Spoofing Detection Performance of Snapshot OSNMA Under Time and Symbol Errors

Husnain Shahid*, Luca Canzian§, Carlo Sarto§, Oscar Pozzobon§,
Joaquín Reyes-González‡, Gonzalo Seco-Granados*, José A. López-Salcedo*
*Universitat Autònoma de Barcelona (UAB), IEEC-CERES, Barcelona, Spain
§Qascom SrL, Bassano del Grappa, Italy
‡EUSPA, Prague, Czech Republic

***Abstract*—In the recent years there have been intense research efforts to protect users of Global Navigation Satellite Systems (GNSS) from spoofing attacks that aim to mislead the user's navigation solution by means of counterfeit signals. Galileo, the European GNSS, has been at the forefront of such efforts and implemented already its so-called Open Service Navigation Message Authentication (OSNMA) protocol, intended to authenticate the navigation message and thus provide a protection layer not existing before in any other GNSS open signal. Rather than message authentication, this paper focuses on a novel technique for detecting the presence of potential spoofers by taking advantage of the unpredictability of some of the OSNMA data. Such opportunistic method is based on processing short sequences of received unpredictable symbols, sending them to a remote trusted server where access to the authentic unpredictable symbols is available, and then comparing both the received and the expected sequences for detecting potential mismatches. This approach is well-suited for receivers operating in snapshot mode, which can only gather and process a short piece of received signal due to their limited wireless connectivity, as in many Internet-of-Things (IoT) devices with low-power constraints. Interestingly, the problem addressed herein has some similarities with frame synchronization in digital networks. At the same time, though, it poses new challenges such as the presence of propagation errors, time uncertainty and different propagation times among different satellites, thus deserving a dedicated study.**

***Index Terms*—Satellite navigation systems, counterfeiting, sequences, detection**

I. Introduction

The widespread adoption of GNSS across many market segments such as agriculture, consumer solutions, finance, logistics or transportation, has enormously improved their operations and users' experience thanks to the provision of accurate position and time, anytime, anywhere on Earth. In recent years, though, the interest on the adoption and deployment of GNSS has been paralleled with efforts to interfere its signals and even to spoof them, as part of fraudulent or criminal activities. The latter has received an increasing attention motivated by the advent of low-cost software-defined radio (SDR) transceivers and open-source software to generate and experiment with GNSS signals. Nowadays, a user with adequate knowledge can put all these ingredients together and be able to implement a GNSS spoofer [1], [2].

This work was supported in part by the OSNMAplus project funded by the European Union Agency for the Space Programme (EUSPA) under contract GSA/GRANT/03/2019/02, and in part by the Spanish Agency of Research (AEI) under the Research and Development projects PID2020-118984GB-I00/ and PDC2021-121362-I00/AEI/10.13039/501100011033.

It is true that GNSS community is aware of it and countermeasures have been proposed from the user's terminal to the system level. At system level, Galileo has been the first GNSS to implement a protection mechanism to its open-service signal, the so-called Open Service Navigation Message Authentication (OSNMA). This mechanism relies on the use of cryptographic data to verify the authenticity of the navigation message (I/NAV) broadcast on the Galileo E1-B signal component. OSNMA is based on the Time Efficient Stream Loss Tolerant (TESLA) protocol, where the authentication data is conveyed in a set of predictable and, most importantly, unpredictable symbols. The key observation is that such unpredictable symbols, can effectively help in identifying a counterfeit signal, as already proposed in [3] and [4].

Existing contributions on the opportunistic use of OSNMA unpredictable symbols, such as the aforementioned ones, rely on the assumption that spoofing detection has access to the correlation between the received signal and the local replica implemented at the receiver. In this way the technique can monitor whether the received symbols, at sample level, experience any abnormal change during the symbol period. Such change would be an indication that a Security Code and Estimation Reply (SCER) attack is taking place. That is, a situation where the attacker is tracking the signals from the authentic satellites, performing an online estimation of the transmitted unpredictable symbols, and then broadcasting the reconstructed signal back to the victim receiver [5].

An alternative to the use of sample-level information is using the estimated symbols already provided by the receiver. This has the advantage that it is transparent to the receiver implementation, but the disadvantage that the information content is much limited and circumscribed to the estimated symbols. Nevertheless, such limited information can still be used, under some assumptions, for spoofing detection. As discussed in [6], a possible approach would be to send the sequence of symbols estimated by the user to a remote server, where access to the authentic unpredictable symbols is available. The server would then compare the unpredictable symbols estimated by the user with the authentic ones, and declare whether potential errors might be due to spoofing.

The aforementioned technique is referred to as snapshot OSNMA because it works with a snapshot (i.e. a short sequence) of OSNMA symbols. It therefore fits well into the paradigm of snapshot GNSS receivers widely adopted in Internet of

Things (IoT) and low-power applications, for instance, the integrated positioning and communication systems where the low power wide area networks are emerging for industrial verticals and also Integrated Sensing and Communication (ISAC) for snapshot-based vehicle localization and the communication simultaneously [7]–[9]. The symbols being sought correspond to the OSNMA unpredictable symbols, which are transmitted within the 40 bits "Reserved" field in the odd pages of the Galileo I/NAV message [10]. These 40 bits are convolutionally encoded at the transmitter at a $1/2$ rate and the resulting 80 symbols, after interleaving, are received by the user. Not all symbols in the I/NAV message are unpredictable, as discussed in [3]. Most of them are predictable and thus carry no information from a spoofing detection point of view. The interest here is on unpredictable symbols, since they are the ones that potential spoofers will struggle to determine their value, and thus, will likely incur in a higher probability of error than in the rest of symbols.

The contribution of the present paper extends the results in [6] by considering the time uncertainty present on the estimated unpredictable symbols, and by analyzing the impact of symbol errors that naturally may arise at the receiver end due to adverse propagation effects. As for the time uncertainty, it is due to several factors. The first one is the unequal propagation time of the signals being received by the user from different satellites, which introduces a misalignment on the sequences of unpredictable symbols that are actually received, as observed in Fig. 1. The second problem is that the user receiver is often not perfectly synchronized with the GNSS time, and this introduces some uncertainty on the exact time at which the received signal is actually gathered by the receiver. On the one hand, this is due to the fact that clocks with modest performance (e.g. TCXO) are typically used in small devices implementing snapshot GNSS positioning. On the other hand, even though the clock offset could be estimated when solving the user's position, the use of a short snapshot of signal and thus coarse-timing navigation algorithms [11], results in time estimation errors on a few tens of ms [12]. Such errors are comparable to those incurred by network synchronization protocols, such as the Network Time Protocol (NTP), which is used by many IoT devices to update their system time.

Therefore, the user receiver must gather a snapshot larger than the one strictly needed to accommodate for such time uncertainty and to make sure that unpredictable symbols are always contained within the gathered snapshot. The problem to be solved at the server side is to decide whether the expected sequence of unpredictable symbols is contained in the sequence of symbols received by the user. Such decision is based on analyzing: i) the number of symbol errors that are incurred when comparing both sequences, and ii) the time offset at which the actual unpredictable symbols are received with respect to the expected GNSS time. Both checks will determine whether the received signal is spoofed or not. The first problem shares some similarities with that of unique word detection or frame synchronization in digital communications. However, while many contributions on this topic do exist, such as [13], [14] or [15], they are mostly based on the additive white Gaussian noise channel. The problem here is a different

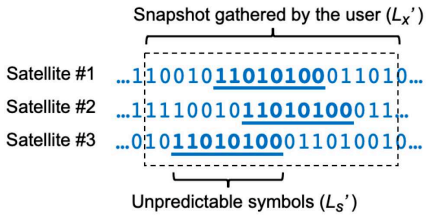

Fig. 1: Illustration of the symbols received by the user due to the time uncertainty caused by the clock offset and the satellites different propagation times.

one because it is based instead on the estimated symbols, that is, the hard-decisions provided by the user receiver. It is for this reason that the problem needs to be formulated instead as a Binary Symmetric Channel (BSC), for which the work in [16] provides some initial guidelines.

In this context, the present paper is structured as follows. First of all, Section II presents the BSC signal model for the sequences of unpredictable symbols at the user's and server sides. Then Section III introduces the proposed detector and characterizes its detection performance. Section IV provides simulation results to assess the goodness of the proposed detector and to solve the open questions on the impact of the sequence length, time uncertainty and probability of error of the spoofer. Finally, conclusions are drawn in Section V.

II. Signal Model

A. Unpredictable symbols within the Galileo I/NAV message

As previously introduced, unpredictable bits are transmitted by the Galileo satellites in the 40-bits "Reserved" field contained in the odd pages of the I/NAV message. After encoding and interleaving, the unpredictable symbols that are actually received by the user turn out to be spread within a short time frame of 380 ms within the 2 seconds duration of an odd page. Unpredictable symbols are assumed to appear in the form of four batches of consecutive symbols, as discussed in [3], [17]. The first three batches contain 9 unpredictable symbols each, and are separated apart by 21 predictable symbols, thus leading to 30 symbols per batch. Since the symbol period in Galileo E1-B is 4 ms, each of these batches has a duration of 120 ms, and the three of them amount to a total duration of 360 ms. After these three batches, a fourth and last batch of 5 unpredictable symbols is received, thus completing the total time period of 380 ms, containing a total of 32 unpredictable symbols per odd-page of the I/NAV message.

A snapshot receiver willing to gather these 32 unpredictable symbols per odd-page would need to precisely know the exact time at which the first unpredictable symbol is to be received. Unfortunately, this is difficult to be known in practice due to the presence of several sources of time uncertainty, such as: i) the satellite clock offset, which is typically smaller than 10 ms; ii) the user receiver clock offset, assumed here to be on the order of 50 ms[1]; and iii) the different propagation times experienced by signals received from different satellites, as

[1]For a short-term uncalibrated clock, even though less than 10 ms have been reported for small devices connected to current 4G/LTE networks [18].

observed in Fig. 1, which typically introduces an additional 20 ms time uncertainty.

All these time uncertainties amount to at least 80 ms (i.e. 20 symbols), which together with the 36 ms to gather the 9 unpredictable symbols per batch, leads to a snapshot length of roughly 120 ms, containing 30 symbols in total. This snapshot length is on the order of that typically used by small portable devices operating in snapshot mode and adopted in low-power IoT applications. While this snapshot length is enough for providing a position fix, it may not for detecting potential spoofers that incur in a low probability of error when trying to estimate the unpredictable symbols. For instance, for a spoofer incurring in a probability of symbol error of 0.01, at least 100 symbols would be needed to observe at least one of such errors. Thus, several snapshots may need to be gathered in order to collect a large enough number of unpredictable symbols to reliably detect the presence of spoofing.

B. Signal model for the sequence of received symbols

Let us stack the sequence of L'_x symbols contained in the n-th gathered snapshot into the $(1 \times L'_x)$ vector $\mathbf{x}_n$ as follows,

$$\mathbf{x}_n = \{[\hat{\mathbf{s}}_n \ \mathbf{d}_n]\}_m \tag{1}$$

where $\hat{\mathbf{s}}_n$ is the $(1 \times L'_s)$ sequence containing the unpredictable symbols estimated by the receiver,

$$\hat{\mathbf{s}}_n = [\hat{s}_{1,n}, \hat{s}_{2,n}, \ldots, \hat{s}_{L'_s,n}] \tag{2}$$

corresponding to the actual sequence of unpredictable symbols, $\mathbf{s}_n = [s_{1,n}, s_{2,n}, \ldots, s_{L'_s,n}]$, with $s_{i,n} = \{0, 1\}$, transmitted by the satellites. Since the latter are assumed to be all unpredictable, and thus no a-priori information on their value is available [17], it is reasonable to assume that they are all equiprobable and *iid*. In turn, $\mathbf{d}_n$ is the $(1 \times L_u)$ sequence containing the remaining $L_u = L'_x - L'_s$ predictable symbols appearing as a result of the aforementioned time uncertainty,

$$\mathbf{d}_n = [d_{1,n}, d_{2,n}, \ldots, d_{L_u,n}]. \tag{3}$$

The notation $\{\cdot\}_m$ in (1) denotes the circular rotation of a sequence of symbols by m positions. Thus, since the sequence of unpredictable symbols $\mathbf{s}_n$ can appear anywhere within the long sequence $\mathbf{x}_n$, either at the beginning, at the middle or at the end, so we have that $m = 0, 1, \ldots, L_u$.

The present work is focused on determining whether the true sequence $\mathbf{s}_n$ is contained within the long sequence $\mathbf{x}_n$. As a byproduct, this returns as well what the value for m is, which can be used later on to determine the reception time of unpredictable symbols. It is important to remark that since (2) is a sequence containing estimates (i.e. decisions) on the actual unpredictable symbols, each estimated symbol may incur in some probability of error or symbol error rate (SER),

$$\begin{aligned} P_e \ = \ & p(\hat{s}_{i,n} = 1 | s_{i,n} = 0) p(s_{i,n} = 0) \\ & + p(\hat{s}_{i,n} = 0 | s_{i,n} = 1) p(s_{i,n} = 1) \end{aligned} \tag{4}$$

for any i and n. Since the unpredictable symbols are assumed to be iid with equiprobable values, the SER just simplifies to $P_e = \text{prob}(\hat{s}_{i,n} \neq s_{i,n})$.

Finally, it worth mentioning that several snapshots may need to be gathered in order to collect a large enough number of unpredictable symbols to reliably detect a spoofer. Assuming that a total of N snapshots are gathered, they can be stacked into a $(N \times L'_x)$ matrix $\mathbf{X}$ as follows,

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \left\{ \begin{bmatrix} \hat{\mathbf{s}}_1 \ \mathbf{d}_1 \\ \hat{\mathbf{s}}_2 \ \mathbf{d}_2 \\ \vdots \\ \hat{\mathbf{s}}_N \ \mathbf{d}_N \end{bmatrix} \right\}_m \Rightarrow \quad \mathbf{X} = \left\{ \left[\hat{\mathbf{S}} \ \mathbf{D} \right] \right\}_m \tag{5}$$

where the circular rotation operator $\{\cdot\}_m$ is applied column-wise. The total number of unpredictable symbols is now $L_s = NL'_s$.

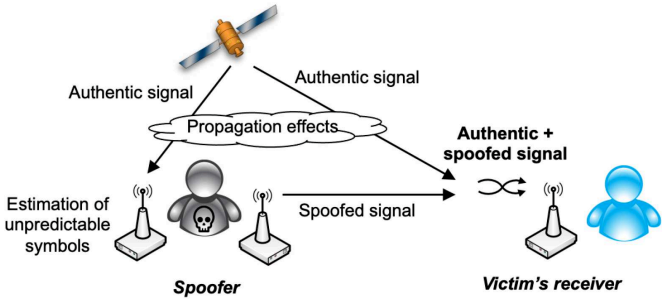

Fig. 2: Illustration of the spoofing scenario where the spoofer estimates the unpredictable symbols from the authentic signal and then transmits a spoofed signal to the victim's receiver.

C. Symbol error rate for typical working scenarios

In the absence of spoofer, $P_e = P_{e,0}$, and it is given by the nominal SER incurred by the user's receiver when estimating the symbols received from the authentic satellites. For perfect clear sky outdoor conditions, the channel can be modeled as an AWGN one with typical carrier-to-noise spectral density ratios, C/N_0, on the order of $40 - 45$ dBHz. In that case, the SER turns out to be negligible due to the relatively long symbol period, i.e. $T = 4$ ms, of Galileo E1-B. However, one may consider a pessimistic case with $C/N_0 \sim 30$ dBHz, which then results in $P_{e,0} = \frac{1}{2}\text{erfc}(T \cdot C/N_0) \sim 0.001$. In urban environments the situation significantly worsens due to the presence of multipath reflections. Assuming satellites at medium elevation angles such as 60°, $P_{e,0} \sim 0.01$ as shown in [6] using the land mobile satellite (LMS) channel. Assuming satellites at lower elevation angles such as 40°, the SER then further worsens up to $P_{e,0} \sim 0.1$ instead [6].

In the presence of a spoofed signal, $P_e = P_{e,1}$, and it becomes the result of various sources of error. On the one hand, there is the SER incurred by the spoofer when trying to estimate the unpredictable symbols. This probability depends on the working conditions and the limitations of the spoofer. On the other hand, there is the SER incurred by the user when trying to estimate the received spoofed symbols.

At this point it is important to note that the symbols received by the user are the aggregation of both the spoofed and the authentic ones. This is because both the spoofed and the authentic signals are received simultaneously at the user receiver, as illustrated in Fig. 2. Nevertheless, even if both signals were perfectly aligned in time and frequency, they could hardly ever be aligned in phase. This involves that

the phase misaligned superposition of both the spoofed and authentic signals likely results in a high SER, unless the power advantage of the spoofer is large enough to be the dominant signal and thus dominate the SER as well. For the sake of simplicity, and without loss of generality, three different cases have been considered in the present work for the SER in the presence of spoofing: i) a spoofer randomly guessing the unpredictable symbols and thus leading to $P_{e,1} = 0.5$; ii) a pessimistic case with $P_{e,1} = 0.01$, thus being difficult to detect; and iii) an optimistic one with $P_{e,1} = 0.1$, being easier to detect as far as the difference with respect to $P_{e,0}$ is large enough to make both hypotheses distinguishable.

III. Proposed snapshot spoofing detection

A. Detector Formulation

The proposed spoofing detector is inspired by the problem of detecting a short sequence of binary symbols within a longer one, a task frequently encountered in frame synchronization for digital communications. One of the most widely adopted detectors is the binary symmetric channel (BSC) frame synchronization algorithm in [16]. This algorithm is based on computing the Hamming distance between the short sequence of interest and portions of the same length from the time-shifted longer sequence. The pseudo-code implementation of such detector adapted to the problem at hand would be as follows,

BSC frame detection and synchronization

1: $\mu \leftarrow 0$
2: **if** $\sum_{n=1}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \mu) \leq \gamma$ **then**
3: $\quad$ $\mathbf{S}$ contained in the columns of $\mathbf{X}$,
4: $\quad$ $\hat{m} \leftarrow \mu$.
5: **else**
6: $\quad$ $\mu \leftarrow \mu + 1$
7: $\quad$ **if** $\mu < L_u$ **then**
8: $\quad\quad$ Jump to line 2.
9: $\quad$ **end if**
10: **end if**

where γ is a predefined detection threshold and $H(\mathbf{s}_n, \mathbf{x}_n; \mu)$ stands for the Hamming distance between $\mathbf{s}_n$ and $\mathbf{x}_n$ when the latter is time-shifted by μ positions. That is,

$$H(\mathbf{s}_n, \mathbf{x}_n; \mu) = \sum_{i=1}^{L_s'} s_{i,n} \oplus x_{i+\mu,n} \tag{6}$$

with $\oplus$ the XOR operator. Since the latter provides an output equal to 1 when the elements being compared are different, the result in (6) is actually providing the number of errors between both sequences, in line with the concept of Hamming distance. The threshold γ is therefore the maximum number of errors that are allowed to occur when comparing $\mathbf{s}_n$ and $\mathbf{x}_n$ for $n = 1, \ldots, N$.

Note that when the received sequence is affected by binary errors occurring with probability P_e, the average number of errors asymptotically tends to NP_e in virtue of the law of large numbers. This value can then be used as a reference for detecting the presence of spoofers by comparing the actual number of errors with the expected one. In other words, by comparing the measured P_e with the expected one.

Based on these observations, and for the problem at hand, let us formulate the two hypotheses under analysis as follows,

$$\begin{array}{lcll} \mathcal{H}_0 & : & P_e = P_{e,0} & \Rightarrow \ \mathbf{S} \subset \mathbf{X}, \\ \mathcal{H}_1 & : & P_e = P_{e,1} \neq P_{e,0} & \Rightarrow \ \mathbf{S} \not\subset \mathbf{X}. \end{array} \tag{7}$$

Under $\mathcal{H}_0$ we would allow both sequences to have up to $NP_{e,0}$ different symbols to account for errors normally occurring in the symbols reception at the user's terminal (e.g. thermal noise, shadowing, etc.). If more errors than expected are incurred, then $\mathbf{S} \not\subset \mathbf{X}$ would be declared and this would be an indication that a spoofing attack is potentially taking place.

The detection problem under analysis is therefore formulated using the following detector,

$$T(\mathbf{S}, \mathbf{X}) = \min_{0 \leq \mu \leq L_u} \sum_{n=0}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \mu) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma. \tag{8}$$

The starting position of $\mathbf{s}_n$ within $\mathbf{x}_n$, which is assumed constant across n, can be used to determine the reception time of the unpredictable symbols and thus, to detect replay attacks introducing a time offset of the spoofed signal. It is obtained as a byproduct of the detector as follows,

$$\hat{m} = \arg \min_{0 \leq \mu \leq L_u} \sum_{n=1}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \mu) \leq \gamma. \tag{9}$$

At this point it is important to emphasize that the threshold γ plays an important role in (8) for accommodating potential errors that could be incurred at the user side. For a negligible probability of error at the user receiver, one could safely set $\gamma = 0$ because no errors are expected to be incurred by the user. Any arising error should then be due to the spoofer, and thus be detected. In contrast, if some errors were expected to be incurred by the user due to shadowing or fading effects, one could set for instance, $\gamma = 1$, to allow one error to be incurred by the received symbols. For a total of $L_x = 50$ received symbols, for instance, this would mean allowing a SER of $1/50 = 0.02$, which may suffice to accommodate expected errors with $P_{e,0} < 0.02$ and thus allow detecting spoofers with $P_{e,1} > 0.02$. As can be seen, there is a tradeoff in setting γ because the larger it is, the more inherent errors at the user side can be accommodated (i.e. due to fading or shadowing), thus reducing the probability of false alarm, but at the same time, a larger $P_{e,1}$ is needed by the spoofer for being detected.

B. Probability of Detection

According to (8), the spoofer detection is successful when the spoofer is present and $T(\mathbf{S}; \mathbf{X}) > \gamma$. The probability of detection can thus be mathematically formulated as follows,

$$P_{\mathrm{d}} = p\left(T(\mathbf{S}, \mathbf{X}) > \gamma | \mathcal{H}_1\right) \tag{10}$$

$$= \prod_{0 \leq \mu \leq L_u} p\left(\sum_{n=1}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \mu) > \gamma | \mathcal{H}_1\right) \tag{11}$$

$$= \prod_{0 \leq \mu \leq L_u} \left(1 - p\left(\sum_{n=1}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \mu) \leq \gamma | \mathcal{H}_1\right)\right) \tag{12}$$

Note that the step from (10) to (11) is done in virtue of the fact that $p(\min_\alpha U(\alpha) \leq \gamma)$ for some positive semi-definite function $U(\cdot)$ and $\alpha = \{0, 1, \ldots, M-1\}$, can be expressed as the joint probability $p(U(0) \leq \gamma, U(1) \leq \gamma, \ldots, U(M-1) \leq \gamma)$. Furthermore, the set of values $U(\alpha) \doteq \sum_{n=1}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \alpha)$ for $\alpha = \{0, 1, \ldots, M-1\}$ can be assumed to be independent one from each other due to the unpredictability (i.e. uncorrelatedness) of the underlying symbols in $\mathbf{s}_n$ and $\mathbf{x}_n$.

At this point, two different regions can be considered for $p\left(\sum_{n=1}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \mu) \leq \gamma | \mathcal{H}_1\right)$ in (12) as a function of μ:

1) The first region corresponds to $\mu = m$, when the correct time-shift is applied in (6) so that the expected unpredictable symbols are aligned with those actually received by the user. The probability we are looking for is then given by,

$$\begin{aligned} P_1(L_s, \gamma, P_{e,1}) &\doteq p\left(\sum_{n=1}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \mu = m) \leq \gamma | \mathcal{H}_1\right) \\ &= \sum_{k=0}^{\gamma} \binom{L_s}{k} (1 - P_{e,1})^{L_s - k} P_{e,1}^{k} \end{aligned} \quad (13)$$

2) The second region comprises the values $\mu \neq m$, when the expected and received unpredictable symbols are misaligned. Since we are dealing with sequences of unpredictable symbols, it is reasonable to assume that a sequence and its time-shifted version will be uncorrelated. As such, the errors γ that are sought in (12) are just the errors arising when comparing two independent sequences of random symbols. As such, we can define

$$\begin{aligned} P_2(L_s, \gamma) &\doteq p\left(\sum_{n=1}^{N} H(\mathbf{s}_n, \mathbf{x}_n; \mu \neq m) \leq \gamma | \mathcal{H}_1\right) \\ &= \frac{1}{2^{L_s}} \sum_{k=0}^{\gamma} \binom{L_s}{k} \end{aligned} \quad (14)$$

Merging both results together, the probability of success in detecting the sequence of unpredictable symbols becomes,

$$P_{\mathrm{d}}(L_s, L_u, \gamma, P_{e,1}) = \big(1 - P_1(L_s, \gamma, P_{e,1})\big)\big(1 - P_2(L_s, \gamma)\big)^{L_u} \quad (15)$$

C. Probability of False Alarm

Apart from the probability of success, it is of interest to determine the probability of false alarm. Such probability is defined as that of incurring more than γ errors when comparing both the expected and received unpredictable symbols, when actually no spoofer is present. That is

$$P_{\mathrm{fa}} = p\left(T(\mathbf{S}, \mathbf{X}) > \gamma | \mathcal{H}_0\right) \quad (16)$$

which is computed in the same manner as the probability of detection except for the fact that $P_{e,1}$ is now replaced by $P_{e,0}$. The result is given by

$$P_{\mathrm{fa}}(L_s, L_u, \gamma, P_{e,0}) = \big(1 - P_1(L_s, \gamma, P_{e,0})\big)\big(1 - P_2(L_s, \gamma)\big)^{L_u}. \quad (17)$$

IV. Simulation Results

The goal of this section is two-fold. On the one hand, to assess the goodness of the theoretical expressions derived in Section III-B and III-C. On the other hand, once these expression are validated, to analyze the performance of the proposed detector in Section III-A.

A. Goodness of the Theoretical Expressions for P_{d} and P_{fa}

The first step in the simulation results is to confirm the validity of the theoretical expressions for P_{d} and P_{fa} presented in (15) and (17), respectively. This is important in order to ensure that the underlying assumptions and subsequent conclusions are consistent with the real behavior of the proposed technique. The way to proceed is by comparing the results provided by the theoretical expressions with those obtained empirically. The latter are obtained by generating random sequences of unpredictable symbols, introducing random errors on their reception, and adding random symbols before and after this sequence to emulate the effect of time uncertainty.

The results are depicted in Fig. 3 for P_{d} and P_{fa}. Without loss of generality, and example is shown for the case of $P_{e,1} = 0.1$ and $P_{e,0} = 0.001$, respectively. A time uncertainty of 20 symbols is considered as per Section II-A and a threshold $\gamma = \{0, 1\}$. The results show that both theoretical and empirical curves closely match one with each other, thus validating the goodness of theoretical expressions. As can be seen, P_{d} increases with the length of the sequence of unpredictable symbols. This is because for a fixed $P_{e,1}$, the longer such sequence, the more errors will be encountered, and the easier will be to detect the spoofer. Increasing the threshold γ provides more tolerance to errors and dramatically reduces P_{fa}, as observed in the right hand side plot of Fig. 3. This may be needed in cases when the user receiver is subject to shadowing or fading, and thus it is prone to non-spoofing errors that need to be accommodated in order not to be confused with spoofing ones. The price to be paid is a reduction as well on P_{d}. The tradeoff betwen P_{d} and P_{fa} will be assessed later on in Section IV-C.

Another aspect to be highlighted is the influence of the time uncertainty onto the detection performance. Interestingly, the impact is only noticeable when the number of unpredictable symbols, L_s, is comparable to the number of symbols introduced by the time uncertainty, L_u. The reason is that the detector in (8) relies on taking the minimum of the Hamming distances among all tentative time shifts μ. This will tend to identify the time shift at which the unpredictable symbols are located, $\mu = m$, because it is at this position where the least number of errors will likely be observed. Falling outside of this region, namely selecting $\mu \neq m$, and therefore being misled by the presence of time uncertainty, becomes more unlikely as L_s grows. This is because the larger L_s, the more difficult it is to find a sequence of random symbols that might incur in less errors than the received sequence of unpredictable symbols. Actually, it can be seen in (14) that for a fixed and finite threshold (e.g. $\gamma = \{0, 1\}$ as in Fig. 3), P_2 rapidly goes to 0 as L_s grows, and then P_{d} does not depend on L_u anymore. This effect can be observed in Fig. 3 for $L_s > 10$ with $\gamma = 0$

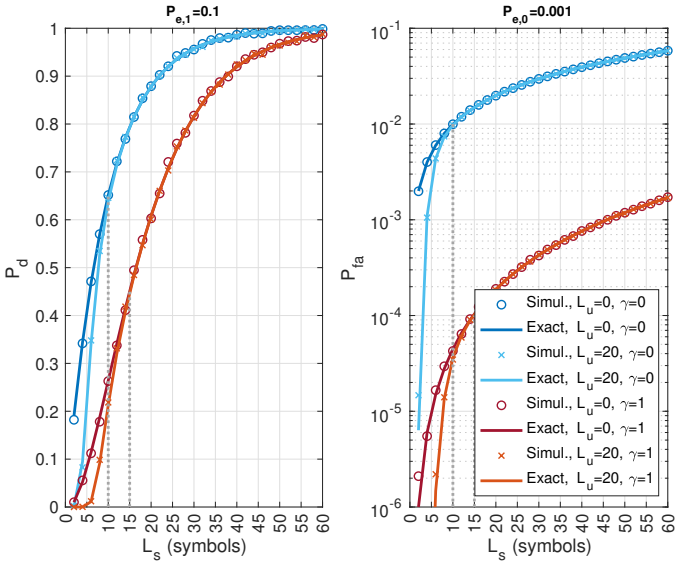

Fig. 3: Comparison between the empirical and theoretical $P_{\rm d}$ for $P_{e,1} = 0.1$ (left), as well as $P_{\rm fa}$ for $P_{e,0} = 0.001$ (right), as a function of the unpredictable symbols L_s, using $\gamma = \{0, 1\}$.

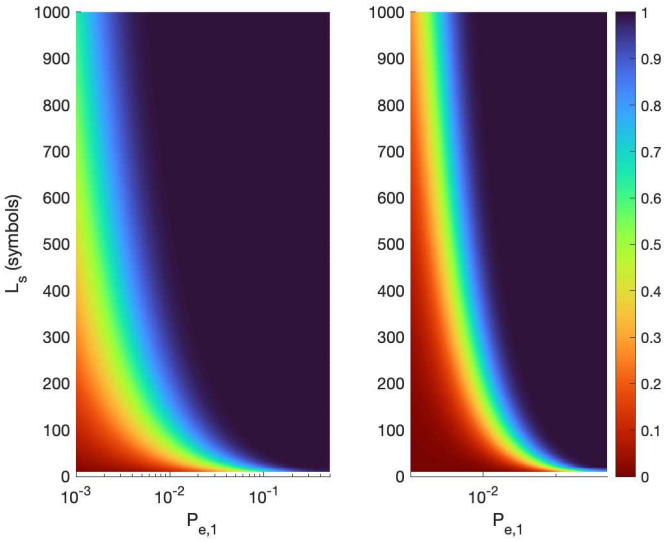

Fig. 4: Probability of detection as a function of the probability of symbol error, $P_{e,1}$, for $L_u = 20$ symbols with $\gamma = 0$ (left) and $\gamma = 1$ (right).

and $L_s > 15$ with $\gamma = 1$, thus suggesting that time uncertainty does note have a relevant impact provided that $L_s > L_u$.

B. Probability of Detection as a Function of P_e

Once the theoretical expressions have been validated, it is of interest to analyze the impact of the spoofer probability of error $P_{\rm e,1}$ onto the spoofer detection. This is shown in Fig. 4 for a wide range of lengths L_s of the unpredictable symbols sequence. For illustration purposes, the detection threshold is set to $\gamma = 0$ and $L_u = 20$ symbols introduced by the time uncertainty are considered.

As expected, $P_{\rm d}$ increases as $P_{e,1}$ does so, since the latter is ultimately unveiling the presence of the spoofer. Conversely, as $P_{e,1}$ decreases, it becomes more difficult to detect the spoofer and the only way to counteract this situation is by increasing the length of the unpredictable symbols sequence. Indeed, it can be observed that roughly $L_s \sim 2/P_{e,1}$ symbols are needed in order to ensure that the spoofer is detected with a high enough probability of detection. This makes challenging to detect spoofers with a low probability of error, since many unpredictable symbols will be needed, thus incurring in a high time to spoofing detection since only ~ 30 unpredictable symbols are available every 2 seconds of the I/NAV message.

C. Probability of Detection vs Probability of False Alarm

The results presented so far consider either the probability of detection or probability of false alarm. In practice, though, both must be assessed simultaneously in order to determine what the price is (i.e. false alarms) for achieving the largest amount of successful detections. To this end, the receiver operating curve (ROC) is a useful tool representing probability of detection versus probability of false alarm. Unfortunately, the proposed detector is strongly influenced by many parameters such as L_s, $P_{e,0}$ and $P_{e,1}$, thus making the ROC curve to be multi-dimensional, and thus difficult to be analyzed.

In these circumstances it is often preferred to summarize the information content of the ROC curve into a lower dimensional metric. One of such metrics is the area under the curve (AUC), which is a threshold-independent metric that summarizes the content of the overall ROC curve. Another metric is the Youden index [19], which measures the vertical distance from a point of the ROC curve down to the diagonal line corresponding to the ROC of a random detector. As such, the Youden index is a threshold-dependent metric and consequently, it is well-suited for analyzing the detection performance of the proposed detector, whose detection threshold γ is playing a central role.

The Youden index for the detector in (8) is shown in Fig. 5 for different pairs of $P_{e,1}$ and $P_{e,0}$, as well as for different detection thresholds γ. The Youden index ranges from 0, which means the worst possible performance (i.e. $P_{\rm d} = P_{\rm fa}$), up to 1, which means the best possible performance (i.e. $P_{\rm d} = 1$ and $P_{\rm fa} = 0$). It is interesting to observe that the more different $P_{e,1}$ and $P_{e,0}$ are, the easier it is to achieve the best detection performance. This can be seen in the upper left hand side subplot of Fig. 5 for $P_{e,1} = 0.5$ and $P_{e,0} = 0.001$. In contrast, the more similar $P_{e,1}$ and $P_{e,0}$ are, the more difficult it is to succeed in detecting the spoofer. This can be seen in the upper right hand side subplot of Fig. 5 for $P_{e,1} = 0.5$ and $P_{e,0} = 0.1$, where the Youden index struggles to reach its maximum value equal to one.

For the rest of intermediate combinations of $P_{e,1}$ and $P_{e,0}$, the detection threshold γ can be adjusted to provide the maximum Youden index. As a result, the optimal required number of unpredictable symbols can be obtained. For instance, in the bottom middle subplot of Fig. 5 for $P_{e,1} = 0.1$ and $P_{e,0} = 0.01$, the smallest threshold achieving the maximum Youden index is $\gamma = 5$, thus confirming that optimal detection is possible for this working scenario. To do so, the required number of unpredictable symbols amounts to 150, thus providing a valuable guideline to properly configure the detection

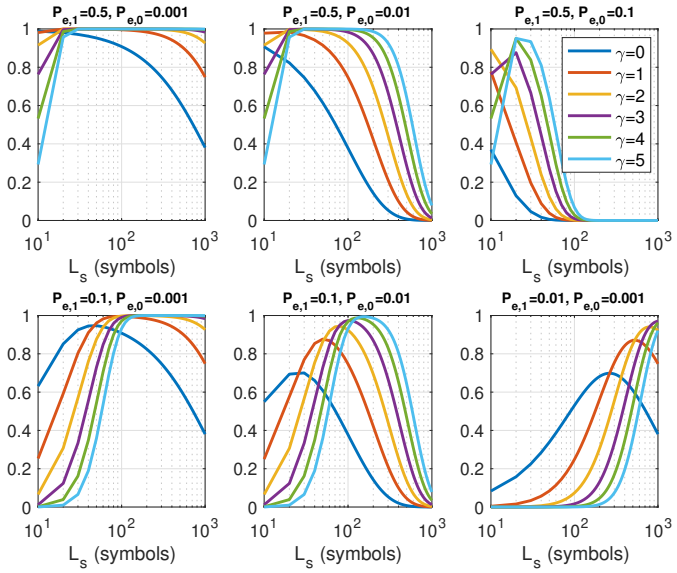

Fig. 5: Youden index for the proposed detector in various working conditions for $P_{e,0} = \{0.001, 0.01, 0.1\}$, $P_{e,1} = \{0.5, 0.1, 0.01\}$, and time uncertainty of $L_u = 20$ symbols.

technique for a given working scenario.

V. Conclusions

This paper has analyzed the spoofing detection performance of the so-called snapshot OSNMA, which relies on comparing the received sequence of unpredictable OSNMA symbols with the expected sequence available at a remote trusted facility. Since unpredictable symbols can hardly be guessed by a spoofer and thus, must be actively estimated instead, errors are likely to occur in such estimation process. This is due to the spoofer inner constraints and limitations, as well as its operating working conditions (e.g. potentially adverse propagation conditions). Thus, by comparing both sequences of unpredictable symbols, an abnormal number errors becomes may indicate the presence of a potential spoofer. This problem, though, is hindered by the misalignment between the time at which the user started to collect the received symbols, and the time at which unpredictable symbols actually arrive at the receiver after being transmitted by the Galileo satellites.

A detector for such problem has been proposed in this paper along with the mathematical characterization of its probability of detection and probability of false alarm. Special emphasis has been placed on characterizing the detection performance as a function of the unpredictable symbols length, L_s, time uncertainty L_u and probability of error in the presence, $P_{e,1}$, and in the absence of spoofer, $P_{e,0}$. The results herein help in understanding the advantages and limitations of OSNMA symbol-level spoofing detection. Additionally, they showcase the potential use of snapshot OSNMA in applications driven by an on-demand or snapshot-mode operation, such as in IoT positioning implemented by battery-operated devices.

Disclaimer

This activity is performed in the scope of the OSNMAplus project, funded by the European Union (EUSPA) under grant GSA/GRANT/03/2019 and coordinated by Qascom. The content of this publication reflects only the authors' view. EC/EUSPA is not responsible for any use that may be made of the information it contains. Snapshot OSNMA is the subject of the Italian patent with application number 102022000025806 filed on 16/12/2022.

Acknowledgment

The authors would like to thank Ignacio Fernandez-Hernandez, from DG DEFIS, European Commission, for his insightful comments and suggestions.

References

[1] T. E. Humphreys, B. M. Ledvina, M. L. Psiaki, B. W. O'Hanlon, P. M. Kintner *et al.*, "Assessing the spoofing threat: Development of a portable GPS civilian spoofer," in *Proc. ION GNSS*, 2008, pp. 2314–2325.

[2] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen, and S. Capkun, "On the requirements for successful GPS spoofing attacks," in *Proc. ACM CCS conference*, 2011, pp. 75–86.

[3] I. Fernández-Hernández and G. Seco-Granados, "Galileo NMA signal unpredictability and anti-replay protection," in *Proc. Intl. Conf. on Localization and GNSS (ICL-GNSS)*. IEEE, 2016, pp. 1–5.

[4] G. Seco-Granados, D. Gómez-Casco, J. A. López-Salcedo, and I. Fernandez-Hernandez, "Detection of replay attacks to GNSS based on partial correlations and authentication data unpredictability," *GPS Solutions*, vol. 25, no. 2, pp. 1–15, 2021.

[5] S. Cancela, J. Navarro, D. Calle, E. Göhler, A. Dalla Chiara, G. da Broi, I. Fernández-Hernández, G. Seco-Granados, and J. Simon, "Testing receiver resilience against signal replay attacks," in *Proc. Intl. Tech. Symp. on Navigation and Timing (ITSNT)*, 2018.

[6] H. Shahid, S. Locubiche, L. Canzian, C. Sarto, O. Pozzobon, I. Fernandez-Hernandez, J. Reyes-González, G. Seco-Granados, and J. A. López-Salcedo, "Feasibility of snapshot OSNMA for spoofing detection in urban scenarios," in *Proc. European Navigation Conference (ENC)*, 2023, pp. 1–9.

[7] K. Borre, I. Fernandez-Hernandez, J. A. López-Salcedo, and M. Z. H. Bhuiyan, *GNSS software receivers. Ch. 9: Snapshot Receivers*. Cambridge University Press, 2022, p. 210–244.

[8] S. Guo, B. Lu, M. Wen, S. Dang, and N. Saeed, "Customized 5g and beyond private networks with integrated urllc, embb, mmtc, and positioning for industrial verticals," *IEEE Communications Standards Magazine*, vol. 6, no. 1, pp. 52–57, 2022.

[9] D. Cong, S. Guo, H. Zhang, J. Ye, and M.-S. Alouini, "Beamforming design for integrated sensing and communication systems with finite alphabet input," *IEEE Wireless Communications Letters*, vol. 11, no. 10, pp. 2190–2194, 2022.

[10] European Union, "European GNSS (Galileo) Open Service, signal-in-space interface control document," January 2021.

[11] F. Van Diggelen, *GPS: Assisted GPS, GNSS, and SBAS*. Artech House, 2009.

[12] K. Muthuraman, J. Brown, and M. Chansarkar, "Coarse time navigation: Equivalence of algorithms and reliability of time estimates," in *Proc. ITM of the Institute of Navigation (ION GNSS)*, 2012, pp. 1115–1138.

[13] J. Massey, "Optimum frame synchronization," *IEEE Trans. on Commun.*, vol. 20, no. 2, pp. 115–119, 1972.

[14] R. Mehlan and H. Meyr, "Optimum frame synchronization for asynchronous packet transmission," in *Proc. IEEE Intl. Conf. on Commun. (ICC)*, vol. 2, 1993, pp. 826–830.

[15] M. Chiani and M. G. Martini, "Analysis of optimum frame synchronization based on periodically embedded sync words," *IEEE Trans. on Commun.*, vol. 55, no. 11, pp. 2056–2060, 2007.

[16] R. Scholtz, "Frame synchronization techniques," *IEEE Trans. on Commun.*, vol. 28, no. 8, pp. 1204–1213, 1980.

[17] C. O'Driscoll and I. Fernández-Hernández, "Mapping bit to symbol unpredictability with application to Galileo Open Service Navigation Message Authentication," *Navigation*, vol. 69, no. 2, pp. 1–19, 2022.

[18] R. Miškinis, D. Jokubauskis, D. Smirnov, E. Urba, B. Malyško, B. Dzindzelėta, and K. Svirskas, "Timing over a 4G mobile network," in *Proc. European Freq. and Time Forum (EFTF)*, 2014, pp. 491–493.

[19] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.