Tampere University

Alejandro Adolfo Sanz Abia

IMPLEMENTATION OF A RADIO FREQUENCY FINGERPRINT DETECTOR BASED ON GNSS SIGNALS

Tampere University (TAU)
Bachelor's Thesis
January 2020

ABSTRACT

Geolocation is one of the most significant manifestations of the current development of information technologies and it is used for multiple applications, such as mobile networks, military systems, or in the stock market. For that reason, it is important to verify the source of this type of signals, as they could be susceptible to being tricked by spoofing attacks, namely fake transmitters. This thesis is based on the development of a GNSS signal type classifier based on radio frequency (RF) fingerprinting methods that will determine if a signal belongs to an authorized transmitter or if it comes from a non-authorized GNSS signal generator/repeater. First, a total of 620 signals have been recorded in lab environments, follows: 40 different scenarios of real GNSS signal (with antennas located on the roof of the university) and 580 scenarios of the generated signal (using a GNSS signal generator). Each of the scenarios contains different types of signals (different GNSS constellations and/or bands, different satellites, etc.). Then, using a MATLAB-based simulator, the recorded signal is read, a certain time-frequency transform is applied (in this case the discrete Wavelet Transform), and an image of the wavelet transform of each sample is saved. These images include the features of the signal's RF fingerprinting. Next, a machine learning algorithm called SVM, also designed in MATLAB, is used. This algorithm classifies two or more different signal classes, and finally evaluate the classification accuracy. We used 80% of the images in each category for training and the remaining 20% for testing. Finally, a confusion matrix is obtained showing the accuracy obtained by the SVM algorithm in the testing phase.

The analysis of the results has shown that the SVM classification algorithm can be a very effective model for the identification of GNSS transmitters through the use of fingerprinting features. It has been observed that when the Spectracom scenario is configured with more than one satellite, accuracy is lower compared to being configured with only one. This is because the signal obtained when more than one satellite is configured is more similar to the signal obtained from the antenna in comparison to the single satellite configuration, and for that reason, SVM has more difficulty in classifying it correctly. Another observation is that accuracy is also reduced when more than two categories are classified at the same time compared to a binary classification. Despite this, the accuracy is very high in the scenarios used, with 99.47% being the lowest value obtained and 100% the highest. Therefore, this implementation of RF fingerprinting methods is very promising in the context of determining whether a signal belongs to the actual GNSS satellite constellation or to a signal generator with a high level of accuracy.

Keywords: Spoofing; radio frequency fingerprinting; GNSS; Spectracom; wavelet transform; SVM.

PREFACE

This Bachelor's Thesis has been written during my stay in Tampere as an exchange student at Tampere University during Fall 2019. It represents the final work of my bachelor's degree in Telematics Engineering at the Polytechnic University of Madrid.

I would like to thank my thesis supervisors Elena-Simona Lohan and Ruben Morales for their dedication and help during the development of this thesis. Also, to my parents for their unconditional encouragement during all these years, specially to my father, to whom I dedicate the end of this challenging journey and all my future achievements.

Tampere, 10th of January 2020

Alejandro Adolfo Sanz Abia

CONTENT

FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ARNS	Aeronautical Radio Navigation Service
ADC	Analog-to-Digital Converter
CDMA	Code Division Multiple Access
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
ESA	European Space Agency
GLONASS	Global Navigation Satellite System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
ITU	International Telecommunications Union
ML	Machine Learning
PPD	Privacy Protection Devices
PRN	Pseudorandom Noise
PVT	Position Velocity and Time
RF	Radio Frequency
RNSS	Radio navigation Satellite Service
SDR	Software-Defined Radio
SV	Space Vehicle
SVM	Support Vector Machine
USRP	Universal Software Radio Peripheral

1. INTRODUCTION

This chapter gives a brief introduction to the addressed topic. In first place, the motivation and the objectives of the thesis are discussed, then the Author's contribution is emphasized. After that, the structure of the thesis is described.

1.1 Motivation

Radio frequency fingerprinting presents a new model of transmitter identification based on the fingerprint that characterizes the signal they emit. This fingerprint is unique to each device, which normally occurs due to hardware imperfections causing a number of signal effects such as I/Q imbalance, phase imbalance, frequency error or signal strength. Such imperfections, combined, form the radiometric fingerprint of the transmitter. By implementing RF fingerprinting methods, it is possible to identify whether a signal is coming from a transmitter of the GNSS constellation or whether it is coming from a GNSS signal generator with possible malicious intentions such as spoofing or Jamming attacks. This will ensure more secure communications minimizing the risk of being attacked.

1.2 Thesis objectives

The main purpose of this thesis has been the implementation of RF fingerprinting methods based on recorded Global Navigation Satellite System (GNSS) signals in order to identify and classify features of the different GNSS signals. The purpose of an RF fingerprinting algorithm is to determine if a certain received GNSS signal belongs to the real constellation of satellites or rather to an attacker (e.g., spoofer, jammer, etc.) generating and transmitting fake GNSS-like signals. Therefore, the final goal is to obtain a detector that by using the features extracted from both real and fake signals is capable to perform this distinction.

1.3 Author's contributions

The main contributions of this thesis work have been:

- Co-defining a set of target measurement scenarios in collaboration with the supervisors. The scenarios were defined based on available equipment at TAU, namely two GNSS roof antennas and a Spectracom GNSS simulator.
- Literature review on RF fingerprinting approaches in the context of GNSS.
- Setting the measurement set-up and detailed data collection based on the pre-defined scenarios.

- Data classification with machine learning algorithms and analysis of the results.

1.4 Thesis structure

The document is organized as follows: Section 2 contains an introduction to GNSS and to the main concepts that will help to understand better next sections. Section 3 contains explanations about interferences in GNSS frequency bands, as for example, spoofing or space weather interferences. Section 4 is dedicated to the explanations of what Radio Frequency Fingerprinting (RFF) is, by explaining some of the properties that can be found during the radio transmission. Section 5 goes deeper in the main topic of this thesis. It is explained in Section 5 how the transmitter-specific features can be extracted from the recorded signal and how different machine learning extractor methods can be applied to classify the collected data.

Sections 6 and 7 describe the measurement campaigns and the measurement-based results. First, an explanation of the setup where the measurements were recorded is given in Section 6. Section 7 analyses the measurement-based results according to the selected machine learning classifiers. This chapter will include the use of a learning algorithm, in particular linear support vector machine (SVM) which will help us to compare the large number of signal recordings. In addition, methods and particular signal features will be explained. Finally, conclusions and future works are addressed in Sections 8 and 9, respectively.

2. GNSS SIGNAL CONCEPTS

GNSS term refers to four global satellite constellations, that by transmitting specific signals to a terrestrial GNSS receiver, can help in determining the receiver's position at any location around the world. By solving the position as a result of the reception of signals from different constellations of artificial satellites we can determine the geographical coordinates (including the altitude) of the given receiver.

Currently, the four global GNSS systems that exist are:

- GPS (Global Positioning System). It is a service which belongs to the United States and provides users with information on positioning, navigation and chronometry.
- GLONASS (Global Navigation Satellite System). This system was developed by the Soviet Union and today belongs to the Russian Federation. The deployment of satellites began in 1982.
- Galileo. It is the European satellite positioning system, developed by the European Union together with the European Space Agency (ESA). Developments are still on-going.
- BeiDou. It is a project developed by China to obtain its own navigation system for its country and neighboring regions. The first generation of the system dates from the year 2000. Developments are still on-going.

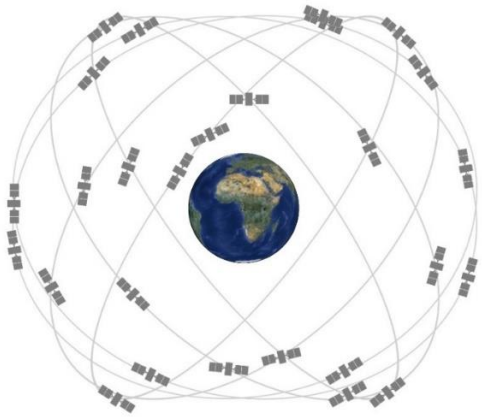

Figure 1. Graph of the six orbital planes of the GPS constellation [1].

For each GNSS constellation there is a frequency-band assignment. This process can be complex as there may be services operating in the same frequency ranges. Depending on the country, the same frequency bands can be used for different purposes. There is an institution in charge of coordinating this issue at a global level called ITU. Figure 2 shows the different frequency bands in which the most important constellations work. This figure shows the two frequency bands marked ARNS (Aeronautical Radio Navigation Service) for GNSS signals; and RNSS (Radio navigation Satellite Service), which is used for terrestrial services.

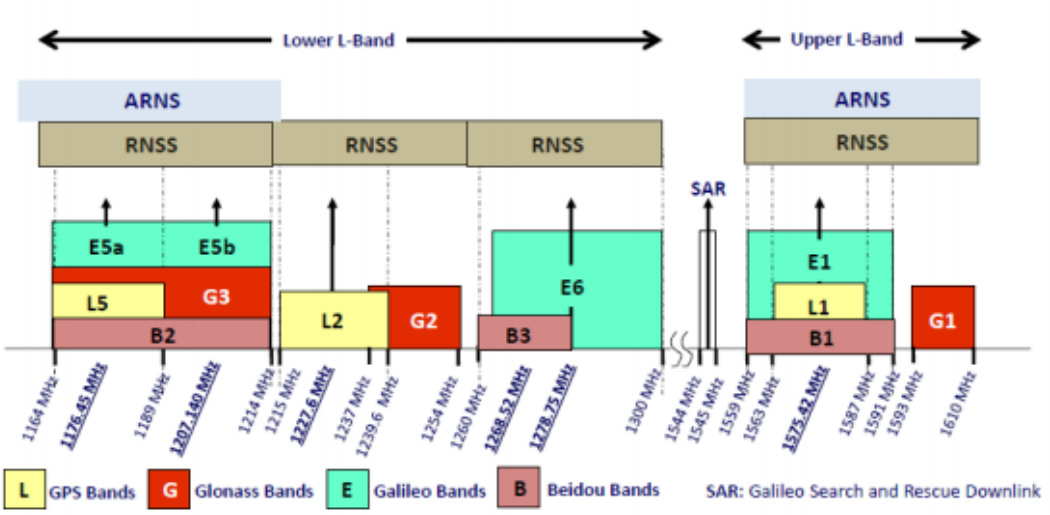

Figure 2. Frequency bands of GNSS most important constellations [2].

Each GNSS signal is composed of:

- **Carrier Frequency**: It is a radio frequency sinusoidal signal whose function is to carry information on a given frequency through space.

- **Ranging Code**: It is a binary code, called pseudorandom noise codes (PRN). PRN codes work as an identifier for each satellite by using Code Division Multiple Access (CDMA) technique. These codes must be known by the receiver in order to decrypt the message that the signal contains.

- **Navigation Data**: It contains information about the satellites such as the position of the satellite at a given time, which is called ephemeris, condition of the satellite, a reduced-precision ephemeris called almanac, and the satellite clock bias.

The stages that a receiver performs in order to ultimately obtain its position are: acquisition, tracking, navigation data extraction, and position resolution. The purpose of the first stage is to identify the visible satellites from the position where the receiver is located. The signal from the satellites is received, the carrier frequency is removed and then a correlation between the received signal and the individual codes (PRN sequences) of the different satellites is performed. After this, it must be determined if a satellite has been acquired or not by comparing if the higher value of the correlation obtained is higher than a certain chosen threshold. If the satellite is considered acquired, a rough estimation of parameters such a Doppler frequency or code delay is performed. Once this acquisition stage is finished, it is followed by the tracking stage that refines the previously estimated parameters to keep track of these as the signal changes over time [3]. After this, when enough data of the navigation message has been received, and the track of the satellites is stable enough, the position of the receiver can be determined by combining the information received from at least four satellites (three satellites to solve the x, y and z coordinates, and one more to solve the time unknown). The mechanism is called trilateration where a range called pseudo range $\rho^{(i)}$ is measured from i-th satellite and it can be expressed by

$$\rho^{(i)} = \sqrt{(x^{(i)} - x_u)^2 + (y^{(i)} - y_u)^2 + (z^{(i)} - z_u)^2} + c * \tau_u + \gamma$$

where $x^{(i)}, y^{(i)}, z^{(i)}$ stands for the satellite position and x_u, y_u, z_u for the receiver position. Additionally, c stands for the speed of light, τ_u for the receiver clock bias (due to relativity effects and to a less stable receiver clock compared to the satellites' clocks) and γ for the sum of errors during the transmission such as atmosphere effects, interferences or background noise. Using pseudo ranges from at least four satellites we end up with a set of non-linear equations which can be solved using closed form solutions (e.g. Least Squares), iterative techniques based on linearization (e.g. Iterative Least Squares) or various types of Kalman filters (e.g. Kalman extended filter). Lastly, position of the receiver is given in Cartesian coordinates that are then transformed into geodetic coordinates specifying its latitude, longitude, and altitude [4].

3. GNSS INTERFERENCES

Signal interference represents one of the most notorious weaknesses of GNSS. One of the reasons why GNSS is so vulnerable to interferences is because of the low power the signal is received (approximately around -130 dBm). As it can be seen from the Table 1, there are two main groups that include the different types of interferences, and they can be distinguished by the source of the interference: i) artificially produced interference by wireless transmitters and ii) interference produced due to the wireless channel effects.

Within the interference produced in an artificial way, there are two different types, called unintentional (e.g. adjacent channel interferences due to harmonics of systems transmitting in nearby frequency bands to GNSS bands) and intentional interference (e.g. spoofing or jamming). In addition, intentional interference is formed by two categories named: adjacent channel and co-channel interferences. Adjacent channel interferences take place when, during a RF transmission, an amount of power is transmitted to those channels which are adjacent to the assigned channel in the transmission (e.g. intermodulation products). Differently, co-channel interferences are produced when two transmitters use the same channel at the same time, therefore creating an interference (e.g. cross-talk).

Table 1. Interference level classification.

Interferences						
Artificially Produced				**Channel-based**		
Intentional		Unintentional		Space Weather	Multipaths	Other
Jamming	Spoofing	Adjacent Channel	Co-channel			

The other interference category is the one based on natural interferences over the wireless channels. Multipath-type interference can be classified into this category and it takes place when one or more paths are received from the same antenna, e.g, due to reflections, refractions or scattering on obstacles. As *Space weather* type, two ionospheric effects should be considered: fast and large ionospheric changes and ionospheric scintillation. The first one can cause changes in telemetry and they occur because large changes in the ionosphere occur near the geomagnetic equator. The second, in contrast, occurs in the equatorial regions and may cause momentary loss of the signal emitted by one or several satellites. Finally, as *Other* type of channel-based interference, we can include those produced by Doppler shifts, fading, and shadowing effects.

Having briefly explained the different types of interference, it is important to point out that this thesis focus on identifying intentional interference such as those mentioned before, namely Jamming and Spoofing.
The first one, called Jamming is typically a synonym for intentional (narrowband) interference, which is the deliberate radiation of the electromagnetic signals at GNSS frequencies. The aim is to overpower the extremely weak GNSS signals so that they cannot be acquired and tracked anymore by the GNSS receiver. There are two main types of jamming: military jammers that are used in part of the military strategies, disabling civil GNSS, and Privacy Protection Devices (PPD)s. PPDs are a type of device, with very small dimensions and despite being banned in most countries, they are easily available via online for only a few euros.

The second intentional interference type, the Spoofing type of interference, is a more complex attack compared to Jamming. It consists in the transmission of a fake GNSS-like signal with the intention of fooling a GNSS receiver into providing false position, velocity, and time. In the most common examples, an attacker would position a broadcast antenna and point it at the target's GNSS receiver antenna in order to interfere with signals of proximate buildings, ships, or aircraft. To execute these attacks, attackers take advantage of the fact that the structure of civil GNSS signals is publicly known at the processing level, and therefore it can be replicated by a (cheap) GNSS receiver. GNSS spoofing can be accomplished with cheap and portable software-defined radio running open source software or with more powerful and expensive transmitters for wide-scale attacks. The awareness about this type of intentional interference dates among 2001 and 2003, and recently, a research group from United states called C4ADS published a report detailing nearly 10,000 instances in which Russia interfered with satellite navigation of more than 1,300 civilian vessels in ten different locations around Russia, Ukraine, and Syria. Other organizations have also reported widespread examples of interference with GNSS signals [5].

4. RADIO FREQUENCY FINGERPRINTING

RF (Radio frequency) fingerprinting is a process that identifies the device from which a radio transmission was originated by looking at the properties or features of its transmission [6]. Like people, each radio transmitter has unique fingerprints called RF fingerprints, which are based on its location and configuration. It can be said that the features that can be found on each transmitter are different from each other due to differences in hardware between different transmitters. The main components that are part of the structure of the radio signal transmitter and therefore cause different fingerprints between devices are e.g. filters, amplifiers, and oscillators. These differences on the electronic components are randomly generated and usually due to imperfections in the material of the component itself [7]. These imperfections cause a number of signal effects such as I/Q imbalance, phase imbalance, frequency error or signal strength, which combined form the fingerprint of the transmitter.

This RF fingerprinting technique is very interesting for authentication of GNSS signals as it is based on the physical layer, which is very difficult or impossible to replicate. It can provide a superior performance than traditional higher-layer encryption solutions [8,9]. One of the problems of using RF fingerprinting localization is the creation and correct maintenance of a fingerprinting database. This process is slow and complex because it requires a large number of measures of the same device to obtain an adequate statistical value. That is why for the study of the fingerprinting methods of this thesis and its consequent development of a product classifier detector, a large number of signals from different satellites of different frequency bands and different constellations GNSS has been recorded. Once this is done, the features that the classifier will use for the training are extracted. Classification algorithms such as SVM (Support Vector Machine) [10] or CNN (Convolutional neural network) [11] are used for this type of classification techniques, where after training with a large number of extracted features, these trained models are able to predict if a certain feature corresponds to a specific type of GNSS signal.

5. FINGERPRINTING EXTRACTION METHODS

Different features can be obtained from RF fingerprinting such as I/Q imbalance, phase imbalance or other hardware-related features. Such features can be extracted by using various transforms of the received I/Q signal. The transforms may reveal the behavior of the data in time and frequency, which may allow one to obtain many useful features for machine learning classification methods. Wavelet transform also allows representations of functions in which are retained both the scale as space information. Many functions can be approximated with great accuracy using only a small number of wavelet coefficients. In the case of thesis, this type of transformation serves to decompose the signal and identify the features for its later extraction. There are different types of transforms suitable for this case, such as the Continuous Wavelet Transform (CWT) and Discrete Wavelet Transforms (DWT) [12]. DWT will be used as based on some preliminary studies they proved to be adequate for the specific case of GNSS RF fingerprinting.

The process would be as follows in Figure 3. First the signals are collected for study, converted into a digital signal through the Analog-to-Digital (ADC) block and the most important characteristics are extracted through a feature extractor transform. Once these characteristics are extracted, two different processes are distinguished. The first would be the construction of a training database where the fingerprints reside (i.e., 'fingerprint creator' block), and the second is the evaluation of the new fingerprints to check if they match those from the database (i.e., 'fingerprint matcher' block).

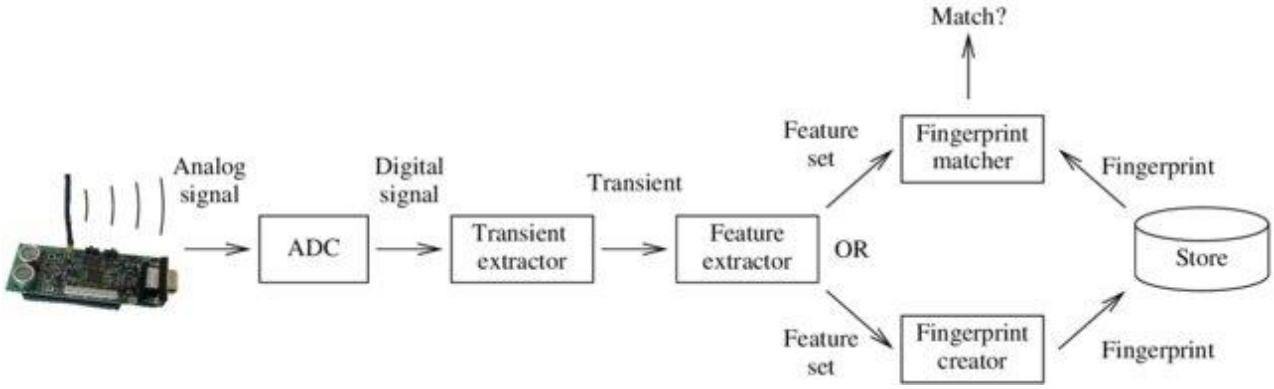

Figure 3. Radio fingerprinting process [13].

In order to classify extracted features, there are different classification algorithms, for example Support Vector Machine (SVM) or Convolutional Neural Network (CNN). In SVM, there are two stages: training and classification. Classification is done by first drawing hyperplanes as it can be seen in Figure 4 using training data. The stablished hyper-planes will split the different classes to classify. The hyper-plane with the maximum distance between Support Vectors, which are the closest data points to the hyper-plane, is the hyper-plane used to classify, called optimal hyperplane. It can be observed that a margin is now defined as being the distance between hyperplanes and the closest support vector. The goal is to obtain the largest margin distance, therefore minimizing the possibility of error. In the classification stage, the characteristics of the new data will be then used to predict the group to which the new input belongs.

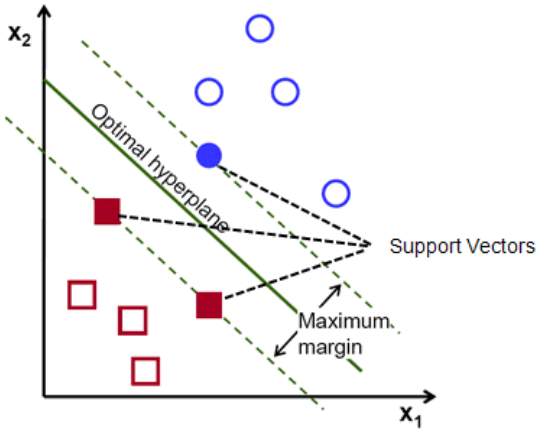

Figure 4. SVM binary classification [14].

Classification is simple when dealing with two-dimensional cases, but the algorithm must deal with more difficult cases according to real cases. These cases can be: Dealing with more than two predictor variables, non-linear separation curves, chaos where data sets cannot be completely separated, or classifications into more than two categories. For this type of cases, the SVM uses the Kernel functions, mapping the input space to a new space of characteristics of greater dimensionality. Most used types of Kernel Functions are enumerated below:

- **Polynomial**. Usually used in image processing. Equation is:
  $k(x_i, x_j) = (x_i \cdot x_j)^n$

- **Gaussian.** It is a general-purpose kernel; used when there is no prior knowledge about the data. Equation is:
  $k(x_i, x_j) = \exp(-\gamma||x_i \cdot x_j||^n)$

- **Sigmoid.** It can be use it as the proxy for neural networks. Equation is: k(x,y) = $tanh(\alpha x^T y + c)$

Kernel types are different when making the hyperplane decision boundary between the classes. Normally, linear and polynomial kernels are faster in decision making but provide less accuracy than Gaussian kernels. For that reason and given that Gaussian kernel is the most common type used with SVM classifiers, this type has been chosen for the analysis of the work.

In section 7, there will be a study of the performance of a learning algorithm such as SVM using the I/Q samples that were recorded in the measurements-based setup. A more detailed explanation of the process that is going to be carried out is; once I/Q samples are obtained, methods to obtain signal transforms will be executed by using the MATLAB program in order to set up a database of the RF fingerprinting features. After that, with the help of the learning algorithms and selecting the most suitable type of classification of SVM, unlabeled plots from the transforms will be compared to the database and we will be able to get final accuracy results. In the next chapter, the setup used to obtain the measurements will be explained in detail.

6. MEASUREMENTS-BASED SETUP

We can split the RF fingerprinting implementation into four steps: Signal collection, features extraction, set-up database and training/classification. In order to accomplish these steps, a setup has been designed, which is shown in Figure 5. The process would be as follows: the signals from the antennas and the Spectracom are collected obtaining a *.bin file that later will be used to extract the characteristics of the signal. Then, after extracting those features, a percentage of the extracted characteristics will be used for the training of the machine learning algorithm and the rest will be used for the classification test to check the accuracy of this algorithm.

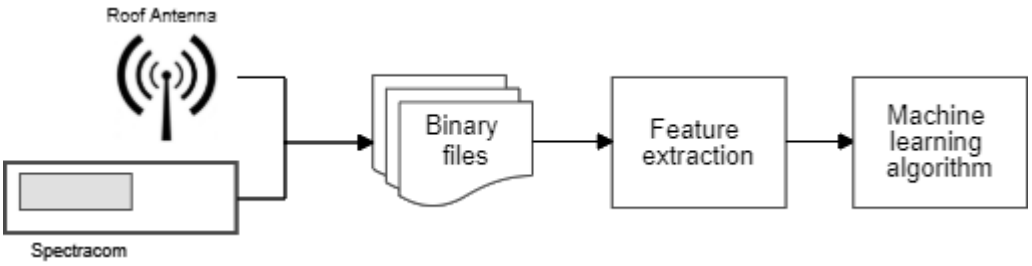

Figure 5. Block diagram showing processing steps

For the signal collection phase, a measurement setup has been designed. First device is a GNSS signal generator called Spectracom, in particular the model Spectracom GSG-6 Series. The Spectracom model used for the thesis is able to collect the signal from any of the following constellations: GPS, GLONASS, Galileo, BeiDou, QZSS and IRNSS. We have from 1 to 64 channels available where one can use any of the GNSS frequency bands L1, L2, L2C, L5, E1, E5, B1 or B2. The output signal power can be selected from -65 to -160 dBm.

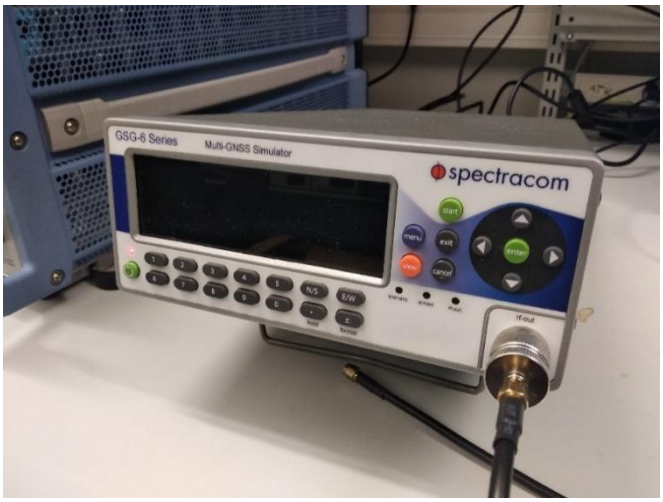

Figure 6. GSG-6 Series Spectracom used in the measurements.

The Spectracom generator has to connect its output signal to a Universal Software Radio Peripheral (USRP) device which will be also connected to the computer. In that way it will be able to record the signal from Spectracom. USRP is a type of software-defined radio (SDR) which can transmit and receive RF signals in several bands. USRPs are connected to a host computer through a high-speed link. A specific software can be used to control the USRP hardware to transmit/receive data. These devices are quite accessible for everyone and can be used from teaching to advanced wireless research, including dynamic spectrum access, whitespace, and PHY- and MAC-layer research.

Figure 7. USRP model 2954R used for recording the signal.

Just as the Spectracom signal is recorded, the same process will be done to capture signal from the GNSS antennas. The antennas are located on the roof of one of the buildings at Tampere University. There are two different antennas. The first antenna model is a Tallysman TW3872 and the second antenna model is a Novatel GPS-703-GGG. Both of them provides triple band functionalities. It means that both antennas are able to record any of the following constellations/frequency bands: GPS L1/L2/L5, GLONASS G1/G2/G3, BeiDou B1/B2, Galileo E1/E5a+b and L-band corrections services. The two antenna models used for the study can be observed in Figure 8 and Figure 9.

Figure 8. Tallysman model antenna.

Figure 9. Novetell model antenna.

Using a Labview interface, as it is shown in Figure 10, different parameters of the recording can be chosen, such as, channels of the USRP that we want to record, IQ rate, carrier frequency, or gain. This Labview interface is used to acquire and record the signal received by the USRP on the selected frequency. In Figure 11 it can be seen the frequency spectrum of channel 0 and channel 1 that it is being recorded. It can be observed how in the channel 0 the spectrum is the typical GNSS spectrum shape because it comes from the Spectracom, in which we have not included any noise. On the contrary, the signal that is observed from channel 1 has much more noise. This is due to the fact that the received power is really low since it comes from the satellites, which are tens of thousands km away. This is the main reason why practically only noise is seen.

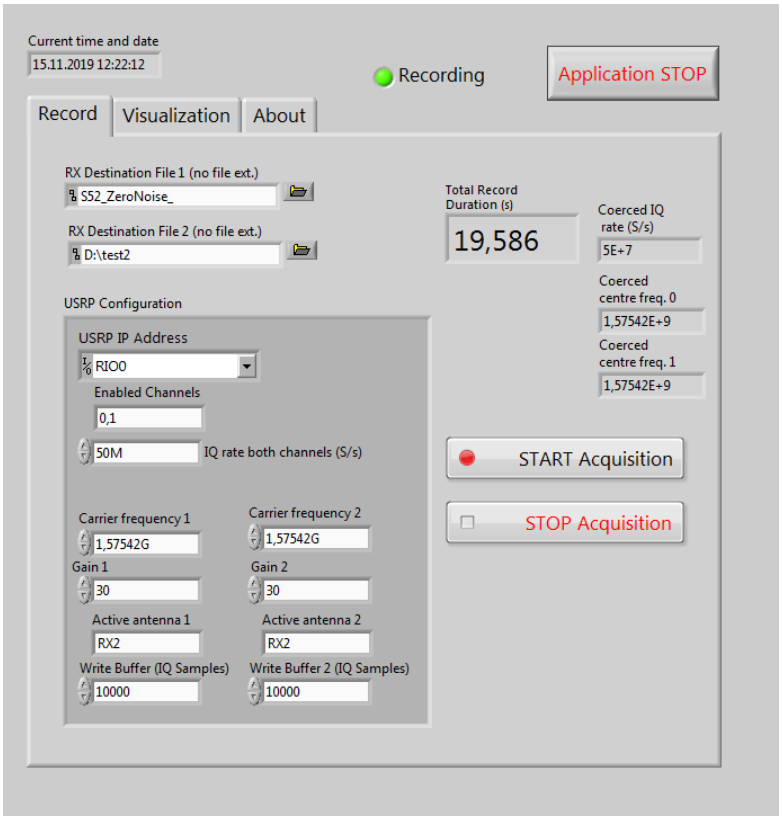

Figure 10. Labview interface.

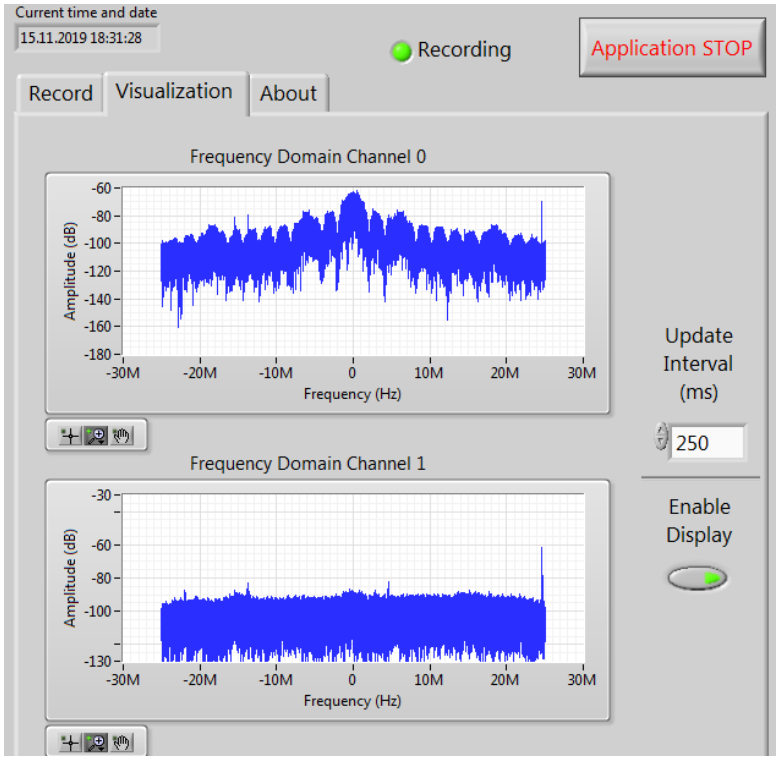

Figure 11. Frequency spectrum of channel 0 and 1 of the Labview interface.

In order to train and evaluate the classifier, a large number of independent recordings were needed to be realized in order to determine statistical patterns of each signal. Moreover, by performing them at different times, we will be able to obtain more reliable statistics of them. Two scenarios were used for the antenna and over a hundred recordings at different times and days were collected from each one. Fifty-eight scenarios were used for the Spectracom recordings as it can be observed in Table 2. Despite the large number of data collected, only scenarios from GPS L1 and GALILEO E1 were finally used in the data analysis due to time limitations of the thesis analysis.

Table 2. List of recorded scenarios, each having been repeated 10 times.

Device	Scenario	Constellation	Number of Satellites
Antenna Tallysman	A1	GPS L1, GALILEO E1, BeiDou B1	Based on the time of recording
	A3	GPS L5, GALILEO E5, BeiDou B2	Based on the time of recording
Antenna Novatell	A2	GPS L1, GALILEO E1, BeiDou B1	Based on the time of recording
	A4	GPS L5, GALILEO E5, BeiDou B2	Based on the time of recording
Spectracom	S1-S9	GPS L1	1 in scenario S1 to S7, 5 in S8, 10 in S9
	S10-S18	GPS L5	1 in scenario S10 to S16, 5 in S17, 10 in S18
	S19-S27	GALILEO E1	1 in scenario S19 to S25, 5 in S26, 10 in S27
	S28-S36	GALILEO E5	1 in scenario S28 to S34, 5 in S35, 10 in S36
	S37-S45	GLONASS G1	1 in scenario S37 to S43, 5 in S44, 10 in S45
	S46-S54	BeiDou B1	1 in scenario S46 to S54, 5 in S55, 10 in S56
	S55-S56	GPS L1, GALILEO E1, BeiDou B1	1 in scenario S55, 5 in S56
	S57-S58	GPS L1, GALILEO E1, BeiDou B1, GLONASS G1	1 in scenario S57, 5 in S58

The parameters used in the USRP during the recordings were:

- Sampling frequency: 50MSamples/s
- Quantization bits: 16 bits
- Time interval for data collected per each scenario: 20 seconds
- Intermediate Frequency (IF): 0 Hz
- Antenna Gain: 30dBm

The parameters used in the Spectracom during the GNSS signal generation:

- Transmit Power: -70dBm
- CN0: No-Noise (i.e., infinite CN0)
- Movement: Static receiver

7. MEASUREMENT ANALYSIS

After recording a comprehensive number of signals, an analysis was performed. Before using the recorded signals, the recordings were checked to verify if they included the satellites that we were expecting. For that purpose, we use a MATLAB-based GNSS Software Defined Receiver (SDR). The receiver is able to acquire and track the GNSS signals in order to determine which satellites are present and track them in order to provide a Position Velocity and Time (PVT) solution. During the acquisition, in order to determine if a certain satellite was present or not a certain threshold was set (in our case to 1.5). As illustrated in Figure 12 and Figure 13, only when the acquired metric is higher than this threshold, the Space Vehicle (SV) is considered acquired (it is present in the signal).

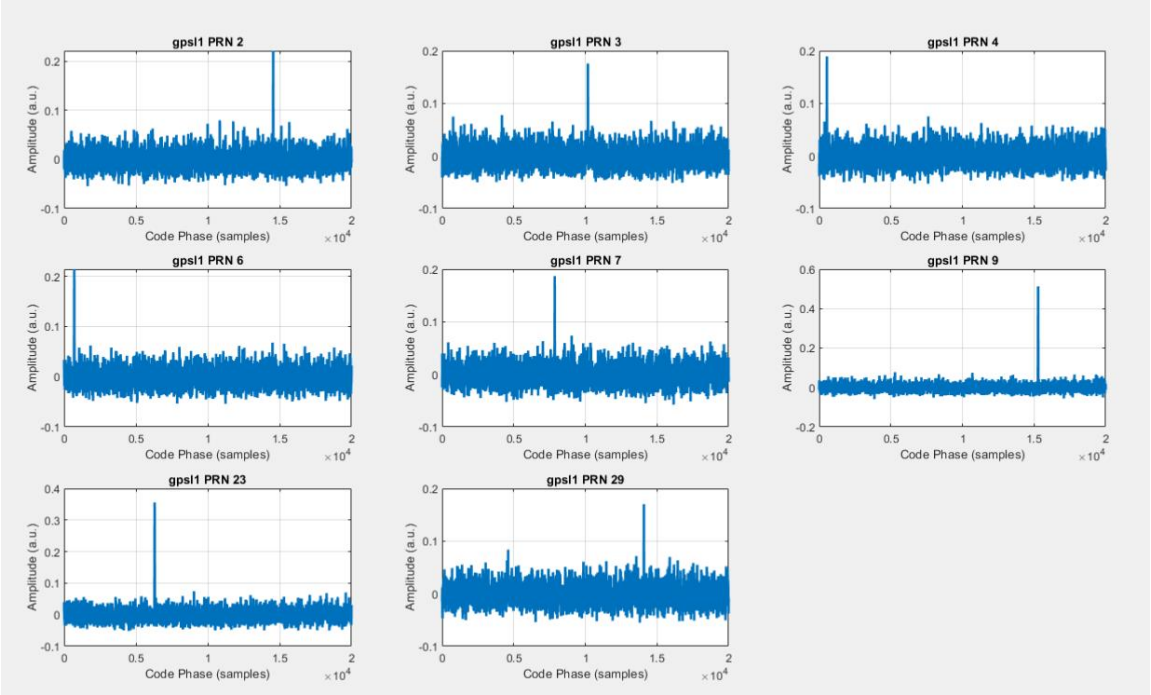

Figure 12. PRN number of acquired satellites from GPS constellation.

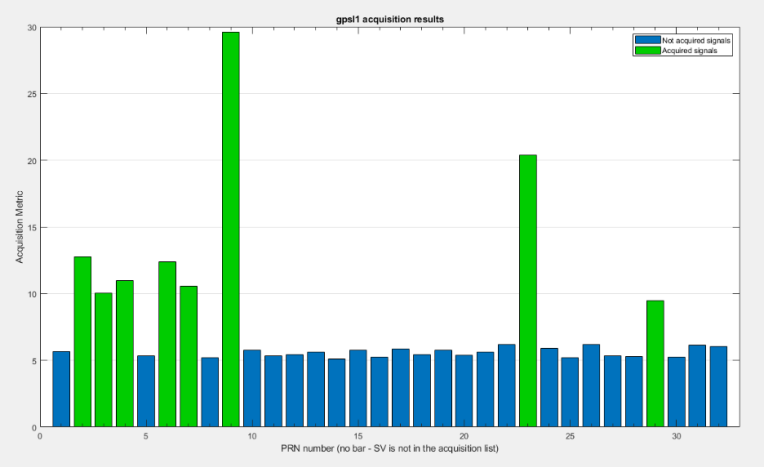

Figure 13. Acquisition plot results from GNSS Software receiver

After this process we have to make sure that the acquisition is correct, we use a website called GNSSplanning [15], provided by Trimble, that tells us the visible satellites from every constellation in any day or time and. Then we finally compare the satellites that the MATLAB receiver found and the ones that are supposed to be visible. This process was only undertaken for the signals collected from the antennas, as they are real signals from satellites and it is important to verify the satellites included in it. In the case of the signals collected by the Spectracom, it is not strictly necessary since it is assumed to be error free, but it would also be interesting since the Spectracom can replicate the signal at any time.

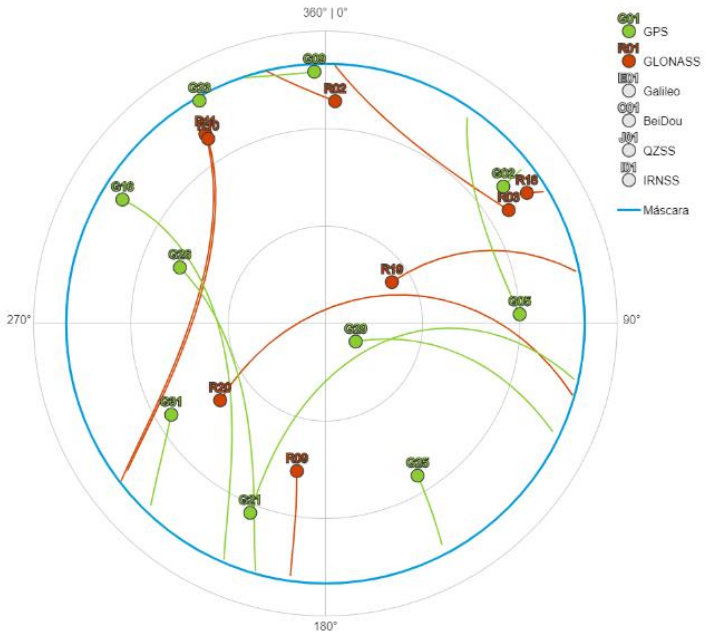

Figure 14. Visible satellites on November 26 at 00:00 located in Tampere using GNSSplanning service.

After being sure that the recorded signals are correct, the extraction of fingerprinting features procedure is conducted. For this purpose, we used a MATLAB script. The main procedure is that it reads the binary files in which the recordings were saved. Then it reads the number of milliseconds that we choose from it and we get the IQ (phase and quadrature components, which corresponds to real and complex part of the recorded signal) data from each millisecond. After that, we performed the Discrete Wavelet Transform for each 1 ms IQ data and we plot the transform (the imaginary versus the real part of the DWT, see an example in Figure 16). We saved the images in two different directories, one containing images that will be used during the training of the Machine Learning (ML) algorithm and another containing the images that will be used during testing. We set 80% of the images to be saved in the training folder and 20% to the test folder. So as we decided to read 15000 milliseconds, 12000 will go for training and the remaining 3000 will go for testing the SVM algorithm.

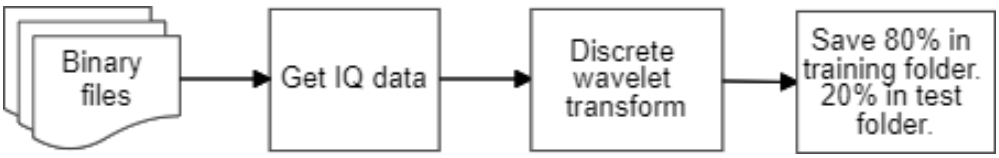

Figure 15. Block diagram showing analysis steps.

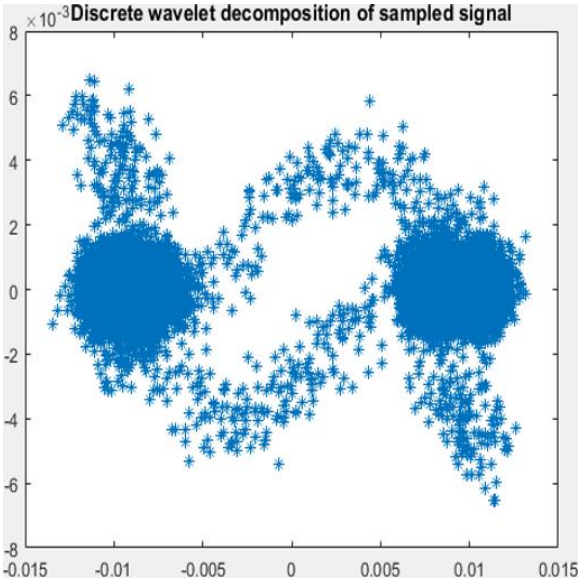

Figure 16. Image of discrete wavelet transform of one millisecond from GPS L1 with one satellite recording.

The classification was performed with GPS L1 signal when only one and ten satellites were present and with Galileo E1 as well with 1 satellite and with 10. These signals will be directly compared to the ones from the Novatel antenna at the same band frequency (1.57542 GHz). The purpose is to use SVM algorithm that will be executed in MATLAB and to introduce the training samples first so the algorithm classifies between signals from the Spectracom and from the antennas. Then the algorithm uses unlabeled test samples to classify them, after that we evaluate how many of the testing samples were right classified and which weren't.

The SVM algorithm implementation works as follows:

1. Path definition of the testing and training datasets.
2. As SVM is a supervised method, it must be informed manually, that an image from a specific type belongs to that category. Then, the bag of features is used to extract features from the image in order to have information to classify.
3. These features are used to train the SVM model, which will learn to associate the different features extracted to the specific label type.
4. After the model is trained, testing data is used to evaluate the classifier accuracy performance.
5. Finally, it represents a confusion matrix showing the results and accuracy of the classifier for the used data. Being the y axis, the different categories in the analysis, and x axis, the number of times that the algorithm classified test data in any of the available categories. I they match it indicates that the accuracy is good.

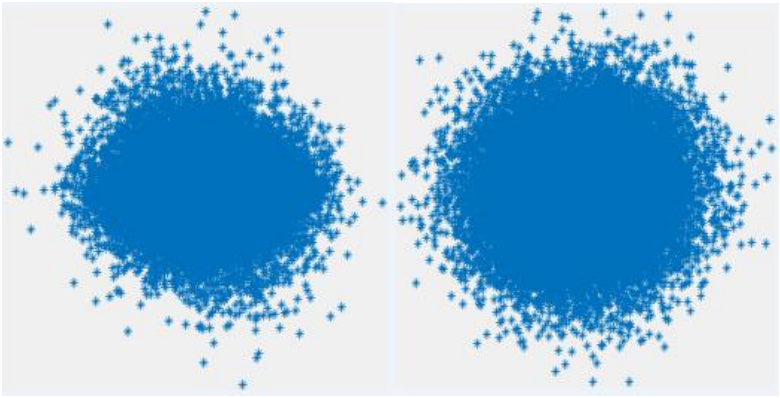

Figure 17. Image of discrete wavelet transform from GalileoE1 with ten satellites on the left and discrete wavelet transform from antenna on the right. Showing the good performance of the SVM, despite being a complicated task when evaluating a signal with 10 satellites, due to its similarity with the one coming from the antenna.

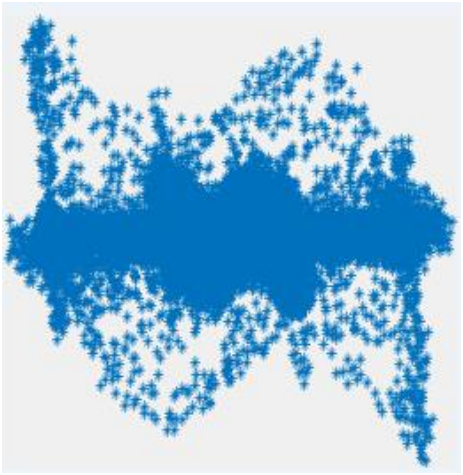

Figure 18. Image of discrete wavelet transform from GalileoE1 with one satellite.

Finally, a matrix is created with the number of times the algorithm has placed a type of category in the available categories and thus determine the accuracy of each one and the total of the evaluation if it has placed it correctly. Here we have final accuracy values from the SVM algorithm for different scenarios:

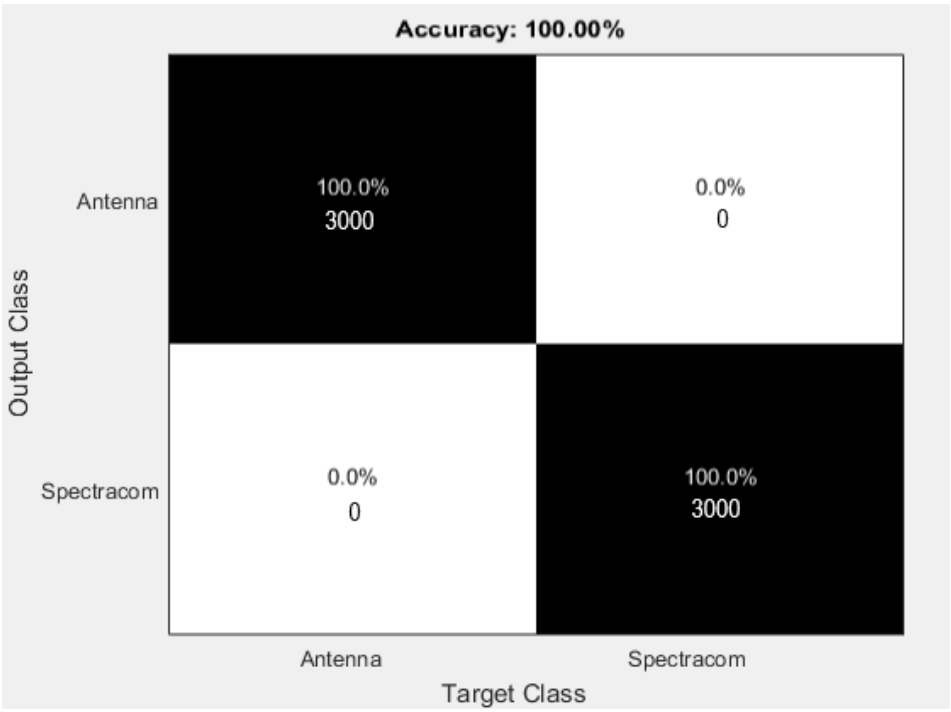

Figure 19. Accuracy result between GPS L1 and Antenna recordings.

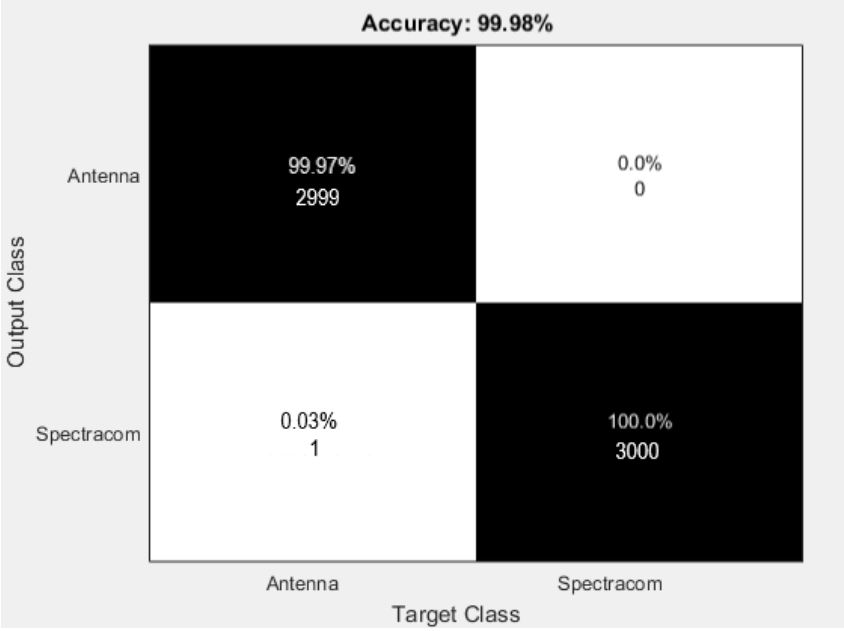

Figure 20. Accuracy result between GPS L1 with 10 satellites and Antenna recordings.

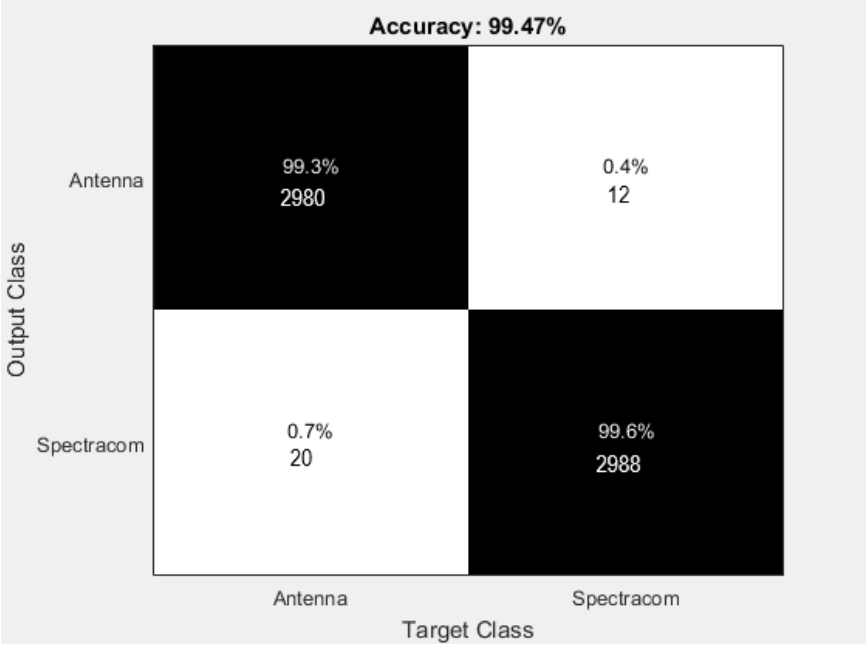

Figure 21. Accuracy result between Galileo E1 with 10 satellites and Antenna recordings.

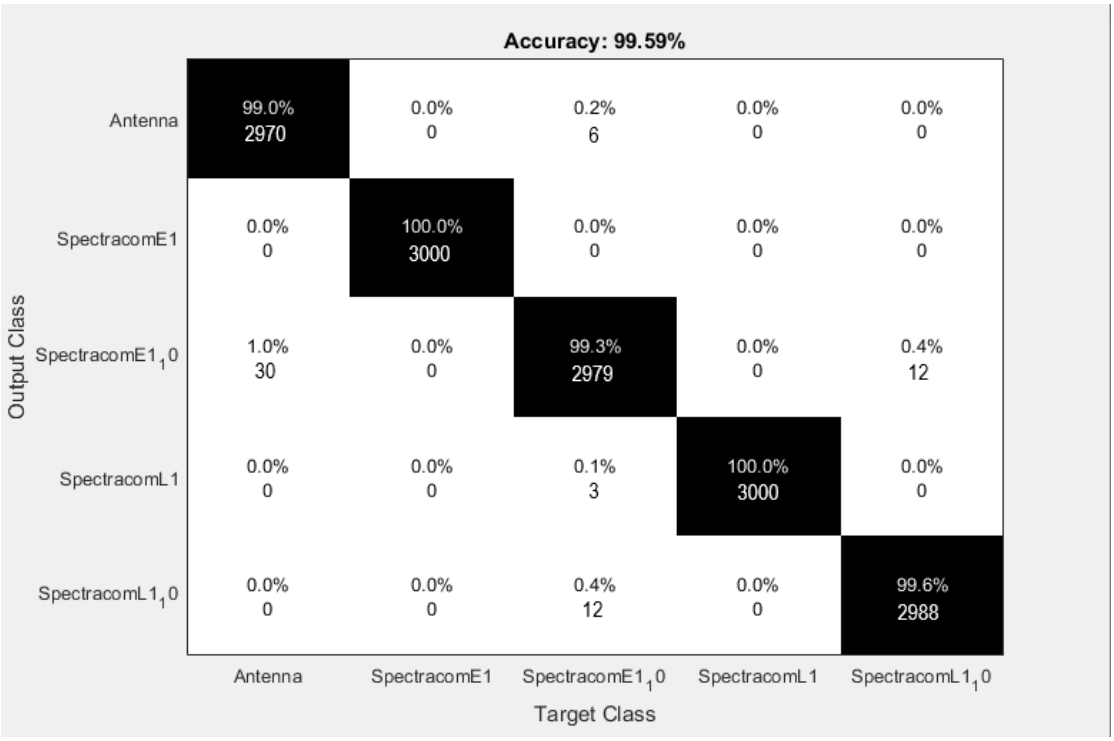

Figure 22. Five classes comparison.

The results obtained are very accurate when classifying all scenarios. This high accuracy is to a certain degree due to the fact that signals received from Spectracom are being captured without noise, which makes them recognisable to extract the features in comparation to those coming from the antennas. Binomial training is quicker compared to working with several categories, since the SVM algorithm uses less information for training and testing. This also usually results in a higher accuracy of the algorithm.

8. CONCLUSIONS

After obtaining the results using different scenarios, it can be highlighted that the accuracy has reached more than 99% in all cases. This shows that SVM was a good choice for the classification of signal features from a real antenna and signals generated by a GNSS signal generator. It is also important to mention that the signal coming from the Spectracom does not include noise or channel effects, that is why it is easier to obtain better the features when performing the wavelet transform, and therefore, to obtain a better result by the SVM algorithm than what one would expect in the presence of noises and multipath. This does not happen with the signals coming from the antenna, which includes noise. Another conclusion that we can emphasize is that when it is intended to classify a GNSS signal generated by the Spectracom configured with more than one satellite, the percentage of accuracy is lower compared to when it has only one satellite. This happens because the image of the spectrum generated by the discrete transformation of the signal with a satellite is easier to distinguish with the first glance, than the one generated with more satellites, which is more similar to the one generated by the antenna. Another highlight of the study is the image collection process, which takes a considerably high time, approximately 5 hours for a binary study of 15000 samples of each category. It is fair to mention that this time used is only necessary for training and testing the algorithm, once this is done, the images can be classified online as soon as raw data is received. On the other hand, execution time of the SVM algorithm in the MATLAB program is considerably shorter, approximately 1.5 hours for a binary study. One solution to reduce this time is to change parameters such as the window size or the number of features that can be extracted per image. However, it could decrease algorithm's performance.

The main conclusions for the thesis could be summarized as:

- SVM classification algorithm demonstrated to be a very effective model for the extracted radio fingerprinting features for identifying GNSS transmitters.
- When a binary classification is made, the SVM algorithm obtains practically perfect accuracy data, but when it comes to a classification of several categories, the accuracy is slightly reduced, although still rather high.

- The accuracy of the classifier is nearly perfect when comparing a signal generated with the Spectracom configured with a single satellite with

the signal from the antenna. But this accuracy is reduced when the signal is configured with more satellites. In our case the comparison was made with a signal containing 10 satellites, both for GPS and Galileo.

Part of our measured data has been uploaded on Zenodo open-access repository under CC BY 4.0 license [16].

9. FUTURE WORK

Once the development and initial validation of the classifier is finished, and verified its performance using RF fingerprinting feature extraction methods succeeded, there are still several aspects that must be worked on in the future. The first would be to check the precision data with more GNSS systems, such as GLONASS and Beidou, which were also recorded but, due to lack of time, the analysis could only be performed on GPS L1 and E1. Despite obtaining satisfactory results with the GNSS GPS and Galileo systems, other systems could cause more difficulties in classifying them. Especially considering that different GNSS constellations transmit at the same frequency and at the same time. Another future work would be to use signals coming from more combinations of GNSS systems with other configurations, for example with different number of satellites and also different satellites in each configuration. By doing this we would obtain more reliable and comprehensive results. In addition to these additional tests, it would also be helpful to use a larger number of samples for each scenario. In these results 15000 ms were used; a future objective would be to use much more data. The problem of increasing the amount of data used for training is that this will increase also the time needed for the generating the training images and to train the SVM algorithm model. This increment of time could also be a future work for the project, trying to compensate it by using different parameters such as lower resolution images, reduced window size for analyzing the image features or reduce the maximum number of features to extract.

Once all this type of new testing of the classifier product has been done using the recorded scenarios, we would proceed to the most important future work, this would be to use real Spoofing signals generated to fool the GNSS receiver. Testing the product with this type of signal would verify the true potential of the classifying algorithm. Finally, once the necessary checks are made and the verification that the product is optimal for this type of signals, a last future job would be to select a minimum training time for the execution of the SVM algorithm where we would obtain reasonable precision data, with the saving of processing time that this would imply.

REFERENCES

[1] Offical U.S. government information about the Global Positioning System (GPS) and related topics. Available: https://www.gps.gov/systems/gps/space/.

[2] Sanz Subirana J.; Juan Zornoza J.M.; Hernández-Pajares M. GNSS DATA PROCESSING. *Volume I: Fundamentals and Algorithms, European Space Agency*, 2013.

[3] Borre K.; Akos D. M.; Barteselen N.; Rinder P.; Soren Holdt J. A Software-Defined GPS and Galileo Receiver (ISBN 0-8176-4390-7). P. 70-72.

[4] Morales-Ferre, R.; Richter, P.; Falleti, E.; de la Fuente, A.; Lohan, E.S. A survey on coping with intentional interference in satellite navigation for manned and unmanned aircraft. *IEEE COMMUNICATIONS SURVEYS AND TUTORIALS*.

[5] C4ADS. (2019). Above us only starts: Exposing GPS Spoofing in Russia and Syria. *C4ADS Innovation for peace*. Texas: USA.

[6] Sankhe, K.; Belgiovine, M.; Zhou, F.; Angioloni, L.; Restuccia, F.; D'Oro, S.; Melodia, T.; Ioannidis, S.; Chowdhury, K. No Radio Left Behind: Radio Fingerprinting Through Deep Learning of Physical-Layer Hardware Impairments. *IEEE Transactions on Cognitive Communications and Networking* 2019, pp. 1–1.254 doi:10.1109/TCCN.2019.2949308

[7] Baldini G.; Steri G.; Giuliani R. (2017). *Identification of wire-less devices from their physical layer radio-frequency finger-prints*. Encyclopedia of Information Science and Technology p. 6136-6146.

[8] Shi, Y.; Jensen, M. A. (2011). Improved Radiometric Identification of Wireless Devices Using MIMO Transmission. *IEEE Transactions on Information Forensics and Security*, 6(4), 1346–1354. doi:10.1109/tifs.2011.2162949

[9] Brik, Vladimir & Banerjee, Suman & Gruteser, Marco & Oh, Sangho. (2008). Wireless device identification with radiometric signatures. *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*. 116-127. 10.1145/1409944.1409959.

[10] Ma, Y.; Guo, G. (2014). Support Vector Machines Applications. Springer International Publishing.

[11] Wang, L.; Zang, J.; Zhang, Q.; Niu, Z.; Hua, G.; Zheng, N. (2018-06-21). "Action Recognition by an Attention-Aware Temporal Weighted Convolutional Neural Network".

[12] Klein, R. W.; Temple, M. A.; Mendenhall, M. J. (2009). Application of wavelet-based RF fingerprinting to enhance wireless network security. *Journal of Communications and Networks*, 11(6), 544–555. doi:10.1109/jcn.2009.6388408.

[13] Rasmussen, Kasper & Capkun, Srdjan. (2007). Implications of radio fingerprinting on the security of sensor networks*. Proceedings of the 3rd International Conference on Security and Privacy in Communication Networks, SecureComm*. 331 - 340.

[14] Low, R: (2019). Kaggle: credit risk (Model: Support Vector Machines). New York, USA. Phytonic Finance. Retrieved from: https://randlow.github.io/posts/machine-learning/kaggle-home-loan-credit-risk-model-svm/.

[15] An open source Global Navigation Satellite Systems software-defined receiver. utrlhttps://gnss-sdr.org/. TRIMBLE GNSS Planning Online. https://www.gnssplanning.com/. Accessed: 2019-11-26

[16] R. Morales Ferre; W, Wang; A. Sanz Abia; E. S. Lohan, “Identifying GNSS transmitters based on their RadioFrequency (RF) features - a dataset with GNSS roofantenna and Spectracom-based GNSS signals, 10.5281/zenodo.3588392.