Datasets Release: An IEEE 802.11 a/g (WiFi) massive-scale and labeled datasets for Radio Fingerprinting

Institute for the Wireless Internet of Things,
Northeastern University, Boston, MA 02115, USA

June 22, 2020

1 Download our datasets

To download our datasets please use the following links:

- **Arena Wireless Different Antennas** "Setup 1"

- **Arena Wireless Single Antenna** "Setup 2"

- **Arena Wired** "Setup 3"

- **Anechoic Chamber Wireless Single Antenna** "Setup 4"

These datasets were used for the paper **"Exposing the Fingerprint: Dissecting the Impact of the Wireless Channel on Radio Fingerprinting"**. Please use this link ShawabkaINFOCOM2020.pdf to download the paper.

Any use of this dataset which results in an academic publication or other publication which includes a bibliography should include a citation to our paper. Here is the reference for this work:

Conference version PDF:
wineslab_papers/ShawabkaINFOCOM2020.pdf.

Copy Bibtex:
wines_bibtex/ShawabkaINFOCOM2020.txt.

2 Description

Recent research proved that by applying convolutional neural networks with their multidimensional mappings we can achieve much better fingerprinting accuracy that we will not be

able to achieve using the traditional low-dimensional algorithms. But since we are investigating the fingerprinting in the wireless domain we should analyze the wireless channel to understand its impact on CNN-based radio fingerprinting algorithms.

Thinking of what radio fingerprinting is missing and if we look at the computer vision field, we can find that this domain has a massive scale of labeled datasets such as MNIST, Cifar and Imagenet. All these datasets are available for research benchmark activities. But the radio fingerprinting domain lacks such large-scale waveform databases, that can provide a common Benchmark for researchers working in this field.

The key contributions of [1] is to report and publicly share the largest dataset collection campaign and experimental evaluation ever conducted to evaluate radio fingerprinting algorithms using rich and diverse channels and testing environments.

The experiment was done in two different environments: (i) "in the wild" and (ii) an Anechoic chamber. Examining different channel conditions: (i) wireless with different antennas, (ii) wireless with the same antenna and (iii) wired using SMA cable.

3   Experimental Setup

In our experiment we generate and transmit WiFi signals from 20 National Instruments Software Defined Radios running Gnuradio. Figure 1 summarizes the I/Q data collection procedure implemented using Gnuradio at the receiver side. We collect one transmission at a time using one single receiver.

In our campaign we stored the received waveforms at three different WiFi demodulation process stages: (i) raw IQ samples before applying FFT operation, (ii) raw IQ after applying FFT operation and (iii) the equalized IQ samples.
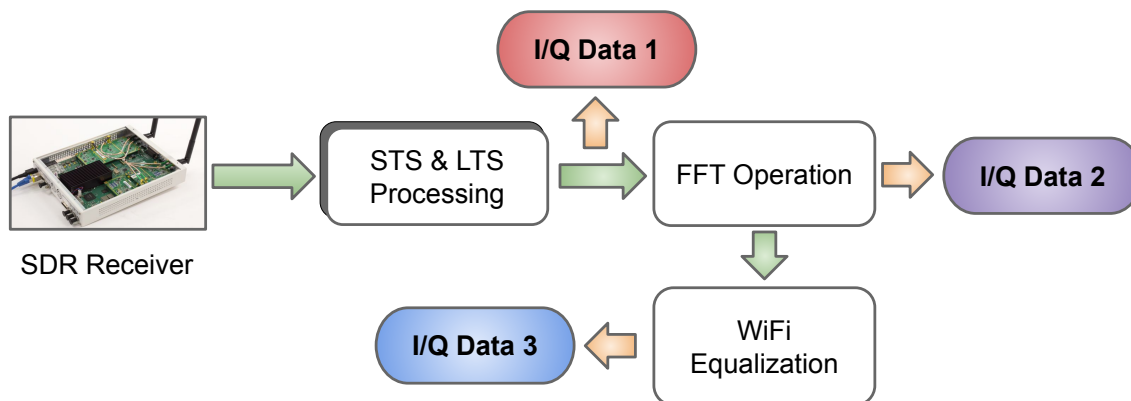

Figure 1: I/Q data collection methodology [1].

Analyzing the RF fingerprinting using such different IQ samples will allow us to understand the impact of each stage on the fingerprinting accuracy and to define the best data type that can help the CNN model in extracting the device unique hardware impairments to boost the classification accuracy.

3.1 "Setup 1" Arena testbed: wireless with different antennas

3.1.1 Transmitter side:

- 12 NI N210 and 8 NI X310 SDRs running GnuRadio.

- One device transmitting at a time for 30 seconds.

- Transmitting the same IEEE 802.11 a/g (WiFi with 2.432 GHz and 20 MS/s).

- Using BPSK modulation.

- 20 Ettus VERT2450 antenna, one per device.

3.1.2 Receiver side:

- One fixed receiver with Ettus VERT2450 antenna for all cases.

- Pilots used to perform channel estimation and offset correction.

- In this setup we release I/Q Data 3. I/Q Data 3: is a WiFi equalized data collected after applying the equalization process, collected and stored at the receiver side as specified in Figure 1.

3.1.3 Channel condition:

- Arena "in the wild" testbed.

- Wireless with different antennas.

3.1.4 Download "Setup 1" dataset:

- Wireless with different antennas Arena dataset is stored in:
  https://repository.library.northeastern.edu/collections/neu:gm80j757w.

- In this dataset, each recorded transmission consists of two files: (i) a dataset file which is a binary file of the recorded digital samples and (ii) a meta file which contains information that describes the dataset in plain-text JSON format [2]. Our binary and meta format is an extension of, and compatible with the SigMF specifications.

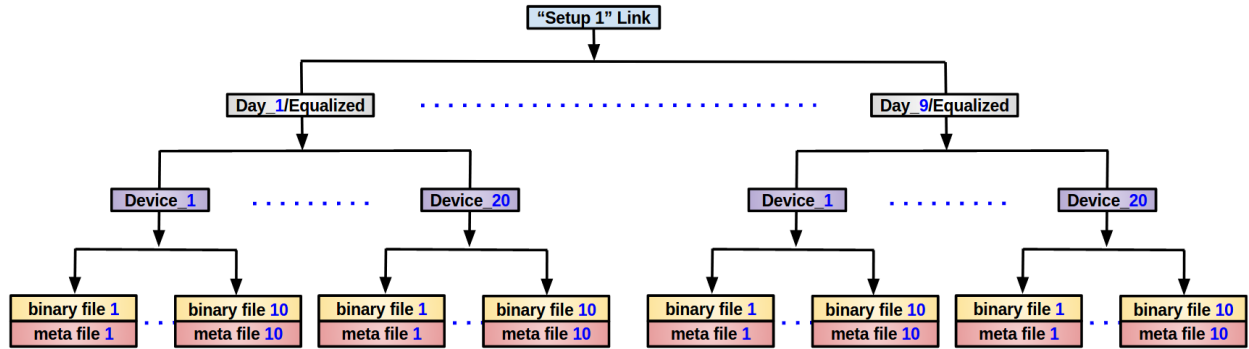- Chart 2 describes "Setup 1" online link organization structure.

Figure 2: "Setup 1" online link organization.

3.2 "Setup 2" Arena testbed: wireless with single antenna

3.2.1 Transmitter side:

- 12 NI N210 and 8 NI X310 SDRs running GnuRadio.

- One device transmitting at a time for 30 seconds.

- Transmitting the same IEEE 802.11 a/g (WiFi with 2.432 GHz and 20 MS/s).

- Using BPSK modulation.

- One fixed Ettus VERT2450 antenna, used for all devices.

3.2.2 Receiver side:

- One fixed receiver with Ettus VERT2450 antenna for all cases.

- Pilots used to perform channel estimation and offset correction.

- In this setup we release I/Q Data 3. I/Q Data 3: is a WiFi equalized data collected after applying the equalization process, collected and stored at the receiver side as specified in Figure 1.

3.2.3 Channel condition:

- Arena "in the wild" testbed.

- Wireless with single antennas.

3.2.4 Download "Setup 2" dataset:

- Wireless with single antennas Arena dataset is stored in:
  https://repository.library.northeastern.edu/collections/neu:gm80j756m.

4

- In this dataset, each recorded transmission consists of two files: (i) a dataset file which is a binary file of the recorded digital samples and (ii) a meta file which contains information that describes the dataset in plain-text JSON format [2]. Our binary and meta format is an extension of, and compatible with the SigMF specifications.

- Chart 3 describes "Setup 2" online link organization structure.

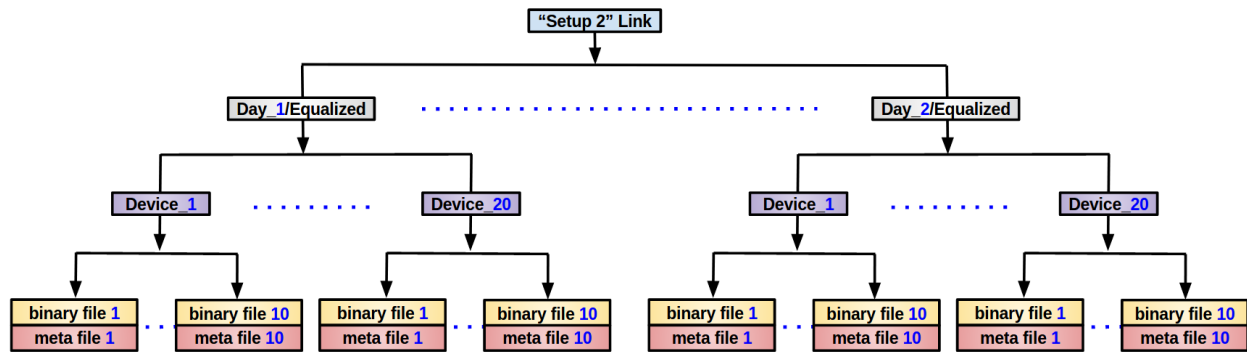
Figure 3: "Setup 2" online link organization.

3.3 "Setup 3" Arena testbed: wired connection

3.3.1 Transmitter side:

- 12 NI N210 and 8 NI X310 SDRs running GnuRadio.

- One device transmitting at a time for 30 seconds.

- Transmitting the same IEEE 802.11 a/g (WiFi with 2.432 GHz and 20 MS/s).

- Using BPSK modulation.

- SMA cable to stream the WiFi signal from the transmitter to the receiver.

3.3.2 Receiver side:

- One fixed receiver with Ettus VERT2450 antenna for all cases.

- Pilots used to perform channel estimation and offset correction.

- Data collected and stored at the receiver side on three different stages as specified in Figure 1

- I/Q Data 1: WiFi raw data collected before applying FFT operation, but after frame detection and time synchronization operations.

- I/Q Data 2: WiFi raw data collected after applying FFT operation directly, and before any equalization process.

- I/Q Data 3: WiFi equalized data collected after applying equalization process.

3.3.3   Channel condition:

- Arena "in the wild" testbed.

- The wired antenna setup.

- SMA cable used to connect the transmitter and the receiver.

3.3.4   Download "Setup 3" dataset:

- The wired Arena dataset is stored in:
  https://repository.library.northeastern.edu/collections/neu:gm80j7542.

- In this dataset, each recorded transmission consists of two files: (i) a dataset file which is a binary file of the recorded digital samples and (ii) a meta file which contains information that describes the dataset in plain-text JSON format [2]. Our binary and meta format is an extension of, and compatible with the SigMF specifications.

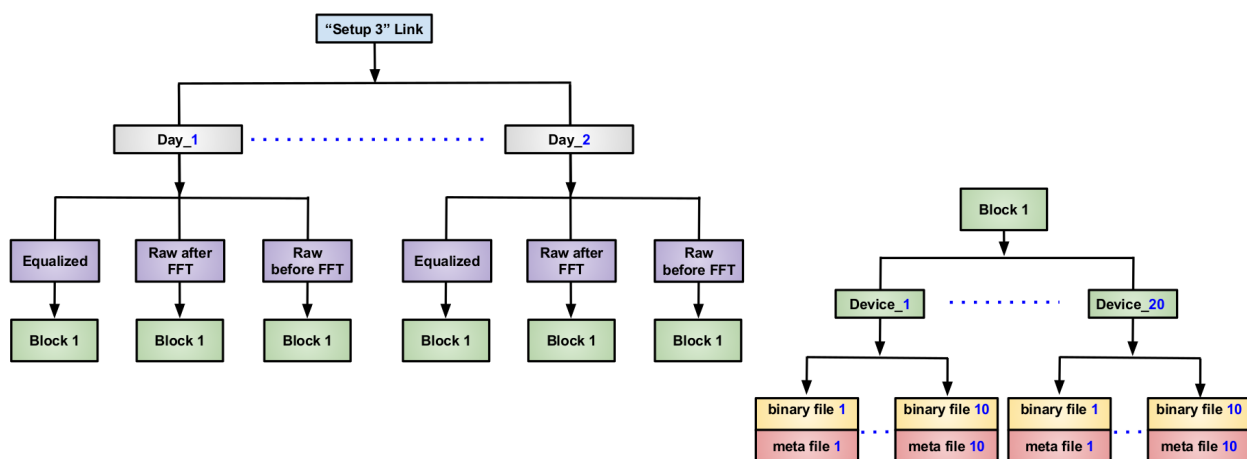- Figure 4 describes "Setup 3" online link organization structure.

Figure 4: "Setup 3" online link organization.

3.4   "Setup 4" Anechoic chamber testbed: wireless with single antenna

3.4.1   Transmitter side:

- 4 NI N210 and 6 NI X310 SDRs running GnuRadio.

- One device transmitting at a time for 30 seconds.

- Transmitting the same IEEE 802.11 a/g (WiFi with 2.432 GHz and 20 MS/s).

- Using BPSK modulation.

- One fixed Ettus VERT2450 antenna, used for all devices.

6

3.4.2 Receiver side:

- One fixed receiver with Ettus VERT2450 antenna for all cases

- Pilots used to perform channel estimation and offset correction

- Data collected and stored at the receiver side on three different stages as specified in Figure 1

- I/Q Data 1: WiFi raw data collected before applying FFT operation, but after frame detection and time synchronization operations.

- I/Q Data 2: WiFi raw data collected after applying FFT operation directly, and before any equalization process

- I/Q Data 3: WiFi equalized data collected after applying equalization process.

3.4.3 Channel condition:

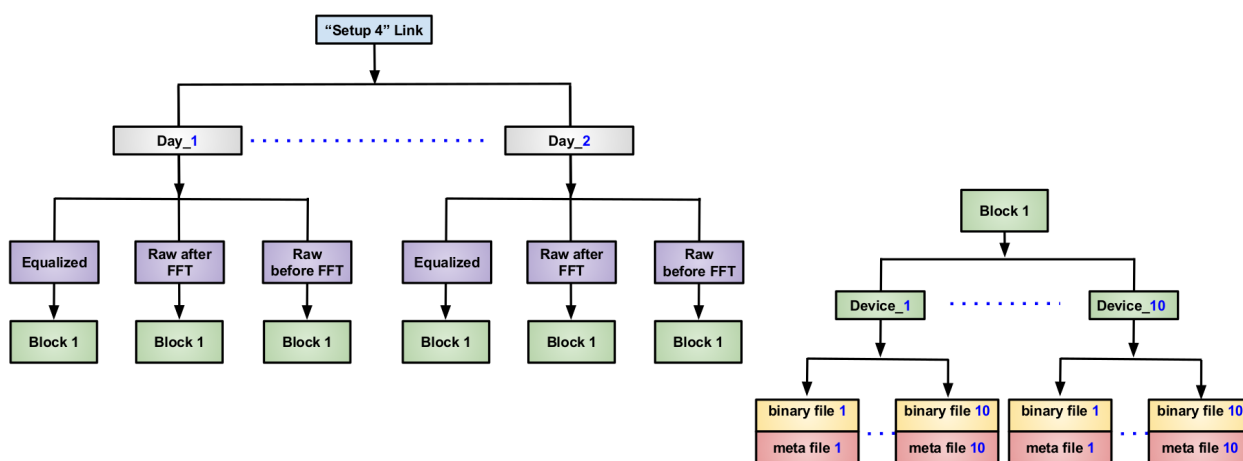- Anechoic chamber testbed.

- Wireless with single antennas.

Figure 5: "Setup 4" online link organization.

3.4.4 Download "Setup 4" dataset:

- The Anechoic Chamber dataset is stored in:
  https://repository.library.northeastern.edu/collections/neu:gm80j486q.

- In this dataset, each recorded transmission consists of two files: (i) a dataset file which is a binary file of the recorded digital samples and (ii) a meta file which contains information that describes the dataset in plain-text JSON format [2]. Our binary and meta format is an extension of, and compatible with the SigMF specifications.

- Figure 5 describes "Setup 4" online link organization structure.

4 Sigmf Description

4.1 Global Object:

The global object consists of name/value pairs that provide information applicable to the entire dataset. It contains the information that is minimally necessary to open and parse the dataset file, as well as general information about the recording itself. The following names are specified in the core name space, please refer to Table 1 for more details:

TABLE 1: SIGMF GLOBAL DATA.

Name	Required	Type	Description
datatype	true	string	The format of the stored samples in the dataset file.
sample_rate	true	double	The sample rate of the signal in samples per second.
version	true	string	The version of the SigMF specification used to create the metadata file.
sha512	false	string	The SHA512 hash of the dataset file associated with the SigMF file.
description	false	string	A text description of the SigMF recording.
hw	false	string	A text description of the hardware used to make the recording.
recorder	false	string	The name of the software used to make this SigMF recording.
author	false	string	The author's name

4.2 Captures:

As per the SigMF specifications, the captured value is an array of capture segment objects that describe the parameters of the signal capture. It MUST be sorted by the value of each capture segment's core:sample_start key, ascending. The following names are specified in the core namespace, please refer to Table 2 for more details :

TABLE 2: SIGMF CAPTURE DATA.

Name	Required	Type	Description
sample_start	true	unit	The sample index in the dataset file at which this segment takes effect.
global_index	false	double	The center frequency of the signal in Hz.
datetime	true	string	TAn ISO8601 string indicating the timestamp of the sample index specified by sample_start.

4.3 Annotations:

According to the SigMF specifications, the Annotations value is an array of annotation segment objects that describe anything regarding the signal data not part of the global and captures objects. Each SigMF annotation segment object must contain a core:sample_start name/value pair, which indicates the first index at which the rest of the segment's name/value pairs apply. Please refer to Table 3 for more details :

4.4 Transmitter Object:

The Transmitter object contains the following name/value pairs. Please refer to Table 4 for more details :

TABLE 3: SigMF annotation data.

Name	Required	Type	Unit	Description
environment	true	double	N/A	A description of the environment where the antenna is mounted.
transmitter_identification	false	object	N/A	Transmitter identification parameters. See Transmitter Object definition.
receiver_identification	false	object	N/A	Receiver identification parameters. See Receiver Object definition.
distance	false	string	feet	Distance between transmitter and receiver.

TABLE 4: SigMF Transmitter Object.

Name	Required	Type	Unit	Description
model	true	string	N/A	Make and model of the transmitter. E.g., "Ettus N210", "Ettus B200", "Keysight N6841A", "Tektronix B206B".
serial_number	false	string	N/A	Globally unique identifier.
low_frequency	false	float	Hz	Low frequency of operational range of the transmitter.
high_frequency	false	float	Hz	high frequency of operational range of the transmitter.
noise_figure	false	float	dB	Noise figure of the transmitter.
max_power	false	float	dBm	Maximum input power of the transmitter.
hw	true	object	N/A	See Antenna Object definition.

4.5 Receiver Object:

The Receiver object contains the following name/value pairs. Please refer to Table 5 for more details :

TABLE 5: SigMF Receiver Object.

Name	Required	Type	Unit	Description
model	true	string	N/A	Make and model of the Receiver. E.g., "Ettus N210", "Ettus B200", "Keysight N6841A", "Tektronix B206B".
serial_number	false	string	N/A	Globally unique identifier.
low_frequency	false	float	Hz	Low frequency of operational range of the receiver.
high_frequency	false	float	Hz	high frequency of operational range of the receiver.
noise_figure	false	float	dB	Noise figure of the receiver.
max_power	false	float	dBm	Maximum input power of the receiver.
hw	true	object	N/A	See Antenna Object definition.

4.6 Antenna Object:

The Antenna object contains the following name/value pairs. Please refer to Table 6 for more details :

TABLE 6: SigMF Antenna Object.

Name	Required	Type	Unit	Description
model	true	string	N/A	Antenna make and model number. E.g. "ARA CSB-16".
type	false	string	N/A	Antenna type. E.g. "dipole", "biconical", "monopole".
low_frequency	false	float	Hz	Low frequency of operational range.
high_frequency	false	float	Hz	High frequency of operational range.

References

[1] A. Al-Shawabka, F. Restuccia, S. D'Oro, T. Jian, B. C. Rendon, N. Soltani, J. Dy, K. Chowdhury, S. Ioannidis, T. Melodia, Exposing the Fingerprint: Dissecting the Impact of the Wireless Channel on Radio Fingerprinting, Proc. of IEEE Conference on Computer Communications (INFOCOM) (2020).

[2] B. Hilburn, N. West, T. O'Shea, T. Roy, SigMF: The Signal Metadata Format, in: Proceedings of the GNU Radio Conference, Vol. 3, 2018.