RRF: A Robust Radiometric Fingerprint System that Embraces Wireless Channel Diversity

Wenqing Yan
Uppsala University, Sweden
wenqing.yan@it.uu.se

Thiemo Voigt
Uppsala University, Sweden
thiemo.voigt@it.uu.se

Christian Rohner
Uppsala University, Sweden
christian.rohner@it.uu.se

ABSTRACT

Radiometric fingerprint schemes have been shown effective in identifying wireless devices based on imperfections in their hardware electronics. The robustness of fingerprint systems under complex channel conditions, however, is a critical challenge that makes their application in real-world scenarios difficult. We systematically evaluate the wireless channel's impact on radiometric fingerprints and find that the channel impacts fingerprint features in a very particular way that depends on the channel's properties. Based on these insights, we present RRF, a system that provides a robust identification/authentication service even under complex channel fading disturbance. Our design deploys a hybrid architecture that combines wireless channel simulation, signal processing and machine learning. In this pipeline, RRF first utilizes a series of structured channel simulations to strategically improve system tolerance towards multipath channel interference. On top of that, in the identification phase, RRF relies on noise compensation and a feature denoising filter to augment the system's stability in noisy conditions with weak signals. Our experimental results show that RRF achieves an average accuracy consistently above 99% in empirical scenarios with complex channels, where the baseline approach from previous work rarely exceeds 50%.

CCS CONCEPTS

• **Networks** → **Mobile and wireless security**; • **Security and privacy** → **Mobile and wireless security**.

KEYWORDS

Physical-layer security, Radio frequency fingerprint, Identification, Authentication

ACM Reference Format:
Wenqing Yan, Thiemo Voigt, and Christian Rohner. 2022. RRF: A Robust Radiometric Fingerprint System that Embraces Wireless Channel Diversity. In *Proceedings of the 15th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec '22), May 16–19, 2022, San Antonio, TX, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3507657.3528542

We are grateful to our shepherd, Matthias Hollick, for his guidance on how to improve this paper and clarify its arguments. We would like to thank our anonymous reviewers for their insightful feedback. This work is supported by the Swedish Foundation for Strategic Research (SSF), the Swedish Science Foundation (2018-05480) and the Swedish Civil Contingencies Agency (2018-12526).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

WiSec '22, May 16–19, 2022, San Antonio, TX, USA.
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9216-7/22/05.
https://doi.org/10.1145/3507657.3528542

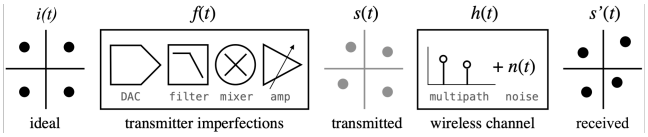


Figure 1: Challenge: Both transmitter imperfections $f(t)$ and the wireless channel $h(t)$ impact the signal (illustrated as constellation). Deviations from the ideal signal $i(t)$ in the transmitted signal $s(t) = i(t) * f(t)$ are unique for each device, regarded as the fingerprint. However, only the received signal $s'(t) = s(t) * h(t)$ is observable. As a consequence, channel perturbations challenge radiometric fingerprint system. *In this paper, we address the challenge with signal processing and classification optimization.*

1 INTRODUCTION

Radiometric fingerprint schemes leverage analog and digital hardware imperfections to distinguish radio frequency (RF) transmitters. Transmitter imperfections manifest themselves in the radiated signal through deviation from the standardized ideal signal, which can be used to identify individual devices. Radiometric fingerprints are regarded as promising physical-layer identifications to augment or replace cryptographic authentications, for the reason that bit-level identification credentials are easy to replicate and commonly exploited by hackers [38]. The purely passive nature of this technique requires no extra resources for end devices which fits well with resource-constrained applications, e.g., wireless sensor networks.

Radiometric fingerprint systems are appealing but difficult in practice. The characteristic properties of the received signal are a compound of both the characteristic imperfections of the transmitter and the transient impact of the wireless channel. Figure 1 illustrates the transmitter and wireless channel components that impact the ideal signal. Imperfections of different transmitter components are manifested in a characteristic radiated signal that, unfortunately, is not directly received by the receiver. Instead the received signal is impacted by multipath, noise and Doppler effects from the wireless channel. It is challenging to perfectly isolate transmitter imperfections from the received signal, particularly when the impact of the wireless channel dominates and hides the transmitter imperfections. This problem is exacerbated with the number of devices to be distinguished as the likelihood to have devices with similar characteristic imperfections increases.

In realistic scenarios, wireless channel properties change due to user movement and environment dynamics. Already small changes in the environment can lead to considerable distortions in magnitude, phase and even frequency of the received signal [10]. We show with simple experiments in the beginning of this paper that these distortions indeed challenge the accuracy of radiometric fingerprint systems severely. Wenhao et al. investigate the real-world constraints of radiometric fingerprint techniques with a theoretical model and experiments [36] and come to similar conclusions.

Most existing fingerprint schemes emphasize the ability to differentiate devices but omit the necessary design to enhance the robustness under diverse wireless channel conditions [2, 6, 25, 28, 34].

A few systems have incorporated components to handle certain impacts of the wireless channel [3, 40, 41], but do not systematically analyse and address different channel impacts specifically. Besides, channel estimation and equalization, applied to compensate the impact of the wireless channel, cannot fully distinguish between transmitter imperfections and the wireless channel either [12]. To avoid this, we choose to embrace channel interference through classifier optimization. Towards this end, RRF first uses carefully-selected channel simulations to strategically improve classifier tolerance towards channel interference.

This paper focuses on the robustness of radiometric fingerprint system under diverse wireless channels. Our system design considers pertinent strategies to deal with different channel impacts. The key insight of this work is that the channel impacts fingerprint features in a very particular way that depends on channel properties. We use this insight in the training phase and during feature extraction to systematically adjust the decision boundaries of the fingerprint classifier to improve its robustness against the wireless channel. We demonstrate that already a simple machine learning (ML) model, combined with an established feature representation and modest signal processing, is sufficient to take the robustness of radiometric fingerprint system to a new level. We call our approach RRF: a Robust Radiometric Fingerprint system and make the following contributions:

(1) *Fundamental insights:* We use three properties of wireless channels, namely multipath, signal to noise ratio (SNR) and Doppler effect, to systematically evaluate the wireless channel's impact on radiometric fingerprint features through controlled wireless channel simulation and real-world experiments.
(2) *System design:* RRF is a system that embraces the wireless channel distortion to enhance fingerprint robustness. It is based on a hybrid pipeline of wireless channel simulation, signal processing and ML-based classification algorithm.
(3) *Systematic evaluation:* We demonstrate RRF's performance in a variety of everyday environments considering small-scale fading that covers spatial variation, multipath, through-wall and channel dynamics. Our experiments with 40 IEEE 802.15.4 compliant devices show that RRF provides a robust identification/authentication service with an average accuracy consistently above 99% while the baseline approach from previous work [3] rarely exceeds 50% in challenging environments.

2 BACKGROUND AND MOTIVATION

In this section, we first give an overview of the common architecture of radiometric fingerprint systems. Then we review the existing fingerprint schemes and reiterate the research gaps. At the end, we perform preliminary experiments to validate the wireless channel challenges for the fingerprint system.

2.1 A Primer on Radiometric Fingerprint

Radiometric fingerprint systems aim to leverage the unique characteristics of wireless signal transmitted by the devices as evidence to verify their identities. The key enabler is that transmitter imperfections introduced by the manufacturing process are manifested in the signal, which serve as unique identity [33]. This is a promising approach to enhance the security of wireless networks. A typical system architecture consists of multiple deployed wireless devices and a dedicated fingerprint identifier/authenticator (gateway).

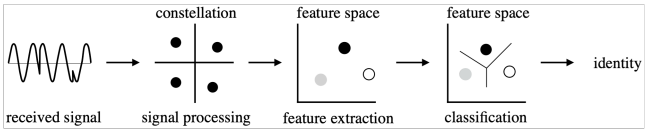


Figure 2: A typical fingerprint pipeline at the receiver. In feature space, different shades of gray represent the fingerprint from different devices.

The system works as below: The transmitters encapsulate data into the frame and generate RF signals for communication that follow their specific physical-layer standard. After acquiring the signals, the receiver refines the received signal's physical-layer characteristics as so-called features and forms fingerprints to differentiate them. Figure 2 illustrates an example of such a pipeline. In general, the core of fingerprint systems is a typical pattern classification problem that is divided into training (enrollment) and testing (identification) procedures. The first time a device joins the network, the fingerprint of the device is collected at the receiver and enrolled as a device profile linked with its ID. The profiles of all registered devices are stored in a library that is used to train a supervised classifier. In the identification procedure, the receiver extracts the fingerprint for the incoming frame. Then the well-trained classifier judges the corresponding identity. The state-of-the-art methods to implement the classification functionality are data-driven ML algorithms, such as support vector machines (SVMs) [3], k-nearest neighbors (kNNs) [14], or neural network (NN) [29].

2.2 Security and Robustness

To measure the security level of a radiometric fingerprint system, the accuracy to uniquely identify a target among a set of possible candidates is used as evaluation standard [31]. In this paper, we use true accept rate (TAR) and false accept rate (FAR) for this purpose: TAR is the probability that a system correctly authenticates an authorized device, and FAR is the probability of incorrectly authenticating an unauthorized or malicious device. Security applications have strict accuracy requirements. An empirical study about widely deployed biometric fingerprint systems considers an average TAR below 80% as poorly performing [37].

Another important aspect of a radiometric fingerprint system is its **robustness** to maintain the same identification accuracy over multiple measurements in different wireless environments. In realistic scenarios, wireless channel properties change unpredictably over time due to user movement and environment dynamics. If a fingerprint system is not robust, already slight changes in the environment make the identification accuracy decline significantly. This will be expressed in a decrease of the TAR and an increase of the FAR. Authentic devices would suffer considerably as they are likely to fail identification. Moreover, unauthorized devices may be mistakenly treated as legitimate and gain access. Hence to be practical and secure, a fingerprint system must be robust. This paper focuses on enhancing the robustness of radiometric fingerprint system.

2.3 Previous Attempts to Enhance Robustness

Previous work attempts to enhance robustness of radiometric fingerprint systems along three lines of research: (1) more advanced

ML techniques, e.g., deep learning directly operating on raw data samples [32], (2) better features, e.g., targeting the modulation error with a sliding window to compensate for random noise [3], and (3) signal processing, e.g., similar to equalization, eliminating channel impacts with function models [28, 40]. In other words, the approaches cover the entire fingerprint pipeline, but do not systematically analyse and address the channel impact specifically.

Brik et al. build on modulation-based features and take an interesting approach by aggregating data to remove the variance of the channel impact [3]. We will show that aggregation is effective against statistical noise but not structural channel properties. Their approach inspired us to have a closer look at the impact of the channel on the modulation-based features. We also use this work as the baseline in our evaluation.

2.4 Feature Definitions

As first, we introduce the fingerprint features we use in this paper. It is not our goal to introduce a new radiometric fingerprint. Rather, we focus on investigating the robustness of such a scheme against wireless channel diversity. We implement the feature extraction and classifier to model as closely as possible the approach described in earlier work [3]. We introduce the definitions as following and include more details in Appendix A.

2.4.1 Synchronization-based Features.

Carrier Frequency Offset (CFO). As for the imperfections of the local oscillator (LO), the actual signal frequency between transmitters might vary up to $\pm 10ppm$ (i.e., up to $\pm 24.5kHz$ for the $2.45GHz$ frequency band)[16]. For successful demodulation, the receiver needs to align its LO frequency with the transmitter as part of the carrier synchronization process. CFO is a measure for the magnitude of this alignment.

2.4.2 Constellation-related Features.

SFD Correlation (SFDC). After synchronization, the similarity between observed I/Q values and the ideal symbol sequence is another feature that is quantified with a correlation metric. To calculate this feature, we use a field in the frame header that is identical for all transmissions, such as the one-byte start of frame delimiter (SFD). We define the feature as below:

$$SFDC = \frac{1}{len(SFD)} \sum_{k \in \text{SFD}} I[k] \cdot S[k] \tag{1}$$

where $I[k]$ is the ideal symbol and $S[k]$ is the imperfect symbol at timestamp k.

Magnitude Error. This metric is defined as the absolute magnitude difference between ideal symbol and detected symbol, which can be estimated as:

$$MagError = \frac{1}{N} \sum_{k=1}^{N} ||S[k]| - |I[k]|| \tag{2}$$

where N is the number of symbols in the payload.

Phase Error. This metric is the absolute phase deviation between ideal symbol and detected symbol, which can be estimated as below:

$$PhaseError = \frac{1}{N} \sum_{k=1}^{N} |\angle S[k] - \angle I[k]| \tag{3}$$

I/Q Offset. Some hardware imperfections such as I/Q imbalance result in the asymmetric character of phase and magnitude errors, so the center of the I/Q axes will deviate from the origin [24]. The I/Q offset is used to quantify the deviation, which is written as:

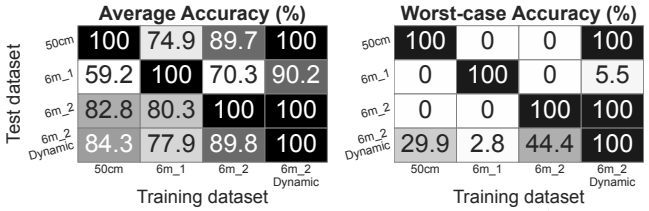


Figure 3: *Fingerprinting accuracy when model is trained and tested on different datasets collected in four setups.* The method proposed in previous research is sensitive to minor environmental dynamics such as rearranging the furniture or human mobility.

$$I/Q\ offset = \sum_{k=1}^{N} S[k] \tag{4}$$

2.5 Preliminary Experiments

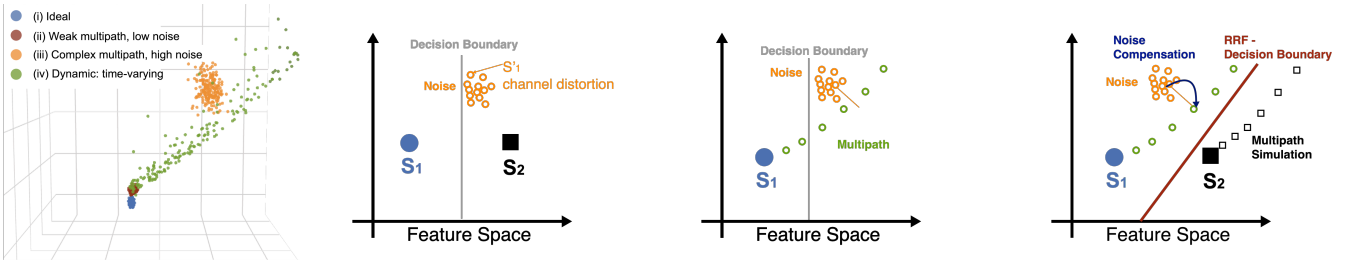
With the following experiments we show that changes in an environment common in daily life impact the wireless channel sufficiently to challenge the accuracy of radiometric fingerprint systems.

Scenarios. We present measurements from ten devices (transmitters) in an indoor meeting room with four line-of-sight (LOS) setups covering both location and environment variations: A static LOS setup with receiver (USRP) $50cm$ away from the transmitters, two static setups for $6m$ separation with the same locations for the transmitters and receiver but different furniture arrangements (moved tables and marked as $6m_1$ and $6m_2$), and dynamic movement of people in the vicinity of the receiver in the same $6m_2$ office LOS environment. We collect 2000 fingerprint samples (one sample per received frame) for each device and scenario.

Basic System Setup. In our implementation, we chose IEEE 802.15.4 as the simplest yet widely used I/Q-based modulation with a single carrier and a long symbol duration to scale away unnecessary aspects such as advanced mechanisms like channel equalization. We use SVM as the classifier and choose the same sliding window strategy to align with the configuration of previous work [3].

Evaluation Metrics. The TAR of a given device is the ratio of identified labels consistent with the actual labels. We present the results with two metrics. *Average accuracy* is defined as the average TAR over the entire device set. *Worst-case accuracy* is the lowest value among the TAR for any device, that captures the distribution of misclassifications among a population. For multi-class classification problem with unbalanced data, it is wildly used to evaluate the classification performance with this definition [21].

Poor Robustness. We use the data collected to train four classification models separately and test their performance. Figure 3 shows the results. The high accuracy along the diagonal indicates the models trained based on each dataset can provide high accuracy when we test on the data collected from the same setup. However, the accuracy drops when we test in other setups. For example, the model trained with $6m_1$ does not perform as well on the data collected in $6m_2$ and vice versa. Even worse, it is difficult to guarantee the identification quality for every single device. When testing in different environments, the worst-case accuracy is significantly lower than 50% or even falls to 0%. We argue that the difference between the three $6m$ environments is common in daily life and a robust fingerprint system should be able to handle them. In Section 6 we perform experiments with 40 devices in more challenging environments where the average accuracy of state-of-the-art approaches barely reaches 50%.



(a) *Measurements* - fingerprint samples extracted from measurements in four scenarios, projected into 3D space.

(b) *Illustration* - fingerprints affected by wireless channel. $S'1$ are likely to be outside the decision boundary.

(c) *Illustration* - multipath and noise move the fingerprints in distinct directions.

(d) *Illustration* - RRF compensates for noise and strategically adjusts the decision boundaries.

Figure 4: The key insight of this paper is that the wireless channel impacts the modulation-based features in a very particular way. The approach of RRF is to compensate for (measurable) noise and use simulation to include the multipath trajectory into the training set of the classifier. $S1$ and $S2$ indicate the collected fingerprint samples of two devices in the **enrollment environment.** The decision boundary is trained with enrolled fingerprints.

Research gap. The inconsistency of classification accuracy in various environments and locations clearly shows that wireless channel diversity is a fundamental challenge that threatens the modulation-related fingerprint system's robustness. One thing to note is that the model trained with a dynamic setup shows better robustness in two of the scenarios ($50cm$ and $6m_2$) but not in the other. The dynamic experiment investigates the time-varying channels which might cover a series of similar wireless channels as those in other setups, but there is no guarantee that it can cover all potential cases. In a nutshell, we may use the fingerprints collected under a specific wireless channel condition to train a good classification model. Unfortunately, relying on collecting fingerprints in all possible scenarios to train a robust model is impractical in many cases because measuring fingerprints systematically covering a large variety of wireless channels is labor-intensive and time-consuming.

3 INSIGHTS AND DESIGN PRINCIPLES

In this section, we present high-level insights about the key impact of the wireless channel on the radiometric features. Then we provide an overview of the design principles that guide RRF's system design.

3.1 Channel Impact on the Feature Space

In this section, we use four representative setups to demonstrate different impacts of the wireless channel: (i) an ideal environment in the anechoic chamber, (ii) weak multipath and low noise environment, two challenging setups including (iii) one static with a complex multipath channel and high noise setup and (iv) another dynamic setup with a continuously time-varying multipath channel. We randomly pick one device and look into the details of its fingerprint features. Figure 4a shows fingerprint features of 200 frames from each of the four experiments. For the visualization of the overall feature space, we use linear Principal Component Analysis (PCA) for dimension compression. PCA linearly projects the original data into the reduced space (3D) [18].

In Figure 4a, the channel's impact on the measurements in two challenging setups ((iii) and (iv)) mainly stems from multipath and noise. The samples from the measurement with weak multipath and low noise (ii) are close to the ideal samples (i). For challenging measurements, the magnitude of the deviation from the ideal sample risks to exceed the decision boundaries among different devices and thus ruin the robustness of a fingerprint system as shown in the schematic illustration Figure 4b. Our *key insight* is that the channel impacts radiometric features not at random but in a very distinct way. Through a simulation study presented in Section 4, we identify that multipath and noise are two major impact factors which deteriorate the system performance. They distort the samples into different distinct directions in the feature space. The fact that features capture different channel properties is not something new in itself [41], but the different impacts of multipath and noise is. We schematically illustrate the details in Figure 4c. In the feature space, there is an extent of the orthogonality of the two channel properties' impacts. The manner how they distort fingerprint features are uncorrelated and additive. Therefore, we can use two independent design principles to tackle multipath and noise in succession.

Multipath. The *dynamic experiment* (iv) investigates the continuously time-varying channels. In the feature space, the measured samples align in a trajectory as illustrated in Figure 4c. We hypothesize that this is mainly due to multipath effects and verify it in Section 4. Diverse multipath environments result in varying degrees of influence on the fingerprint. Inspired by this observation, we propose our first system design principle and illustrate it in Figure 4d: *By systematically modeling the fingerprint feature distortions with statistical wireless channel models, the system can optimize the decision region of the classifier for each device and improve the tolerance towards multipath channel distortion.*

Signal to Noise Ratio. We make the second observation by comparing the experiment (ii) and (iii). The measured samples with high noise setup are more spread out and diverge in an orthogonal direction from the multipath trajectory as shown in Figure 4c. We hypothesize this is due to the difference in SNR of the received signals. When the SNR is low, higher random noise in the measured signal leads to a high variance in features and the deterioration of some features. As the SNR is easy to measure, based on this phenomenon, we introduce our second design principle: *With measured SNR as an indicator, RRF compensates the affected feature and reduces the sample variance with a denoising filter.*

3.2 RRF: Robust Radiometric Fingerprint

Ideally, a robust radiometric fingerprint system would compensate for channel impacts to arrange the spread of samples in the feature space as compact as possible compared to the distance to other devices' features. This is non-trivial to achieve because it would require accurate channel estimation. As discussed previously and illustrated in Figure 1, transmitter imperfections and the wireless

channel are indistinguishable for a receiver. An attempt to compensate for the channel unavoidably involves a risk of compensating the transmitter imperfection and thus the characteristic fingerprint.

Instead of optimizing the features, RRF takes a different approach and looks for solutions from the perspective of the classifier. For data-driven ML classifiers, the decision boundary highly depends on the training data. Existing fingerprint classifiers trained with enrolled profiles only learn fingerprint characteristics under the enrollment channel condition and meet formidable challenges in practical scenarios under complex and dynamic wireless conditions.

RRF combines the insights from Section 3.1 and strategically adjusts the decision boundaries among devices by including simulation-based multipath samples into the training set during device enrollment. This results in a much more robust alignment of the decision boundaries among devices. Our approach is positively supported by the fact that the trajectories of multipath samples relatively align among all devices. Second, we compensate for noise during the identification phase by projecting the feature samples onto the narrow multipath trajectory. This is possible because first, the SNR is easy to measure and second, there is a deterministic relation between the amount of noise and the magnitude of the relevant feature to be corrected (demonstrated in Section 5). As a result, feature samples are likely to fall into the decision area independent of channel conditions. In the following, Section 4 formalizes our observations and Section 5 explains the design and implementation details.

4 WIRELESS CHANNEL DISTORTIONS

This section systematically introduces insights into how the wireless channel distorts the radiometric fingerprint. The wireless channel is characterized by radio propagation properties that impact the wireless signal. We consider multipath propagation, noise and Doppler shift, as they are the most common propagation properties in typical environments.

4.1 Methodology

To discuss the channel impact on fingerprint features, we use a hybrid simulation to give a systematic overview of the insights. We input the raw I/Q samples from ten devices measured in an ideal environment (i.e., anechoic chamber) into Matlab simulations using the Rician and Additive White Gaussian Noise (AWGN) channel models [17, 19]. The multipath simulation parameters are reflecting the two-ray channel used in the following analytical model:

$$h(t) = \rho_0 \delta(t) + \rho_1 \delta(t - \tau) \tag{5}$$

where the first ray arrives with amplitude ρ_0 and the second ray arrives τ seconds later with amplitude ρ_1. With the help of automatic gain control (AGC), the amplitudes of the two rays are normalized by $\rho = \frac{\rho_1}{\rho_0}$. The range of the parameters is chosen according to the empirical indoor channel statistics by Saleh et al. [30]. Then fingerprint features are extracted from the simulated I/Q samples and constellation-related features (except SFDC) are averaged over all symbols of a frame payload. We expect the results to give us insights about fingerprint system robustness concerning the channel; the lower the variation of fingerprint features throughout the parameter range, the more robust the fingerprint. This is the first analysis of its kind that provides details of different channel aspects.

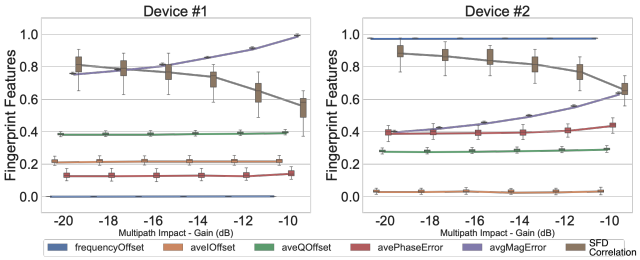


Figure 5: Multipath Gain: The side ray strength disturbs Magnitude Error and SFDC. Channel parameters are $[\rho, \tau] = [-20\ dB, ..., -10\ dB, 0.1\mu s]$. After normalization, the close-to-zero value (e.g., carrier frequency offset (CFO) for device #1) means that the corresponding feature of the device is the smallest among the tested devices.

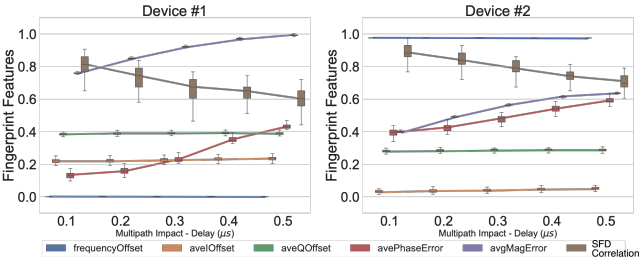


Figure 6: Multipath Delay: The side ray delay has significant impacts on Phase Error, Magnitude Error and SFDC. The channel parameters are $[\rho, \tau] = [-20\ dB, 0.1\mu s, ..., 0.5\mu s]$.

4.2 Channel Impacts

In the following we discuss the impact of multipath, signal to noise ratio and Doppler on the radiometric features based on the hybrid simulation. The interested reader finds an analytical model to introduce insights into how the wireless channel distorts the radiometric features in Appendix B.

4.2.1 Multipath. Figure 5 and Figure 6 show the (min-max) normalized features from simulation results for two devices using the Rician channel model with varying multipath gain ρ and delay τ, respectively. Our first observation is that three features are more stable than the others under multipath interference, i.e., the two I/Q offset features and carrier frequency offset (CFO). The stability of CFO indicates that the phase noise introduced by the multipath channel does not impact carrier synchronization, at least not for the range of parameters used. The other three features, magnitude error, phase error and SFDC, are susceptible to channel disturbance. A more severe multipath channel leads to a higher magnitude and/or phase error as well as lower SFDC. The phase error is less sensitive to the multipath gain than it is to the delay. The variance over the samples is low in general, except for the SFDC feature. Comparing the results between the two devices, we see a similar trend in the feature changes. The level of the features differs significantly, emphasizing that the channel impact on fingerprint features is a combination of intrinsic imperfections and the residual channel response.

4.2.2 Signal to Noise Ratio. To study the impact of noise, we add noise (i.e., AWGN) to the measured I/Q samples to adjust their SNR in a range of $[5\ dB, 40\ dB]$. Figure 7 (a) shows a similar picture as the multipath case; many of the features are constant throughout the SNR range. The big exception is the SFDC below 20 dB, and some decrease in the frequency offset below 15 dB, as the low SNR inevitably challenges the demodulation process and results in

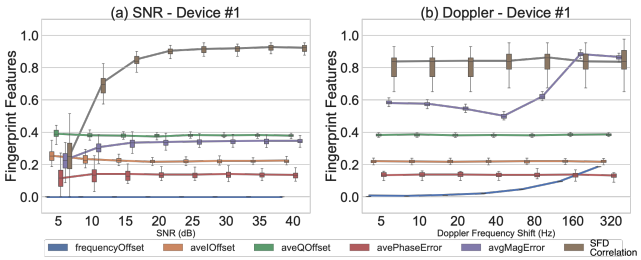


Figure 7: Noise and Doppler: (a) The AWGN noise spreads out all features except CFO and decreases SFDC. SNR of received frame varies from 5 dB to 40 dB. (b) A large Doppler frequency shift has an impact on Magnitude Error and CFO.

higher random noise left in the synchronized constellation. SFDC is a metric that directly quantifies the difference between measured samples and the ideal signal and is therefore more susceptible to noise. Noise also increases the variance across the samples for all constellation-based features.

4.2.3 Doppler. To investigate the impact of the moving transmitter/receiver on the features, we specify the value of the Doppler shift in the Rician fading channel for the dominant path and vary it from 5 Hz to 320 Hz. The Doppler spread challenges the resolution of the recovery algorithm in the coherent demodulator, which naturally disturbs the CFO. As shown in Figure 7(b), a small Doppler spread that cannot be fully resolved by the carrier recovery algorithm mainly leads to a residual phase and magnitude noise for synchronized symbols, which causes the small fluctuations of susceptible features. For a significant Doppler shift, the receiver will synchronize with the frequency involving the Doppler shift, which directly causes a noticeable deviation of the CFO feature. However, RRF's system design does not consider Doppler as a major impact factor. For common indoor deployment scenarios, the object moving speed is uncommon to meet the level which leads to fingerprint features deviate significantly. Even when the frequency shift even reaches 80 Hz which corresponds to a speed of $\approx 9.8m/s$ for 2.45 GHz, most features are stable.

5 SYSTEM DESIGN

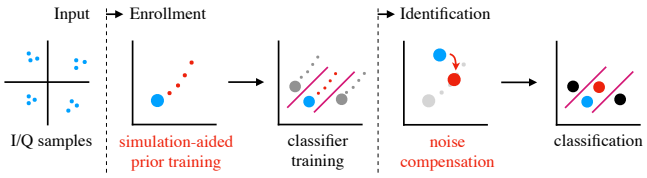
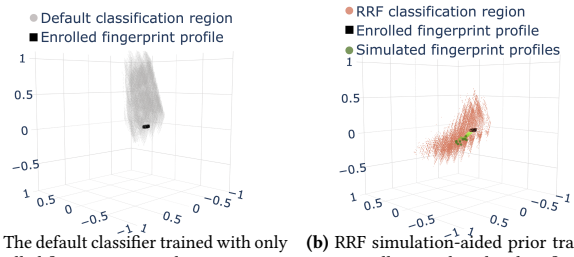


Figure 8: RRF system architecture overview. Simulation-aided prior training strategically adjusts the decision boundary to anticipate multipath samples, while noise compensation projects the noise samples onto the multipath trajectory and thus into the decision region.

The goal of RRF is to provide robust radiometric fingerprint classification even under challenging channel conditions. RRF combines the insights from the previous sections, namely that wireless channels do not impact fingerprint features in a purely random way, but rather result in distinct trajectories for multipath and noise, respectively. As outlined in Section 3, the first approach in RRF system design, namely simulation-based prior training, is to support the classifier in adjusting the decision boundaries such that they embrace multipath impacts. We implement this idea by including simulation-based samples into the training set during



(a) The default classifier trained with only enrolled fingerprint samples cannot generalize to samples measured in different environments.

(b) RRF simulation-aided prior training strategically stretches the classifier decision boundaries towards the direction of simulated fingerprint samples.

Figure 9: Comparison of classification regions (represented as point clouds) with and without prior training. For visualization in three dimensions, we use linear PCA as in Section 3.

device enrollment. The other approach is the noise compensation for scenarios with weak signals. We leverage the distinct relation between the amount of (measurable) noise and the magnitude of the SFDC feature to project the feature samples onto the narrow multipath trajectory — or in other words, into the decision area of the node in question. Figure 8 presents RRF's design.

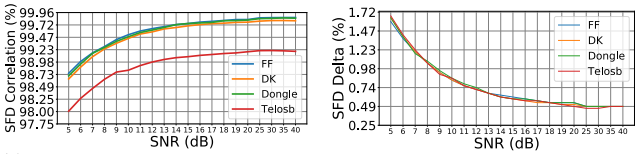
5.1 Simulation-aided Prior Training

The core idea of prior training is to utilize wireless channel knowledge to simulate fingerprints under different multipath channel conditions before practical testing. This is a type of data augmentation. Simulating fingerprints under diverse wireless channel conditions provide distinct advantages over experimenting on a physical setup. A simulation takes less time with less overhead and allows for a variety of configurations that is difficult to cover with practical measurements. Our system generates a hybrid profile for each enrolled device in the prior training, including enrolled fingerprints and simulated versions under different multipath channels, which is used to train the classifier later.

This approach is positively supported by the fact that the trajectories of the multipath samples are relatively narrow and aligned among all devices. With the simulated fingerprints, the classifier can generalize the fundamental hardware-related differences between devices better, even in different environments. In the identification phase, the classifier can implement the transferred knowledge learned with simulation into the real world.

Multipath Simulation Channel Selection. The multipath parameters in Equation 5 are not arbitrary. In a realistic indoor wireless environment, the channel parameters follow certain statistical distributions. Saleh et al. use radar-like pulses to measure an office building multipath propagation and report the following statistics [30]: the maximum delay spread τ is about $0.1\mu s$ to $0.2\mu s$ within the same room and occasionally reaches $0.3\mu s$ in the hallway. The non-line-of-sight (NLOS) path signal attenuation varies over a 60 dB (corresponding to the relative gain ρ in our model). Considering the potential difference of building structure and material, we expand the delay range up to the half symbol duration time $0.5\mu s$ [1] to achieve a higher tolerance in complex multipath propagation scenarios. With hybrid simulations, we find that weak rays (-20 dB of the dominate path) only have a negligible impact on radiometric features. Thus we only choose the gain ρ between $[-20\ dB, 0\ dB]$.

[1] IEEE 802.15.4 uses O-QPSK with half-symbol-period offset. [15]



(a) SFDC changes with SNR, raw data without normalization.

(b) The relation between SFDC Delta with SNR aligns among devices.

Figure 10: Four types devices SFDC feature changes with frame SNR, simulated with AWGN channel.

Finally, **the hybrid profiles are based on the enrolled fingerprints and their 50 simulated versions** covering the channel setups with delay τ between $[0.1\mu s, 0.5\mu s]$ with $\tau = 0.1\mu s$ as interval and relative gain ρ between $[-20\ dB, 0\ dB]$ with $-2\ dB$ as a step. The goal of selecting multipath parameters is not equal to exhausting possible multipath channels shown in the physical deployment environment. Rather, it aims to provide a guideline to the data-driven classification algorithm based on the insights we presented in Section 4. Therefore, we include extreme parameter pairs less common in realistic scenarios, which ensures a much more robust alignment of the decision boundaries among devices.

Decision Boundary Optimization. To demonstrate the benefit of prior training, we first use the hybrid profiles of ten devices with the aforementioned channel setups. Figure 9 compares the decision zone of two classifiers trained with/without the simulated profiles for one device and shows that the decision zone is optimized in the direction of the multipath trajectory. When extending the hybrid versions for more devices, Section 6 demonstrates that our designed channel region can fit 40 additional devices well. With a series of carefully selected multipath channel parameters mentioned above, the simulated fingerprints diverge away from the enrolled fingerprints to a limited extent. Moreover, as mentioned in Section 4, for all devices, along with the increase of the multipath side-ray strength, the directions of the extensions of the trajectories are roughly aligned. Thus, in most cases, the simulated profiles are in parallel with each other in different hyperplanes, rather than opposed to each other in the same hyperplane. Hence, the probability that the hybrid profile of one device interferes with another device's profile is low.

5.2 Noise Compensation

As shown in Section 4, improving the fingerprint robustness at low SNR is a two-fold challenge. Firstly, noise increases the variance of all features. Second, when the SNR is low, the SFDC feature decreases distinctly.

Denoising Filter. To enhance the system's capability to handle inputs with high variance, we have two possible options: either augment the prior training with noisy fingerprints with low SNRs; or using preprocessing methods to decrease the fingerprint feature variance in preprocessing. In our case, adding noisy fingerprints is not a wise choice as highly noisy inputs are undesirable for a stable classifier and degrade performance [13]. In order to reserve more space to tolerant multipath distortions, we apply a sliding window filter on the extracted fingerprint features before passing them to the classifier. Each received frame generates one fingerprint sample. Then a filter averages all features of the fingerprint samples in the window as one input data for the fingerprint classifier. We will discuss suitable window sizes in Section 6.

SFDC Calibration. A low signal strength typically also implies a significant deterioration of the SFDC feature value as shown in Section 4. Eliminating this feature hurts the fingerprint system accuracy and we demonstrate this in Section 6. To address the SFDC feature deterioration problem, we propose a compensation approach to calibrate the SFDC, which is based on the correlation between SFDC and SNR. We first investigate it with AWGN simulations for four different types of boards as shown in Figure 10a. We observe that the impact of AWGN is consistent among the different devices. When using the SFDC value at 40 dB SNR as reference and calculating the SFDC difference Δ for each SNR level, we get the results shown in Figure 10b. The trends of the different boards overlap with each other. Therefore, we propose a compensation approach to calibrate the SFDC according to the received frame's SNR. RRF uses a simple discrete mapping table, but it also can be implemented with a continuous function of $\Delta(snr)$.

6 EVALUATION

In this section, we first present the performance of the basic system configuration for benchmark fingerprints collected in the anechoic chamber. We then evaluate our system to demonstrate RRF's performance in a range of diverse channel setups with both simulations and experiments in a series of indoor environments.

Fingerprint System Implementation. For data acquisition, we use a USRP B210 as the receiver to capture signals in the I/Q format with a sampling rate of 16 MHz and 50 dB fixed gain. The fingerprint pipeline has two parts: the feature extractor is implemented in Matlab using the Communication toolbox, and the fingerprint classifier is written in Python using the `Sciki-learn` [4] and `Keras` [5] libraries. Matlab is also used to simulate wireless channels for prior training and evaluation.

Devices. In our evaluation, we fingerprint four types of off-the-shelf sensor boards as shown in Appendix C, in total 4×10 motes including ten TelosB boards with Texas Instruments CC2420 radio chips and integrated PCB antennas, ten Zolertia Firefly boards with Texas Instruments CC2538 system on chips (SoCs) and ceramic antennas, ten nRF52840 dks and ten nRF52840 dongles with Nordic Semiconductor multi-protocol SoCs and PCB antennas. All motes transmit frames at 2.48 GHz following the IEEE 802.15.4 protocol with 2 MHz bandwidth.

Evaluation Metrics. We use both the *average accuracy* and *worst-case accuracy* metrics introduced in Section 2. Besides, in order to differentiate between false positives and false negatives, we use average *FAR* which is the average FAR over the entire device set.

In the following, we evaluate RRF's performance by comparing different system configurations in different test scenarios. We use the following abbreviation to refer to different system components:

Table 1: System configuration abbreviation.

Component	Radiometric Fingerprints	Simulation-aided Prior Training	Sliding Window Denoising Filter	SFDC Calibration
Abbr.	M	P	W	N

6.1 Benchmark Experiments

Experimental Setup. We conduct benchmark experiments in an anechoic chamber (ideal channel) to avoid any impact of the wireless channel. We fix the USRP on one side of a wooden holder and

place the sensor motes 50 cm away, ensuring the receiver is out of the near field region. The frame transmission rate is 200 $frames/s$ with 90 B random payload, and we collect I/Q samples for 10s per device. In our implementation, fingerprint samples are frame-based. For the following evaluation, we refer to the fingerprint samples extracted from the benchmark I/Q data as **enrolled profiles**.

Classifier Selection. Our method is not designed with a target classification algorithm in mind. To select a classifier and build a solid preparation for the following evaluation, we compare four classifiers with different properties that are commonly used, namely SVM, random forest (RFC), kNN and NN. For practical reasons, before applying the classifier, we rescale the fingerprint features with a min-max normalizer into the interval $[0, 1]$. The normalized parameters extracted with the samples are used in the following experiments. We use a basic configuration with the radiometric fingerprint only (M) and exclude other components of our system design for the algorithm selection purpose. We evaluate every model with 10-fold cross-validation that reduces random performance artifacts by evaluating performance over ten non-overlapping subsets of the dataset.

The results in Table 2 show that all classifiers reach an average accuracy of at least 98.11%. The SVM and RFC algorithms slightly outperform the other two classifiers. The selection of the algorithm depends highly on the data provided and system design requirements such as the resource budgets for storage, computation, training, and running time. For our system evaluation, we use an SVM based on its advantages in terms of limited hyper-parameters, short training time and low memory storage requirements. In addition, we tested SVM with varying-size training datasets. The results showed that SVM reaches an accuracy above 94.8% with only five samples per device, 98.75% for 25 samples and 99.58% with 50 samples. For the following evaluation, we choose the SVM algorithm with 50 samples per device as classifier setup.

Table 2: Classification accuracy of four classification algorithms with 10-fold cross-validation. Accuracy% (accuracy variance among 10 fold tests%): SVM with rbf kernel and one-over-rest strategy, RFC with 100 subtrees, kNN with 5 neighbors and 2-layer NN composed of a single hidden layer with 512 rectified linear units (ReLu) neurons and an L2 regularizer.

Classifiers	SVM	RFC	kNN	NN
Accuracy	99.98% (0.02%)	99.99% (0.01%)	98.84% (0.11%)	98.11% (0.58%)

6.2 Simulation Evaluation

In this section, we evaluate the two RRF design approaches, simulation-aided prior training (P) and noise compensation (N), using controlled channel simulations. We measure the device identification robustness in terms of average accuracy and the average FAR.

Different Multipath Channels. As introduced in Section 5, prior training adds simulated fingerprint samples (generated based on benchmark I/Q data) to the training set of the classifier to anticipate fingerprints affected by multipath. To evaluate this approach, we generate a test set of fingerprint samples (500 per device), with randomly picked channel parameters. The channel parameters follow a uniform distribution over the continuous range $\tau = [0.1\mu s, 0.5\mu s]$ and $\rho = [-20\ dB, 0\ dB]$. Compared with the set used for hybrid profile, the channel setup for the test set is wider and includes both seen and unseen multipath channels for the fingerprint classifier.

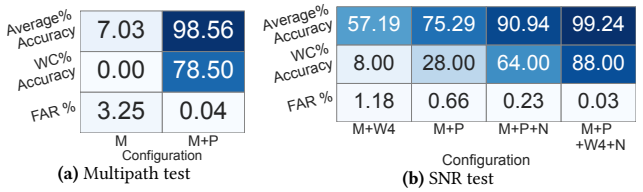


Figure 11: The simulation test sets: (a) multipath, channel parameters follow $\rho \sim U(-20, 0)dB$ and $\tau \sim U(0.1, 0.5)\mu s$, (b) SNR test set, the SNR of frame follows $SNR \sim U(5, 40)dB$. The prior training supports system overcome multipath distortions. The noise compensation eliminates the impacts of AWGN.

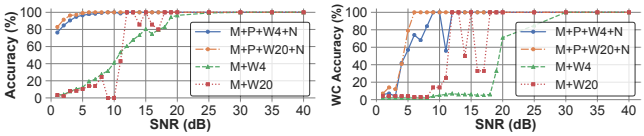


Figure 12: SNR test - classification accuracy comparison for two system configurations: reference ($M + W$) and RRF ($M + P + W + N$) with two window sizes for different frame SNR conditions varying from 1 dB to 40 dB.

Figure 11a compares the robustness of the basic configuration (M) which trains the classifier with enrolled profiles only, and the configuration using prior training ($M + P$). The results clearly show the devastating impact of the multipath channel without prior training, and this is in a parameter range well in line with empirical channel measurements [30]. In contrast, prior training leads to significant improvements, and the system reaches an average accuracy of 98%. The classifier trained with hybrid profiles indeed can arrange the decision boundaries to anticipate the multipath samples and avoid overfitting with enrolled profiles.

In addition, we evaluate the configuration with only relatively stable features (M') (CFO, I/Q offsets). The results (not in the figure) show a poor performance with worst-case accuracy 0.06% for 40 devices. This indicates that the fingerprint with few features hurts the classification accuracy of the system.

Diverse SNR Conditions. In this section, we evaluate the performance of noise compensation and complete RRF under different SNR conditions. The first test set aims to provide an overview of RRF in diverse SNR conditions. We generate the test set by adding Gaussian noise to the benchmark I/Q samples, which includes fingerprint samples extracted from frames with different SNR conditions randomly picked from 5 dB to 40 dB. The system performance is shown in Figure 11b. Comparing the baseline system ($M + W4$, averaging over 4 samples as in [3]), the system configuration with prior training ($M + P$) improves the average accuracy by 18% with only a single sample. Our complete RRF system design ($M + P + W4 + N$) can optimize the average accuracy further and reaches an average accuracy of 99%.

Besides, we also evaluate the configuration without the SFDC feature ($M'' + P + W4$) to demonstrate that this feature should not be ignored. In the results (not in the figure), the poor worst-case accuracy (0.08%) is mainly due to a misclassification between two devices of the same model, which implies that the SFDC feature brings valuable information to differentiate similar devices.

Next, we demonstrate the system performance in different SNR conditions from 1 dB to 40 dB in detail. This test set investigates the lowest SNR bound of RRF and the SNR operating region of the noise compensation design. For each SNR level we generate one test set, evaluate the classification performance and plot them in Figure 12. The results show that the sliding window denoising

filter improves system performance with increasing SNR and larger window size. In the case of AWGN, a larger window can reduce the features' variance or spread further. However, the baseline configurations alone ($M + W4$ and $M + W20$) are not sufficient to reach an acceptable performance below 20 dB, most likely because the filtered features still reach beyond the decision boundaries. RRF significantly outperforms the reference configuration in this range, where even the worst-case accuracy of our approach is strictly superior to the average accuracy of the reference configurations. RRF's classification performance deteriorates below 5 dB. However, this is at a level of SNR when even communication is challenging and most packets cannot be received anymore in IEEE 802.15.4 [11]. We consider 5 dB as the lower bound for RRF's ability to operate reliably. We also conclude that noise compensation is not needed beyond 20 dB where even the worst-case accuracy reaches 100%. We confirm this finding with experimental results in next section.

In addition, compared to the configuration $M + P + W4 + N$, a larger window size configuration ($M + P + W20 + N$) improves the worst-case accuracy to 99% for SNR around 5dB. A large window size gives higher accuracy but requires more samples and therefore implies a longer response time. Therefore, considering a practical implementation with reasonable response time, RRF chooses 20 as a window size in following evaluation (in our setup, it takes 0.1s to receive 20 frames).

6.3 Real World Experiments

In this section, we evaluate RRF in four different indoor environments with both LOS and NLOS setups with data collected over six months. We mainly consider three environmental factors: spatial variation, through-wall scenarios and channel dynamics. All system configurations are trained with 4×10 devices, but we focus the testing on ten Firefly motes of the same model as we believe that identifying devices made of the same components at the same facility is the most challenging scenario. We run a total of 13 experiments and each experiment collects 30000 fingerprint samples. We compare three configurations: radiometric fingerprint with a sliding window ($M + W$) representing state-of-the-art as a **baseline**, radiometric fingerprint with prior training ($M + P$), and ($M + P + W + N$) with prior training and noise compensation.

Preliminary Experiment Scenario. In the first experiment, we test RRF's performance with the same LOS setup as preliminary experiments presented in Section 2. The meeting room is equipped with standard office equipment such as chairs, tables, book shelves and blackboards. With LOS paths the signal maintains as strong as around 20 dB for different 6m setups. Among evaluated environments this is the least challenging, since the measurements are impacted by less multipath and have a strong SNR.

We present the results in Figure 13. In consistent with the preliminary results introduced in Section 2, the baseline configuration shows poor performance. The prior training ($M+P$) alone addresses the major multipath interference and maintains the average accuracy above 99% (95% worst-case accuracy) for all scenarios. The complete RRF system design provides the most reliable performance and consistently achieves a worst-case accuracy over 99%.

Rich Multipath Scenario. We conduct the second experiment in an aisle with many metal tubes suspended from the ceiling along the wall, which leads to complex multipath effects. We test two

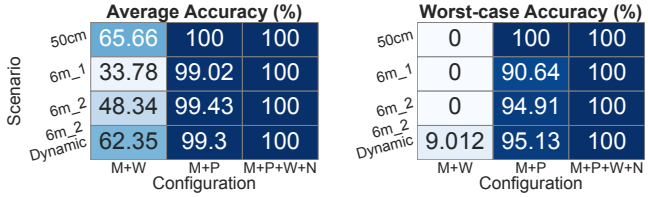


Figure 13: Indoor meeting room scenario - for the LOS setup, the baseline system $M + W$ can not reach sound performance, but RRF demonstrated significant improvements over all scenarios.

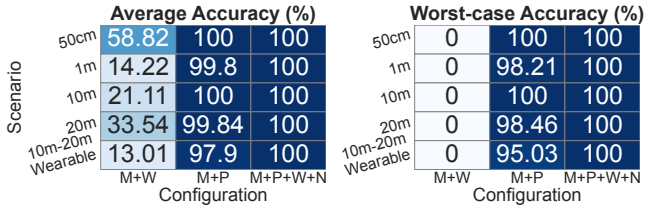


Figure 14: Rich multipath - high SNR - fingerprint performance comparison in an underground aisle with LOS setup. RRF identifies devices with 100% accuracy.

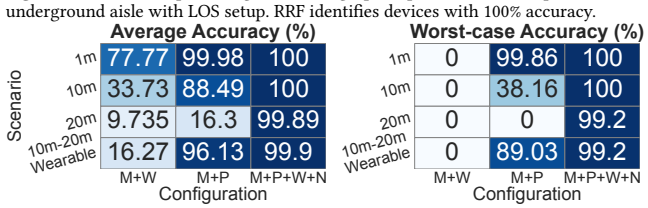


Figure 15: Corridor scenario - fingerprint performance comparison in office corridor environment with a series of systematic NLOS setups. RRF maintains an accuracy above 99% even for the dynamic setup with a moving receiver.

scenarios: a static setup with transmitters at different distances up to 20m away from the receiver, and wearable experiments where a volunteer holds the tested device waving her hand. The scenario in the aisle is to mimic a factory where multiple reflections are caused by surrounding metal equipment. The measurements show that the signal strength is strong in this environment, e.g., at 20m the SNR is around 17 dB.

The results in Figure 14 demonstrate similar results as in the first experiment. The baseline configuration ($M + W$) has consistently low accuracy, while prior training alone ($M + P$) explicitly addressing multipath reaches average an accuracy of 97% and above (92% worst-case accuracy) for all scenarios. It is interesting to note that an increasing distance does not automatically translate into worse performance in multipath environments, as the gain of the reflections is much higher at lower distances. Adding noise compensation (M+P+W+N) additionally improves performance and consistently leads to 100% accuracy in this environment.

Corridor Scenario. In this experiment, we investigate the ability of RRF to identify devices in an office environment with NLOS setups with the signal passing through multiple walls. This setup features complex multipath fading due to signal reflections on walls, glass, standard office furniture and equipment. The walls between the rooms are about 16cm thick and consist of insulated gypsum, so signal attenuation is significant when passing through the walls. The tested devices are fixed in one office room with the door closed. We systematically put the receiver at different locations with the distance to the test device ranging from 1m to 20m. Besides the static scenario, we also create a mobile scenario with higher moving speed than the wearable setup in the previous experiment, in which we put the receiver on a trolley and move it at a speed of about 2m/s back and forth in the corridor.

Figure 15 shows the results for this environment. The figure shows that prior training ($M+P$) is very effective at short distances ($1m$ and $10m$) and in the high-speed wearable scenario, and significantly improves the average accuracy. However, the worst-case accuracy is still low for $10m$ and prior training can hardly embrace the channel at $20m$ (the average SNR for $10m$ is around 13 dB and for 6 dB for $20m$). On the other hand, the complete RRF configuration including noise compensation ($M + P + W + N$) maintains a high average classification accuracy and worst-case accuracy above 94.9% even at $20m$. This indicates that the SNR is the limiting factor for longer distances in this environment.

Summary. The wireless channel in our real-world experiments severely challenges fingerprint robustness. RRF achieves a consistently good performance with both average and worst-case accuracy above 99%.

7 SECURITY ANALYSIS AND LIMITATION

In this section, we discuss security aspects of RRF. We are specifically interested if the prior training with hybrid profiles exposes new attack opportunities. We consider an adversary that can impersonate users' bit-level identification credentials but not their radiometric fingerprints. Advanced clone attacks that allow the adversary to access the low-level circuit and manipulate its fingerprint would require extra defense strategies [1, 7] that are out of scope for this paper.

In the following analysis, we mainly relate to our implementation with min-max normalization preprocessing and SVM as the classifier. We define a normalization zone based on the minimum and maximum values of each feature across all enrolled devices's hybrid profiles. An enrolled device impersonating bit-level credentials would be revealed by RRF because of its accurate identification ability. If this attack came from an unenrolled device, the current implementation would classify the attacker as one of the enrolled devices unless the fingerprint falls outside the normalization zone. This is due to that SVM partitions the feature space into decision zones based on extreme points. Thereby the features with mild to moderate variations can be generously classified as the same class. RRF simulated-aided prior training optimizes SVM partitioning in a way that the size of each individual zone might expand/shrink, but the total normalization zone does not expand. For this reason, RRF does not make the SVM classifier more permissive in favour of the attacker.

To improve the protection against attacks initiated by an unenrolled device further, we propose a distance-based sanity check strategy to make the approach less permissive. We define a zone for each device by the ranges between the minimum and maximum values of each feature in its hybrid profiles. If the incoming fingerprint is not in the zone, it will be identified as an unenrolled device. Sampling of the SVM feature space reveals that these zones are small compared to the decision zones, below 2% for 40 enrolled devices. Adding a sanity check together with RRF thereby would decrease the attacker's success probability by 98% for every attempt.

In general, the RRF system design is independent of the classifier choice. The implementations with different classifiers have distinct security concerns. We leave the performance and security analysis of other classifiers for future work. RRF aims to enhance the robustness of radiometric fingerprint system, which ensures the system can identify legitimate users accurately over multiple measurements in different wireless environments.

8 RELATED WORK

Transient-based Radiometric Fingerprint. These fingerprints are built on unique features extracted from the signal's transient phase to identify individual devices [6, 34, 35]. Boris et al. show that transient phases are sensitive to antenna polarization and device locations [6], and require expensive receivers with high sampling rates (1-2 GS/s). Hence, they are unsuitable for civilian usage.

Coarse-extracted Radiometric Fingerprint. These systems use unprocessed I/Q time series as high-dimensional fingerprints and leverage deep learning algorithms to automate the feature extraction process [32]. Amani et al. confirm that this category of systems suffers significant interference from wireless channels [2]. To enhance the robustness, ORACLE modifies the modulation processing chain of transmitters by introducing additional artificial imperfections to enlarge the difference between devices. However, this method requires access to the modulation processing pipeline, which is not practical for cheap off-the-shelf devices. In comparison, RRF focuses on the receiver classifier and inherently does not require any transmitter modifications.

Fine-extracted Radiometric Fingerprint. Fine-extracted fingerprints are more structured, extracted via well-defined signal processing procedures. The feature extractor can be easily integrated with the receiver demodulation processing chain. These features include AGC-based features [22], channel state information (CSI)-based features [23], and modulation based features [3, 8, 20, 26, 27]. To enhance the fingerprint system's robustness towards channel fading interference, some features are not optional. For example, AGC-based features are highly dependent on the received signal strength that is susceptible to location changes. CSI-based features are limited to protocols supporting channel estimation, such as WiFi [23]. Most works usually assess only the classification accuracy and fail to capture the robustness aspect. Few works specifically augment the robustness of the system towards environment conditions. Xinyu et al. devise a hybrid classifier by adjusting feature weights based on the received signal SNR level [39, 41]. Their solution emphasizes the robustness towards SNR but does not does not consider other factors such as multipath and device mobility.

9 CONCLUSIONS

We propose RRF, a robust radiometric fingerprint system that addresses the fundamental challenge of wireless channel distortions. We demonstrate through experiments that already small changes in the environment can lead to a significant reduction in the probability that a system correctly authenticates an authorized device. We are the first to analyze the impact of the wireless channel on fingerprints through simulation, real-world experiments and analytical modeling. The systematic insights allow us to design a hybrid pipeline that embraces the distortions caused by the wireless channel and enhances the system's robustness. Our evaluation shows that RRF can achieve a high accuracy in a variety of challenging everyday environments, and demonstrates significantly enhanced robustness compared to previous work.

REFERENCES

[1] Luis Fernando Abanto-Leon et al. 2020. Stay connected, leave no trace: Enhancing security and privacy in WiFi via obfuscating radiometric fingerprints. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 4, 3 (2020).

[2] Amani Al-Shawabka et al. 2020. Exposing the fingerprint: Dissecting the impact of the wireless channel on radio fingerprinting. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. 646–655.

[3] Vladimir Brik et al. 2008. Wireless device identification with radiometric signatures. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*. 116–127.

[4] Lars Buitinck et al. 2013. API design for machine learning software: experiences from the scikit-learn project. , 108–122 pages.

[5] François Chollet et al. 2015. Keras. https://github.com/fchollet/keras.

[6] Boris Danev and Srdjan Capkun. 2009. Transient-based identification of wireless sensor nodes. In *International Conference on Information Processing in Sensor Networks*. 25–36.

[7] Jeroen Delvaux and Ingrid Verbauwhede. 2013. Side channel modeling attacks on 65nm arbiter PUFs exploiting CMOS device noise. In *2013 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST)*. IEEE, 137–142.

[8] Clay K Dubendorfer et al. 2012. An RF-DNA verification process for ZigBee networks. In *MILCOM, IEEE Military Communications Conference*. 1–6.

[9] Andrea Goldsmith. 2005. *Wireless Communications*. Cambridge University Press. 192 pages.

[10] Alejandro Gonzalez-Ruiz et al. 2011. A comprehensive overview and characterization of wireless channels for networked robotic and control systems. *Journal of Robotics* (2011).

[11] Mukul Goyal et al. 2010. Evaluating the impact of signal to noise ratio on IEEE 802.15. 4 PHY-level packet loss rate. In *2010 13th International Conference on Network-Based Information Systems*. 279–284.

[12] Andrey Gritsenko et al. 2019. Finding a 'New'needle in the haystack: unseen radio detection in large populations using deep learning. In *2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. 1–10.

[13] Shivani Gupta et al. 2019. Dealing with noise problem in machine learning data-sets: A systematic review. *Procedia Computer Science* 161 (2019).

[14] Guangquan Huang et al. 2016. Specific emitter identification based on nonlinear dynamical characteristics. *Canadian Journal of Electrical and Computer Engineering* 39, 1 (2016).

[15] IEEE. 2006. IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (WPANs). https://doi.org/10.1109/IEEESTD.2006.232110

[16] IEEE. 2020. IEEE Standard for Low-Rate Wireless Networks–Amendment 2: Low Power Wide Area Network (LPWAN) Extension to the Low-Energy Critical Infrastructure Monitoring (LECIM) Physical Layer (PHY).

[17] Cyril-Daniel Iskander et al. 2008. A MATLAB-based object-oriented approach to multipath fading channel simulation. *Hi-Tek Multisystems* 21 (2008).

[18] Gordana Ivosev et al. 2008. Dimensionality reduction and visualization in principal component analysis. *Analytical chemistry* 80, 13 (2008).

[19] C Jeruchim, Michel et al. 2000. *Simulation of Communication Systems: Modeling, Methodology and Techniques*. Springer.

[20] Yu Jiang et al. 2019. Physical layer identification of LoRa devices using constellation trace figure. *EURASIP Journal on Wireless Communications and Networking* 2019, 1 (2019).

[21] Brendan Juba et al. 2019. Precision-recall versus accuracy and the role of large data sets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 4039–4048.

[22] David A Knox et al. 2010. AGC-based RF fingerprints in wireless sensor networks for authentication. In *2010 IEEE International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 1–6.

[23] Pengfei Liu et al. 2019. Real-time identification of rogue WiFi connections using environment-independent physical features. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 190–198.

[24] Kevin McClaning et al. 2012. *Wireless Receiver Design for Digital Communications*. SciTech Publishing. 726 pages.

[25] Kevin Merchant et al. 2018. Deep learning for RF device fingerprinting in cognitive communication networks. *IEEE Journal of Selected Topics in Signal Processing* 12, 1 (2018).

[26] Nam Tuan Nguyen et al. 2011. Device fingerprinting to enhance wireless security using nonparametric Bayesian method. In *2011 Proceedings IEEE INFOCOM*. IEEE, 1404–1412.

[27] Linning Peng et al. 2019. Deep learning based RF fingerprint identification using differential constellation trace figure. *IEEE Transactions on Vehicular Technology* 69, 1 (2019).

[28] Adam C Polak et al. 2011. Identifying wireless users via transmitter imperfections. *IEEE Journal on selected areas in communications* 29, 7 (2011).

[29] Francesco Restuccia et al. 2019. Deepradioid: Real-time channel-resilient optimization of deep learning-based radio fingerprinting algorithms. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*. 51–60.

[30] Adel AM Saleh et al. 1987. A statistical model for indoor multipath propagation. *IEEE Journal on selected areas in communications* 5, 2 (1987).

[31] Iskander Sanchez-Rola et al. 2018. Clock around the clock: Time-based device fingerprinting. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 1502–1514.

[32] Kunal Sankhe et al. 2019. ORACLE: Optimized radio classification through convolutional neural networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 370–378.

[33] Kenneth I Talbot et al. 2003. Specific emitter identification and verification. *Technology Review* 113 (2003).

[34] J Toonstra et al. 1996. A radio transmitter fingerprinting system ODO-1. In *Proceedings of 1996 Canadian Conference on Electrical and Computer Engineering*, Vol. 1. IEEE, 60–63.

[35] Oktay Ureten et al. 2007. Wireless security through RF fingerprinting. *Canadian Journal of Electrical and Computer Engineering* 32, 1 (2007).

[36] Wenhao Wang, et al. 2016. Wireless physical-layer identification: Modeling and validation. *IEEE Transactions on Information Forensics and Security* 11, 9 (2016).

[37] Charles Wilson et al. 2004. Fingerprint vendor technology evaluation 2003: Summary of results and analysis report. *NIST Technical Report NISTIR* 7123 (2004).

[38] Qiang Xu et al. 2015. Device fingerprinting in wireless networks: Challenges and opportunities. *IEEE Communications Surveys & Tutorials* 18, 1 (2015).

[39] Jiabao Yu, et al. 2019. A robust RF fingerprinting approach using multisampling convolutional neural network. *IEEE Internet of Things Journal* 6, 4 (2019).

[40] Tianhang Zheng et al. 2019. FID: Function modeling-based data-independent and channel-robust physical-layer identification. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 199–207.

[41] Xinyu Zhou et al. 2021. A Robust Radio Frequency Fingerprint Extraction Scheme for Practical Device Recognition. *IEEE Internet of Things Journal* 8, 14 (2021).

[42] C Ziomek et al. 1995. Digital i/q demodulator. In *Proceedings Particle Accelerator Conference*, Vol. 4. 2663–2665.

A FINGERPRINT FEATURE MODELING

We consider the complex signal $r(t) = u(t)e^{j2\pi ft}$ where the baseband signal $u(t)$ encodes data in magnitude and phase, and f is the frequency at which the signal is transmitted. This signal can be represented in the constellation diagram with two orthogonal subcarriers, called in-phase (I) and quadrature (Q) components [42], in the form of vectors (phasors). A modulation of order M can thus be represented by a set of ideal symbols $I = \{A_i e^{j\Phi_i}\}, i = 1..M$, where $A_i e^{j\Phi_i}$ is the phasor of symbol i with magnitude A_i and phase Φ_i. For instance, the QPSK modulation used in IEEE 802.15.4 has modulation order $M = 4$ and encodes information only in the phase with the symbols $I = \{e^{j\pi/4}, e^{j3\pi/4}, e^{j5\pi/4}, e^{j7\pi/4}\}$.

The transmitters' hardware brings in different imperfections in both magnitude and phase, as shown for one symbol in Figure 16. The introduced error $ae^{j\phi}$ is called symbol error with magnitude a and phase ϕ. Thereby, the imperfect symbols actually transmitted are the set $S = \{A_i e^{j\Phi_i} + a_i e^{j\phi_i}\}, i = 1..M$. The radiometric fingerprint features mainly reflect the statistical characteristics of the symbol errors by comparing S and I.

We can classify fingerprint features into two categories based on two modes of the coherent demodulator. The first mode is the initial acquisition mode, where **synchronization-based** fingerprint features are extracted. When a frame arrives at the receiver, various signal parameters are measured to support AGC and synchronization processes. After that, the demodulator switches to the decision-direct mode. In this stage, the demodulator achieves synchronization, and the locked constellation is the source of **constellation-based** features. For each frame, the preamble is used for synchronization in the initial acquisition mode, and after synchronization, payload symbols are used to extract constellation-based features.

The constellation-based features refer to the modulation-base feature set introduced by Brik et al. [3].

Synchronization-based Features.
Carrier Frequency Offset (CFO). As for the imperfections of the local oscillator (LO), the actual signal frequency between transmitters might vary up to $\pm 10ppm$ (i.e., up to $\pm 24.5kHz$ for the $2.45GHz$ frequency band)[16].

Constellation-related Features.
SFD Correlation (SFDC). We define the feature as below:

$$SFDC = \frac{1}{len(SFD)} \sum_{k \in \text{SFD}} I_k \cdot S_k \tag{6}$$

Magnitude Error. This metric is defined as the absolute magnitude difference between ideal symbol and detected symbol phasor, which can be estimated as:

$$MagError = \frac{1}{M} \sum_{i=1}^{M} ||S_i| - |I_i|| = \frac{1}{M} \sum_{i=1}^{M} ||A_i e^{j\Phi_i} + a_i e^{j\phi_i}| - A_i| \tag{7}$$

Phase Error. This metric is the absolute phase deviation between ideal symbol and detected symbol phasor, which can be estimated as below:

$$PhaseError = \frac{1}{M} \sum_{i=1}^{M} |\angle S_i - \angle I_i| = \frac{1}{M} \sum_{i=1}^{M} |\angle (A_i e^{j\Phi_i} + a_i e^{j\phi_i}) - \Phi_i| \tag{8}$$

I/Q Offset. The I/Q offset is used to quantify the deviation between the center of the ideal signal I/Q axes and real signal, which is written as:

$$\begin{gathered} I/Q\ offset = \sum_{i=1}^{M} S_i = \sum_{i=1}^{M} (A_i e^{j\Phi_i} + a_i e^{j\phi_i}) \\ I\ offset = real(I/Q\ offset);\ Q\ offset = imag(I/Q\ offset) \end{gathered} \tag{9}$$

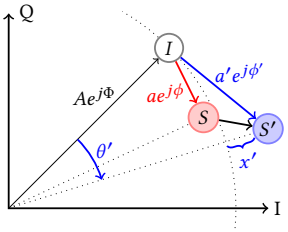


Figure 16: A close look into the modulation constellation: The ideal symbol I is transformed by an error vector $ae^{j\phi}$ into the observed symbols S (fingerprint) respective S' (hardware imperfection plus channel impacts). The resulting modulation error is estimated from the received samples as magnitude error x', phase error θ' and I/Q offsets.

The aforementioned feature definitions are based on the imperfect symbols S actually transmitted. The received signal is a linear combination of the transmitted signal and the channel response. We express the received symbol set as $S' = \{A_i e^{j\Phi_i} + a'_i e^{j\phi'_i}\}, i = 1..M$, where a'_i is the magnitude and ϕ'_i the phase of the resulting symbol error (c.f., Figure 16). Concrete expressions of S' for different channel models are established in the next section. To calculate the constellation-related features for the received signal, a'_i and ϕ'_i we use instead of a_i and ϕ_i, respectively.

B CHANNEL IMPACT MODELING

Multipath. In a multipath environment, the received signal is the sum of the signals propagating through different paths (rays). To simplify and support the analysis of the channel impact towards fingerprint features, we use a two-ray model $h(t) = \rho_0 \delta(t) + \rho_1 \delta(t-\tau)$ to approximate the propagation of electromagnetic waves. Specifically, the first ray arrives with amplitude ρ_0 and the second ray arrives τ seconds later with amplitude ρ_1. In this case, the baseband signal $u(t)$ composed of imperfect symbols S is received as:

$$r(t) = \rho_0 u(t) e^{j2\pi f t} + \rho_1 u(t-\tau) e^{j2\pi f (t-\tau)} \tag{10}$$

The receiver samples the signal as a discrete-time series which is then further processed. With the help of automatic gain control (AGC), the amplitudes of the two rays are normalized by $\rho = \frac{\rho_1}{\rho_0}$. We extract symbols $S'[k]$ at timestamp $kT_S + t^*$ where T_S is the symbol duration. Because of the delay of the second ray, $S'[k]$ might not only be dependent on $S[k]$ but also $S[k-1]$ if $t^* < \tau$:

$$S'[k] = S[k] \cdot e^{j2\pi f t^*} + S[n] \cdot \rho e^{j2\pi f (t^* - \tau)}, \tag{11}$$

where $n = k-1$ if $t^* < \tau$ and $n = k$ else. At frequency synchronization, the receiver aligns with the carrier frequency to compensate $e^{j2\pi f t^*}$. Then the received symbols S' can be statistically estimated by the symbols in the frame payload as below, where $\alpha = \frac{\tau}{T_s}$ is the intersymbol interference coefficient.

$$S' \approx \{S_i + (1-\alpha)\rho e^{-j2\pi\tau} S_i + \rho \frac{\alpha}{M} e^{-j2\pi\tau} \sum_{m=1}^{M} S_m\}, \tag{12}$$

where $S_i = A_i e^{j\Phi_i} + a_i e^{j\phi_i}$. Assuming the symbols having equal probability of being chosen, we have $\sum_{m=1}^{M} A_i e^{j\Phi_i} = 0$ and can rewrite the received symbols as:

$$S' \approx \{(1 + (1-\alpha)\rho e^{-j2\pi\tau})(A_i e^{j\Phi_i} + a_i e^{j\phi_i}) + \frac{\rho\alpha}{M} e^{-j2\pi\tau} \sum_{m=1}^{M} a_m e^{j\phi_m}\} \tag{13}$$

In other words, the observed symbol S'_i is a linear combination of the phasor expressing the imperfect symbol S_i and the symbol error vector from the other symbols due to inter symbol interference (ISI) and the residual channel response.

Based on the insights from the analytic model in Equation 13 we expect the symbol error to increase with a higher second path gain ρ and a longer delay τ. In particular, we have a dominating component scaling the phasor S_i (fingerprint) relative to the gain ρ and a phase shift depending on τ. The impact is thereby dependent on the individual intrinsic hardware imperfection, and the residual term including all symbols has limited impact. Translated into fingerprint features, we can expect the magnitude error to depend on both gain ρ and delay τ of the second multipath ray, while the phase error is expected to be more dependent on the delay τ alone. As the same scaling and phase shift apply to all symbols, we expect their impact to cancel out and not impact the I/Q offset.

Signal to Noise Ratio. The second impact factor is the signal to noise ratio (SNR), which is an essential indicator of signal quality. We use an AWGN to model the interference. Then the received symbol set S' can be written as:

$$S' = \{A_i e^{j\Phi_i} + a_i e^{j\phi_i} + n\}, \tag{14}$$

where $n \sim \mathcal{N}(0, \sigma^2)$ adds to both the I and Q component and thus impacts both the symbol amplitude and phase.

Doppler. The mobility of transmitters/receivers results in a Doppler spread in the observed frequency. The Doppler frequency shift is given by $\Delta f = \frac{v cos(\theta)}{c} f$, where c is the velocity of signal propagation and the object is moving at a spatial angle θ with velocity v [9]. In the analytical model, we consider the Doppler shift for a single ray.

$$r(t) = u(t) e^{j2\pi(f+\Delta f)t} \tag{15}$$

The received symbols S' is affected by the Doppler shift residual phase noise ϕ_d:

$$S' = \{e^{j\Delta\phi_d}(A_i e^{j\Phi_i} + a_i e^{j\phi_i})\} \tag{16}$$

C EXPERIMENT DETAILS

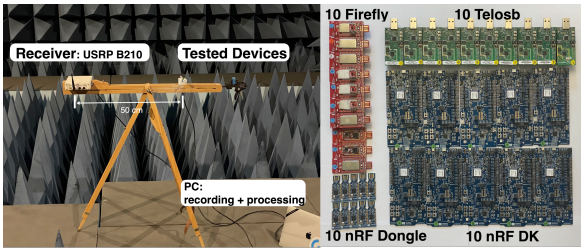


Figure 17: Benchmark experiment in anechoic chamber and 40 devices.