Release Note: Bluetooth RF Fingerprint Datasets: Collected After Hardware Warm-up Period

Haytham Albousayri and Bechir Hamdaoui

August 4, 2025

1 Dataset Download Links

Datasets consist of BLE transmissions from 31 devices over different environments are presented in the document. These datasets are utilized in our papers [1, 2]

- Bluetooth Fingerprint Identification Under Domain Shift Through Transient Phase Derivative, IEEE CNS 2025.
- Neural Network-Driven Estimation of Hardware Impairments for Robust Wireless Device Identification, ICML Workshop on ML4Wireless 2025.

Please cite our work in case the datasets were used.

Copy and paste the bibtext below

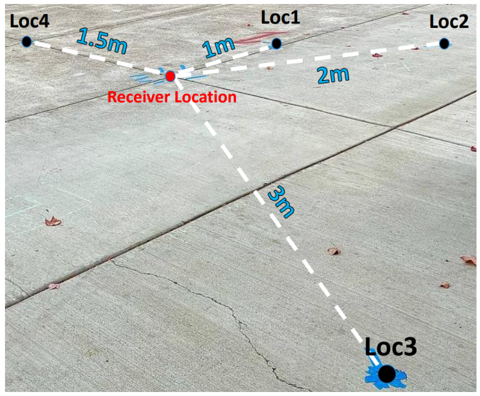
```
@inproceedings{albousayri2025neural,
  title={Neural Network-Driven Estimation of Hardware
  Impairments for Robust Wireless Device Identification},
  author={Albousayri, Haytham and Hamdaoui, Bechir and Wong, Weng-Keen},
  booktitle={ICML 2025 Workshop on Machine Learning for Wireless Communication
  and Networks (ML4Wireless)}
}
@inproceedings{albousayri2025bluetooth,
  title={Bluetooth Fingerprint Identification Under Domain Shift Through
  Transient Phase Derivative},
  author={Albousayri, Haytham and Hamdaoui, Bechir and Wong, Weng-Keen and
  Basha, Nora},
  booktitle={2025 IEEE Conference on Communications and Network Security (CNS)
     },
  year={2025},
  organization={IEEE}
}
```

The links to the different datasets:

- Scenario 1: Outdoors (Wireless) Across Locations
- Scenario 2: Indoors (Wired) Across Receivers
- Scenario 3: Indoors (Wired) Across Channels

2 Dataset Description

These BLE fingerprint datasets were collected at the NetSTAR lab at Oregon State University. We conducted three experiments, namely: *1)* Different channels on the same environment, *2)* Different receivers on the same environment, and *3)* Different environments on the same channel. For all of these experiments, we collected multiple datasets from 31 different Seeed Studio XIAO devices. Seeed Studio



(a) Outdoor setup

(b) ESP32C3

(c) USRP B210

Figure 1: Testbed setup and hardware

XIAO is an IoT mini development board based on the Espressif ESP32-C3 WiFi/Bluetooth dual-mode chip. Two Ettus USRP (Universal Software Radio Peripheral) B210 receivers were employed to sample and collect the RF data in the form of raw IQ values via GNURadio; see Figures 1(b) and 1(c). Before starting the collection process, each device is powered on and allowed a 6-minute warm-up period to ensure hardware stabilization [3] followed by 2 minutes of data collection. Bandwidth was set to 2 MHz, the sampling rate was set to 6 MS/s and power gain was set to 29dB and 8dB for the wireless data collection and wired data collection, respectively. We utilized 1M PHY modulation scheme for this experiment, where each symbol represents 1 bit of information without employing any coding scheme. The frequency channels we used for the above mentioned experiments are: Channel 1 (Ch1), Channel 2 (Ch2), Channel 14 (Ch14) and Channel 32 (Ch32), centered on 2.406GHz, 2.408GHz, 2.434GHz and 2.470GHz, respectively.

2.1 Description of the different scenarios

- **Outdoors (Wireless) across locations:** Here we collected our data across 4 different locations, where the receiver is kept at a fixed location and the transmitters are placed at 1m, 1.5m, 2m and 3m away from the receiver, as shown in Figure 1. For this experiment, we only used receiver 1 (Rx1) for the data collection and the RF signals were transmitted over Ch1.
  Download scenario 1 dataset: Outdoors (Wireless) Across Locations.

- **Indoors (Wired) across channels:** Here we collected our data across 4 frequency channels, namely Ch1, Ch2, Ch14 and Ch32. Wired transmission was used and all of the signals were collected on Rx1.
  Download scenario 2 dataset: Indoors (Wired) Across Channels.

- **Indoor (Wired) across receivers:** The data were collected from Rx1 for all of the devices and then collected again on receiver 2 (Rx2); this was done to ensure that the collection was done after the same warm up time (6 mins), i.e., there will be no variation in the hardware impairments. Channel Ch1 was utilized in this experiment.
  Download scenario 3 dataset: Indoors (Wired) Across Receivers.

2.2 Dataset

The payload inside each frame was kept the same across all transmissions, meaning the frames (equivalently captured signals) were identical when operating on the same channel. When the channel is changed, the frame content is changed as well, since the header contains channel-specific information. The frames (each of duration 308.33μs) were transmitted back to back with short gaps between them. After sampling the signals and removing the silent periods, we obtained 1850 complex IQ samples in the time domain per signal. Once all signals were extracted, we performed power normalization.

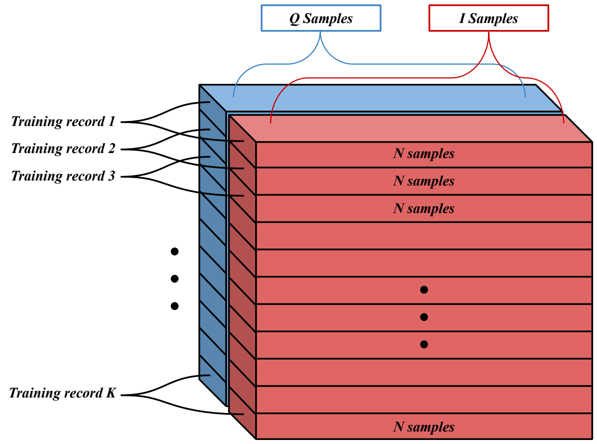


Figure 2: Dataset Tensor

3 File format description

We separated the power-normalized complex IQ signals in the time domain into their real and imaginary components, and stacked them into a 3D tensor of shape $(D, 2, N)$. Here, D is the number of training signals, 2 refers to the I and Q components, and N is the number of samples per frame ($N = 1850$). Figure 2 illustrates the structure of the resulting dataset.

4 Code Example

You can find the source code on our GitHub repository at BLE-RF-Fingerprint-CNN-classifier. The code includes how to load the data, process it, and train a CNN based classifier to identify devices. This classifier was used in [4], [1].

```
# Download the dataset from
# Change Directory to where the data is

# For user to change
Your_Dir = "..."

X_train = np.load(f"{Your_Dir}/X_train.npy")
Y_train = np.load(f"{Your_Dir}/Y_train.npy")
X_test = np.load(f"{Your_Dir}/X_test.npy")
Y_test = np.load(f"{Your_Dir}/Y_test.npy")
```

References

[1] H. Albousayri, B. Hamdaoui, W.-K. Wong, and N. Basha, “Bluetooth fingerprint identification under domain shift through transient phase derivative,” in *2025 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2025, pp. 1–9.

[2] H. Albousayri, B. Hamdaoui, and W.-K. Wong, “Neural network-driven estimation of hardware impairments for robust wireless device identification,” in *ICML 2025 Workshop on Machine Learning for Wireless Communication and Networks (ML4Wireless)*, 2025.

[3] A. Elmaghbub and B. Hamdaoui, “Distinguishable iq feature representation for domain-adaptation learning of wifi device fingerprints,” *IEEE Transactions on Machine Learning in Communications and Networking*, 2024.

[4] B. Johnson, H. Albousayri, B. Hamdaoui, and L. Dunn, “Domain-adaptive device fingerprints for network access authentication through multifractal dimension representation,” *IEEE Network*, 2025.