LoRa Device Fingerprinting in the Wild: Disclosing RF Data-Driven Fingerprint Sensitivity to Deployment Variability

ABDURRAHMAN ELMAGHBUB and BECHIR HAMDAOUI
Oregon State University, Corvallis, OR 97331 USA (e-mail: elmaghba, hamdaoui@eecs.oregonstate.edu)
Corresponding author: Bechir Hamdaoui (e-mail: hamdaoui@eecs.oregonstate.edu).

This research is supported in part by National Science Foundation Award No. 2003273.

ABSTRACT Deep learning-based fingerprinting techniques have recently emerged as potential enablers of various wireless applications. However, their resiliency to time, location, and/or configuration changes in the operating environment undoubtedly remains one major challenge that lies ahead in their deployment pathway. In this paper, we present an experimental framework that aims to disclose, understand and overcome the sensitivity of LoRa device fingerprinting to variations in deployment settings. We first began by presenting our RF fingerprinting datasets, collected from 25 different LoRa devices. The datasets cover a comprehensive set of experimental scenarios, considering both indoor and outdoor environments with varying network deployment settings, such as varying the distance between the transmitters and the receiver, the configuration of the LoRa protocol, the physical location of the conducted experiment, and the receiver hardware used for capturing the fingerprints. We then proposed a new technique that leverages out-of-band spectrum distortions, that are caused by device-specific hardware impairments, to provide unique device signatures that we exploit to improve fingerprinting accuracy. Finally, we conducted an experimental study that discloses the sensitivity of deep learning-based RF fingerprinting to changes in various deployment settings while considering three data representations of the learning model input: time-domain IQ, frequency-domain FFT, and Amplitude/Phase polar-coordinate. We found that the learning models perform relatively well when trained and tested under the same deployment settings, with FFT representation yielding the best performance followed by IQ representation. However, when trained and tested under different settings, the models (i) fail to maintain their high accuracy when the channel conditions change, and (ii) completely lose their ability to classify devices when the LoRa configuration and/or the USRP receiver hardware change. In addition, we interestingly observed that FFT representation performs exceptionally poorly when training and testing are done under different deployment settings, regardless of the type of the setting change.

INDEX TERMS Device fingerprinting, LoRa datasets, deep learning, experimental study, hardware impairments.

I. INTRODUCTION LoRa/LoRaWAN [1] has been adopted as the de facto standard for hundreds of millions of IoT devices, serving many IoT ecosystems and applications, thanks to its long-range, low-power capability. LoRa protocol offers two levels of security: (i) a network-level security that ensures device authentication and message integrity, and (ii) an application-level security that ensures end-to-end confidentiality [2]. The functionality and robustness of these layers rely on generating and storing session keys in a secured manner. Some IDs and keys (e.g., DevEUI and AppKey) used in this process can be found on the device tags or hardcoded in the installed source code. Hence, simple yet common human errors such as failing in removing the tags or replacing the source code before deployment can defeat security mechanisms and expose the network to great risks [3]. Due to the gained popularity of the LoRa protocol, providing robust security mechanisms for LoRa-enabled IoT devices and networks is therefore very crucial to the success and deployment of such IoT applications and services.

While conventional security solutions rely heavily on upper-layer and cryptographic mechanisms, it is essential that these mechanisms be complemented with unclonable,
physical-layer security solutions so that the security robustness of such systems is further increased [4]. RF fingerprinting, which extracts features from received RF signals to provide unique device signatures (aka fingerprints), has emerged as a potential solution for providing physical-level security [5], [6]. Until recently, however, the feature selection process for RF fingerprinting has been done manually, which requires expert knowledge and many trials and error iterations [7]. Following the unprecedented achievements of deep learning models in computer vision and natural language processing in recent years [8], [9], researchers have shown that the function approximation power of deep learning can yield performance improvements also for radio communications applications like device/RF fingerprinting [10].

Deep learning-based RF fingerprinting approaches rely on the layers of a deep neural network to process raw signals and extract the features automatically, so that they can directly infer device identity without resorting to feature engineering or requiring RF domain knowledge [11]–[14]. They exploit wave-level distortions in the received signals—for instance, those caused by hardware impairments—to extract fingerprints that are unique to the devices (e.g., [5], [6]). The uniqueness of these fingerprints comes from the randomness of the collective hardware-specific deviations of the RF analog components’ values from their values specified in the manufacturing process. However, as technology continues to improve, then so is the quality of devices. As a result, the impairment variability among devices—and thus the distance between devices in the feature space—is shrinking, making the device identification task more challenging, as fingerprinting becomes more vulnerable to changes in the wireless channel conditions and/or system deployment settings.

A. RELATED WORK

Among recent efforts aimed to enhancing deep learning-based techniques for classifying high-quality, bit-similar RF devices is Oracle [5], which imposes a set of distinctive artificial impairments on the transmitted signals to increase the variability among the devices and enhance the classification performance of high-quality RF devices. To tackle the sensitivity issue of the learned features to slight variations in the wireless channel, DeepRadioID [15] proposed to implement fine-tuned FIR filters in the transmitter side to alter the baseband signal based on the filter taps received as feedback from the classifier head to negate the impact of the wireless channel. Some other works focused on evaluating different aspects of the RF fingerprinting problem. For example, researchers in [16] presented a massive experimental study using the large-scale DARPA dataset of WiFi and ADS-B signals to investigate the scalability, the channel impact, and the size of the data, among others, on the performance of simple CNN and Resnet models. The large number of devices used to collect the DARPA dataset makes it an appropriate fit to test the scalability and the effect of different factors of the RF fingerprinting techniques. However, its downside is that the devices do not have identical chips, and the dataset is missing several essential scenarios. Researchers in [17] used another large-scale WiFi and ADS-B dataset to investigate the effects of different environmental factors as well as neural network architectures and hyperparameters on the system’s performance. They also showed the impact of some data augmentation techniques and compared the performance of complex- and real-valued CNNs under some channel effects such as spatial implication, temporal implication, and signal-to-noise ratio implication. The closest work to ours is the work of NEU [18] in which the authors released a large-scale WiFi dataset that was captured over 10 days and covers indoor and outdoor setups. Their evaluation showed that the state-of-the-art deep learning models fail significantly when the testing set is drawn from a dataset that is different from that used for training. Moreover, they suggested that these models are likely to learn channel-related features instead of hardware-specific features.

The sensitivity of device fingerprints to the receiver’s front-end presents another major issue that limits the portability of these fingerprints. Experimental results [19] showed that RF fingerprints created with a specific receiver could not be used universally across other receivers. However, researchers in [20] proposed to generate a calibration function, using a golden reference, and apply it to the received signals collected by different receivers to remove the bias introduced by the receiver hardware. Even though the empirical results showed the effectiveness of the proposed method, it still requires further investigation on a more extensive set of devices and a more comprehensive set of experimental scenarios. Also, in order to create universal RF fingerprints, this method requires collecting data from all involved receivers to generate the calibration function, which makes it unpractical.

In this paper, we present an experimental study that aims to disclose, understand and overcome the sensitivity of deep learning-based RF fingerprinting to changes in various deployment and environment settings. We do so by investigating a comprehensive set of experimental scenarios, considering both indoor and outdoor environments while changing essential deployment settings such as the distance between the transmitters and the receiver, the configuration of the LoRa protocol, the physical location of the conducted experiment, and the receiver hardware used for collecting the fingerprints.

B. NEED FOR MORE COMPREHENSIVE RF DATASETS

Recent research on RF fingerprinting has shifted from model-based approaches to deep learning-based (data-driven) approaches [16], [21]–[24]. Although recently proposed deep learning-based approaches show promising results, some rely on synthesized data for evaluation and validation [25], [26] and most of those that rely on real datasets are evaluated for general indoor and outdoor setups, and do not consider the impact of changes in the network deployment settings like the impact of the transmitter-receiver distance, the protocol configuration, the physical location of the experiment, and the receiver hardware [23], [24], [27].

We attribute this evaluation gap to the lack of adequate RF
datasets that cover the essential experimental scenarios that allow for performance robustness assessment under varying channel and network deployment conditions. Indeed, a major challenge the wireless research community has been facing is the lack of comprehensive, publicly available datasets that could serve as benchmarks for device/RF fingerprinting [18]. The availability of dataset benchmarks has undoubtedly played a key role for achieving technology innovation and maturity in fields like image recognition and natural language processing. Therefore, such efforts must be mimicked in the wireless community to achieve similar successes.

Researchers in [28] provided three large LPWAN (Sigfox and LoRAWAN) datasets in an outdoor environment to evaluate location fingerprinting algorithms. The datasets were collected over a three-month period, and consist of thousands of Sigfox-rural, Sigfox-urban, and LoRaWAN messages that include protocol, reception time, and GPS information. These datasets do not, however, include device labels, and hence cannot be used for supervised learning techniques. Another recent work—closely related and complementary to our work—published by a Northeastern University team [18] releases a dataset for IEEE 802.11 a/g (WiFi) standard data obtained from 20 wireless devices over several days (a) in an anechoic chamber, (b) in-the-wild, and (c) with cable connections. The focus of this WiFi dataset is on exploring the channel impact on the performance of the deep learning models used for classification. The dataset is limited in terms of the covered experimental scenarios, and is for WiFi signals only. Unfortunately, to the best of our knowledge, there are still no public datasets for LoRa device fingerprinting nor datasets that include the diverse experimental scenarios we covered in this work. Our datasets, presented and used in this paper, offer a comprehensive LoRa RF fingerprint data, covering a wide range of diverse scenarios useful and essential for enabling thorough technique evaluation and validation.

C. OUR LORA DATASET IN BRIEF

Our RF dataset provides both time-domain IQ samples and their corresponding FFT representations collected using an IoT testbed that consists of 25 identical Pycom devices and USRP B210 receivers, operating at a center frequency of 915MHz and used for collecting the received signals sampled at 1MS/s. Recorded data samples in the form of both the time-domain IQ and frequency-domain FFT are stored into binary files in compliance with SigMF [29] by creating, for each binary file, a metafile written in plaintext JSON to include recording information such as sampling rate, time, and day of recording, and carrier frequency, among other parameters. This dataset, consisting of more than 16K files, covers multiple different experimental scenarios (summarized in Table 1) and can be downloaded at http://research.engr.oregonstate.edu/hamdaouei/datasets.

These scenarios are briefly described next:

- **Setup 1—Different Days Indoor Scenario**, with data collected over 5 consecutive days. For each day, 10 transmissions, each of length 20s, were captured from each of the 25 transmitters. All transmissions took place 1 minute apart from one another. All experiments were run in a typical occupied indoor room environment.

- **Setup 2—Different Days Outdoor Scenario**, with data collected also over 5 consecutive days. For each day, 10 transmissions, each of length 20s, were captured from each of the 25 transmitters. All transmissions took place 1 minute apart from one another. All experiments were run in an outdoor environment.

- **Setup 3—Different Days Wired Scenario**, with data collected over 5 consecutive days. For each day, 10 transmissions were captured from each of the 25 transmitters, with each transmission having a duration of 20s. All transmissions are 1 minute apart from one another. All experiments were run with all devices connected directly to USRP receiver via SMA cables and 30dB attenuator.

- **Setup 4—Different Distances Scenario**, with data collected from 4 different distances: 5m, 10m, 15m, and 20m away from the receiver. For each distance, one transmission was captured from each of the 25 transmitters, with each transmission having a duration of 20s. All transmissions are 1 minute apart from one another.

- **Setup 5—Different Configurations Scenario**, with data collected from 4 different LoRa configurations in an indoor setup. For each configuration, one transmission was captured from each of the 25 transmitters, with each transmission having a duration of 20s. All transmissions are 1 minute apart from one another.

- **Setup 6—Different Locations Scenario**, with data collected in 3 different locations: room, outdoor, and office environments. At each location, one transmission was captured from each of the 25 transmitters, with each transmission having a duration of 20s. All transmissions are 1 minute apart from one another.

- **Setup 7—Different Receivers Scenario**, with data collected using 2 different receivers in indoor setup. For each receiver, one transmission was captured from each of the 25 transmitters, with each having a duration of 20s. All transmissions are 1 minute apart from one another.

D. SUMMARY OF PAPER CONTRIBUTIONS

The specific contributions are summarized as follows:

- **LoRa datasets.** Collected and provided comprehensive LoRa RF datasets using an IoT testbed consisting of 25 Pycom IoT devices. The dataset covers a number of scenarios, considering both indoor and outdoor environments with varying network deployments, such as varying the distance between the transmitters and the receiver, the configuration of the LoRa protocol, the physical location of the conducted experiment, and the receiver hardware used for capturing the fingerprints. All datasets are made freely available to the research community to allow for the development and validation of new and existing RF fingerprinting approaches.

- **Out-of-band fingerprinting technique.** Proposed a new technique that leverages out-of-band spectrum distor-
tions to increase the robustness of deep learning based device fingerprinting. The technique exploits out-of-band distortion information caused by hardware impairments to provide unique device-specific signatures that can be used to identify the devices.

- **Comprehensive experimental study.** Conducted an experimental study, using the collected datasets, with the goal of disclosing as well as understanding the sensitivity of deep learning-based LoRa RF fingerprinting to various deployment settings while also considering three different input data representations: time-domain IQ, frequency-domain FFT, and polar-coordinate A/θ (amplitude/phase). Our experimental results reveal that the learning models perform relatively well when trained and tested under the same deployment settings, with FFT yielding the best performance followed by IQ. But, when trained and tested under different settings, the models:
  1. fail to maintain their high accuracy when channel condition change occurs (either in time or in space),
  2. completely lose their ability to classify devices when the LoRa protocol configuration and/or the USRP receiver hardware change, and
  3. perform exceptionally poorly under FFT input representation, regardless of the type of change that occurs.

The rest of the paper is organized as follows. Sec. II, Sec. III and Sec. IV describe our testbed, the experimental setups and the RF datasets, respectively. Sec. V presents the proposed out-of-band driven fingerprinting approach. Sec. VI presents the comprehensive study and result findings aimed at revealing the sensitivity of RF fingerprinting to channel and deployment changes. Finally, Sec. VII concludes the paper.

## II. TESTBED

In this section, we describe the hardware, software, and protocol components used in building our testbed.

### A. HARDWARE DESCRIPTION

Our testbed, shown in Fig. 1, consists of 25 Pycom devices withSemtech SX1276 LoRa transceivers, 25 Pycom sensor shields, and an Ettus USRP (Universal Software Radio Peripheral) B210 for data sampling, which was configured with a center frequency at 915MHz and a sample rate of 1MS/s. Our collection of Pycom devices is made up of 23 Lopy4 boards and 2 Fipy boards, which are MicroPython-enabled boards with LoRa, Sigfox, WiFi, Bluetooth, and NB-IoT network capabilities. The sensor shield collection is composed of 22 PySense boards equipped with sensing capability, 2 PyScan boards equipped with RFID scanning capability, and 1 PyTrack board equipped with GPS/tracking capability. We used lipo batteries to power the devices. Each Pycom device was connected to a dedicated antenna and configured for Raw-LoRa mode transmissions at the 915MHz US band with: 125KHz bandwidth, a spreading factor (SF) of 7, a preamble of 8, a TX power of 20dBm, and a coding rate of 4/5.

### B. SOFTWARE DESCRIPTION

#### 1) Transmission Subsystem

We programmed and configured our Pycom boards using MicroPython [30], an efficient implementation of Python3 that is composed of a subset of standard Python libraries and optimized to run on microcontrollers and constrained environments. Also, we used Pymakr plugin as a REPL console that connects to Pycom boards to run codes or upload files.

#### 2) Reception Subsystem

We used the GNURadio software [31], a real-time signal processing graphical tool, to set up and configure the USRP receiver to capture LoRa transmissions, plot their time and spectrum domains, implement some preprocessing techniques and store the samples into their files. Fig. 2 shows the general flow graph used for our data acquisition.

### C. LORA PROTOCOL DESCRIPTION

We transmitted/captured LoRa modulation signals, a proprietary physical layer implementation that employs a derivative of the Chirp Spread Spectrum (CSS) technique in the sub-
GHz ISM band and trades data rate for coverage range, power consumption, and/or link robustness. LoRa does so by providing a tunable parameter, called a spreading factor (SF), that varies from 7 to 12 and determines the sequence length of an encoded symbol within a fixed bandwidth. A higher spreading factor means longer ranges with lower data rates. LoRa uses orthogonal SFs to allow packets of different SF values to be in the same channel concurrently, so as to improve network efficiency and throughput [32]. LoRa can also adjust, through its Adaptive Data Rate mechanism, its power level to accommodate the varying data rate and link conditions. LoRa modulator generates both raw (unmodulated) chirp signals that have fixed amplitude and continuously varying frequency with constant rate and a set of modulated chirps that are cyclically time-shifted raw chirps where the initial frequency determines the content of the chirp symbol. LoRa upchirps are used for transmitting data whereas the downchirps are used for synchronization purposes [33]. Unlike other spread spectrum techniques, the chirp-based modulation allows LoRa to maintain the same coding gain and immunity to noise and interference while meeting the low-cost, low-power consumption requirements. In addition, LoRa is agnostic to higher-layer implementations, which enables it to coexist with existing network architectures [32].

III. EXPERIMENTAL SETUPS

We used our testbed, described in Section II, to create and collect LoRa RF fingerprint datasets while considering multiple different experimental scenarios designed specifically to allow thorough evaluations of various deep learning-based wireless techniques, with particular focus on RF/device fingerprinting. For each setup, we considered a total bandwidth of 1MHz that covers in-band as well as an adjacent out-of-band spectrum of LoRa transmissions. The reason for choosing such a wide bandwidth to account for out-of-band spectrum information is explained in Section VI when we present the proposed hardware-impairment-driven fingerprinting technique. In this section, we present each of the seven considered experimental setups. For each setup, we used GNUradio packages to store the sampled raw-IQ values and their corresponding FFT representations into binary files as depicted in Fig. 2. We will be referring to Fig. 3 for describing setups 1, 2, and 3, and to Fig. 4 for describing setups 4, 5, 6 and 7. Table 1 summarizes all seven setups.

A. SETUP 1: DIFFERENT DAYS INDOOR SCENARIO

In order to enable performance evaluation while masking the impact of the outside environment, we created an indoor setup, ran experiments and collected data for this scenario. These indoor experiments were carried out in a typical occupied room environment over 5 consecutive days. All devices transmit the same message from the same location, 5m away from the receiver so that all devices experience similar channel conditions. As shown in Fig. 3, for each day, each transmitter generated 10 transmissions, each of 20s duration, all spaced apart by 1 minute. As a result, we collected about 200M complex-valued samples from each device per day.

B. SETUP 2: DIFFERENT DAYS OUTDOOR SCENARIO

In order to allow for performance evaluation while considering the impact of outdoor wireless channel impairments, in this setup, we carried out the experiments in an outdoor environment at nighttime. Here again, all devices transmitted the same message from the exact location, situated 5m away from the receiver, so that all devices experience similar channel conditions. Like in the indoor setup case and as shown in Fig. 3, for five consecutive days, each transmitter generated 10 transmissions per day, each of 20s duration, all spaced 1 minute apart from one another. This resulted in about 200M complex-valued samples per device per day. We ran this experiment over 5 consecutive days and provided 5-day worth...
of dataset to study the robustness of deep learning models when trained on data collected on one day but tested on data captured on a different day.

C. SETUP 3: DIFFERENT DAYS WIRED SCENARIO
The wireless channel has a notable impact on the performance of deep learning models and presents its unique challenges [18]. Hence, to assess how well these models perform in the absence of the wireless channel impact, we created a wired setup where the Pycom boards are directly connected to the USRP via an SMA cable and 30dB attenuator. Similar to the Indoor and Outdoor cases, we ran this experiment over 5 consecutive days. Each day each device transmits 10 bursts each of 20s duration as shown in Fig. 3. This resulted in a total amount of collected data of about 200M complex-valued samples per device per day.

D. SETUP 4: DIFFERENT DISTANCES SCENARIO
Because devices change positions, it is critical to study the impact of distance on the performance of classifiers and see whether a classifier could still recognize a device when it moves to a position that is different from the one used during training. This experiment was carried out in a typical outdoor environment in a sunny day, with four different considered distances, 5m, 10m, 15m, and 20m, between the transmitters and the receiver. For each distance, we collected 1 transmission of 20s for each of the 25 devices. We kept the receiver at the same location for all the transmissions. The transmissions were captured consecutively in time with only 60s apart from one another as shown in Fig. 4. This resulted in about 80M complex-valued samples per device.

E. SETUP 5: DIFFERENT CONFIGURATIONS SCENARIO
LoRaWAN allows to adjust the data rates, air-time, and energy consumption of end devices to accommodate the varying RF conditions, and does so through a set of end-device parameters like spreading factor, bandwidth, and consumed power. Changing, for example, the spreading factor in LoRa modulation results in a change in the data rate, receiver sensitivity, time in the air, and power consumption. Fig. 5 shows spectrum snapshots of the four LoRa configurations considered in our dataset. Ideally, classifiers should identify a device even if it changes its configuration; i.e., models trained using one configuration but tested on a different one should still perform. Therefore, in order to enable the assessment of how agnostic learning models are to protocol configuration, we captured transmissions using 4 different configurations, as presented in Table 2. For this, we collected a single LoRa transmission of 20s from each device for each configuration in an indoor setup with 5m as the distance between the receiver and transmitters as shown in Fig. 4.

F. SETUP 6: DIFFERENT LOCATIONS SCENARIO
Another practical scenario we consider here allows deep learning models to be trained on data captured in one location but tested on data collected in another location. For this, we captured LoRa transmissions in three different deployment locations/environments—room, office, and outdoor, all taken on the same day. Here, we kept the distance between the receiver and transmitters (i.e., 5m) and the LoRa configuration the same. We captured a single transmission of 20s from each device at each location with a 60s period between the devices, resulting in about 60M complex-valued samples from each device as depicted in Fig. 4.

G. SETUP 7: DIFFERENT RECEIVERS SCENARIO
Like transmitters, receivers also suffer from hardware impairments due to hardware imperfection. Therefore, learning models trained using data collected by one receiver but tested using data collected by a different receiver may not perform well due to the possible additional impairments caused by the receiver. To allow researchers to study the impact of such changes at the receiving side, we provided a dataset for the 25 Pycom transmitting devices, collected using two different USRP B210 receivers. In this experiment, we considered an
indoor setup where the transmitters (the 25 Pycom devices) were located 5m away from the receiver. For each receiver, we captured a single transmission of 20s from each of the 25 Pycom transmitters, as shown in Fig. 4, yielding a total of about 40M samples for each device.

**IV. DATASET: COLLECTION AND RELEASE**

Our dataset contains LoRa transmissions of 25 devices, a total of 43 transmissions on average for each device. An Ettus USRP B210 is used to record each transmission, operating at a center frequency of 915MHz with a sampling rate of 1MS/s. Each recording consists of 20M IQ samples. For each recording, we stored the time-domain IQ samples and their FFT representations into binary files. The binary files are encoded with Float32 and the complex-valued samples are interleaved where the I-values are in the odd indices and the Q-values are in the even ones. In order to be SigMF compliant, we created a metadata file written in plain-text JSON adapting the Signal Metadata Format (SigMF) [29] for each binary file to describe the essential information about the collected samples, the system that generated them, and the features of the signal itself. In our case, we stored in the metadata files information regarding (i) the sampling rate, (ii) time and day of recording, and (iii) the carrier frequency, among others.

Following Fig. 6 for guidance and help with their file system organization, the datasets (16000+ files with 1.2TB+ data) can be downloaded from NetStar laboratory website at http://research.engr.oregonstate.edu/hamdaoui/datasets.

Specifically, the dataset contains:

- **Different Days (Indoor/Outdoor/Wired) Scenario** directories, each having 5 subdirectories, one for each day. Each day subdirectory has 25 subdirectories, one for each device. Each device subdirectory has 20 SigMF recordings, with one dataset file and one metadata file for each of the 10 transmissions.

- **Different Locations Scenario** directory, having 4 subdirectories representing the four distances. Each subdirectory includes 50 SigMF recordings corresponding to one transmission, with one dataset file and one metadata file for each of the 25 devices.

- **Different Configurations Scenario** directory, having 4 subdirectories, representing the four configurations. Each subdirectory includes 50 SigMF recordings corresponding to one transmission, with one dataset file and one metadata file for each of the 25 devices.

- **Different Locations Scenario** directory, having 3 subdirectories, representing the three locations. Each subdirectory includes 50 SigMF recordings corresponding to one transmission, with one dataset file and one metadata file for each of the 25 devices.

- **Different Receivers Scenario** directory, having 2 subdirectories, representing the two receivers. Each subdirectory includes 50 SigMF recordings corresponding to one transmission, with one dataset file and one metadata file for each of the 25 devices.

To visualize the captured LoRa signal, we plotted, in Fig. 7, the time-domain IQ (top figure), frequency spectrum (middle figure), and the spectrogram showing upward chirps (bottom figure) of a device’s LoRa transmissions captured in the Indoor Setup.

**V. EXPLOITING OUT-OF-BAND SPECTRUM DISTORTIONS FOR ENHANCED RF FINGERPRINTING**

We now present our RF fingerprinting technique, which: (i) increases device distinguishability by efficiently identifying minimally-distorted devices, (ii) is robust against signature cloning and modification as it relies on inherently difficult-to-teper with, hardware features, (iii) does not require changes to be made at the transmitters, and (iv) incurs minimal extra processing at the receiver side that can be performed with
existing radio technology. The novelty of our technique lies in the fact that hardware impairments cause distortions in the spectrum surrounding the original band of the transmitted RF signal that are unique to the transmitting device. Our technique essentially exploits these distortions to extract signatures and features that uniquely identify devices.

### A. HARDWARE IMPAIRMENTS AND OOB EMISSIONS

Despite the endless engineering efforts aimed at designing hardware techniques that can eliminate hardware impairments and/or limit their impact on spectrum distortion [34]–[36], there will always be some inevitable amounts of impairments that cause (fortunately tolerable) out-of-band (OOB) spectrum emissions. Since our technique exploits the impact of such impairments to improve device fingerprinting, we begin in this section by taking a closer look at the sources, modeling, and impact of the most significant transmitter-specific impairments, with more emphasis on the OOB distortions that these impairments cause. Fig. 8, showing these impairments, will be used throughout this section for illustration.

1) DC Offset

Direct-conversion transmitters like the one shown in Fig. 8 leverage the quadrature mixer configuration to implement the upconversion of the baseband signal without the need for filtering. It does so by separately (in parallel) upconverting, at the carrier frequency $w_c$, the two in-phase (I) baseband modulated, $S_I(t) = A(t) \cos(\phi(t))$, and quadrature (Q) baseband modulated, $S_Q(t) = A(t) \sin(\phi(t))$, components with two independent mixers fed by a local oscillator (LO) tone shifted by 90° from one another. Here, $A(t)e^{j\phi(t)} = S_I(t) + jS_Q(t)$ represents the baseband modulated signal. Each mixer outputs the product of the baseband signal (I or Q) component and the carrier signal coming from the LO port. For ideal mixers, the output consists of two terms: one term appears at the summation while the other appears at the difference of the frequency of the two multiplied/mixed signals. However, due to hardware impairments, real mixers also produce some other unwanted emissions at different frequencies. Of particular importance is an undesired spike, known as carrier leakage spike, that appears at the center of the desired signal and cannot be easily filtered out. This results in distortion of the signal constellation, as well as in an increase in the error vector magnitude.

There are two main sources of DC offsets: carrier leakage and second-order nonlinearity. Carrier leakage stems from the LO leakage due to the poor isolation between the LO and RF output ports of the mixer. Thus, a strong LO signal can leak through unintended paths toward the mixer output port and appear at the middle of the desired signal’s spectrum [37]. For example, when mixing the in-phase baseband component $S_I(t)$, because of this LO leakage, the mixer output becomes

$$S_{I_{RF}}(t) = S_I(t) \cos(w_c t) + v_{lo} \cos(w_c t)$$

where $v_{lo} \cos(w_c t)$ is the unwanted carrier term due to the LO’s leakage through the mixer output port, and $v_{lo}$ is a hardware-specific quantity that varies from a mixer to another.

The second source of the DC offset is the second-order nonlinearity. When passing single-tone signals through a system with second-order nonlinearity, the output signal contains frequency components at an integer multiple of the input frequency. To illustrate, let’s feed the in-phase baseband component to the mixer while considering only the nonlinearity up to the second-order and ignoring the LO leakage effect. The output of the mixer in this case becomes

$$S_{I_{RF}}(t) = \alpha_1 S_I(t) \cos(w_c t) + \alpha_2 S_I^2(t) \cos^2(w_c t)$$

where $\alpha_1$ and $\alpha_2$ are the coefficients that model and capture the mixer’s first- and second-order nonlinearity. When rewriting $S_I(t)$ as $A(t) \cos(\phi(t))$, the second-order nonlinearity term—the one responsible for the DC component—can be written as

$$\alpha_2 S_I^2(t) \cos^2(w_c t) = \frac{\alpha_2 A^2(t)}{4} + \frac{\alpha_2 A^2(t)}{8} \left[ 2 \cos(2w_c t) + 2 \cos(2\phi(t) + \cos(2(\phi(t) - w_c t)) + \cos(2(\phi(t) + w_c t)) \right]$$

(1)

Note that the first term in Eq. (1) represents the DC component, and it is affected by the nonlinearity distortion captured by the parameter $\alpha_2$. Besides the relatively large carrier leakage component at the center of the signal’s spectrum, the nonlinearity of the mixer also introduces other undesired harmonic spurs in the out-of-band domain. The amplitude of
the carrier leakage spike and its harmonics depend on both the silicon-level circuitry of the mixer and the second-order nonlinearity distortion of the device. This can be clearly observed in Fig. 9, which compares the amplitudes of the carrier leakage spikes shown through the PSD of three simulated devices. Here, Device 1 mimics an ideal mixer (i.e., zero DC offset), while Device 2 and Device 3 mimic real mixers with in-phase DC offset values of 0.9 and 0.5 and quadrature offset values of 0.9 and 0.5, respectively. The figure clearly shows that while the output of the ideal mixer (Device 1) has neither a carrier leakage spike nor harmonic spurs, mixers with simulated DC offset impairments (Devices 2 and 3) cause DC offset spikes (carrier leakage and the harmonic spurs) to appear not only in the center of the message’s spectrum, but also in the adjacent out-of-band regions. Also, observe that the amplitudes of the spikes in Device 2 and Device 3 are quite different from one another even though the difference between their DC offset values is insignificant. Therefore, the carrier leakage and the harmonic spurs caused by the mixer’s impairments can potentially be leveraged for providing unique device signatures that can be used for device classification. Furthermore, providing the classifier with out-of-band information capturing the differences between the DC offset harmonic spurs can increase device separability and classification accuracy.

2) Phase Noise
In RF transmitter architectures, Local Oscillators (LOs) are responsible for generating periodic oscillating signals that can be used by the mixer to upconvert the baseband signal at the carrier frequency. In an ideal LO, this periodic signal can be represented as a pure sinusoidal waveform \( \cos(w_c t) \), which allows to upconvert baseband signals to the carrier frequency \( w_c \) while preserving their original spectrum shape.

This is illustrated in Fig. 10a, which upconverts a baseband tone to 100KHz using an ideal LO. In real LOs, the time domain instability of the generated signals causes random phase fluctuations that result in expansion or regrowth of the signal’s spectrum in both sides of the carrier frequency. The real LO oscillating signal can thus be represented as \( \cos(w_c t + \theta(t)) \), where \( \theta(t) \) is the phase deviation or noise term. The impact of this noise, commonly known as phase noise, is illustrated in Fig. 10b, which shows the upconversion of the same tone—whose upconversion using ideal LO is shown in Fig. 10a—using real LO signal.

The phase noise is manifested in a random rotation in the receiver signal constellation, thereby increasing the symbol detection error [38] as well as the out-of-band noise level. To illustrate this, consider mixing the in-phase baseband signal, \( S_I(t) \), with a real LO signal, \( \cos(w_c t + \theta(t)) \) (here \( \theta(t) \) represents the phase noise). The output of the mixer equals then \( S_I(t) \cos(w_c t + \theta(t)) \). Now applying the Fourier transform to this mixer output gives

\[
\frac{1}{2} \{ \tilde{S}_I(f-f_c) * F[e^{j\theta(t)}] + \tilde{S}_I(f+f_c) * F[e^{-j\theta(t)}] \} \quad (2)
\]

where \( f_c = \frac{w_c}{2\pi} \), \( \tilde{S}_I(f) = F[S_I(t)] \), and \( F[.] \) and * are the Fourier transform and convolution operators. From Eq. (2), we observe that the phase noise \( \theta(t) \) results in a bandwidth expansion beyond the original signal’s spectrum around the carrier frequency \( f_c \), which comes from the convolution of the spectrum of the bandpass (upconverted) signal, \( \tilde{S}_I(f+f_c) \), and that of the phase noise, \( F[e^{-j\theta(t)}] \).

Now since the spectrum expansion (or regrowth) depends on the LO phase noise, different devices exhibit different out-of-band distortions. This can be clearly seen in Fig. 11, where the PSD of three simulated devices, each with a different phase noise value but all with the same frequency offset, are displayed. Device 1 mimics an ideal LO (i.e., zero phase noise value), while Device 2 and Device 3 mimic real LOs with phase noise values of \(-80\) and \(-72\) dBc/Hz, respectively, at the same frequency offset, 1MHz. The figure clearly shows that the out-of-band spectrum shapes for Device 2 and Device 3 are different from one another and from Device 1. Therefore, like DC offsets, a transmitter’s phase noise caused by its LO impairments can potentially be leveraged for providing unique device signatures that can too be used for device classification. Additionally, considering the out-of-band information makes the spectra of devices more discernible and thus enhances the performance of the classifier.

3) Power Amplifier (PA) Nonlinearity Distortion
The majority of circuit nonlinearity is attributed to PAs as they provide the modulated RF signals with the required radiation power to reach their destination. When a PA operates in the linear region, its I/O characteristics is linear and an acceptable performance is ensured. However, operating in that region leads to more power consumption due to the associated lower power efficiency. Since PAs dominate power consumption, transmitters typically drive PAs to work near the
saturation region for higher power efficiency. Unfortunately, power efficiency and linearity conflict one another in that signals would suffer severely from the nonlinearity of the PA when operating in the saturation region. Such nonlinear distortions result in amplitude compression, as well as in high adjacent channel power leakage as a result of the bandwidth expansion, aka spectral regrowth. Although many methods have been proposed to minimize the distortion, PAs still exhibit some nonlinearity behaviors.

PA nonlinearity distortion is typically captured through the instantaneous amplitude and phase output responses to changes in the amplitude of the PA input signal, respectively known as Amplitude-to-Amplitude (AM-AM) and Amplitude-to-Phase (AM-PM) distortion curves. Using complex power series [39], the nonlinear PA output modelling the AM-AM and AM-PM distortions in response to the PA input signal $S_{RF}(t)$ can be expressed as [40]

$$S_{PA}(t) = \tilde{\alpha}_1 S_{RF}(t) + \tilde{\alpha}_3 S_{RF}^3(t) + \tilde{\alpha}_5 S_{RF}^5(t) + \ldots$$

where $\tilde{\alpha}_i$s are model complex coefficients. As we can infer from the equation above, only odd terms can be determined from single-tone complex compression characteristics, but fortunately, these terms are the most important ones because they produce inter-modulation distortions in-band and adjacent to the desired signal [41].

To illustrate the impact of PA nonlinearity on out-of-band spectrum distortions, suppose that the PA input signal is $S_{RF}(t) = A(t) \cos(w_c t + \phi(t))$ and consider looking at the effect of the third-order nonlinearity term only; i.e., the term

$$\tilde{\alpha}_3 S_{RF}^3(t) = \frac{\tilde{\alpha}_3 A^3(t)}{4} [3 \cos(w_c t + \phi(t)) + \cos(3w_c t + 3\phi(t))].$$

Now provided that the out-of-band component at $3w_c$ is located sufficiently far away from the center frequency, $w_c$, and that the bandwidth of the original signal is much less than $w_c$, this out-of-band component can easily be filtered out without causing any bandwidth regrowth around the original message spectrum. However, the first term at frequency $w_c$ can lead to spectrum regrowth. For instance, in the case of constant-envelope modulation schemes such as BPSK where the amplitude $A(t)$ is constant, the spectrum of the modulated signal in the vicinity of $w_c$ remains unchanged. We show this in Fig. 12 for the BFSK modulation, where the spectrum of the BFSK modulated signal has not changed after passing through a nonlinear PA. Note that the shape of the spectrum is the same under both linear and nonlinear PAs. However, for variable-envelope modulation schemes such as 16QAM where the amplitude $A(t)$ varies over time, nonlinearity causes a spectral regrowth of the original signal spectrum in that the $\frac{\tilde{\alpha}_3 A^3(t)}{4}$ term generally exhibits a broader spectrum than $A(t)$ itself. For this case of modulation, the severity of the spectral growth also depends on the nonlinearity model parameter $\tilde{\alpha}_3$. To illustrate, we show in Fig. 13 the case of a 16QAM modulated signal passing through a linear PA (Fig. 13a) and two nonlinear PAs (Figs. 13b and 13c) each under slightly different nonlinearity parameters. Two key observations we make here. First, observe that the nonlinearity of PA leads to an OOB spectrum growth (or distortion). Second, even a slight difference in the nonlinearity impairments causes a considerable differences in the amplitude of the frequency components in the OOB spectrum, as can be observed from the indicated amplitudes of the spikes. That is, different PA nonlinearity impairments cause different OOB spectrum distortions. Therefore, we argue that OOB spectrum distortion information due to PA nonlinearity can potentially be exploited to increase device distinguishability, thereby enhancing the accuracy and scalability of device classification.

4) IQ Mismatch

As shown in Fig. 8, the I and Q baseband components, $S_I(t) = A(t) \cos(\phi(t))$ and $S_Q(t) = A(t) \sin(\phi(t))$, are upconverted at the carrier frequency $w_c$ with two mixers, and the two outputs of the mixers are summed up, yielding, for real mixers, the bandpass modulated signal

$$S_{RF}(t) = A(t) \cos(\phi(t)) \cos(w_c t) - A(t) \sin(\phi(t)) \sin(w_c t)$$

However, DAC and mixer hardware impairments are manifested in amplitude mismatch, $\Delta\alpha$, and phase deviation, $\Delta\theta$, between the I and Q paths. This IQ mismatch, aka IQ imbalance, leads to imperfect image cancellation and results in
residual energy at the mirror frequency \(-w_c\), causing interference and SNR degradation. Considering amplitude and phase imbalances of \(\Delta\alpha\) and \(\Delta\theta\) when upconverting the baseband signal, the distorted bandpass signal can be expressed as:

\[
S_{RF}(t) = (1 - \Delta\alpha)S_I(t)\cos(w_c t) - S_Q(t)\sin(w_c t + \Delta\theta)
\]

Now assuming an ideal power amplifier and an ideal direct-conversion receiver, the distorted complex baseband signal \(\hat{R}(t) = S_{RF}(t)e^{-jw_c t}\) received at the receiver after down-conversion can be expressed as (after some math manipulations and clearing the terms appearing at twice the carrier frequency)

\[
\hat{R}(t) = \frac{1 - \Delta\alpha}{2}S_I(t) + j\left(\frac{\sin(\Delta\theta) - j\cos(\Delta\theta)}{2}\right)S_Q(t)
\]

Clearly, IQ imbalances cause in-band and out-of-band signal distortions that can be extracted and used for increasing device signature separability and device classification.

**B. THE PROPOSED TECHNIQUE: LEVERAGING OOB EMISSIONS FOR ROBUST DEVICE CLASSIFICATION**

We showed in previous sections that the out-of-band (OOB) spectrum surrounding the in-band region exhibits distinctive features due to the hardware impairments. We now propose a technique that leverages these distinctive features extracted from the adjacent OOB regions, along with those extracted from the in-band information, to uniquely identify RF devices. The inclusion of both in-band and OOB information can be achieved via oversampling or extending the filtered band to include the spectral regrowth and other frequency artifacts in the near OOB region. For our 125KHz LoRa transmissions, we used two different low-pass filters to capture (i) a 125KHz in-band spectrum and (ii) a 1MHz spectrum that spans over the in-band and adjacent OOB regions.

To assess the effectiveness of the proposed OOB technique, we implemented a CNN model (see Section VI-A for detail) and tested its classification accuracy, obtained under each of the two techniques, using our datasets capturing RF data from 25 different transmitting devices, for both the Indoor scenario (Setup 1 in Sec. III-A) and the Outdoor scenario (Setup 2 in Sec. III-B). Fig. 14 (indoor scenario) and Fig. 15 (outdoor scenario) show the confusion matrices of the classifier obtained using both "the conventional in-band only" and "the proposed in-band and OOB" techniques when the FFT representation of the data is used as an input to the learning model. In both scenarios, the figures show that the proposed OOB technique outperforms the conventional in-band only technique in terms of their achieved classification accuracy. For more specific insights, the averaged (over 5 transmissions)
indoor and outdoor results are shown in Fig. 16 for both FFT and IQ input representations. Observe that, when using FFT representation as an input, the proposed technique, which extracts features from the spectrum surrounding the original in-band spectrum, significantly enhances the testing accuracy, raising it from about 30% to about 80%. This significant gain in accuracy is achieved in both environment setups, Indoor and Outdoor, as shown in Fig. 16. When using IQ representation as input, both techniques, however, achieve similar results. We therefore conclude that by including both the in-band and OOB spectrum information, the proposed technique enhances device fingerprinting and classification accuracy significantly. More details on the experimental evaluation setup, including deep learning model’s architecture, input representation, parameters, etc., are provided in Sec. VI.

VI. EXPERIMENTAL RESULTS: DISCLOSING THE SENSITIVITY OF LORA DEVICE FINGERPRINTS

In this section, we present comprehensive experimental results that aim to disclose and understand the performance sensitivity of LoRa device fingerprinting and classification to network deployment and channel condition variability. Throughout this section, we only considered evaluating the proposed In-Band & OOB technique discussed in Sec. V-B.

A. CNN ARCHITECTURE & PARAMETERS

Deep Convolutional Neural Network (CNN) is proven to be an effective deep learning model for the classification and fingerprinting tasks of wireless RF signals [5], [6], [15], [42]. As a result, it has been one of the most commonly used architectures in deep learning-based RF fingerprinting, making it an appropriate candidate for studying the sensitivity of RF fingerprinting approaches to various network deployment and channel condition variations. In this work, we implemented a variation of the CNN model used in [42] whose architecture is depicted in Fig. 17. Each data input sequence was represented as a 2D (real/imaginary or amplitude/phase) real-valued tensor. The input was fed to the first convolutional layer (Conv1), which consists of 16 filters, each of size 1x4. Each filter learns 4-sample variations in time over the I and Q dimensions separately to generate 16 distinct feature maps over the complete input sample, except for the last convolution layer where the filter size is 2x4, which covers both the I and Q component simultaneously. The number of the convolutional layers and their filters’ sizes were tuned to provide the best performance. Each ConvLayer was followed by a Batch normalization layer, a Leaky Rectified Linear Unit (LeakyRelU) activation and a maximum pooling (MaxPool) layer with size of 1x2 and stride [1 2] to perform a predetermined non-linear transformation on each element of the convolution output, except for the last ConvLayer, which was followed by an Average Pooling (AP) layer with a dimension 1x256. The output of the AP layer was then provided as an input to the Fully Connected (FC) layer with 25 neurons, followed by another activation layer and dropout layer with drop rate of 0.5. Then, the output of the FC was finally passed to a classifier layer. To overcome overfitting, dropout rate was set to 0.5 at the dense layers. A Softmax classifier was used in the last layer to output the probabilities of each frame being fed to the CNN.

The neural network weights were updated using the Stochastic Gradient Descent technique with a momentum optimizer, as it achieved better performances compared to Adam and RMSProp techniques. Our experiments revealed that the model’s performances were quite sensitive to the learning rates, and hence to choose the appropriate values, we varied the initial learning rate over the (0.01-0.09) and (0.001-0.009) ranges. Consequently, we found that an initial learning rate of 0.07 which drops to 0.007 in the last epoch delivers the highest performances. Similarly, the network with an L2 regularization value of 0.0001 yielded a better performance compared to the values in the (0.001-0.009) and (0.0001-0.0009) ranges. We then minimized the prediction error through back-propagation, using categorical cross-entropy
as a loss function computed on the classifier output. A sliding window of 8192 samples was used to convert the CNN input samples into fixed-length non-overlapping frames (with stride size equaling window-size). Each frame was normalized to have a unit energy. A collection of 25 transmissions with 20M samples each was partitioned into training (about 80%), validation (about 10%), and testing (about 10%) sets. We implemented our CNN architecture in MATLAB using the Deep Learning Toolbox and ran it on the Oregon State University’s high-performance cluster that includes 6 Nvidia dgx2 nodes with 16 V100 GPUs.

### B. INPUT DATA PREPROCESSING & REPRESENTATION

In our experimental evaluation, three different data representations were used as input to the learning model: raw time-domain IQ data (referred to as IQ hereafter), FFT representation of the time-domain IQ data (referred to as FFT hereafter), and the polar coordinates of the IQ data (referred to as A/θ (amplitude/phase) hereafter).

1) IQ (Time-Domain) Representation

IQ representation is complex-valued vectors whose real and imaginary parts correspond to the I and Q components. We generated IQ frames by first creating fixed-size 1D complex-valued vectors, and then converting them into two 1D real-valued vectors, with the first dimension carrying the in-phase (I) information and the second carrying its quadrature (Q) one.

2) FFT (Frequency-Domain) Representation

Fast Fourier Transformer (FFT) converts time-domain samples to their corresponding spectral components in the frequency domain to provide a closer look into the frequency features of a signal capturing the spectrum artifacts that various hardware impairments might generate. We generated FFT representation frames by first converting the time-domain IQ complex-valued frames into the frequency domain using MATLAB `fft` function, then converting the resulted 1D complex-valued vectors into two 1D real-valued vectors.

3) Polar (Amplitude/Phase) Representation

Amplitude/Phase (or A/θ) representation is the polar coordinates of the complex-valued IQ samples, yielding real-valued points each representing the amplitude \( |A[n]| \) and phase \( \theta[n] \) of the corresponding IQ complex sample \((R[n], I[n])\). That is, \( A[n] = |R[R[n]]|^2 + |S[R[n]]|^2 \) \( \frac{1}{2} \) and \( \theta[n] = \arctan(S[R[n]]/R[R[n]]) \), where \( R() \) and \( S() \) are the real and imaginary operators. We created A/θ representation frames by converting complex-valued time-domain frames into two real-valued vectors representing the corresponding amplitude and phase values.

### C. RESULT ANALYSIS

We now present a comprehensive evaluation with the aim to disclose and understand the sensitivity of the achievable device classification accuracy to various channel condition and network deployment variability. We do so by studying the impact of changes in the channel conditions in both indoor (Sec. VI-C1) and outdoor (Sec. VI-C2) environments, in the LoRa protocol configuration (Sec. VI-C3), in the physical location of the experiment (Sec. VI-C4), and in the receiver hardware used for collecting the data samples (Sec. VI-C5), all investigated for each of the three data input representations explained in Sec. VI-B.

For each of these studied scenarios, our analysis of the results considers two cases: (i) when both the training and testing of the learning model are done using data collected under the same deployment and channel conditions and (ii) when the training and testing are done using data collected under different deployment and/or channel conditions.

1) Sensitivity to Channel Variability in Indoor Environment

We began by considering the impact of the time-variability of the channel conditions in an Indoor scenario, using the dataset presented in Sec. III-A, where the experiments took place in a typical occupied Indoor room. The results are presented in Fig. 18a for when training and testing are done using the same-day data (i.e., using data collected on the same day) and in Fig. 18b for when the training and testing are done using data collected on different days. In both figures, the presented testing accuracy is an average over 5 different transmissions. Two trends we observed when using same-day data for both training and testing (Fig. 18a). First, the achieved accuracy was consistent throughout the 4 different days. That is, the observed accuracy trends are similar regardless of the training/testing day. Second, the highest accuracy was obtained when the FFT representation was used as input to the learning model, with an achievable accuracy of more than 80% as opposed to only about 70% and 60% respectively when the IQ and A/θ input representations were used.

Now when the training and testing are done using different-day data (Fig. 18b), we first observed that regardless of the input representation, the accuracy drops substantially when the model is trained and tested on data obtained on different days. As confirmed in other previous studies, these results indicate that the model seems to be latching into channel-related features instead of hardware-specific features, making it not able to maintain its high accuracy when the channel conditions change. A second interesting trend we observed is that unlike the same-day data case, the model with the FFT input representation showed the worst performance when trained and tested using data collected on different days; the accuracy dropped from 84% to about 5% when training on data collected on Day 1, but testing on data collected on Day 1, 2, 3 or 4. The time-domain IQ input representation seems to be more robust to channel condition changes, as the accuracy drops from 70% to only about 45%.

Therefore, we conclude that the learning model is more resilient to channel condition changes when using time-domain IQ or A/θ representation as its input. This suggests that the impact of channel impairment variations is more profound in the frequency domain than in the time domain.
poor accuracy compared to using IQ or A/θ representations when training and testing are done using data collected on different days. As can be seen from Fig. 19b, for both IQ and A/θ representations, the accuracy dropped to about 20% whereas for the FFT representation it dropped to about 4%, an accuracy that is as good as random guesses. Looking at Indoor versus Outdoor, although similar trends were observed, the accuracy achieved in the Outdoor scenario was worse than that achieved in the Indoor scenario when using different-day data for training and testing. This is expected because channel impairments (noise, fading, interference, etc.) are more pronounced in an outdoor environment, causing greater distortions to the device features.

In conclusion, like indoor, in the outdoor scenario, the learning model fails to maintain its high accuracy when faced with varying channel conditions, suggesting again that the model seems to be learning channel-impaired features as opposed to learning device-specific features.

3) Sensitivity to Protocol Configuration Variability

After assessing the sensitivity of the classification model to changes in the wireless channel conditions in both Indoor and Outdoor environments, we took one step further in this investigation by assessing its sensitivity to changes in the operating configuration of the LoRa protocol. For this, we used the dataset presented in Sec. III-E, which considers the four different LoRa configurations shown in Table 2. Recall that, as discussed in Sec. III-E, changing the spreading factor of LoRa modulation changes the data rate, reception sensitivity, air-time, and power consumption, which in turn results in a change of the shape of the spectrum as clearly shown in Fig. 5. As done in the two previous scenarios, we also studied the impact of a configuration change on the accuracy by first training and testing using same-configuration data (Fig. 20a), and then training using one configuration and testing using another configuration (Fig. 20b). First, we observe from Fig. 20a that the model with FFT input representation consistently achieves higher performance compared to using IQ or A/θ representation as an input when same configuration is used during training and testing. Another observation we made is that in the case of IQ and A/θ representations, the accuracy degrades further when a higher spreading factor is used.

Now surprisingly, the model seems unable to recognize LoRa devices when tested on a spreading factor configuration that is different from that used for training. This is clearly depicted in Fig. 20b, which shows that the model performs very poorly and regardless of the input representation when it is tested with data captured with different LoRa configuration, even when the capturing is done during the same day and in the same location. Fig. 20b shows that the accuracy drops drastically to about 4% (what you actually get with random guesses) when the model is trained with configuration 1 but tested with configuration 2, 3 or 4, regardless of the data representation used for input. This drop is even worse to the drop observed when using different-day data as presented in Sec. VI-C1 and Sec. VI-C2.

2) Sensitivity to Channel Variability in Outdoor Environment

Here we consider the exact same experiment conducted in the Indoor scenario presented in Sec. VI-C1, except for an Outdoor environment, where we also consider the cases when the training and testing are done using same-day (Fig. 19a and different-day data (Fig. 19b). This Outdoor scenario was evaluated using the dataset presented in Sec. III-B. Overall, the trends observed in the Indoor scenario are also observed in the Outdoor scenario. For instance, the testing accuracy reported in Fig. 19a also indicate the advantage of using FFT input representation over using IQ or A/θ representations when training and testing on same-day data. Also, as observed in the Indoor scenario, using FFT representation as input yields very poor accuracy compared to using IQ or A/θ representations when training and testing are done using data collected on different days. As can be seen from Fig. 19b, for both IQ and A/θ representations, the accuracy dropped to about 20% whereas for the FFT representation it dropped to about 4%, an accuracy that is as good as random guesses. Looking at Indoor versus Outdoor, although similar trends were observed, the accuracy achieved in the Outdoor scenario was worse than that achieved in the Indoor scenario when using different-day data for training and testing. This is expected because channel impairments (noise, fading, interference, etc.) are more pronounced in an outdoor environment, causing greater distortions to the device features.

In conclusion, like indoor, in the outdoor scenario, the learning model fails to maintain its high accuracy when faced with varying channel conditions, suggesting again that the model seems to be learning channel-impaired features as opposed to learning device-specific features.
outdoor space, and office environment. These results show that both indoor locations (Loc1 and Loc3) demonstrated similar performance, and achieved higher accuracy than that achieved in the Outdoor location case. Also, we observed that the FFT representation continues to prevail over the other two representations. One thing to note is that the testing accuracy corresponding to Loc2 (outdoor) appears to be lower than that attained in the Outdoor Scenario previously discussed in Sec. VI-C2, especially for the FFT representation. This gap is (most likely) contributed to the difference in the surrounding environment of the two outdoor setups, as the outdoor environment in Loc2 enjoys a much broader open space and more crowded wireless activities.

Fig. 21b shows that the model with IQ or A/θ as input experienced a drop in accuracy of about 30% when trained in Loc1 (Room) but tested in Loc2 (Outdoor) or Loc3 (Office). For instance, with the IQ representation, the accuracy dropped from about 70% when model was trained and tested in Loc1 to about 50% when model was trained in Loc1 and tested in Loc2. The model with the FFT representation, on the other hand, continued to show a substantial drop in accuracy when tested in a location different from that used for training. In fact, like in the previous scenarios, the accuracy achieved under FFT representation was almost as good as making random guesses.

We concluded that the FFT representation again performs poorly when training and testing are done in different environments. Whether training vs. testing change occurs in time or location, such a change seems to impact the wireless channel dramatically, with distortions being more profound in the frequency domain than in the time domain.

5) Sensitivity to Receiver Hardware Variability
We now investigate the impact of the receiver’s impairments on the classification performance, and do so by using two different USRP B210 receivers to capture the transmissions of the 25 devices while maintaining the same indoor location, the same protocol configuration, and the same time of the conducted experiments. For this, we used the dataset and the experimental setup described in Sec. III-G.

As before, we first began by looking at the case when training and testing are done on data collected using the same receiver. The results of this case, given in Fig. 22a for two different USRP B210 receivers (Rec1 and Rec2), show that the IQ and FFT input representations still maintain performance superiority to the A/θ representation.

Now when changing the receivers, Fig. 22b shows that the accuracy dropped significantly when the model was tested using data captured on a receiver different from that used for training; this is true for both cases: train on Rec1/test on Rec2 and train on Rec2/test on Rec1. Also, as in all other scenarios, the FFT data representation seems to also suffer significantly when using different receivers for training and testing.

One detail to mention here is that the IQ representation when used under Receiver 1 achieved an accuracy (90%) that is higher than that achieved in the Indoor Scenario presented

VOLUME x, 2020

FIGURE 20. Different Configurations Scenario

FIGURE 21. Different Locations Scenario

Our conclusion here is that the RF fingerprinting model seems to completely lose its ability to classify devices when changes in the LoRa configuration occur, implying that LoRa spreading factor does change the data distribution drastically rendering data-driven classification approaches inadequate.

4) Sensitivity to Physical Location Variability
We now turn our attention to studying the resiliency of the learning model to changes in the experiment location. For this, we used the dataset and the experimental setup described in Sec. III-F. Fig. 21a shows the testing accuracy of the learning model when trained and tested with same-location data captured in three different locations: room environment, open

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3121606, IEEE Access
in Sec. VI-C1 (70%). This is because even though both of these scenarios are carried out in an indoor environment, the data acquisition approach used in each scenario is different. For the Indoor Scenario presented in Sec. VI-C1, 10 transmissions were collected for each device before moving to the next device, and since we are doing this for 25 devices, the time that elapsed between the first transmissions of the first device and the last transmissions of the last device could be significantly high for the channel to be considered the same. On the other hand, for this Different Receivers scenario, only one transmission per each device was collected, yielding lesser time difference between the first and last devices.

In conclusion, these results show that the learning models depend also on the receiver hardware.

VII. CONCLUSION

We proposed a novel RF data-driven device fingerprinting technique that leverages both In-Band and Out-of-Band RF signal distortion information, caused by hardware impairments during device manufacturing, to increase the classification accuracy. We also collected and released comprehensive LoRa RF datasets, covering a set of different indoor and outdoor experimental scenarios under varying network deployment settings. Our motivation behind this work is two-fold. First, it provides the community with a large-scale and diverse datasets, covering a set of different indoor and outdoor settings. Our motivation behind this work is two-fold.

In conclusion, these results show that the learning models depend also on the receiver hardware.

VIII. ACKNOWLEDGEMENT

This research is supported in part by National Science Foundation Award No. 2003273. We would like to thank Intel researchers, Dr. Kathiravetpillai Sivasan, Dr. Lily Yang, and Dr. Richard Dorrance, for their constructive feedback.

REFERENCES

[1] J. P. S. Sundaram, W. Du, and Z. Zhao, “A survey on lora networking: Research problems, current solutions, and open issues,” IEEE Communications Surveys & Tutorials, vol. 22, no. 1, pp. 371–388, 2019.
[2] D. Basu, T. Gu, and P. Mohapatra, “Security issues of low power wide area networks in the context of lora networks,” arXiv preprint arXiv:2006.16554, 2020.
[3] M. S. Cesar Cerrudo, Esteban Martinez Fayo, “LoRaWAN Networks Susceptible to Hacking: Common Cyber Security Problems, How to Detect and Prevent Them,” https://ioac.tv/30fZuZe, 2020, [Online; accessed 28-April-2021].
[4] Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S. C.-H. Huang, and H.-H. Chen, “Physical layer security in wireless networks: a tutorial,” IEEE Wireless Communications, vol. 18, no. 2, pp. 66–74, 2011.
[5] K. Sankhe, M. Beligiovia, F. Zhou, S. Riyaz, S. Ioannidis, and K. Chowdhury, “Oracle: Optimized radio classification through convolutional neural networks,” in IEEE INFOCOM 2019-IEEE Conference on Computer Communications, IEEE, 2019, pp. 370–378.
[6] A. Elmaghoub, B. Hamdaoui, and A. Natarajan, “Widescan: Exploiting out-of-band distortion for device classification using deep learning,” in GLOBECOM 2020-2020 IEEE Global Communications Conference, IEEE, 2020, pp. 1–6.
[7] Y. Li, X. Chen, Y. Lin, G. Srivastava, and S. Liu, “Wireless transmitter identification based on device imperfections,” IEEE Access, vol. 8, pp. 59 305–59 314, 2020.
[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, vol. 25, pp. 1097–1105, 2012.
[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in Advances in neural information processing systems, 2017, pp. 5998–6008.
[10] T. J. O’Shea, J. Corgan, and T. C. Clancy, “Convolutional radio modulation recognition networks,” in International conference on engineering applications of neural networks. Springer, 2016, pp. 213–226.
[11] D. Roy, T. Mukherjee, M. Chatterjee, E. Blasch, and E. Pasiliao, “Rfai: Adversarial learning for rf transmitter identification and classification,” IEEE Transactions on Cognitive Communications and Networking, vol. 6, no. 2, pp. 783–801, 2019.
[12] M. Cekić, S. Gopalakrishnan, and U. Madhow, “Robust wireless fingerprinting: Generalizing across space and time,” arXiv e-prints, pp. arXiv–2002, 2020.
[13] L. Peng, J. Zhang, M. Liu, and A. Hu, “Deep learning based rf fingerprint identification using differential constellation trace figure,” IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 1091–1095, 2019.
[14] Q. Xu, R. Zheng, W. Saad, and Z. Han, “Device fingerprinting in wireless networks: Challenges and opportunities,” IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 94–104, 2015.
[15] F. Restuccia, S. D’Oro, A. Al-Shawabka, M. Beligiovia, L. Angioloni, S. Ioannidis, K. Chowdhury, and T. Melodia, “Deepradiofi: Real-time channel-resilient optimization of deep learning-based radio fingerprinting algorithms,” in Proc. of the 20th ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2019, pp. 51–60.
[16] T. Jian, B. C. Rendon, E. Ojbaa, N. Soltani, Z. Wang, K. Sankhe, A. Gritsenko, J. Dy, K. Chowdhury, and S. Ioannidis, “Deep learning for rf fingerprinting: A massive experimental study,” IEEE Internet of Things Magazine, vol. 3, no. 1, pp. 50–57, 2020.
[17] I. Agadakos, N. Agadakos, J. Polakis, and M. R. Amer, “Deep complex networks for protocol-agnostic radio frequency device fingerprinting in the wild,” arXiv preprint arXiv:1909.08703, 2019.
[18] A. Al-Shawabka, F. Restuccia, S. D’Oro, T. Jian, B. C. Rendon, N. Soltani, J. Dy, S. Ioannidis, K. Chowdhury, and T. Melodia, “Exposing the fingerprint: Dissecting the impact of the wireless channel on radio fingerprinting,”
in IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, 2020, pp. 646–655.

[19] S. U. Rehman, K. W. Sowerby, S. Alam, and I. Ardekani, “Portability of an rf fingerprint of a wireless transmitter,” in 2014 IEEE Conference on Communications and Network Security. IEEE, 2014, pp. 151–156.

[20] G. Baldini, R. Giuliani, C. Gentile, and G. Steri, “Measures to address the lack of portability of the rf fingerprints for radiometric identification,” in 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS). IEEE, 2018, pp. 1–5.

[21] Z. Chen, L. Peng, A. Hu, and H. Fu, “Generative adversarial network-based rogue device identification using differential constellation trace figure,” EURASIP Journal on Wireless Communications and Networking, vol. 2021, no. 1, pp. 1–27, 2021.

[22] B. Li and E. Cetin, “Waveform domain deep learning approach for rf fingerprinting,” in 2021 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2021, pp. 1–5.

[23] J. Li, Y. Ying, C. Ji, and B. Zhang, “Differential contour stellar-based radio frequency fingerprint identification for internet of things,” IEEE Access, vol. 9, pp. 53 745–53 753, 2021.

[24] B. Li and E. Cetin, “Design and evaluation of a graphical deep learning approach for rf fingerprinting,” IEEE Sensors Journal, 2021.

[25] Y. Pan, S. Yang, H. Peng, T. Li, and W. Wang, “Specific emitter identification based on deep residual networks,” IEEE Access, vol. 7, pp. 54 425–54 434, 2019.

[26] J. Persons, L. J. Wong, W. C. Headley, and M. C. Fowler, “Classification of radio signals using truncated gaussian discriminant analysis of convolutional neural network-derived features,” arXiv preprint arXiv:2008.04 874, 2020.

[27] G. Shen, J. Zhang, A. Marshall, L. Peng, and X. Wang, “Radio frequency fingerprint identification for lora using spectrogram and cnn,” in IEEE INFOCOM 2021-IEEE Conference on Computer Communications. IEEE, 2021, pp. 1–10.

[28] M. Aernouts, R. Berkvens, K. Van Vlaenderen, and M. Weyn, “Sigfox and lorawan datasets for fingerprint localization in large urban and rural areas,” Data, vol. 3, no. 2, p. 13, 2018.

[29] B. Hilburn, N. West, T. O’Shea, and T. Roy, “Signf: the signal metadata format,” in Proceedings of the GNU Radio Conference, vol. 3, no. 1, 2018.

[30] N. H. Tollervey, B. Hilburn, N. West, T. O’Shea, and T. Roy, “Sigmf: the signal metadata format,” in Proceedings of the GNU Radio Conference, vol. 3, no. 1, 2018.

[31] B. Hilburn, N. West, T. O’Shea, and T. Roy, “Sigmf: the signal metadata format,” in Proceedings of the GNU Radio Conference, vol. 3, no. 1, 2018.

[32] A. Semtech, “Lora modulation basics,” Semtech Corporation, Tech. Rep.

[33] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, “A generalized memory polynomial model for digital predistortion of rf power amplifiers,” IEEE Tran. on signal processing, vol. 54, no. 10, pp. 3852–3860, 2006.

[34] L. Ding, “Digital predistortion of power amplifiers for wireless applications,” Ph.D. dissertation, Georgia Institute of Technology, 2004.

[35] M. Li, B. Selim, S. Muhaidat, P. C. Sofotasios, M. Dianati, P. D. Yoo, J. Liang, and A. Wang, “Effects of residual hardware impairments on secure noma-based cooperative systems,” IEEE Access, vol. 8, pp. 2524–2536, 2019.

[36] R. Svitak and S. Raman, “De offsets in direct-conversion receivers: Characterization and implications,” IEEE Microwave Magazine, vol. 6, no. 3, pp. 76–86, 2005.

[37] M. R. Khanzadi, D. Kaylenstierna, A. Panahi, T. Eriksson, and H. Zirath, “Calculation of the performance of communication systems from measured oscillator phase noise,” IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 61, no. 5, pp. 1553–1565, 2014.

[38] N. Blachman, “Band-pass nonlinearities,” IEEE Transactions on Information Theory, vol. 10, no. 2, pp. 162–164, 1964.

[39] K. G. Gard, L. E. Larson, and M. B. Steer, “The impact of rf front-end characteristics on the spectral regrowth of communications signals,” IEEE Transactions on Microwave Theory and Techniques, vol. 53, no. 6, pp. 2179–2186, 2005.

[40] K. G. Gard, H. M. Gutierrez, and M. B. Steer, “Characterization of spectral regrowth in microwave amplifiers based on the nonlinear transformation of a complex gaussian process,” IEEE Transactions on Microwave Theory and Techniques, vol. 47, no. 7, pp. 1059–1069, 1999.

[41] X. Liu, D. Yang, and A. El Gamal, “Deep neural network architectures for modulation classification,” in 2017 51st Asilomar Conference on Signals, Systems, and Computers. IEEE, 2017, pp. 915–919.

ABDURRAHMAN ELMAGHBUB received the B.S. degree with summa cum laude distinction, and MS in Electrical and Computer Engineering from Oregon State University in 2019, 2021, respectively, and is currently pursuing his Ph.D. degree in the School of Electrical Engineering and Computer Science at Oregon State University. His research interests are in the area of wireless communication and networking with a current focus on applying deep learning to wireless device classification.

BECHIR HAMDAOUI (S’02-M’05-SM’12) is a Professor in the School of Electrical Engineering and Computer Science at Oregon State University. He received M.S. degrees in both ECE (2002) and CS (2004), and the Ph.D. degree in ECE (2005) all from the University of Wisconsin-Madison. His research interests are in the general areas of computer networking, network security, and wireless communication, with a current focus on edge cloud computing, network analytics, autonomous systems, dynamic spectrum management, 5G systems, and datacenters. He won several awards, including the ISSIP 2020 Distinguished Recognition Award, the ICC 2017 Best Paper Award, the IWCMC 2017 Best Paper Award, the 2016 EECS Outstanding Research Award, and the 2009 NSF CAREER Award. He serves/served as an Associate Editor for several journals, including IEEE Transactions on Mobile Computing, IEEE Transactions on Wireless Communications, IEEE Network, and IEEE Transactions on Vehicular Technology. He also chaired/co-chaired many IEEE conference programs/symposia, including the 2017 INFOCOM Demo/Posters program, the 2016 IEEE GLOBECOM Mobile and Wireless Networks symposium, and many others. He currently serves as the Chair for the IEEE Communications Society’s Wireless Communication Technical Committee (WTC). He served as a Distinguished Lecturer for the IEEE Communication Society for 2016 and 2017. He is a Senior Member of IEEE, IEEE Computer Society, IEEE Communications Society, and IEEE Vehicular Technology Society.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3121606, IEEE Access