IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 2, 2020, accepted April 23, 2020, date of publication May 4, 2020, date of current version May 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992119

SCER Spoofing Attacks on the Galileo Open Service and Machine Learning Techniques for End-User Protection

FRANCISCO GALLARDO[1] AND ANTONIO PÉREZ YUSTE[2], (Senior Member, IEEE)

[1]DLR GfR mbH, Universidad Politécnica de Madrid, 28040 Madrid, Spain

[2]Department of Communications and Audio and Video Engineering, Universidad Politécnica de Madrid, 28040 Madrid, Spain

Corresponding author: Francisco Gallardo (francisco.gallardo@dlr-gfr.de)

This work was possible due to the agreement between DLR GfR mbH and the Technical University of Madrid for the development of an industrial PhD for researching SCER OS-NMA anti-spoofing protection techniques.

ABSTRACT Spoofing attacks pose a clear cybersecurity risk for all systems relying on Global Navigation Satellite Systems (GNSS) for time synchronization or positioning. Secure Code Estimation and Replay (SCER) spoofing attacks are the most challenging type of spoofing attacks, as these may be problematic even for future GNSS protection systems, like Navigation Message Authentication (NMA) or Spreading Code Authentication (SCA). This is one of the reasons that make the development of complementary protection techniques, like the one proposed in this work, necessary. In the first part of the paper, the spoofing SCER attacks are analyzed in detail for GPS and, particularly, for Galileo. The role of the Galileo Pseudorandom Noise (PRN) intra-satellite non-orthogonality distortion term in hindering the attacks is discussed and a detailed comparison between GPS and Galileo expected quality curves for the SCER attack is provided. A complementary detection method for end-user receivers (assuming NMA is used) against SCER attacks is proposed, based on the application of machine learning and a proposed set of features extracted from the receiver search space, assuming the attacker was not able to null the satellite signal.

INDEX TERMS Cybersecurity, Galileo, GNSS authentication, GNSS security, machine learning, SCER.

I. INTRODUCTION

A cryptographic protection system for the Galileo Open Service Navigation Message of the E1B signal is currently under development, based on TESLA (Timed Efficient Stream Loss-tolerant Authentication) protocol. It is expected to be available by 2020 [1]. The TESLA protocol is a symmetric cryptographic system that provides some level of asymmetry by means of a delayed provision of keys [2].

The Galileo Open Service signature solution for E1B, known as the Open Service Navigation Message Authentication (OS-NMA), is intended to protect GNSS users against attacks based on generating false GNSS signals. This technique is called Spoofing. There are two main groups of spoofing attacks, as detailed in [3]: Based on the source of the GNSS signal:

1) Simplistic Attack: A GNSS simulator is used to generate the false GNSS signal used in the attack.
2) Meaconing: Recording and rebroadcasting a GNSS signal while adding a time delay, with the intention of diverting the real position of the victim.

And based on the used resources:

1) Intermediate Attack: This attack implies knowing the victim's receiver antenna's position and velocity. This is required to properly place the counterfeit signals with respect to the real signals, at the victim's search space. In order to do so, the attacker will be receiving the real signals from the actual satellites.
2) Sophisticated Attack: This attack is conceived to overcome defenses based on the Angle Of Arrival (AOA) of the received signals. It implies the use of several Spoofers with a common oscillator. All of these Spoofers will use the real satellite signals, as in the case of the Intermediate attack.

Other types of attacks are described in [4], like the Selective Delay attack, consisting in the isolation of each spacecraft signal components (e.g. by the use of directive antennas tracking each satellite) and the addition of extra delays to each

The associate editor coordinating the review of this manuscript and approving it for publication was Ana Lucila Sandoval Orozco.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

signal components. For other definitions of Spoofing attacks, please refer to [4].

Regardless of the used sources, the simplistic attack, which relies on using a signal generator to create the fake signal, does not imply any knowledge of the original Navigation Message, so it should be prevented by any authentication technique like NMA. Meaconing, on the other hand, since it implies the rebroadcasting of a real signal, makes such protection, in some cases, unsuccessful. Nonetheless, this type of attacks could be detectable by the victim with a trustable time source, if the time delay introduced by the spoofer is big enough [3]. This imposes an upper limit to the allowed delay added by the spoofer. In order to be able to control the victim's Position-Velocity and Time (PVT), the spoofer will need to add different delays to each satellite signal; this forces the attacker to estimate the symbols transmitted by the satellite (particularly the unpredictable symbols). Here, we are assuming that the spoofer does not have enough resources to use isolated channels, including antennas and RF equipment, per satellite. This particular situation is evaluated in Section II-E.

The spoofer approach of estimating the unpredictable symbols transmitted by the satellite and synthesizing a fake signal based on the estimated symbols is known as Secure Code Estimation and Replay (**SCER**).

Therefore, if we assume that breaking the cryptographic security is impossible for an attacker, the only available solution would be the estimation of the real unpredictable symbol while it is being transmitted by the satellite and adding this information to the fake signal. Each GNSS can support their users protection by including unpredictable symbols in the Navigation Message, like those of the Galileo OS-NMA. There is currently an active discussion in the space industry on the NMA role in protecting users against SCER.

The paper is structured as follows:

1) In Section II, the SCER attack for both Galileo and GPS is reviewed in detail, providing comparative results. The Intra-satellite PRN non-orthogonality distortion term is also defined.
2) In Section III, The Intra-satellite PRN non-orthogonality distortion term's impact on the SCER on Galileo NMA is analyzed. A case study, centered in Galileo OS-NMA SCER attacks simulations is considered.
3) In Section IV, a Spoofing detection complementary technique, applicable for NMA and SCA, but based on the fact that NMA is used, is proposed.
4) In Section V, a number of different machine learning algorithms are analyzed and their expected accuracies are presented based on simulations.
5) In Section VI, the expected conclusions are discussed.

Note that the suggested detection method in Section IV, relies on the use of NMA and on the fact that the Spoofer was not able to null the original signal. If the conditions and type of attacks defined in Section II are not met, then it is impossible to ensure that the Navigation Message was not modified, making the use the detection method used in Section IV risky.

II. SCER ATTACK

Two different types of SCER attacks are considered, from the point of view of the delay [3]:

1) Zero-latency SCER attack: The delay of the spoofed signal is considered to be 0 at the beginning of the attack and then gradually increased, avoiding effects easily noticeable in the tracking loops of the victim.
2) Non-zero-latency SCER attack: A significant delay is present in the spoofer-generated signal. In order to avoid being detected, due to the tracking jumps in the victim's receiver, at the beginning of the attack, the spoofer may try to generate jamming signals that could temporarily ''blind'' the victim's receiver.

It is also true that it will be impossible to perform a zero-latency SCER attack when the signal arrives to the victim first. This particular point is also analyzed in [4], suggesting the idea of transmitting any symbol value until the necessary number of samples are processed by the matched filter, getting at that instant a good estimate of the unpredictable symbol. This issue will depend greatly on the geometry of the satellites constellation and the arrangement of the victim and the spoofer.

A. BAYESIAN ESTIMATORS

Following a similar approach to the one described in [5], it seems reasonable to assume that the spoofer will use some sort of Bayesian estimator to determine the value of the unpredictable symbol transmitted by the satellites. These Bayesian estimators are based on the output of a matched filter which can be modeled, during a single symbol of the unpredictable pattern, as:

$$Z_l\left(n\right)=\frac{2}{n}\sum_{k=k_l}^{k_l+n-1}Y_k s_k \tag{1}$$

where Y_k is the sampled sequence of the signal received by the spoofer and s_k is the sampled sequence of the local replica of the signal, generated by the spoofer. Note that n indicates the number of samples of the unpredictable symbol used for the estimation, while k_l represents the first sample to be used for the integration. $Z_l\left(n\right)$ is the output of the matched filter after processing ''n'' samples. It is up to the Spoofer to determine what sampling frequency should be used, as long as it meets the needed sampling frequency recommended for the different GNSS. For the results derived in Section V, a sampling frequency of 50 MHz was used.

At this stage, it is assumed that the Spoofer performed a good estimation of the signal delay $\left(\hat{\tau}\right)$ and the Doppler frequency $\left(\hat{f_{dop}}\right)$ by means of acquisition and tracking blocks.

An analysis of the random variable that can be found at the output of the matched filter for Galileo, and GPS will be analyzed later.

As decribed in [5], the proposed Bayesian estimators (MAP, ML, MMSE) are used to provide a real value to replace the unpredictable binary symbol, instead of picking one out from the two binary values of the unpredictable symbol. As in [4], we will follow the approach of using the MAP estimator, using the sign of the output of the matched filter, as described in [5].

B. GPS SIGNAL

1) GPS SIGNAL MODEL WITH UNITARY POWER IN INTERMEDIATE FREQUENCY

As considered in [5], the received GPS L1 C/A signal can be modeled as follows in Intermediate Frequency (IF):

$$Y_k' = w_k c_k cos\left(2\pi f_{IF} t_k + \theta_k\right) + N_k \tag{2}$$

where c_k is the NRZ (Non-Return to Zero) Spreading code, w_k is the estimate of the NRZ unpredictable symbol, θ_k is the carrier phase and f_{IF} is the Intermediate Frequency. N_k is the AWGN (Additive White Gaussian Noise) at the input of the receiver.

In order to allow later comparison with Galileo, we will consider unitary power:

$$Y_{K_{GPS-IF}} = \sqrt{2} w_k c_k cos\left(2\pi f_{IF} t_k + \theta_k\right) + N_k \tag{3}$$

Therefore, defining the matched filter as:

$$Z_l\left(n\right)_{GPS-IF} = \frac{\sqrt{2}}{n} \sum_{k=k_l}^{k_l+n-1} Y_{k_{GPS-IF}} s_k \tag{4}$$

where s_k is the sampled sequence of the GPS local copy signal in the spoofer receiver.

This leads to:

$$E\left[Z_l\left(n\right)_{GPS-IF}\right] = W_L \tag{5}$$

$$Var\left[Z_l\left(n\right)_{GPS-IF}\right] = \frac{\sigma^2}{n} \tag{6}$$

where W_L is the true value of the NMA unpredictable symbol.

2) GPS SIGNAL MODEL WITH UNITARY POWER IN BASE BAND (BB)

We will also consider the GPS L1 C/A signal in Base Band (BB). Then, the received GPS L1 C/A signal can be modeled as follows:

$$Y_k' = w_k c_k + N_k \tag{7}$$

Therefore, defining the matched filter as:

$$Z_l\left(n\right)_{GPS-BB} = \frac{1}{n} \sum_{k=k_l}^{k_l+n-1} Y_{k_{GPS-BB}} s_k \tag{8}$$

Which leads to:

$$E\left[Z_l\left(n\right)_{GPS-BB}\right] = W_L \tag{9}$$

$$Var\left[Z_l\left(n\right)_{GPS-BB}\right] = \frac{\sigma^2}{n} \tag{10}$$

Note that both results are the same in terms of expectation and variance, regardless whether we consider the GPS L1 C/A signal in IF or BB.

As it is stated in [5], equation(1) (or in equations (4) or (8)) can be used to estimate the value of the unpredictable symbol. Depending mainly on the received C/N_0, we can obtain a good estimation of the unpredictable code symbol under analysis after 6 μs of integration.

The chipping period in GPS L1 is approximately 0.978 μsec. Leading to the evaluation of the signal during less than 6 Chips [6].

If a Spoofer is trying to perform a SCER attack, using a single antenna to receive all the signals, a linear combination of different satellites signals will be present at the input. Those signals will be modulated with different spreading codes, which will be orthogonal among each other. This means that, in order to estimate the symbol, it will be necessary to evaluate the signal in an interval big enough, to start taking advantage of the sequences orthogonality. A way to overcome this delay (imposed by the sequences orthogonality not present in the very short term) could be using directional antennas in order to provide extra gain to the signal coming from the satellite under evaluation (note that it will be necessary to use several antennas in order to track different satellites), then the differential delay needed for properly controlling the victim's position could be performed by means of having isolated channels -one per satellite-, and applying differential delays to each channel.

In this attack, we assume that the attacker is close to the victim and a mobile environment is considered due to the remarks in [7], regarding detecting the spoofing attack based on the channel behavior. This will imply that the antennas gain could not be extremely big. If this assumption does not hold, then the spoofer could consider not regenerating the signal but just using a channel per satellite, applying a differential delay as needed. See Section II-E for further considerations on this type of attack.

C. GALILEO SIGNAL

As per [8] and [2], the Galileo E1B Signal will include a NMA based on TESLA, providing with unpredictable symbols, forcing an attacker to follow the SCER schema. We can define the E1 (excluding PRS (Public Regulated Service)) Galileo signal as (based on [9], using the syntax from [5]):

$$GAL_k = \frac{1}{\sqrt{2}}\left(w_k e_{1B_k} sub_B - e_{1C_k} sub_C\right) \tag{11}$$

$$sub_B = \left(\alpha SC_{E1B,a_k} + \beta SC_{E1B,b_k}\right) \tag{12}$$

$$sub_C = \left(\alpha SC_{E1C,a_k} - \beta SC_{E1C,b_k}\right) \tag{13}$$

where w_k is the estimate of the NRZ unpredictable symbol, the e_{1B_k} term is the NRZ PRN sequence for E1B, e_{1C} is the NRZ PRN sequence for E1C, $\alpha = \sqrt{\frac{10}{11}}$ and $\beta = \sqrt{\frac{1}{11}}$.

The $SC_{E1A|B,a|b_k}$ is defined as follows:

$$SC_{E1A|B,a|b_k} = sign\left(sin\left(2\pi t_k R_x\right)\right) \tag{14}$$

where $R_a = 1.023MHz$ and $R_b = 6.138MHz$.

We can consider the generation of the full signal (both E1B and E1C) for the local copy in the receiver, or just the E1B part [10].

The received signal will also contain Additive White Gaussian Noise (AWGN), therefore the spoofer will receive $Y_k = GAL_k + N_k$, where $N_k \approx \mathcal{N}\left(0, \sigma^2\right)$.

As for the GPS case, we will also make some general considerations regarding the output of the matched filter in IF and BB. We will consider, for now, that only the E1B PRN is present in the local copy of the spoofer receiver and that a large number of samples are taken for integration ("n" is large), although we will later consider these points in detail, in Section II-D1 and Section II-D2 for BB.

1) GALILEO SIGNAL MODEL WITH UNITARY POWER IN INTERMEDIATE FREQUENCY

The received Galileo signal can be modeled as follows, in Intermediate Frequency (IF):

$$\begin{aligned} Y_{k_{GAL-IF}} = & \left(w_k e_{1B_k} sub_B - e_{1C_k} sub_C\right) \\ & cos\left(2\pi f_{IF} t_k + \theta_k\right) + N_k \end{aligned} \tag{15}$$

Therefore, defining the matched filter as:

$$Z_l\left(n\right)_{GAL-IF} = \frac{2}{n}\sum_{k=k_l}^{k_l+n-1} Y_{k_{GAL-IF}} s_k \tag{16}$$

Leading to:

$$E\left[Z_l\left(n\right)_{GAL-IF}\right] = W_L \tag{17}$$

$$Var\left[Z_l\left(n\right)_{GAL-IF}\right] = \frac{2\sigma^2}{n} \tag{18}$$

Note that equations (17), (18), (21) and (22) are considering that the Spoofer is only using E1B PRN in the local copy in the receiver used to estimate the unpredictable symbol. For more details on this, please refer to Sections II-D1 or II-D2.

2) GALILEO SIGNAL MODEL WITH UNITARY POWER IN BASE BAND (BB)

The received Galileo signal can be modeled as follows in BB:

$$Y_{k_{GAL-BB}} = \frac{1}{\sqrt{2}}\left(w_k e_{1B_k} sub_B - e_{1C_k} sub_C\right) + N_k \tag{19}$$

Therefore, by defining the matched filter as:

$$Z_l\left(n\right)_{GAL-BB} = \frac{\sqrt{2}}{n}\sum_{k=k_l}^{k_l+n-1} Y_{k_{GAL-BB}} s_k \tag{20}$$

Leads to:

$$E\left[Z_l\left(n\right)_{GAL-BB}\right] = W_L \tag{21}$$

$$Var\left[Z_l\left(n\right)_{GAL-BB}\right] = \frac{2\sigma^2}{n} \tag{22}$$

Note that, here, we are considering that the Spoofer is only using E1B PRN for generating the local copy in the receiver, used to estimate the unpredictable symbol. For more details on this, please refer to Sections II-D1 or II-D2.

Therefore, Galileo is providing the same results for both BB and IF, and it is providing with higher variance than in GPS, leading to a predicted reduction in the effective C/N_0 of 3dB.

Based on these results (equations: (9), (10), (21) and (22)), assuming both symbols transmitted by the satellite have the same probability, then:

$$p_{e_{GPS}} = \frac{1}{2} erfc\left(\sqrt{\frac{C}{N_0} T_s n}\right) \tag{23}$$

$$p_{e_{GAL}} = \frac{1}{2} erfc\left(\sqrt{\frac{C}{2N_0} T_s n}\right) \tag{24}$$

Regardless whether we consider the signals in BB or IF. Therefore, for the sake of simplicity, we will further evaluate the output of the matched filter for Galileo in BB.

D. AUTHENTICATION TECHNIQUES AND SCER FOR GALILEO

1) GALILEO SIGNAL WITH FULL LOCAL COPY OF E1 WITH NMA

Let's first consider the full local copy, assuming that the spoofer already knows the delay and Doppler of the incoming signal (by means of previous acquisition and tracking stages). In such case, the spoofer is going to generate (25) as s_k in (1).

Where:

$$s_k = \left(e_{1B_k} sub_B - e_{1C_k} sub_C\right) \tag{25}$$

Which leads to (26).

$$Z_{l_{E_{full}}}\left(n\right) = \frac{1}{n}\sum_{k=k_l}^{k_l+n-1}\left[w_k - \iota_k + 1\right] + \frac{\sqrt{2}}{n}\sum_{k=k_l}^{k_l+n-1}\left[N_k s_k\right] \tag{26}$$

where (with the current Galileo SIS (Signal In Space) ICD (Interface Control Document) [9]):

$$\iota_k = e_{1C_k} e_{1B_k}\left(1 + w_k\right)\left[\alpha^2 - \beta^2\right] \tag{27}$$

Note that, as already mentioned, it is assumed that the Spoofer is able to perfectly align the local replica of the signal and the satellite signal. In this context, it has to be considered that the average of the products of both Galileo subcarriers with themselves is one, in order to get to the expression in (26).

In this case, we can identify the term ι_k which will be present in our matched filter output. The term ι_k, the intra-satellite **PRNs non-orthogonality distortion term**, will affect the spoofer estimation, as it will be discussed later in Section III-A.

This term could be eliminated by means of extending the evaluated signal length in the matched filter (the spoofer will have to wait until the inner product of both spreading sequences is close to 0).

The expectation of the output of the matched filter will be:

$$E\left[Z_{l_{E_{full}}}(n)\right] = W_L + 1 - E\left[\frac{9}{11}\left[W_L+1\right]\xi(n)\right] \tag{28}$$

where:

$$\xi(n) = \frac{1}{n}\sum_{k=k_l}^{k_l+n-1} e_{1C_k}e_{1B_k}$$

And the variance will be:

$$Var\left[Z_{l_{E_{full}}}(n)\right] = \frac{4\sigma^2}{n}\left[1-\frac{9}{11}\xi(n)\right] \tag{29}$$

where W_L is the true value of the unpredictable symbol.

This means that the output of the matched filter will follow a non-stationary Gaussian (note that the noise at the receiver input is AWGN). As the number of samples ("n") increases, the Gaussian expectation will tend to be stationary, and $Z_{l_{E_{full}}}$ will follow:

$$Z_{l_{E_{full}}}(n) \approx \mathcal{N}\left(W_L+1, \frac{4\sigma^2}{n}\right) \tag{30}$$

2) GALILEO WITH JUST E1B SPREADING CODE IN THE LOCAL COPY OF E1 WITH NMA

We will now consider a local copy with only the E1B spreading code. This option is interesting for the spoofer, compared to the full local copy, as it is not so computationally expensive.

We assume, again, that the spoofer already knows the delay and Doppler of the incoming signal and the estimations are perfect. We will denote s_k' to the local copy of the signal, properly aligned to the incoming signal and without the E1C PRN.

Then:

$$Z_{l_{E_{partial}}}(n) = \frac{\sqrt{2}}{n}\sum_{k=k_l}^{k_l+n-1} s_k' Y_{k_{GAL-BB}} \tag{31}$$

where $s_k' = \left(e_{1B_k}sub_B\right)$

$$Z_{l_{E_{partial}}}(n) = \frac{1}{n}\sum_{k=k_l}^{k_l+n-1}\left[w_k - \iota_k'\right] + \frac{\sqrt{2}}{n}\sum_{k=k_l}^{k_l+n-1}\left[N_k s_k\right] \tag{32}$$

where (with the current Galileo SIS ICD [9]):

$$\iota_k' = e_{1C_k}e_{1B_k}\left(\alpha^2-\beta^2\right) \tag{33}$$

Note that, as already mentioned, it is assumed that the Spoofer was able to perfectly align the local replica of the signal and the satellite signal. In this context, it has to be considered that the average of the products of both Galileo subcarriers with themselves is one, in order to get to the expression in (32).

In this case, we can identify the term ι_k', which will be present in our matched filter output. By evaluating the output of the matched filter over a long-enough period, we can eliminate this term, called the **intra-satellite PRN non-orthogonality distortion term**.

The expectation of the output of the matched filter will be:

$$E\left[Z_{l_{E_{partial}}}(n)\right] = W_L - E\left[\frac{9}{11}\xi(n)\right] \tag{34}$$

where:

$$\xi(n) = \frac{1}{n}\sum_{k=k_l}^{k_l+n-1} e_{1C_k}e_{1B_k}$$

And the variance will be:

$$Var\left[Z_{l_{E_{partial}}}(n)\right] = \frac{2\sigma^2}{n} \tag{35}$$

where W_L is the real value of the unpredictable symbol.

As the number of samples ("n'') increases, the output of the matched filter will follow a Gaussian with stationary expectation.

$$Z_{l_{E_{partial}}}(n) \approx \mathcal{N}\left(W_L, \frac{2\sigma^2}{n}\right) \tag{36}$$

3) SCER ON GALILEO WITH SPREAD CODE AUTHENTICATION (SCA)

The Spread Code Authentication (SCA) is a protection method which is currently under discussion for its future implementation in Global Navigation Satellite Systems (GNSS). Currently, an implementation for GPS known as "Chips-Message Robust Authentication" (CHIMERA) is under evaluation [11].

The techniques, based on including the unpredictable signature in the Spreading Codes, can be considered as an evolution of the techniques based solely on the use of unpredictable symbols in the Navigation Message (like the NMA).

As detailed in [12], the SCER attack on SCA protection approach relies on a similar method to the one presented in Section II, namely: applying Bayesian estimators to the output of a matched filter, then, after comparing against a threshold, the polarity of the unpredictable chip is obtained. The main difference between both cases is that, in the NMA, the Spoofer can use longer integration times, while in the SCA case, the integration has to be limited to the chip length. Note that, in the NMA case, the result was the unpredictable symbol polarity, while in the SCA the chip polarity is obtained. This implies that the symbol must be previously known by the attacker, either because the symbol is unpredictable (e.g. NMA) and is firstly estimated (implying a larger delay to start transmitting the false signal) or because the symbol is predictable and well known in advance. Note that if the spoofer is not able to include the proper polarity values in the unpredictable chips, then, when the end user (the victim) came to know, via the Navigation Message, the actual chip authentication sequence, correlation losses will be present, so making the SCA test fail. Moreover, the use of NMA in combination with SCA can make the GNSS systems even more secure since it ensures that the Navigation Message has not been tampered. The Navigation Message will allow us to determine the real values of the unpredictable chips, hence this point is of high importance.

Every time the attacker needs to estimate a chip, (1) will be used. Assuming the positions of the unpredictable chips are known by the attacker and assuming the unpredictable chips are included in the E1B spreading code, then the following local copy would be used:

$$s_{kSCA} = (w_k sub_B) \tag{37}$$

Note that the design decisions on the distribution of the unpredictable chips (and whether this information is made available to the users in advance or not) will also have an impact in the SCER attack, as described in [12].

Although operations performed by the Spoofer in the SCA case are very similar to those presented in Section II, the expectation result will differ sightly, unlike in (34), the distortion term will not depend on e_{1B_k} but on the transmitted symbol and the pilot spreading code.

$$E\left[Z_{l_{E_{partial_{SCA}}}}(n)\right] = e_{1B_k} - E\left[\frac{9}{11}\xi_{SCA}(n)\right] \tag{38}$$

where:

$$\xi_{SCA}(n) = \frac{1}{n}\sum_{k=k_l}^{k_l+n-1} e_{1C_k} w_k \tag{39}$$

The variance will be the same as in the NMA case:

$$Var\left[Z_{l_{E_{partial_{SCA}}}}(n)\right] = \frac{2\sigma^2}{n} \tag{40}$$

This implies that the spoofer will have to take the term $\xi_{SCA}(n)$ into account. For the Galileo E1B case, the chip length is approximately 1 μs, therefore the maximum integration time available for the Spoofer will be 1 μs. Taking the expectation and variance obtained into account, the theoretical expression in (24) and the Galileo curves presented in Fig. 1 are still applicable. These curves provide low detection probabilities, with integration times of 1 μs ($P_d < 0.6$) for C/N_0 between 35dBHz to 50dBHz, which are good C/N_0 for normal GNSS equipment.

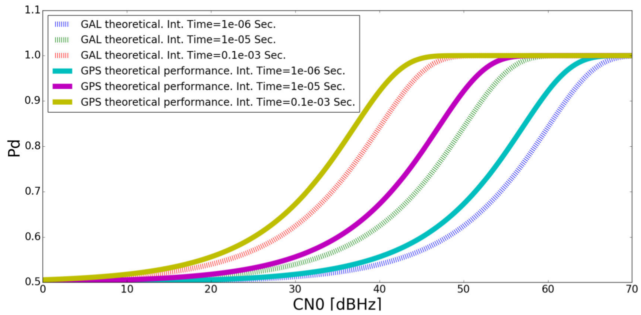


FIGURE 1. Detection Probability theoretical performance for Galileo E1 Signal compared to the theoretical GPS performance. As it can be seen, the Galileo modulation plays against the spoofer.

This implies that the SCA technique will be a good protection method against SCER, combined with NMA, as the SCER Spoofer detection probabilities will make the current attack approach very complicated: it will require working with very high C/N_0. On top of that, even in case that the attacker is able to work under such circumstances, the Galileo modulation will impose a distortion term, present in (38). Hence, if the spoofer is not able to estimate $\xi_{SCA}(n)$, it will have a negative impact on top of the already poor detection probability.

On the other hand, retro compatibility problems are yet to be fully understood for all types of GNSS users. Note that Galileo OS-NMA is just using spare bits in the current navigation message (therefore, it does not impact the expected performance or how users are currently using the Galileo signal). Nonetheless, SCA implies an initial reduction in the Auto Correlation Function (ACF) peaks that users will obtain in the search space. Suggestions on the use of the pilot channel to track the signal, while SCA unpredictable chips are being transmitted, are provided in [13]. GPS will require a significant change in their modulation to provide with an open service signal with a pilot channel. This implies that receivers manufacturers will anyway need to upgrade their receivers to track such new signals. Still, when applied to Galileo, this may imply changes on how receivers are tracking the Galileo signals, which are already providing with the pilot channels. Hence, a bidirectional discussion should be put in place between the Galileo programme and receivers manufacturers to fully understand the impact for end users of applying SCA. Moreover, SCA will impose some computing requirements to the end user's receivers, as it is necessary to store pre-correlation raw signal samples to be used at a later stage, when authentication message with the details on the used puncturing of the Spreading Code is received.

E. ONE SATELLITE - ONE ANTENNA - ONE CHANNEL CASE

We will also consider the scenario where the spoofer, instead of trying to estimate the symbol transmitted by the satellite, has enough resources to use a directive antenna per satellite (we will assume that the spoofer is able to track each satellite in the sky and the antenna directivity is such that the signals of the rest of the satellites are completely eliminated in the output of the antenna). The spoofer is assumed, then, to provide an independent channel per satellite. Instead of estimating the symbol, as described in the first part of the paper (Section I), the spoofer could use an independent antenna and channel to track each real satellite, leading to the application of a different delay per channel (hence, per satellite too). This will not require estimating each symbol.

If we assume that the Noise Figure of the spoofer $NF = 8dB$ (e.g. National Instrument USRP UBX Daughter Board, configured gain of 19.50dB, see [14]), then:

$$\frac{SNR_{NoSpoofer}}{SNR_{WithSpoofer}} = \frac{\frac{S_i}{BK300}}{\frac{S_i}{BK(300+1592.87)}} \tag{41}$$

$$\frac{SNR_{NoSpoofer}}{SNR_{WithSpoofer}} = 8dB \tag{42}$$

Implying that the spoofer may introduce, at least, an 8 dB sanction to the SNR of the generated signal, with respect to the case where the spoofer is estimating the symbol transmitted by the satellite. Note that the Spoofer will need to track each satellite and use a group of directive antennas and isolated channels. Such setup for a spoofer following the victim is not simple, although it represents a serious threat if the attacker has enough resources to use such a complicated setup.

Even if we assume that the spoofer has enough resources to use this setup, techniques like those proposed in Section IV can still protect critical infrastructure standoff victims.

III. CASE STUDY: THE INTRA-SATELLITE PRNs NON-ORTHOGONALITY DISTORTION TERM AND ITS ROLE IN HINDERING THE SCER ATTACKS TO THE GALILEO NMA

Different simulations were performed in order to study the influence the intra-satellite PRNs non-orthogonality distortion term could have in the SCER attacks on Galileo NMA. The NMA case was further evaluated, in order to better characterize the impact of SCER on the Galileo NMA which will be available by 2020, as per [1].

In terms of the PRN distortion in SCER for Galileo, the key differentiator, with respect to the GPS L1 C/A case, is that these PRN non-orthogonality distortion terms are caused by the **very same satellite** the spoofer is trying to falsify, while in the GPS case, the distortion will be caused only by other satellites. The Spoofer cannot get rid of the $\iota_k{}'$ distortion term in Galileo *even if only one single satellite is present at the receiver input*.

We can conclude here that the modulation of the Galileo system is making the Spoofer estimation of the symbol harder, which is good for the Galileo users. We will discuss the intra-satellite non-orthogonality distortion term ι_k and its impact in the SCER attacks, and we will compare it to the GPS case and its impact in the performance of the attack. Future work will also compare this intra-satellite effect to the effect due to the non-orthogonality between different satellite PRNs, for short integration times. The present work is only focused on the intra-satellite effect, as it is considered more daunting for Spoofers. Note that, regardless whether a very directive antenna is used (if possible) by the spoofer, the intra-satellite non-orthogonality distortion term effect will still be present, as it is introduced by the same satellite the spoofer is trying to use for the SCER attack. On the other hand, attenuating the signals coming from other satellites will reduce the effect between satellites.

The attacker has two main options available in order to overcome this problem appearing in the matched filter output due to the Galileo CBOC modulation in E1:

1) Extending the integration of the matched filter long enough, so the ι_k parameter tends to 0.
2) Estimating the PRN non-orthogonality distortion term (ι_k) parameter. This case will not be further detailed, but, taking into account different channel models and studies [15], [16], the channel will only be challenging for low elevation satellites in urban areas, particularly for unpredictable symbols that may be transmitted together [9]. Note that this may make the Spoofer estimation of the intra-satellite PRN non-orthogonality distortion term invalid, after some symbols were transmitted.

A. EXTENDING THE INTEGRATION OF THE MATCHED FILTER WITH NMA

In order to get rid of the distortion terms, the obvious way forward will be to extend the integration time of the matched filter. Clearly, if we extend the integration time to the symbol period, we will be maximizing the C/N_0 and we will completely eliminate the distortion terms. Nonetheless, the spoofer cannot wait until the end of the symbol period is reached. Instead, and depending on the received C/N_0, the spoofer will extend the integration time until a valid symbol estimation is available, based on an output, as clean as possible, of the matched filter.

It can be seen, in Fig. 1, that the theoretical result for Galileo is worse than the theoretical expression for GPS (complementary of (23)), mainly due to the Galileo modulation.

In Fig. 2 the results of the Galileo simulation are compared to the Galileo theoretical curve (43). The results in Fig. 2 were obtained by analyzing Galileo E1 simulated signal (1 second of data per C/N_0), generated with the workbench described in Section III-B. No quantization (signal generator was configured to work using directly float numbers), sampling frequency of 50MHz, spoofer working with a local copy with just E1B (as described in Section II-D2), no acquisition or tracking errors were included and MAP was used as the Bayesian estimator. One single satellite under analysis. A known pattern of alternating ones and zeros was used for error estimation.

$$P_d = 1 - \left(\frac{erfc\left(\sqrt{nT_s\left(\frac{C}{2N_0}\right)}\right)}{2} \right) \tag{43}$$

In (43), P_d is the Spoofer probability of detection of an unpredictable symbol. As it can be observed in the Fig. 2, the simulation confirms the 3dB reduction. The solid-line shows the theoretical P_d result for Galileo, derived from (43). The dotted-line shows the Galileo simulation results.

On the other hand, the non-orthogonality distortion term $(\iota_k{}')$ effect is obvious for high C/N_0 and short integration times (see Fig. 2, for high values of C/N_0 and short integration times: the results differ from the theoretical response for Galileo, the one in (43)). Nonetheless, the effect is smaller compared to the impact of the 3 dB reduction, with respect to GPS, due to the Galileo modulation. For more realistic values of C/N_0, like 53 dBHz and integration times of 1 μs or 10 μs, the difference between the simulated results probability of detection and the theoretical probability

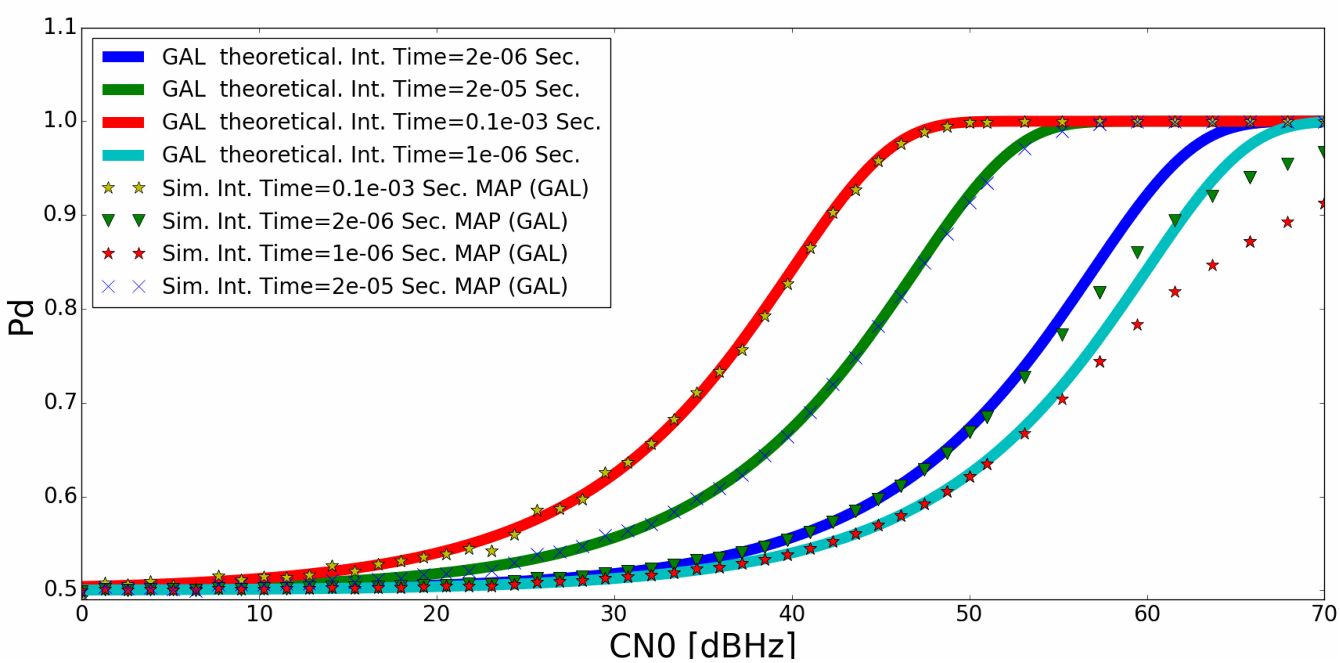


FIGURE 2. Impact of the Intra-satellite PRN non-orthogonality distortion term ($\iota_k{}'$) effect in the Detection Probability of the Spoofer. The effect is visible for low integration times with high SNR (lower right part of the figure). The simulation results divert from the Galileo theoretical results.

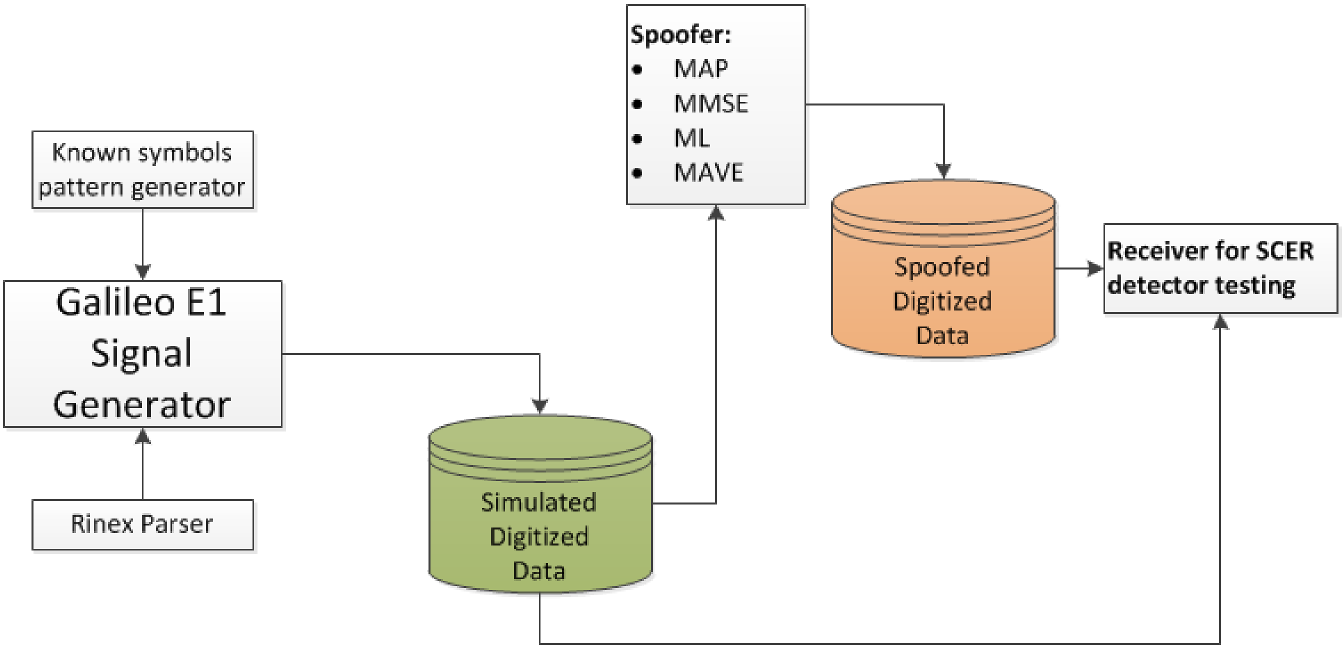

Spoofer:
• MAP
• MMSE
• ML
• MAVE
Known symbols pattern generator
Galileo E1 Signal Generator
Rinex Parser
Simulated Digitized Data
Spoofed Digitized Data
Receiver for SCER detector testing

FIGURE 3. Python workbench for SCER testing on Galileo E1.

of detection is about 1%. For higher values of C/N_0, like 67 dBHz and integration times of 1 μs, the difference is close to 10%. Therefore, this effect may not play a major role against SCER protection, although it could make the SCER spoofer work slightly harder. In Fig. 2, a comparison of the corrected theoretical Galileo P_d, (43), and the simulation results can be found. It can be seen how, for low integration times and high C/N_0, the simulation results divert more from the theoretical expected values (rightmost part of the figure, integration times of one and two microseconds).

Note that the effects of the Intra-satellite PRN non-orthogonality distortion terms are not modeled in the provided theoretical quality curves, as it only accounts for the complementary error function (erfc).

B. WORKBENCH FOR GALILEO SIGNAL SPOOFING

A complete workbench for testing the overall Galileo SCER approach was generated in Python, following the Galileo SIS ICD [9].

A signal generator was developed, capable of reading RINEX 3.0 files with Galileo Navigation Messages or including a known pattern of symbols that are considered unpredictable, as if TESLA completely unpredictable chains were used. This module is able to read the text files with the Galileo PRNs, annexed to [8] and receives an input from the user with the desired signal C/N_0, the simulated satellite name (as of now, only one satellite is simulated at once), the sampling frequency, the Navigation message content to be modulated (in hexadecimal format), and the length of the simulation.

The signal simulator generates a binary file with I/Q samples with the Galileo E1 signal, as requested by the user (sampling rate, number of bits for quantization, delay, Doppler, length of the resulting file, Galileo satellite PRN and SNR are configurable). The spoofer module performs the estimation of the symbol and allows the use of the three Bayesian Estimators, described in Section II-A.

The receiver module will allow the benchmarking of different spoofer detection methods, by means of adding modules in the victim's receiver.

At the time of this paper submission, only the Galileo Signal generator, the spoofer and the acquisition step of the receiver are fully developed. Nonetheless, for the analysis performed for this article, the workbench capabilities are sufficient, as it is not necessary to really generate the unpredictable symbol in order to analyze the performance of the different SCER spoofer estimators with the Galileo signal. Indeed, a controlled combination of 1 and 0 symbols were used, in order to have a reliable source to quickly determine whether the spoofer was wrongly estimating the symbol.

IV. COMPLEMENTARY MACHINE LEARNING TECHNIQUES FOR SCER PROTECTION

As described in Section II, the SCER attack is a risk for GNSS users, even those relying on techniques like NMA. Nonetheless, Galileo OS-NMA forces the attackers to use the SCER schema and prevents them to follow other approaches like modifying the navigation message. This implies that, if the attacker wants to divert the victim's PVT, it is mandatory to generate a fake signal (including the unpredictable symbols estimated from the real signal), with a different Doppler and/or delay.

Therefore, in the victim's Search Space (a very detailed analysis on the search space can be found in [15]), two correlation peaks will be found: one due to the spoofer signal and one from the original satellite signal, if the spoofer is present and if the spoofer was not able to null the original signal in the receiver input.

If the spoofer signal is superposed with the real signal and the navigation message was not modified by the attacker, then the effect on the victim will be negligible. If it is not superposed, then, assuming enough resolution in the search space is available, two separated peaks shall be present.

It is straightforward to conclude that a full branch of protection methods could rely on identifying abnormal search space distributions. We will evaluate the use of machine learning techniques to protect users against SCER attacks on Galileo OS-NMA, based on the analysis of features extracted from the search space.

Note that, if NMA techniques, analyzed in detail in Section II, are not used by the victim, the technique discussed in this Section will not, by any means, guarantee the navigation message was not modified. The detection method proposed in this Section is a complementary method to NMA, particularly designed against SCER on GNSS with NMA and only applicable if the original signal was not nulled.

The Search Space implemented in the workbench, defined in Section III-B and visible in Fig. 5, was calculated using the Parallel Acquisition in Time Domain method. The current method, as defined, is very heavy in terms of computing load. Future work will be focused on implementing a demonstrator and reducing the computing load. Note that the current resolution implies the use of powerful FPGAs implementing parallel correlators in order to generate the Search Space. The Spoofer signal was generated using the SCER method with MAP as Bayesian estimator, particularly using (1) with a local copy of E1B only, therefore the output of the Spoofer matched filter was following a random variable with expectation defined in (34) and variance defined in (35).

A. THE SEARCH SPACE WITH SPOOFER PRESENCE

Each cell of the Search Space is calculated by performing the following operation (based on syntax from [15]):

$$S\left(\tau, F_D\right) = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} r\left[n\right] c\left[n-\tau_D\right] e^{(-2j\pi F_D n)} \tag{44}$$

where $c\left[n-\tau_D\right]$ is the local copy used by the victim and $r\left[n\right]$ is the signal received by the victim's receiver (where τ_D is the delay used by the receiver in each search space cell, F_D is the Doppler frequency used by the receiver in each search space cell and N is the number of samples to be integrated to calculate each cell of the search space).

If any spoofer is present and the original signal was not nulled by the spoofer, then the received signal will follow:

$$r\left[n\right] = Y_{SAT_{GAL-BB}} + Y_{Spof\,GAL-BB} \tag{45}$$

where $Y_{SAT_{GAL-BB}}$ is the real Galileo signal, in baseband, with AWGN noise and $Y_{Spof\,GAL-BB}$ is the Spoofed generated signal, in baseband, with AWGN noise.

For the sake of simplicity, we will consider that the victim is using a local copy with only the E1C PRN (the Galileo pilot signal) for generating the search space.

Therefore, each search space cell in the victim's receiver will follow:

$$E\left[S\left(\tau_D, F_D\right)\right] = E\left[\frac{1}{N}\sum_{n=0}^{N-1} \Upsilon_{sat}\,\Omega_{sat}\,\Phi_{sat} + \Upsilon_{spf}\,\Omega_{spf}\,\Phi_{spf}\right] \tag{46}$$

where we have the terms coming from the real satellite signal:

$$\Upsilon_{sat} = e_{1C_{sat}} e_{1C}\left[n-\tau_D\right] \tag{47}$$

$$\Omega_{sat} = subc_{C_{sat}} subc_C\left[n-\tau_D\right] \tag{48}$$

$$\Phi_{sat} = e^{(-2j\pi F_D n)} e^{(2j\pi F_{sat} n)} \tag{49}$$

And those terms coming from the spoofed signal:

$$\Upsilon_{spf} = e_{1C_{spof}} e_{1C}\left[n-\tau_D\right] \tag{50}$$

$$\Omega_{spf} = subc_{C_{spof}} subc_C\left[n-\tau_D\right] \tag{51}$$

$$\Phi_{spf} = e^{(-2j\pi F_D n)} e^{(2j\pi F_{spof} n)} \tag{52}$$

And the variance of each cell will be:

$$Var\left[S_{real}\left(\tau, F_D\right)\right] = 2\frac{\sigma_{sat}^2 + \sigma_{spoofer}^2}{N} \tag{53}$$

where N is the number of samples used in the matched filter of (44). Then, it is quite straightforward to conclude:

1) If $F_{sat} = F_D$, $\tau_{sat} = \tau_D$ and $F_{spof} \neq F_D$, $\tau_{spof} \neq \tau_D$, then:

$$E\left[S\left(\tau, F_D\right)\right] = 1 \tag{54}$$

2) If $F_{sat} \neq F_D$, $\tau_{sat} \neq \tau_D$ and $F_{spof} = F_D$, $\tau_{spof} = \tau_D$, then:

$$E\left[S\left(\tau, F_D\right)\right] = 1 \tag{55}$$

3) If $F_{sat} \neq F_D$, $\tau_{sat} \neq \tau_D$ and $F_{spof} \neq F_D$, $\tau_{spof} \neq \tau_D$, then:

$$E\left[S\left(\tau, F_D\right)\right] = 0 \tag{56}$$

4) And if $F_{sat} = F_D = F_{spof}$ and $\tau_{sat} = \tau_D = \tau_{spof}$ then:

$$E\left[S\left(\tau, F_D\right)\right] = 2 \tag{57}$$

Note that the case in (57) will not pose a risk to the user at all, if the OS-NMA is used and the cryptographic protection is not broken (e.g. SCER attack). As the OS-NMA cryptographic protection is not broken and the Doppler and delay are the same as the ones of the authentic satellite, the victim's computed solution shall not differ with respect to the real one.

B. FEATURES EXTRACTION

As in any other machine learning problem, the first step is the feature extraction. We need to evaluate what information we are going to feed into the classification algorithms. The proposed features extraction is based on fitting the correlation peaks in the search space as 2D Gaussians and detecting RFIs during a time analysis window previous to the beginning of the signal sequence used for computing the search space, refer to Fig. 4 for details on this. As it is known [17], the ACF

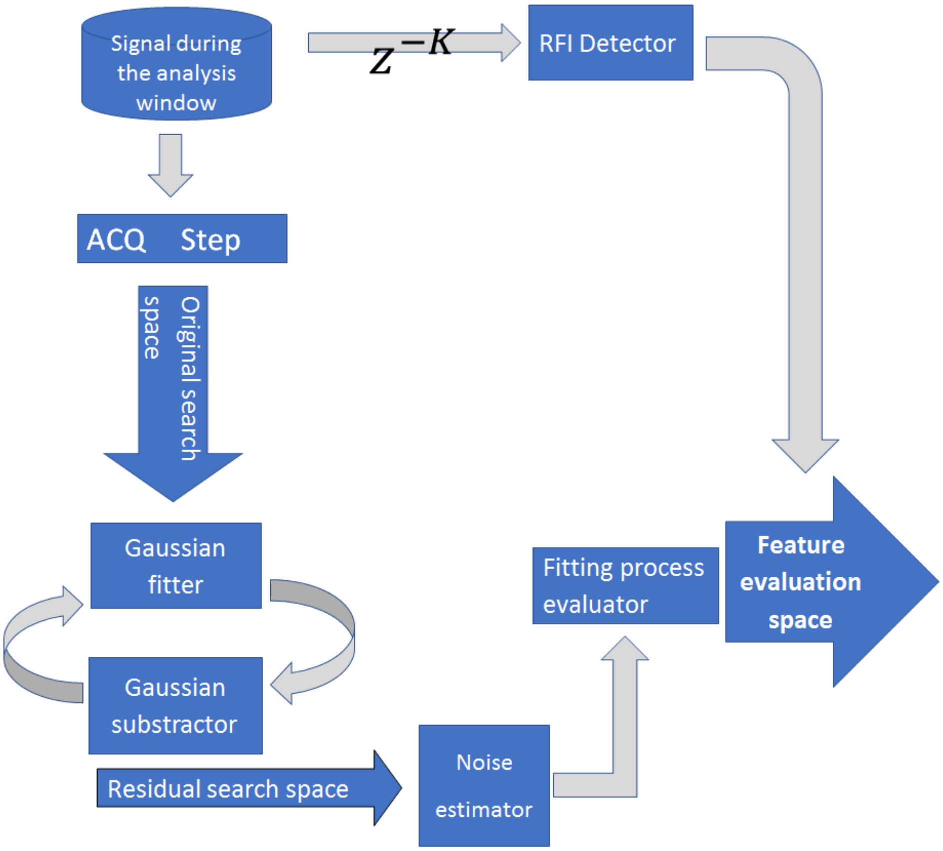


FIGURE 4. Overall features extraction process.

of the Galileo signal is not following a Gaussian waveform, although, for the purpose of increasing the computing efficiency and, at the same time, capturing the relevant features of the autocorrelation peaks, it is deemed sufficient for the purpose of detecting Spoofing signals. Further work will evaluate other waveforms that can further improve the computing efficiency, while retaining the necessary information for the classification algorithms.

1) GAUSSIAN EXTRACTION

In order to properly characterize the location and shape of the peaks in the search space, the algorithm will:

1) Adjust 2D Gaussians around the maximum peaks in the search space.
2) After successfully fitting a Gaussian in the search space, the fitted Gaussian is substracted.
3) Repeat the process N times.
4) Estimate the residual noise.

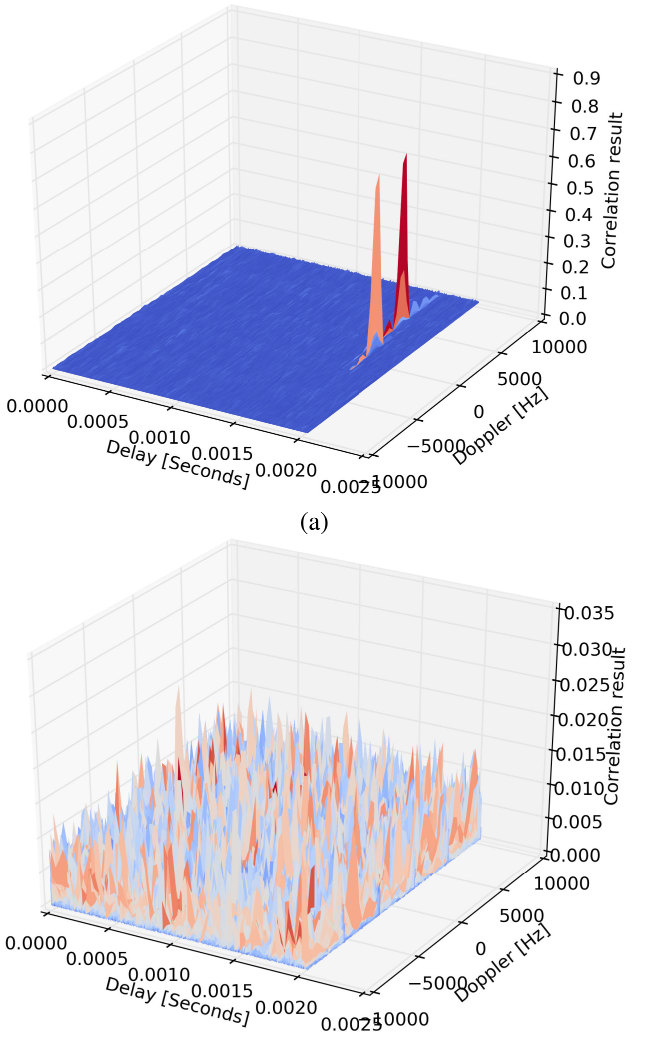
In the upper part of Fig. 5, the search space with the real satellite signal (Galileo E1) and the spoofer signal, generated using symbols estimated by using SCER with MAP Bayesian estimator, can be seen. No channel attenuation was introduced.

In the lower part of Fig. 5, the search space, after the Gaussian extraction process, can be seen. This is the resulting Search Space after applying the algorithm that can be seen in Fig. 4. As it can be appreciated, only the residual noise after the Gaussian subtraction remains in the search space. The value of this residual noise is also estimated and fed into the classification algorithms, so the Machine Learning techniques can have the information related to the relationship between the peak amplitude values and the noise in the Search Space. The workbench described in III-B was used. Note that the Fig. 5 is not showing one of the cases used for the algorithm training nor the exact same configuration of the workbench used for the testing of the machine learning algorithms, but just as an example to show the concept.

2) RFIs DETECTION

As described in Section II, attackers may try to blind the victim's receiver before starting the attack. Due to this reason, we will also look for RFIs during the time window previous to the reception of data used to generate the search space in the victim's receiver.

The initial analysis proposed in this paper is based on simply finding outliers, assuming that an Automatic Gain Controller (AGC) is present, although future work will be performed in order to use a more sophisticated RFI detection schema. The RFI presence will be a feature to be considered for the machine learning algorithms.



(b)

FIGURE 5. Search Space with Spoofer. Pre (a) and Post (b) Gaussian extraction.

V. CASE STUDY. SIMULATION WITH GALILEO E1 SIGNALS

In order to benchmark several machine learning algorithms, the data extraction method described in Section IV-B was implemented in the workbench from Section III-B. Different machine learning algorithms were analyzed using the workbench, based on Python Scikit-learn [18] library. Particularly: RBF SVM, Ada Boost, Decision Trees, Nearest Neighbors and Random Forests.

A. DATASET GENERATION

The datasets were generated using several combinations of Doppler shifts and time delays:

1) In the Spoofer case, for all the attacks, delay of 300 μsec and Dopplers of -5KHz and -2KHz were used.
2) For the real satellite the delays were of 200 μsec and 240 μsec. Dopplers: 5KHz and 10KHz.

These configurations were deemed sufficient for the used search space resolution, as the detection results were correct. The resolution in the victim's search space should always be high enough to properly allow the 2D Gaussian functions fitting.

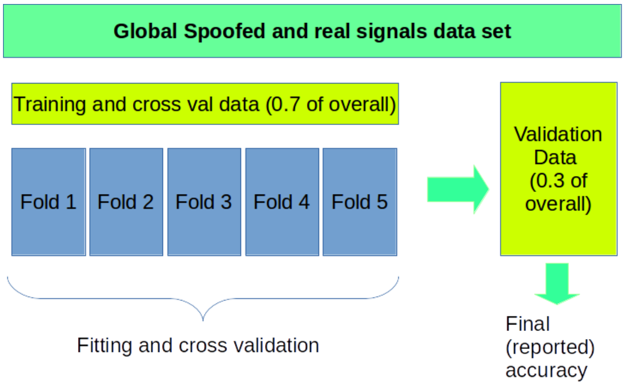

Global Spoofed and real signals data set
Training and cross val data (0.7 of overall)
Fold 1
Fold 2
Fold 3
Fold 4
Fold 5
Validation Data (0.3 of overall)
Fitting and cross validation
Final (reported) accuracy

FIGURE 6. Performed K-folds and reported accuracy. Based on [23].

Due to the reduced amount of peaks positions in the dataset, overfitting may occur. In order to rule out this possibility, the accuracy results were calculated, both feeding the positions to the algorithms and without feeding them with this information. This means that, in the Fig. 7, the results were obtained when the algorithm did not know where the peaks were located in the Search Space.

The used dataset was generated by applying the proposed feature extraction algorithm to datasets with Galileo signals from $C/N_0 = 0dBHz$ to $C/N_0 = 50dBHz$.

The integration time in the victim's receiver was 16ms and the local signal was only the E1C PRN. The dataset was composed of 381 cases with Spoofer, and 1074 without Spoofer. Note that the dataset was not balanced. This had a clear impact on the false alarm probability and the missed detection probability. It can also be seen in the F1 [19] scores in Fig. 10. Depending on the system final application in which the Machine Learning complementary protection algorithm will be deployed, it will be necessary to tailor the dataset to reduce the false alarm probability or the missed detection probability. In order to reduce the probability of missed detection of a particular class (e.g. Spoofer present in the received signal), firstly, such class should be over-represented in the input dataset and secondly, the algorithms should be evaluated to find the fitting parameters that maximize the accuracy and F1 scores for the class of interest. Confusion Matrices [19] should also be considered when evaluating the results. As it can be seen in other state-of-the-art techniques that rely on the analysis of the Search Space [20], the use of adaptive thresholds that depend on the location of the peak is already proposed. Nonetheless, the innovative and beneficial point of using Machine Learning techniques for the Search Space analysis is that these techniques allow the redefinition of the thresholds by just modifying the used training dataset. This approach will allow a reconfiguration of an operational deployment of the system by means of feeding the operational system with a known dataset that may include new Spoofing techniques that were not conceived at the moment of the deployment of the system (as long as these new spoofing techniques imply a detectable signature in the Search Space). This also implies that, in order to not allow the Machine Learning techniques to fit into non-relevant features, the training dataset must be carefully designed and curated. If, for instance, the training dataset does not contain the sufficient

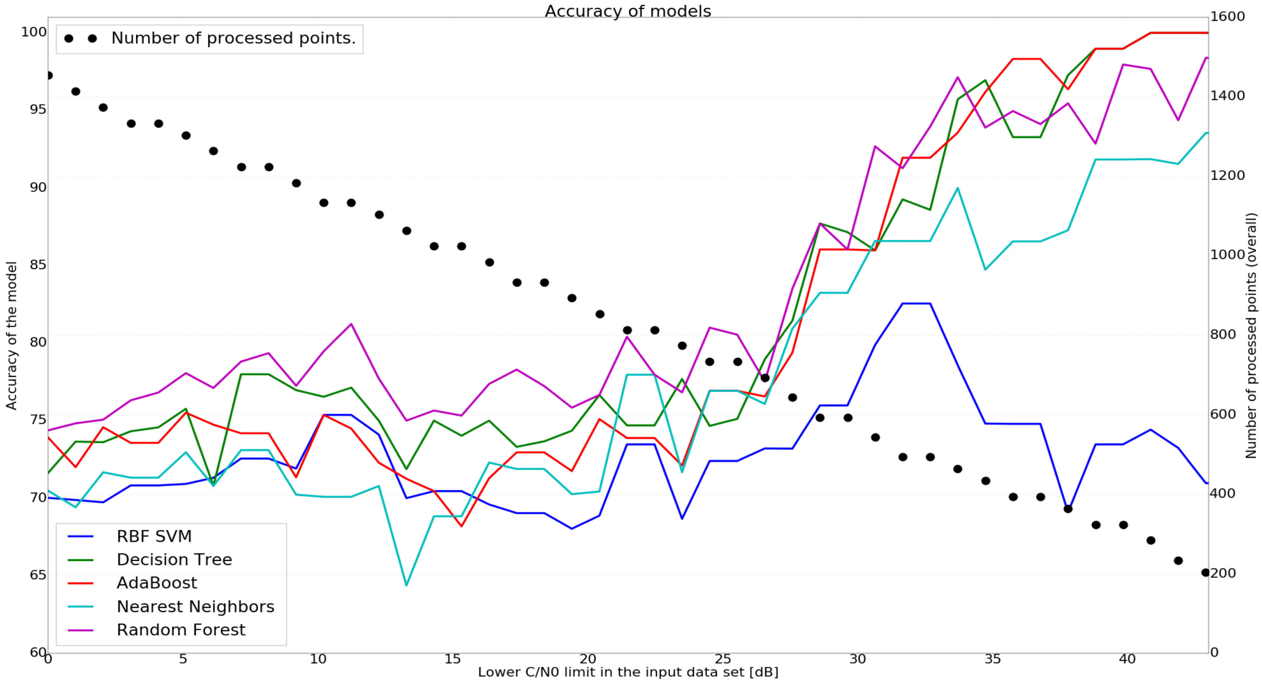


FIGURE 7. Accuracy (correct classifications) with E1C and 16ms of integration. No position fed into the algorithms.

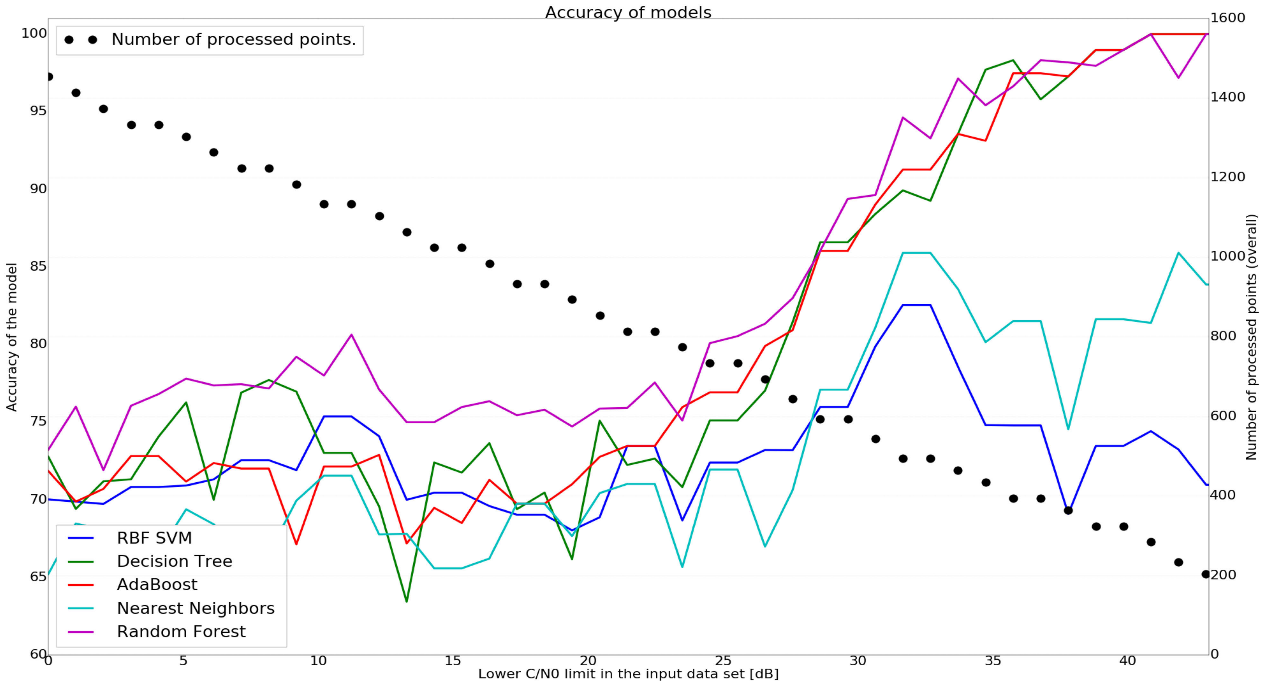


FIGURE 8. Accuracy (correct classifications) with E1C and 16ms of integration. Position fed into the algorithms.

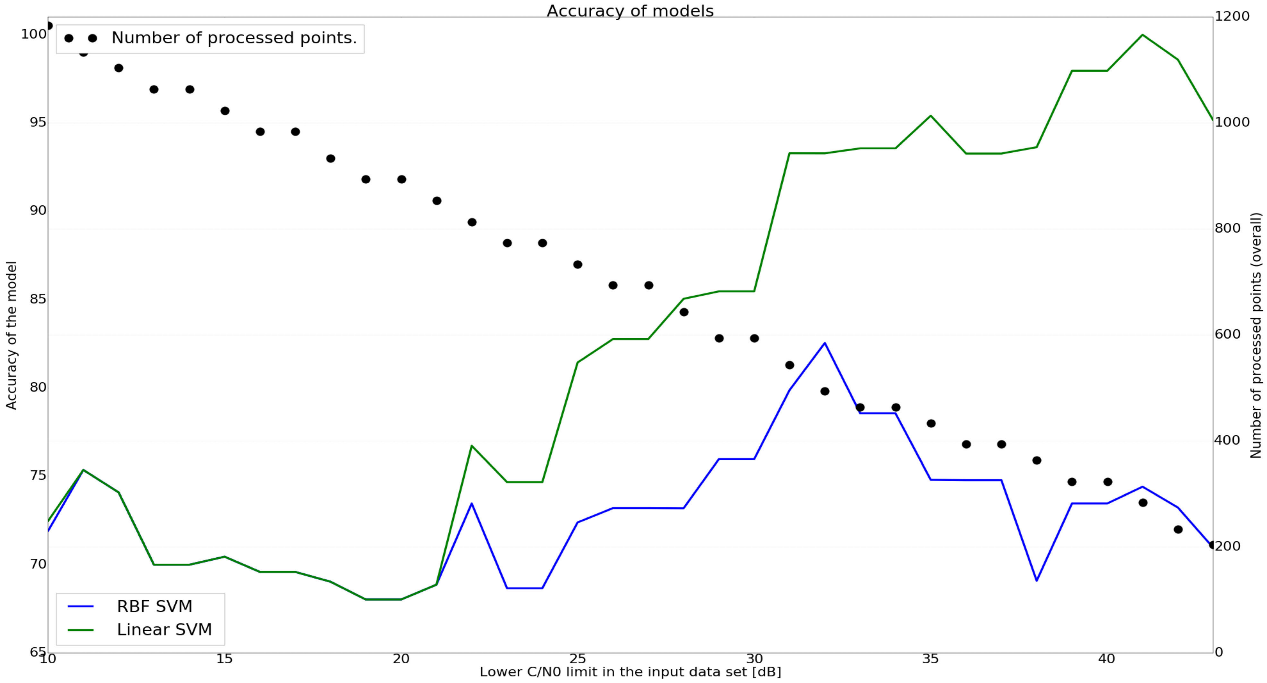


FIGURE 9. Accuracy (correct classifications) of SVM with RBF and Linear kernels, with E1C only 16 ms integration. No position fed.

distribution of spoofing signals over the Search Space, and positions are fed into the Machine Learning algorithms, then these algorithms may consider that the reduced number of positions in the Search Space that were fed into them are relevant for the detection, while it may not be the case. But, on the other hand, if the considerations in [20] regarding the relative position of the Spoofing signal with respect to the time delay are relevant, that can also be modeled into the system by accounting that situation into the training dataset. In other words: the detection capabilities of the deployed

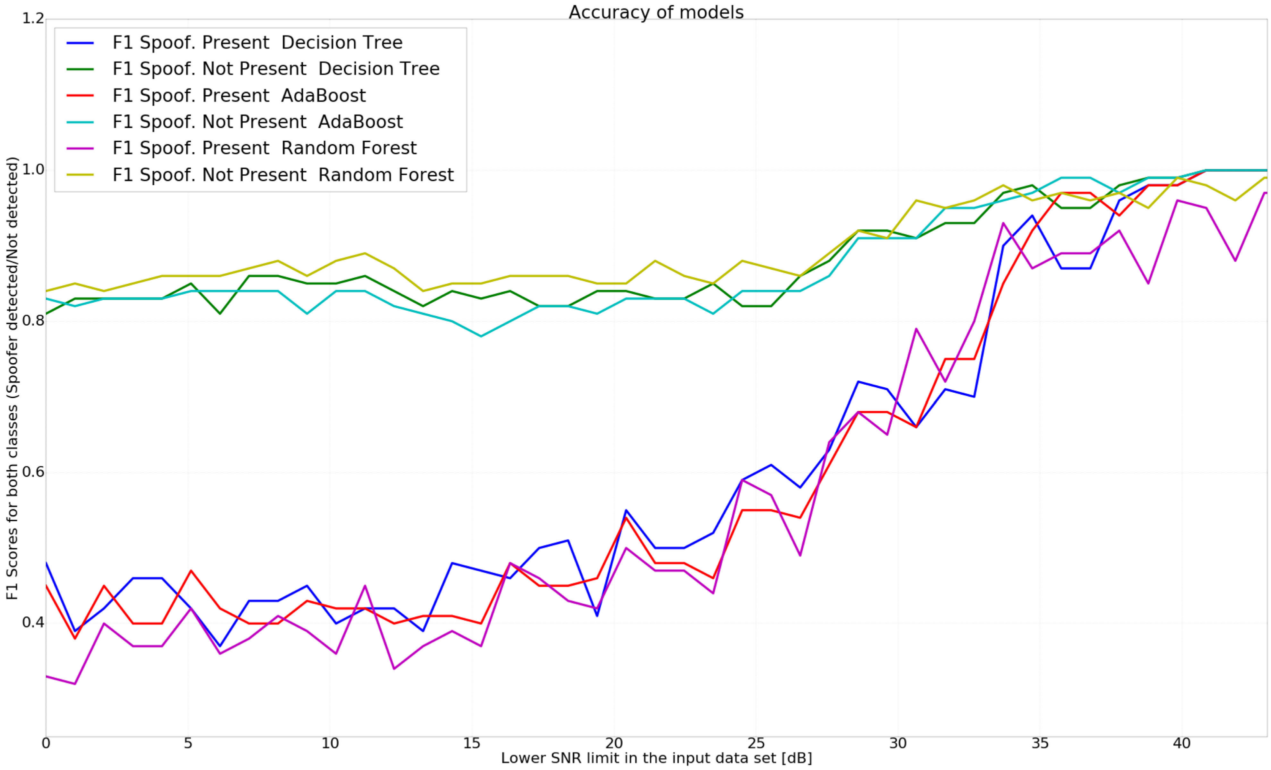


FIGURE 10. F1 Score with E1C and 16ms of integration. No position fed into the algorithms.

system can be further improved by means of a simple reconfiguration, without major modifications of the system. The relevancy of the training dataset is not only limited to this future evolution of the system and its detecting capabilities. As the PFA (Probability of False Alarm) of the system is also modeled by means of the over/under-representation of the spoofing cases in the training dataset. The same applies to the PMD (Probability of Missed Detection).

The result that should be considered for analyzing the PFA (Probability of False Alarm) is not the accuracy, which considers both classes (Spoofer present and spoofer not present), but can be derived from the reported confusion matrices. The caption of such matrices show all the ROC (Receiver Operating Characteristics) values, namely: True Positives, False Negatives (PMD), True Negatives and False Positives (PFA). Please refer to tables: 1, 2, 3 and 4. On the other hand, one of the best scores that can be used to derive the quality of the models, per class, is the F1 score. This score is typically used for evaluating data mining algorithms. It has been reported for the classes Spoofer not present and Spoofer present. The F1 score is based on the following calculation, per class (''Spoofer present''/''Spoofer not present''):

1) Calculate the **Precision**, which is number of True Positives (i.e. For the class ''Spoofer present'', this means that the Spoofer was there and the model determined that the Spoofer was there) divided by the number of True Positives, plus the number of False Positives (i.e. For the class ''Spoofer present'', this means that the Spoofer was not there, but the system determined that the Spoofer was there). For the class ''Spoofer present'', we can consider this as the number of correct predictions of Spoofer present, divided by the total number of predictions of spoofer present. So it is a ratio, where 1 means that there is no False Alarm (for the class ''Spoofer present'', for the class ''spoofer not present'' it would imply no Missed Detection) at all.
2) Calculate the **Recall**, which is the number of True Positives (i.e. For the class ''Spoofer present'', this means that the Spoofer was there and the model determined that the Spoofer was there) divided by the number of True Positives and the number of False Negatives (i.e. For the class ''Spoofer present'' the Spoofer was there and the model determined it was not there). For the class ''Spoofer Present'', we can consider this as the number of times the model detected the Spoofer when it was there, divided by the number of times the Spoofer was there (regardless whether the model detected it or not). It is a ratio, where 1 means that there is no Missed Detection (for the class ''Spoofer Present'', for the class ''Spoofer not present'' it would imply no False Alarm (False Alarm)).
3) Then, compute **F1**, which is the harmonic average of the Precision and Recall, and it is defined as per (58).

$$F_1 = 2\frac{PrecisionRecall}{Precision + Recall} \tag{58}$$

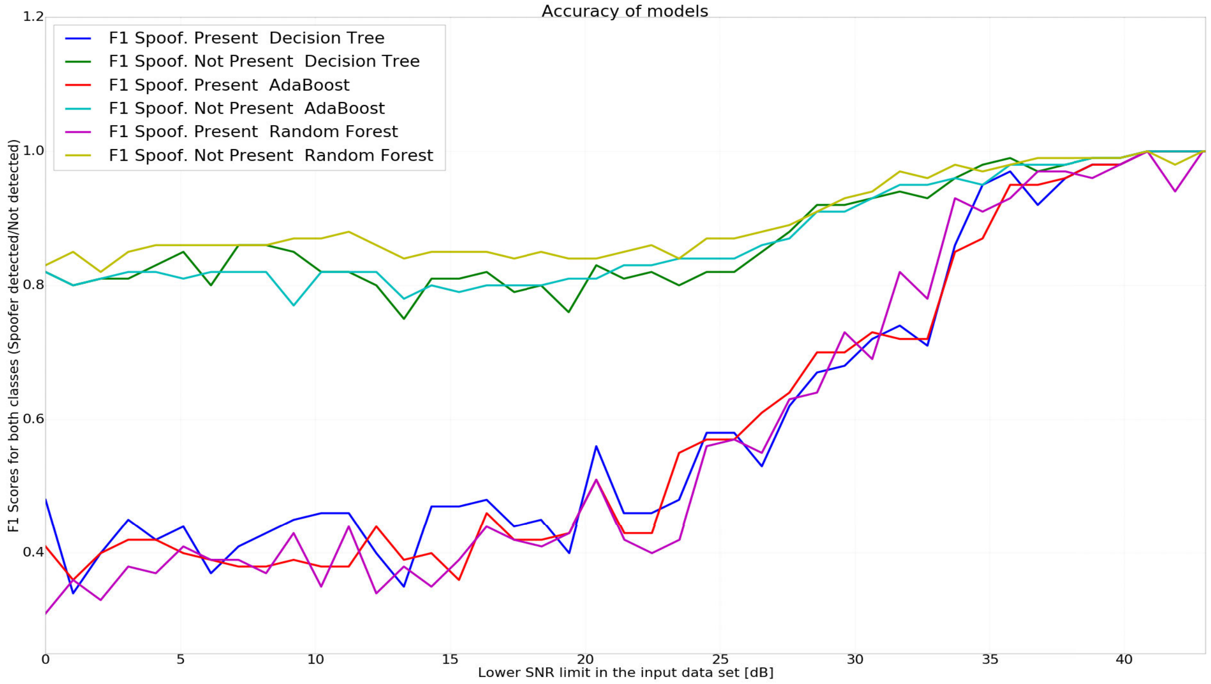


FIGURE 11. **F1 Score with E1C and 16ms of integration. Position fed into the algorithms.**

Both Recall and Precision give a rate of the number of correctly predicted elements of a class against the number of wrongly labeled elements (elements that, in reality, belong to a class and were predicted to the other class (Precision), or elements that don't belong to the class and were predicted as part of the class (Recall)). This means that F1 will be a value between precision and recall, per class, providing with a score which will be one for a perfect case and 0 for a model performing terribly. The F1 scores can be found in Fig. 10 and Fig. 11.

Note that modifying the number of spoofing signal cases in the training dataset will modify the values for F1 for both types of classes, too.

Other authors, as in [21] suggest the application of Neural Networks for the analysis of features extracted from the output of early and late correlators tracking the signal. In the present work, we are analyzing the entire search space for the generation of the features. This is an important detail, as otherwise if the victim tracks the false signal and that signal is far from the real peak in the Search Space, the attack may go undetected. In [22], the use of Support Vector Machines (SVM) on sensor fusion data is suggested for UAVs. Such approach makes sense in a dynamic environment like the one of a moving vehicle but, as claimed by the authors, if the Spoofer has absolute knowledge of the victim's trajectory, the attack will go undetected if this protection approach is followed. Note that this implies that critical infrastructure standoff victims (e.g. standoff timing users) are particularly in risk because of that, so for such receivers, the analysis of the Search Space should be advised.

As it can be seen in Figs. 8 and 7, the results were generated with two different setups in the victim receiver:

1) No Gaussian positions in the search space were fed into the algorithms, only using a local copy of E1C PRN and with a coherent integration time of 16ms.
2) Gaussian positions in the search space were fed into the algorithms, only using a local copy of E1C PRN and with a coherent integration time of 16ms.

Multipath was not simulated in the dataset, hence clear sky conditions are assumed at the victim's receiver antenna.

The calculated Search Space has a resolution of 392.16Hz in the Doppler axis and 20 ns in the delay axis, which was sufficient for the cases under analysis. Parallel Acquisition in Time Domain was used in the victim's receiver in order to compute the search space.

The workbench was configured to work with comma floating numbers. This allows disregarding effects related to fixed-point precision.

The results can be found in the Figs. 8 and 7.

The algorithm configurations were as follows:

1) Nearest Neighbors: groups (K) = 5, uniform weights, ball tree algorithm, leaf size = 20.
2) RBF SVM: Radial Kernel, Regularization parameter (C) = 0.025.
3) Decision Tree: Max depth = 7, Minimum number of samples per split = 3.
4) Random Forest: Max depth = 7, Minimum number of samples per split = 3, Number of trees in the random forest = 10.
5) AdaBoost: Number of estimators = 50.

The accuracy results were obtained using K-folds technique (K = 5). From the overall amount of samples, 30% of them were used for validation, deriving the accuracy results reported in this paper. The other 70% were used to train the models, using K-folds technique, dividing the dataset into five groups (K = 5). See Fig. 6 for details on the performed cross-validation.

The C/N_o lower limit for the used datasets can be seen in the horizontal axis of Figs. 7 and 8. For instance, a point in 20 dBHz means that all input data samples used for the training and validation are extracted from signal records with C/N_0 of 20dBHz or higher. The used data samples reduction can be seen on the right vertical axis in all these Figs. and the black points in the figures.

The best results were obtained with algorithms based on Decision Trees, namely: Decision Trees, ADA Boost and Random Forest. These algorithms perform in a remarkable manner, when signal C/N_0 is above 30dBHz. As it can be seen in Fig. 7, RBF SVM algorithm does not achieve such great performance in that C/N_0 range, providing low (compared to the results provided by the Decision Trees based algorithms) accuracy, around 75%. It is expected that the multipath will affect the accuracy results in a negative manner, although the solution could already be applicable to critical applications where a standoff receiver is in full open sky conditions and with no multipath. Further work will evaluate the multipath impact in the accuracy results. As per the impact of the algorithms not knowing the location of the peaks in the search space, a small accuracy reduction can be seen for the best performers: Decision Trees, Ada Boost and Random Forest. The reduction is very small, but it is still present. In the RBF SVM case, the algorithm seems to be very stable, with respect to the inclusion of the position of the peaks, as no difference is seen when allowing the algorithm to know the location of the peaks. As per the Nearest Neighbors case, the accuracy is reduced when the location of the peaks is introduced, particularly for high C/N_0. This can be explained due to the fact that, for the Nearest Neighbors case, an intense optimization was performed with the dataset without including the location of the peaks. As it can be seen in the Confusion Matrices (tables: 1, 2, 3 and 4), the impact is lower, but it is still present. This seems to imply that the results of these algorithms are more reliable and less prone to over-fitting, considering the proposed features and the simulated dataset.

The RBF SVM results do not improve as C/N_0 increases. It provides poor results, with respect to other algorithms. These results are not improved by the inclusion of the positions of the detected Gaussians. The reason for that is that the Radial Base Function kernel is not able to properly separate the data in the proposed features space. Indeed, results with other kernels for SVM, e.g. Linear kernels, are better and improve with C/N_0, confirming the fact that RBF is not properly separating the data. The results of the Linear SVM are not reported in order to not clutter the results graphs.

TABLE 1. Random Forest with E1C, 16 ms of integration time and no Gaussian positions in the search space fed to the algorithm. Lower limit: C/N_0>=31dBHz. True Positives: 94.62%. False Negatives: 5.38%. True Negatives: 96.15%. False Positives: 3.85%.

	True class: Spoofer present	True class: Spoofer not present
Classified as Spoofer present	88	1
Classified as Spoofer not present	5	25

TABLE 2. Random Forest with E1C, 16 ms of integration time and no Gaussian positions in the search space fed to the algorithm. Lower limit: C/N_0>=39dBHz. True Positives: 97.30%. False Negatives: 2.70%. True Negatives: 100%. False Positives: 0%.

	True class: Spoofer present	True class: Spoofer not present
Classified as Spoofer present	72	0
Classified as Spoofer not present	2	24

TABLE 3. Random Forest with E1C only, 16ms of integration time and Gaussian positions in the search space fed to the algorithm. Lower limit: C/N_0>=35dBHz. True Positives: 96.43%. False Negatives: 3.57%. True Negatives: 96.43%. False Positives: 3.57%.

	True class: Spoofer present	True class: Spoofer not present
Classified as Spoofer present	81	1
Classified as Spoofer not present	3	27

TABLE 4. Random Forest with E1C only, 16ms of integration time and Gaussian positions in the search space fed to the algorithm. Lower limit: C/N_0>=39dBHz. True Positives: 98.63%. False Negatives: 1.37%. True Negatives: 100%. False Positives: 0%.

	True class: Spoofer present	True class: Spoofer not present
Classified as Spoofer present	72	0
Classified as Spoofer not present	1	25

Just as a reference, in Fig. 9, the SVM with linear kernel and RBF kernel are compared. In the linear case, as the C/N_0 improves the results also improve, eventually providing similar values to Decision Trees, when C/N_0 is greater than 30 dBHz. This demonstrates that while the RBF kernel is not able to properly separate the provided dataset, the Linear kernel is. The results in Fig. 9 were generated with E1C only 16 ms of integration and without feeding the position into the algorithm. In order to get more relevant accuracy figures, the best algorithms in terms of performance will be evaluated showing the F_1 score against the lower C/N_0 limit for both classes (Spoofer present/Spoofer not present). Confusion Matrices are shown (tables 1, 2, 3 and 4) for the best performer (as it can be derived from Figs. 10 and 11), the Random Forest algorithm. In general, for Random Forest, as reported in the Confusion Matrices in table 2 and table 4, for $C/N_0 >= 39$ dBHz some classification errors are reduced to zero. This cannot be understood as a perfect result but as a very low error rate that, due to the size of the dataset, is not shown. The proposed technique should be only used to detect attacks (mainly for stand-off critical application receivers). Once the attack is detected, other auxiliary navigation systems shall be used (e.g. Inertial navigation systems, alternative stable clocks, etc.)

VI. CONCLUSION

Spoofing attacks represent a very serious threat for GNSS systems. To fight against that risk, some authentication techniques, like NMA in Galileo, will be available. However, this may not be enough to counteract the spoofing attacks based on SCER. For this particular case, a complementary approach based on the application of machine learning methods to the receiver search space has been introduced in this paper.

Two main aspects have been studied: the Intra-satellite PRN non-orthogonality distortion term due to Galileo modulation and the use of machine learning techniques for end-user protection.

Such Intra-satellite PRN non-orthogonality distortion term was found and a quality curve that allows direct comparison between GPS and Galileo for SCER Spoofer symbol estimation, was provided (refer to Section II). A deep examination of the operations performed by a spoofer to produce a SCER attack was provided, too.

Simulations in Section III confirmed the presence of the Intra-satellite PRN non-orthogonality distortion term for Galileo attacks on NMA, with low integration times and high C/N_0. Appreciable impact was only found for integration times of 1 μs and C/N_0 greater than 67 dBHz. The impact was around 10%, or more, in terms of Spoofer P_d. This effect may be of particular relevance for attacks on systems using SCA while using high-gain antennas. The Galileo quality curves proposed in Section II were also confirmed by the simulations.

Theoretical calculations suggest that a different distortion term, as seen in (38), appears in the SCER attack on Galileo with SCA, hindering the estimation of the unpredictable chip. Moreover, the Galileo theoretical quality curve, calculated in Section II, is very challenging for the attack on SCA as the integration time will need to be below 1 μs. In the second part of the paper, a new Machine Learning technique for SCER spoofer detection was proposed. Based on the simulation results, with classifiers based on Decision Trees and the proposed features extraction method, the models obtained performance ratios (correct classifications) greater than 98.48%, for C/N_0 between 40 dBHz and 50 dBHz. The False alarm rates get a significant improvement for decision tree-based algorithms for $C/N_0 >= 39$ dBHz, as seen in tables 2 and 4. Therefore, this seems to be a promising complementary solution for detecting spoofing attacks which, otherwise, may not be detected (assuming NMA is used, as the Spoofer may modify the navigation message without being detected by the proposed method if NMA is not used). Note that it was assumed that the attacker was not able to null the original satellite signal. Hence, the proposed Machine Learning technique could be applied for critical applications standoff receivers in two cases: when using GNSS signals that do not support NMA or when using NMA signals to protect against SCER, and always after a successful check of the Navigation Message authenticity.

Further work will evaluate other signal features, trying to reduce the computational load of the extraction step. Multipath simulation will be considered and more sophisticated RFI detection methods will be evaluated, too. The workbench will be updated to simulate SCA, and steps will be performed to start implementing a demonstrator on FPGAs.

ACKNOWLEDGMENT

The authors would like to thank Dr. Ignacio Fernandez for his support. They would also like to thank Prof. Christian Arbinger (DLR GfR mbH) for his support for the development of this article.

REFERENCES

[1] *Commission Implementing Decision*, document (EU) 2017/224, European Commission, 2017.

[2] I. Fernández-Hernández, V. Rijmen, G. Seco-Granados, J. Simon, I. Rodríguez, and J. D. Calle, "A navigation message authentication proposal for the galileo open service," *Navigation*, vol. 63, no. 1, pp. 85–102, Mar. 2016.

[3] T. E. Humphreys, "Assessing the spoofing threat: Development of a portable GPS civilian spoofer," in *Proc. 21th Int. Tech. Meeting Satell. Division Inst. Navigat. (ION)*, Savannah, GA, USA, 2008.

[4] G. Caparra, "Feasibility and limitations of self-spoofing attacks on GNSS signal with message authentication," in *Proc. 30th Int. Tech. Meeting The Satell. Division Inst. Navigat. (ION)*, Portland, OR, USA, 2017, pp. 3968–3984.

[5] T. E. Humphreys, "Detection strategy for cryptographic GNSS antispoofing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 2, pp. 1073–1090, Apr. 2013.

[6] *Interface Specification IS-GPS-800 Navstar GPS Space Segment/User Segment L1C Interface. Global Positioning Systems Directorate Systems Engineering and Integration*, Directorate Syst. Eng. Integr., El Segundo, CA, USA, 2013.

[7] P. L. Vahid, J. Nielsen, and G. Lachapelle, "GNSS spoofing detection based on receiver C/N0 estimates," in *Proc. ION GNSS12 Conf.*, Nashville, TN, USA, 2012, pp. 2878–2884.

[8] I. Fernandez-Hernandez and G. Seco-Granados, "Galileo NMA signal unpredictability and anti-replay protection," in *Proc. Int. Conf. Localization GNSS (ICL-GNSS)*, Barcelona, Spain, Jun. 2016, pp. 1–5.

[9] *European GNSS (Galileo) Open Service Signal In Space Interface Control Document*, document, European Commission, 2015.

[10] E. S. Lohan, "Limited bandwidths and correlation ambiguities: Do they co-exist in galileo receivers," *Positioning*, vol. 2, no. 1, pp. 14–21, 2011.

[11] J. M. Anderson, K. L. Carroll, N. P. DeVilbiss, J. T. Gillis, J. C. Hinks, B. W. O'Hanlon, J. J. Rushanan, L. Scott, and R. A. Yazdi, "Chips-message robust authentication (Chimera) for GPS civilian signals," in *Proc. 30th Int. Tech. Meeting Satell. Division Inst. Navigat. (ION GNSS)*, Nov. 2017, pp. 2388–2416.

[12] G. Caparra, "On the achievable equivalent security of GNSS ranging code encryption," in *Proc. IEEE/ION Position Location Navigat. Symp. (PLANS)*, Monterey, CA, USA, 2018, pp. 956–966.

[13] L. Scott, "The role of civil signal authentication in trustable systems," in *Proc. Presentation PNT Advisory Board*, Alexandria, VA, USA, 2019.

[14] *RF Characterization Data UBX USRP Daughterboard Rev 001*. Ettus Res., 2015.

[15] D. Borio, "A statistical theory for GNSS signal acquisition," Ph.D. dissertation, Politecnico Di Tori, Turin, Italy, 2008.

[16] F. P. Fontan, M. Vazquez-Castro, C. E. Cabado, J. P. Garcia, and E. Kubista, "Statistical modeling of the LMS channel," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1549–1567, Nov. 2001.

[17] K. Borre, *A Software-Defined GPS Galileo receiver. A Single-Frequency Apprach*. Boston, MA, USA: Birkhäuser, 2007.

[18] F. Pedregosa, "Scikit-learn: Machine learning in python," in *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[19] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[20] C. Hegarty, B. O'Hanlon, A. Odeh, K. Shallberg, and J. Flake, ''Spoofing detection in GNSS receivers through cross-ambiguity function monitoring,'' in *Proc. 32nd Int. Tech. Meeting Satell. Division Inst. Navigat. (ION GNSS)*, Oct. 2019, pp. 920–942.

[21] E. Shafiee, M. R. Mosavi, and M. Moazedi, ''Detection of spoofing attack using machine learning based on multi-layer neural network in single-frequency GPS receivers,'' *J. Navigat.*, vol. 71, no. 1, pp. 169–188, Jan. 2018.

[22] G. Panice, S. Luongo, G. Gigante, D. Pascarella, C. Di Benedetto, A. Vozella, and A. Pescape, ''A SVM-based detection approach for GPS spoofing attacks to UAV,'' in *Proc. 23rd Int. Conf. Autom. Comput. (ICAC)*, Huddersfield, U.K., Sep. 2017, pp. 1–11.

[23] Scikit-Learn Developers. *Cross-Validation*. Accessed: Dec. 1, 2019. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html

FRANCISCO GALLARDO received the B.Eng. degree in radio communications from the Technical University of Madrid (UPM), Spain, in 2012, the M.Eng. degree in astronomy and astrophysics from Valencian International University (VIU), in 2015, and the Ph.D. degree from UPM. He is currently a GNSS Engineer at DLR GfR mbH, Technical University of Madrid. He joined the NASA Madrid Deep Space Communication Complex as a Systems Engineer, in 2012. Since 2017, he has been working as a GNSS Engineer with DLR GfR mbH in the Galileo Programme. He received the ROHDE and SCHWARZ Price for the best final degree projects in radio frequency.

ANTONIO PÉREZ YUSTE (Senior Member, IEEE) received the B.Eng. degree in radio communications, the M.Eng. degree in telecommunications, and the Ph.D. degree *(cum laude)* in electrical engineering from the Universidad Politécnica de Madrid (UPM), Spain, in 1991, 1996, and 2004, respectively. He was the Vice Director of the School of Telecommunications Engineering, Technical University of Madrid, from 1997 to 2001, where he was the Director, from 2001 to 2004. He was the Chief of the Technical University of Madrid President's Cabinet, from 2004 to 2012, and the UPM Sino-Spanish Campus Director, Shanghai, China, from September 2012 to February 2014. He was responsible for the UPM Sino-Spanish Cooperation Office, Madrid, from March 2014 to July 2014. In 2014, he was appointed under the Chinese Government's prestigious National High-End Foreign Expert Program to a three-year Guest Professorship at Tongji University, Shanghai. He is currently a Professor of telecommunications engineering with the Technical University of Madrid (Spanish: UPM). He is currently a Guest Professor with the Tongji College of Electronics and Information Engineering, Shanghai. His research interests are related to wireless propagation and channel modeling. In addition, he is also a Specialist in the History of Electrical and Electronic Engineering, with special emphasis in the History of Telecommunications and in the development of the Information Society. Since 2012, he has been a member of the IEEE Spain Section Executive Committee. Since 2015, he has been a member and a Current Correspondent Member of the IEEE History Committee.