GENERALIZED IMPROVED SPREAD SPECTRUM WATERMARKING ROBUST AGAINST TRANSLATION ATTACKS

Neslihan Gerek *M. Kıvanç Mıhçak*

Electrical - Electronics Engineering Dep.
Boğaziçi University İstanbul, 34342, Turkey
Email:*neslihan.gerek,kivanc.mihcak@boun.edu.tr*

ABSTRACT

In this work, we consider the ISS (improved spread spectrum) watermarking [1] framework, and propose a generalized version of it, termed "Generalized Improved Spread Spectrum" (GISS), where we achieve both host-interference cancelation and robustness to "translation" attacks up to some tolerance. In particular, we reduce the correlation between the watermark and the host, not only at the embedding location, but also within an a-priori-defined neighborhood around it. We show that the resulting framework leads to a constrained quadratic optimization problem, where the cost function and the constraint represent the amount of host interference on the watermark and the norm of the resulting "host interference cancelation sequence" (HICS), respectively. We provide a closed-form analytical solution to this optimization problem and experimentally demonstrate its effectiveness for 1D signals.

Index Terms – **security, spread spectrum communication, signal processing, decoding, optimization methods**

1. INTRODUCTION

In this paper, we concentrate on spread spectrum (SS) watermarking [2], which is a practically popular approach, due to its performance in the low watermark-to-noise ratio regime in the presence of a wide-class of geometry-preserving attacks. However, one major well-known concern for SS approach is the potential correlation between the unmarked host and the watermark which leads to host interference resulting in performance degradation[1]. In [1], Malvar et. al. introduce the ISS (improved spread spectrum) approach to overcome this problem, where the main idea is to achieve host interference cancelation via modulating the watermark amplitude to attenuate the norm of the projection of the host to the subspace spanned by the watermark, thereby effectively decorrelating the pseudo-randomly-generated watermark and the host. The idea of "host interference cancelation" forms the backbone of our approach, as well.

In particular, we propose a generalized version of the ISS algorithm, termed "GISS (generalized improved spread spectrum)", where we concentrate on (i) host interference cancelation, and (ii) robustness to translation, jointly. In a nutshell, the idea is to modify the host at the encoder side, such that the norm of the projection of the host to the subspace, spanned by several shifted versions of the watermark, is decreased as much as possible, subject to a constraint on the amount of modification of the host (this constraint aims to preserve the perceptual quality of the modified host). In this way, host interference within a neighborhood of interest (specified by the tolerance limit against translation attacks, which introduces a performance vs. perceptual quality tradeoff) is obtained. This neighborhood is termed as the "tolerance region", and its center is the watermark-embedding location. Hence, unlike ISS, approximate decorrelation between the host and the watermark is attained not only at the embedding location, but also in the greater "tolerance region" around it. At the receiver side, the correlation detector is applied at all locations within the tolerance region; the final decision on the decoded bit is based on the maximum absolute correlation value. Consequently, robustness against local translations is obtained.

In Sec. 2, we briefly introduce the necessary background regarding SS and ISS methods. In Sec. 3, we introduce the resulting optimization problem of GISS method and provide the analytical solution. In Sec. 4, we present a practical digital watermarking algorithm, constructed based on the GISS approach. In Sec. 5, we show the experimental results of the proposed approach and discuss them briefly. In Sec. 6, we summarize the contribution of the paper.

Notation: Bold upper- and lower-case letters represent matrices and vectors, respectively. Corresponding regular letters with subscripts represent individual elements. For example, $\mathbf{a} \in \mathbb{R}^n$ is a vector and $a_i \in \mathbb{R}$ is its i^{th} element; given the matrix $\mathbf{A}$, A_{ij} is its $(i,j)^{th}$ element; $\langle\, .\, ,\, .\, \rangle$ represents the inner product which induces the Euclidean (L_2) norm; $(.)^T$ represents the transpose operation; $\mathbf{I}_k$ denotes the identity matrix of size $k \times k$.

2. BACKGROUND

2.1. Basic Spread Spectrum Watermarking Method

In its most basic form, we assume that one bit of information is embedded in a vector of N coefficients, achieving a bit rate of $1/N$ bits/sample. In this case, embedding is

[1] Although the watermark may be generated statistically independent of the host, this does not guarantee the absence of the undesired correlation between the watermark and the host in the deterministic sense.

performed by:

$$\mathbf{s} = \mathbf{x} + b\mathbf{u},$$

where, $\mathbf{s}$, $\mathbf{x}$ and $\mathbf{u} \in \mathbb{R}^n$ are the watermarked, host and watermark signals, respectively; $b \in \{\pm 1\}$ represents the embedded bit. If the attack channel can be modeled as independent additive noise, then the received signal at the decoder is given by

$$\mathbf{y} = \mathbf{s} + \mathbf{n} = \mathbf{x} + b\mathbf{u} + \mathbf{n}.$$

Decoding is performed by checking the sign of the normalized statistics produced by the correlation detector:

$$\gamma \triangleq \frac{\langle \mathbf{y}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} = \frac{\langle b\mathbf{u} + \mathbf{x} + \mathbf{n}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2}; \quad \hat{b} = \text{sign}(\gamma), \qquad (1)$$

where γ and $\hat{b}$ denote the detection statistics and the decoded bit, respectively. See [2] for further details.

2.2. Improved Spread Spectrum (ISS) Watermarking Method

The main idea in this method is to reduce the correlation between the host signal and the watermark by modulating the energy of the watermark at the embedding process using the projection of the host signal on the watermark. The resulting embedding method, ISS [1] is a slightly modified version of the conventional SS embedding method:

$$\mathbf{s} = \mathbf{x} + \lambda(x, b)\mathbf{u}.$$

The value x represents the projection coefficient of the host $\mathbf{x}$ on the watermark $\mathbf{u}$: $x \triangleq \langle \mathbf{x}, \mathbf{u} \rangle / \|\mathbf{u}\|^2$. Amplitude of the embedded watermark is controlled by the function $\lambda(x, b)$, which can be defined in various ways, including a linear mapping:

$$\mathbf{s} = \mathbf{x} + (\alpha b - \lambda_{ISS} x)\,\mathbf{u};$$

α controls the distortion level and λ_{ISS} controls the removal of the host interference on the detection statistics. We refer the interested reader to [1] for further details.

3. PROBLEM DEFINITION AND ANALYTICAL SOLUTION

ISS reduces the error probability in the presence of most geometry-preserving attacks, compared to the conventional SS. However, robustness against geometric attacks is not improved in ISS. We propose "Generalized Improved Spread Spectrum (GISS)" to address the "translation type" geometric attacks.

The original location of the center of watermark (also termed as the "embedding location") in the received signal is known by the decoder if the key is known in the absence of geometric attacks. However, if the signal is exposed to a (even mild) geometric attack, then the center of the embedded watermark is shifted. Therefore, we expect that, after a potentially-applied geometric attack, the watermark's location may have been shifted within a reasonably-large area, termed as the "tolerance region". Tolerance region may potentially be larger than the spatial support of the watermark itself; the encoder first performs host interference cancelation (with a particular tolerance region in mind), subsequently followed by watermark embedding. Thus, in the GISS framework, at every point within the "tolerance region", the interference between the watermark and the host is decreased as much as possible at the encoder. At the decoder, the center of the embedded watermark is assumed to be the point where the correlation between the received signal, $\mathbf{y}$, and the watermark $\mathbf{u}$ is the highest.[2]

Let $\mathbf{c}$ denote the host interference cancelation sequence (HICS). At the encoder side, our aim is to design $\mathbf{c}$ so as to minimize both its norm, $\|\mathbf{c}\|$, and the correlation between the watermark, $\mathbf{u}$, and the "HICS-embedded host signal", $\mathbf{x} + \mathbf{c}$ everywhere within the tolerance region, jointly. The resulting optimization problem can be formulated as follows:

$$\min_{\mathbf{c} \in \mathbb{R}^n} \|\mathbf{H}(\mathbf{x} + \mathbf{c})\|^2, \qquad (2)$$

$$\text{s.t.} \quad \|\mathbf{c}\|^2 \leq \text{A}, \qquad (3)$$

where, the "linear correlation transform matrix", $\mathbf{H}$ of size $m \times n$, is a function of the watermark, $\mathbf{u}$, such that rows of $\mathbf{H}$ consist of cyclic-shifted versions of $\mathbf{u}$, denoted by $\{\mathbf{u}^i\}$, i.e., $(\mathbf{H} \cdot \mathbf{x})_i = \langle \mathbf{u}^i, \mathbf{x} \rangle$, where $\mathbf{u}^i_j = \mathbf{u}_{\left(j - \frac{m}{2} + i\right) \, mod \, n + 1}$; $1 \leq i \leq m$, $1 \leq j \leq n$. Note that, the number of rows of $\mathbf{H}$ (i.e., m) is equal to the size of the tolerance region, which is smaller than the size of the host signal, n. We also assume that the support of the watermark is sufficiently small to ensure that while sliding the watermark in the tolerance region, it will remain inside the support of the host signal. Moreover, we assume that $\mathbf{H}$ is full-rank, i.e., rank $(\mathbf{H}) = m$.

Parameter A controls the distortion, induced by adding $\mathbf{c}$ to $\mathbf{x}$. The Lagrangian corresponding to (2,3) is given by

$$\begin{aligned} \mathfrak{L} &= \|\mathbf{H}(\mathbf{x} + \mathbf{c})\|^2 + \lambda \|\mathbf{c}\|^2, \\ &= \left(\mathbf{x}^T + \mathbf{c}^T\right) \mathbf{Q} (\mathbf{x} + \mathbf{c}) + \lambda \mathbf{c}^T \mathbf{c} \end{aligned} \qquad (4)$$

where λ is the Lagrange multiplier and $\mathbf{Q} \triangleq \mathbf{H}^T\mathbf{H}$, which is of size $n \times n$, rank m. Note that, since $\mathfrak{L}$ is quadratic (and hence convex) in $\mathbf{c}$, the necessary and sufficient condition for optimality is given by

$$\left[\nabla_c \mathfrak{L}|_{\mathbf{c}=\mathbf{c}_{opt}} = 0\right] \iff \left[\mathbf{c}_{opt} = -\mathbf{R}^{-1}\mathbf{Q}\mathbf{x}\right], \qquad (5)$$

where $\mathbf{R} \triangleq \mathbf{Q} + \lambda_{opt}\mathbf{I}_n$ (of size $n \times n$, rank n for $\lambda_{opt} \neq 0$) and λ_{opt} is the value of the Lagrange multiplier λ at optimality.

Practically, it may be a computationally-challenging task to compute $\mathbf{R}^{-1}$ for large n which can be simplified via using the "Singular Value Decomposition (SVD)" of the matrix $\mathbf{H}$. Let the SVD of $\mathbf{H}$ be denoted by

$$\mathbf{H} = \mathbf{U}_{m \times m} \Sigma_{m \times m} (\mathbf{V}_{n \times m})^T,$$

where Σ is the singular value matrix; $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular vector matrices, respectively. Using $\sigma_i \triangleq \Sigma_{ii}$, and (5), we get

$$\mathbf{c}_{opt} = -\mathbf{V}\widetilde{\mathbf{S}}(\mathbf{V})^T\mathbf{x}, \qquad (6)$$

where $\widetilde{\mathbf{S}}$ is a diagonal matrix of size $m \times m$, such that

$$\widetilde{S}_{ii} \triangleq \frac{\sigma_i^2}{\lambda_{opt} + \sigma_i^2}, \quad 1 \leq i \leq m.$$

[2] In the developments below, we assume that we work with $1D$ signals; however, the approach can be extended to $2D$ signals with little or no difficulty.

At optimality, the constraint (3) is active, i.e., λ_{opt} is non-zero and chosen such that (3) is satisfied with equality; λ_{opt} determines the tradeoff between the power of the HICS, $\mathbf{c}$, (represented by the constraint (3)) and the norm of the correlation vector between the HICS-embedded signal, $\mathbf{x}+\mathbf{c}$, and the watermark, $\mathbf{u}$ (represented by the cost (2)). Increasing λ_{opt} jointly reduces the effect of the cost function in the Lagrangian, and increases the effect of the constraint function. Thus, it can be shown that λ_{opt} is non-negative and monotonic decreasing in A. Asymptotically, we have

$$\lim_{\lambda_{opt}\mapsto 0}\mathbf{c}_{opt}=-\mathbf{V}\mathbf{V}^T\mathbf{x} \quad \text{and} \quad \lim_{\lambda_{opt}\mapsto\infty}\mathbf{c}_{opt}=0.$$

Note that, when $\lambda=0$, complete host interference rejection is possible.

Using $\mathbf{x}'\triangleq\mathbf{V}^T\mathbf{x}$ ($\mathbf{x},\mathbf{x}'\in\mathbb{R}^m$), we can define two different quantities that are important in interpreting the solution to this problem. The first one is the norm of HICS at optimality (denoted by $C_{1,opt}$), which amounts to the magnitude of the distortion added to signal by HICS. It is given as follows:

$$\begin{aligned} C_{1,opt} &\triangleq \mathbf{c}_{opt}^T\mathbf{c}_{opt}=\mathbf{x}^T\mathbf{V}\tilde{\mathbf{S}}\mathbf{V}^T\mathbf{V}\tilde{\mathbf{S}}\mathbf{V}^T\mathbf{x}, \\ &= (\mathbf{x}^T\mathbf{V})\tilde{\mathbf{S}}^2(\mathbf{V}^T\mathbf{x}), \\ &= \sum_{i=1}^{m}\left(\frac{\sigma_i^2}{\lambda_{opt}+\sigma_i^2}\cdot x_i'\right)^2. \end{aligned}$$

Asymptotically, the behavior of $C_{1,opt}$ is given by,

$$\lim_{\lambda\mapsto 0}C_{1,opt}=\sum_{i=1}^{m}\left(x_i'\right)^2=\left\|\mathbf{V}^T\mathbf{x}\right\|^2, \quad \lim_{\lambda\mapsto\infty}C_{1,opt}=0. \quad (7)$$

The second important quantity is the norm of correlation between watermark and the decorrelated (i.e., HICS-embedded) signal (denoted by $C_{2,opt}$), which can be interpreted as the value that we want to minimize in the cost function. It is given as follows:

$$\begin{aligned} C_{2,opt} &\triangleq \|\mathbf{H}\left(\mathbf{x}+\mathbf{c}_{opt}\right)\|^2=\left(\mathbf{x}+\mathbf{c}_{opt}\right)^T\mathbf{H}^T\mathbf{H}\left(\mathbf{x}+\mathbf{c}_{opt}\right), \\ &= \mathbf{x}^T\left(\mathbf{I}_n-\mathbf{V}\tilde{\mathbf{S}}\mathbf{V}^T\right)^T\mathbf{Q}\left(\mathbf{I}_n-\mathbf{V}\tilde{\mathbf{S}}\mathbf{V}^T\right)\mathbf{x}, \quad (8) \\ &= \left(\mathbf{x}'\right)^T\left(\mathbf{I}_m-\tilde{\mathbf{S}}\right)\Sigma^2\left(\mathbf{I}_m-\tilde{\mathbf{S}}\right)^T\mathbf{x}', \quad (9) \\ &= \sum_{i=1}^{m}\left[\frac{\lambda_{opt}\sigma_i^2}{\lambda_{opt}+\sigma_i^2}x_i'\right]^2 \quad (10) \end{aligned}$$

where (8) follows from (6) and the definition of $\mathbf{Q}$, (9) follows from the definition of $\mathbf{x}'$, and the rest from straightforward algebra. Asymptotically, the behavior of $C_{2,opt}$ is given by,

$$\begin{aligned} \lim_{\lambda\mapsto 0}C_{2,opt} &= 0, \\ \lim_{\lambda\mapsto\infty}C_{2,opt} &= \sum_{i=1}^{m}\left(\sigma_i x_i'\right)^2=\|\Sigma\mathbf{V}^T\mathbf{x}\|^2=\|\mathbf{H}\mathbf{x}\|^2 \quad (11) \end{aligned}$$

Results in (7) and (11) exhibit the tradeoff between visual distortion added to the signal and the detection ratio, since decreasing the correlation between the watermark and the host signal improves the performance of the decoder.

Remark: "Exhaustive search" has been proposed as a potential countermeasure to geometric attacks in the literature. It has also been pointed out that, even though computational resources may be sufficient, such an approach would yield a significant increase in the probability of false positives [4]. In spirit, the GISS approach aims to overcome this difficulty via nullifying the correlation between the host and the watermark as much as possible prior to embedding, which may potentially lead to the practical usage of exhaustive-search-based approaches to achieve robustness against geometric attacks.

4. PROPOSED WATERMARKING METHOD

In the proposed method, we suggest adding a single-bit watermark (spatially-limited to a predefined region of the signal) to the host in the spatial domain. The approach can be extended via using multiple watermarks to increase the payload.

4.1. Embedding

Embedding consists of three steps. In the first step, a watermark is generated using a secret key. The key is used as the seed of a pseudo-random number generator, which produces a Gaussian-distributed sequence with zero mean and unit norm. The next step is the decorrelation of the "tolerance region" of the host signal with the watermark, via adding the optimal HICS to the host. Using the method explained in Sec. 3, with the given value of parameter A, the optimal HICS, $\mathbf{c}_{opt}$ is found and subsequently added to the determined region of the host signal, producing the decorrelated signal. The last step is the actual watermark embedding. The strength of the watermark is adjusted to match a targeted Signal to Watermark Ratio (SWR) value. If the bit value is 1 (resp. 0), then the watermark is added to (resp. subtracted from) the decorrelated host signal.

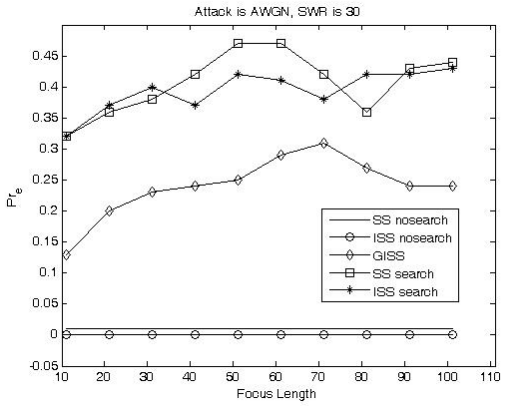
4.2. Decoding

In order to decode the value of the embedded bit in the received signal, watermark is generated at the receiver with the secret key using the same procedure as in the embedding process. The tolerance region is filtered by the watermark and a "correlation map" is generated. If the absolute maximum point of this map is positive (resp. negative), then the decoded bit value is assigned to 1 (resp. 0). The correlation map is composed of $\{\gamma_i \triangleq < \mathbf{y},\mathbf{u}^i >\}$, where γ_i is the correlation value corresponding to the i-translated watermark. Then, denoting the tolerance region by $\mathcal{R}$, the decoded bit value is given by

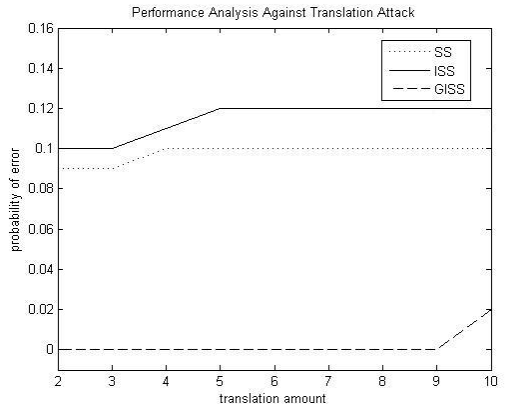
$$i^*=\arg\max_{i\in\mathcal{R}}|\gamma_i|, \quad \text{and} \quad \hat{b}=sign(\gamma_{i^*}).$$

5. EXPERIMENTAL RESULTS

The experimental setup is designed to compare the performances of three embedding methods (GISS, ISS, SS) when the watermarked signal is attacked by additive white Gaussian noise (AWGN) (Attack I) and translation (Attack II). In all experiments, we use $n=1000$. Operational error probabilities over 1000 realizations are used as the performance measure. The AWGN noise variance is chosen such



(a)



(b)

Figure 1: Performance plots comparing the proposed GISS-based watermarking algorithm with the SS and ISS methods; (a) additive white Gaussian noise attack, and (b) translation attack. In both panels, the vertical axis represents the operational probability of error, whereas the horizontal axis represents the additive noise variance and the translation amount, in (a) and (b), respectively. In the upper panel, solid, circle-solid, square-solid, star-solid, and diamond-solid lines correspond to no-search-SS, no-search-ISS, search-SS, search-ISS and GISS methods, respectively.In the lower panel, dotted, solid, and dashed lines correspond to SS, ISS and GISS algorithms, respectively.

that SNR = 10 dB. Within Attack I, while the other parameters remain constant, length of the tolerance region (denoted by l), is changed such that $0 \leq l \leq 101$. For SS and ISS methods, two different decoding approaches are applied. For the "no-search"-SS and -ISS methods, the sign of the embedded bit is checked only at the location where the watermark is embedded. For the "search"-SS and -ISS methods, the decoded bit is decided by checking the sign of maximum absolute value of the correlation map within the tolerance region. In case of Attack II, the signal is translated by k pixels, such that $2 \leq k \leq 10$. Decoding is performed only for the "search"-SS, "search"-ISS and GISS methods. Same set of signals and keys are used to compare all methods.

In Fig. 1, the operational probability of error values are depicted. For the GISS and ISS cases, λ and λ_{ISS} are chosen to be 1 and 0.95, respectively. SWR's for all cases are 30 dB to ensure to introduce the same amount of distortion. The experimental results show that the performance of the GISS method against AWGN attack is worse than SS and ISS methods, when decoding is performed only on the embedding location. However, when SS and ISS methods are also decoded by searching in the tolerance region, GISS method outperforms the others. Increasing the focus radius of the search region increases the probability of error for all methods on average (more simulations are required to get a smoother error probability behavior). When the focus radius is 0, the search region has only one pixel; therefore GISS method reduces to the ISS method. As expected, results of Fig. 5b reveal that the performance of the GISS method in the presence of the translation attack is superior to those of SS and ISS.

6. CONCLUSION

In spread-spectrum watermarking systems, it is well-known that the correlation between the host signal and the watermark degrades the performance of the system. Improved Spread-Spectrum (ISS) [1] aims to provide a solution to this problem via modifying the host prior to mark embedding, such that the correlation between the host signal and the watermark at the embedding location is reduced as much as possible. However, even in this case, geometric attacks are still problematic.

In this paper, we propose a new watermarking strategy, termed "Generalized Improved Spread Spectrum" (GISS), where the strategy we follow is similar to, but an extended version, of the one of ISS, so as to include robustness against translation-type geometric attacks. As a result of our method, we introduce the optimal design of the local host interference cancelation sequence in order to reduce the correlation between the host signal and the watermark, not only at the embedding location, but also within a certain neighborhood around it. Therefore, robustness against translation-type attacks is attained. Next, we apply the resulting scheme to signals via a practical algorithm and experimentally demonstrate its effectiveness. As a part of our future research, we plan to derive closed-form analytical results regarding the error probability of the proposed watermarking algorithm as well as accomplish the design and analysis of more advanced decoding techniques.

Acknowledgments: We thank Onur Özyeşil for his useful comments and discussions. We also thank the reviewers for their constructive comments.

7. REFERENCES

[1] H. S. Malvar and D. A. F. Florencio, "Improved Spread Spectrum: A New Modulation Technique for Robust Watermarking", *IEEE Trans. Signal Proc.*, Vol. 51, No. 4, 2003.

[2] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia", *IEEE Trans. Image Proc.*, Vol. 6, No. 12, pp. 1637–1687, 1997.

[3] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[4] J. Lichtenauer, I. Setyawan, T. Kalker, and R. Lagendijk, "Exhaustive Geometrical Search and the False Positive Watermark Detection Probability," *SPIE EI, Security and Watermarking of Multimedia Contents V*, San Jose, CA, Jan. 2003.