2013

Spoof Detection Using Multiple COTS Receivers in Safety Critical Applications

Peter F. Swaszek
University of Rhode Island, swaszek@uri.edu

Richard J. Hartnett

Citation/Publisher Attribution

Swaszek, Peter F. and R. J. Hartnett. "Spoof detection using multiple COTS receivers in safety critical applications," Proc. ION GNSS+ 2013, Nashville TN, Sept. 2013.
Available at: https://www.ion.org/publications/abstract.cfm?articleID=11221

Spoof Detection Using Multiple COTS Receivers in Safety Critical Applications

Peter F. Swaszek, *University of Rhode Island*
Richard J. Hartnett, *U.S. Coast Guard Academy*

Biographies

Peter F. Swaszek is a Professor in the Department of Electrical, Computer, and Biomedical Engineering at the University of Rhode Island. His research interests are in statistical signal processing with a focus on digital communications and electronic navigation systems.

Richard J. Hartnett is a Professor of Electrical Engineering at the U.S. Coast Guard Academy, having retired from the USCG as a Captain in 2009. His research interests include efficient digital filtering methods, improved receiver signal processing techniques for electronic navigation systems, and autonomous vehicle design.

Abstract

While the GPS is well known to be an accurate provider of position information across the globe, its low power level makes it susceptible to spoofing. Given its status as the primary (perhaps only) provider of position in many safety critical applications, this susceptibility is of great concern. Several possible methods to detect a spoofing event at a single GPS receiver have been proposed in the literature. We note, however, that almost all of this prior work has been on the conceptual level; there has been very little analysis of the resulting detection performance.

Recognizing that redundant equipment may already exist for some users, we have proposed to detect spoofing by comparing the position solutions from two or more COTS receivers mounted on the same platform (ION ITM, Jan. 2013). The concept is that the existence of a spoofer would make the statistical relationship of the observed positions different than it would be during normal, non-spoofed, operation. The primary advantage of such an approach is that its implementation does not require receiver hardware modification or even access to software GPS methods; a separate processor could easily monitor the positions generated by each of the receivers and decide spoof versus no spoof. Our earlier paper initiated a performance analysis of the approach; this paper continues and extends the investigation.

Introduction

The GPS is well known to be an accurate provider of position information across the globe. As such, it is commonly used to locate and navigate vessels in various transportation modes (e.g. land vehicles, boats and ships, and aircraft). GPS (or, more generally, GNSS) spoofing refers to intentional, and usually considered malicious, interference of a GPS user's inputs so as to distort that position information. This is in contrast to GPS jamming which attempts to make position information unavailable to the user. Depending upon the cargo and/or mission of the transport, the integrity of the provided position information could be safety critical. Examples that directly come to mind include National Airspace System traffic separation, aircraft approaches, restricted visibility harbor entrance and approach, positioning of buoys, automated port container loading, and truck transport of hazardous cargo.

GPS spoofing is a hot topic of late; technical discussions can vary widely based upon the assumed capabilities and a priori knowledge of the spoofer. In 2003 Warner and Johnston [1] suggested several possible methods to detect a spoofing event at a single GPS receiver: monitoring the power levels of the GPS signals (absolute, relative, and across satellites), checking that the constellation itself is correct for the given time (e.g. number and IDs of the satellites), testing the accuracy of the clock component, and even checking against some non-GPS source (e.g. an IMU). Since then various authors have experimented with spoofing and suggested detectors including correlating the P(Y) code at the RF level [2], looking for vestigial peaks in the correlator outputs [3], comparing to trusted reference signals [4], and using antenna arrays to spatially identify signals [5]. Much of this prior work has been on the conceptual level with limited analysis of the resulting detection performance.

Recently, at the ION's ITM 2013, we proposed a simple spoofing detection concept based on the use of multiple COTS receivers and attempted to assess its performance under nominal assumptions on the signal environment [6].

Specifically, the detector monitors the GPS signals using not one, but two or more receivers with their antennae at known relative positions. With no spoofer present, each antenna would receive a unique RF signal consistent with its position in space. Under the assumption that the spoofer is present, and has only one broadcast antenna, these multiple receivers would receive nearly identical spoofer RF signals, perhaps with a time delay. (We assume that each receiver tracks the spoofer signal, ignoring any true GPS signal that is present.) The presence of spoofing is thus discernible from the near equivalence of the receivers' receptions. While one could compare these multiple receptions at the RF level, we proposed comparing the position solutions across receivers, declaring a spoofing event if the resulting position solutions are too close to each other as compared to the (known) relative locations of the antennae. The primary advantage of such an approach is that an implementation of the hypothesis test does not require receiver hardware modification (hence, no recertification is necessary) or even access to software GPS methods; a separate processor could easily monitor the positions generated by each of the receivers. We note that [7] briefly describes this same approach, but without providing any analysis. Our January 2013 work developed several different detection algorithms (based on differences in the knowledge of the receivers' locations; e.g. known relative position with and without orientation information) and analyzed each detector from a Neyman-Pearson perspective assuming Gaussian statistics on the position measurement errors. This prior work assumed quite general Gaussian models; the result was an inability to formulate the optimum tests which led us to selecting ad hoc detectors for the 2 and 3 antennae cases. In this paper we examine a simpler model that allows us to develop optimum tests for any number of receivers. Examples are then presented. We focus on two dimensional solutions (latitude and longitude), commenting on the extension to altitude in the future works section.

Notation

Imagine a configuration of m GPS antennae/receivers, each of which provides a two dimensional position solution based upon its observed RF signals (while latitude and longitude are the nominal coordinates, we will assume that they are converted to East and North in a local reference frame). For simplicity of the resulting analysis, we will parameterize the position of each antenna as a point on the complex plane relative to some fixed origin. Specifically, the k^{th} antenna, $k = 1, \dots m$, is at position $d_k = d_{k,r} + jd_{k,i}$ (in this decomposition into real and imaginary components, we will think of the real part as the East component and the imaginary part as the North component of the position). Further, and without loss of generality, we will assume that the origin of our reference frame is such that the centroid of these antennae positions is zero, so that

$$\sum_{k=1}^{m} d_k = 0$$

Our interest is in mounting this array of antennae onto a moving platform; hence, relative to the location of the centroid, the array could have a random orientation. Keeping the array horizontal, we model this as an angular rotation by angle θ (in radians) on the complex plane. As such, the position of the k^{th} antenna is now $d_k e^{j\theta}$. Further, we note that even with the rotation, the centroid is still zero

$$\sum_{k=1}^{m} d_k e^{j\theta} = e^{j\theta} \sum_{k=1}^{m} d_k = e^{j\theta} \cdot 0 = 0$$

For our spoof detector each antenna processes the RF signals it receives, yielding an estimate of its position; this position is a complex number (also in East/North coordinates) which we will denote x_k. We will assume that the error in this estimate is dominated by additive white Gaussian noise, so will employ complex Gaussian distributions when describing the statistics of these positions. For simplicity we will assume that these noise effects are mutually independent, extending the discussion later.

The Hypotheses

We consider two situations, the null hypothesis, H_0, in which no spoofer is present and the alternative hypothesis, H_1, in which a spoofer is present:

H_0: With no spoofer present we assume that each individual antenna is giving an accurate estimate of its actual positions. For notation, let b represent the true position of the centroid of the antennae array; including this position offset, the rotation for each antenna, and an additive complex Gaussian noise term (n_k), we have a model for the position observations of

$$x_k = b + d_k e^{j\theta} + n_k$$

for $k = 1, 2, \dots m$.

H_1: With a spoofer present we assume that the individual antennae all receive identical RF signals; hence, all would provide noisy estimates of the same constant position (with only one radiator, a spoofer can create only one possible position solution [7]). Letting c represent this spoofed position, we have the observation model

$$x_k = c + n_k$$

for $k = 1, 2, \dots m$. Note that this is independent of the antennae offsets and the rotation angle.

Hypothesis Testing

We imagine a Neyman-Pearson formulation for this problem and wish to develop a binary hypothesis test with fixed probability of false alarm (the probability of deciding H_1 when H_0 is true) and maximum power or probability of detection (the probability of deciding H_1 when H_1 is true). Hypothesis testing is usually implemented by computing a scalar function of the observation data, $T(x_1, \ldots x_m)$, called the test statistic and comparing this value to a constant called the threshold. If the test statistic exceeds the threshold, we decide H_1; if not, we decide H_0. Symbolically, we write this as

$$T\left(x_1, \ldots x_m\right) \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda$$

The optimum test statistic for the Neyman-Pearson formulation is well known to be the likelihood ratio test [8]

$$T\left(x_1, \ldots x_m\right) = \frac{f\left(x_1, \ldots x_m | \mathrm{H}_1\right)}{f\left(x_1, \ldots x_m | \mathrm{H}_0\right)}$$

which is the ratio of the conditional probability density functions (pdfs) of the data under the two hypotheses. Usually, one simplifies the algebraic form of this test by taking monotonic functions of the result (e.g. the natural logarithm is very common for independent observations) and ignoring any additive and positive multiplicative terms that are independent of the data. As noted in the section above, we will assume that the pdfs are complex Gaussian.

If one has a complete characterization of the two hypotheses, then the development of the test statistic is usually quite straightforward. The work, then, is the development of the expressions for the probability of false alarm (so that the threshold can be selected) and the probability of detection, the resulting performance. If some of the parameters are unknown, additional analysis is required.

All Parameters Known

As stated above, the observation consists of position measurements, x_k, $k = 1, 2, \ldots m$, each with independent complex Gaussian statistics. Under hypotheses H_0 and H_1, these are

$$x_k \sim \mathcal{CN}\left(b + d_k e^{j\theta}, 2\sigma^2\right) \quad \text{and} \quad x_k \sim \mathcal{CN}\left(c, 2\sigma^2\right)$$

respectively. The notation $x \sim \mathcal{CN}\left(\mu, \Gamma\right)$ implies that the vector consisting of the real and imaginary parts of x_k has a bivariate Gaussian distribution with mean vector equal to the real and imaginary parts of μ, respectively, and covariance matrix

$$\frac{1}{2}\left[\begin{array}{cc} \Re\left\{\Gamma\right\} & -\Im\left\{\Gamma\right\} \\ \Im\left\{\Gamma\right\} & \Re\left\{\Gamma\right\} \end{array}\right]$$

($\Re\left\{\cdot\right\}$ and $\Im\left\{\cdot\right\}$ are the real and imaginary parts of the argument, respectively.) Under these assumptions, the likelihood ratio test is

$$T = \prod_{k=1}^{m} \frac{\dfrac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x_k - c)(x_k - c)^*}}{\dfrac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left(x_k - b - d_k e^{j\theta}\right)\left(x_k - b - d_k e^{j\theta}\right)^*}}$$

(Note that we have dropped the explicit dependence on measurement in the notation T for brevity of the notation. Further, for this Gaussian model, σ^2 is the variance of both the real and imaginary components of the noise, assumed independent, and the superscript * represents complex conjugate. This formulation is usually called *proper* complex Gaussian.) Taking the natural logarithm, simplifying the algebra, and ignoring any additive or positive multiplicative constants yields an equivalent test statistic (denoted T') with a different threshold (denoted λ')

$$T' = \sum_{k=1}^{m} \left[\begin{array}{c} (c - b - d_k e^{j\theta})^* x_k \\ + \left(c - b - d_k e^{j\theta}\right) x_k^* \end{array}\right] \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda'$$

or since the sum of a complex number and its conjugate is twice its real part

$$T' = \sum_{k=1}^{m} 2\Re\left\{\left(c - b - d_k e^{j\theta}\right)^* x_k\right\} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda'$$

The Statistics of the Test

First, we note that the test statistic is a real scalar (as it should be). Also, as a linear combination of Gaussian variables, the test statistic itself is also Gaussian distributed under both hypotheses. Specifically,

$$T' \sim \mathcal{N}\left(\mu_0, \sigma_T^2\right) \quad \text{and} \quad T' \sim \mathcal{N}\left(\mu_1, \sigma_T^2\right)$$

for H_0 and H_1, respectively (here the notation $\mathcal{N}\left(\mu, \sigma^2\right)$ refers to a *real* Gaussian random variable with mean μ and variance σ^2). To specify the parameters of these distributions, recognize that the test statistic is of the form

$$T' = \sum_{k=1}^{m} \left(a_k^* x_k + a_k x_k^*\right)$$

with

$$a_k = c - b - d_k e^{j\theta}$$

Thus, each term in the sum has mean

$$a_k^* \mu_k + a_k \mu_k^*$$

and variance

$$2|a_k|^2 \sigma^2$$

(see Appendix A). Since the mean of a sum is the sum of the means, then the test statistic has distinct means under the two hypotheses of

$$\mu_0 = \sum_{k=1}^{m} \left[a_k^*(b + d_k e^{j\theta}) + a_k(b + d_k e^{j\theta})^*\right]$$

and

$$\mu_1 = \sum_{k=1}^{m} \left[a_k^* c + a_k c^*\right]$$

Furthermore, under the assumption that the x_k were independent random variables, then the variance of the sum formed in T' is equal to the sum of the variances, and the test statistic has common variance under the two hypotheses

$$\sigma_T^2 = 2 \sum_{k=1}^{m} \left|c - b - d_k e^{j\theta}\right|^2 \sigma^2$$

All three of these parameters can be simplified. Expanding the product for μ_0 and combining terms, we have

$$\begin{aligned} \mu_0 \;\; &= \;\; \sum_{k=1}^{m} \left[(c-b)^* b + (c-b) b^*\right] - 2 \sum_{k=1}^{m} |d_k|^2 \\ &+ \;\; \sum_{k=1}^{m} d_k e^{j\theta} \left(c^* - 2b^*\right) + \sum_{k=1}^{m} d_k^* e^{-j\theta} \left(c - 2b\right) \end{aligned}$$

To further simplify μ_0 we use the following two facts:

- The terms in the first sum are independent of k.
- Since the centroid of the antennae positions is identically zero, multiplying each term by a complex constant s has no effect

  $$\sum_{k=1}^{m} d_k s = s \sum_{k=1}^{m} d_k = s \cdot 0 = 0$$

  Letting s be $e^{j\theta}(c^* - 2b^*)$, then the third sum is zero. The fourth sum is just the conjugate of the third, so it is zero as well.

The result is that μ_0 reduces to

$$\begin{aligned} \mu_0 \;\; &= \;\; m\left[(c-b)^* b + (c-b) b^*\right] - 2 \sum_{k=1}^{m} |d_k|^2 \\ &= \;\; 2m \Re\left\{(c-b)^* b\right\} - 2 \sum_{k=1}^{m} |d_k|^2 \end{aligned}$$

Following a similar argument, μ_1 reduces to

$$\mu_1 = 2m \Re\left\{(c-b)^* c\right\}$$

Finally, using the same set of facts and methods, the variance becomes

$$\sigma_T^2 = 2m\left|c-b\right|^2 \sigma^2 + 2\sigma^2 \sum_{k=1}^{m} \left|d_k\right|^2$$

The Performance of the Test

For any test with Gaussian statistics the false alarm probability is

$$\mathrm{P_{fa}} = \mathrm{Prob}\left(T' > \lambda' | \mathrm{H}_0\right) = Q\left(\frac{\lambda' - \mu_0}{\sigma_T}\right)$$

($Q(\cdot)$ being the Gaussian tail probability). If $\mathrm{P_{fa}}$ is fixed (which is typical for a Neyman-Pearson formulation), then we can solve for the threshold as

$$\lambda' = \sigma_T Q^{-1}(\mathrm{P_{fa}}) + \mu_0$$

The power, or the detection probability, of the test is then

$$\begin{aligned} \mathrm{P_d} &= \mathrm{Prob}(T' > \lambda' | \mathrm{H}_1) = Q\left(\frac{\lambda' - \mu_1}{\sigma_T}\right) \\ &= Q\left(Q^{-1}(\mathrm{P_{fa}}) + \frac{\mu_0 - \mu_1}{\sigma_T}\right) \end{aligned}$$

Substituting in our expressions for the means and standard deviation, and simplifying, yields

$$\mathrm{P_d} = Q\left(Q^{-1}(\mathrm{P_{fa}}) - \frac{\sqrt{2m|c-b|^2 + 2\sum_{k=1}^{m} |d_k|^2}}{\sigma}\right)$$

Unknown Location Parameters, b and c

The expression for $\mathrm{P_d}$ above depends upon knowledge of the actual location variables b and c. For a moving platform, b changes with its position in space. If the value of b is fixed, then the spoofer can minimize his/her probability of detection by making the second argument of the Q function as small as possible; hence, selecting $c = b$. In other words, by placing the spoofed position at the centroid of the actual antennae positions.

With this selection, the test statistic reduces to one invariant to the actual positions

$$T' = \sum_{k=1}^{m} 2\Re\left\{-d_k^* e^{-j\theta} x_k\right\} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \lambda'$$

with threshold

$$\lambda' = \sigma_T Q^{-1}(\mathrm{P_{fa}}) - 2 \sum_{k=1}^{m} |d_k|^2$$

This resulting test follows the expected form for a Gaussian scenario: first, correct for any known scaling (in this case, by the rotation) and then correlate against the known signal (in this case, with the antennae positions). Its performance is

$$\mathrm{P_d} = Q\left(Q^{-1}(\mathrm{P_{fa}}) - \frac{\sqrt{2\sum_{k=1}^{m} \left|d_k\right|^2}}{\sigma}\right)$$

Unknown Angle of Rotation, θ

There are several scenarios we could consider with respect to the angle θ:

- In a few applications both the actual position b and the rotation angle are known and fixed; one such example is an antennae installation at a GPS reference site in which the angle can be selected by the system designer. An obvious question to consider is how to place the array for best performance. We defer this problem to a future study.

- In a second scenario, information on the rotation might be available to the system; for example, through real time data from a high quality compass. In this case, the results presented above hold and obvious questions include how to select the array positions. We defer examining examples under this case as well, recognizing that we would want to place the array to limit the worst case performance.

- Finally, if the angle is unknown, a common approach, called the *generalized likelihood test* (GLRT) [8], involves first estimating the unknown parameter and then using that value in the likelihood ratio test. We pursue that approach below, developing the maximum likelihood estimate of θ.

Normally one would estimate θ under both hypotheses, using the two solutions in their corresponding portions of the likelihood ratio. We note that under H_1 the data is independent of θ, so need only to estimate it under H_0.

Conditioned on H_0 and θ, the pdf of the observed data is a product of bivariate complex Gaussian pdfs

$$\begin{aligned}&f\left(x_1,\dots x_m|\mathrm{H}_0,\theta\right)\\&\quad=\prod_{k=1}^{m}\frac{1}{2\pi\sigma^2}e^{-\frac{1}{2\sigma^2}\left(x_k-b-d_ke^{j\theta}\right)\left(x_k-b-d_ke^{j\theta}\right)^*}\end{aligned}$$

Maximizing this function over the angle θ can be accomplished by setting the derivative to zero

$$\frac{\partial f\left(x_1,\dots x_m|\mathrm{H}_0,\theta\right)}{\partial\theta}=0$$

After some tedious algebra we find the necessary condition

$$e^{-j\theta}\sum_{k=1}^{m}(x_k-b)d_k^*=e^{j\theta}\sum_{k=1}^{m}d_k(x_k-b)^*$$

Expanding both sides, and using the fact that the sum of the d_k is zero, yields

$$e^{j2\theta}=\frac{\sum_{k=1}^{m}d_k^*x_k}{\sum_{k=1}^{m}d_kx_k^*}=\frac{\sum_{k=1}^{m}d_k^*x_k}{\left(\sum_{k=1}^{m}d_k^*x_k\right)^*}$$

The second form of this expression makes it very clear that the ratio has unit magnitude (being a number divided by its conjugate) and that its phase angle is twice that of the numerator; hence, we can write

$$e^{j\theta}=\frac{\sum_{k=1}^{m}d_k^*x_k}{\left|\sum_{k=1}^{m}d_k^*x_k\right|}$$

so

$$\theta=\tan^{-1}\left(\frac{\Im\left\{\sum_{k=1}^{m}d_k^*x_k\right\}}{\Re\left\{\sum_{k=1}^{m}d_k^*x_k\right\}}\right)$$

Plugging this result back into the test statistic and simplifying, the GLRT is

$$T'(x_1,\dots x_m)=-\left|\sum_{k=1}^{m}d_k^*x_k\right|\ \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}}\ \lambda'$$

Sanity Check

Consider the case of m sensors forming a regular m-gon inscribed within a circle of radius r so that

$$d_k=re^{j2\pi k/m}=r\cos\frac{2\pi k}{m}+jr\sin\frac{2\pi k}{m}$$

for $k=1,\dots m$. Then the GLRT test statistic is

$$T'=-r\left|\sum_{k=1}^{m}x_ke^{j2\pi k/m}\right|$$

Without loss of generality we can drop the r, so

$$T'=-\left|\sum_{k=1}^{m}x_ke^{j2\pi k/m}\right|$$

If $m=2$, this is

$$\begin{aligned}T'(x_1,x_2)&=-\left|x_1e^{j\pi}+x_2e^{j2\pi}\right|\\&=-\left|-x_1+x_2\right|\\&=-\left|x_2-x_1\right|\end{aligned}$$

the negative of the distance between the two position estimates; in other words, nearly equal positions implies H_1 while a large spacing implies H_0. We note that this is the test statistic that we considered for $m=2$ in [6].

If $m=3$, the GLRT test statistic is

$$\begin{aligned}T'(x_1,x_2,x_3)&=-\left|x_1e^{j2\pi/3}+x_2e^{j4\pi/3}+x_3e^{j6\pi/3}\right|\\&=-\left|x_1e^{j2\pi/3}+x_2e^{j4\pi/3}+x_3\right|\end{aligned}$$

a test of size and "triangularness".

Statistics of the GLRT

Our test is

$$T' = -\left|\sum_{k=1}^{m} d_k^* x_k\right| \underset{H_0}{\overset{H_1}{\gtrless}} \lambda'$$

To start an analysis of its statistics, let's consider the inner summation

$$y = \sum_{k=1}^{m} d_k^* x_k$$

As a linear combination of complex Gaussian random variables, y is also complex Gaussian with the following statistics:

H_0: Since in this case

$$x_k = b + d_k e^{j\theta} + n_k$$

then

$$\begin{aligned} y &= \sum_{k=1}^{m} d_k^* \left(b + d_k e^{j\theta} + n_k\right) \\ &= b\left(\sum_{k=1}^{m} d_k\right)^* + e^{j\theta}\sum_{k=1}^{m} |d_k|^2 + \sum_{k=1}^{m} d_k^* n_k \end{aligned}$$

Invoking the centroid assumption, the first of these terms is identically zero. The second term is a known real constant (since the array locations are known) with an unknown rotation angle (θ). The third term, a sum of scaled, independent, zero mean and variance $2\sigma^2$ complex Gaussian variates, is a Gaussian variate with zero mean and variance scaled by the sum of the squares of the magnitudes of the individual weights. Letting

$$\beta^2 = \sum_{k=1}^{m} |d_k|^2$$

we have

$$y \sim \mathcal{CN}\left(e^{j\theta}\beta^2, 2\beta^2\sigma^2\right)$$

Specifically, under H_0, y is a complex Gaussian variate with mean somewhere on a circle of radius β^2 centered about the origin.

H_1: Now our assumed observation model is

$$x_k = c + n_k$$

so

$$\begin{aligned} y &= \sum_{k=1}^{m} d_k^* \left(c + n_k\right) \\ &= c\left(\sum_{k=1}^{m} d_k\right)^* + \sum_{k=1}^{m} d_k^* n_k \end{aligned}$$

The first of these terms is again zero and the second is a zero mean Gaussian variate; hence, under H_1,

$$y \sim \mathcal{CN}\left(0, 2\beta^2\sigma^2\right)$$

At this point it is convenient to reverse the direction of the test (to remove the negative sign); the result is a new, but equivalent test which we will denote with two primes

$$T''(x_1, \ldots x_m) = |y| \underset{H_1}{\overset{H_0}{\gtrless}} \lambda'$$

In words, if the test variable $|y|$ is within λ' units of the origin we decide spoofing; if $|y|$ is outside of a circle of this radius, we decide no spoofing.

The performance of this test is as follows (see Appendix B for details):

- The power of the test is the probability under H_1 that the test statistic is smaller than the threshold

  $$\begin{aligned} P_d &= \text{Prob}_{H_1}\left(|y| < \lambda'\right) \\ &= 1 - e^{-\frac{\lambda'^2}{2\beta^2\sigma^2}} \end{aligned}$$

  Note that we can solve this expression for the threshold λ'

  $$\lambda' = \sqrt{-2\beta^2\sigma^2 \ln\left(1 - P_d\right)}$$

- The false alarm of the test is the probability under H_0 that the test statistic is smaller than the threshold

  $$\begin{aligned} P_{fa} &= \text{Prob}_{H_0}\left(|y| < \lambda'\right) \\ &= 1 - Q\left(\frac{\beta}{\sigma}, \frac{\lambda'}{\sigma\beta}\right) \end{aligned}$$

  in which $Q(\cdot,\cdot)$ is Marcum's Q function [8, pp.344-346]

We note that by inserting the expression above for λ' into this last expression, we have

$$P_{fa} = 1 - Q\left(\frac{\beta}{\sigma}, \sqrt{-2\ln\left(1 - P_d\right)}\right)$$

We acknowledge that this expression is backwards, that it is more usual to write the detection probability as a function of the false alarm probability. However, the utility of this closed-form expression is that for a fixed P_d and noise standard deviation σ, the known monotonically of Marcum's Q function in its arguments implies that our test's performance improves with increasing β. Below we will focus on maximizing β (actually, for simplicity, maximizing β^2) over the antennae positions

An Example

It was noted above that the performance of the GLRT improves with larger β^2. Now it seems, for a moment, that the performance is not a function of the antennae array geometry (since β^2 is just the sum of the squares of the distances to each antenna). However, we must recall that the d_k must also satisfy the centroid condition, so the geometry does come into play. Consider, as an example, the case of $m = 3$ antennae and let's require that each antenna fall within a circle of radius r about the origin. Without loss of generality, place the first antenna on the negative real axis at $d_1 = -r$. Also, for sake of argument, place the other two at conjugate positions of $p \pm jq$ such that

$$\sqrt{p^2 + q^2} \leq r$$

(since the first antenna's location is real, the second and third must have opposite imaginary parts so as to satisfy the centroid condition). Then it appears that β^2 is

$$\begin{aligned} \beta^2 &= \sum_{k=1}^{m} |d_k|^2 \\ &= r^2 + 2(p^2 + q^2) \end{aligned}$$

However, the centroid condition requires that

$$\begin{aligned} \sum_{k=1}^{m} d_k &= -r + p + jq + p - jq \\ &= 2p - r = 0 \end{aligned}$$

or p must equal $r/2$, so

$$\beta^2 = \frac{5}{4} r^2 + 2q^2$$

Maximizing this expression over q subject to the constraint of being within the radius r circle yields three points at the vertices of the equilateral triangle inscribed within the circle.

Optimum Antennae Configurations

Let's consider the general case with $m \geq 2$ antennae, recalling that our goal is to understand the relationship between the antennae locations and the resulting performance. Since the performance is optimized by maximizing β^2, we have the optimization problem

$$\beta^2_{\max} = \max_{\{d_k\}} \sum_{k=1}^{m} |d_k|^2$$

subject to

$$\sum_{k=1}^{m} d_k = 0$$

To yield interesting results, we will further constrain all of the antennae to fall within a circle of radius r on the complex plane, so that

$$|d_k| \leq r$$

for all k. Normally, we would attempt to directly solve the optimization problem with the two constraints (this one of bounded radius plus the centroid constraint), but this direct approach appears difficult. Instead, we proceed by constructing an upper bound to β^2 and then showing that since the result is achievable, it is, in fact, the optimum result.

To develop the upper bound we relax the centroid constraint on the d_k and solve the optimization problem with the antennae locations limited to the circle of radius r. Without the centroid constraint, we can write

$$\begin{aligned} \beta^2_{\max} &= \max_{\{d_k\}} \sum_{k=1}^{m} |d_k|^2 \\ &\leq \sum_{k=1}^{m} \max_{\{d_k\}} |d_k|^2 = \sum_{k=1}^{m} r^2 = mr^2 \end{aligned}$$

We observe that this upper bound occurs when all of the antennae are located precisely on the circle of radius r. We also note that this is an upper bound since we did not require that the d_k meet the centroid constraint. However, if we distribute the m antennae uniformly about the circle of radius d so that

$$d_k = re^{j2\pi k/m}$$

then $|d_k| = r$ and $\sum_{k=1}^{m} d_k = 0$, so the maximum is achievable! In general, we have for this problem

$$\beta^2 \leq mr^2$$

The corresponding performance expressions are

$$\mathrm{P_{fa}} = 1 - Q\left(\frac{\sqrt{mr^2}}{\sigma}, \frac{\lambda'}{\sigma\sqrt{mr^2}}\right)$$

and

$$\mathrm{P_d} = 1 - e^{-\frac{\lambda'^2}{2mr^2\sigma^2}}$$

These expressions allow us to compute Receiver Operating Characteristic (ROC) curves [8], a standard way to show the performance of a hypothesis test, to show the dependence of the performance upon both the number of antennae, m, and the spacing ratio, r/σ. As a first example, we set r/σ equal to 3 and vary m from 2 to 5. Figure 1 shows the resulting ROC curve, limiting the false alarm probability to 1% to show the detail. As expected, the performance is better for more sensors. As a second example, we set m at 3 sensors and vary r/σ from 1 to 4. Figure 2 shows the ROC curve, again limiting the false alarm probability to 1% to show the detail. As expected, the performance is better for more separation between sensors. From these curves, 3 or 4 antennae spaced on a circle of radius 4σ or more yields very good performance. To get vanishingly small $\mathrm{P_{fa}}$ and $\mathrm{P_d}$ very near to unity, we need only space the antennae more.

We close this section with several comments on this optimized configuration:

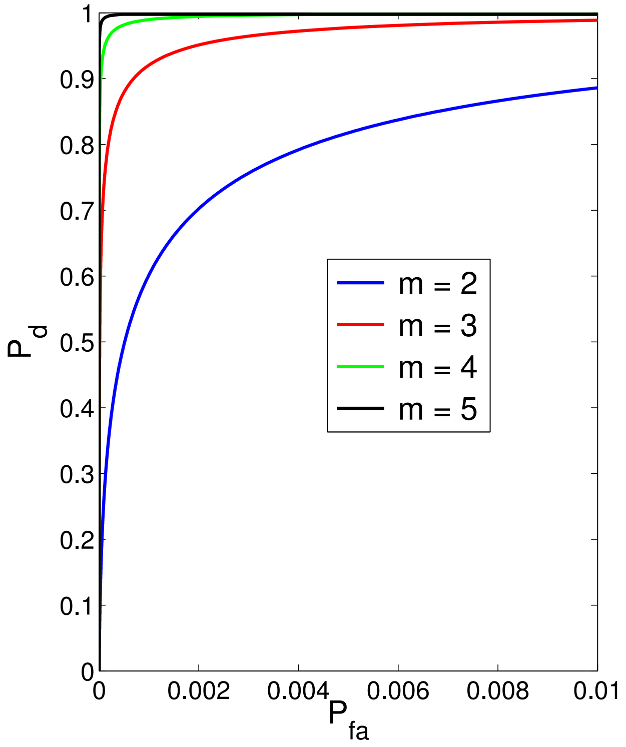

Figure 1: ROC curves with $r/\sigma = 3$ and different numbers of sensors (m).

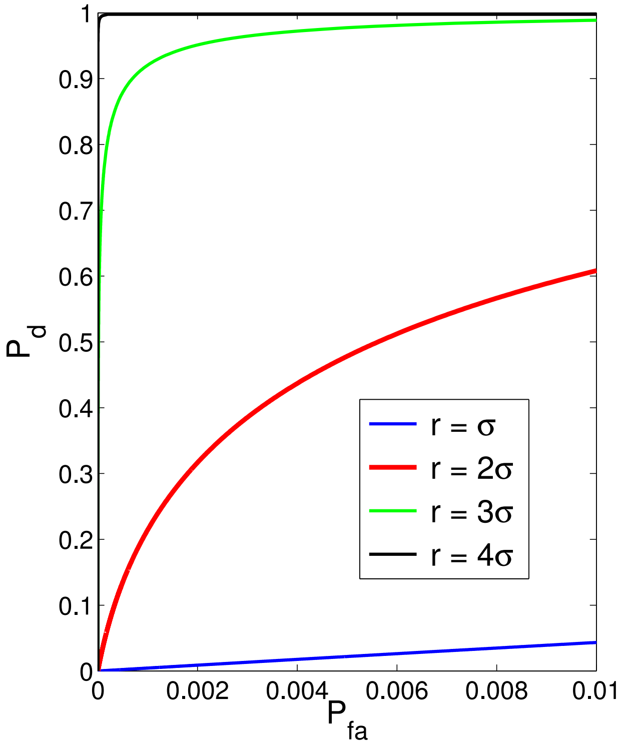

Figure 2: ROC curves with $m = 3$ and different sensor spacing (r).

- The maximum β^2 is achieved by spreading the antennae on the circle of radius r; specifically, placing any of the antennae inside the circle is suboptimum!

- Consider the case of $m = 4$. While the maximum is clearly achieved with the antennae forming a square, the same value of β^2 results from placing the antennae on the corners of any rectangle inscribed within the circle (e.g. at locations $\pm p \pm jq$ with $p^2 + q^2 = r^2$); equivalently, two pairs of diametrically opposed antennae. Theoretically, we could let $q = 0$ and use collocated antennae, but assume that our model of independent x_k would break down in this case.

- For $m = 5$ we could space the antennae 72° apart to achieve the maximum β^2. It appears from our results that we could also place three on an equilateral triangle and the other two diametrically opposed on the circle.

- For larger m we could either uniformly space the antennae, or decompose m into the sum of smaller integers and combine the optimum patterns (with arbitrary rotations) of those smaller counts. Intriguing!

Conclusions

We have presented an approach to GPS spoof detection based upon using multiple COTS receivers. The contributions of the current work include:

- Under the assumptions of Gaussian measurement errors and a two-dimensional (horizontal) restriction of the positions, we formulated the problem in a Neyman-Pearson sense and provided an exact analysis of performance.

- The development was then extended to allow for an unknown platform rotation; exact performance expressions were again developed.

- The resulting performance was optimized over the antennae locations.

- Examples were presented to show that with just a few antennae (3 or 4) at relatively close spacing (within a circle of radius 4 times the measurement error standard deviation, e.g. 6 meter radius) the performance would be quite good.

Future Work

There are several obvious areas for improvement of the theoretical discussion:

- Better statistical models – the assumption of circular error distributions for each antenna makes sense when averaging over a long time period; however, in an instantaneous sense the sky view presented to the user changes, which means that the HDOP is changing, and that a non-circular pdf is appropriate. One question is how much that variation impacts performance. While we could try to optimize over the sky view, it is possibly better to keep the model insensitive to the sky and bound the loss of performance.
- Extend to three dimensions – for simplicity we assumed that the antennae were located on a horizontal plane; in some applications, adding height to the array could be easily accomplished and its effect on performance should be studied.
- Sequential processing – our approach could be called a snapshot method, only looking at the positions from the receivers at one instant of time. In our experimentation we notice significant geometric similarity of the receivers' positions under H_0 and little geometric similarity under H_1; hence, we expect that sequential tests could be developed to exploit this existence or lack of correlation.
- Simulation/experimentation – our obvious next step is to use a sophisticated GNSS simulator, mimicking both regular GPS and a spoofer.

Disclaimer

The views expressed herein are those of the authors and are not to be construed as official or reflecting the views of the U.S. Coast Guard or any agency of the U.S. Government.

References

[1] J. S. Warner and R. G. Johnston, "GPS spoofing countermeasures," *Homeland Security Jour.*, Dec. 2003.

[2] M. L. Psiaki, B. W. OHanlon, J. A. Bhatti,. D. P. Shepard, and T. E. Humphreys, "Civilian GPS spoofing detection based on dual-receiver correlation of military signals," *Proc. ION GNSS*, Portland, OR, Sept. 2011.

[3] K. D. Wesson, D. P Shepard, J. A. Bhatti, and T. E. Humphreys, "An evaluation of the vestigial signal defense for civil GPS anti-spoofing," *Proc. ION GNSS*, Portland, OR, Sept. 2011.

[4] B. M. Ledvina, W. J. Bencze, B. Galusha, and I. Miller, "An in-line anti-spoofing device for legacy civil GPS receivers," *Proc. ION ITM*, San Diego, CA, Jan. 2010.

[5] S. Daneshmand, A. Jafarnia-Jahromi, A. Broumandon, and G. Lachapelle, "A low-complexity GPS anti-spoofing method using a multi-antenna array," *Proc. ION GNSS 2012*, Nashville, TN, Sept. 2012.

[6] P. F. Swaszek, R. J. Hartnett, M. V. Kempe and G. W. Johnson, "Analysis of a simple, multiple receiver GPS spoof detector," *Proc. ION NTM*, San Diego, CA, Jan. 2013.

[7] N. O. Tippenhauer, C. Pöpper, K. B. Rasmussen and S. Čapkun, "On the requirements for successful GPS spoofing attacks," *Proc. ACM CCS 2011*, Chicago, IL, Oct. 2011.

[8] H. L. Van Trees, **Detection, Estimation, and Modulation Theory, Part I**, New York: Wiley, 1968.

Appendix A

Let's recall some basics facts about manipulating complex random variables. Let x be a complex random variable with mean μ (itself a complex number) and variances σ^2 for both the real and imaginary components. Also, assume that the real and imaginary components are independent. Define a new random variable y by

$$y = a^* x + a x^* = 2\Re\left\{a^* x\right\}$$

where a is a complex constant. Then using $E\{\cdot\}$ as the expectation operator, the mean of y is

$$\begin{aligned} \mu_y &= E\left\{a^* x + a x^*\right\} = a^* E\left\{x\right\} + a E\left\{x^*\right\} \\ &= a^*\mu + a\mu^* = 2\Re\left\{a^*\mu\right\} \end{aligned}$$

and its variance is

$$\begin{aligned} \sigma_y^2 &= E\left\{\left(y-\mu_y\right)\left(y-\mu_y\right)^*\right\} \\ &= 2aa^* E\left\{(x-\mu)(x-\mu)^*\right\} \\ &\quad + (a^*)^2 E\left\{(x-\mu)(x-\mu)\right\} \\ &\quad + a^2 E\left\{(x-\mu)^*((x-\mu))^*\right\} \end{aligned}$$

Now for a *proper* complex random variable x (equivalently, statistically independent real and imaginary components), the first expectation in this last expression is just the variance of x and the second and third expectations are the pseudocovariances which are identically zero, so

$$\sigma_y^2 = 2aa^*\sigma^2 = 2|a|^2\sigma^2$$

Appendix B

The probability of detection of the GLRT is the probability under H_1 that the test statistic is smaller than the threshold

$$\mathrm{P_d} = \mathrm{Prob}_{\mathrm{H}_1}\left(|y| < \lambda'\right)$$

Since y is complex Gaussian

$$y \sim \mathcal{CN}\left(0, 2\beta^2\sigma^2\right)$$

its real and imaginary parts y_r and y_i are jointly Gaussian, so

$$\mathrm{P_d} = \iint\limits_{\Omega} \frac{1}{2\pi\beta^2\sigma^2} e^{-\frac{1}{2\beta^2\sigma^2}\left(y_r^2+y_i^2\right)} dy_r dy_i$$

in which Ω is the disk about the origin of radius λ'. Changing to polar coordinates of r and ϕ

$$y_r = r\cos\phi \qquad y_i = r\sin\phi \qquad dy_r dy_i = r dr d\phi$$

yields

$$\mathrm{P_d} = \int_0^{2\pi}\int_0^{\lambda'} \frac{r}{2\pi\beta^2\sigma^2} e^{-\frac{r^2}{2\beta^2\sigma^2}} dr d\phi$$

in which we have explicitly described the limits of integration of Ω. Integrating

$$\mathrm{P_d} = 1 - e^{-\frac{\lambda'^2}{2\beta^2\sigma^2}}$$

The false alarm of the test is the probability under H_0 that the test statistic is smaller than the threshold

$$\mathrm{P_{fa}} = \mathrm{Prob}_{\mathrm{H}_0}\left(|y| < \lambda'\right)$$

Again, y is complex Gaussian

$$y \sim \mathcal{CN}\left(e^{j\theta}\beta^2, 2\beta^2\sigma^2\right)$$

although this time with non-zero means for y_r and y_i. Writing the integral

$$\mathrm{P_{fa}} = \iint\limits_{\Omega} \frac{1}{2\pi\beta^2\sigma^2} e^{-\frac{\left(y_r-\beta^2\cos\theta\right)^2+\left(y_i-\beta^2\sin\theta\right)^2}{2\beta^2\sigma^2}} dy_r dy_i$$

Changing to polar coordinates yields

$$\begin{aligned}
\mathrm{P_{fa}} &= \int_0^{2\pi}\int_0^{\lambda'} \frac{r}{2\pi\beta^2\sigma^2} e^{-\frac{\left(r\cos\phi-\beta^2\cos\theta\right)^2}{2\beta^2\sigma^2}} \\
&\qquad \cdot e^{-\frac{\left(r\sin\phi-\beta^2\sin\theta\right)^2}{2\beta^2\sigma^2}} dr d\phi \\
&= \int_0^{2\pi}\int_0^{\lambda'} \frac{r}{2\pi\beta^2\sigma^2} e^{-\frac{r^2+\beta^4}{2\beta^2\sigma^2}} \\
&\qquad \cdot e^{\frac{r(\cos\phi\cos\theta+\sin\phi\sin\theta)}{\sigma^2}} dr d\phi \\
&= \int_0^{\lambda'} \frac{r}{\beta^2\sigma^2} e^{-\frac{r^2+\beta^4}{2\beta^2\sigma^2}} \left[\int_0^{2\pi} \frac{1}{2\pi} e^{\frac{r}{\sigma^2}\cos(\phi-\theta)} d\phi\right] dr
\end{aligned}$$

Now, the inner integral in brackets can be manipulated by changing variables to $s = \phi - \theta$, $d\phi = ds$, using the periodicity of the cosine function to shift the integration limits, and recognizing the definition of the modified Bessel function of the first kind

$$\begin{aligned}
\int_0^{2\pi} \frac{1}{2\pi} e^{\frac{r}{\sigma^2}\cos(\phi-\theta)} d\phi &= \int_{-\theta}^{2\pi-\theta} \frac{1}{2\pi} e^{\frac{r}{\sigma^2}\cos s} ds \\
&= \int_0^{2\pi} \frac{1}{2\pi} e^{\frac{r}{\sigma^2}\cos s} ds \\
&= 2\int_0^{\pi} \frac{1}{2\pi} e^{\frac{r}{\sigma^2}\cos s} ds \\
&= \frac{1}{\pi}\int_0^{\pi} e^{\frac{r}{\sigma^2}\cos s} ds \\
&= I_0\left(\frac{r}{\sigma^2}\right)
\end{aligned}$$

The result for the false alarm probability is then

$$\mathrm{P_{fa}} = \int_0^{\lambda'} \frac{r}{\beta^2\sigma^2} e^{-\frac{r^2+\beta^4}{2\beta^2\sigma^2}} I_0\left(\frac{r}{\sigma^2}\right) dr$$

To simplify this expression, we first change variables to

$$z = \frac{r}{\sigma\beta}$$

so

$$\begin{aligned}
\mathrm{P_{fa}} &= \int_0^{\frac{\lambda'}{\sigma\beta}} z e^{-\frac{1}{2}\left[z^2+\frac{\beta^2}{\sigma^2}\right]} I_0\left(\frac{z\beta}{\sigma}\right) dz \\
&= \int_0^{\frac{\lambda'}{\sigma\beta}} z e^{-\frac{1}{2}\left[z^2+d^2\right]} I_0\left(zd\right) dz
\end{aligned}$$

with $d = \beta/\sigma$. This final form can be written in terms of Marcum's Q function [8, pp.344-346]

$$\mathrm{P_{fa}} = 1 - Q\left(d, \frac{\lambda'}{\sigma\beta}\right)$$

Substituting for d

$$\mathrm{P_{fa}} = 1 - Q\left(\frac{\beta}{\sigma}, \frac{\lambda'}{\sigma\beta}\right)$$