Release Note: WiFi (802.11b/n) RF Fingerprints with Corresponding Hardware Impairments Collected During & After Hardware Warm-up

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

Version 1, June 2024

1 Quick Dataset Download Links

This document describes four datasets: two WiFi 802.11b datasets (two-day wired and two-day wireless) and two WiFi 802.11n datasets (one-day wired and one-day wireless). These datasets were collected from the same Pycom devices via the Keysight PXA spectrum analyzer. For each device, we saved the time-domain I/Q values and the corresponding hardware impairments of more than 2500 packets spanning the initial 30 minutes of the device's operational life, covering both the hardware warm-up period and stable period (after hardware warm-up ends).

These datasets are detailed and used in the paper titled No Blind Spots: On the Resiliency of Device Fingerprints to Hardware Warm-Up Through Sequential Transfer Learning for RF device fingerprinting. This dataset can also be used for other applications, like deep learning-based impairment estimation and compensation.

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:

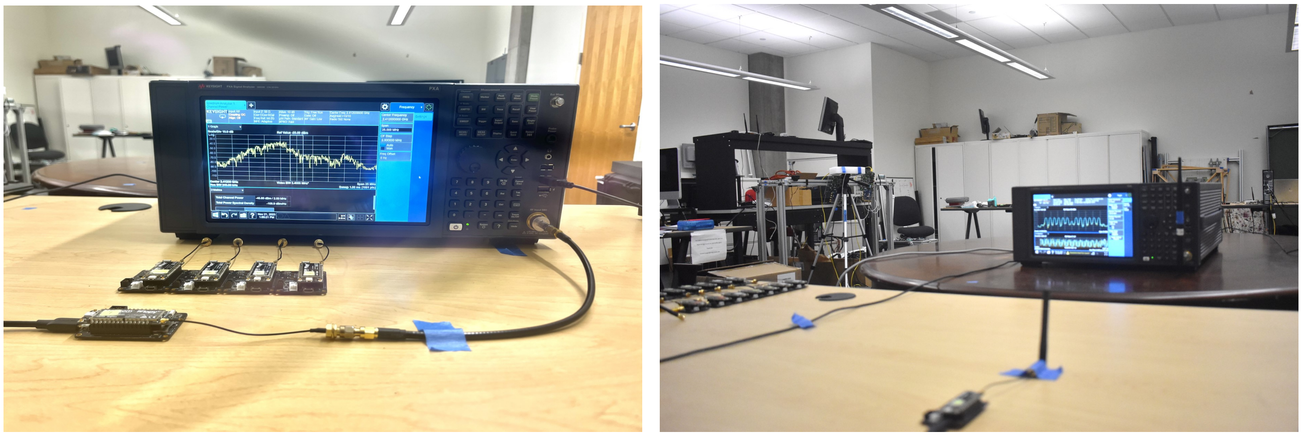
Copy and paste the BibTeX below:
@inproceedings{elmaghbub2024no,
title={No Blind Spots: On the Resiliency of Device Fingerprints to Hardware Warm-Up Through Sequential Transfer Learning},
author={Elmaghbub, Abdurrahman and Hamdaoui, Bechir},
booktitle={In Proceedings of the 17th ACM Conference on Security and Privacy in Wireless and Mobile Networks},
pages={134144},
year={2024}}

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

- Scenario 1: Two-Day WiFi B Wired Scenario
- Scenario 2: Two-Day WiFi B Wireless Scenario
- Scenario 3: One-Day WiFi N Wired Scenario
- Scenario 4: One-Day WiFi N Wireless Scenario

2 Brief Dataset Description

These WiFi fingerprint datasets have been collected at the 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, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 5, 22]. These WiFi datasets contain WiFi 802.11b & WiFi 802.11n transmissions from the same 15 Pycom devices captured in both wired and wireless scenarios across several days. The datasets contain both the time-domain I/Q samples and the corresponding hardware impairments of the collected packets.



(a) Wired Setup (b) Wireless Setup

Figure 1: The Hardware Impairment Measurement Setup

- **WiFi Type B Datasets:** We captured both wired and wireless datasets of 15 Pycom devices transmitting WiFi 802.11b using the HR/DSSS physical-layer mode. The initial data collection involved gathering a wireless dataset (Day 1 dataset), followed by a deliberate two-week gap before resuming data collection to obtain the second wireless dataset (Day 2 dataset). Wireless data offers the advantage of capturing real-world environmental interactions, providing a comprehensive view of signal behavior in typical usage scenarios. Additionally, we collected 2 wired datasets, which eliminates environmental variables, on two different days thus offering a controlled setting to focus on hardware-specific impairments. Within each dataset, we captured 3000 frames per device, spanning the initial 30 minutes of each device's operation, ensuring the inclusion of both the warm-up and stable phases across various days. All devices used the same antenna which is positioned 1 meter away from the spectrum analyzer, as visually depicted in Fig. 1. The resulting datasets encompass over 180k frames, where each frame is characterized by its time-domain I/Q samples, represented as (2x17550) dimensions, and, notably, includes the corresponding 8 key hardware impairments (EVM, CFO, Symbol Clock Error, IQ Offset, Magnitude Error, Phase Error, Carrier Suppression Error, and Average Burst Power). They can be downloaded from NetSTAR Lab at:
  - Scenario 1: Two-Day WiFi B Wired Scenario.
  - Scenario 2: Two-Day WiFi B Wireless Scenario
- **WiFi Type N Datasets:** We used the same transmitters to send the same message but using the WiFi 802.11n protocol, which employs different modulation and data rates from that employed by WiFi 802.11b. Our WiFi 802.11n transmitters utilize OFDM modulation with QAM-16 for subcarrier modulation and a 20MHz bandwidth. Similar to the previous datasets, we captured data in both wired and wireless setups, using the same measurement setup. The resulting datasets encompass over 96k frames, where each frame is characterized by its time-domain I/Q samples, represented as (2x1014) dimensions, and notably include the corresponding 8 key hardware impairments (EVM, CFO, Symbol Clock Error, IQ Offset, IQ Gain Imbalance, Quadrature Error, IQ Timing Skew, and Pilot EVM). These datasets can also be found and downloaded from the NetSTAR Lab page at:
  - Scenario 3: One-Day WiFi N Wired Scenario
  - Scenario 4: One-Day WiFi N Wireless Scenario

3 Brief Setup Description

We established a dedicated measurement setup, as depicted in Fig. 1, to closely monitor the behavior of various impairments of 15 Pycom devices both during and after the warm-up period over one month. Specifically, we established a wired connection for each Pycom device to interface with a Keysight PXA signal analyzer N9030B, running the WLAN 802.11 X-Series Measurement Application. Our Pycom devices were programmed to consistently transmit identical WiFi 802.11b & 802.11n packets at fixed intervals. Simultaneously, we configured the signal analyzer to sample incoming RF bursts at a rate of 35MSps and a bandwidth of 20MHz. Over the initial 30 minutes, encompassing both the warm-up and stabilization periods, the spectrum analyzer receives more than 2500 packets, extracts IQ samples, and calculates the corresponding impairments for each packet.

4 File Format Description

The WiFi packets, alongside instrument configuration details, were archived in CSV files. Raw time-domain I/Q data and impairments were extracted and stored in NumPy files for convenience. For WiFi Type B, each file comprises an Nx35141 data array (N is the number of captured packets, 35141 is the number of samples per packet), encompassing the impairments and the I/Q values. Specifically, the 1st sample/column contains the frame index/number, the following 40 samples (col 2 to col 41) contain the impairments values, the subsequent samples (42 to 17591) contain the In-phase (I) component values, and the last (col 17592 to 35141) samples contain the Q component values. The set of 8 impairments used in our evaluation is {EVM, CFO, SCE, IQ offset, Mag Error, Phase Error, Carrier Suppression Error, Avg Burst Power} whose impairments have the following indices: [2, 8, 12, 14, 21, 27, 34, 36]. For WiFi Type N, the dataset is (Nx2086) data array, with the first sample/column encompassing the frame number and the next 61 indices (from col 2 to col 62) hardware impairments. Samples/col 63 to 1074 contain the In-phase (I) component values, while the remaining columns contain the Quadrature (Q) component values. The set of 8 impairments used in our evaluation is {EVM, CFO, SCE, IQ offset, IQ Gain Imb, Quadrature Error, Pilot EVM, IQ Timing Skew} with indices: [2, 8, 12, 14, 16, 18, 27, 32]. First index is 0.

5 Code Example

This is an example of using Python to read the files in our dataset (one WiFi B packet and one WiFi N packet):

```python
import numpy as np
# For WiFi Type B
# Change the file path
B_data = np.load('.../RFFP-dataset/WiFi-B-N-Dataset/WiFi-B/Wired/Day1/Device1-15Jan.npy')
# for packet #1
# Impairments: ["EVM", "CFO", "SCE", "IQ offset", "Mag Error", "Phase Error", "Carrier Suppression
     Error", "Avg Burst Power"]
pkt_nbr, imps, data = B_data[0,0], B_data[0, [2, 8, 12, 14, 21, 27, 34, 36]], B_data[0, 42:]

# For WiFi Type N
# Change the file path
N_data = np.load('.../RFFP-dataset/WiFi-B-N-Dataset/WiFi-N/Wired/Device4-31Jan.npy')
# for packet #1
# Impairments: ["EVM", "CFO", "SCE", "IQ offset", "Gain Imb", "Q Imb", "Pilot EVM", "IQ Timing
    Skew"]
pkt_nbr, imps, data = N_data[0,0], N_data[0,  [2, 8, 12, 14, 16, 18, 27, 32]], N_data[0, 62:]
```

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. In *2022 IEEE conference on communications and network security (CNS)*, pages 1–9. IEEE, 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] Jun Chen, Weng-Keen Wong, and Bechir Hamdaoui. Unsupervised contrastive learning for robust rf device fingerprinting under time-domain shift. *arXiv preprint arXiv:2403.04036*, 2024.

[10] Benjamin Johnson and Bechir Hamdaoui. On the domain generalizability of rf fingerprints through multifractal dimension representation. In *2023 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9. IEEE, 2023.

[11] 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.

[12] Abdurrahman Elmaghbub and Bechir Hamdaoui. A needle in a haystack: Distinguishable deep neural network features for domain-agnostic device fingerprinting. In *2023 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9. IEEE, 2023.

[13] Luke Puppo, Weng-Keen Wong, Bechir Hamdaoui, and Abdurrahman Elmaghbub. Hinova: A novel open-set detection method for automating rf device authentication. In *2023 IEEE Symposium on Computers and Communications (ISCC)*, pages 1122–1128. IEEE, 2023.

[14] Abdurrahman Elmaghbub, Bechir Hamdaoui, and Weng-Keen Wong. Adl-id: Adversarial disentanglement learning for wireless device fingerprinting temporal domain adaptation. In *ICC 2023-IEEE International Conference on Communications*, pages 6199–6204. IEEE, 2023.

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

[16] Abdurrahman Elmaghbub and Bechir Hamdaoui. No blind spots: On the resiliency of device fingerprints to hardware warm-up through sequential transfer learning. In *Proceedings of the 17th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 134–144, 2024.

[17] Bechir Hamdaoui, Nora Basha, and Kathiravetpillai Sivanesan. Deep learning-enabled zero-touch device identification: Mitigating the impact of channel variability through mimo diversity. *IEEE Communications Magazine*, 61(6):80–85, 2023.

[18] Nora Basha, Bechir Hamdaoui, Kathiravetpillai Sivanesan, and Mohsen Guizani. Channel-resilient deep-learning-driven device fingerprinting through multiple data streams. *IEEE Open Journal of the Communications Society*, 4:118, 2023.

[19] Bechir Hamdaoui and Abdurrahman Elmaghbub. Uncovering the portability limitation of deep learning-based wireless device fingerprints. *arXiv preprint arXiv:2211.07687*, 2022.

[20] Jiaqi Bao, Bechir Hamdaoui, and Weng-Keen Wong. Iot device type identification using hybrid deep learning approach for increased iot security. In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pages 565–570. IEEE, 2020.

[21] Luke Puppo, Weng-Keen Wong, Bechir Hamdaoui, Abdurrahman Elmaghbub, and Lucy Lin. On the extraction of rf fingerprints from lstm hidden-state values for robust open-set detection. *ITU Journal on Future and Evolving Technologies*, 5(1), 2024.

[22] Abdurrahman Elmaghbub and Bechir Hamdaoui. Wireless device classification apparatus and method, March 26 2024. US Patent 11,943,003.