*3rd URSI AT-AP-RASC, Gran Canaria, 29 May – 3 June 2022*

# Detection and Classification of Interference affecting LoRaWAN communications in Railway environment

Jonathan Villain(1), Virginie Deniau(1), Eric Pierre Simon (2), Christophe Gransart(1), Artur Nogueira de São José(2), Florent Valenti (3), Norbert Becuwe (3)
(1) COSYS-LEOST, Univ Gustave Eiffel, IFSTTAR, Univ Lille, F-59650 Villeneuve d'Ascq, France
(2) Univ. Lille, CNRS, UMR 8520-IEMN-Institut d'Electronique de Microélectronique et de Nanotechnologie, F-59000 Lille, France
(3) SNCF Voyageurs - Direction du matériel ingeniérie du matériel - Cluster NORD - Groupe Systèmes Informatiques Embarqués, 59260 HELLEMMES, France

## Abstract

The French national railway company (SNCF) is deploying Internet of Things (IoT) technologies using the LoRaWAN communication protocol to centralize and transmit the data measured by the onboard sensors to the railway maintenance centers. They recently developed a communication interface connected to the different sensors called MARTI. A gateway called MELI then switches data from the LoRa protocol to the 4G network to centralize them to SNCF's IoT platform. However, the reception of the gateway MELI can be affected by the transient electromagnetic interference occurring with the catenary-pantograph contact losses. Moreover, in a security context, these communications can also be intentionally disturbed by the use of jammers. This work aims to detect the presence of this intentional and non-intentional interference and to distinguish them. This should allow sending the LoRa signal at instants without interference to guaranty the good reception of the LoRa communications by the gateway. We performed experiments in the laboratory to analyze the performance of a Support Vector Machine classification (SVMc) approach to detect and separate such interference.

## 1 Introduction

Connected sensors are more and more present in our society, and the french national railway company (SNCF) is not left out of this technology. The SNCF is seeking to connect different sensors and centralize their information on servers to be consulted remotely. The wireless connexion of the objects and the availability of a large amount of information make it possible to envisage evolutions in the maintenance and management of railway components.

Onboard SNCF trains, many sensors collect information on the state of each wagon. In order to centralize the sensors information over the IoT plateforms, SNCF decided to employ the LoRaWAN technology [1]. The use cases already implemented in commercial service are numerous: detection of the water level in sanitary facilities, supervision of train doors, pressure in air conditioning units, etc.

These connected sensors help prevent failures and guide maintenance actions so that predictive maintenance can replace routine maintenance at regular intervals through continuous information monitoring. The solution chosen to make these sensors communicate is integrating a LoRa communication system. SNCF has thus developed its own solution of certified railway-connected sensors, called MARTI (Agile Module for Reception and Transmission of Information), which communicate in LoRa with a gateway called MELI (Information Link Embedded Modem).

Deniau et al. [2] studied the susceptibility of the LoRa communication for an application in a railway environment, in the context of the LoRa-R project. Such project aims to study the LoRa communication system's vulnerabilities and detect any disturbances that could disrupt communication in the railway context. Indeed, the SNCF implementation of the LoRaWAN communications can be particularly vulnerable to interference, since it does not include acknowledgement messages in order to save the battery life duration. The disturbances analyzed in [2] are intentional and non-intentional disturbances such as jamming attacks and transient electromagnetic interference, the last one being typical of railways environment resulting from the losses of contact between the catenary and the pantograph.

The jamming attacks and transient electromagnetic interference are the main threat to LoRa communication in the railway environment. In this paper, we propose a method to build a classification model able to detect these signals on LoRa communication channels while being robust to the variability of the surrounding electromagnetic (EM) noise. We proceed in two steps. First, we build the classification model based on conducted measurements. This controlled environment ensures to take into account only the signals of interest, i.e., the LoRa communication, jamming, and transient signals. In this way, the surrounding EM noise does not affect the construction of the classification model, which makes it more general than if it has been built in a given specific EM environment. Then, in a second step, the

obtained model is tested in a radiated mode to assess its validity in real environments. The results from the second step presented in this paper are still preliminary since a more realistic evaluation would require on-site measurements.

## 2 Experimentation

### 2.1 Bench description

The experiment was carried out in a laboratory of the Gustave Eiffel University, in which a specific test bench for the LoRaWAN communication study was installed. The first experiment is carried out with a fully conducted bench (see Figure 1) and the second with a radiated mode for the LoRa communication (see Figure 2). In this second test bench, the LoRa reception is potentially affected by surrounding EM noise. For these experiments, we use a Kerlink gate-

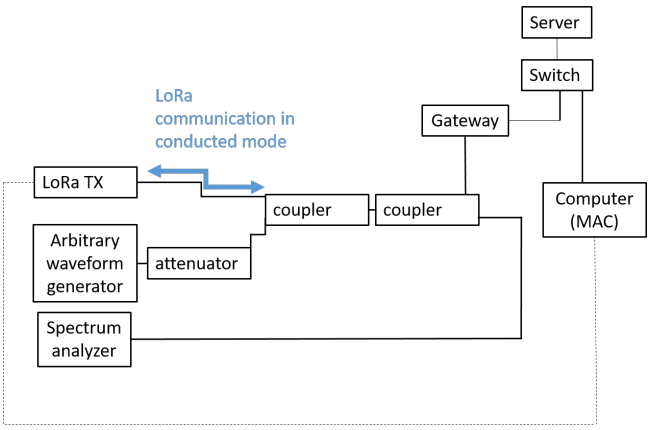


**Figure 1.** LoRaWAN communication deployed in a conducted bench.

way configurated with a bandwidth of 125 kHz and a center frequency of 868.3 MHz. The LoRa communication is implemented using a Spreading Factor of 7 and a Coding Rate of 4/5. The signal analyzer used for the spectra acquisition is the PXA Agilent N9030 A. It is configured with a frequency range of 10 MHz, a center frequency of 868 MHz,

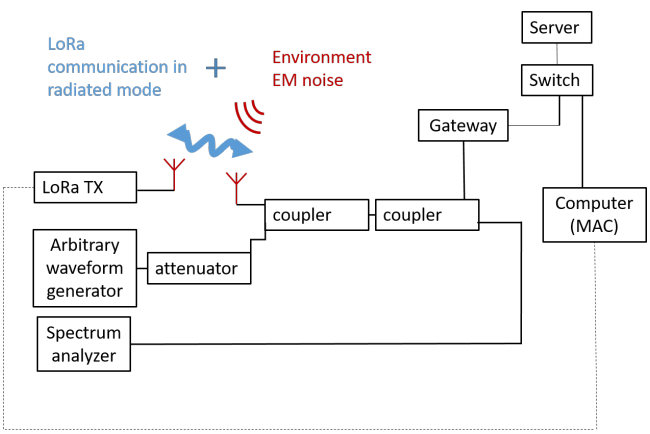


**Figure 2.** Radiated LoRaWAN communication bench.

a resolution bandwidth of 100 kHz. The interference signal is generated by an arbitrary waveform generator Tektronix AWG70001A linked to an variable attenuator and a 27.5 dB amplifier.

### 2.2 Interference signal

We generate three types of interference, two jamming signals, and a transient signal. Different types of jamming signals can be used [3], but the majority of commercial jammers produces a cosine wave sweeping a frequency band $[f_1, f_2]$ over a period of time $T$ and can be given by

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

with $A$ the amplitude of the interference signal. The first interference signal that we are considering sweeps the frequencies between $[840; 980]$ MHz in 5 $\mu s$ (see Figure 3). This frequency band is generally covered by the jammers designed to disturb the 2G and 3G communications and it also includes the LoRa frequency band. The second jam-

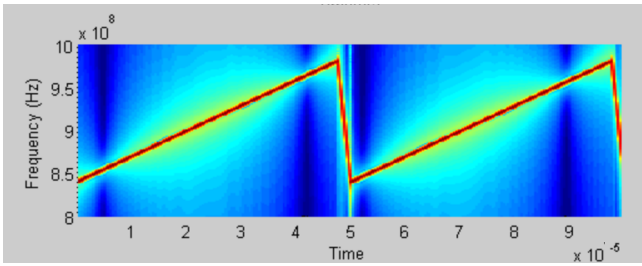


**Figure 3.** Representation of the jamming signal (840-980 MHz): time vs frequency.

ming signal that we are considering sweeps the frequencies between $[863; 873]$ MHz, corresponding to the LoRa frequency band, in 5 $\mu s$ (see Figure 4). It emulates a jamming signal which is specifically designed to disturb the LoRa Communications.

Finally, we consider certain interference that is specific to railway environment. We emulate, by a transient signal (see Figure 5), the interference generated by the contact losses between the catenary and the train pantograph. The transient signal is a broadband signal covering the frequencies

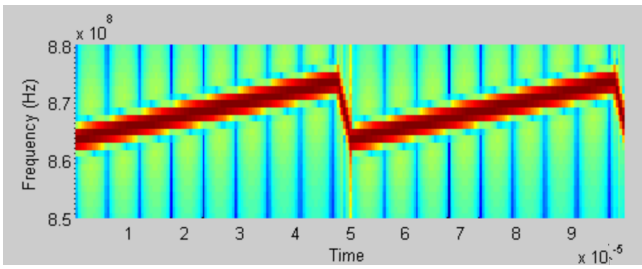


**Figure 4.** Representation of the jamming signal (863-873 MHz): time vs frequency.

between [863;873] MHz and is emitted by sequence with an average 5 $\mu s$ time interval (see Figure 6). The model of this transient interference is given in [4].

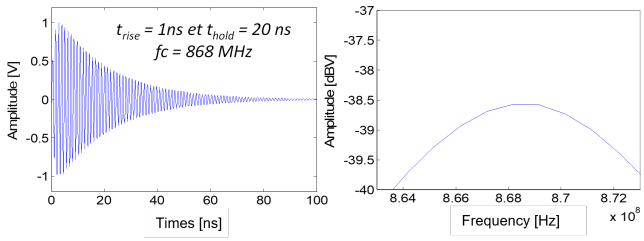


**Figure 5.** Representation of a transient signal.

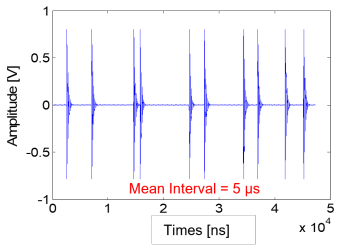


**Figure 6.** Transient signal sequence.

## 3 Characterization and detection results

The data recovered during the experiments are spectra covering the 10 MHz frequency band dedicated to the LoRa communications. For each spectrum, the power level on 801 frequency points within this bandwidth is measured. The results presented are obtained by studying these spectra. We record the spectra obtained with conducted and radiated LoRaWAN communications, using respectively the test benches of the figures 1 and 2. 99 spectra for each measurement case are recorded. The measurements scenarios are: **(A)** a communication without interference, **(B)** a communication disturbed by a jamming signal sweeping the frequencies between [840;980] MHz with an attenuation of 60 dB, 50 dB and 40 dB, **(C)** a communication disturbed by a jamming signal sweeping the frequencies between [863;873] MHz with an attenuation of 60 dB, 50 dB and 40 dB, **(D)** a communication disturbed by a transient signal which appeared with an average time interval of 5 $\mu$s with an attenuation of 0 dB, 10 dB and 20 dB.

### 3.1 Comparison of conducted and radiated communications

To visualize the different spectra, we transform the space of representations of the spectra by conducting a principal component analysis (PCA). To keep a maximum of variability on a minimum of axis, we represent the spectra on the two eigenvectors (also called components) associated with the two highest eigenvalues. These eigenvalues and eigenvectors are obtained from the matrix of correlations of the spectra obtained for the conducted communication (see Figure 7). To be able to compare the spectra obtained

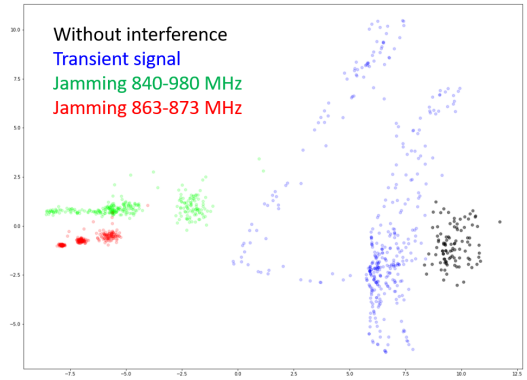


**Figure 7.** Projection of the spectra on the first two components: conducted case

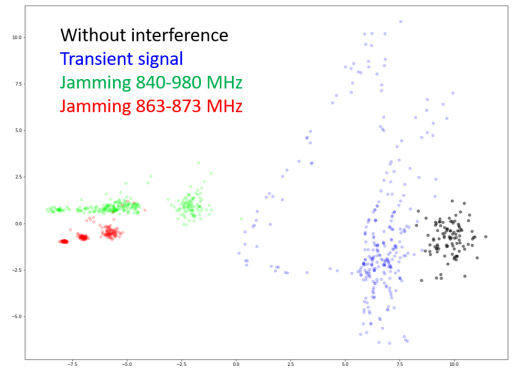


**Figure 8.** Projection of the spectra on the first two components: radiated case

so far with those obtained in the radiated mode, we project these latter on the eigenvectors obtained in the case of conducted communication. The result can be seen in Figure 8. By comparing the dispersion of the projections on the first two components, we notice that the variability caused by the surrounding EM noise is not large enough to change the characteristics of the communication. Whether through the study of conducted communication or radiated communication, the characteristics of the spectra remain unchanged, and the difference between each configuration of communication are not mixed. This validates our approach to build a model based only on the conducted mode, since it shows good generalization ability in radiated mode. However, this should be confirmed by further on-site measurement campaigns.

### 3.2 Detection

In this section, we aim to construct a classification model to detect the studied interference. The built model is based on the characteristics of LoRa communications observed at the physical layer. In fact, using the data collected in conducted mode permits to delete the surrounding noise on the observation. Then, we check the validity of the model on data obtained in radiated mode (with the presence of surrounding noise).

**Table 1.** Confusion matrix of the prediction of the testing set.

| label \| prediction | A | B | C | D |
|---|---|---|---|---|
| A | 25 | 0 | 0 | 0 |
| B | 12 | 27 | 0 | 0 |
| C | 10 | 0 | 65 | 0 |
| D | 5 | 0 | 0 | 70 |
| error | 12.4% | | | |

**Table 2.** Confusion matrix of the prediction of the wireless records.

| label \| prediction | A | B | C | D |
|---|---|---|---|---|
| A | 99 | 0 | 0 | 0 |
| B | 56 | 241 | 0 | 0 |
| C | 71 | 0 | 226 | 0 |
| D | 32 | 0 | 0 | 265 |
| error | 16.1% | | | |

To set up the learning process, we separated the conducted communication recording into two data sets. A training set comprises 75% of the data and a test set comprising the remaining 25%. Both data sets are obtained by proportional random draw without replacement. Due to the separability of the different configurations, the classification algorithm used for the detection must be non-linear and requires a separator hyperplane. For these reasons, the Support Vector Machine classification (SVMc) [5] is the algorithm retained for the detection system. After testing different kernels, the selected kernel is the radial basis function (RBF). Since each configuration is represented in the same proportion, the cost matrix is fixed to 1 (so the parameter C of the algorithm is set to 1). The width of the RBF kernel is selected using the sigma estimation function (sigest) [6]. The classification learns to predict the four classes presented before.

The result of the detection for the test set is presented in Table 1 and in Table 2 for the radiated communication bench. The model correctly predicts the majority of configurations; it has an error of 12.4 % on the test set and 16.1 % on the data from wireless communications. The stability of the model confirms the similarity of the spectra obtained by the two bench configurations. By looking at the type of error brought by the model, we notice that the errors relate only to undetected interference, and more precisely, these spectra relate to times when no communication is established (due to the small amount of data transmitted).

## 4 Conclusion

This paper focuses on the conception of a monitoring system able to detect and classify jamming and transient signals. In this study, the LoRaWAN communications and this interference are conducted using a test bench. To ensure that the detection model only learns the characteristics of the interference, the learning is performed on data obtained from a conducted test bench. The resulting model shows promising results in predicting attacks and loses very little quality when it comes to detect attacks from a test bench with a radiated communication mode. The results obtained in radiated mode constitute a first step in the validation of the model. In order to confirm these results, we will then study the validity of the model under various surrounding conditions.

## 5 Acknowledgements

This work was performed in the framework the LoRa-R project which is co-financed by the European Union with the European Regional Development Fund, the Hauts de France Region Council and the SNCF railway company.

## References

[1] L. Vangelista, A. Zanella, M. Zorzi, (2015, *Long-range IoT technologies: The dawn of LoRa*™, In Future access enablers of ubiquitous and intelligent infrastructures, Springer, pp. 51-58, 2015.

[2] V. Deniau, T. Vantrovs, N. Becuwe, C. Gransart, A. N. De Sao Jose, A. Boe, E. P. Simon, O. Vmamynck, F. Valenti, J. Villain, Q. Rivette, *Analysis of the Susceptibility of the LoRa Communication Protocol in the Railway Electromagnetic Environment*, In 2021 XXXIVth General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS), pp. 1-4, IEEE, 2021.

[3] 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. 1625–1633, 2017.

[4] G. Romero, V. Deniau, E.P. Simon, C. Gransart, *Evaluation of the Transient EM Interferences Impact on the Clear Channel Assessment in Wi-Fi Communications*, 2020 XXXIIIrd General Assembly and Scientific Symposium of the International Union of Radio Science, pp. 1-4, 2020.

[5] V. Vapnik, *The nature of statistical learning theory*, Red Bank: Springer, vol. 2, 2000.

[6] B. Caputo, K. Sim, F. Furesjo, 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.