Revisiting Combinatorial Watermarking under SCER Adversarial Models

Jason Anderson, Sherman Lo, Todd Walter

Abstract

Combinatorial Watermarking Signal Authentication can help establish trust in a GNSS signals. In Combinatorial Watermarking, the GNSS provider elects to invert a subset of spreading code chips secretly and then later distribute those perturbations to receivers. The receivers can use statistics of the signal to make determinations of the signal authenticity. Previous work demonstrated how to design a Combinatorial Watermarking scheme and derive the distributions of receiver statistics to ensure small probabilities of missed detection and false alarm under assuming an adversary does not attempt to estimate the watermarked chips and replay. In this work, we extend the analysis of Combinatorial Watermarking to adversaries capable of engaging in Security Code Estimation and Replay ("SCER") attacks. We derive the distributions of our statistics under these models and assemble a collection of statistics needed to defend against SCER-capable adversaries. Provided a bound on the estimation capability of the SCER-capable adversary, one can use this work to design a Combinatorial Watermarking scheme that meets security requirements.

I. INTRODUCTION

GNSS remains vulnerable to spoofing attacks. For civilian users, watermarking the signal could provide a pathway to utilize cryptography for receivers to determine a signal's authenticity [Scott, 2003]. In Watermarking Signal Authentication, the spreading code of the GNSS signal is watermarked cryptographically. Several proposals and studies are underway [Anderson et al., 2017, Hinks et al., 2021, O'Hanlon et al., 2022]. Watermarking Signal Authentication, together with Navigation Message Authentication, could allow receivers to assert authenticity of the entire GNSS signal.

The security of Watermarking Signal Authentication is limited with adversaries capable of estimating and replaying the watermarked spreading code. These attacks are called Security Code Estimation and Replay ("SCER") attacks [Humphreys, 2013, Caparra and Curran, 2018, O'Driscoll et al., 2022]. These attacks are the most sophisticated and complicated attacks against Watermarking Signal Authentication, but they are nevertheless possible with sophisticated technical equipment and know how.

In Combinatorial Watermarking, the GNSS provider elects to pseudorandomly invert a fixed-number combination of chips within each spreading code. Combinatorial watermarking presents several design advantages, including deriving the distributions of receiver-observable statistics in the presence of spoofing. In this work, we examine the security of watermarks under attack by SCER-capable adversaries.

1. Combinatorial Watermarking

In this section, we provide an introduction to Combinatorial Watermarking. We refer to our previous work for additional details, including the mathematical and cryptographic derivations [Anderson et al., 2023b]. For the reader's convenience, Table 1 includes the variable notation definitions, adapted from [Anderson et al., 2024].

With a combinatorial watermark, the provider selects a combination of r chips among the n total chips. The construction from [Anderson et al., 2023b] exploits the properties of cryptographic functions to ensure several necessary security properties. Pertinent to this work, the properties ensure that the chips selection is unbiased and there is no efficient algorithm to predict which chips are inverted or any underlying structure among chips selected.

In [Anderson et al., 2023b], we construct a radio observable and bound the probabilities of missed detection and false alarm. The construction of the watermark and the radio observable lend a way to compute the distribution under spoofing conditions with certain adversarial modelling assumptions. Of the most significant consequence, [Anderson et al., 2023b] assumed that the adversary was not listening to the authentic signal to directly estimate the watermark. Rather, the adversary could only make an exhaustive guess. In this work, we allow the adversary to engage in an SCER attack [Humphreys, 2013].

In the non-SCER case, with our derivation of g to Equation (1) and our judicious selection of K in Equation (2), we arrive at

Table 1: A Table defining the variable notation for Combinatorial Watermarking, adapted from [Anderson et al., 2024].

Variable	Definition
n	The number of chips in a single watermark. For example, with SBAS, $n = 1023$.
r	The number of chips inverted in a single watermark. r can vary to meet specific design concerns, but in [Anderson et al., 2024], we suggest $r = 15$.
s	The number of chips an adversary may elect to invert when attempting to spoof a receiver. s may be any integer from 0 to $n/2$.
$\mathcal{H}(n, r, s)$	The Hypergeometric Distribution. An adversary engaged in a spoofing attack generating false signals with s randomly selected chips inverted will guess $h \sim \mathcal{H}(n, r, s)$ correctly for any one watermark.
R	The spreading code replica. R^w refers to the watermarked replica. R_- refers to the reversed replica, which is convolved with the signal to enact correlation.
H	The number of individual watermarks over R.
F	The sampling rate of the radio measuring the receiver observable used to determine the authenticity of a watermarked signal. For SBAS, this should be greater than 2 MHz.
T	The coherent integration of a single watermark measurement. In this work, we use T = 1ms.
P	The power of the signal within a receiver radio immediately proceeding correlation.
σ^2	The noise power within a receiver radio immediately proceeding correlation.
$\mathcal{N}$	The normal distribution.
S	The signal measured over time T at sampling rate F immediately proceeding correlation.
$Y, \mathcal{Y}$	The receiver observable statistic and its distribution, respectively.
$y = g(h \mid n, r, s)$	The linear function g transforms the support of $h \sim \mathcal{H}$ into the support of the radio observable Y.

Equations (3) and (4) for the distribution $\mathcal{Y}$ under the authentic and spoofing hypotheses.

$$y = g(h, n, r) = \frac{1}{n}(2h - r) \tag{1}$$

$$K = \frac{1}{2}\frac{1}{\sqrt{P}}\frac{1}{||R||_1} \tag{2}$$

$$\mathcal{Y} \mid \text{authentic}, H = \mathcal{N}\left(\frac{r}{n}, \frac{r}{n}\frac{\sigma^2}{P}\frac{1}{FTH}\right) \tag{3}$$

$$\mathcal{Y} \mid \text{spoof}, H = \sum_H g(\mathcal{H}(n, r, s), n, r) + \mathcal{N}\left(0, \frac{r}{n}\frac{\sigma^2}{P}\frac{1}{FT}\right) \tag{4}$$

$$\text{PDF}_{\mathcal{Y}|\text{spoof, }H}(y) = \text{PDF}_{\mathcal{H}(n,r,s)}(g^{-1}(y, n, r))^{*H} * \text{PDF}_{\mathcal{N}\left(0, \frac{r}{n}\frac{\sigma^2}{P}\frac{1}{FTH}\right)}(y) \tag{5}$$

From the distributions from Equation (3) and (4) (the PDF of Equation (4) is Equation (5)), one can select a boundary on Y and the scheme parameters to achieve desired probabilities of missed detection and false alarm.

2. A Potential Scheme

In [Anderson et al., 2024], we use the derivations from [Anderson et al., 2023b] to create a scheme that meets reasonable design requirements. [Anderson et al., 2024] suggests that the provider elect to flip $r = 15$ chips among the $n = 1023$ for L1, and that a receiver observe 6000 individual watermarks over 6 seconds. The 6 second timeline comes from a potential TESLA distribution strategy via WAAS [Anderson et al., 2023a]. The $r = 15$ selection fulfills a 10^{-9} missed detection and false alarm requirement for a worst case 2 MHz receiver operating at a $\frac{\text{C}}{\text{N}_0} = 30$. These conditions are intentionally worst case to accommodate a wide

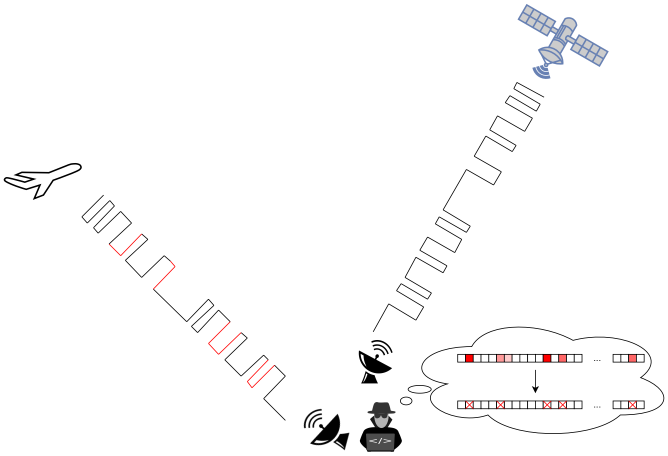

Figure 1: A conceptual diagram of an adapted SCER attack. The adversary attempts to observe the watermark directly in the signal, and then replay a watermarked signal to spoof a receiver. The thought bubble of the adversary portrays the adversary attempting to use its measurements of the true signal to construct a single watermark likely to spoof without detection by the receiver. The top row of boxes represents a collection of inverted-chip likelihoods among a single watermarked spreading code. The varying hues of red represent the soft information provided by likelihood (i.e., the darker the red, the higher the inverted likelihood). The bottom row of boxes represents the watermark decision by the adversary. The adversary can elect not to invert all or invert additional chips.

breadth of receivers and operating conditions. As a case study, we will extend our adversarial model and attack this scheme and predict requirements of the SCER adversary to predictably spoof a receiver.

3. Extending Adversarial Models

In [Anderson et al., 2023b, Anderson et al., 2024], we assumed that the adversary did not listen to the signal to estimate the watermark and replay a signal with the observed watermark. Rather, the adversary made a random guess watermark and transmitted a spoofed signal. In this work, we now examine an adversary attempting to observe the watermark and replay a signal.

In the literature, attacks that listen for security chips and replay are called SCER attacks [Humphreys, 2013]. Figure 1 provides a conceptual diagram of an SCER attack for our combinatorial-watermarking context. Among GNSS spoofing adversaries, SCER-capable adversaries are considerably more sophisticated and complicated, and succeeding is considerably more difficult. In some contexts, schemes that prohibit all but SCER-capable adversaries are sufficient spoofing deterrents. However, once cryptography is incorporated into GNSS signals, GNSS signals will still remain vulnerable to SCER attacks.

SCER attacks are difficult for multiple reasons. One reason is that the GNSS signal is below the thermal noise floor, and estimating security chips requires sophisticated (and likely arduous) radio equipment (e.g., high-gain antennae). A second reason is that the adversary must transport the estimate the security chips to a transmitting antennae within a sufficiently short time as to avoid detection my the receiver's onboard clock. A third reason is that the cryptographic construction limits the effectiveness of advanced decision algorithms beyond exhaustive search among an enormous search space. Like with [Humphreys, 2013], we will assume that our adversary has access to advanced radio equipment, has no delay among its observation antenna, the watermark decision-making algorithm (though a practical computer must be capable of computing the decision), and the replaying antenna.

4. Chip Estimating Model

Under the standard σ^2-noise-power AWGN assumption for the Binary Phase Shift Key ("BPSK") constellation, the constellation points are separated by $2\sqrt{P}$, distributed normally with standard deviation σ, and 0 is halfway between them, as in Figure 2. Without loss of generality, lets suppose that a non-inverted chip center at $\sqrt{P}$ and an inverted chip will center at $-\sqrt{P}$. This is practically achieved by wiping off the spreading code by element-wise multiplying the signal S by the replica R. The chip-estimating adversary must select a decision boundary. Halfway would be a good choice assuming a uniform prior between inverted and non-inverted. However, that will never be the case for a watermarked signal because the number of watermarked chips must be less than the number of non-watermarked chips so that receivers can track the signal. Given a boundary α, the

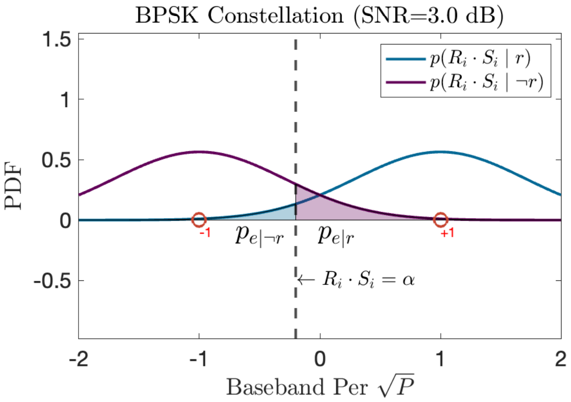

Figure 2: A conceptual figure of the chip estimation model for a single chip after element-wise multiplying my the unwatermarked replica. For the non-watermarked chip hypothesis, the constellation point will be 1. For the watermarked chip hypothesis, the constellation point will be -1. The diagram includes the probability density functions for an example SNR of 3dB. The adversary may elect a decision boundary α (e.g., a Maximum Likelihood or Maximum A Posterior Decision). The probabilities of errors are labeled, given the decision boundary and noise model.

probabilities of error $p_{\mathrm{e}|r}$ and $p_{\mathrm{e}|\neg r}$ given whether the chip is inverted or not inverted, respectively, will be the following.

$$p_{\mathrm{e}|r} = \int_{-\infty}^{\alpha} \mathrm{PDF}_{\mathcal{N}(\sqrt{P},\sigma)}(x)dx \tag{6}$$

$$p_{\mathrm{e}|\neg r} = \int_{\alpha}^{\infty} \mathrm{PDF}_{\mathcal{N}(-\sqrt{P},\sigma)}(x)dx \tag{7}$$

II. THE HARD DECISION ADVERSARY

To spoof, an adversary will have have access to a collection of measurements over the spreading code. Given the cryptographic construction of the watermark, the only structure present in the watermark is that it is composed of exactly r chips. An optimal maximum likelihood detector would evaluate the likelihood of its measurement among all the $\binom{n}{r}$ hypotheses. In this imagined detector, the number of hypotheses is too enormous for any practical detector, following the standard cryptographic security approach of limiting attacks to brute force on an enormous search space. This section discusses an initial decision simplification similar to an error correction code decoder that ignores soft information (i.e., a hard decision decoder).

For the hard decision adversary, lets suppose that the adversary uses a hard decision on each chip, ignoring potentially useful likelihood information from the measurements. The adversary will consider whether each chip is inverted independently without knowledge of the r structure of the watermark (except with its election of α, which can account for r). The adversary will make a hard decision without regard to other measurements of the spreading code. For a particular chip, suppose that the probability of error are $p_{\mathrm{e}|r}$ and $p_{\mathrm{e}|\neg r}$ given whether the chip is inverted or not inverted, respectively. Whichever chips it observes are inverted, it will invert in its spoof signal, even if the adversary flips more or less than the actual known number (e.g., $r = 15$).

For a moment, lets suppose the receiver samples once per chip, the adversary outputs unity power, and there is no noise. The adversary measures a chip i over a watermark spreading code, and makes its decision. From the n decisions, the adversary forms a replica $S^{\text{spoof}} \in \{-1, 1\}^n$, and after the TESLA distribution, the receiver forms $R, R^w \in \{-1, 1\}$. We compute the following *valid* convolutions.

$$b_r \sim \mathcal{B}(r, 1 - p_{\mathrm{e}|r}) \tag{8}$$

$$b_{\neg r} \sim \mathcal{B}(n - r, p_{\mathrm{e}|\neg r}) \tag{9}$$

$$R^w_- * S^{\text{spoof}} = n - 2r + 2b_r - 2b_{\neg r} \tag{10}$$

$$R_- * S^{\text{spoof}} = n - 2b_r - 2b_{\neg r} \tag{11}$$

For Equation (10), suppose that $S^{\text{spoof}} = R$, the convolution within (i.e., with R^w would be $n - 2r$, which is the case where the spoofer broadcasts the original spreading code without any attempt to incorporate the watermark. However, according to

$\mathcal{B}(r, 1-p_{\mathrm{e}|r})$, the adversary will measure watermarked chips and broadcast them correctly increasing $R^w * S$. Simultaneously, the adversary will incorrectly invert a non-watermark chips according to $\mathcal{B}(n-r, p_{\mathrm{e}|\neg r})$ and broadcast them incorrectly, decreasing $R^w * S$. Equation (10) follows the addition and subtraction of two binomial distributions because each of the component chips are measured independently. For Equation (11), suppose that $S^{\text{spoof}} = R$, the convolution within would be n. Each watermarked chip the adversary measures correctly will subtract to the convolution and each non-watermarked chip the adversary measures incorrectly will subtract from the convolution. Equation (11) again follows from the independently-measured chips. We can relax the unity power assumption by multiplying each equation by $\sqrt{P}$. And if we assume that the receiver evenly samples at F over spreading code time T, we can adjust each equation by multiplying by $\frac{FT}{n}$.

In [Anderson et al., 2023b], we suggest the filter $R^w - R$ (with an additional constant gain K from Table 1) and derive that the distribution in the non-SCER case of the statistic in spoofing conditions is the hypergeometric distribution. In the SCER case, repeating this argument, we show in Section II.1 that the distribution is a binomial distribution (rather than a hypergeometric distribution). This follows from subtracting Equations (10) and (11), where the $b_{\neg r}$ will cancel out.

The statistic $R^w - R$ is mostly 0, except where the chips are inverted, meaning the statistic only looks at the data where the watermark should be present and ignores the rest of the spreading code. The adversary could set α to sensitively invert more chips in its spoofed signal; therefore, the adversary could sacrifice some of the receivers tracking ability in favor of ensuring the adversary identifies the inverted chips. This motivates the need for a two statistics to detect spoofing.

1. A Filter Set

From Equations (10) and (11) we suggest two filters of the forms of Equations (12) and (13). The first filter is the subtraction filter from previous work, and the second is the sum. This symmetric pair poses several advantages related to analysis by separating the two binomial distributions. That separation makes following mathematics derivations easier and symmetric, provides an intuitive interpretation, and constrains the adversary's election over α.

$$(R^w - R) * S^{\text{spoof}} = -2r + 4b_r \tag{12}$$

$$(R^w + R) * S^{\text{spoof}} = 2(n-r) - 4b_{\neg r} \tag{13}$$

The statistic of Equation (12) measures how well the adversary can predict where the chip inversions exist. The statistic of Equation (13) measures how well the receiver will tracked the spoofed signal. When we judicious pick the following gains for these filters, the distribution of the statistics is simplified and easier to intuitively understand.

$$k_\Delta = \frac{1}{||R^w - R||_1} \frac{1}{\sqrt{P}} = \frac{1}{2r} \frac{n}{FT} \frac{1}{\sqrt{P}} \tag{14}$$

$$k_\Sigma = \frac{1}{||R^w + R||_1} \frac{1}{\sqrt{P}} = \frac{1}{2(n-r)} \frac{n}{FT} \frac{1}{\sqrt{P}} \tag{15}$$

Finally, the final filters are defined with Equations (16) and (17) and diagrammed with Figure 3.

$$Y_\Delta = k_\Delta \cdot R_\Delta = k_\Delta \cdot (R^w - R) \tag{16}$$

$$Y_\Sigma = k_\Sigma \cdot R_\Sigma = k_\Sigma \cdot (R^w + R) \tag{17}$$

To compute the distributions $\mathcal{Y}_\Delta^{\text{spoof}}$ and $\mathcal{Y}_\Sigma^{\text{spoof}}$ for the statistics Y_Δ and Y_Σ under spoofing conditions, respectively, we could take a direct computational approach via convolution, like in [Anderson et al., 2023b]. However, the adversary model is a group of adversaries with varying observation capability (via $p_{\mathrm{e}|r}$ and $p_{\mathrm{e}|\neg r}$). Therefore, for this work, we think that examining trends in the statistic expectation over varying adversary error is more useful.

We note that from several authentication designs, the receiver will be looking at multiple watermarks [Air Force Research Laboratory, 2019, Anderson et al., 2024]. For this work, we adopt the 6-second watermark observation window from [Anderson et al., 2024]; therefore, the receiver will look at the average of 6000 individual Y_Δ and Y_Σ, handily allowing us to apply the Central Limit Theorem to our results below.

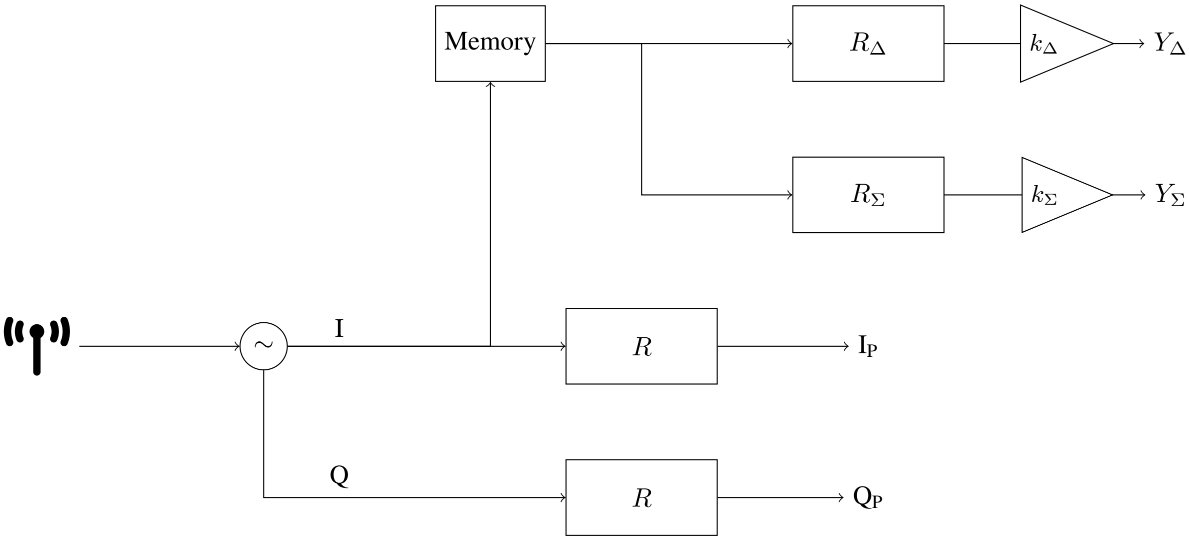

Figure 3: Diagram of a radio that checks the watermark for authentication. The bottom includes the standard tracking loop. From a converged and tracking tracking loop, I base band samples are stored in memory to await the cryptographic seed that determines the watermark. After the watermark seed distribution, the I base band samples are processed through the R_Δ and R_Σ filters. The diagram uses I base band samples assuming the C/A signal; other samples would be required depending on the signal design.

2. Deriving Mean and Variance of the Filter Set

From Section II.1, we now add AWGN to each of the statistics. Over a single spreading code with a total number of samples FT, the noise added to a single Y_Δ and Y_Σ is the following.

$$N \sim \mathcal{N}(0, \sigma^2) \tag{18}$$

We now derive the expectations and variances of the two output filter statistics under authentic and spoofing conditions. The signal is $S = \sqrt{P}R_i + N_i$, where R is R_w in the authentic case, and R is the SCER-estimated R^{spoof} with the chip estimation errors are $p_{\mathrm{e}|r}$ and $p_{\mathrm{e}|\neg r}$. To compute the authentic distributions, one can ignore the binomial distributions entirely and repeat the below. But to avoid unnecessary math, we note that we can use our derivations below and simply set $p_{\mathrm{e}|r} = p_{\mathrm{e}|\neg r} = 0$ for the authentic case. First, we derive the expectation for $\mathcal{Y}_\Delta$ with Equations (19) and (20).

$$\begin{aligned}
\mathbb{E}\left[\mathcal{Y}_\Delta^{\text{spoof}}\right] &= \mathbb{E}\left[\sqrt{P}\cdot k_\Delta \cdot (R_-^w - R_-) * R^{\text{spoof}} + k_\Delta \cdot (R_-^w - R_-) * N\right] \\
&= \sqrt{P}\cdot k_\Delta \cdot \mathbb{E}\left[(R_-^w - R_-) * R^{\text{spoof}}\right] \\
&= \sqrt{P}\cdot \frac{1}{2r}\frac{n}{FT}\frac{1}{\sqrt{P}} \cdot \mathbb{E}\left[(R_-^w - R_-) * R^{\text{spoof}}\right] \\
&= \frac{1}{2r}\cdot \mathbb{E}\left[\frac{n}{FT}\cdot (R_-^w - R_-) * R^{\text{spoof}}\right] \\
&= \frac{1}{2r}\cdot \mathbb{E}\left[-2r + 4\mathcal{B}(r, 1 - p_{\mathrm{e}|r})\right] \\
&= \frac{1}{2r}\cdot \left(-2r + 4r(1 - p_{\mathrm{e}|r})\right) \\
&= -1 + 2(1 - p_{\mathrm{e}|r}) \\
\mathbb{E}\left[\mathcal{Y}_\Delta^{\text{spoof}}\right] &= 1 - 2p_{\mathrm{e}|r} && (19) \\
\mathbb{E}\left[\mathcal{Y}_\Delta^{\text{auth}}\right] &= 1 && (20)
\end{aligned}$$

Next, we derive the expectation for $\mathcal{Y}_\Sigma$ with Equations (21) and (22).

$$\begin{aligned}
\mathbb{E}\left[\mathcal{Y}_\Sigma^{\text{spoof}}\right] &= \mathbb{E}\left[\sqrt{P}\cdot k_\Sigma\cdot(R_-^w+R_-)*R^{\text{spoof}}+k_\Sigma\cdot(R_-^w+R_-)*N\right]\\
&= \sqrt{P}\cdot k_\Sigma\cdot\mathbb{E}\left[(R_-^w+R_-)*R^{\text{spoof}}\right]\\
&= \sqrt{P}\cdot\frac{1}{2(n-r)}\frac{n}{FT}\frac{1}{\sqrt{P}}\cdot\mathbb{E}\left[(R_-^w+R_-)*R^{\text{spoof}}\right]\\
&= \frac{1}{2(n-r)}\cdot\mathbb{E}\left[\frac{n}{FT}\cdot(R_-^w+R_-)*R^{\text{spoof}}\right]\\
&= \frac{1}{2(n-r)}\cdot\mathbb{E}\left[2(n-r)-4\mathcal{B}(n-r,p_{\text{e}|\neg r})\right]\\
&= \frac{1}{2(n-r)}\cdot\big(2(n-r)-4(n-r)p_{\text{e}|\neg r}\big)\\
\mathbb{E}\left[\mathcal{Y}_\Sigma^{\text{spoof}}\right] &= 1-2p_{\text{e}|\neg r} \qquad (21)\\
\mathbb{E}\left[\mathcal{Y}_\Sigma^{\text{auth}}\right] &= 1 \qquad (22)
\end{aligned}$$

This formulation poses the convenience that we can directly connect the expectation of the statistics to the efficacy of the adversaries radio equipment. $p_{\text{e}|r}$ and $p_{\text{e}|\neg r}$ are constrained over α. In Appendix A, we derive the level sets over α to connect them conveniently to the adversaries estimation SNR via Equation (23). In Equation (23), the SNR is the SNR of the SCER adversary, which is a function of the adversary's radio equipment.

$$\operatorname{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_\Sigma\right])+\operatorname{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_\Delta\right])=\sqrt{2\text{SNR}_{\text{SCER}}} \tag{23}$$

Next, we derive the corresponding variances.

$$\begin{aligned}
\mathbb{V}\left[\mathcal{Y}_\Sigma^{\text{spoof}}\right] &= \mathbb{V}\left[\sqrt{P}\cdot k_\Sigma\cdot(R_-^w+R_-)*R^{\text{spoof}}+k_\Sigma\cdot(R_-^w+R_-)*N\right]\\
&= \mathbb{V}\left[\sqrt{P}\cdot k_\Sigma\cdot(R_-^w+R_-)*R^{\text{spoof}}\right]+\mathbb{V}\left[k_\Sigma\cdot(R_-^w+R_-)*N\right]\\
&= \mathbb{V}\left[\sqrt{P}\cdot\frac{1}{2(n-r)}\frac{n}{FT}\frac{1}{\sqrt{P}}\cdot(R_-^w+R_-)*R^{\text{spoof}}\right]+k_\Sigma^2\cdot||R^w+R||^2\cdot\mathbb{V}\left[N\right]\\
&= \frac{1}{4(n-r)^2}\mathbb{V}\left[\frac{n}{FT}\cdot(R_-^w+R_-)*R^{\text{spoof}}\right]+\frac{1}{||R^w+R||_1^2}\frac{1}{P}\cdot||R^w+R||^2\cdot\sigma^2\\
&= \frac{1}{4(n-r)^2}\mathbb{V}\left[2(n-r)-4\mathcal{B}(n-r,p_{\text{e}|\neg r})\right]+\frac{||R^w+R||^2}{||R^w+R||_1^2}\cdot\frac{\sigma^2}{P}\\
&= \frac{1}{4(n-r)^2}\mathbb{V}\left[4\mathcal{B}(n-r,p_{\text{e}|\neg r})\right]+\frac{1}{(n-r)}\frac{n}{FT}\frac{\sigma^2}{P}\\
&= \frac{4}{(n-r)^2}(n-r)p_{\text{e}|\neg r}(1-p_{\text{e}|\neg r})+\frac{1}{(n-r)}\frac{n}{FT}\frac{\sigma^2}{P}\\
\mathbb{V}\left[\mathcal{Y}_\Sigma^{\text{spoof}}\right] &= \frac{4}{(n-r)}p_{\text{e}|\neg r}(1-p_{\text{e}|\neg r})+\frac{1}{(n-r)}\frac{n}{FT}\frac{\sigma^2}{P} \qquad (24)\\
\mathbb{V}\left[\mathcal{Y}_\Sigma^{\text{auth}}\right] &= \frac{1}{(n-r)}\frac{n}{FT}\frac{\sigma^2}{P} \qquad (25)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{V}\left[\mathcal{Y}_\Delta^{\text{spoof}}\right] &= \mathbb{V}\left[\sqrt{P}\cdot k_\Delta\cdot(R_-^w - R_-) * R^{\text{spoof}} + k_\Delta\cdot(R_-^w - R_-) * N\right] \\
&= \mathbb{V}\left[\sqrt{P}\cdot k_\Delta\cdot(R_-^w - R_-) * R^{\text{spoof}}\right] + \mathbb{V}\left[k_\Delta\cdot(R_-^w - R_-) * N\right] \\
&= \mathbb{V}\left[\sqrt{P}\cdot\frac{1}{2r}\frac{n}{FT}\frac{1}{\sqrt{P}}\cdot(R_-^w - R_-) * R^{\text{spoof}}\right] + k_\Delta^2\cdot||R^w - R||^2\cdot\mathbb{V}\left[N\right] \\
&= \frac{1}{4r^2}\mathbb{V}\left[\frac{n}{FT}\cdot(R_-^w - R_-) * R^{\text{spoof}}\right] + \frac{1}{||R^w - R||_1^2}\frac{1}{P}\cdot||R^w - R||^2\cdot\sigma^2 \\
&= \frac{1}{4(n-r)^2}\mathbb{V}\left[-2r + 4\mathcal{B}(r, 1 - p_{\text{e}|r})\right] + \frac{||R^w - R||^2}{||R^w - R||_1^2}\cdot\frac{\sigma^2}{P} \\
&= \frac{1}{4r^2}\mathbb{V}\left[4\mathcal{B}(r, 1 - p_{\text{e}|r})\right] + \frac{1}{r}\frac{n}{FT}\frac{\sigma^2}{P} \\
&= \frac{4}{r^2}rp_{\text{e}|\neg r}(1 - p_{\text{e}|r}) + \frac{1}{r}\frac{n}{FT}\frac{\sigma^2}{P}
\end{aligned}
$$

$$
\mathbb{V}\left[\mathcal{Y}_\Delta^{\text{spoof}}\right] = \frac{4}{r}p_{\text{e}|r}(1 - p_{\text{e}|r}) + \frac{1}{r}\frac{n}{FT}\frac{\sigma^2}{P} \tag{26}
$$

$$
\mathbb{V}\left[\mathcal{Y}_\Delta^{\text{auth}}\right] = \frac{1}{r}\frac{n}{FT}\frac{\sigma^2}{P} \tag{27}
$$

To apply to a receiver examining the average of more than one of these statistics, provided the number is large (e.g., 6000 in [Anderson et al., 2024]), we can apply the Central Limit Theorem. Suppose the receiver is averaging among $H = 6000$ statistics, then the expectation value would not change, but the variance would be $1/H$ the variance derived above.

3. Adversarial Spoofing Efficacy

Now that we have derived the mean and variance of the filter distribution under authentic and spoofing conditions (as a function of the adversaries chip estimation error probability) in Section II.2, in this section, we discuss the receiver decision problem. First, we discuss an intuitive example, and then we discuss design implications.

In Figure 4, we conceptually connect the adversary's decision for α to the sum and difference statistics. On the left, we have the security code estimation model with a decision boundary $\alpha = -0.75\sqrt{P}$ selected by the adversary (as an example for intuition). The decision boundary α directly relates to the prior probability that a chip is flipped. However, without the inclusion of R_Σ filter, the adversary could adjust α to better spoof R_Δ without detection. Therefore, we now consider α to be a hyper parameter for which the security scheme must account all α. On the right, we have the 1-sigma confidence interval for 1 single 1ms watermark with at an SNR of 0dB. The green is probability distribution under authentic conditions, and red for spoofing conditions, from the derivations of Section II.2. The dashed line a trajectory defined by Equation (23). As the adversary changes their selection α, the red ellipse traverses the dashed trajectory. As the SCER SNR increases, the spoofing probability distribution trajectory moves closer to intersecting the authentic case, and vice versa.

Figure 4 is meant to provide an intuitive visual on the dynamics of how α and $\mathcal{Y}_\Delta$ and $\mathcal{Y}_\Sigma$ relate. When the receiver applies the Central Limit Theorem over the observation of 1000s of watermarks, the ellipses will shrink substantially (e.g., a factor of $\sqrt{6000}$). Ultimately, as the SCER adversary's $p_{\text{e}|r} \to 0$ and $p_{\text{e}|\neg r} \to 0$ with better radio and computational equipment, the adversary will be able to approach perfectly estimating and replaying the watermark. The combination of the Central Limit Theorem distribution narrowing, and knowledge that a better and better SCER adversary could exist, motivates exclusively designing based on the expectation value (and ignore the spread of the distribution). Therefore, we now provide Figure 5 based on the SNR-level sets computed via Equation (23).

From a mathematical concise point of view, the upper tail distributions and other effects (such as advantages from Section III) would be better accounted for by adjusting the adversaries actual SNR. For instance, rather than computing the false-alarm and missed-detection probabilities from integration of the repeatedly-convolved distributions of a receiver-decided decision boundary on Y_Δ, Y_Σ, one could compute the dB-width of sigma, adjust the adversary's SNR, and continue design with the formulations of this section. To design a scheme, similar to [Anderson et al., 2024], it is now the problem of selecting n and r under an SCER model (i.e., how big of a dish can the SCER-capable adversary use yielding a specific SNR) that yields acceptable missed detection probabilities. However, it is possible to consider the distribution tails (rather than just the expectation) via repeated convolution of the the binomial distributions.

Efficacy of receiver-decisions on the Y_Δ, Y_Σ can be evaluated integrating over the joint decision space of Figure 5. Or with the

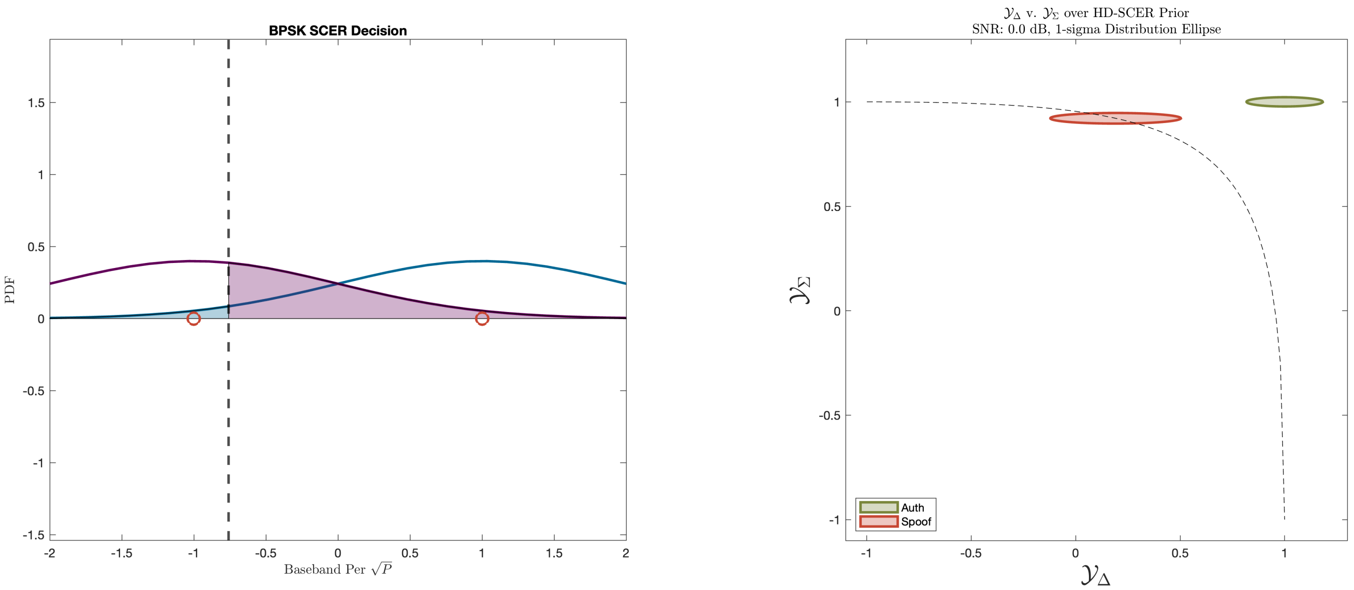

Figure 4: A conceptual figure that relates the adversary's choice of α to the probability distribution of $\mathcal{Y}_\Delta$ and $\mathcal{Y}_\Sigma$, as explained in Section II.3.

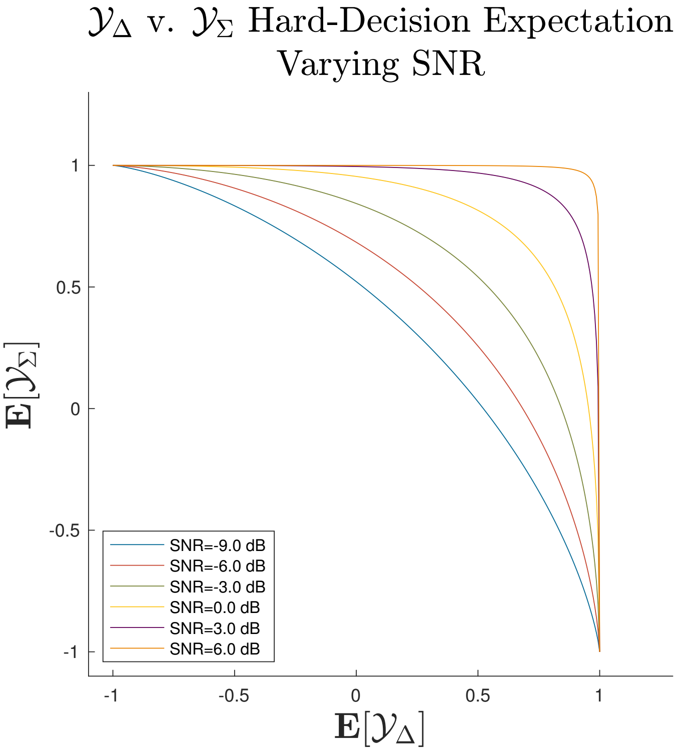

Figure 5: The hard-decision expectation trajectory (along α) for varying levels of SNR.

SNR adjusted by a 3-sigma dB width or with other adjustments from Section III. An interesting consequence is a suggestion to use Equation (23) as the decision boundary to have more favorable probability of missed detection and false alarms compared to a linear decision boundary. Note that as Y_Σ decreases, the receivers ability to track the signal rapidly decreases, informing a reasonable decision area over Y_Δ, Y_Σ.

III. SOFT DECISION SCER

Whereas the previous section considered a hard-decision adversary, in this section we consider a soft-decision adversary that beats the performance of the hard-decision adversary. The hard-decision adversary poses a scheme where the adversary spoofing distributions can be computed for the purpose of design. However, in this section, we find a better adversary, but we are only able to show its advantage via Monte Carlo simulation (without knowledge of a concise pathway to repeat the distribution derivations).

In the hard-decision adversary from Section II.3, the adversary make a hard decision on the security code estimation problem. This ignores potentially useful soft-information, for instance, with the measurement likelihood from the BPSK model. Moreover, the hard-decision adversary employs a constant chip power. We propose the following soft-decision adversary without any claim about whether this adversary is the best obtainable. With our soft-decision adversary, the adversary will set the chip power to be proportional to the hypothesis likelihood ratio.

$$P_i \propto \begin{cases} p(r) & \text{if SCER adversary does not invert chip } i \\ p(\neg r) & \text{if SCER adversary does invert chip } i \end{cases} \tag{28}$$

Our choice for P_i is simply a judicious, first-guess choice, inspired by [O'Driscoll et al., 2022], that serves our intuitive purpose. When the adversary is very confident that a chip is inverted, it will place more power on the particular chip (and the same with a chip highly believed to be not inverted). When the adversary is not confident that a chip is inverted or not inverted, the adversary places less power on that particular chip. For our adversary, we re normalize the signal so that it contains the same aggregate average power over the entire spreading; hence our use of $\propto$ for Equation (28). This accounts for tracking loop automated gain control and establishes a fair comparison in the Y_Δ and Y_Σ space. We tried a couple of other functions that ensure more power on more confidence measurements (e.g., having P_i be a function of the likelihood ratio) with varying advantage.

Like with the order of Section II.3, we will first provide a Monte Carlo experiment for the purpose of intuition, and then a second experiment for design implication. Figure 6 provides the result of Monte Carlo simulation of the soft decision adversary against the hard decision adversary. The dashed line is where the statistic expectation for the hard-decision adversary. The distributions provided are for the aggregation of $H = 6000$ watermarks. The green is the 3σ authentic distribution ellipse, and the red is the 100 spoofing Monte Carlo trials. The soft-decision advantage is demonstrated by the spoofing ellipse being to the right of the hard-decision trajectory line.

In a typical scenario, the adversary and receiver observe different SNRs for the GNSS signal because the adversary is likely using a better antennae. The actual spoofing distribution must account for both variances. In Figure 7 does not distributions of $\mathcal{Y}$; rather, they are distributions of $\mathbf{E}[\mathcal{Y}]$ under the central limit theorem. Figure 7 shows the trend of the expectations when the adversary elects different α, which affects whether r or $\neg r$ election applies. The trajectories follow the general trend, except that the soft-decision is slightly beating the hard-decision.

1. Better SCER Adversaries and Design Implications

At the time of this work, we do not yet observe a pathway to mathematically derive the advantage for the soft-decision adversary, to find the best soft-decision adversary, and bound the advantage of any soft decision adversary. Given the convenience of the mathematically concise derivations for the hard-decision adversary, and the conventions of error-correction code, it is likely appropriate to attempt to find a soft-information advantage bound or correction for use in designing a system with the hard-decision derivations. For instance, suppose one could show that a soft-decision adversary performs no better than a hard-decision adversary with x more SNR dB. Then one could design using the hard decision formulae with a simple correction.

Because an adversary could continually achieve a better radio for the security code estimation, the GNSS designer should focus on ensuring that the system design requires an antennae that is reasonable arduous on the spoofer and easy for someone in the area to detect. For instance, one could design the system to require a large dish antennae that should be likely visible in a protection area (e.g., in the vicinity of an airport). Noting that the $r = 15$ design from [Anderson et al., 2024], was created before this work, we can derive the gain required to spoof a receiver. In [Anderson et al., 2024], we suggested a decision boundary of $\mathcal{Y}_\Delta^{\text{spoof}} > 0.5$ for 10^{-9} missed-detection and false-alarm rates for non-SCER adversaries. To spoof a receiver *on*

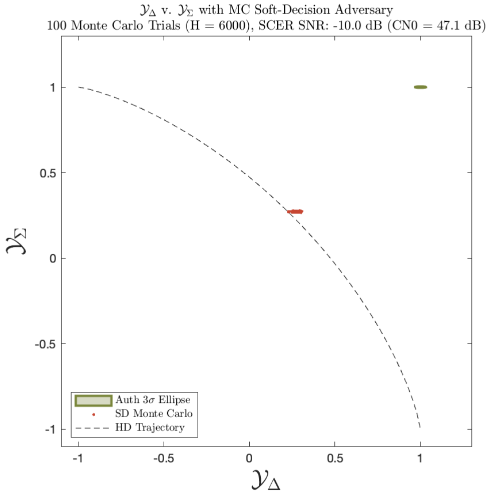

Figure 6: Monte Carlo Experiment showing the advantage of our Soft-decision Adversary. The adversary is this figure chose $\alpha = 0$. For the hard-decision Adversary, the expectation will be along the Equation (23) trajectory (dashed). Monte Carlo simulation demonstrates a small advantage by having the power of each chip be a function of the confidence of the chip estimation.

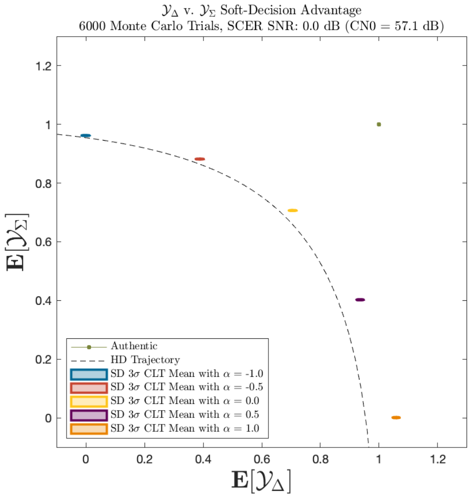

Figure 7: A diagram generated with Monte Carlo simulation that shows the trend of the expectations of $\mathcal{Y}_\Delta$ and $\mathcal{Y}_\Sigma$ under spoofing conditions with the soft-decision adversary of Section III with varying α. The soft-decision ellipses are the 3σ Central Limit Theorem confidence ellipses of where the expectation should be. For differing SCER SNRs (only 0dB depicted), the soft-decision adversary poses a small advantage over the hard-decision trajectory line. In the case of this figure, an SCER adversary would need to have an antennae of about 10 dB gain to achieve this performance. Note that the receiver will lose the ability to track the signal when $\mathcal{Y}_\Sigma$ decreases.

expectation, the adversary would need an antennae array or a high-gain antennae until the spoofing ellipses from Figure 7 cross past the receiver's decision boundary (e.g., $\mathcal{Y}_{\Delta}^{\text{spoof}}, \mathcal{Y}_{\Sigma}^{\text{spoof}} > 0.5$).

Deriving a rigorous answer to the advantage of a soft-decision adversary poses a difficult challenge for both deriving an answer and defining a model. For instance, in the model of this work, and adversary could put an enormous power on a single chip (and zero out the other chips). Among the entire spreading code measurements, suppose the adversary only placed power on two chips: the one with the highest measured likelihood of being inverted and the one with the highest measured likelihood of not being inverted. It is very likely that these two measurements (e.g., among the r and $n - r$) are correct. With a perfectly tracking receiver, the adversary could spoof Y_{Δ} and Y_{Σ} by placing max power on those two chips. However, this represents a degenerate case, motivating a more sophisticated receiver and spoofing radio models (e.g., where the power of this chips are saturated in the 2-chip power spoof). As the model becomes more complicated and realistic, it is unlikely there exists a mathematical concise answer, relegating our best answer to Monte Carlo methods and direct experimentation.

IV. CONCLUSION

In this work, we extend Combinatorial Watermarking analysis to SCER-capable adversaries. We provide a set of receiver statistics that can be used to detect SCER attacks, provided a limitation on how well an adversary can estimate watermarked chips and a hard-decision watermark detection strategy. We derive the distributions of the receiver statistics in the presence of an hard-decision SCER spoofing attack and provide a pathway to design a Combinatorial Watermarking scheme to meet security requirements in the presence of an SCER-capable adversary. We propose a soft-decision SCER spoofing attack with an advantage over the hard-decision SCER spoofing attack with Monte Carlo simulation. From this work, a GNSS designer can approximately predict how well a Combinatorial Watermark is resistant to an SCER adversary.

REFERENCES

[Air Force Research Laboratory, 2019] Air Force Research Laboratory (2019). IS-AGT-100: Chips Message Robust Authentication (Chimera) Enhancement for the L1C Signal: Space Segment/User Segment Interface. Technical report, Air Force Research Laboratory, Space Vehicles Director.

[Anderson et al., 2023a] Anderson, J., Lo, S., Neish, A., and Walter, T. (2023a). Authentication of Satellite-Based Augmentation Systems with Over-the-Air Rekeying Schemes. *NAVIGATION: Journal of the Institute of Navigation*, 70(3).

[Anderson et al., 2023b] Anderson, J., Lo, S., and Walter, T. (2023b). Authentication Security of Combinatorial Watermarking for GNSS Signal Authentication. In *Proceedings of the 36th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2023)*, pages 495–509.

[Anderson et al., 2024] Anderson, J., Lo, S., and Walter, T. (2024). Combinatorial Watermarking for GNSS Signal Authentication. In *Proceedings of the 2024 International Technical Meeting of The Institute of Navigation*.

[Anderson et al., 2017] Anderson, J. M., Carroll, K. L., DeVilbiss, N. P., Gillis, J. T., Hinks, J. C., O'Hanlon, B. W., Rushanan, J. J., Scott, L., and Yazdi, R. A. (2017). Chips-message robust authentication (chimera) for GPS civilian signals. pages 2388 – 2416.

[Caparra and Curran, 2018] Caparra, G. and Curran, J. T. (2018). On the achievable equivalent security of GNSS ranging code encryption. In *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pages 956–966. IEEE.

[Hinks et al., 2021] Hinks, J., Gillis, J. T., Loveridge, P., Miller, S., Myer, G., Rushanan, J. J., and Stoyanov, S. (2021). Signal and Data Authentication Experiments on NTS-3. pages 3621–3641.

[Humphreys, 2013] Humphreys, T. (2013). Detection Strategy for Cryptographic GNSS Anti-Spoofing. *Aerospace and Electronic Systems, IEEE Transactions on*, 49:1073–1090.

[O'Hanlon et al., 2022] O'Hanlon, B., Rushanan, J. J., Hegarty, C., Anderson, J., Walter, T., and Lo, S. (2022). SBAS Signal Authentication. In *Proceedings of the 35th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2022)*, pages 3369–3377.

[O'Driscoll et al., 2022] O'Driscoll, C., Scuccato, T., DallaChiara, A., Pany, T., Diez, M., and Hameed, M. (2022). The Attack Agnostic Defence: a spoofing detection metric for secure spreading sequences.

[Scott, 2003] Scott, L. (2003). Anti-Spoofing & Authenticated Signal Architectures for Civil Navigation Systems. In *Proceedings of the 16th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS/GNSS 2003)*, pages 1543 – 1552.

A. HARD DECISION TRAJECTORY EQUATION

First, we substitute the probability of errors with their functions of α from Equations (6) (7) and isolate α for both statistics.

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{Y}_{\Delta}^{\text{spoof}} \mid \alpha\right] &= 1 - 2p_{\text{e}|r,\alpha} \\
&= 1 - 2 \cdot \int_{\alpha}^{\infty} \text{PDF}_{\mathcal{N}(-\sqrt{P},\sigma^2)}(y)dy \\
&= 1 - 2 \cdot \Big(1 - \text{CDF}_{\mathcal{N}(-\sqrt{P},\sigma^2)}(\alpha)\Big) \\
&= -1 + 2 \cdot \text{CDF}_{\mathcal{N}(-\sqrt{P},\sigma^2)}(\alpha) \\
&= -1 + 2 \cdot \left(\frac{1}{2}\left(1 + \text{erf}\left(\frac{\alpha+\sqrt{P}}{\sigma\sqrt{2}}\right)\right)\right) \\
&= \text{erf}\left(\frac{\alpha+\sqrt{P}}{\sigma\sqrt{2}}\right) \\
\alpha &= -\sqrt{P} + \sqrt{2}\sigma \cdot \text{erf}^{-1}(\mathbb{E}\left[Y_{\Delta} \mid \alpha\right])
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{Y}_{\Sigma}^{\text{spoof}} \mid \alpha\right] &= 1 - 2p_{\text{e}|\neg r,\alpha} \\
&= 1 - 2 \cdot \int_{\infty}^{\alpha} \text{PDF}_{\mathcal{N}(\sqrt{P},\sigma^2)}(y)dy \\
&= 1 - 2 \cdot \text{CDF}_{\mathcal{N}(\sqrt{P},\sigma^2)}(\alpha) \\
&= 1 - 2 \cdot \left(\frac{1}{2}\left(1 + \text{erf}\left(\frac{\alpha-\sqrt{P}}{\sigma\sqrt{2}}\right)\right)\right) \\
&= -\,\text{erf}\left(\frac{\alpha-\sqrt{P}}{\sigma\sqrt{2}}\right) \\
&= \sqrt{P} + \sqrt{2}\sigma \cdot \text{erf}^{-1}(-\mathbb{E}\left[Y_{\Sigma} \mid \alpha\right]) \\
\alpha &= \sqrt{P} - \sqrt{2}\sigma \cdot \text{erf}^{-1}(\mathbb{E}\left[Y_{\Sigma} \mid \alpha\right])
\end{aligned}
$$

Then, we set the α equal to each other.

$$
\begin{aligned}
\sqrt{P} - \sqrt{2}\sigma \cdot \text{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_{\Sigma} \mid \alpha\right]) &= -\sqrt{P} + \sqrt{2}\sigma \cdot \text{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_{\Delta} \mid \alpha\right]) \\
\sqrt{2}\sigma \cdot \text{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_{\Sigma} \mid \alpha\right]) + \sqrt{2}\sigma \cdot \text{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_{\Delta} \mid \alpha\right]) &= 2\sqrt{P} \\
\text{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_{\Sigma} \mid \alpha\right]) + \text{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_{\Delta} \mid \alpha\right]) &= \sqrt{2P/\sigma^2} \\
\text{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_{\Sigma} \mid \alpha\right]) + \text{erf}^{-1}(\mathbb{E}\left[\mathcal{Y}_{\Delta} \mid \alpha\right]) &= \sqrt{2\text{SNR}_{\text{SCER}}}
\end{aligned}
$$

Note that the SNR here is the SNR of the SCER adversary, which is a function of the adversary's radio equipment.