EM Monitoring and classification of IEMI and protocol-based attacks on IEEE 802.11n communication networks

Jonathan Villain, Virginie Deniau, Anthony Fleury, Eric Pierre Simon, Christophe Gransart, Raouf Kousri

EM Monitoring and classification of IEMI and protocol-based attacks on IEEE 802.11n communication networks

Jonathan Villain[1], Virginie Deniau[2,1], Anthony Fleury[3,1], *Member, IEEE,*, Eric Pierre Simon[4,1], Christophe Gransart[2,1] and Raouf Kousri[1],

***Abstract*—The development of connected devices and their daily use are today at the origin of the omnipresence of Wi-Fi wireless networks. However, these Wi-Fi networks are often vulnerable, and can be used by malicious people to disturb services, intercept sensitive data or to gain access to system. In railways, trains are now equipped with wireless communication systems for operational purposes or for passenger services. In both cases, defense strategies have to be developed to prevent misuses of the networks. The first objective of this study is to propose a monitoring solution, which is independent of the communication networks, to detect the occurrence of attacks. The second objective is to develop a method able to classify attacks of different types: the intentional electromagnetic interference (IEMI), i.e., jamming attacks, and the protocol-based attacks. This study focuses on the IEEE 802.11n Wi-Fi protocol. To perform these analyses, we propose to monitor and to analyze electromagnetic (EM) signals received by a monitoring antenna and a receiver collecting EM spectra. After that, we build a classification protocol following two steps: the first consists in the construction of a Support Vector Machine (SVM) classification model using the collected spectra and the second step uses this SVM model to predict the class of the attack (if any). A time-based correction of this prediction using the nearest neighbors is also included in this second step.**

***Index Terms*—IEMI, Intentional ElectroMagnetic Interference, Classification, Wlan, Wi-Fi, communication network journal, IEEE 802.11n.**

I. Introduction

Communications based on radio wave propagation, that cannot be confined and emit in all directions, can be victim of various cyberattacks. The main consequence of this "wild propagation" of radio waves is that unauthorized persons may listen to the network communications and possibly from outside a building. The risks of poor protection of a wireless network are multiple: data interception, diversion of connection for illicit access to a local network, jamming signals or dummy commands for the denials of service etc. Different approaches can be studied to protect these wireless communications from such attacks. In this article, we consider that the first step to countermeasure attacks consists in detecting them and recognizing the type of attack in order to adapt the action to envolve.

Intrusion Detection Systems (IDS) [1] can detect an abnormal activity on an analyzed target. There are three major IDS families. The Network Intrusion Detection System (NIDS) [2] which monitors the security state at the network level, the HostBased Intrusion Detection System (HIDS) [3] which monitors the security state at the host level and an hybrid IDS which combines NIDS and HIDS. The major difference between NIDS and HIDS is that HIDS is particularly effective in determining whether a host is contaminated whereas a NIDS can monitor an entire network. However, IDS mainly work on the upper layers of the OSI model and do not protect the wireless communication links. Moreover, the intrusion detection principles used need to be deployed on all terminals. In this study, we work on a solution which outsources the attack detection function by analyzing the wireless electromagnetic (EM) activity. It consists in taking data from antennas and receivers perfectly independent from the protected communication networks, then applying classification algorithms on the data.

In [4], the authors have already considered improving IDS approaches through EM measurements with the same objective of outsourcing the monitoring. However, the context was different as their work dealt with the analysis of processor EM emissions with the view to detecting software compromise in the context of protecting industrial control systems.

In [5], the detection of jamming attacks on wireless link networks was studied by analyzing the Received Signal Strength Indication (RSSI) received by a station. The RSSI is used by the IEEE 802.11 standard to measure the relative quality of the received signal. In [6], the detection is based on a synchronization indicator together with an adaptive signal to noise plus jammer power ratio. In our case, as the monitoring solution is outsourced, we are not limited to the indicators of the standard. We selected the EM spectra collected by an independent bench, since they provide more information than the RSSI for the monitoring of the physical link.

Our article focuses on the Wi-Fi system. Nowadays Wi-Fi is not only used to access to internet. There are growing numbers of applications using Wi-Fi, including critical applications in

Manuscript received February 22, 2019.

* Designates the corresponding author: Jonathan Villain, PhD, IRT Railenium, jonathan.villain@railenium.eu

[1] Railenium Test and Research Center, F-59540 Valenciennes, France. e-mail: jonathan.villain@railenium.eu and raouf.kousri@railenium.eu

[2] IFSTTAR, French Institute of Science and Technology for Transport, Development and Networks, Villeneuve dAscq, 59650 France. e-mails: virginie.deniau@ifsttar.fr christophe.gransart@ifsttar.fr

[3] IMT Lille Douai, Univ. Lille, Unité de Recherche Informatique et Automatique (URIA), F-59000 Lille, France. e-mail: anthony.fleury@imt-lille-douai.fr

[4] IEMN lab, TELICE group, University of Lille, France. e-mail: eric.simon@univ-lille.fr

terms of security. For instance, in the railway sector, Wi-Fi is increasingly used to ease the maintenance. Certain trains are now equipped with on-board systems that provide maintenance checks and report to a center via Wi-Fi transmissions. Hence, a monitoring system with attack detection functions can help strengthen the Wi-Fi network when used for critical applications.

[7] recently analyzed the impact of jamming attacks on the performance of a Wi-Fi 802.11n transmission but it did not consider the detection and the classification of attacks. In [8] the authors studied different EM interference sources but they did not consider attacks. However, in [8], the interference signals were classified, using a combination of SVM binary classifications to determine if a channel is free from the interference source, if a microwave oven is active during the sensing period or if another network is overlapping the channel.

In our article, the considered attack scenarios correspond to attackers who would use jammers to provoke denial of services and/or who would send deauthentication frames on a public access point. The attack by deauthentication frames is generally applied to disconnect a station from a licit access point in order to benefit from all of the Wi-Fi ressource or to force a station to connect to an illicit access point in order to intercept private data. The attacks by deauthentication frames correspond to protocol-based attacks. In this work, we want to develop a single approach able to detect both jamming attacks and protocol-based attacks, and to distinguish them.

To achieve this goal, the proposed system first collects EM spectra of the frequency band of interest via an antenna. Then, the classification based on these extracted data from the spectra is carried out. Moreover, we propose the following two steps based on classification protocol: the first step uses the Support Vector Machine (SVM) for the classification, and the second step applies a time correction using the nearest neighbors. This time correction takes advantage of the fact that the attacks range on several spectra. In this way, both dimensions, the frequency dimension via the spectra, and the time dimension via the corrections are exploited. To sum up, the contributions of the paper are :

- The original approach for detecting the attacks based on the physical layer via the spectra - this approach makes it possible to outsource the monitoring system.
- The classification protocol adapted to this context - both the ability to detect jamming attacks as well as protocol-based attacks.

This paper is organized as follows. Section II presents the considered EM attack experimentation configuration. Section III reports the state of the art about the SVM and the nearest neighbour approaches. Section IV is a detailed analysis of the spectra in a classification perspective. Finally, Section V gives classification results and the conclusion is drawn in Section VI.

II. EM ATTACK EXPERIMENTATION CONFIGURATION

The considered communication protocol is the IEEE 802.11n, which employs an OFDM modulation scheme. We consider two main attack modes: attack by deauthentication frames which corresponds to a protocol attack and attack by jamming signals.

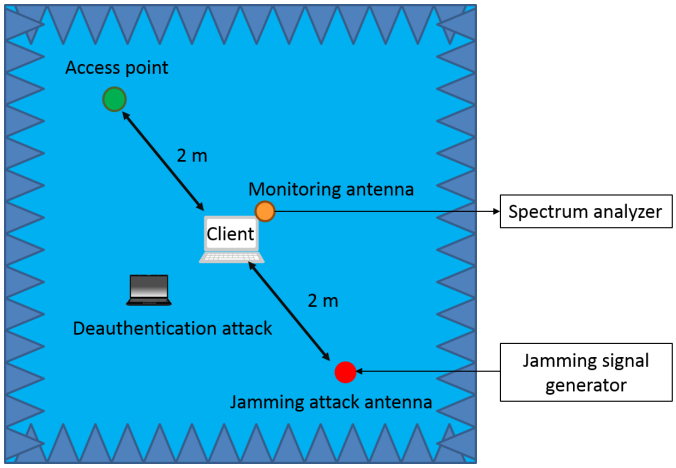


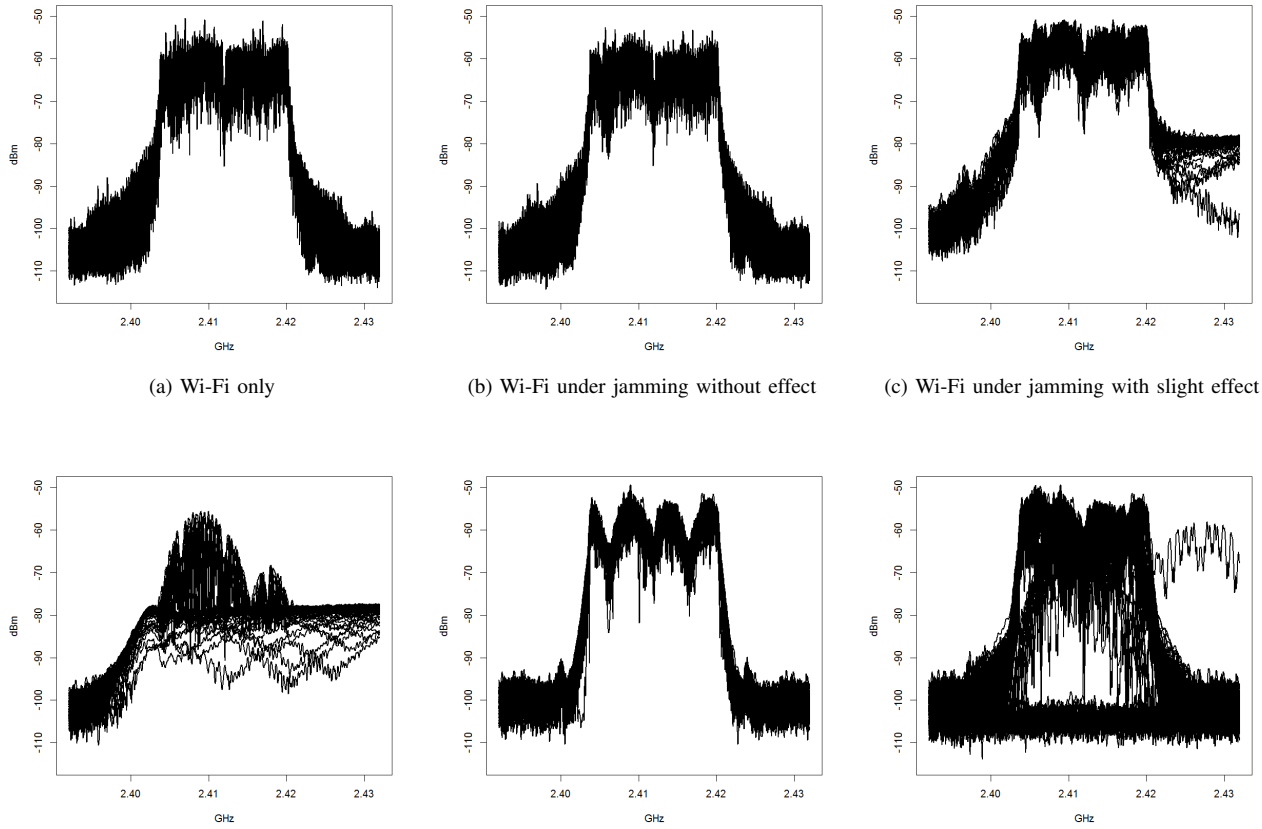
Fig. 1: Experimentation with a 802.11n communication in the presence of attack

a) IEMI attack: The attack by jamming signals consists in intentionally emitting a signal which covers the frequency bands employed by a communication system in order to disturb the reception of a communication device. A jamming signal is then an IEMI. The difference with IEMIs considered in [9] is the power level. The power levels of jamming signals are similar to communication signal power levels. The jamming signals can degrade the performance of the communication networks without damaging the communication devices. Different types of jamming signals can be used [7]. The vast majority of commercial jammers use a frequency-sweeping interference signal, which sweeps a frequency band $[f_1, f_2]$ in a time duration T. It can be expressed as:

$$s(t) = A\cos\left(2\pi\left(\frac{f_2 - f_1}{2T}t + f_1\right)t\right), \quad 0 < t < T, \quad (1)$$

where A is the interference signal amplitude. Here, the jamming signal that we consider sweeps the $[2.4\text{ GHz}, 2.5\text{ GHz}]$ frequency band in $T = 10\ \mu$s A time-frequency representation of the jamming signal is given Fig. 3.

b) Protocol attack: The attack by deauthentication frames uses management frames defined into the IEEE 802.11 standard. In a network infrastructure composed of several access points, when a client station (STA) is moving, the power of the Wi-Fi signal evolves. As for cellular networks (3G, 4G), a roaming principle has been specified in the standard. When a STA is connected to an Access Point (AP) and moves away from this AP, the Wi-Fi received signal power decreases. Due to the moving of the STA, it can detect the Wi-Fi beacon signal from another AP with an increasing power. In that case, a roaming procedure is launched. The roaming procedure consists in disconnecting the STA from the first AP and in reconnecting the STA to the second AP using IEEE 802.11 management frames deauthentication and authentication. The attack by deauthentication sends to a STA a frame of deauthentication even if the STA is not moving. Then, the STA



(a) Wi-Fi only

(b) Wi-Fi under jamming without effect

(c) Wi-Fi under jamming with slight effect

(d) Wi-Fi under jamming at the limit of loss of connection

(e) Wi-Fi in the presence of absorbers

(f) Wi-Fi under deauthentication attack

Fig. 2: Spectra of IEEE 802.11n communication plus attacks collected by the monitoring antenna

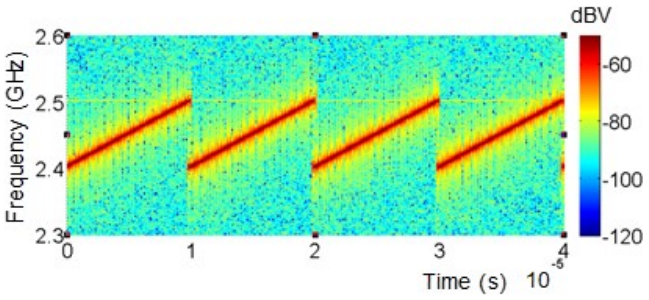


Fig. 3: Time-Frequency representation of the jamming signal.

executes the command and is disconnected from the wireless network. This attack can be easily implemented using tools like aircrack-ng (airplay-ng) [10]. Using particular parameters, the attacker creates a crafted frame that contains the real MAC address of the AP and the MAC address of the STA target. This attack is called directed deauthentication. For directed deauthentications, aireplay-ng sends out a total of 128 packets for each specified deauthentication. 64 packets are sent to the AP and 64 packets are sent to the client. After the attack, according to the configuration of the STA, the client can stay offline and the user has to reconnect the computer manually to the Wi-Fi network or the device can try by itself to reconnect after a few seconds. Both attacks with jamming signals or deauthentication frames, are generally the first step of attacks aiming the interception of data. Indeed, these attacks permit to disconnect the STA from a licit access point in order to reconnect it to a fake access point.

c) Device setting: To study the detection and the distinction of the different attack modes by using a classification-based approach, a Wi-Fi communication is set-up in an anechoic chamber by installing a server, an access point and a client computer. The Wi-Fi channel 1, centered on the 2.412 GHz frequency, is used. We include a monitoring antenna nearby the client. The monitoring antenna is connected to a spectrum analyzer which is outside the chamber. A 40 MHz frequency band, centred on 2.412 GHz, is monitored by the spectrum analyzer. The classification approach is performed on the collected spectra by the spectrum analyzer. Another antenna connected to an arbitrary waveform generator is placed in the chamber to emit the jamming signal. For the protocol-based attack, another computer is present in the anechoic chamber to send the deauthentication frames.

d) Attack configuration: To evaluate the performances of the classification approach, we implement 6 distinct configurations. The first configuration is without attack: spectra acquisitions are carried out with a Wi-Fi communication only. Three jamming attack configurations are tested: one configuration with a low powerful jamming signal that has no impact on the communication quality. The bit rate is still at the maximal level (about 95 Mbits/s) and the noise on the channel does not significantly change with or without the jamming signal. A second configuration uses a jamming signal power level which sligthly degrades the communication quality. The bit rate is reduced at about 75 Mbits/s. The third configuration uses a jamming signal power level 1 dB inferior to the required power to totally interrupt the communication. So, the three configurations whith three different jamming signal power levels represent a jammer which would be at three different distances from the client and the AP. Another configuration consists in degrading the communication quality but without any attack. For that, electromagnetic absorbing materials are placed around the access point in order to degrade the signal quality. Finally, the last configuration corresponds to the deauthentication attack for which a dedicated computer has been used. This dedicated computer sends deauthentication requests to the STA to force it to leave the network. In this work, these different attacks were applied permanently during the acquisitions.

e) Acquisition: To illustrate the different configurations, we represent 99 spectra (see Fig. 2) collected by the spectrum analyzer from the monitoring antenna which is nearby the client. The spectrum analyzer configuration is as follows: a 40 MHz frequency span, 2.412 GHz center frequency, a 100 kHz resolution bandwidth and 1601 points. The sweep time of the spectrum analyzer is 38.2 μs. By observing the figure, we notice different curves for each attack configuration.

III. Statistical approach

For the classification of data, it is essential to know the relationships that bind a variable (here the type of attack) that one seeks to classify with some observable variables called *explanatory variables* (in this work the spectra corresponding to the attacks).

The purpose of the statistical modeling is to predict an output from explanatory variables. In our work, this output is the state (attack or not and if attack, which one) of the communication network at a precise time. To perform this task, we first learn to model the six profiles of communication presented in figure 2.

A. Support Vector Machine (SVM) Basics

In 1964, Vapnik and Chervonenkis [11] set up a method to determine an optimal margin separator hyperplane for the separation of two classes in a Hilbertian space [12] associated with a scalar product denoted by $\langle .,. \rangle$. To begin with, let us consider a binary classification problem. The multi-class problem is investigated at the end of this section. For two classes, the goal is to find a classifier that will separate the data by maximizing the distance between the margin and the closest points of each class.

a) Main idea: In SVM classification, a linear classifier, called hyperplane, performs this task. The closest points, which are used for the determination of the hyperplane, are called support vectors (points $\mathbf{x}_1$ and $\mathbf{x}_2$ in Figure 4). There is

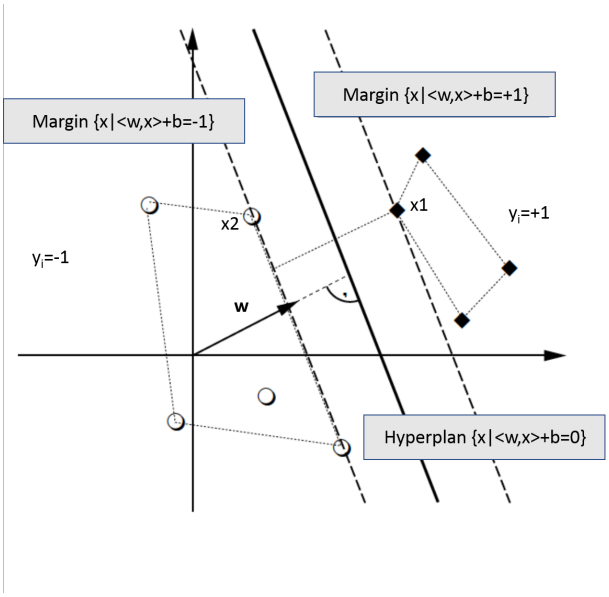


Fig. 4: Example of Separating hyperplane and support vectors in 2 dimensions.

a large variety of possible hyperplanes, but SVM chooses the optimal one. The intuitive idea is to look for an hyperplane for which the minimal distance in the learning data is maximal. To separate two classes, the greater the distance, the easier the class separation is. This distance is named "margin", as we seek to maximize it, it is then called a wide margin separator.

b) Separating Hyperplane: We have the following training data set : n couples $\{(\mathbf{x}_i, y_i), i = 1, \ldots, n\}$ where the label $y_i \in \{-1, 1\}$ indicates to which class the data $\mathbf{x}_i \in \mathbb{R}^p$ belongs. A canonical separating hyperplane H is given by:

$$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0. \tag{2}$$

We can choose $\mathbf{w}$ and b such that the closest point to H satisfies:

$$\langle \mathbf{w}, \mathbf{x} \rangle + b = \begin{cases} 1 & \text{if } \mathbf{w}^\top \mathbf{x}_i + b \geq 0, \\ -1 & \text{if not.} \end{cases} \tag{3}$$

we deduce that $\mathbf{w}^\top \mathbf{x}_1 + b > 0$ and $\mathbf{w}^\top \mathbf{x}_2 + b < 0$ as:

$$\langle \mathbf{w}, \mathbf{x}_1 - \mathbf{x}_2 \rangle = \langle \mathbf{w}, \mathbf{x}_1 \rangle - \langle \mathbf{w}, \mathbf{x}_2 \rangle = (1 - b) - (-1 - b) = 2, \tag{4}$$

and consequently the margin towards H, denoted by M is given by:

$$M = \left\langle \frac{\mathbf{w}}{\|\mathbf{w}\|}, \mathbf{x}_1 - \mathbf{x}_2 \right\rangle = \frac{2}{\|\mathbf{w}\|}, \tag{5}$$

where $\|\mathbf{w}\| = \sqrt{w_1^2 + \ldots + w_p^2}$.

So, smaller $\|\mathbf{w}\|$ is, bigger the margin is. For this reason, in order to find the hyperplane which offers the best separation of the data, we must find the one that respects the conditions

of a canonical hyperplane and for which $\|w\|$ is minimal. In order to find this hyperplane, we need to minimize:

$$\frac{1}{2}\|\mathbf{w}\|^2 \tag{6}$$

under the constraints

$$y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) \geq 1, i = 1, \ldots, n \tag{7}$$

which ensures that the hyperplane separates well the data and that it is canonical.

c) Decision function: The decision function determines to which class belongs $\mathbf{x}_i$. It is based on the sign of $\langle \mathbf{w}, \mathbf{x}_i\rangle + b$. The mathematical function that extracts the sign is noted sgn and returns the values 1 or -1. For $i = 1, \ldots, n$, $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) > 0$ only if $\text{sgn}\left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right)\right) = y_i$ and so if $\mathbf{x}_i$ is in the right side of the hyperplane. Then, we can show that if $\langle \mathbf{w}, \mathbf{x}_i\rangle + b \geq 0$, then $y_i = 1$ is assigned to $\mathbf{x}_i$, if $\langle \mathbf{w}, \mathbf{x}_i\rangle + b \leq 0$, $y_i = -1$. The hyperplane is said canonical.

A property of this problem is that

$$f(\mathbf{w}) = \|\mathbf{w}\|^2 = w_1^2 + \ldots + w_p^2 \tag{8}$$

is a convex function. That ensures the absence of local minimum and the presence of an unique solution.

d) Lagrangian: In the current context, the problem of maximization becomes a problem of maximization under constraints. The minimization problem of (6) under the constraints (7) can be solved with the Lagrange multipliers. We denote, for $i = 1, \ldots, n$, $\alpha_i \geq 0$ the Lagrange's multipliers associated to the constraints. The Lagrangian function is given by:

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i \left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1\right). \tag{9}$$

e) Dual problem: Integrating the constraints, using Lagrange multipliers, the minimization problem can be formulated in its dual form. Thus, we try to minimize $\mathcal{L}(\mathbf{w}, b, \alpha)$ with respect to the primal variables $\mathbf{w}$, b and to maximize it with respect to the dual variables α_i. When we reach the optimal point of this minimization problem, we have $\frac{\partial \mathcal{L}(\mathbf{w},b,\alpha)}{\partial b} = 0$ and $\frac{\partial \mathcal{L}(\mathbf{w},b,\alpha)}{\partial \mathbf{w}=0}$ and so $\sum_{i=1}^{n} \alpha_i y_i = 0$ and $\sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i = w$. By replacing these values in the Lagrangian, we obtain the dual problem:

$$\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle \tag{10}$$

under constraints $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$. In the dual problem, the b coefficient does not appear. It is known that when $\langle \mathbf{w}, \mathbf{x}\rangle + b > 1$ the coefficient $\alpha_i = 0$ and that $\langle \mathbf{w}, \mathbf{x}\rangle + b = 1$ for the support vector.

f) Resolution of the dual problem: To solve the dual problem, an average of these parameters is then calculated for the carrier vectors to estimate $M = \frac{2}{\|\mathbf{w}\|} = \left(\sum_{i \in SV} \alpha_i\right)^{-1/2}$ where SV (Support Vector) is the set of the support vectors. The decision boundary is then written:

$$\langle \mathbf{w}, \mathbf{x}\rangle + b = \sum_{i \in SV} \alpha_i y_i \langle \mathbf{x}_i, \mathbf{x}\rangle + b \tag{11}$$

g) Non-linear hyperplane: Separation of classes by a linear hyperplane is a special case. Generally, the definition of the boundary between the different classes requires to determine a non linear hyperplane. The definition of a such hyperplane leads to the introduction of a release variable ξ_i for the constraints. The release variable ξ_i penalizes the slacking in the objective function. The problem then becomes

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \xi_i \tag{12}$$

under constraints $y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) \geq 1 - \xi_i$ and $\xi_i \geq 0$ with C the penalisation term which has to be defined. The Lagragian is then given by

$$\begin{aligned}\mathcal{L}(\mathbf{w}, b, \xi_i, \alpha, \nu) =& \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \\ &- \sum_{i=1}^{n} \alpha_i \left(y_i\left(\langle \mathbf{w}, \mathbf{x}_i\rangle + b\right) - 1 + \xi_i\right) \\ &- \sum_{i=1}^{n} \nu_i \xi_i.\end{aligned} \tag{13}$$

Under optimal conditions we have $\sum_{i=1}^{n} \alpha_i y_i = 0$, $w = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i = 0$ and $C - \alpha_i - \nu_i = 0$ for $i = 1, \ldots, n$. By replacing these values in the Lagrangian, we obtain the following dual problem:

$$\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle \tag{14}$$

under constraints $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$. We consider now a nonlinear separation. In this case, the complexity of the separator hyperplane used in the linear case is not sufficient to correctly classify the data. In order to overcome this problem, we consider a nonlinear transformation $\phi(.)$ which projects the data in a larger dimension space so that the data in the transformed space are linear. The separating hyperplane is then written:

$$\langle \mathbf{w}, \phi(\mathbf{x})\rangle + b. \tag{15}$$

Replacing the data by the transformation $\phi(.)$ in the Lagrangian we obtain:

$$\begin{aligned}\mathcal{L}(\mathbf{w}, b, \xi_i, \alpha, \nu) =& \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \xi_i \\ &- \sum_{i=1}^{n} \alpha_i \left(y_i\left(\langle w, \phi(\mathbf{x}_i)\rangle + b\right) - 1 + \xi_i\right) \\ &- \sum_{i=1}^{n} \nu_i \xi_i.\end{aligned} \tag{16}$$

under constraints $\sum_{i=1}^{n} \alpha_i y_i = 0$, $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i = 0$ and $C - \alpha_i - \nu_i = 0$, $i = 1, \ldots, n$ we deduce:

$$\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle \tag{17}$$

under constraints $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$.

h) Kernel trick: It is difficult to build the hyperplane in the transformed space ($\phi(.)$). In 1992, Boser et al. [13], using Mercer theorem [14] have found a way to build the optimal hyperplane in the projection space without using an explicit form of it. Applying this theorem, the problem can be written:

$$\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \qquad (18)$$

under constraints $0 \leq \alpha_i \leq C$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$. In this study, we consider the Radial Basis Function kernel (RBF), given by:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\,\sigma^2}\right), \qquad (19)$$

where $\frac{1}{2\sigma}$ set the width of the bell-shaped curve which should be chosen a priori. The larger the value of $\frac{1}{2\sigma}$, the narrower the bell is.

i) Multi-class problem: In this study, the problem is a multi-class classification case. Indeed, the problem is not limited to one class for "attack" and one for "normal communication". Firstly, we consider different kinds of attacks in order that the classification allows us to identify the type of attacks. Secondly, we consider the possibility to have configurations in which the communication quality is degraded due to bad propagation conditions. We want to be able to distinguish this last case in order to avoid false positive attack detections. In a multi-class classification case, two stategies can be deployed to adapt the SVM algorithm. The first is One-against-one strategy and the second is One-against-all strategy. In the following, the strategy selected to classify the spectra is the One-against-all strategy [15]. One-against-all strategy involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negative. To build the L-class classifiers, it is common to construct binary classifiers $f^1, f^2, \ldots, f^L$ and combine them. The combination of these classifiers is carried out by ajusting the maximal output before applying it to the function sgn. The combination is then given by

$$\underset{j=1,\ldots,L}{\operatorname{argmax}} \sum_{i=1}^{l} y_i \alpha_i^j K(\mathbf{x}, \mathbf{x}_i) + b^j. \qquad (20)$$

This value can also be used as a rejecting decision when we consider the difference between the two largest values and as a confidence-building-measure in the classification of x.

B. Nearest Neighbour Approach

The nearest neighbor classifier does not require any pre-processing of the labelled samples prior to its use. The crisp nearest-neighbor classification rule assigns an input sample vector $\hat{\mathbf{y}}$ to the class of its nearest neighbor [16]. The vector $\hat{\mathbf{y}}$ contains the labels predicted by the SVM model. This idea can be extended to the K-nearest neighbors with the vector $\hat{\mathbf{y}}$ being assigned to the class that is represented by a majority amongst the K-nearest neighbors. When more than one neighbor is considered, we can have a tie among classes with a maximum number of neighbors in the group of K-nearest neighbors. One simple way of handling this problem is to restrict the possible values of K. For example, given a two-class problem, if we restrict K to odd values only, no tie will be possible. Of course, when more than two classes are possible, this technique is not useful. A way of handling the occurrence of a tie is as follows. If these classes are tied, the sample vector is assigned to the class for which the sum of distances from the sample to each neighbor in the class is a minimum. This can still lead to a tie. In that case, the assignment is to the last class encountered. Consequently, there are cases where a classification vector becomes an arbitrary assignment, no matter what additional procedures are included in the algorithm.

C. General protocol

In this section, we describe the protocol which is applied to classify attacks by focusing on the relationship between the attack and EM spectra (classification) as well as the time correlation (correction). To estimate the profile of the attack (and as a consequence learn to recognize it), the protocol contains two main parts.

a) Step 1: The first part of the protocol (Figure 5) performs a learning step of a classification based on SVM algorithm [11] (see Section III-A) with a radial basic function [17] as kernel. The SVM classification is performed using an One-against-all classification approach [15]. In order to choose the parameter σ, we minimize the validation error for a sigma value included in a selected interval obtained with the procedure developed by Caputo and al. [18]. To estimate C, we considered the approach of Cherkassy and Ma [19].

b) Step 2: The second step predicts the class of a new data using two steps. In a first time, using the model estimated in step 1, a prediction is made for the new data. In a second time, with a lag of $k/2$, a correction is applied on the predicted class using K nearest temporal neighbors. The correction is possible due to the duration of the attacks over time which cannot be focused on a single spectrum [16]. In the figure 5, Step 2 presents the general architecture of the proposed system that allows us to test the quality of the prediction.

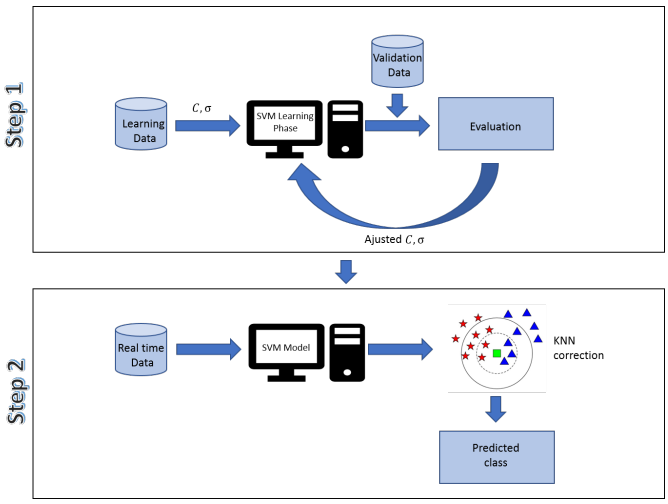


Fig. 5: Schematization of the Protocol.

IV. Data analysis

A. Frequency band selection

To analyse the importance of the frequencies, we study the linear relationship between the frequencies of a spectrum. From the spectra obtained previously, we aim to find models that could distinguish classes corresponding to known attacks or to a normal use of the network.

The spectrum acquisitions are performed over a 40 MHz frequency band. The Wi-Fi channel bandwidth being 20 MHz, the spectrum characteristics before and after the 20 MHz channel may significantly vary if there are other access points or not in the vicinity. Thus, it is preferable for the classification to only exploit data belonging to the frequency band of the channel so that the detection is less dependent on the electromagnetic activity over the other channels. In learning algorithms, before any modeling we try to reduce the number of variables and to only select variables in the 20 MHz channel frequency band, for example by deleting the ones providing identical information. In the case of the spectra, due to the type of acquisition and the set up of the spectrum analyzer, we need to check the correlation between the different frequencies composing the spectra (see Fig. 6). Figure 6 highlights the high level of correlation between the frequencies in the interval $[2.402; 2.422]$ GHz as well as the frequencies in the interval $[2.392; 2.402[U]2.422; 2.432]$ GHz. In addition, the central frequency corresponding to the zone D is more than 60% correlated with the whole band. Thus, only the data included in the band 2.402 and 2.422 GHz, corresponding to the zones B, D and E, are retained to perform a classification.

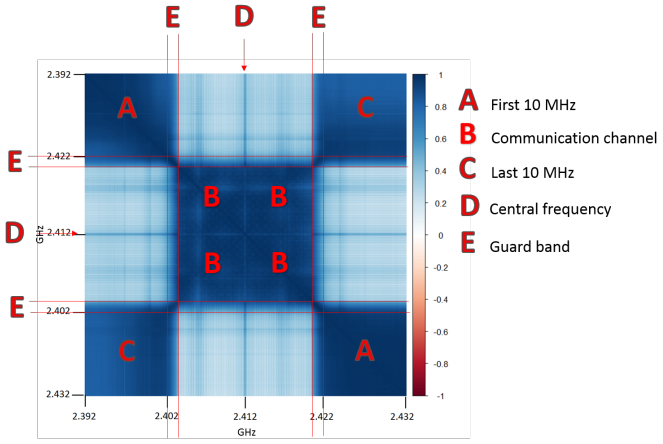


Fig. 6: Frequencies correlation.

B. Principal Component Analysis

A way to check if the different classes can be separated is to compute a Principal Component Analysis (PCA) on all data, i.e. 99 spectra for the 6 configurations [20]. These components correspond to the axes obtained from the eigenvectors constructed from the spectra. Projecting the spectra on the two components associated to the eigenvector possessing the higher eigenvalues (see Fig. 7), we check if the different configurations can be discriminated by classification. In this new representation, each point corresponds to one spectrum. The six different colors represent the spectra collected in the six different configurations: Wi-Fi communication alone in blue, Wi-Fi in the presence of absorbers in yellow, Wi-Fi with jamming signal without effect in red, Wi-Fi with jamming signal with lightly effect in green, Wi-Fi with jamming signal at the limit of the break in purple and Wi-Fi communication with a deauthentication attack in brown.

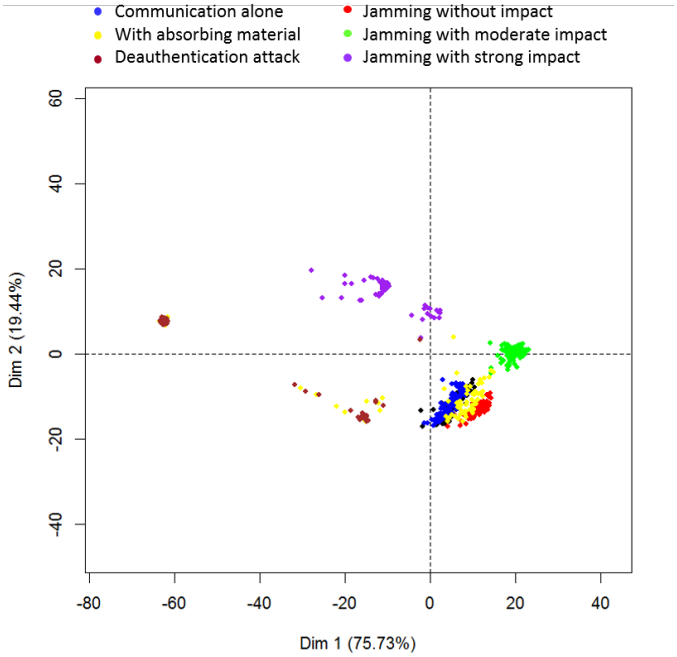


Fig. 7: Representation of the spectra on the two first components of the PCA.

In Fig. 7, we notice that the attack by deauthentication is significantly separated from the other classes. This result is very satisfying because this attack is a protocol-based attack and the nature of the signal is not really different from a normal communication. This result is encouraging to develop a detection approach able to distinguish a large number of attacks. Nevertheless, the separation is less obvious between the Wi-Fi communication alone, the jamming without impact and with moderate impact. To improve the separation between these three classes, we have to take into account other components beyond the first two and we must adopt a nonlinear separation. As a consequence, the relevance of the retained variables (zones B, D et E in Fig. 6) has to be verified in the case of a non linear analysis.

C. Identification of the most discriminating frequencies

The relevance of the retained variables has been demonstrated in a linear analysis. While conducting a nonlinear analysis, it must be verified that the most discriminative frequencies belong to sections B, D and E of Fig 6.

To identify the most discriminating frequencies in the spectrum, we use sensitivity response aggregation functions [21]. The sensitivity method works by varying an input variable x_a through its range with S levels. This method works by

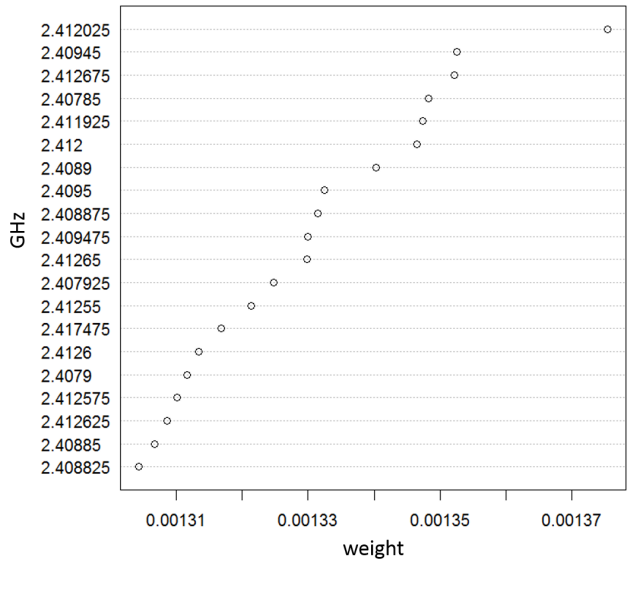


Fig. 8: Weight of the 20 first relevant frequencies.

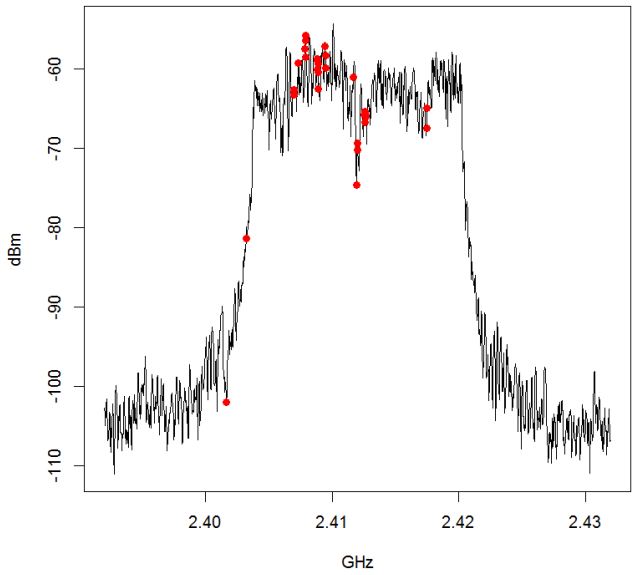


Fig. 9: Best Frequency.

considering a given baseline vector $\mathbf{d}$ of size S which generally contains the statistical values (mean, median, quartile, ...) of each input. Then, it cycles through all $\{x_a : a \in 1, \ldots, p\}$ inputs. For each input, S inputs are built using all d values except $\{x_{aj} : j \in \{1; \ldots; S\}\}$. We denote the respective model responses by $\hat{y_a} = \{\hat{y_{aj}} : j \in \{1; \ldots; S\}\}$, where $\hat{y_{aj}}$ represents the response for x_{aj}. Based on these values, we compute a sensitivity measure of the input. Thus, highest is the score obtained by a variable, the more it discriminates the studied cluster. To study the discrimination capacity of the different frequencies, we use these sensitivity response aggregation functions. Fig. 8 represents the 20 most discriminating frequencies, with in abscissa, their respective weight issued from the sensitivity measure. These most discriminating frequencies are located between 2.4 GHz and 2.42 GHz (see red points in Fig. 9) and the best frequency is 2.412025 GHz. Therefore, the central 20 MHz frequency band is sufficient to discriminate the spectra associated to the different types of attack, even using a non linear analysis. In the following, only the central frequency band of the spectrum is considered.

Moreover, in this paper the acquisitions are performed in anechoic chamber, using a single Wi-Fi channel. However, in realistic conditions, the adjacent Wi-Fi channels can be used or not. Therefore, the spectra over 40 MHz can significantly evolve according to the activity of the Wi-Fi network. As a consequence, centering the analysis on the 20 MHz of the used channel can make the classification results less dependent on the activity of the adjacent communication channels.

The identification of the most discriminative frequencies is also interesting to optimize the acquisition process. Indeed, the perspective of this work, is also to develop attack detection tools on Software Defined Radio (SDR) platforms. The acquisition process could be optimized by focusing on a reduced frequency band including the most relevant frequencies for the classification.

V. Classification results

The classification is then performed only on the central 20 MHz frequency band. We want to identify whether or not the communication network is facing an attack at a precise time using the spectra corresponding to this time. We have to identify 6 attack profiles presented in section II. We estimate the attack profile using the protocol presented in Fig. 5. In the second step of the protocol, to perform the K-nearest neighbors on the SVM predictions, we use the 10 nearest neighbors. To verify the quality of the prediction, the general architecture of the proposed system is tested on sampling data constructed as presented in Fig. 5.

The learning phase is the phase where the SVM model learns the separating hyperplanes. The learning data set is composed of 49 spectra for each configuration (294 spectra in total). The validation phase is a verification of the parametrization of the SVM model. The validation data set is composed of 29 spectra for each configuration (174 spectra). Finally, the testing phase is the phase in which we compute and evaluate the correction. The real time data is replaced by a testing data set composed of 126 spectra organized in a temporal way simulating a succession of configurations (see figure 10).

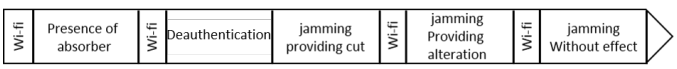


Fig. 10: Test set organization.

The classification quality is evaluated using the classification error. Table I represents this error on the training, the

TABLE I: Classification error

	model error (%)	sample size n
Cross-validation	15	468
Training	7	294
Validation	9	174
Test	14	126

validation, the testing sets and by a cross-validation. Cross-validation is a validation technique to assess how the results of a statistical analysis can be generalized to an independent data set. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset and validating the analysis on the other subset. The number of folds determines the number of rotations used to estimate the combined measures. In the following, we use the cross-validation on the combined training, validating and testing sets to estimate the classification error using a three-fold validation.

Table II represents the confusion matrix between the predicted classes and the real one (the predicted classes are obtained by cross-validation). 1) corresponds to a Wi-Fi communication only, 2) corresponds to a communication in the presence of absorbers, 3) corresponds to a communication in the presence of jamming without effect, 4) corresponds to a communication with a light impact jamming, 5) corresponds to a communication in the presence of jamming that provokes short communication interruptions and 6) corresponds to a communication with a deauthentication attack (see Table II).

Observing the confusion matrix, the majority of the errors are concentrated on two different cases : (1) between normal communication and the presence of absorbing materials, and (2) between strong jamming creating short connection losses and deauthentication attacks. It is important to note that there is no confusion between a degraded propagation situation by the presence of absorbing materials and the presence of an attack signal. This shows that the degraded propagation situation is not likely to produce a false alarm. The errors between normal communication and the presence of absorbing materials can be explained by the proximity in their spectra profiles as it can be observed in Fig. 7. The confusion between strong jamming creating short connection losses and deauthentication attacks, comes from restarts on the Wi-Fi communication after interruptions. To overcome this confusion, we compute a correction using the 9 nearest neighbors which permits to take into account the proximity in time of the spectra to be analysed. To evaluate this correction, we apply it on the testing set.

In Table III, to analyze the weakness of the initial model and the potential relation between the attacks, we also analyze the confusion matrix established over the test data respectively without and with correction by the nearest neighbors.

Here, the correction is based on the fact that an attack cannot be implemented on only one spectrum, it lasts a certain time and covers several spectra. In our configuration, this correction suppresses all the errors. But, the correction does not work when an error occurs on a transition phase (transition between two states of the network). Moreover, the lack of good prediction in a transition phase can create a new error in the case where the classification correctly predicts the attack or produces a wrong correction of the error. The choice of the K parameter is crucial for the correction. On one hand, it is important to take a sufficiently large value to have a good correction. On the other hand, higher the value is, later the prediction arrives.

TABLE II: Confusion matrix obtained by cross-validation between predicted classes and real configuration.

real attack \ predicted attack	1	2	3	4	5	6
1: Wi-Fi alone	80	19	0	0	0	0
2: With absorbers	25	74	0	0	0	0
3: Low jamming	0	0	97	2	0	0
4: Moderate jamming	0	0	4	95	0	0
5: Strong jamming	0	0	0	0	63	36
6: Deauthentication	0	0	0	0	2	97

TABLE III: The confusion matrix obtained on the testing set. In each line of the table, the two results correspond respectively to the result without and with correction

real attack \ predicted attack	1	2	3	4	5	6
1: Wi-Fi alone	21	5	0	0	0	0
	26	0	0	0	0	0
2: With absorbers	7	19	0	0	0	0
	0	26	0	0	0	0
3: Low jamming	0	1	23	2	0	0
	0	0	26	0	0	0
4: Moderate jamming	0	0	1	25	0	0
	0	0	0	26	0	0
5: Strong jamming	0	0	0	0	20	6
	0	0	0	0	26	0
6: Deauthentication	0	0	0	0	2	24
	0	0	0	0	0	26

VI. Conclusion

This paper focuses on the conception of a monitoring system able to detect and classify jamming and protocol based attacks. To achieve this goal, we propose to outsource the attack detection function from the network to protect and we use an antenna to monitor the spectrum over the time.

In this study, the Wi-Fi network and the attacks are carried out in an anechoic chamber to avoid disturbing other Wi-Fi communication networks in the vicinity.

A study of the spectra highlights that the frequencies of interest belong to the communication channel between 2.402 GHz and 2.422 GHz. Focusing the analysis on this 20 MHz frequency band permits to construct a classification model to overcome the problems induced by the utilization of the adjacent channels which can be or not occupied by other Wi-Fi communications.

On these frequencies, the proposed estimation model shows good results in the prediction of attacks. In addition, the correction using the K spectra nearest in time permits to correct most of the miss classification.

In our future work, we plan to verify the behavior of our model on data acquired outside of the anechoic chamber, in realistic situations. Another important point is knowing how our model can evolve in the case where unknown attack occurs. Finally, as new (unpresented/unlearned) attacks can appear very quickly, we aim to use machine learning techniques. These techniques include adaptive classification algorithms, able to change the models after their creation and the number of classes over time. By learning upstream the standard behavior of the communication, the algorithm analyzes the data as they arrive and try to classify them as one standard communication or not. If the communication is not standard, a new class is created and considered by defaults as an unknown attack.

ACKNOWLEDGMENT

This work was performed in the framework of the X2Rail-1 project (Shift2Rail Joint Undertaking) and in the framework of the ELSAT2020 project which is co-financed by the European Union with the European Regional Development Fund, the French state and the Hauts de France Region Council.

REFERENCES

[1] C. H. Rowland, *Intrusion detection system*, Jun. 11 2002, uS Patent 6,405,318.
[2] M. Sun and T. Chen, *Network intrusion detection system*, Sep. 30 2010, uS Patent App. 12/411,916.
[3] L. Vokorokos and A. Balaz, *Host-based intrusion detection system*,in 14th International Conference on Intelligent Engineering Systems (INES), IEEE, 2010, pp. 4347.
[4] P. Van Aubel, K. Papagiannopoulos, L. Chmielewski, and C. Doerr, *Side-channel based intrusion detection for industrial control systems*, arXiv preprint arXiv:1712.05745, 2017.
[5] W. Xu, W. Trappe, Y. Zhang, and T. Wood, *The feasibility of launching and detecting jamming attacks in wireless networks*, in Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing, ACM, 2005, pp. 4657.
[6] R. Bhojani and R. Joshi, *An integrated approach for jammer detection using software defined radio*, Procedia Computer Science, vol. 79, pp. 809816, 2016.
[7] V. Deniau, C. Gransart, G. L. Romero, E. P. Simon, and J. Farah, *IEEE 802.11n Communications in the Presence of Frequency-Sweeping Interference Signals*, IEEE Transactions on Electromagnetic Compatibility, vol. 59, no. 5, pp. 16251633, 2017.
[8] S. Grimaldi, A. Mahmood, and M. Gidlund, *An svm-based method for classification of external interference in industrial wireless sensor and actuator networks*, Journal of Sensor and Actuator Networks, vol. 6, no. 2, p. 9, 2017.
[9] *Electromagnetic compatibility (EMC) - part 2-13: Environment - high-power electromagnetic (HPEM) environments - radiated and conducted*, IEC Standard, Tech. Rep. 61000-2-13 Ed. 1, 2005.
[10] R. Vinek, *BackTrack 5 Wireless Penetration Testing Beginners Guide*, Packt Publishing Ltd., 2011, ISBN : 978-1-849515-58-0.
[11] V. Vapnik, *The nature of statistical learning theory*, Red Bank: Springer, vol. 2, 2000.
[12] P. Halmos, *Introduction to Hilbert space and the theory of spectral multiplicity*, Chelsea Pub. Co., 1957.
[13] B. E. Boser, I. M. Guyon, and V. N. Vapnik, emphA training algorithm for optimal margin classifiers, in Proceedings of the fifth annual workshop on Computational learning theory, ACM, 1992, pp. 144152.
[14] B. J Mercer, *XVI. Functions of positive and negative type, and their connection the theory of integral equations*, Phil. Trans. R. Soc. Lond. A, vol. 209, no. 441-458, pp. 415446, 1909.
[15] Y. Liu and Y. F. Zheng, *One-against-all multi-class SVM classification using reliability measures*, in Neural Networks, IJCNN05. Proceedings. 2005 IEEE International Joint Conference on , vol. 2. IEEE, 2005, pp. 849854.
[16] L. E. Peterson, *K-nearest neighbor*, Scholarpedia, vol. 4, no. 2, pp. 1883, 2009.
[17] J. Park and I. W. Sandberg, *Approximation and radial-basis-function networks*, Neural computation, vol. 5, no. 2, pp. 305316, 1993.
[18] B. Caputo, K. Sim, F. Furesjo, and A. Smola, *Appearance-based object recognition using SVMs: which kernel should I use?*, in Proc of NIPS workshop on Statistical methods for computational experiments in visual processing and computer vision, Whistler, vol. 2002, 2002.
[19] V. Cherkassky and Y. Ma, *Practical selection of SVM parameters and noise estimation for SVM regression*, Neural networks, vol. 17, no. 1, pp. 113126, 2004.
[20] S. Wold, K. Esbensen, and P. Geladi, *Principal component analysis*, Chemometrics and intelligent laboratory systems, vol. 2, no. 1-3, pp. 3752, 1987.
[21] P. Cortez and M. J. Embrechts, *Using sensitivity analysis and visualization techniques to open black box data mining models*, Information Sciences, vol. 225, pp. 117, 2013.

Jonathan Villain received the Ph.D. degree in applied Mathematics from the University of Bretagne-Sud, Vannes, France, in 2016 in a French Laboratory of mathematic (LMBA, Vannes) and a Pharmaceutical Laboratory (CERMN, Caen). During 2015 and 2017, he was a Teaching Assistant at the IUT of Vannes. His research field include machine learning and pattern recognition modeling in pharmaceuticals toxicity and cyber-security. He is actually attached to the Railenium test and research center as a Post-doctoral fellow and participates to WP8 Shift2Rail programs by studying classification models of electromagnetic disturbances in the railway.

Virginie Deniau received the M.S. and Ph.D. degrees in electronics from the University of Lille in 2000 and 2003, respectively. Since 2003, she is Researcher in electromagnetic compatibility (EMC) for the French Institute of Science and Technology for Transport, Development, and Networks (IFSTTAR). She conducts works on electromagnetic compatibility (EMC) for land transport. Her research interests include EMC test facilities and methodologies, characterization and modeling of electromagnetic transport environments and the immunity test methodologies for embedded systems. Currently, she works in the hardening of land transport systems regarding cyberattacks, such electromagnetic attacks. She has participated in numerous national and European projects and she was scientic coordinator of the FP7 project SECRET for SECurity of Railways against Electromagnetic aTtacks. She is also vice-chair of the URSI Committee E (Electromagnetic Interference).

Anthony Fleury (IEEE S2005, M2008) is associate professor at IMT Lille Douai. He received an Engineer (Computer Science) and a M.Sc. (Signal Processing) degree in 2005 in Grenoble and a PhD degree in Signal Processing from the University Joseph Fourier of Grenoble in 2008 for his work on Health Smart Homes and activity recognition. He joined then the LMAM team at Swiss Federal Institute of Technology and is now, became, in sept. 2009, Associate Professor at Ecole des Mines de Douai (now IMT Lille Douai). His research interests include modeling human behaviors and activities, machine learning and pattern recognition with applications to biomedical engineering and to security.

Eric P. Simon received the Masters degree in electronics engineering from the Superior School of Electronics (ESCPE), Lyon, France, in 1999, and the Ph.D. degree in signal processing and communications from the National Polytechnic Institute of Grenoble (INPG), France, in 2004. During 2005, he was a Teaching Assistant at the INPG and the following year he joined one of France Telecom R&D Laboratories as a Postdoctoral Fellow. He is currently an Associate Professor at the Institute of Electronics, Microelectronics and Nanotechnology (IEMN), TELICE (Telecommunications, Interference and Electromagnetic Compatibility) Group, University of Lille, France. His main research interests are in mobile communications and carrier and symbol synchronization.

Christophe Gransart received the Ph.D. degree from the University of Lille, Villeneuve-dAscq, France, in 1995. He is a Senior Researcher with French Institute of Science and Technology for Transport, Development, and Networks, Villeneuve dAscq, with 15 years experience in participating in industrial and academic research projects dealing distributed systems and middleware for transportation systems, V2V and V2I communications, adaptive middleware and cybersecurity. He was involved in various national and European projects. The main competencies are computer science, distributed architecture design, and middleware expertise. He participated to FP6, FP7, H2020, Shift2Rail programs.

Raouf Kousri Raouf Kousri received a Master's Degree in Electronic Embedded Systems in 2012 from the University Pierre & Marie Curie in Paris, and received his PhD in EMC in 2016 from the University of Sciences and Technologies in Lille. He worked with the IFSTTAR (French Institute of Science and Technology for Transport, Development, and Networks) and the Railenium Reasearch Institute on the detection and caracterization of intentionnal and non intentionnal electromagnetic disturbances in the railway domain and the susceptiblity of several wireless communication protocols regarding these disturbances. He is currently an EMC expert with Alten Technologies, Paris, France.