GPS Solut (2017) 21:577–589
DOI 10.1007/s10291-016-0548-7

ORIGINAL ARTICLE

Hypothesis testing methods to detect spoofing attacks: a test against the TEXBAT datasets

Micaela Troglia Gamba[2] · Minh Duc Truong[1] · Beatrice Motella[2] · Emanuela Falletti[2] · Tung Hai Ta[1]

Received: 8 April 2015 / Accepted: 25 May 2016 / Published online: 14 June 2016

Abstract The hazardous effects of spoofing attacks on the global navigation satellite system (GNSS) receiver are well known. Technologies and algorithms to increase the awareness of GNSS receivers against such attacks become more important and necessary. We present the validation of two statistical spoofing detection methods, namely the Chi-square goodness of fit (GoF) test and the Sign test applied to pairwise correlator differences, for each satellite tracked by the receiver. The test bench for the algorithms, both implemented in a software receiver, is the public database produced by the University of Texas at Austin, which reproduces various representative cases of spoofing attacks (the so-called TEXBAT). The algorithms show a very promising capability of detecting the attack, in particular when an aggregate decision is taken based on a joint detection upon all the tracked satellites. Furthermore, the GoF test appears also reliable in dynamic conditions and in case of a huge power advantage of the spoofing signal. The response of the receiver to the attacks confirms the spoofing signal represents an "extraneous agent" which, before taking control of the receiver, can be recognized by properly combined strategies of signal quality monitoring.

Keywords GNSS receivers · Spoofing · Goodness of fit · Sign test · Software defined radio · TEXBAT

Introduction

In the last decade, global navigation satellite system (GNSS) receivers were introduced in a large variety of applications. It is apparent that the mass market sector is the most active one, thanks to the wide spread use of smartphones and tablets with embedded GNSS chipsets which open the way for new position, navigation and timing (PNT) applications such as road tolling, pay as you drive, location based services, communication networks synchronization, financial transactions, transports and fleet management. GNSS plays a vital role also in other more established fields, like aeronautical and maritime, where satellite localization is often a matter of safety.

Unfortunately, the civil GNSS receivers are vulnerable to radio-frequency interference (RFI), either intentional or not. The open GNSS signal does not inherently carry any "built-in" anti-interference method; therefore, it is vulnerable to radio-frequency attacks that aim either at blinding the receiver or fooling it. The first condition is typically indicated as jamming, as in the case of Grant et al. (2009), while the second is known as spoofing (Heng et al. 2015).

Spoofing is a particularly malicious attack, whose effect can be devastating for the system containing the GNSS receiver, because the PNT information reported by a

✉ Emanuela Falletti
falletti@ismb.it

Micaela Troglia Gamba
trogliagamba@ismb.it

Minh Duc Truong
duc.truongminh@hust.edu.vn

Beatrice Motella
Motella@ismb.it

Tung Hai Ta
tung.tahai@hust.edu.vn

[1] NAVIS Centre, Hanoi University of Science and Technology, No. 1 Dai Co Viet, Hanoi 10000, Vietnam

[2] Research Area on Navigation Technologies, Istituto Superiore Mario Boella, Via P. C. Boggio 61, 10138 Turin, Italy

receiver under attack can be highly misleading. The major hazard in this situation is that the receiver is typically not aware of being fooled; therefore, it does not raise any alarm to the hosting system, which is induced to make wrong and possibly hazardous decisions based on spoofed PNT information. The recent literature has highlighted the risk of intentional attackers willing to interfere with the correct GNSS receiver functions (Heng et al. 2015; Akos 2012).

In this context, it is clear that anti-spoofing capabilities, or at least spoofing detection capabilities, become a desirable feature of commercial-grade GNSS receivers. The effectiveness of spoofing attacks against civil GNSS receivers, proved in Psiaki et al. (2013) and Humphreys et al. (2008), led to the development of several techniques for spoofing detection over the last decade. A detailed survey of the most promising ones can be found in Jafarnia-Jahromi et al. (2012a), where several methods are described and compared in terms of complexity and effectiveness. Comprehensive references are also Wesson (2013, 2014—Chapter 4). The most robust technique is represented by the use of antenna arrays (Magiera and Katulski 2013; Montgomery et al. 2009; Konovaltsev et al. 2013), which allow for the detection of the angle of arrival (AOA) or the signal phase difference, but their use in mass market applications is still difficult due to the additional hardware and cost. Techniques based on power measurements are more affordable than those using the automatic gain control (AGC) to indicate an increase in the incoming signal power (Akos 2012; Jafarnia-Jahromi et al. 2012b). Nonetheless, such techniques might not be effective when the power of the spoofing signal and the true signal match. Moreover, they require front ends providing the AGC gain information. Methods based on multiple tracking loops (Moon et al. 2013) or vector tracking loops are also under investigation (Jafarnia-Jahromi et al. 2012c), but they are still too complex to be implemented. Among the proposed approaches there are countermeasures that belong to the class of signal quality monitoring (SQM) techniques and work at the correlation level by monitoring the quality of the GNSS correlation peak (Wesson et al. 2011; Troglia Gamba et al. 2013; Jin et al. 2013; Manfredini et al. 2014). A well-known method, belonging to this class, is the ratio test (Ledvina et al. 2010). The GoF test and the Sign test, proposed here, belong to the same family of statistical methods of the ratio test (Ali et al. 2014): They all work at the correlator output and monitor the correlation function shape. The low complexity, in terms of computational burden, is one of the main features, which make such tests attractive to be implemented in a software receiver (Pini et al. 2013).

Since spoofing detection is almost always a matter of accurate signal processing, receivers, in which early signal processing stages are accessible and raw data are observable, are the best candidates to host and test spoofing detection approaches. Receiver architectures featuring such characteristics are by nature the software receivers, in which signal processing is delegated to programmable devices. Software receivers offer the necessary flexibility and accessibility to support the inclusion of algorithms and functions, without the need of a redesign of the processing architecture (Lo Presti et al. 2014).

We present the results of a test suite developed on one of our software receivers, in which we implemented a promising interference detection strategy based on the observation of the distortion of the code cross-correlation function during the signal tracking stage. The strategy relies on the construction of a test statistic from a set of correlators, which enables an event of detection formulated in terms of binary hypothesis testing: interference absent/interference present. In Troglia Gamba et al. (2013) and Pini et al. (2013), two different tests are presented, used to discriminate whether the correlation function is distorted or not: the Chi-square goodness of fit (GoF) test and the Sign test.

Troglia Gamba et al. (2013) and Pini et al. (2013) applied hypothesis testing for interference detection to unintentional interference and simulated spoofing attack, showing encouraging results. However, the method was never tried thus far on real data. We show the results of the two statistical methods implemented in a complete software receiver architecture and applied to a database which is a de facto standard for testing anti-spoofing barriers at the signal processing level. The datasets are parts of the "Texas Spoofing test Battery" (TEXBAT), a database of recorded GPS signal samples available on the University of Texas at Austin Web site at the time of writing. The database currently covers six different realistic reference spoofing attacks, against which anti-spoofing enabled receivers can be tested (Humphreys et al. 2012). Some preliminary results of this experiment have been presented by the authors in Truong et al. (2014).

The next section recalls the major features of the TEXBAT database. We then summarize the main theoretical points of the interference detection algorithms based on the two test statistics, followed by a detailed analysis of the processing output obtained by our software receiver enabled by the spoofing detection functionality, and a discussion on the detection capability of the two methods applied to the reference datasets. Finally, we provide conclusions and options for future research.

The TEXBAT datasets

The Radionavigation Laboratory at the University of Texas at Austin produced in 2012 the first public database of signals affected by several types of spoofing attacks concerning GPS satellites (Humphreys et al. 2012). They view their database, referred to as TEXBAT, as “the data component of an evolving standard meant to define the notion of spoof resistance for commercial GPS receivers.” The database includes six GPS data collections representing six different spoofing attacks, constructed in the laboratory so as to be as much representative as possible of the currently known attack techniques. They are based on two “clean datasets,” recorded by a wideband, 25 Msps sampling rate data grabber attached in one case to a static antenna and in the other case to an antenna mounted on a car, which travelled across the city.

Each clean dataset was then played back in a laboratory setup and mixed with a counterfeit signal generated with appropriate characteristics; the mix at the radio-frequency level of the true and counterfeit signals was recorded by the same data grabber to produce the six datasets made public for research use. The first four datasets, Cases 1–4, are generated from the static data collection, while the fifth and sixth ones refer to the dynamic collection. As the authors clearly point out, detecting a spoofing menace is far tougher for a receiver in motion than for a static one, since the “deviations from a nominal condition” observed in a moving receiver can be often ascribed to natural environmental effects such as multipath and temporary signal blockage, which unfortunately tend to mask the effect of a spoofing attempt.

The addressed statistical methods aim at detecting distortions of the code cross-correlation function due to an anomalous signal mixed to the signal generated by the satellite. Consequently, the “Static Switch” attack contained in the dataset 1 (Case 1) is not likely detectable by such techniques, because it does not generate any significantly distorted waveform in the receiver. On the other hand, cases in which both authentic and counterfeit signals are simultaneously present are more likely to generate detectable distortions in certain metrics observed in the receiver tracking loops. For these reasons, our analysis concentrates on Cases 2–6.

In Case 2, an “overpowered” attack is conducted, in which the counterfeit signal has a 10 dB power advantage on the true one. In this condition, an AGC at the early stages of a receiver could react to the appearance of such a strong signal with a proportional gain reduction, which in practice would induce the true signal to be definitively buried below the noise floor. This could limit the interaction between the authentic and counterfeit signals in the receiver and consequently the detection capability of the anti-spoofing techniques. In fact, as the spoofing-to-authentic power ratio increases, the detection power of the method declines. For this reason, it is essential to adopt complementary countermeasures, able to simultaneously monitor the correlation function and the received signal power (Wesson 2014).

In Case 3, after about 100 s of clean reception, a counterfeit signal is superimposed to the authentic one, with a power advantage of 1.3 dB. The spoofer works in “frequency lock” mode, meaning that it strives to keep a nearly constant phase offset between the authentic and counterfeit signal carrier; in this way, the receiver cannot detect any suspicious carrier phase ramp during carrier phase tracking. In such a case, the rates of change of the code and carrier phases do not maintain a constant proportionality, but the relative code phase offset between counterfeit and authentic signal shifts with respect to the constant carrier phase offset, inducing a timing deviation in the receiver.

Case 4 is similar to 3, apart from the fact that the spoofer’s power advantage is further reduced to 0.4 dB and the attack is developed on the receiver’s position instead of receiver’s time. The response of the statistical detection methods to these two cases is expected to be similar and effective in both cases.

Cases 5 and 6 are similar, respectively, to 2 and 4, except that the receiver platform is dynamic rather than static. In fact, an overpowered time push attack is performed in Case 5, while a matched-power position push attack is performed in 6. In these cases, the spoofing detection capability is exposed to the additional challenge of discriminating between the spoofing effects and natural fading and multipath, experienced in a dynamic environment.

Spoofing detection techniques based on statistical analysis

Hypothesis testing or significance testing is defined in Albright (1977) as a family of methods for testing a hypothesis, the null hypothesis H_0, about a parameter in a population, using observed data. The data sample is examined, through the computation of a test statistic, in order to see whether a parameter is consistent with the hypothesized one or not. The decision on H_0 is taken comparing the test statistic t, extracted from the observed data, against a critical value t_α, representing the threshold for which the observed data confirm or do not confirm the null hypothesis H_0, with a required level of significance. This is equivalent to compare the statistical metric called p-value, against the significance level of the test. The p-value is defined in Moore and McCabe (1993) as “the probability, computed assuming H_0 is true, that the test statistic would take a value as extreme or more extreme than that actually observed.” The smaller the p-value, the stronger the evidence against H_0 provided by the data. The significance level α represents the probability of rejecting a true null hypothesis (Cohen 1988).

Troglia Gamba et al. (2013) and Pini et al. (2013) describe and compare two different tests applied at the output of the correlators. The goodness of fit test, belonging to the family of Chi-square tests, makes a statement concerning the nature of the distribution for the whole population, while the Sign test focuses on a single parameter of the distribution. In the problem at hand, i.e., the detection of a counterfeit signal, the data sample is represented by consecutive early and late correlations. With such outputs, the test statistic can be built and used to discriminate whether the code correlation function is distorted or not by the presence of a tracked signal replica. Both tests are based on the fact that, if there are no disturbances, the code correlation in a locked loop is an even function (null hypothesis) and the output of two correlators (early, E, late, L) equally spaced from the "prompt" one can be modeled as two normally distributed random variables with the same mean $\mu_\mathrm{E} = \mu_\mathrm{L} = \mu$ and variance, which depends on the noise floor N_0. Assigning two specific early–late correlators to build the test statistic, with an early–late spacing $d_\mathrm{EL} > 1$ chip, it is possible to show that E and L are statistically independent and $D = E - L$ results to be a normally distributed random variable with zero mean $\mu_D = 0$. An observation of M pairs of correlators outputs is taken as data sample, i.e.,

$$\mathcal{D} = \{D_1, D_2, \cdots\ D_M\} \tag{1}$$

where the subscript represents a temporal index counting the consecutive correlators outputs. Both tests are applied on the set of differences $\mathcal{D}$ and take a decision on the distortion of the correlation function. The Sign test verifies whether the mean of the population, represented by the observed data $\mathcal{D}$, is zero, i.e., $\mu_D = 0$, while the Chi-square GoF test also checks whether D is normal distributed, or, more in general terms, whether it follows a given distribution F. For the two tests, the null hypothesis H_0 can be formulated as

$$H_0 : \begin{cases} \mu_D = 0 & \text{Sign Test} \\ D \sim F & \chi^2 \text{ GoF Test} \end{cases} \tag{2}$$

where F is a given distribution function. The procedure for the tests implementation is described in Albright (1977) and recalled in Pini et al. (2013). In our case, both the Chi-square GoF and Sign tests are applied to a pair of in-phase extra correlators dedicated to the detection technique and not used in the tracking loops. Such a choice allows for a proper setting of the correlators spacing for the detection test, without impacting on the tracking performance. Another option would be using also the quadrature component, adding a further pair of correlators and looking for example at the magnitude of the complex difference. In this work, we opted for a solution that kept the complexity limited, by adding a single correlator pair on the in-phase output.

The tests are performed in two phases: The former, generally referred to as calibration phase, has the scope of tuning the algorithm parameter, while the latter is the actual application of the test. In Pini et al. (2013), the Chi-square GoF and the Sign test are also compared in terms of performance and computational burden: Though the former test obtains better performance in specific interference and spoofing situations, the latter requires lower number of operations.

The performance of these tests depends on a few parameters that are worth to be briefly discussed here.

- The significance level α is the probability of rejecting a true null hypothesis and is typically a design parameter, often put equal to 0.05.
- The assignment of α and the knowledge of the theoretical distribution of the test statistic under H_0 fix the critical value t_α and allow the computation of the p-value associated with the test statistic. For example, the test statistic for GoF is known to be χ^2 distributed with d_f degrees of freedom, with d_f equal to the number of non-empty categories k of the test minus 1. In some cases, the null hypothesis involves fitting a model with parameters estimated from the data sample. By estimating a parameter, a degree of freedom in the Chi-square test statistic is lost. In general, the degrees of freedom has to be adjusted with the number of estimated parameters d_p. Thus, $d_f = k - 1 - d_p$. On the other hand, for the Sign test, the distribution of the test statistic can be well approximated to a standard normal distribution, if sample size is moderately large, i.e., $M > 30$.
- The probability of accepting the null hypothesis, even though it is false (miss detection), is typically indicated with β. It is related to the so-called power of the test, which is the probability of rejecting the null hypothesis when it is false, defined as power $= 1 - \beta$. The power is clearly a function of the distribution of the test statistic t under the alternative hypothesis H_1 and increases with the separation of the distributions of t under either H_0 or H_1. In turn, such a separation is measured in terms of minimum deviation from the null hypothesis that the test hopes to detect, namely the "effect size" w. In other words, w is the difference between the null hypothesis and the alternative one that the test aims at discriminating. This means that, in order to evaluate the power of the test, an assumption on the effect size targeted by the test must be done, in addition to α and the test statistic distribution. An example of power analysis for the Chi-square GoF test is shown hereafter.

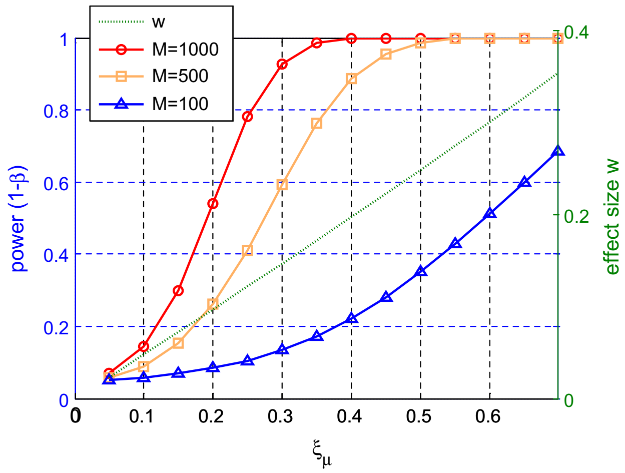


Fig. 1 Power (*left*) and effect size (*right*) of the GoF χ^2 test as a function of the normalized value of μ_D, under the alternative hypothesis. Curves are parameterized on the sample size M. Here, we set $\alpha = 0.05$ and $d_f = 10$

- Also the size of the observed data M plays a fundamental role in determining the power of the test. The ability to reject the null hypothesis, when it is false, is strongly dependent on the sample size: The smaller the size of the observed data, the weaker the test is.

Assume that the significance level for the Chi-square GoF test is set to $\alpha = 0.05$ and that the number of categories for the test are set to 11, so that $d_f = 10$. The distribution of t under H_1 is a non-central χ^2 with $d_f = 10$ and non-centrality parameter $\lambda = w^2 M$. A distortion in the code correlation function determines the E–L difference D therefore the amount of distortion is able to detect can be measured as a function of μ_D, under H_1, or equivalently as a function of the effect size induced by μ_D. Let define $\xi_\mu = \frac{\mu_D}{\Delta_D}$, where Δ_D is the size of each category in which the observations in D are allocated by the Chi-square test. The power of the Chi-square GoF test is reported in Fig. 1, as a function of the code correlation normalized distortion ξ_μ and for different values of M. Clearly, the power increases with M and it is possible to see that, for $M = 1000$ and $d_f = 10$, the test is able to reach a power >90 % for any code correlation distortion exceeding one third of the category size. The green dotted line represents the effect size w, which increases with μ_D; it determines the non-centrality parameter of the alternative distribution of t through the size M, which is the reason for a greater power associated with a higher M.

Implementation of the techniques in the SW receiver

The detection methods mentioned have been included in one of our software receivers, as C language routines with 32-bits floating point precision for both the mathematical operations and the look-up tables (LUTs) used to represent the standard normal and χ^2 cumulative density functions (CDFs). The TEXBAT datasets offer the chance of testing the detection capability of the proposed algorithms in real situations, in the presence of spoofing attacks. Results obtained validated the methodologies on real data and proved their detection capabilities.

The partial but promising results presented in Truong et al. (2014) encouraged us to refine the detection methods. In this work, both test methods are preliminary calibrated over the "clean" portion of each dataset, before the attack starts and various implementation details have been improved. In this way, the reference distribution function for the Chi-square GoF and the mean and the standard deviation of the reference Gaussian distribution for the Sign test are computed and stored, to be used subsequently during the test phase. The calibration phase is fundamental to fit both algorithms to the target receiver: In fact, it allows compensating for the reshaping effects of radiofrequency front-end filter on the correlation function and the reduced accuracy due to the implementation choices. In the implementation at hand, the E–L correlators spacing for the test was set to $d_{\mathrm{EL}} = 1.5$ chips and the integration time was $T_{\mathrm{int}} = 1$ ms. In addition, the size of the sample used to take a single decision was set equal to 1000 ($M = 1000$), resulting in a test rate of one per second.

Both tests are then completed with a decision criterion in order to warn the user about the presence of a potential spoofing attack. The next subsection describes how an aggregate decision can be taken over a fraction of the tracked satellites.

Taking an aggregate decision on the presence of counterfeit signals

The mentioned detection methods apply to the correlators output of each satellite being tracked by the receiver, i.e., for each satellite,

$$\text{if} \quad \begin{cases} p\text{-value} \le \alpha \\ p\text{-value} > \alpha \end{cases} \quad \text{then} \quad \begin{cases} H = 1 \; (H_0 \text{ rejected}) \\ H = 0 \; (H_0 \text{ accepted}) \end{cases} \tag{3}$$

where H is the test decision. However, definitely an aggregate decision has to be taken in order to warn the host system about the detection of a potential spoofing attack. In this way, the resilience against false alarms is improved. To do this, the hypothesis test decisions H of all the satellites must be considered as a whole. A simple majority rule has been employed in our receiver: Indicating with $N_{S,i}$ the number of satellites tracked at the time instant i, we define the aggregate test

$$\hat{h}_i = \sum_{s=1}^{N_{S,i}} H_i^{(s)} \tag{4}$$

where s indicates the satellite ID and i is the time instant. Thus, indicating with $\theta_{q,i}$ the decision threshold at the time instant i, we decide for the aggregate decision $\hat{H}_i$ as follows

$$\text{if} \quad \begin{cases} \hat{h}_i < \theta_{q,i} \\ \hat{h}_i > \theta_{q,i} \end{cases} \quad \text{then} \quad \begin{cases} \hat{H}_i = 0: & \text{Normal condition} \\ \hat{H}_i = 1: & \text{Attack detected} \end{cases} \tag{5}$$

where the threshold is simply set as

$$\theta_{q,i} = \max\left\{2, \frac{N_{S,i}}{q}\right\} \quad (\text{e.g.,} \quad q = 2, 3) \tag{6}$$

where q is chosen on the basis of the maximum percentage of $N_{S,i}$ the user accepts to raise a warning, before declaring to be under attack.

The decision threshold $\theta_{q,i}$ offers a simple discrimination mechanism between signal-level attack and propagation-induced distortions (mainly, multipath): However, it would be too simplistic to say that "an attack is present when most of the signals hypothesis test decisions are 1 simultaneously," since such a statement does not take into account that a sophisticated attack could spoof a few signals at a time, in variable number and with different levels of induced distortion. For this reason, any optimization on $\theta_{q,i}$, i.e., on q, would be intrinsically a function of the probability of having a certain fraction of satellites simultaneously spoofed, but this depends on the type of attack and cannot be set a priori.

The aggregate decision improves the resilience of the method against false alarms; nonetheless, it does not reduce the vulnerability to attacks conducted one channel at a time. For this reason, the per-channel decision could be paired to a receiver autonomous integrity monitoring (RAIM) approach to detect single-channel attacks, as proposed for example in Ledvina et al. (2010). However, this technique is not developed here.

Spoofing detection tests: analysis of the results

In this section, we discuss the results of our tests against Cases 2–6 of the TEXBAT datasets. The results are obtained by our software receiver enabled with spoofing detection capabilities and therefore are intrinsically dependent on its characteristics.

Cases 2, 3 and 4: static receiver

In the "Static Overpowered Time Push" attack (Case 2) and in the "Static Matched-Power Time Push" attack (Case 3), the spoofer induces an error along the time dimension, i.e., an offset of 2 μs, equivalent to 600 m. On the other hand, in the "Static Matched-Power Position Push" attack (Case 4) the attack leads to a position error of 600 m along the Z dimension.

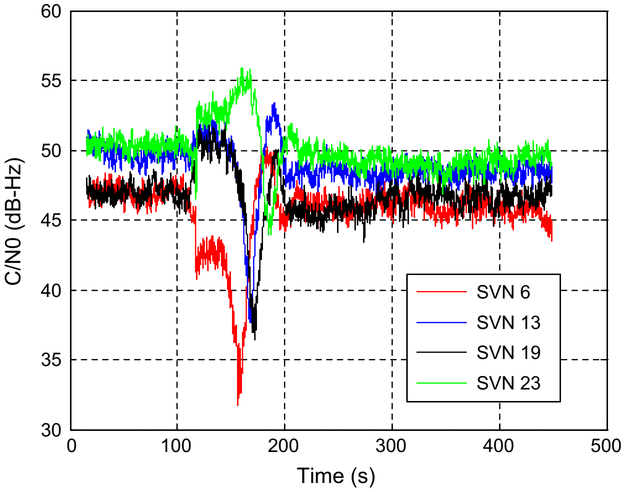


Fig. 2 Case 3: time history of C/N_0 measurements for different SVNs

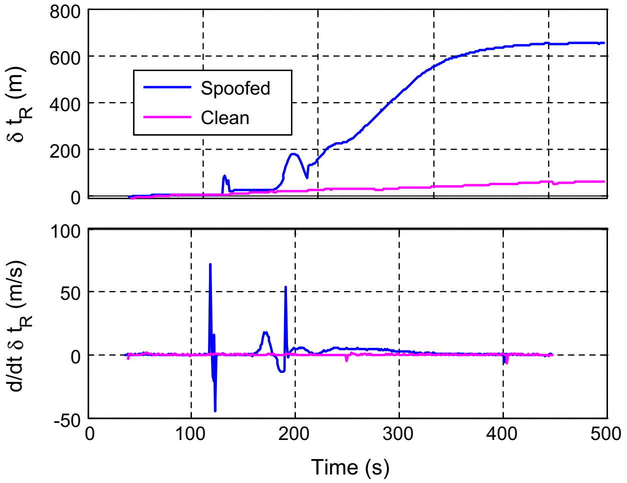


Fig. 3 Case 3: receiver clock offset δt_R and clock offset rate $\dot{\delta t}_R$ for both spoofed and clean datasets

As expected, in all three cases all satellites in lock present variations in the C/N_0 level: As an example, Fig. 2 shows the trend of carrier-to-noise ratio for four PRNs in Case 3, clearly highlighting a critical time interval between 120 and 210 s. Such fluctuations are due to the existing residual differential Doppler in the spoofing signal, as explained in Humphreys et al. (2012). In fact, the imprecise frequency lock of the spoofer to the Doppler shift causes the counterfeit and real phasors to slowly rotate with respect to each other, determining a power leakage from in-phase to quadrature and a consequent loss of the C/N_0 estimate (Falletti et al. 2011).

Considering the time push attacks, Fig. 3 shows the clock offset δt_R and clock offset rate $\dot{\delta t}_R$ measured by the

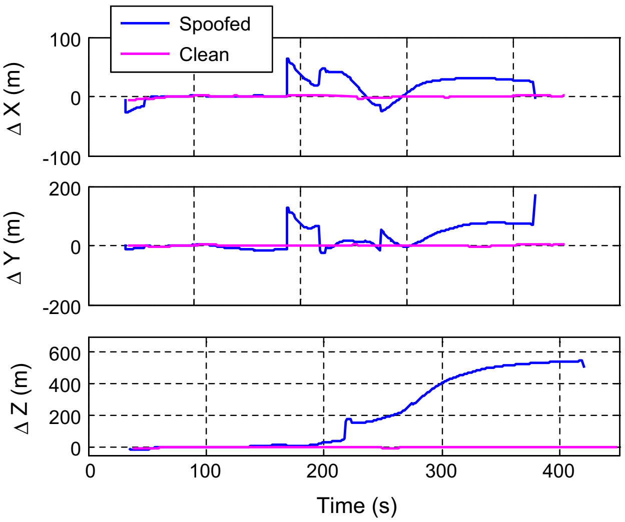


Fig. 4 Case 4: ECEF position deviation from mean for both spoofed and clean datasets

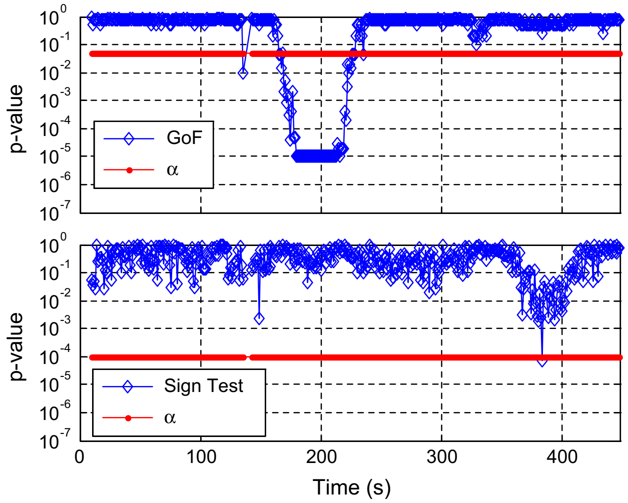


Fig. 5 Case 2: time sequences of the *p*-values for the GoF test (*top*) and the Sign test (*bottom*) obtained for GPS PRN 23

receiver in Case 3. The results are in line with those reported by Humphreys et al. (2012, Figure 16): In Case 3, the spoofer successfully induces a 600 m error in δt_R, and the fluctuations in the clock offset rate $\dot{\delta t}_R$ clearly indicate that the drag-off from the true signal is not smooth.

On the other hand, Fig. 4 reports the earth centered earth fixed (ECEF) receiver position output during the position push attack of Case 4: Substantial errors arise during the interval between 150 and 250 s and are persistent in time, inducing an offset of about 600 m in the *Z*-coordinate. As pointed out in Humphreys et al. (2012), the residual unexpected offset from the true position also along the *X* and *Y* dimensions suggests that only a subset of the receiver channels has locked to the spoofed signals, likely because of the low power advantage of the attack.

On the basis of the theoretical description in the section of Spoofing detection techniques based on statistical analysis, the Chi-square GoF and the Sign test have been applied at the correlators output for spoofing detection purposes. The time sequence of the *p*-values for the spoofed datasets and the relative threshold α is shown in Figs. 5, 6 and 7 for the two methods in the three mentioned cases, respectively, applied to one of the tracked satellites taken as an example. In Case 2, as shown in Fig. 5 (top), the GoF test is able to perform a clear spoofing detection between 150 and 250 s, corresponding to the initial time interval when the spoofer induces a common offset in the Doppler frequency of all signals, as reported in Humphreys et al. (2012, Figure 11). The GoF detection capability stops when the receiver is completely locked to the fake signal, and the authentic and counterfeit signals correlation peaks are no longer overlapped. On the contrary, the *p*-values produced by the Sign test are always above the significance level (bottom panel), meaning that the test is not able to notice the attack. This is due to the fact that the Sign test just checks if μ_D is equal to zero, so that in case of an overpowered attack the signal correlators remain aligned and the detection capability of the Sign test is limited.

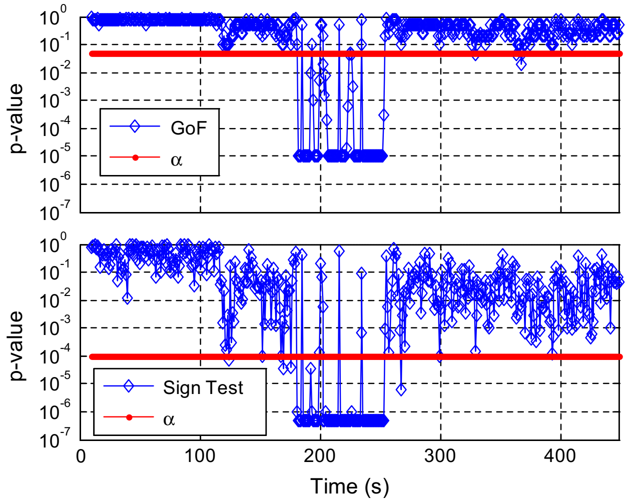


Fig. 6 Case 3: time sequences of the *p*-values for the GoF test (*top*) and the Sign test (*bottom*) obtained for GPS PRN 23

On the other hand, in Cases 3 and 4 (Figs. 6, 7), both GoF and Sign test have the ability to detect an abnormal distortion of the output of the correlators in the test statistic, corresponding to the appearance of the counterfeit signal.

The time series of the decisions H_i for all the tracked satellites of the spoofed dataset in Case 2 are reported in Fig. 8. For the GoF (top panel), the first 100 s of the dataset are clean and the hypothesis on the absence of distortion is

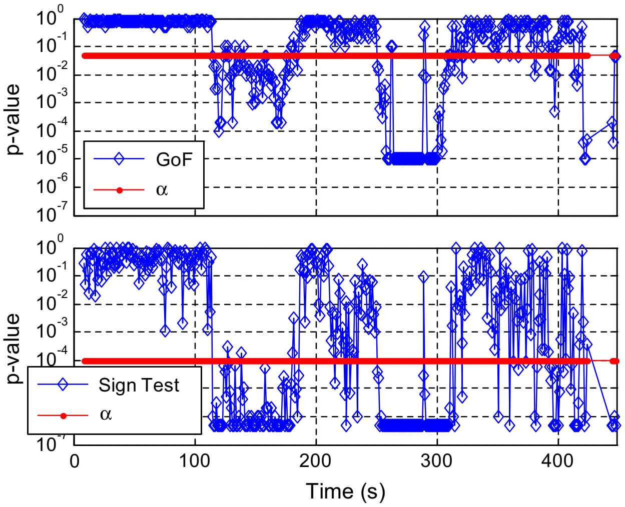


Fig. 7 Case 4: time sequences of the *p*-values for the GoF test (*top*) and the Sign test (*bottom*) obtained for GPS PRN 23

always accepted, $H_i^{(s)} = 0$, while between 150 and 250 s all the satellite correlation functions present detected anomalies. In the rest of the dataset, the decisions H_i are equal to zero for all satellites, since the receiver is completely locked to the spoofed signal and no distortions on the correlation functions are present. On the other hand, for the Sign test $H_i^{(s)} = 0$ almost always, apart for a few false alarms in the first clean part of the dataset, confirming that it is not able to detect an overpowered attack.

On the contrary, in Cases 3 and 4 both methods show the capability of detecting the attack on most channels simultaneously.

Figure 9 reports the time series of the aggregate test metrics $\hat{h}_i$ for the two methods in the three static cases, where they are compared with a possible decision threshold $\theta_{3,i} = \frac{N_{S,i}}{3}$; the corresponding aggregate decision $\hat{H}_i$ is reported in the lower subplot of each panel. It can be noticed that in the clean part of the dataset, i.e., the first 100 s approximately, no false alarms are ever detected, while overall both methods are able to the rise an alert on the possible presence of undesired signals during the attack phase in Cases 3 and 4 (middle and bottom panels) while only the GoF is able to perform the detection in Case 2 (top panel), as already discussed above.

The delay between the attack onset and the instant in which anomalies become detectable is on the order of some tens of seconds. In Cases 2 and 3 shown in Fig. 9 (top and middle), signal anomalies are detected during the initial phase of the time push, particularly between 150 and 250 s in which the clock offset starts increasing at the receiver output (Fig. 3). This is in line with the observation that, once the attack is completed, the receiver definitely unlocks the true signal in favor of the counterfeit one. Due to the huge advantage of the spoofed signal, in the overpowered attack the GoF decisions are clearer for all satellites than in the matched-power case. On the contrary, in Case 4, shown in Fig. 9 (bottom), the anomaly detection starts earlier and is more persistent in time, likely because of the nearly negligible power advantage of the false signal which is not enough to completely unlock the true signal. In particular, two situations can be identified: The aggregate decision test metrics $\hat{h}_i$ is always above the threshold $\theta_{3,i}$ for both methods between 250 and 350 s, when the position error along the *Z*-coordinate is becoming relevant (Fig. 4), while before and after that interval the final decisions $\hat{H}_i$ fluctuate between 0 and 1, likely due to the fact that some channels remain locked to the authentic signal.

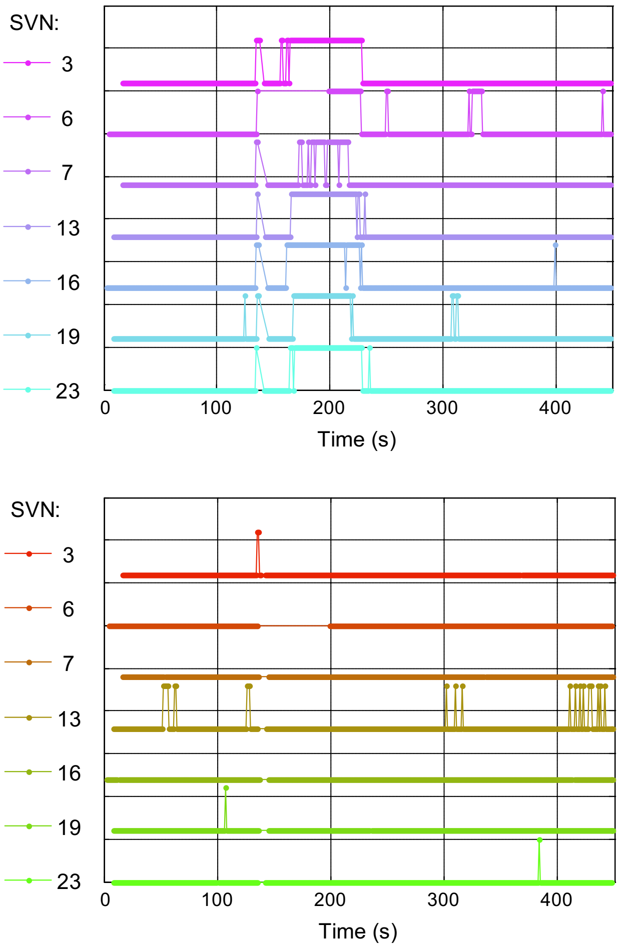
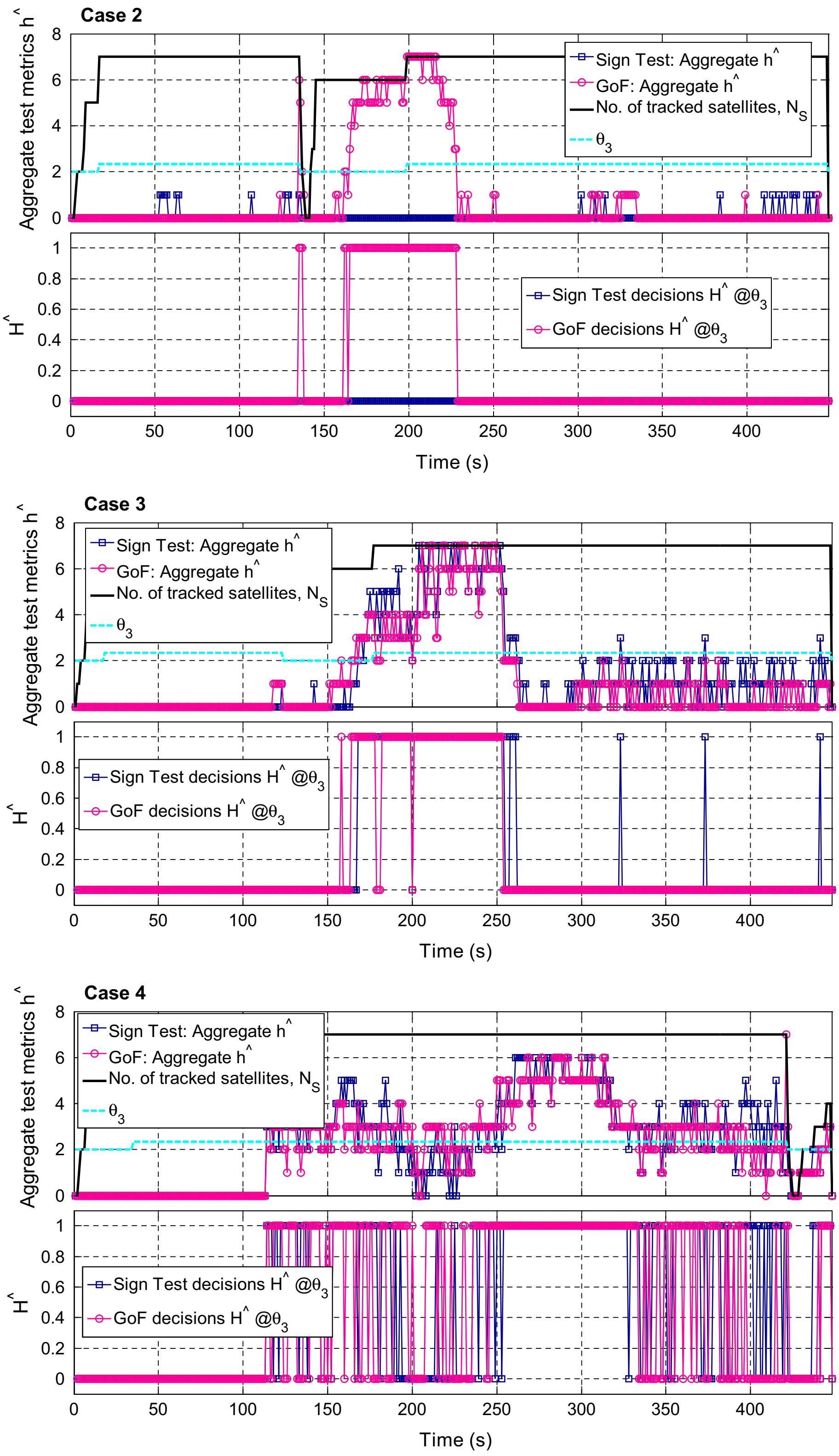


Fig. 8 Case 2: time sequences of the hypothesis test decisions $H_i^{(s)}$ for all the tracked satellites ($s = 1, \ldots, N_{S,i}$) for the GoF test (*top*) and the Sign test (*bottom*)

Cases 5 and 6: dynamic receiver

The same analysis has been conducted on datasets 5 and 6, in which the receiver travels along an urban path and the

Fig. 9 Time sequences of the GoF and Sign test aggregate test metrics $\hat{h}_i$ and aggregate decision $\hat{H}_i$ in Case 2 (*top*); Case 3 (*middle*); and Case 4 (*bottom*)



attack is an “Overpowered Time Push” in Case 5 and a “Matched-Power Position Push” in Case 6. The processing of the static datasets has shown a higher reliability of the GoF test with respect to the Sign test, which has proved to fail in detecting an overpowered attack. For this reason, we decided to proceed with the GoF test only in Cases 5 and 6.

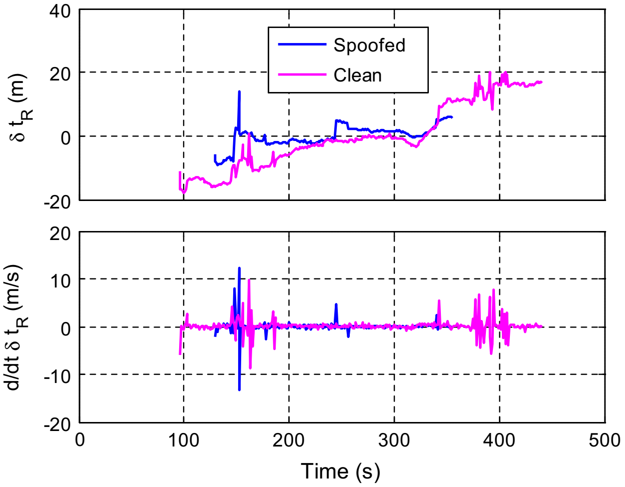


Fig. 10 Case 5: receiver clock offset δt_R and clock offset rate $\dot{\delta t}_R$ for spoofed and clean datasets

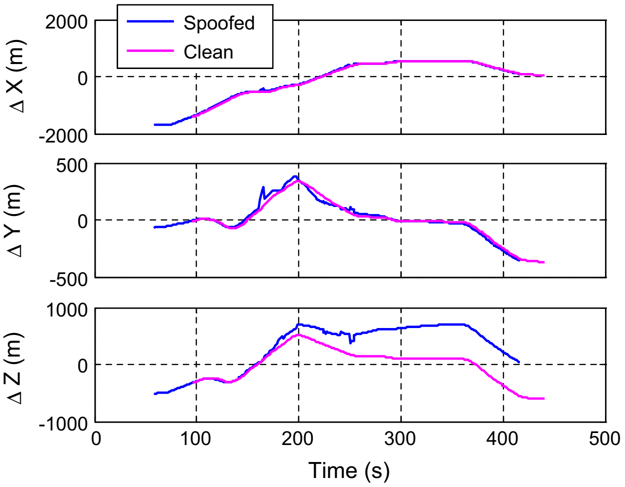


Fig. 11 Case 6: ECEF position deviation from mean for spoofed and clean datasets

Figure 10 shows the clock offset δt_R and clock offset rate $\dot{\delta t}_R$ measured by the target receiver in Case 5. In this case, the receiver raises a “bad time estimation” warning and stops producing PVT outputs after about 350 s.

Results of the ECEF position deviation for Case 6, reported in Fig. 11, are in line with those in Humphreys et al. (2012, Figure 30): This situation is similar to that of Case 4, where the spoofer induces an offset of about 600 m in the Z-coordinate, except that the receiver platform is dynamic.

In order to address the false alarm probability in the presence of multipath, Fig. 12 shows the time series of the decisions H_i for all the tracked satellites of the clean dynamic dataset. The false alarm rate is about 0.007, thus acceptable, because it is smaller than the significance level $\alpha = 0.05$, as expected from the theory.

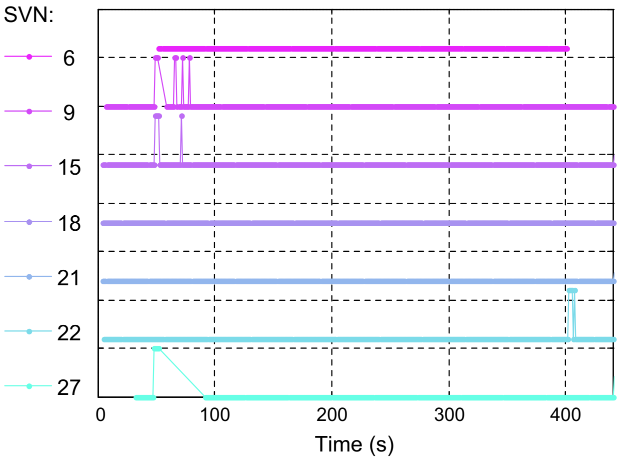


Fig. 12 Case clean dynamic: time sequences of the hypothesis test decisions $H_i^{(s)}$ for all the tracked satellites ($s = 1, \ldots, N_{S,i}$) for the GoF test

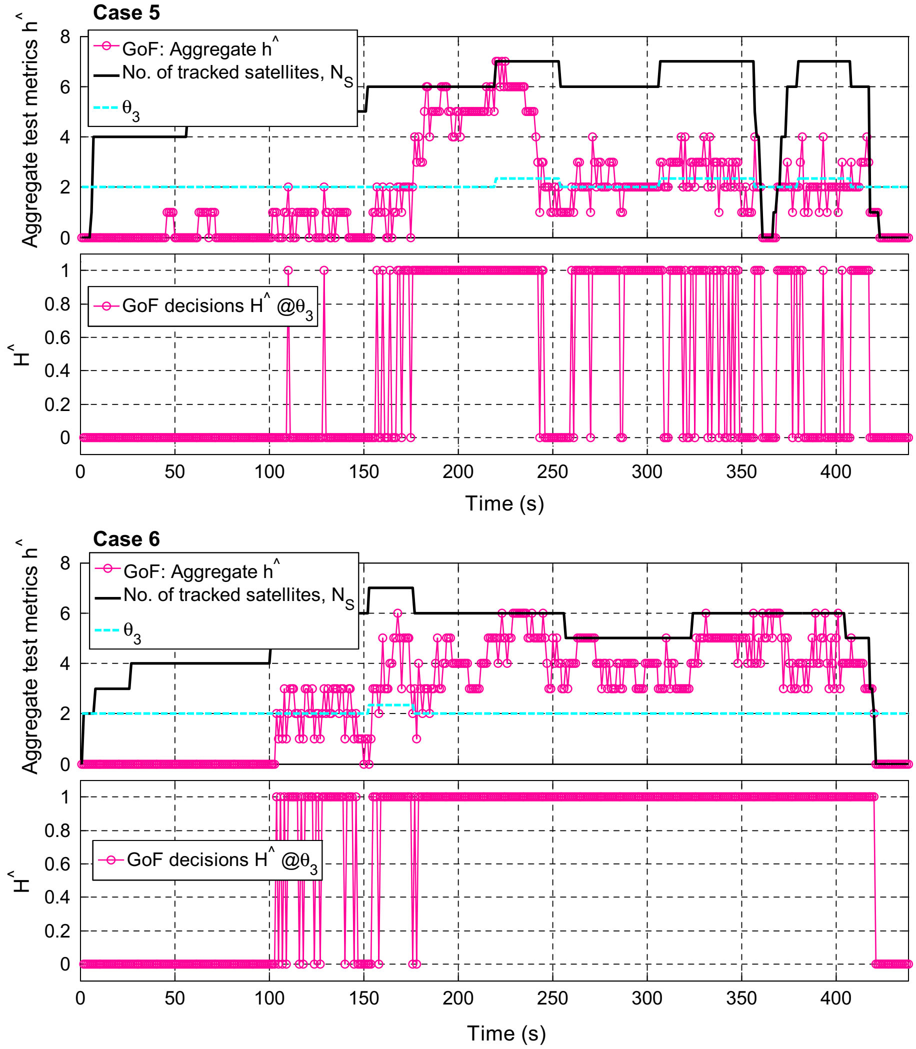
The aggregate test metrics $\hat{h}_i$ and decisions $\hat{H}_i$ for the GoF are shown in Fig. 13. These results are perfectly in line with those obtained for the similar static Cases 2 and 4 in Fig. 9 (top and bottom): No false alarms rise in the clean part of the dataset, while alerts are produced during the attack phase. In Case 5, a clean spoofing detection is performed between 180 and 250 s, but many alarms rise also after 250 s (Fig. 13 top panel): Decisions are more fluctuating than in Case 2 due to the dynamic environment. In Case 6, the aggregate decision $\hat{H}_i$ is persistently stable to 1 from 180 s onward, when the Z-coordinates starts deviating (Fig. 11), till the end of the dataset.

Conclusions and perspectives

We presented the validation of two spoofing detection methods implemented in a software receiver, tested against three static and two dynamic datasets included in the TEXBAT database. The results demonstrate their detection capability: Both the Sign test and Chi-square GoF test proved to be able to correctly warn the user about the presence of a matched-power spoofing signal, although the Sign test has proved to fail in case of an overpowered attack.

These methods and the GoF test in particular demonstrate to be powerful basic tools to implement complete spoofing detection functionality in a receiver. However, this requires a receiver logic that integrates statistical tests, signal quality monitoring, RAIM decisions and observations over time, to encompass various possible types of attack. It is clear that only with a mix of different observations along the receiving chain and suitable consistency

Fig. 13 Time sequences of the GoF aggregate test metrics $\hat{h}_i$ and aggregate decision $\hat{H}_i$ in Case 5 (*top*) and Case 6 (*bottom*)



checks among them, it is possible to make the receiver robust against, or at least aware of, the reception of counterfeit signals.

Future work will be focused on possible strategies to be adopted for the optimization of the test parameters. The study will cover both the implementation of the tests on the single receiver channel and the fine-tuning of the aggregate decision.

Acknowledgments The work of HUST staff is supported by the Project No. VT/CN-02/13-15 of the National S&T Program on Space Technology of Vietnam. Minh Duc Truong's stay in ISMB in 2014 was supported by the Growing NAVIS project, funded by the European Commission under the FP7 Call Galileo.2011.4.3-1—International Activities (Grant Agreement No. 287203).

References

Akos DM (2012) Who's Afraid of the Spoofer? GPS/GNSS spoofing detection via automatic gain control (AGC). Navigation 59(4):281–290

Albright B (1977) Essentials of mathematical statistics. Jones & Burtlett Learning, Burlington

Ali K, Manfredini E, Dovis F (2014) Vestigial signal defense through signal quality monitoring techniques based on joint use of two metrics. In: Proceedings of the ION PLANS 2014. Institute of Navigation, Monterey, CA, May, pp 1240–1247. doi:10.1109/PLANS.2014.6851499

Cohen J (1988) Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, Hillsdale

Falletti E, Pini M, Lo Presti L (2011) Low complexity carrier-to-noise ratio estimators for GNSS digital receivers. IEEE Trans Aerosp Electron Syst 47(1):420–437. doi:10.1109/TAES.2011.5705684

Grant A, Williams P, Ward N, Basker S (2009) GPS jamming and the impact on maritime navigation. Navigation 62(2):173–187

Heng L, Work DB, Gao GX (2015) GNSS signal authentication from cooperative peers. IEEE Intell Transp Syst 16(4):1794–1805

Humphreys TE, Ledvina BM, Psiaki ML, O'Hanlon W, Kintner PM (2008) Assessing the spoofing threat: development of a portable GPS civilian spoofer. In: Proceedings of the ION GNSS 2008. Institute of Navigation, Savannah, GA, September, pp 2314–2325

Humphreys T, Bhatti J, Shepard D, Wesson K (2012) The texas spoofing test battery: toward a standard for evaluating GPS signal authentication techniques. In: Proceedings of the ION

GNSS 2012. Institute of Navigation, Nashville, TN, September, pp 3569–3583

Jafarnia-Jahromi A, Broumandan A, Nielsen J, Lachapelle G (2012a) GPS vulnerability to spoofing threats and a review of antispoofing techniques. Int J Navig Obs. doi:10.1155/2012/127072

Jafarnia-Jahromi A, Broumandan A, Nielsen J, Lachapelle G (2012b) GPS spoofer countermeasure effectiveness based on signal strength, noise power, and C/N0 measurements. Int J Satell Commun Network 30:181–191

Jafarnia-Jahromi A, Broumandan A, Nielsen J, Lachapelle G (2012c) Detection and mitigation of spoofing attack on a vector based tracking GPS receiver. In: Proceedings of the ION ITM 2012. Institute of Navigation, Newport Beach, CA, January, pp 790–800

Jin MH, Choi YS, Choi HH, Lee SJ, Park C (2013) A multiple lock detector for the signal abnormality detection in the GPS receiver. In: Proceedings of the ION GNSS + 2013. Institute of Navigation, Nashville, TN, September, pp 1577–1582

Konovaltsev A, Cuntz M, Haettich C, Meurer M (2013) Autonomous spoofing detection and mitigation in a GNSS receiver with an adaptive antenna array. In: Proceedings of the ION GNSS + 2013. Institute of Navigation, Nashville, TN, September, pp 2937–2948

Ledvina BM, Bencze WJ, Galusha B, Miller I (2010) An in-line antispoofing device for legacy civil GPS receivers. In: Proceedings of the ION ITM 2010. Institute of Navigation, San Diego, CA, January, pp 698–712

Lo Presti L, Falletti E, Nicola M, Troglia Gamba M (2014) Software defined radio technology for GNSS receivers. In: Proceedings of the MetroAeroSpace 2014, Benevento, Italy, May, pp 314–319. doi:10.1109/MetroAeroSpace.2014.6865941

Magiera J, Katulski R (2013) Accuracy of differential phase delay estimation for GPS spoofing detection. In: Proceedings of the 36th international conference on telecommunications and signal processing 2013, September, pp 695–699. doi:10.1109/TSP.2013.6614026

Manfredini EG, Dovis F, Motella B (2014) Validation of a signal quality monitoring technique over a set of spoofed scenarios. In: Proceedings of the 7th ESA workshop on satellite navigation technologies and european workshop on GNSS signals and signal processing 2014, NAVITEC 2014, ESA/ESTEC, Noordwijk, The Netherlands, December. doi:10.1109/NAVITEC.2014.7045136

Montgomery PY, Humphreys TE, Ledvina BM (2009) Autonomous spoofing detection: experimental results of a multiantenna receiver defense against a portable civil GPS spoofer. In: Proceedings of the ITM 2009. Institute of Navigation, Anaheim, CA, January, pp 124–130

Moon GB, Im S-H, Jee G-I (2013) A civil GPS anti-spoofing and recovering method using multiple tracking loops and an adaptive filter technique. In: Proceedings of the ION GNSS + 2013. Institute of Navigation, Nashville, TN, September, pp 2916–2920

Moore DS, McCabe GP (1993) Introduction to the practice of statistics. W. H. Freeman and Company, New York

Pini M, Motella B, Troglia Gamba M (2013) Detection of correlation distortions through application of statistical methods. In: Proceedings of the ION GNSS + 2013. Institute of Navigation, Nashville, TN, September, pp 3279–3289

Psiaki ML, O'Hanlon BW, Bhatti JA, Shepard DP, Humphreys TE (2013) Civilian GPS spoofing detection based on dual-receiver correlation of military signals. IEEE Trans Aerosp Electron Syst 49(4):2250–2267. doi:10.1109/TAES.2013.6621814

Troglia Gamba M, Motella B, Pini M (2013) Statistical test applied to detect distortions of GNSS signals. In: Proceedings of the international conference on localization and GNSS 2013, June, pp 1–6. doi:10.1109/ICL-GNSS.2013.6577267

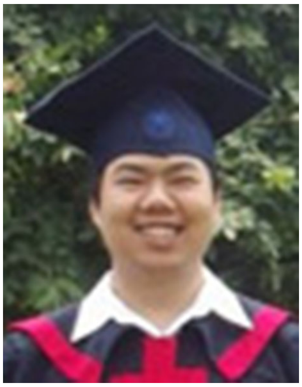
Truong MD, Troglia Gamba M, Motella B, Falletti E, Hai TT (2014) Enabling GNSS software receivers with spoofing detection techniques: a test against some TEXBAT datasets. In: Proceedings of the conference on maritime communication and navigation 2014, Hanoi, Vietnam, October

Wesson KD (2014) Secure navigation and timing without local storage of secret keys. PhD thesis, The University of Texas at Austin, May

Wesson KD, Shepard DP, Bhatti JA, Humphreys TE (2011) An evaluation of the vestigial signal defense for civil GPS anti-spoofing. In: Proceedings of the ION GNSS 2011. Institute of Navigation, Portland, OR, September, pp 2646–2656

Wesson KD, Evans BL, Humphreys TE (2013) A combined symmetric difference and power monitoring GNSS anti-spoofing technique. In: Proceedings of the IEEE global conference on signal and information processing, Austin, TX, December

Micaela Troglia Gamba is a researcher in the Navigation Technologies Area of the Istituto Superiore Mario Boella, Torino, Italy. She received her Co-tutelle Ph.D. in Electronics and Communication Engineering from Politecnico di Torino and Télécom Bretagne, France, in 2011. Her main research activity focuses on implementation and prototyping aspects of GNSS software receivers, both for general purpose processors and for ARM-based embedded platforms.



Minh Duc Truong graduated in June 2012 from the School of Information and Communication Technology, Hanoi University of Science and Technology. He has started to work at Navis Centre as a researcher since July 2012. Currently, he is focusing on researching and developing multi-GNSS software receiver.

Beatrice Motella is a researcher at the Istituto Superiore Mario Boella, Torino, Italy, working within the Navigation Technologies area. She received the Master degree in Communications Engineering in January 2003 at the Politecnico di Torino and the Ph.D. degree in Electronics and Communications Engineering at the same university in 2007. Her main research topic is the radio-frequency interference (RFI) monitoring for GNSS receivers and local components, specifically designed for safety-of-life applications.

Emanuela Falletti is the head of the GNSS core algorithms unit in the Navigation Technologies research area of the Istituto Superiore Mario Boella, Torino, Italy. She obtained her Ph.D. degree in Electronics and Communications Engineering from Politecnico di Torino in 2004. Her current research focuses on digital signal processing techniques and algorithms for advanced GNSS software receivers, in particular for interference and multipath detection and mitigation, as well as signal simulation.

Tung Hai Ta received his M.Sc. degree (2005) in Information Technology from Hanoi University of Science and Technology (HUST), Vietnam, a Master degree in Navigation and Related Application (2006), and PhD degree in Information and Communication Technologies (2010) from Politecnico di Torino, Italy. Since December of 2011, he has been the Director of the NAVIS Centre for Research and Development of Satellite Navigation Technology in Southeast Asia, Hanoi, Vietnam. His research focuses on GNSS signal processing and Nav/Com integration technologies.