qi Duan. 2016. Multi-Dimensional Host Identity Anonymization for Defeating Skilled Attackers. In *Proceedings of the 2016 ACM Workshop on Moving Target Defense*. 47–58.

[20] Steven M. Kay. 1993. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall.

[21] Guyue Li, Jiabao Yu, Yuexiu Xing, and Aiqun Hu. 2019. Location-Invariant Physical Layer Identification Approach for WiFi Devices. *IEEE Access* 7 (Aug 2019), 106974–106986.

[22] Hongbo Liu, Yang Wang, Jie Yang, and Yingying Chen. 2013. Fast and Practical Secret Key Extraction by Exploiting Channel Response. In *IEEE INFOCOM*. 3048–3056.

[23] P. Liu, P. Yang, W. Song, Y. Yan, and X. Li. 2019. Real-time Identification of Rogue WiFi Connections Using Environment-Independent Physical Features. In *IEEE INFOCOM*. 190–198.

[24] R. Miller and C. B. Chang. 1978. A modified Cramér-Rao bound and its applications (Corresp.). *IEEE Transactions on Information Theory* 24, 3 (May 1978), 398–400.

[25] Sangho Oh, Tam Vu, Marco Gruteser, and Suman Banerjee. 2012. Phantom: Physical Layer Cooperation for Location Privacy Protection. In *IEEE INFOCOM*. 3061–3065.

[26] Yue Qiao, Ouyang Zhang, Wenjie Zhou, Kannan Srinivasan, and Anish Arora. 2016. PhyCloak: Obfuscating Sensing from Communication Signals. In *USENIX NSDI*. 685–699.

[27] Hanif Rahbari and Marwan Krunz. 2014. Friendly CryptoJam: A Mechanism for Securing Physical-layer Attributes. In *Proceedings of the 2014 ACM conference on Security and privacy in wireless & mobile networks*. 129–140.

[28] Hanif Rahbari and Marwan Krunz. 2015. Secrecy Beyond Encryption: Obfuscating Transmission Signatures in Wireless Communications. *IEEE Communications Magazine* 53, 12 (2015), 54–60.

[29] Hanif Rahbari and Marwan Krunz. 2015. Secrecy beyond encryption: obfuscating transmission signatures in wireless communications. *IEEE Communications Magazine* 53, 12 (2015), 54–60.

[30] Kasper Bonne Rasmussen and Srdjan Capkun. 2007. Implications of Radio Fingerprinting on the Security of Sensor Networks. In *EAI SecureComm*. 331–340.

[31] Pieter Robyns, Bram Bonné, Peter Quax, and Wim Lamotte. 2017. Noncooperative 802.11 MAC Layer Fingerprinting and Tracking of Mobile Devices. *Security and Communication Networks* (2017).

[32] T M Schmidl and D C Cox. 1997. Robust frequency and timing synchronization for OFDM. *IEEE Transactions on Communications* 45, 12 (1997), 1613–1621.

[33] Matthias Schulz, Jakob Link, Francesco Gringoli, and Matthias Hollick. 2018. Shadow Wi-Fi: Teaching Smartphones to Transmit Raw Signals and to Extract Channel State Information to Implement Practical Covert Channels over Wi-Fi. (Jun 2018), 256–268.

[34] Matthias Schulz, Daniel Wegemer, and Matthias Hollick. 2017. Nexmon: The C-based Firmware Patching Framework. https://nexmon.org

[35] IEEE Computer Society. 2016. *802.11-2016: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. Technical Report. IEEE.

[36] Mathy Vanhoef, Célestin Matte, Mathieu Cunche, Leonardo S Cardoso, and Frank Piessens. 2016. Why MAC Address randomization Is Not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. In *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*. 413–424.

[37] Mathy Vanhoef and Frank Piessens. 2017. Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2. In *CCS ACM SIGSAC*. 1313–1328.
[38] Yaxiong Xie, Zhenjiang Li, and Mo Li. 2018. Precise Power Delay Profiling with Commodity Wi-Fi. *IEEE Transactions on Mobile Computing* 18, 6 (Sep 2018), 1342–1355.
[39] Qiang Xu, Rong Zheng, Walid Saad, and Zhu Han. 2015. Device Fingerprinting in Wireless Networks: Challenges and Opportunities. *IEEE Communications Surveys & Tutorials* 18, 1 (2015), 94–104.
[40] Yao Yao, Yan Li, Xin Liu, Zicheng Chi, Wei Wang, Tiantian Xie, and Ting Zhu. 2018. Aegis: An Interference-negligible RF Sensing Shield. In *IEEE INFOCOM*. 1718–1726.
[41] Yiwei Zhuo, Hongzi Zhu, Hua Xue, and Shan Chang. 2017. Perceiving Accurate CSI Phases with Commodity WiFi Devices. In *IEEE INFOCOM*. 1–9.

A KRONECKER PRODUCT PROPERTIES

Property 1 *(Transpose of a Kronecker product)*: Let $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{B} \in \mathbb{C}^{r\times s}$, then $(\mathbf{A}\otimes\mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T$.

Property 2 *(Product of two Kronecker products)*: Let $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{B} \in \mathbb{C}^{r\times s}$, $\mathbf{C} \in \mathbb{C}^{n\times p}$, and $\mathbf{D} \in \mathbb{C}^{s\times t}$, then $\mathbf{AB}\otimes\mathbf{CD} = (\mathbf{A}\otimes\mathbf{C})(\mathbf{B}\otimes\mathbf{D})$.

Property 3 *(Trace of a Kronecker product of matrices)*: Let $\mathbf{A} \in \mathbb{C}^{m\times m}$, $\mathbf{B} \in \mathbb{C}^{n\times n}$, then $\mathrm{Tr}(\mathbf{A}\otimes\mathbf{B}) = \mathrm{Tr}(\mathbf{A})(\mathbf{B})$.

Property 4 *(Cyclic permutation of the trace)*: Let $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{B} \in \mathbb{C}^{n\times m}$, then $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$.

Property 5 *(Trace of a Kronecker product of vectors)*: Let $\mathbf{a} \in \mathbb{C}^{m\times 1}$, $\mathbf{b} \in \mathbb{C}^{m\times 1}$, then $\mathrm{Tr}(\mathbf{a}\otimes\mathbf{b}^T) = \mathrm{Tr}(\mathbf{a}\mathbf{b}^T)$.

B CRAMER-RAO BOUND OF RF-SCOPE

We analyze the performance of RF-Scope and compare it to the Cramer-Rao bound (CRB) bound. We show that RF-Scope is a near-optimal estimator of the CSI, as defined in (B22). For notation simplicity and without loss of generality, in the sequel, we drop the subcarrier index k and consider the analysis for a single subcarrier for which N measurements are available.

Let m_n be a measurement (or observation) in a given subcarrier defined as

$$m_n = he^{jZ_n} + w_n, \tag{B1}$$

where Z_n is a random phase rotation and h is the complex-valued channel. Recalling Section 6, Z_n is introduced by our proposed approach RF-Veil to prevent attackers from acquiring the channel accurately. Thus, let $p(m_n \mid Z_n; h)$ denote the joint likelihood function of Z_n and h, given the observation m_n

$$p(m_n \mid Z_n; h) = \frac{1}{\sqrt{\pi\sigma^2}} e^{-\frac{1}{\sigma^2}\left|m_n - he^{jZ_n}\right|^2}. \tag{B2}$$

For N uncorrelated measurements, we have the likelihood function

$$p(m_1, \cdots, m_N \mid Z_1, \cdots, Z_N; h) = \Pi_{n=1}^{N} \frac{1}{\sqrt{\pi\sigma^2}} e^{-\frac{1}{\sigma^2}\left|m_n - he^{jZ_n}\right|^2}, \tag{B3}$$

which can be equivalently recast as

$$p(\mathbf{m} \mid \mathbf{Z}; h) = \frac{1}{(\pi\sigma^2)^{N/2}} e^{-\frac{1}{\sigma^2}\left\|\mathbf{m} - he^{j\mathbf{Z}}\right\|_2^2}, \tag{B4}$$

where $\mathbf{m} = [m_1, \cdots, m_N]^T$ and $\mathbf{Z} = [Z_1, \cdots, Z_N]^T$. To compute the CRB of h, we require the likelihood function $p(\mathbf{m}; h)$. Note that this function can be obtained through averaging $p(\mathbf{m} \mid \mathbf{Z}; h)$ over the random nuisance variables $\mathbf{Z}$. Thus, the likelihood function $p(\mathbf{m}; h)$ is computed as

$$\begin{aligned} p(\mathbf{m}; h) &= \mathbb{E}_{\mathbf{Z}}\left[\frac{1}{(\pi\sigma^2)^{N/2}} e^{-\frac{1}{\sigma^2}\left\|\mathbf{m} - he^{j\mathbf{Z}}\right\|_2^2}\right], \\ &= \int_{\mathcal{D}_{\mathbf{Z}}} \frac{1}{(\pi\sigma^2)^{N/2}} e^{-\frac{1}{\sigma^2}\left\|\mathbf{m} - he^{j\mathbf{Z}}\right\|_2^2} p_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z}, \end{aligned} \tag{B5}$$

where $\mathbf{z} = [z_1, \cdots, z_N]^T$ denote the integration variables, and $\mathcal{D}_{\mathbf{Z}}$ is the domain of the random variables $\mathbf{Z}$. In addition, $\mathbb{E}_{\mathbf{Z}}$ denotes statistical expectation with respect to $\mathbf{Z}$, which has a priori probability density function $p_{\mathbf{Z}}(\mathbf{z})$. Assuming that the random phases are independent, then $p_{\mathbf{Z}}(\mathbf{z}) = \prod_{n=1}^{N} p_{Z_n}(z_n)$.

Thus, (B5) can be expressed as

$$p(\mathbf{m};h) = \int_{\mathcal{D}_{Z_1}} \cdots \int_{\mathcal{D}_{Z_N}} \frac{1}{(\pi\sigma^2)^{N/2}} e^{-\frac{1}{\sigma^2}\sum_{n=1}^{N}|m_n - he^{jZ_n}|^2} p_{Z_1}(z_1)\cdots p_{Z_N}(z_N)\, dz_1 \cdots dz_N. \tag{B6}$$

The CRB of any unbiased estimator $\widehat{h}$ of the channel h is given by

$$\text{CRB}\left(\widehat{h}\right) = \mathbb{E}_{\mathbf{m}}\left[\frac{\partial}{\partial h}\ln p(\mathbf{m};h)\frac{\partial}{\partial h^*}\ln p(\mathbf{m};h)\right]^{-1}, \tag{B7}$$

where $\mathbb{E}_{\mathbf{m}}$ denotes statistical expectation with respect to $\mathbf{m}$ [20]. Nonetheless, the computation of this expression is analytically intractable due to the embedded integration with respect to the random variables $Z_1, \cdots, Z_N$. As a result, a simpler (but looser) bound called the modified CRB (MCRB) has been derived in [10, 24]. Specifically, the MCRB is a lower bound of the CRB, i.e., $\text{MCRB}\left(\widehat{h}\right) \leq \text{CRB}\left(\widehat{h}\right)$ is defined as

$$\text{MCRB}\left(\widehat{h}\right) = \mathbb{E}_{\mathbf{Z}}\left[\mathbb{E}_{\mathbf{m}|\mathbf{Z}}\left[\frac{\partial}{\partial h}\ln p(\mathbf{m}\mid\mathbf{Z};h)\frac{\partial}{\partial h^*}\ln p(\mathbf{m}\mid\mathbf{Z};h)\right]\right]^{-1}. \tag{B8}$$

From (B6), we compute the derivatives with respect to h and h^*,

$$\frac{\partial}{\partial h}\ln p(\mathbf{m}\mid\mathbf{z};h) = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(e^{jZ_n}m_n^* - h^*\right) = \frac{1}{\sigma^2}\sum_{n=1}^{N} w_n^* e^{jZ_n}, \tag{B9}$$

$$\frac{\partial}{\partial h^*}\ln p(\mathbf{m}\mid\mathbf{z};h) = \frac{1}{\sigma^2}\sum_{n=1}^{N}\left(e^{-jZ_n}m_n - h\right) = \frac{1}{\sigma^2}\sum_{n=1}^{N} w_n e^{-jZ_n}, \tag{B10}$$

Upon replacing (B9) and (B10) in (B8), we obtain that

$$\begin{aligned}
\text{MCRB}\left(\widehat{h}\right) &= \mathbb{E}_{\mathbf{Z}}\left[\mathbb{E}_{\mathbf{m}|\mathbf{Z}}\left[\frac{1}{\sigma^2}\sum_{n=1}^{N} w_n^* e^{jZ_n}\frac{1}{\sigma^2}\sum_{l=1}^{N} w_l e^{-jZ_l}\right]\right]^{-1},\\
&= \mathbb{E}_{\mathbf{Z}}\left[\mathbb{E}_{\mathbf{w}|\mathbf{Z}}\left[\frac{1}{\sigma^2}\sum_{n=1}^{N} w_n^* e^{jZ_n}\frac{1}{\sigma^2}\sum_{l=1}^{N} w_l e^{-jZ_l}\right]\right]^{-1},\\
&= \mathbb{E}_{\mathbf{Z}}\left[\mathbb{E}_{\mathbf{w}|\mathbf{Z}}\left[\frac{1}{\sigma^4}\sum_{n=1}^{N}\sum_{l=1}^{N} w_n^* w_l e^{jZ_n}e^{-jZ_l}\right]\right]^{-1},\\
&= \mathbb{E}_{\mathbf{Z}}\left[\mathbb{E}_{\mathbf{w}|\mathbf{Z}}\left[\frac{1}{\sigma^4}\sum_{n=1}^{N}|w_n|^2 + \frac{1}{\sigma^4}\sum_{n=1}^{N}\sum_{l\neq n}^{N} w_n^* w_l e^{jZ_n}e^{-jZ_l}\right]\right]^{-1},\\
&= \mathbb{E}_{\mathbf{Z}}\left[\mathbb{E}_{\mathbf{w}|\mathbf{Z}}\left[\frac{1}{\sigma^4}\sum_{n=1}^{N}|w_n|^2\right] + \mathbb{E}_{\mathbf{w}|\mathbf{Z}}\left[\frac{1}{\sigma^4}\sum_{n=1}^{N}\sum_{l\neq n}^{N} w_n^* w_l e^{jZ_n}e^{-jZ_l}\right]\right]^{-1},\\
&= \mathbb{E}_{\mathbf{Z}}\left[\frac{1}{\sigma^4}\sum_{n=1}^{N}\mathbb{E}_{w_n|\mathbf{Z}}\left[|w_n|^2\right] + \frac{1}{\sigma^4}\sum_{n=1}^{N}\sum_{l\neq n}^{N}\mathbb{E}_{w_n|\mathbf{Z}}\left[w_n^*\right]\mathbb{E}_{w_l|\mathbf{Z}}\left[w_l\right]e^{jZ_n}e^{-jZ_l}\right]^{-1},\\
&= \mathbb{E}_{\mathbf{Z}}\left[\frac{N\sigma^2}{\sigma^4}\right]^{-1}.
\end{aligned} \tag{B11}$$

In the second step of (B11), $\mathbb{E}_{\mathbf{m}|\mathbf{Z}}$ has been changed to $\mathbb{E}_{\mathbf{w}|\mathbf{Z}}$ due to the direct dependence of $\mathbf{m}$ on $\mathbf{w}$ (when $\mathbf{Z}$ is fixed). Note that $\mathbb{E}_{w_n|\mathbf{Z}}[w_n] = 0$ and $\mathbb{E}_{w_n|\mathbf{z}}\left[|w_n|^2\right] = \sigma^2$ since $\mathbf{w} \sim \mathcal{CN}(0,\sigma^2\mathbf{I})$. Thus, $\sum_{n=1}^{N}\mathbb{E}_{w_n|\mathbf{z}}\left[|w_n|^2\right] = N\sigma^2$ and $\sum_{n=1}^{N}\sum_{l\neq i}^{N}\mathbb{E}_{w_n|\mathbf{z}}\left[w_n^*\right]\mathbb{E}_{w_l|\mathbf{z}}\left[w_l\right]e^{jZ_n}e^{-jZ_l} = 0$ yielding

$$\text{MCRB}\left(\widehat{h}\right) = \frac{\sigma^2}{N}. \tag{B12}$$

From (B12), we realize that the performance of an optimal estimator $\hat{h}$ improves with N. Essentially, as more measurements become available, the estimation error decreases. From Section 5.1, the channel estimated by RF-Scope for a single subcarrier was found to be $u = \frac{1}{N}\sum_{n=1}^{N} m_n$. To evaluate the performance of RF-Scope we compute its mean square error (MSE). To this purpose, we assume that the random phase rotations are distributed according to a Gaussian probability density function

defined as $p_{Z_n}(z_n) = \frac{1}{\sqrt{2\pi\xi^2}} e^{-\frac{(z_n-\mu)^2}{2\xi^2}}$ with mean μ and variance ξ^2. Note that ξ^2 and μ are the same for all the measurements because these are collected for a single subcarrier. Thus,

$$\begin{aligned}
\text{MSE}\,(u) &= \mathbb{E}\left[(u-h)^*\,(u-h)\right], \\
&= \mathbb{E}\left[\frac{1}{N^2}\sum_{n=1}^{N}\sum_{i=1}^{N} m_n^* m_i - \frac{h^*}{N}\sum_{n=1}^{N} m_n - \frac{h}{N}\sum_{n=1}^{N} m_i^* + |h|^2\right], \\
&= \underbrace{\mathbb{E}\left[\frac{1}{N^2}\sum_{n=1}^{N}\sum_{i=1}^{N} m_n^* m_i\right]}_{S_1} - \underbrace{\mathbb{E}\left[\frac{h^*}{N}\sum_{n=1}^{N} m_n\right]}_{S_2} - \underbrace{\mathbb{E}\left[\frac{h}{N}\sum_{n=1}^{N} m_i^*\right]}_{S_2^*} + \mathbb{E}\left[|h|^2\right],
\end{aligned} \tag{B13}$$

Now, by using (B1), we expand S_1

$$\begin{aligned}
S_1 &= \mathbb{E}\left[\frac{1}{N^2}\sum_{n=1}^{N}\sum_{i=1}^{N} m_n^* m_i\right], \\
&= \frac{|h|^2}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\sum_{i=1}^{N} e^{-j(Z_n-Z_i)}\right] + \frac{h}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\sum_{i=1}^{N} w_n^* e^{jZ_i}\right] + \frac{h^*}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\sum_{i=1}^{N} w_i e^{-jZ_n}\right] + \frac{1}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\sum_{i=1}^{N} w_n^* w_i\right] \\
&= \frac{|h|^2}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\sum_{i=1}^{N} e^{-j(Z_n-Z_i)}\right] + \underbrace{\frac{h}{N^2}\sum_{n=1}^{N}\sum_{i=1}^{N}\mathbb{E}\left[w_n^*\right]\mathbb{E}\left[e^{jZ_i}\right]}_{0} + \underbrace{\frac{h^*}{N^2}\sum_{n=1}^{N}\sum_{i=1}^{N}\mathbb{E}\left[w_i\right]\mathbb{E}\left[e^{-jZ_n}\right]}_{0} + \underbrace{\frac{1}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\sum_{i=1}^{N} w_n^* w_i\right]}_{N\sigma^2}, \\
&= \frac{|h|^2}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\sum_{i=1}^{N} e^{-j(Z_n-Z_i)}\right] + \frac{\sigma^2}{N}
\end{aligned} \tag{B14}$$

The sum of complex exponentials in (B14) can be equivalently expressed as,

$$\begin{aligned}
\sum_{n=1}^{N}\sum_{i=1}^{N} e^{-j(Z_n-Z_i)} &= N + \sum_{n=1}^{N}\sum_{i\neq n}^{N} e^{-j(Z_n-Z_i)} \\
&= N + 2\sum_{i=1}^{N-1}\sum_{n=i+1}^{N}\cos(Z_i - Z_n) \\
&= N + 2\sum_{l=1}^{\frac{N(N-1)}{2}}\cos(X_l) \\
&= N + N(N-1)\cos(X).
\end{aligned} \tag{B15}$$

In (B15), $X = Z_i - Z_n$, $\forall i, n$ denotes the difference of two Gaussian random variables. The resulting random variable X is also Gaussian, which can be obtained by means of the convolution theorem. Specifically, X has mean zero and twice the variance of Z_i, i.e., the probability density function of X is given by $p_X(x) = \frac{1}{\sqrt{4\pi\xi^2}} e^{-\frac{x^2}{4\xi^2}}$. Replacing (B15) in (B14), S_1 can be recast as

$$\begin{aligned}
S_1 &= \frac{|h|^2}{N^2}\mathbb{E}\left[\sum_{n=1}^{N}\sum_{i=1}^{N} e^{-j(Z_n-Z_i)}\right] + \frac{\sigma^2}{N}, \\
&= \frac{|h|^2}{N^2}\left(N + N(N-1)\mathbb{E}\left[\cos(X)\right]\right) + \frac{\sigma^2}{N}, \\
&= \frac{|h|^2}{N} + \frac{|h|^2\,(N-1)}{N}\mathbb{E}\left[\cos(X)\right] + \frac{\sigma^2}{N}, \\
&= \frac{|h|^2}{N} + \frac{|h|^2\,(N-1)}{N}\int_{-\infty}^{\infty}\cos(x)\,\frac{1}{\sqrt{4\pi\xi^2}} e^{\frac{-x^2}{4\xi^2}}\,dx + \frac{\sigma^2}{N}, \\
&= \frac{|h|^2}{N} + \frac{|h|^2\,(N-1)}{N} e^{-\xi^2} + \frac{\sigma^2}{N},
\end{aligned} \tag{B16}$$

where $e^{-\xi^2} = \int_{-\infty}^{\infty} \cos(x) \frac{1}{\sqrt{4\pi\xi^2}} e^{\frac{-x^2}{4\xi^2}} dx$. Besides, the term S_2 collapses to

$$\begin{aligned} S_2 &= \mathbb{E}\left[\frac{h^*}{N}\sum_{n=1}^{N} m_n\right] \\ &= \frac{|h|^2}{N} \cdot \mathbb{E}\left[\sum_{n=1}^{N} e^{jZ_n}\right] + \frac{h^*}{N}\mathbb{E}\left[\sum_{n=1}^{N} w_n\right] \\ &= |h|^2 \, \mathbb{E}\left[e^{jZ_n}\right] + \underbrace{\frac{h^*}{N}\mathbb{E}\left[\sum_{n=1}^{N} w_n\right]}_{0} \\ &= |h|^2 \int_{-\infty}^{\infty} e^{jZ_n} \frac{1}{\sqrt{2\pi\xi^2}} e^{\frac{-(Z_n-\mu)^2}{2\xi^2}} dZ_n \\ &= |h|^2 \, e^{-\xi^2/2} e^{j\mu}. \end{aligned} \tag{B17}$$

Collecting the results in (B16) and (B17), the MSE collapses to

$$\begin{aligned} \text{MSE}\,(u) &= \frac{|h|^2}{N} + |h|^2 e^{-\xi^2} - \frac{|h|^2 e^{-\xi^2}}{N} + \frac{\sigma^2}{N} - |h|^2 e^{-\xi^2/2} e^{j\mu} - |h|^2 e^{-\xi^2/2} e^{-j\mu} + |h|^2, \\ &= |h|^2 + |h|^2 e^{-\xi^2} - 2\,|h|^2 \cos(\mu) e^{-\xi^2/2} + \frac{|h|^2}{N} - \frac{|h|^2 e^{-\xi^2}}{N} + \frac{\sigma^2}{N}. \end{aligned} \tag{B18}$$

By definition, the MSE of any estimator consists of the bias and the variance as shown in

$$\text{MSE}\,(u) = \text{bias}\,(u)^2 + \text{var}\,(u)\,. \tag{B19}$$

The bias of the estimator is computed as

$$\begin{aligned} \text{bias}\,(u) &= \mathbb{E}\left[u - h\right], \\ &= \mathbb{E}\left[u\right] - h, \\ &= h\mathbb{E}\left[e^{jZ}\right] - h, \\ &= h\int_{-\infty}^{\infty} e^{jZ} \frac{1}{\sqrt{2\pi\xi^2}} e^{-\frac{(z-\mu)^2}{2\xi^2}} dz - h, \\ &= h e^{-\xi^2/2} e^{j\mu} - h. \end{aligned} \tag{B20}$$

Thus, the squared bias is

$$\begin{aligned} \text{bias}\,(u)^2 &= (h e^{-\xi^2/2} e^{j\mu} - h)^* (h e^{-\xi^2/2} e^{j\mu} - h), \\ &= |h|^2 + |h|^2 e^{-\xi^2} - 2\,|h|^2 \cos(\mu) e^{-\xi^2/2}. \end{aligned} \tag{B21}$$

By comparing (B18), (B19) and (B21), we can extract the variance of the estimator. Therefore,

$$\text{var}\,(u) = \frac{|h|^2}{N} - \frac{|h|^2}{N} e^{-\xi^2} + \frac{\sigma^2}{N}. \tag{B22}$$

Upon comparing (B12) and (B22), we note that a large variance ξ^2 (ξ^2 in radians) of the random variables Z_n leads to a high estimation error according to (B22). In such a case, $\text{var}\,(u) \approx \frac{|h|^2}{N} + \frac{\sigma^2}{N}$. However, for small values of ξ^2, the variance collapses to $\text{var}\,(u) \approx \frac{\sigma^2}{N}$, thus showing the equivalence between (B12) and (B22). While this observation demonstrates that the estimation error of RF-Scope is near-optimal in the variance sense, we also need to consider the bias in (B21), which is nonzero. Ideally, the estimator needs to be unbiased, i.e., $\text{bias}\,(u)^2 = 0$. As explained in Section 6.2, a legitimate user is aware of the synchronization index and key, and can therefore generate the same sequence of random numbers that yield μ (i.e., shifts of the probability density functions). As a result, a legitimate user can remove the additional shift, thus making $\mu = 0$. In contrast, for an attacker, $\mu \neq 0$. The bias for legitimate users and attackers are respectively defined as

$$\text{bias}_l\,(u)^2 = |h|^2 + |h|^2 e^{-\xi^2} - 2\,|h|^2 e^{-\xi^2/2}, \tag{B23}$$

$$\text{bias}_a\,(u)^2 = |h|^2 + |h|^2 e^{-\xi^2} - 2\,|h|^2 \cos(\mu) e^{-\xi^2/2},\ \mu \neq 0, \tag{B24}$$

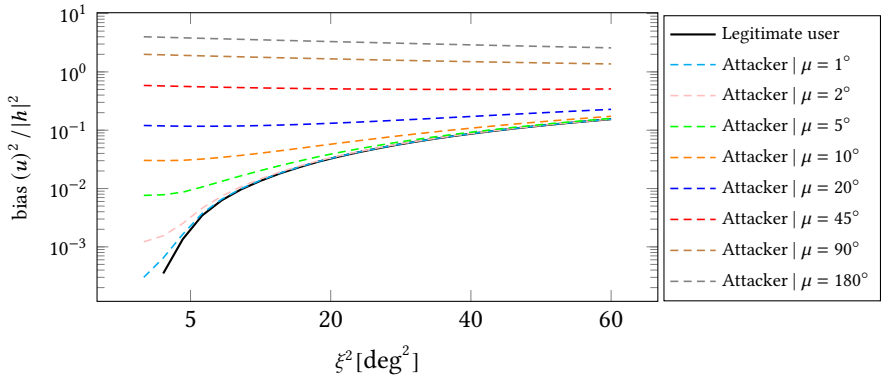


Fig. 19. Comparison of normalized biases between legitimate users and attackers considering various configurations of μ and ξ^2. *In the case of attackers, the bias increases since the additional shift μ cannot be removed. Specifically, this occurs due to the impossibility to attackers of generating the sequence of random numbers that renders μ, which can only obtained by legitimate users.*

showing that $\text{bias}_l\,(u)^2 \leq \text{bias}_a\,(u)^2$.

To illustrate the difference between (B23) and (B24), we show in Fig. 19 the biases for several configurations of μ and ξ^2. We observe that for only small $\mu = \{1°, 5°\}$ the biases of the attacker and the legitimate users are similar. However, for sufficiently large μ the difference between the two biases becomes noticeable. In our approach, `RF-Veil`, μ is not fixed but is instead randomly generated for every subcarrier using the randomization index and the key. Therefore, for potential attackers—not aware of this information—the bias for each subcarrier varies within the range of values shown in Fig. 19, hindering accurate CSI acquisition. Further, for small ξ^2 we observe that $\text{bias}_l\,(u)^2 \approx 0$, thus indicating that `RF-Scope` can be seen as an unbiased estimator in the case of legitimate users when the variance of the phase rotations is low. To clarify this aspect, in Fig. 20a we show the MCRB bound and the variance of the estimator `RF-Scope` when `RF-Veil` is used to conceal the CSI. In Fig. 20a we have neglected the effect of bias and assumed that $|h|^2 = \sigma^2 = 1$. We realize that even for large values of ξ^2, `RF-Scope` is capable of performance similarly to the MCRB bound in terms of its variance. In Fig. 20b, we consider the overall effect of bias and variance in channel estimation for both legitimate users and attackers. We observe that for legitimate users, `RF-Scope` performs near-optimally when ξ^2 is relatively small (i.e., $\xi^2 = 5$) whereas the error increases for larger ξ^2. However, as demonstrated in Section 7, even a small value of ξ^2 is effective in hindering the CSI acquisition by attackers. Thus, $\xi^2 = 5$ can be used to successfully protect the CSI and the radiometric fingerprint while assuring near-optimality. For potential attackers, the errors between 10 and 100 times higher for the shown setting.

As complementary discernment, we show non-tight MCRBs bounds for the channel magnitude, channel phase, and variance of the disturbance that reveal relations that can be used to guide the design of alternative estimators.

$$\text{MCRB}\left(\widehat{|h|}\right) = \frac{\sigma^2}{2N} \tag{B25}$$

$$\text{MCRB}\left(\widehat{\phi}\right) = \frac{\sigma^2}{2N\,|h|^2} = \frac{\text{MCRB}\left(\widehat{|h|}\right)}{|h|^2} \tag{B26}$$

$$\text{MCRB}\left(\widehat{\mathbf{z}}\right) = \frac{1}{\frac{2N|h|^2}{\sigma^2} + \frac{N}{\xi^2}} = \frac{1}{\text{MCRB}\left(\widehat{\phi}\right)^{-1} + \frac{N}{\xi^2}} \tag{B27}$$

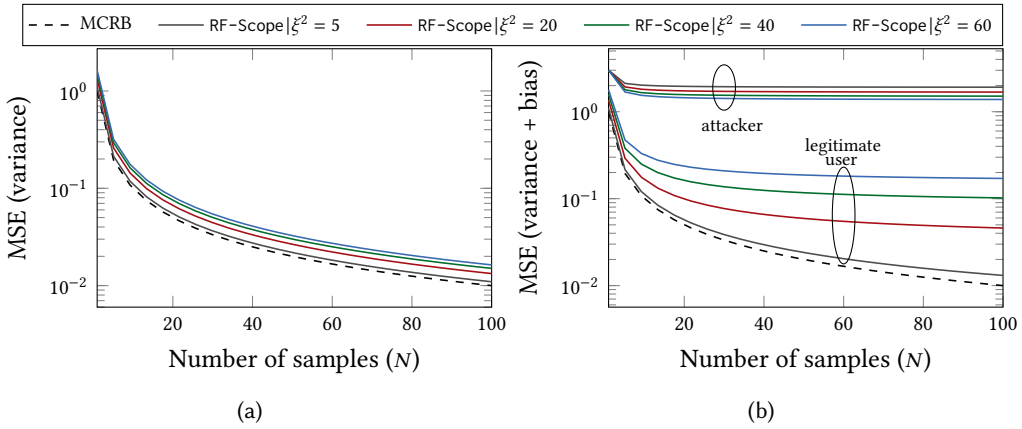


Fig. 20. Comparison of MCRB and RF-Scope when using RF-Veil as a tool to conceal the CSI from potential attackers.

C EXPERIMENTAL RESULTS FOR 802.11A

We analyze the performance of naive randomization in 802.11a[6] and show the results in Fig. 21. The results suggest that naive randomization performs similarly in 802.11a and 802.11ac, i.e., through RF-Scope the MAE error diminishes with larger N, which allows an adversary to restore the original fingerprint.

The results of RF-Veil over 802.11a are shown in Fig. 22, which are reminiscent of the behavior observed with 802.11ac in Fig. 16. In particular, the MAE is 10-fold the error achieved with naive randomization. This experiment corroborates that RF-Veil is not only feasible in 802.11ac but also in other technologies with a similar underlying structure.

D NI 802.11 APPLICATION FRAMEWORK

The implementation of the NI 802.11 AFW is separated into a host and an FPGA module, as depicted in Fig. 23a. The host module of the NI 802.11 AFW mainly implements middle MAC layer functionalities as well as a MAC high abstraction layer. The latter allows third party 802.11 higher MAC applications, such as the ns-3 network simulator, to connect to the stack. The MAC high abstraction layer is connected to MAC middle layer via UDP, which is responsible for duplicate detection in RX and synchronization index assignment in TX direction.

Note that the higher MAC abstraction layer does not implement association or authentication procedures, so the NI 802.11 AFW cannot complete the connection setup to COTS hardware. To still allow data streams for demo and measurement purposes between two instances of the 802.11 AFW, simplistic data sinks and sources (random data / User Datagram Protocol (UDP)) are available.

The FPGA module includes the implementation of the lower MAC layer as well as the whole physical layer. Even though both layers are running on the FPGA, they are executed in different clock domains, as the timing requirements are different. This results in clock rates of 100 MHz (10 nanoseconds) and 250 MHz (4 nanoseconds) for MAC and physical layer, respectively.

The main building blocks of the physical layer, as implemented on the FPGA, are depicted in Fig. 23b. In TX direction, the lower MAC layer passes the digital data as bits through the TX PHY service access point (SAP) to the physical layer. The first block (PSDU Discard) removes the PHY service data unit (PSDU) and verifies the consistency of the packet before passing it to the TX

[6]The main difference between these two technologies is the number of subcarriers, which is 52 and 56 for 802.11a and 802.11ac, respectively.

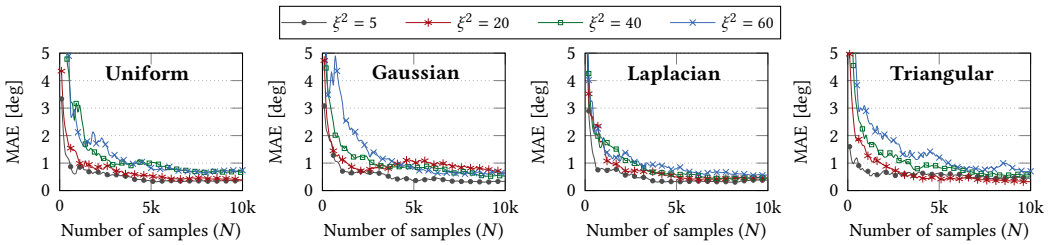


Fig. 21. Mean absolute error (MAE) for different symmetric zero-mean distributions using RF-Scope in 802.11a. *The MAE values with* 10000 *samples are below* 1° *for all distributions.*

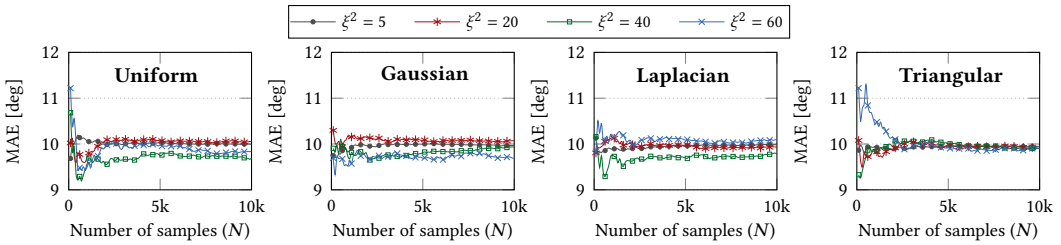
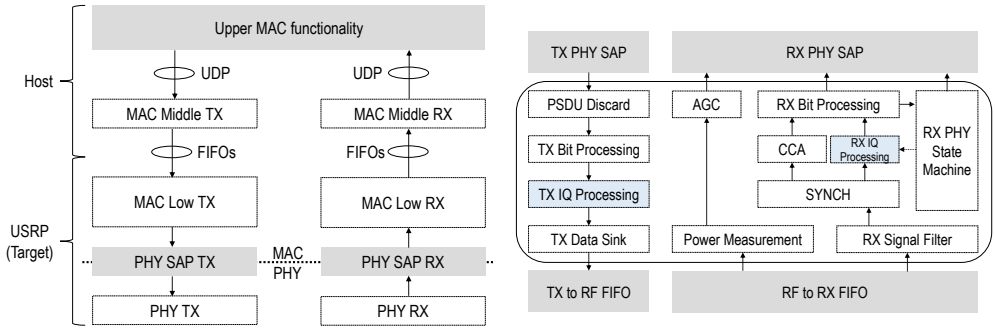


Fig. 22. Mean absolute error (MAE) of transmissions using RF-Veil in 802.11a. *The observed errors are in the order of* 10°, which shows that fingerprint extraction by adversaries is not feasible.

Bit Processing block. This following block takes care of serialization, scrambling, convolutional encoding, puncturing, and interleaving. The subsequent TX IQ Processing block modulates the data and creates the OFDM symbols, including all training fields. While the STF is pre-calculated in time-domain, all other training fields, such as the LTFs are modulated in frequency-domain. After that, the Inverse Fast Fourier Transform (IFFT) transforms the data form the frequency domain into the time-domain. The time domain samples are written to the *TX to RF FIFO* in the last block (TX Data Sink). The final steps, consisting of up-conversion to carrier frequency and sending out the up-converted samples over the air, are left out of Fig. 23b for brevity.

The PHY code in RX direction involves more steps to accurately and correctly identify and retrieve a packet. The RX Signal Filter reads the already down-converted samples in the baseband and filters for 40 or 20 MHz channels. Upon that, the synchronization detects the packet start using the Schmidl and Cox algorithm [32], which also estimates and compensates the CFO. The Clear Channel Assessment (CCA) in the subsequent step calculates the received signal power and compares it against a given CCA threshold. The RX IQ Processing module then transforms the samples from time-domain into frequency-domain using FFT. Furthermore, it equalizes the channel and detects the format of the frame (non-HT, HT, VHT). The last step consists of transforming the IQ samples into bits, creating field assignments, bit deinterleaving, decoding, and descrambling. The RX Bit Processing writes the finished frame into the RX PHY SAP for the lower MAC layer to further process it. The RX PHY State Machine generates control information for the IQ and bit processing modules based on the bitstream. It keeps track of meta-information of the packet, such as the number of OFDM symbols inside a packet, bandwidth, and PSDU length, and signals the lower MAC layer when a packet is completely received. Parallel to this whole process, power measurements on the baseband are performed (Power Measurement) and provided to the Automatic Gain Control (AGC), which dynamically determines the gain of the amplifiers.



(a) (b)

Fig. 23. NI 802.11 implementation details. Fig. 23a shows the interfaces used to connect the different layers of the WiFi implementation. Fig. 23b provides an overview of the physical layer implementation on the FPGA. We mainly adjusted the IQ processing blocks (marked in blue) in both, RX and TX directions.