Release Note: Comprehensive LoRa RF Datasets for Device Fingerprinting Using Deep Learning

Abdurrahman Elmaghbub and Bechir Hamdaoui
School of EECS, Oregon State University

July 2021

1 Quick Dataset Download Links

LoRa RF datasets (download links given below) are described and used in the freely accessible IEEE Access paper titled LoRa Device Fingerprinting in the Wild: Disclosing RF Data-Driven Fingerprint Sensitivity to Deployment Variability. IEEE Access, pp: 142893–142909, October 2021.

The datasets can be downloaded and used for research, but we would like to request that any use that results in technical or other publications should include a citation to the IEEE Access paper [1, 2] above whose Bibtex is:

Copy and paste the bibtex below:

```
@article{elmaghbub2021lora,
 title={{LoRa} Device Fingerprinting in the Wild: Disclosing {RF} Data-Driven Fingerprint
  Sensitivity to Deployment Variability},
  author={Elmaghbub, Abdurrahman and Hamdaoui, Bechir},
  journal={IEEE Access},
  volume={9},
  pages={142893--142909},
  year={2021},
  publisher={IEEE}}
```

Click on the link corresponding to the setup you would like to download the dataset for:

- Setup 1: Different Days Indoor Scenario.
- Setup 2: Different Days Outdoor Scenario
- Setup 3: Different Days Wired Scenario
- Setup 4: Different Distances Scenario.
- Setup 5: Different Configurations Scenario
- Setup 6: Different Locations Scenario.
- Setup 7: Different Receivers Scenario.

2 General Description

This comprehensive LoRa fingerprint dataset has been collected at Oregon State University, as part of an NSF project. The dataset provides both time-domain I/Q samples and their corresponding FFT samples of LoRa transmissions collected using an IoT device testbed. The testbed consists of 25 identical Pycom IoT devices and USRP B210 receivers, operating at a center frequency of 915MHz, used for recording the received signals sampled at 1MS/s. The listed scenarios, summarized in Table 1, are specifically designed to evaluate the performance of Deep

Learning-Based RF fingerprinting algorithms, but can also be used for other research purposes.

We used the GNURadio software to set up and configure the USRP receivers 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.

We created raw I/Q data and FFT-based representation files for each transmission in ".dat" format. The binary files are encoded with Float32 and the complex-valued samples are interleaved, with the I values in the odd indices and the Q values in the even indices. For each binary file, we created a metadata file written in plain-text JSON adapting the Signal Metadata Format (SigMF) 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 file information regarding (*i*) the sampling rate, (*ii*) time and day of recording, and (*iii*) the carrier frequency, among others. Table 1 shows a summary of the dataset.

The datasets, provided in the links above, are composed of $16,300$ files with more than 1.2TB of data, which are described next (refer to Fig. 1 for help with the system organization and notation of the files):

- **Setup 1: Diff_Days_Indoor_Setup** directory has 5 subdirectories, one for each day. Each day subdirectory has 25 subdirectories, one for each device. Each device subdirectory has 20 SigMF recordings corresponding to the 10 transmissions (with each transmission having a dataset binary file and a metadata file).
- **Setup 2: Diff_Days_Outdoor_Setup** directory has 5 subdirectories, one for each day. Each day subdirectory has 25 subdirectories, one for each device. Each device subdirectory has 20 SigMF recordings corresponding to the 10 transmissions (with each transmission having a dataset binary file and a metadata file).
- **Setup 3: Diff_Days_Wired_Setup** directory has 5 subdirectories, one for each day. Each day subdirectory has 25 subdirectories, one for each device. Each device subdirectory has 20 SigMF recordings corresponding to the 10 transmissions (with each transmission having a dataset binary file and a metadata file).
- **Setup 4: Diff_Distances_Setup** directory has 4 subdirectories, representing the four distances. Each subdirectory includes 50 SigMF recordings, corresponding to one transmission (with each transmission having a dataset binary file and a metadata file) for each of the 25 devices.
- **Setup 5: Diff_Configurations_Setup** directory has 4 subdirectories, representing the four configurations. Each subdirectory includes 50 SigMF recordings, corresponding to one transmission (with each transmission having a dataset binary file and a metadata file) for each of the 25 devices.
- **Setup 6: Diff_Locations_Setup** directory has 3 subdirectories, representing the three locations. Each subdirectory includes 50 SigMF recordings, corresponding to one transmission (with each transmission having a dataset binary file and a metadata file) for each of the 25 devices.
- **Setup 7: Diff_Receivers_Setup** directory has 2 subdirectories, representing the two receivers. Each subdirectory includes 50 SigMF recordings, corresponding to one transmission (with each transmission having a dataset binary file and a metadata file) for each of the 25 devices.

An example MATLAB code to retrieve the binary files and convert them into complex-valued data:

```
fid = fopen('~\IQ_1.dat');
date_size = 40000000;
real =  1:2:date_size-1;
imag =  2:2:date_size;
[val, count] = fread(fid, date_size, 'float');
data = complex(val(real), val(imag));
```

Setups	Number of Devices	Number of Receivers	Protocol	Number Days	Transmissions per Device	Duration per Transmission	Distances	Environment	Representation
1) Diff Days Indoor	25	1	LoRa	5	10	20s	5m	Indoor	IQ/FFT
2) Diff Days Outdoor	25	1	LoRa	5	10	20s	5m	Outdoor	IQ/FFT
3) Diff Days Wired	25	1	LoRa	5	10	20s	5m	Wired	IQ/FFT
4) Diff Distances	25	1	LoRa	1	4	20s	5,10,15,20m	Outdoor	IQ/FFT
5) Diff Configurations	25	1	LoRa	1	4	20s	5m	Indoor	IQ/FFT
6) Diff Locations	25	1	LoRa	1	3	20s	5m	2 Indoor, 1 Outdoor	IQ/FFT
7) Diff Receivers	25	2	LoRa	1	2	20s	5m	Indoor	IQ/FFT

Table 1: Summary of Experimental Setups/Scenarios.

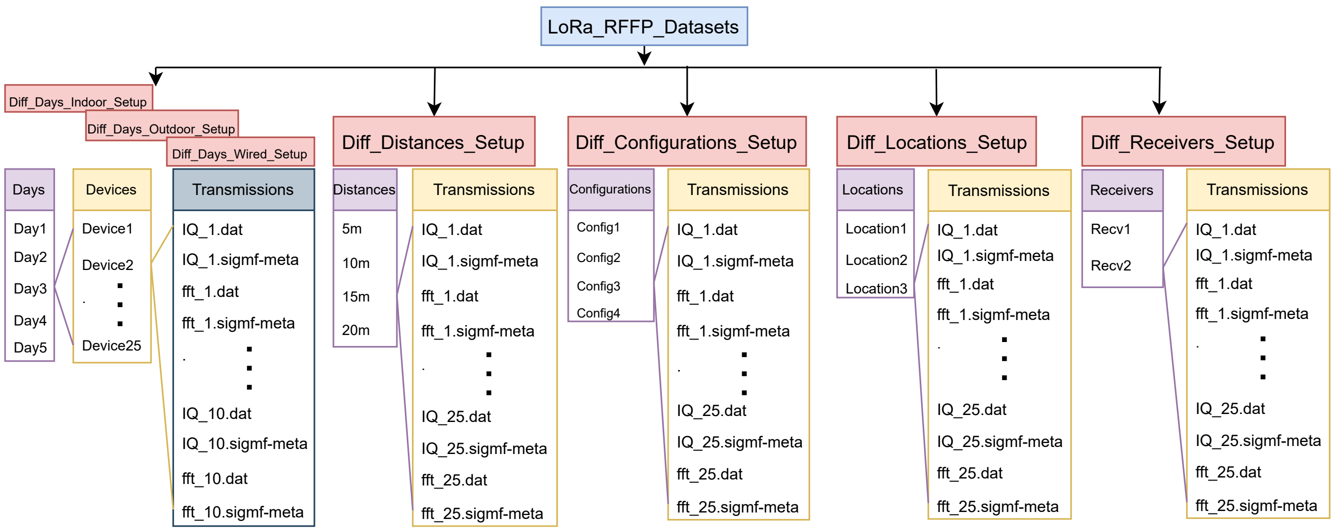

Figure 1: File/link structure and organization of the dataset. Notice that Diff_Days_(Indoor, Outdoor, and Wired)_Setup directories have the same file system architecture: 5 days subdirectories → 25 devices subdirectories → 20 data files and 20 metadata files.

3 Experimental Setups

3.1 Setup 1: Different Days Indoor Scenario

3.1.1 Transmitter Side

- 23 Lopy4 boards and 2 Fipy boards on top of 22 Pysense sensor shields, 2 Pytrack sensor shields, and 1 Pyscan sensor shield.
- Each device uses its own LoRa Whip, Tilt RF Antenna, and they all are powered via Lithium Ion Battery −3.7v 2000mAh.
- Protocol of operation: LoRa Modulation.
- LoRa configuration: Raw-LoRa mode, a channel bandwidth of 125KHz, a spreading factor of 7, a preamble of 8, a TX power of 20dBm, and a coding rate of 4/5.

3.1.2 Receiver Side

- One USRP B210 receiver configured with a center frequency at 915MHz, a sample rate of 1MHz, and an observed bandwidth of 1MHz.
- One Vert900 Ettus Antenna.
- The USRP is connected to a host computer with GNURadio software with some signal processing packages.

3.1.3 Download Dataset:

Different Days Indoor Scenario.

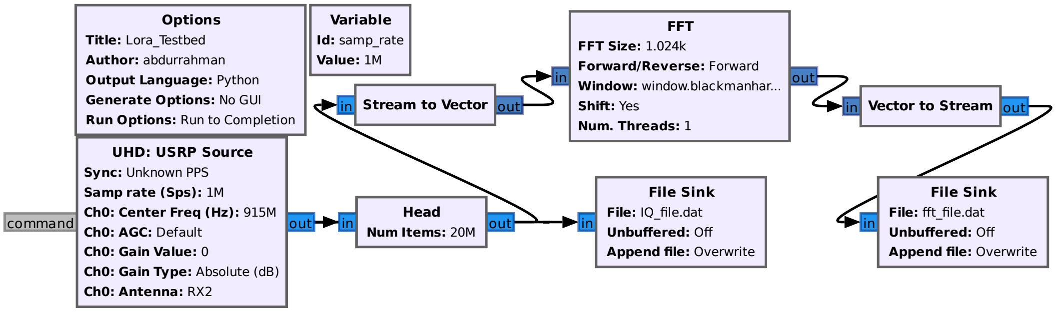

Figure 2: The flowgraph of our data collection.

3.2 Setup 2: Different Days Outdoor Scenario

3.2.1 Transmitter Side

- 23 Lopy4 boards and 2 Fipy boards on top of 22 Pysense sensor shields, 2 Pytrack sensor shields, and 1 Pyscan sensor shield.
- Each device uses its own LoRa Whip, Tilt RF Antenna, and they all are powered via Lithium Ion Battery −3.7v 2000mAh.
- Protocol of operation: LoRa Modulation.
- LoRa configuration: Raw-LoRa mode, a channel bandwidth of 125KHz, a spreading factor of 7, a preamble of 8, a TX power of 20dBm, and a coding rate of 4/5.

3.2.2 Receiver Side

- One USRP B210 receiver configured with a center frequency at 915MHz, a sample rate of 1MHz, and an observed bandwidth of 1MHz.
- One Vert900 Ettus Antenna.
- The USRP is connected to a host computer with GNURadio software with some signal processing packages.

3.2.3 Download Dataset:

Different Days Outdoor Scenario

3.3 Setup 3: Different Days Wired Scenario

3.3.1 Transmitter Side

- 23 Lopy4 boards and 2 Fipy boards on top of 22 Pysense sensor shields, 2 Pytrack sensor shields, and 1 Pyscan sensor shield.
- 30dB attenuator and an SMA Female to RP-SMA Male Cable.
- Protocol of operation: LoRa Modulation.
- LoRa configuration: Raw-LoRa mode, a channel bandwidth of 125KHz, a spreading factor of 7, a preamble of 8, a TX power of 20dBm, and a coding rate of 4/5.

3.3.2 Receiver Side

- One USRP B210 receiver configured with a center frequency at 915MHz, a sample rate of 1MHz, and an observed bandwidth of 1MHz.
- One Vert900 Ettus Antenna.
- The USRP is connected to a host computer with GNURadio software with some signal processing packages.

3.3.3 Download Dataset

Different Days Wired Scenario. (Will become available soon)

3.4 Setup 4: Different Distances Scenario

3.4.1 Transmitter Side

- 23 Lopy4 boards and 2 Fipy boards on top of 22 Pysense sensor shields, 2 Pytrack sensor shields, and 1 Pyscan sensor shield.
- Each device uses its own LoRa Whip, Tilt RF Antenna, and they all are powered via Lithium Ion Battery −3.7v 2000mAh.
- Protocol of operation: LoRa Modulation.
- LoRa configuration: Raw-LoRa mode, a channel bandwidth of 125KHz, a spreading factor of 7, a preamble of 8, a TX power of 20dBm, and a coding rate of 4/5.

3.4.2 Receiver Side

- One USRP B210 receiver configured with a center frequency at 915MHz, a sample rate of 1MHz, and an observed bandwidth of 1MHz.
- One Vert900 Ettus Antenna.
- The USRP is connected to a host computer with GNURadio software with some signal processing packages.

3.4.3 Download Dataset:

Different Distances Scenario.

3.5 Setup 5: Different Configurations Scenario

3.5.1 Transmitter Side

- 23 Lopy4 boards and 2 Fipy boards on top of 22 Pysense sensor shields, 2 Pytrack sensor shields, and 1 Pyscan sensor shield.
- Each device uses its own LoRa Whip, Tilt RF Antenna, and they all are powered via Lithium Ion Battery −3.7v 2000mAh.
- Protocol of operation: LoRa Modulation.

3.5.2 Receiver Side

- One USRP B210 receiver configured with a center frequency at 915MHz, a sample rate of 1MHz, and an observed bandwidth of 1MHz.
- One Vert900 Ettus Antenna.
- The USRP is connected to a host computer with GNURadio software with some signal processing packages.

3.5.3 Download Dataset:

Different Configurations Scenario

3.6 Setup 6: Different Locations Scenario

3.6.1 Transmitter Side

- 23 Lopy4 boards and 2 Fipy boards on top of 22 Pysense sensor shields, 2 Pytrack sensor shields, and 1 Pyscan sensor shield.
- Each device uses its own LoRa Whip, Tilt RF Antenna, and they all are powered via Lithium Ion Battery −3.7v 2000mAh.
- Protocol of operation: LoRa Modulation.

3.6.2 Receiver Side

- One USRP B210 receiver configured with a center frequency at 915MHz, a sample rate of 1MHz, and an observed bandwidth of 1MHz.
- One Vert900 Ettus Antenna.
- The USRP is connected to a host computer with GNURadio software with some signal processing packages.

3.6.3 Download Dataset:

Different Locations Scenario.

3.7 Setup 7: Different Receivers Scenario

3.7.1 Transmitter Side

- 23 Lopy4 boards and 2 Fipy boards on top of 22 Pysense sensor shields, 2 Pytrack sensor shields, and 1 Pyscan sensor shield.
- Each device uses its own LoRa Whip, Tilt RF Antenna, and they all are powered via Lithium Ion Battery −3.7v 2000mAh.
- Protocol of operation: LoRa Modulation.

3.7.2 Receiver Side

- Two USRP B210 receiver configured with a center frequency at 915MHz, a sample rate of 1MHz, and an observed bandwidth of 1MHz.
- Two Vert900 Ettus Antenna.
- The USRP is connected to a host computer with GNURadio software with some signal processing packages.

3.7.3 Download Dataset:

Different Receivers Scenario.

References

[1] Abdurrahman Elmaghbub and Bechir Hamdaoui. LoRa device fingerprinting in the wild: Disclosing RF data-driven fingerprint sensitivity to deployment variability. *IEEE Access*, 2021.

[2] Abdurrahman Elmaghbub and Bechir Hamdaoui. Comprehensive RF dataset collection and release: A deep learning-based device fingerprinting use case. In *2021 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2021.