Horizon 2020
European Union Funding
for Research & Innovation

Database of RF fingerprinting on use case IoT devices

RadioSpin Project

H2020-FETPROACT09-2020

Author: Alice Mizrahi, THALES

Deliverable D6.2
Work Package 6

Version: Draft/Final, 11/07/2022

Project number: 101017098

Horizon 2020
European Union Funding
for Research & Innovation

Deliverable No. 6.2

Dissemination Level		
PU	Public	x

Nature		
DATA	Data sets, microdata, etc.	X

Deliverable Details			
Due date:	31/07/2022		
Submission date:			
Involved participants:	Thales, CNRS		
WP no. and title:	WP6 Large scale simulations		
WP leader:	THALES		
Task no. and title:	Task T6.2 Database and RF fingerprinting use case evaluation		
Version/Date			
Keywords:	Dataset, RF fingerprinting, drones, neural network		

Summary

We have collected a database of 1000 RF signals in order to evaluate the spintronic neural network simulator on an RF Fingerprinting task. The chosen task is to identify commercial drones and their radio-controllers out of ten classes. The signals were adapted from real data recorded from the devices in an anechoic chamber. The resulting signals are suitable for the RadioSpin simulator and will be used for neural networks learning and inference. We demonstrate the suitability of the dataset by performing classification with a simple neural network on the simulator and achieving 94.1 % accuracy. Further work will focus on achieving higher accuracy, using more complex neural networks, and benchmarking the performance.

Contents

1. Goal of the deliverable

This deliverable is part of *Task 6.2 – Database and RF fingerprinting use case evaluation*.

In the Grant Agreement, this task is described as follows. *Thales will collect real or realistic RF signatures from drones or other IoT devices in order to constitute a training dataset. Thales and CNRS will assess the most suitable algorithm for this use case, among those developed in WP4 and use it to train large-scale neural networks for this task. Thales will then benchmark RadioSpin technology in terms of accuracy, size, power consumption and speed.*

The goal of deliverable *D6.2 Database of RF fingerprinting on use case IoT devices* is to present the database that we have constituted. This database has been adapted from an open database, to suit our requirements. The requirements are the following:

- The database should compose a meaningful task, linked to a real-life application and deriving from real data.
- The database should be compatible with the simulator in terms of functional capacity. This limitation is not too constraining since we are doing simulations. The only thing to be mindful of is that the networks required to solve the task should not be so large that their simulation time is prohibitive.
- The database should be compatible with the simulator in terms of system parameters. This means that the amplitudes and frequencies of the RF signals must match realistic parameters of spintronic devices.

Link to Deliverable D5.2 RF fingerprinting database

As described below, we chose the same original signal database as for the D5.2. However, the dataset we have developed for D5.2 is aimed at the physical demonstrator, which requirements are much more constraining than for the simulator. In consequence the dataset we develop here, for the simulator, is a different modification of the original database.
The main differences are the number of inputs and the number of classes.

2. Description of the database

2.1. Chosen task: identification of commercial drones

The RadioSpectrum is full of different radiofrequency (RF) signals. It is critical for the good functioning of the internet of things that smart connected objects are able to identify the origin of these RF signals, i.e. to perform RF fingerprinting.
We choose to focus on RF signals emitted by commercial drones and their radio-controllers.

We use the data collected by Basak et al. in “Drone classification from RF fingerprints using deep residual nets” (IEEE COMSNETS conference, 2021). This database has been identified and investigated in WP6. Basak et al. collected signals in an anechoic chamber, using a universal software radio peripheral (USRP X310) placed seven meters apart from the devices (as shown in Figure 1). The signals were all in the 2.4 GHz ISM band and the whole 100 MHz band was received instantaneously using a receiving sampling rate of 100 MSps (i.e. the system down-converted the signal frequencies to the 0-100 MHz band to sample them correctly).

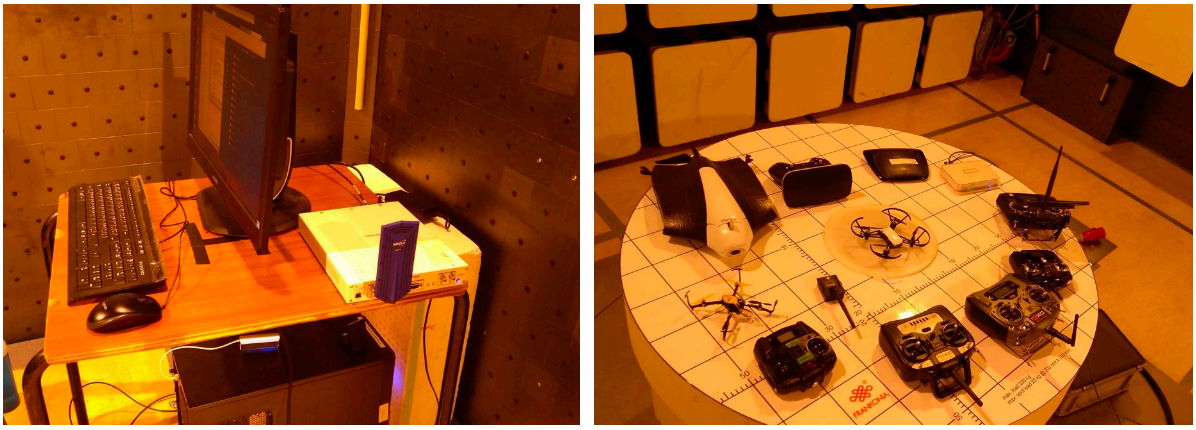

Figure 1. Left: the USRP recording system. Right: the drones and radio-controllers. Both photos are taken in the anechoic chamber and reproduced from Basak et al., 2021.

Basak et al. have recorded signals from ten devices. Here, because the dataset is aimed at the simulator, we are able to keep the ten classes. The different devices are described in Figure 2. The task is to identify from a received RF signal which of the ten devices emitted it.

Figure 2: Different devices for the RF fingerprinting database.

2.2. Description of the signals

Each recording of a signal is a spectrogram of size 256 x 256 (i.e. 256 frequency bins and 256 time frames). Figure 3 shows one example of spectrogram of each class. In the experiment of Basak et al., the signals frequencies where down converted from the WiFi band (around 2.4 GHz) to the 0 – 100 MHz range so that they could be sampled and digitized. However, when applied to RadioSpin technology, the signals would not require digitization: the spintronics devices are able to process analogue RF signals because of their intrinsic RF dynamics. In consequence, the lowest frequency bin can be set to an arbitrary value. In a future application, this value would correspond to the chosen down-conversion frequency, or to the frequency of the raw signals.

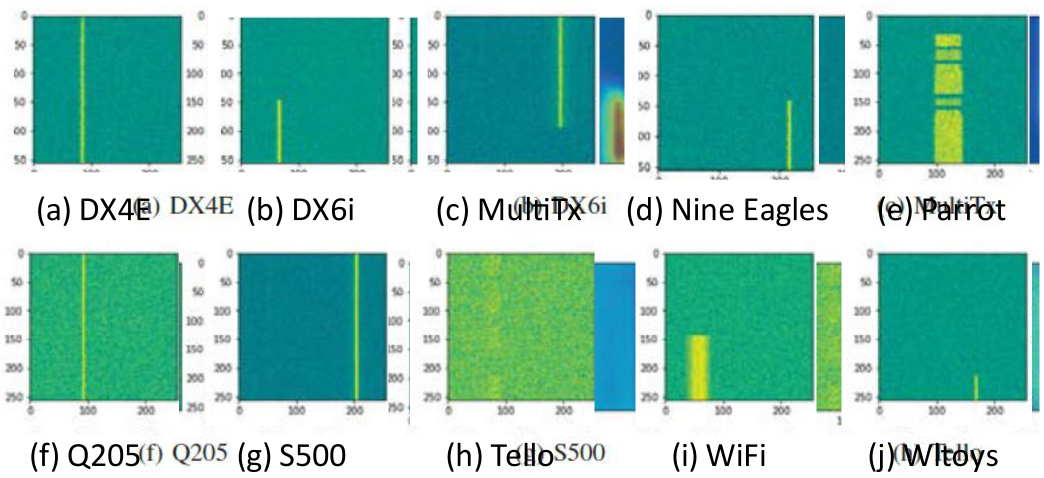

Figure 3: Spectrogram of each class

3. Preliminary test: classification with a feed-forward perceptron

To test that this database is adequate, we perform a preliminary test with the simulator. Note that the simulator is still under construction so this uses a simple first version of it. The first test is to perform classification of the signals with a feed-forward perceptron, which is the simplest neural network.

3.1. Adaptation of the database for feed-forward neural networks

Feed-forward neural networks are networks where information flows only one way, from the inputs to the outputs. There are no recurrences (backward connections) or connections within one neuron layer. Algorithms on how to compose feed forward neural networks with spintronic oscillators are studied in *Task 4.1 Networks with semi-independent oscillators*. Feed-forward networks are the most used in machine learning and in applications. Therefore, we chose to focus first on such networks for the use case of RF fingerprinting.

Feed forward neural networks do not have a state memory or feedback loop. In consequence, we are first not interested in the temporal evolution of the amplitude of the RF signals. We transform the spectrograms into simple spectra (power density versus frequency) by averaging over the time dimension. In the original database, white noise was added. For now, we chose to restrict ourselves to the lowest noise level (i.e. the noise that is naturally present), but might use the whole dataset later. The resulting inputs constitute an “adapted dataset”. However, in further stages of the project, we will study more complex neural networks (in *Task*

4.2 Deep reservoir computing and *Task 4.3 Synchronization based computing*), and might apply the resulting algorithms on this RF fingerprinting use case.

Figure 4 shows the resulting signals of the adapted dataset. We have renormalized the power to be compatible with realistic parameters of the spintronic devices. Here the frequencies are between 20 MHz and 120 MHz, corresponding to a down-conversion, as done by Basak et al. to collect the signals. In further stages of the project, we will perform classification of the signal before down-conversion, i.e. at 2.4 GHz.

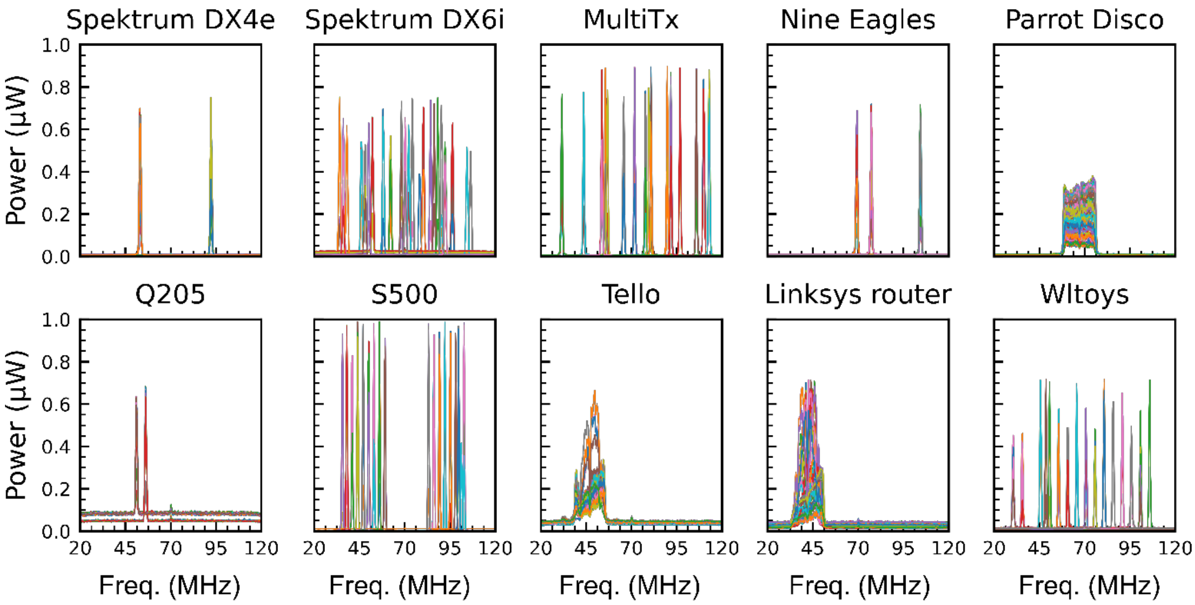

Figure 4: Resulting signals of the dataset adapted for feed-forward networks. Each color is one example of signal.

3.2. Classification results

We perform our first test on a perceptron, i.e. a single layer neural network where each input is connected to each output. Here the inputs are the powers at each of the 256 frequency bins between 20 and 120 MHz. The outputs are the classes 0 to 9, corresponding to the ten different devices. In consequence, there are 256 x 10 = 2560 synaptic weights.

In our simulator, these weights are implemented by ten synaptic chains of 256 spintronic resonators each. In each chain, the resonators initial frequencies range linearly from 20 to 120 MHz. During training, the frequencies are adjusted progressively in order to achieve the desired weights and perform classification (see *Deliverable D4.1 Shallow networks with semi-independent oscillators* for more details on how the simulator implements spintronic networks).

Figure 5 shows the accuracy (i.e. the percentage of correctly classified signals) versus the number of epochs (i.e. the number of times that the training set is presented to the network). **We reach 94.1 % accuracy on the validation set. This result is promising and validates the suitability of the database.** In further work, we will improve this accuracy by using more complex neural networks with more layers.

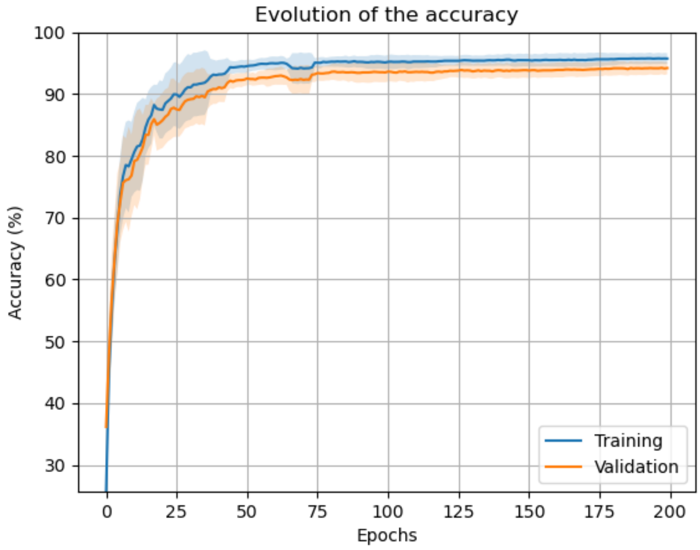

Figure 5: Accuracy versus the number of epochs, fort he training set (blue) and validation set (orange). Ten trials are performed, the solid lines are the mean values and the shaded zone represent the standard deviation.

4. How to use the database

The database is composed of 1000 signals. There are 702 signals in the train set and 298 in the validation set. Each signal is a vector of 256 elements (i.e. power at each frequency bin). To each signal corresponds a target, which is the number of the class it belongs to.

Device	Target
Parrot Disco	0
Q205	1
Tello	2
MultiTx	3
Nine Eagles	4
Spektrum DX4e	5
Spektrum DX6i	6
Wltoys	7
S500	8
Linkys router	9

The database is stored in an h5 file, a format adapted to databases. Inside the file there are two datasets: the signals ('Signals') and the targets ('Targets'). The data can then easily be extracted and fed to a machine learning data-loader (from the Pytorch library for instance).

The database file is ***RadioSpin_D62_RF_fingerprinting.h5***

5. Conclusion

We have chosen an open database of real RF signals collected from commercial drones and controllers. We have adapted the signals into a dataset suitable for our spintronic neural network simulator. We have conducted a preliminary test on a simple neural network and achieved 94.1 % accuracy, a promising result, which validates the suitability of the dataset. Further work will focus on achieving higher accuracy, using more complex neural networks, and benchmarking the performance.

6. ACKNOWLEDGEMENT

This deliverable is part of the RadioSpin project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101017098.