# The DSI dataset for Wi-Fi fingerprinting using mobile devices

*Adriano Moreira, Ivo Silva, Joaquín Torres-Sospedra, v1, 30.04.2020*

## 1. Contents

This document describes one dataset with samples taken from a Wi-Fi network interface for experiments with indoor positioning based on Wi-Fi fingerprinting.

To reference this dataset, please use:

> *Moreira, A.; Silva, I. and Torres-Sospedra, J. The DSI dataset for Wi-Fi fingerprinting using mobile devices, Zenodo 2020. http://dx.doi.org/10.5281/zenodo.3778646*

For further information, please contact adriano.moreira@algoritmi.uminho.pt.

## 2. Dataset format

This dataset includes two sets of samples:
(i) radio map – a set of Wi-Fi samples collected at a grid of points (reference points);
(ii) trajectory – a set of Wi-Fi samples collected along a pedestrian trajectory.

Both sets are described in the same format, and all files are CSV – Comma Separated Values plain text (UTF-8).

For the radio map, the provided files have their names starting with "rm_"; for the trajectory, the trajectory files, "tj_" is used.

Three files are provided for each set:
- rm_crd.csv: hold coordinates (x,y) where the samples were collected;
- rm_rss.csv: hold the measured RSSI values from each of the Access Points (AP) detected in each sample;
- rm_tms.csv: hold the timestamps of the instant each sample was collected.

**Coordinates:** Each sample is associated to a pair of coordinates in a 2D Euclidean reference system. The origin of the reference system was chosen arbitrarily for convenience. The units are meters. Therefore, distances between points can be easy calculated.

**RSSI values:** The RSSI values provided as read from the Wi-Fi network interface through the Android API. In each sample, a value of -150 was assigned to each AP not detected during a measurement. No information is provided about the MAC addresses of the APs. However, in the files, the same order is used for all samples, meaning that the values in each column are all associated to the same AP.

**Timestamps:** The provided timestamps are in the UNIX format, and are represented in seconds.

None of the provided files include an identifier for each sample. The values in the three provided files are associated by the line number, meaning that the coordinates, RSSI values and timestamps in the same line, in each file, refer to the same sample.

## 3. Data collection process

The provided data were collected at the first floor of building 11 of the University of Minho, Portugal, Azurém Campus (Department of Information Systems – DSI), between 2 and 3 of May 2016.
Data was collected by a single person, using a Nvidia Shield tablet running Android 6.0.

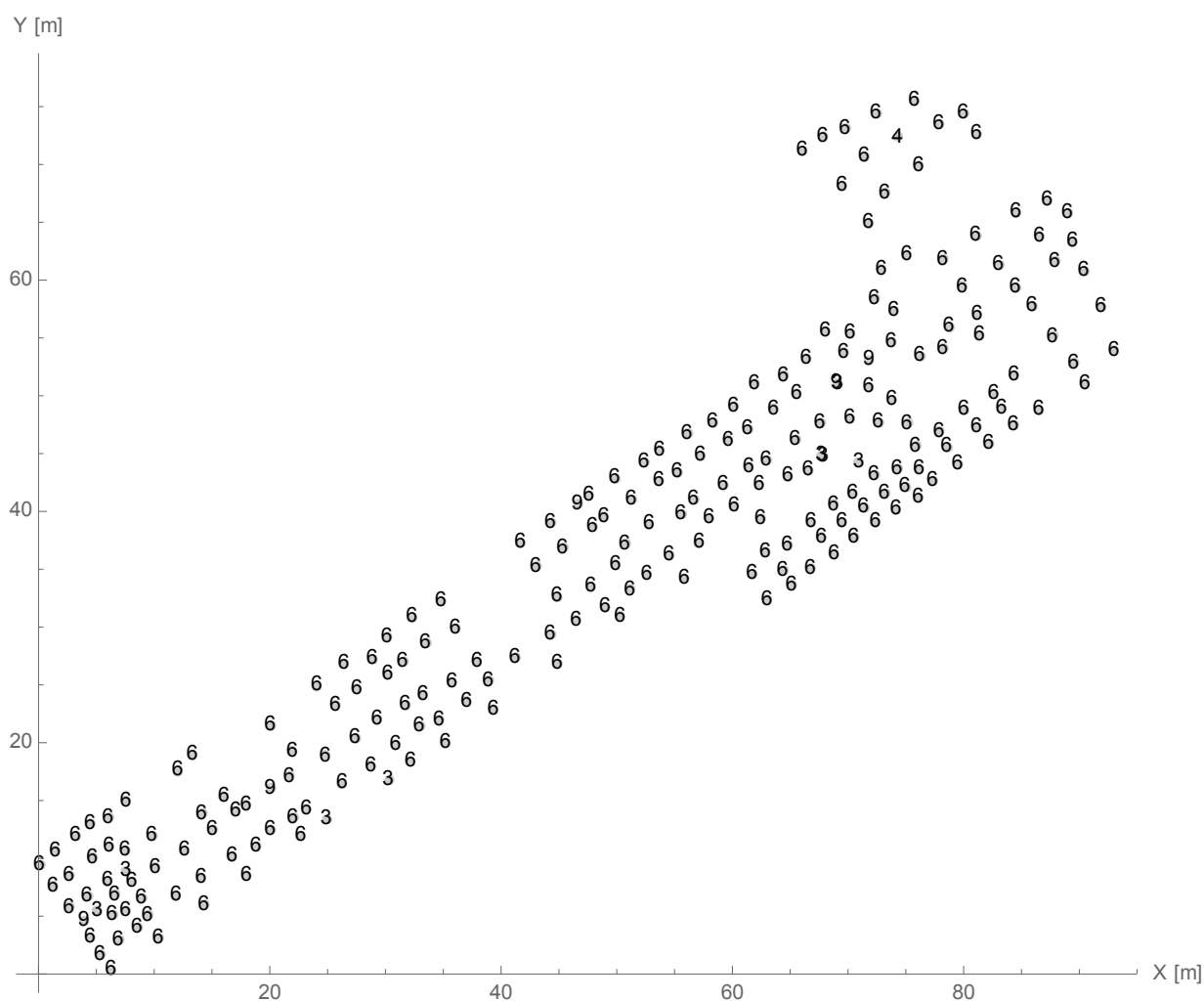Figure 1, below, shows the reference points and the number of samples collected at each one of them.



Figure 1 – Number of samples collected at each reference point.

The radio map is made of 6 samples per reference point (there are a few cases with a different number of samples). A summary of the collected data follows:

- total number of samples: 1369
- number of distinct APs: 157
- number of distinct buildings: 1
- number of distinct floors: 1
- number of distinct rooms: 25 (room information is not provided in this dataset)
- number of reference points: 230
- number of distinct devices: 1
- number of distinct users: 1

The user's trajectory is made of a set of samples collected manually while a user was walking. The device was hold in the hand, with the screen pointing upwards.

Figure 2, below, shows the trajectory and the points at which samples were collected.
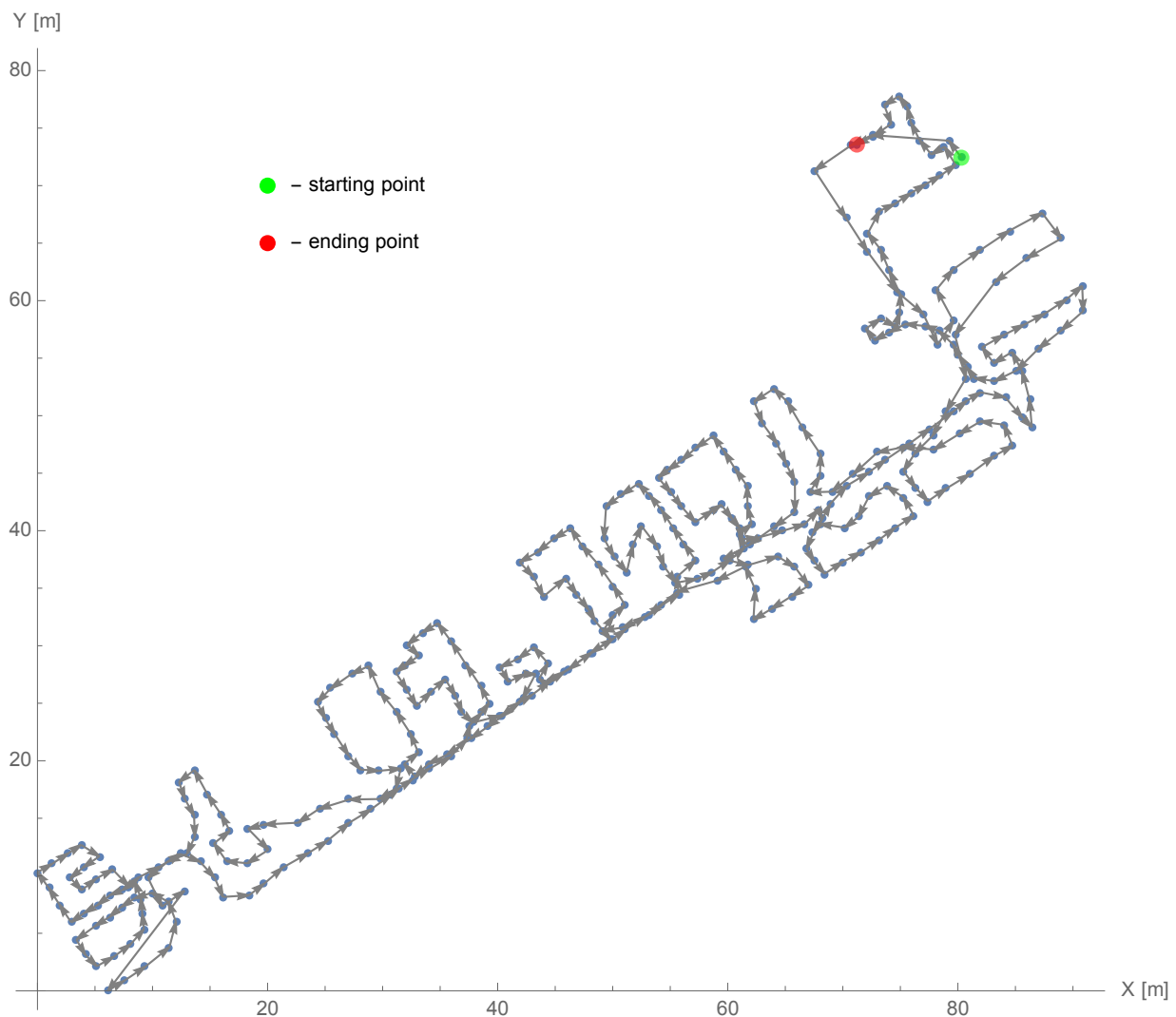


Figure 2 – Collected samples along the trajectory.

A summary of the collected data follows:
- number of samples: 348
- number of distinct APs: 157
- number of distinct buildings: 1
- number of distinct floors: 1
- number of distinct rooms: 22
- number of distinct positions: 348
- number of distinct devices: 1
- number of distinct users: 1

The sequence of samples taken along the user's trajectory exhibit spatial and temporal properties as described below. Figure 3 shows the distribution of the displacement values (distance between consecutive sample points) (distance in meters):
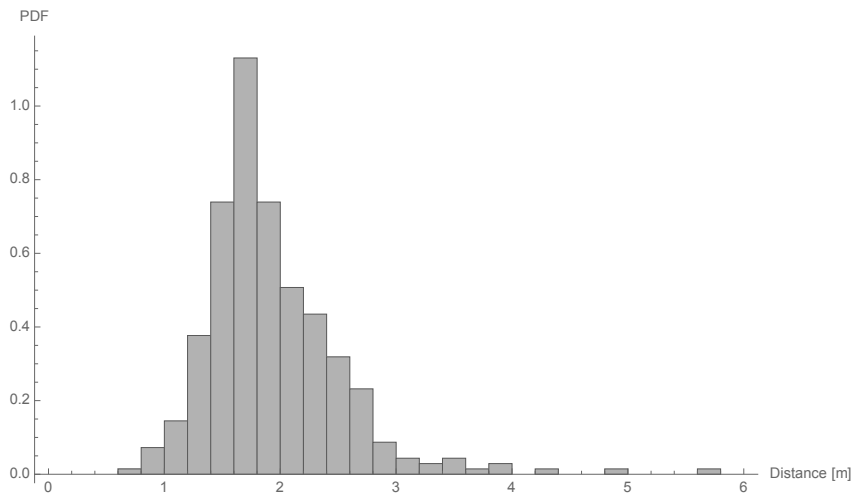


Figure 3 – Distribution of the distance between consecutive samples along the trajectory.

Figure 4 shows the distribution of the time elapsed between consecutive samples (time in seconds):
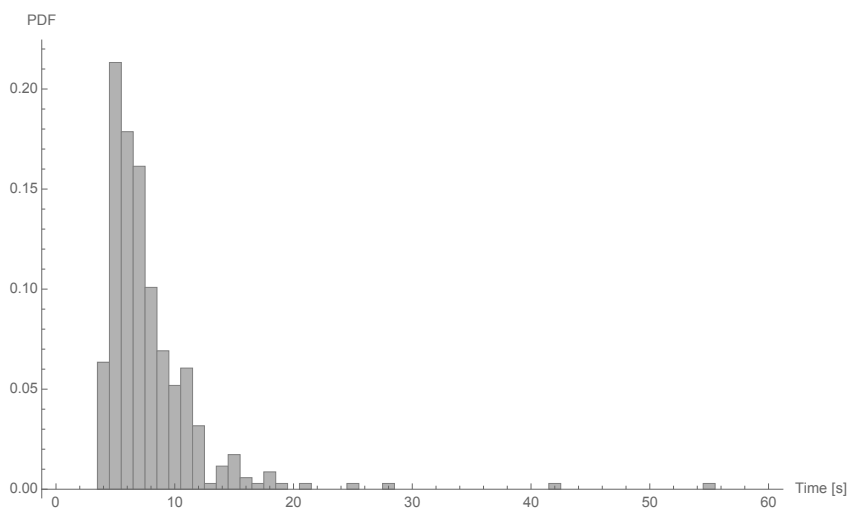


Figure 4 – Distribution of the time elapsed between consecutive samples along the trajectory.