Release Note: WiFi RF Datasets for Device Fingerprinting Using Deep Learning

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

Version 2, October 2023

1 Quick Dataset Download Links

WiFi datasets (download links given below) are described and used in the paper titled ADL-ID: Adversarial Disentanglement Learning for Wireless Device Fingerprinting Temporal Domain Adaptation.

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 following paper:

Copy and paste the bibtex below:

```
@article{elmaghbub2023adl,
title={ADL-ID: Adversarial Disentanglement Learning for Wireless Device Fingerprinting Temporal Domain Ada
author={Elmaghbub, Abdurrahman and Hamdaoui, Bechir and Wong, Weng-Keen},
journal={arXiv preprint arXiv:2301.12360},
year={2023}}
```

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

2 General Description

This WiFi fingerprint dataset has been collected at NetSTAR lab at Oregon State University, as part of an NSF project. in which we published several works in the effort of solving the RFFP problem [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]. The dataset provides time-domain I/Q samples of WiFi IEEE802.11b transmissions collected using our extended testbed. The testbed consists of 50 Pycom devices (25 LoPy & 25 FiPy boards connected to PySense extension boards) and an Ettus USRP B210 receivers, operating at a center frequency of 2.412GHz, used for recording the received signals sampled at 25MS/s. Our WiFi dataset contains 1.2TB of WiFi transmissions of 50 Pycom devices captured over 5 consecutive days in both indoor (laboratory) and outdoor environments. For each day, we capture 5 transmissions from each device in a round-robin fashion, where each capture consists of 50M complex-valued I/Q samples. The time-gap between two consecutive captures of the same device is 5mins. Transmitters were located about 5 meters away from the access point and the receiver, with a clear line of sight.

The two 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. 1 shows the general flow graph used for our data acquisition.

We created raw I/Q representation files for each transmission in“.dat” format. The binary files are encoded with

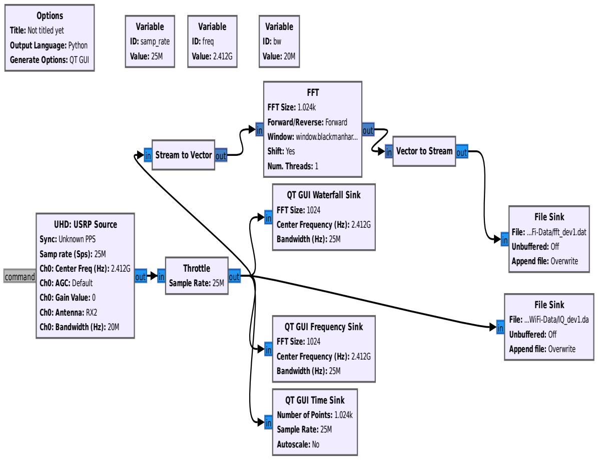

Figure 1: The GNURadio flowgraph of our data collection.

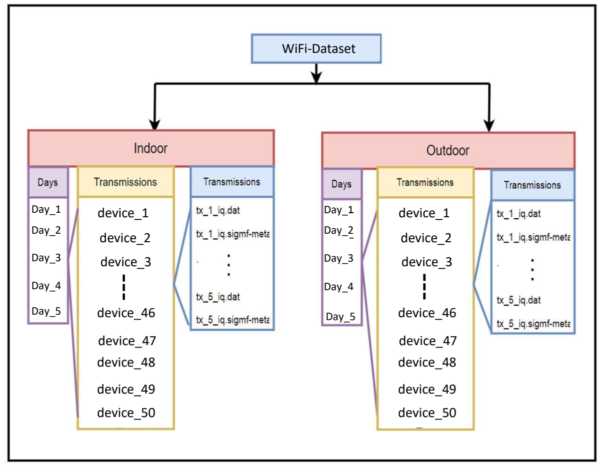

Figure 2: File/link structure and organization of the dataset. Notice that wifi-dataset directory is under rf-datasets directory.

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 2000 files with more than 1.2TB of data, which are described next (refer to Fig. 2 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 50 subdirectories, one for each device. Each device subdirectory has 5 SigMF recordings corresponding to the 5 transmissions (with each transmission having a dataset binary file and a metadata file).
  Download Setup 1 dataset: Different Days Indoor Scenario.
- **Setup 2: Diff_Days_Outdoor_Setup** directory has 5 subdirectories, one for each day. Each day subdirectory has 50 subdirectories, one for each device. Each device subdirectory has 5 SigMF recordings corresponding to the 5 transmissions (with each transmission having a binary file and a metadata file).
  Download Setup 2 dataset: Different Days Outdoor Scenario.

Our testbed for the two scenarios consist of the following components:

2.0.1 Transmitter Side

- 50 Pycom development boards: 25 LoPy boards, 25 FiPy boards, and 50 PySense extension boards.
- Pycom boards' networks: LoRa, WiFi, Bluetooth, Sigfox, and Cellular LTE.
- Hardware:
  - ESP32-D0WDQ6 Chip(WiFi & Bluetooth capabilities).
  - Semtech SX1276 Chip (LoRa & Sigfox).
  - Sequans Monarch Chip (LTE).
  - Nano Usim slot (Sim card).
- Software:
  - FREERTOS (Operating system).

Setups	Number of Devices	Number of Receivers	Protocol	Number Days	Transmissions per Device	Duration per Transmission	Distances	Environment	Representation
1) Diff Days Indoor	50	1	WiFi	5	5	2s	5m	Indoor	IQ
2) Diff Days Outdoor	50	1	WiFi	5	5	2s	5m	Outdoor	IQ

Table 1: Summary of Experimental Setups/Scenarios.

  - Micropython (Development).
  - Pymakr (REPL Console).

- Each device uses its own external WiFi Antenna, and they all are powered via Lithium Ion Battery −3.7v 1100mAh, 4.1Wh.
- Protocol of operation: WiFi IEEE 802.11b.
- WiFi configuration: Direct Sequence Spreading Spectrum (DSSS) modulation, 1Mbps data rate, WiFi channel 1, a channel bandwidth of 20MHz, a TX power of 14dBm.

2.0.2 Receiver Side

- One Ettus USRP B210 (Software-Defined Radio).
- Lenovo ThinkPad laptop.
- Features:
  - Frequency range: 70MHz-6GHz.
  - Real-time bandwidth: 57MHz.
  - Fully coherent 2 x 2 MIMO, and Full-duplex capabilities.
  - SuperSpeed USB 3.0.
- : Hardware:
  - AD9361 RFIC direct-conversion transceiver.
  - Open Xilinx Spartan®-6 FPGA.
- Software
  - USRP Hardware Driver (UHD).
  - GNURadio.
  - C/C++ and Python APIs.
- RF Front-end Parameters: f_c: 2.412GHz, F_s: 25MSps, and B_w: 20MHz.

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, 'float32');
data = complex(val(real), val(imag));
I = data(real);
Q = data(imag);
```

References

[1] Abdurrahman Elmaghbub, Bechir Hamdaoui, and Arun Natarajan. Widescan: Exploiting out-of-band distortion for device classification using deep learning. In *GLOBECOM 2020-2020 IEEE Global Communications Conference*, pages 1–6. IEEE, 2020.

[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.

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

[4] Nora Basha, Bechir Hamdaoui, and Kathiravetpillai Sivanesan. Leveraging mimo transmit diversity for channel-agnostic device identification. In *ICC 2022-IEEE International Conference on Communications*, pages 2254–2259. IEEE, 2022.

[5] Jared Gaskin, Bechir Hamdaoui, and Weng-Keen Wong. Tweak: Towards portable deep learning models for domain-agnostic lora device authentication. *arXiv preprint arXiv:2209.00786*, 2022.

[6] Bechir Hamdaoui and Abdurrahman Elmaghbub. Deep-learning-based device fingerprinting for increased lora-iot security: Sensitivity to network deployment changes. *IEEE network*, 36(3):204–210, 2022.

[7] Bechir Hamdaoui, Abdurrahman Elmaghbub, and Siefeddine Mejri. Deep neural network feature designs for rf data-driven wireless device classification. *IEEE Network*, 35(3):191–197, 2020.

[8] Jun Chen, Weng-Keen Wong, Bechir Hamdaoui, Abdurrahman Elmaghbub, Kathiravetpillai Sivanesan, Richard Dorrance, and Lily L Yang. An analysis of complex-valued cnns for rf data-driven wireless device classification. *arXiv preprint arXiv:2202.09777*, 2022.

[9] Jared Gaskin, Abdurrahman Elmaghbub, Bechir Hamdaoui, and Weng-Keen Wong. Deep learning model portability for domain-agnostic device fingerprinting. *IEEE Access*, 11:86801–86823, 2023.

[10] Abdurrahman Elmaghbub and Bechir Hamdaoui. Eps: Distinguishable iq data representation for domain-adaptation learning of device fingerprints. *arXiv preprint arXiv:2308.04467*, 2023.