

PLL-BASED SYNCHRONIZATION OF DITHER-MODULATION DATA HIDING

K.M. Whelan, F. Balado, G.C.M. Silvestre, N.J. Hurley

University College Dublin, Belfield, Dublin 4 – Ireland

ABSTRACT

A new approach to synchronization recovery for signals watermarked using the Dither Modulation data hiding scheme is presented. The strategy followed involves the use of a digital phase-locked loop to track the offsets applied by an attacker to the sampling grid of the watermarked signal. The main element in this synchronization loop is the timing error detector which is responsible for generating an error signal, used to update the estimates of the applied offsets. It is shown how a timing error detector which has been used in digital communications may be easily adapted to extract timing information from DM watermarked signals. The performance of the proposed synchronizer is evaluated using the probability of decoding error under different models for the sampling grid offsets.

1. INTRODUCTION

Reliable watermark extraction following a desynchronization attack is one of the most challenging problems facing designers of data hiding systems. This attack first gained attention in the field of image watermarking, where early attacks consisted of simple global affine transformations (e.g. rotation, scaling, translation) of the watermarked signal. These attacks were usually coarse, i.e., the deviation of the sampling grid of the transformed image from that of the original image was relatively large.

In contrast, fine desynchronization attacks are those which apply a relatively small offset to each point of the original sampling grid. Such offsets may arise as the result of a specific desynchronization attack or as residual effects resulting from an inexact attempt at inverting a coarse desynchronization. Pertinent examples of a specific fine desynchronization attacks are the random bending attack (RBA) [1] for images or the grid-warping used in [2]. In this paper will focus on tackling fine desynchronization attacks on one dimensional watermarked signals.

The approach we take to this problem is the one followed in typical digital communications receivers, which consists of estimating the synchronization parameters. An interesting approach to estimating these parameters involves the use of phase-locked loop (PLL). This approach has proved successful, especially when the channel has time-varying synchronization parameters – a scenario where PLL can be used to track their variations over time.

Funded under Enterprise Ireland ATRP program, research grant ATRP2002/230 and by the European Commission through the IST Programme under Contract IST-2002-507609 SIMILAR.

To the best of our knowledge PLL-based techniques have never been investigated as a solution to the synchronization problem in data hiding. One of the possible reasons for this is that in early Spread Spectrum methods, the signal-to-noise ratio at the receiver is so low (due to the consideration of the relatively large variance host signal as interfering noise) that an estimation approach would most likely fail. With the advent of quantization-based data hiding methods offering host signal interference rejection, the resulting signal-to-noise ratio gain at the receiver allows this previously unexplored estimation strategy to be considered as a possible solution to the synchronization problem. In this paper we will present a synchronization method for signals watermarked using the quantization method Dither Modulation [3], and employs a PLL to track the sampling grid offsets applied by an attacker. The performance of the proposed scheme is examined when the offsets remain constant and when they vary across the sampling grid.

2. PROBLEM FORMULATION

Scalar random variables are denoted by capital letters, e.g., X , while their realizations are indicated with lowercase types, e.g., x . The host signal will be denoted by the N -length vector \mathbf{x} whose elements are independent with $X_k \sim \mathcal{N}(0, \sigma_x^2)$. We will assume that the information symbols to be embedded are statistically independent and equally likely.

2.1. Dither Modulation (DM)

In binary DM employing scalar uniform quantizers each sample of the watermarked signal carries one information symbol $b_k \in \{\pm 1\}$. This symbol is hidden by quantizing a sample of the host signal x_k to the nearest centroid $Q_{b_k}(x_k)$ of the shifted lattice $\Lambda_{b_k} \triangleq 2\Delta \mathbb{Z} + \Delta(b_k + 1)/2 + d_k$, with \mathbb{Z} the integer lattice, 2Δ the quantization step size and d_k a key-dependent pseudorandom dither value deterministically known to both the encoder and the decoder. The watermarked signal at sample k when b_k is embedded is given by $y_k = x_k - e_k = Q_{b_k}(x_k)$ where e_k is the quantization error with respect to Λ_{b_k} , i.e., $e_k \triangleq x_k - Q_{b_k}(x_k)$. Therefore the watermark $w_k \triangleq y_k - x_k$ is simply the quantization error e_k . The relative distortion introduced by the watermark is measured using the host-to-watermark power ratio (HWR = σ_x^2/σ_w^2).

The minimum Euclidean distance decoder in DM acts by quantizing a received vector \mathbf{z} , which is a distorted version of \mathbf{y} , yielding decisions

$$\hat{b}_k = \arg \min_{b \in \{\pm 1\}} |z_k - Q_b(z_k)|. \quad (1)$$

2.2. Modelling Desynchronization Attacks

We assume that the attacker interpolates \mathbf{y} using a given sampling period T_a and interpolation filter $h_A(t)$. The resulting continuous signal $y(t)$ is then resampled at points $\{kT_a + \tau_k\}$ where τ_k is an offset from the k^{th} original sampling point. The offsets $\{\tau_k\}$ may remain constant, vary deterministically or randomly over the sampling points. It is assumed that the attacker also adds some noise signal \mathbf{n} , independent of \mathbf{y} , to the resampled signal. One sample of the desynchronized signal arriving at the receiver is then given by

$$z_k = \sum_j y_j \cdot h_A((k-j)T_a + \tau_j) + n_k, \quad (2)$$

where $N_k \sim \mathcal{N}(0, \sigma_n^2)$. The relative strength of the Gaussian noise attack is measured using the watermark-to-noise power ratio (WNR = σ_w^2/σ_n^2). On the left of Figure 1 we may see the desynchronizing attack channel described by (2). Expanding (2) we have

$$z_k = y_k \cdot h_A(\tau_k) + \sum_{j \neq k} y_j \cdot h_A((k-j)T_a + \tau_j) + n_k. \quad (3)$$

The first term in (3) represents the information-bearing sample of interest. The second term represents Intersymbol Interference (ISI) introduced on y_k . As will be verified in Section 5, even very small values of τ_k can lead the probability of bit error at the decoder $P_b \triangleq \frac{1}{N} \sum_{k=1}^N \Pr(\hat{b}_k \neq b_k)$ to reach 1/2. This is due to the relatively large variance of the ISI term in (3).

Without any *a priori* knowledge of the offsets $\{\tau_k\}$, the receiver assumes that the samples of \mathbf{z} were taken at points $\{kT_a\}$. Clearly, to remove the ISI and scaling on the sample y_k the received signal must be resampled at points $\{kT_a - \tau_k\}$. The synchronization problem then, is to estimate the offsets $\{\tau_k\}$ applied by the attacker and use them to resample the received signal at the correct sampling points. Finally, notice that the actual sampling periods assumed by the attacker and receiver, respectively, are unimportant.

3. THE PHASE-LOCKED LOOP

A PLL can operate in either the continuous or discrete-time domain [4]. As the signals we deal with are discrete, synchronization will be performed entirely in this domain. In this scenario the term digital PLL is more appropriate however we will use the term PLL in what follows where the assumption of discrete domain operation is understood.

A first order PLL adjusts its estimate of the offset τ_k according to the recursion

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \nu \cdot \hat{\epsilon}_k, \quad (4)$$

where $\hat{\tau}_k$ is the receiver's estimate of τ_k , $\epsilon_k = \tau_k - \hat{\tau}_k$ is the error in this estimate and $\hat{\epsilon}_k$ is the receiver's estimate of ϵ_k . ν is the PLL gain parameter which is chosen to tradeoff agility of the tracking loop versus attenuation of the noise in the estimate $\hat{\epsilon}_k$. The estimates $\{\hat{\tau}_k\}$ are used in the loop to resample the received signal at points $\{kT_a - \hat{\tau}_k\}$. The device responsible for computing $\hat{\epsilon}_k$ is the timing error detector (TED) which we discuss next.

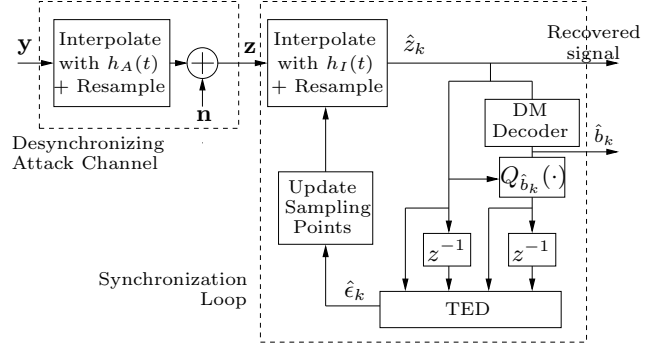


Fig. 1. PLL for synchronization of DM with TED operating in DD mode.

3.1. The Mueller and Müller (M&M) TED

We will focus on a particular TED proposed by Mueller and Müller [5] for timing recovery of baseband Pulse Amplitude Modulation (PAM) systems, which uses one sample per symbol along with the transmitted symbols to generate $\hat{\epsilon}_k$. In this scenario the output of a channel with overall impulse response $h(t)$ may be described by $s(t) = \sum_k a_k \cdot h(t - kT) + n(t)$, where $n(t)$ is additive white Gaussian noise, $1/T$ is the symbol rate and $\{a_k\}$ are the data symbols. Assuming a loss of synchronization, the receiver samples $s(t)$ at times $\{kT + \tau_k\}$

$$s(kT + \tau_k) = \sum_j a_j \cdot h((k-j)T + \tau_j) + n(kT + \tau_k). \quad (5)$$

To recover synchronization, the authors show in [5] how to build a timing error estimator,

$$\hat{\epsilon}_k = (s_k \cdot a_{k-1} - s_{k-1} \cdot a_k) / (2 \cdot E[a_k^2]), \quad (6)$$

where $s_k = s(kT + \tau_k)$. Notice that the statistics of the estimate in (6) are, by construction, dependent on $h(t)$. For example, in [5] it is shown that when $h(t)$ is a root raised cosine filter, the output of the TED in (6) exhibits some desirable statistical properties.

Finally, this TED can operate in either Data-Aided (DA) or Decision-Directed (DD) modes. In DA mode the transmitted symbols $\{a_k\}$ are known *a priori* by the receiver (usually as a preamble/training sequence in a practical system). In DD mode the symbols $\{a_k\}$ are replaced by estimates made by the receiver $\{\hat{a}_k\}$.

4. PLL-BASED SYNCHRONIZATION OF DM

We will show next how to apply the M&M TED to the data hiding scenario described. Comparing (2) and (5) the similarities between the system model in the our problem and that studied by M&M are evident. The watermarked samples $\{y_k\}$ in (2) are analogous to the symbols $\{a_k\}$ in (5). The impulse response of the interpolation filter $h_A(t)$ plays the role of the channel impulse response $h(t)$ in (5) and the sampling period T_a is analogous to the symbol period T .

We assume initially that the dither sequence $\mathbf{d} = \mathbf{0}$. Under this assumption, notice that the DM watermarked signal may be seen as a multilevel PAM signal with infinite

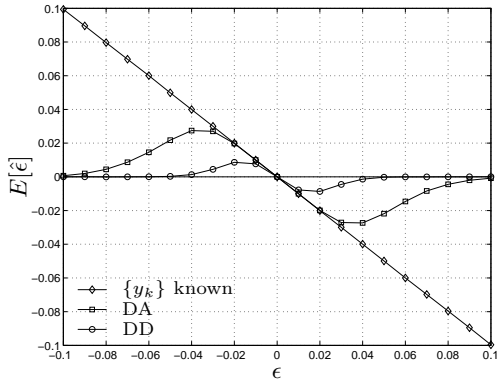


Fig. 2. $E\{\hat{\epsilon}\}$ of M&M TED operating on DM watermarked signal assuming $h_A(t) = h_I(t) = \text{sinc}(t)$. HWR = 20 dB, WNR = 10 dB.

equidistant levels $\{\Delta Z\}$. The encoder selects one of these levels to be sent over the attack channel based on symbol b_k and host sample x_k . Using this analogy, we can reformulate (6) to extract timing information from DM watermarked signals.

Let us assume initially that $\{y_k\}$ are known to the receiver, then (6) becomes

$$\hat{\epsilon}_k = (\hat{z}_k \cdot y_{k-1} - \hat{z}_{k-1} \cdot y_k) / (2 \cdot E[y_k^2]), \quad (7)$$

where \hat{z} is the signal produced by interpolating and resampling the received signal \mathbf{z} using the current information about the offsets. A single element of \hat{z} is given by

$$\hat{z}_k = \sum_j z_j \cdot h_I((k-j)T_s + \hat{\tau}_j), \quad (8)$$

where $h_I(t)$ is the impulse response of the interpolation filter used at the receiver. The expectation in (7) is the variance σ_y^2 of the watermarked signal with pmf $P_Y(Y = y) = \sum_{i=-\infty}^{\infty} w_i \delta(y - i\Delta)$, where $w_i = \frac{1}{2} \int_{(i-1)\Delta}^{(i+1)\Delta} p_X(x) dx$, $\forall i \in \mathbb{Z}$, are the weights on each delta function or the probability of level $i\Delta$. The required expectation is therefore computed as $E[y_k^2] = \sum_{i=-\infty}^{\infty} (i\Delta)^2 \cdot w_i$.

Finally, knowledge of y_k at the receiver is not a realistic assumption. We therefore examine strategies by which estimates of $\{y_k\}$ are formed by the receiver and used in place of their actual values in (7). There are two approaches to do this depending on the mode of operation of the TED.

4.1. DA & DD Modes of Operation

In DA mode the embedded symbol b_k is known to the receiver. In this case the receiver simply quantizes \hat{z}_k to the closest centroid of Λ_{b_k} to form an estimate of y_k

$$\hat{y}_k = Q_{b_k}(\hat{z}_k). \quad (9)$$

Notice that, differently to PAM in DA mode, *a priori* knowledge of b_k will only indicate to the receiver which lattice Λ_{b_k} the watermarked sample y_k belongs to but there is a probability of error in choosing the correct level.

In DD mode the embedded symbols are unknown to the receiver. Then, a hard decision is first made on the

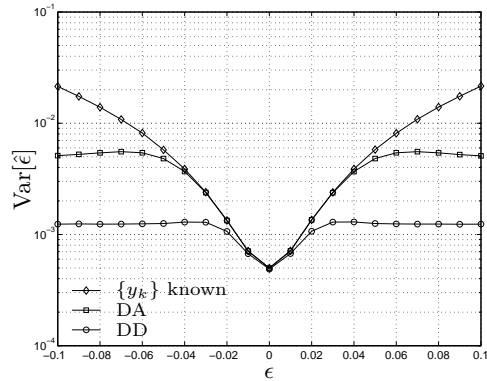


Fig. 3. Experimental variance of TED output assuming $h_A(t) = h_I(t) = \text{sinc}(t)$. HWR = 20 dB, WNR = 10 dB.

embedded symbol b_k according to (1) and used in place of b_k in (9). Figure 1 shows the proposed synchronization loop with the M&M TED operating in DD mode.

4.2. TED Performance Assessment

To assess the performance of the proposed timing recovery loop, we examine the statistical properties of the TED output when the feedback loop in Figure 1 is open. To this end the offset applied to the sampling grid of the input signal is constant $\tau_k = \tau$, $\forall k$. If the receiver assumes samples of the received signal taken at times $\{kT_a\}$, the error in its estimate will be $\epsilon = -\tau$. One of the desirable statistical properties that TED output is that $E[\hat{\epsilon}]$ should ideally be a straight line of negative unit slope passing through the origin. Then, for a given ϵ the timing function evaluates (in expectation) to the $-\epsilon$ which is the value required to correct the offset. The variance of $\hat{\epsilon}$ will also be examined in Section 5.

As mentioned in Section 3.1, $E[\hat{\epsilon}]$ will depend on the overall impulse response of the channel. In this case the impulse response is given by the convolution of $h_A(t)$ and $h_I(t)$. In general, $h_A(t)$ is not known to the receiver; however we assume that an attacker employs a $h_A(t)$ closely approximating a filter with ideal response $\text{sinc}(t)$ in order to minimize the perceptual degradation. If this assumption is valid and we choose $h_I = \text{sinc}(t)$, then the impulse response resulting from their convolution will be a close approximation to $\text{sinc}(t)$. In all the experiments which follow we will assume for simplicity that $h_A(t) = h_I(t) = \text{sinc}(t)$. The interpolation filter $h_A(t)$ may be different from the ideal case, but more in-depth studies of the associated distortion are necessary.

5. EXPERIMENTAL RESULTS

Figure 2 shows $E[\hat{\epsilon}]$ as a function of ϵ for different modes of TED operation. For each mode, when operating in their respective linear regions, the TED (and the PLL) will track the offset well. Although the linear range achieved seems small, the continuity constraint on the offsets discussed in Section 1 means that the PLL can still track a large accumulated offset. This is under the assumption that the dif-

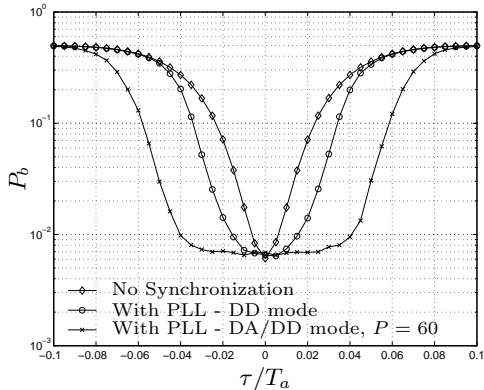


Fig. 4. Probability of decoding error for DM under desynchronization attack with $\tau_k = \tau \forall k$. HWR = 20 dB, WNR = 10dB, $\nu = 0.01$, $N = 600$.

ference between offsets applied to adjacent sampling points is in the linear range. Figure 3 shows the variance of the timing error estimate as a function of ϵ for all three modes of operation. Although we have assumed for notational simplicity that the dither sequence $\mathbf{d} = \mathbf{0}$, we have verified experimentally that the performance of the TED is unaffected by the inclusion of a dither sequence.

Next we close the loop and perform synchronization using P_b as the performance measure. The PLL parameter ν is chosen empirically. We start with a simple case where $\tau_k = \tau, \forall k$. Figure 4 shows P_b for DM under this attack. Starting with case where no synchronization is attempted we can see that even very small values of τ lead to a large increase in P_b . We can see in this figure the reductions in P_b offered by the proposed PLL synchronization scheme with the TED operating in DD mode. Also shown is the case where the TED operates in DA mode for the first P samples of the received signal, i.e., the receiver knows the first P embedded data symbols. The TED is then switched to DD mode. Figure 4 shows the improvement offered by this strategy when $P = 60$. It should be noted the P_b is measured over N samples. During the initial period when the PLL is converging to the offset, the decoder is more likely to make incorrect decisions as synchronization has not been achieved. Asymptotically in N this effect on P_b will be negligible so we can expect further reductions in P_b over those shown in Figure 4 for larger N .

Next we consider a more complex desynchronization attack where the sampling grid offsets are modelled using a random walk, i.e., $\tau_{k+1} = \sum_{j=1}^k g_j$ where $G_j \sim \mathcal{N}(0, \sigma_g^2)$ and $\tau_1 = 0$. This model takes into account the usual continuity constraints on the applied offsets, also observed in [1, 2]. Figure 5 shows the P_b performance of the proposed synchronization scheme in DD mode as a function of σ_g/T_a for this attack. From this we can see that if the offsets $\{\tau_k\}$ applied by an attacker are slowly varying then the proposed system achieves the almost same P_b as the case where no desynchronization is performed.

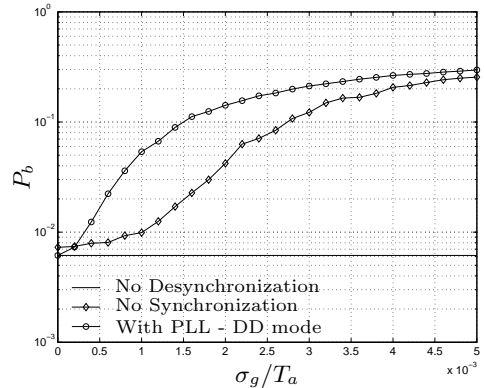


Fig. 5. Probability of decoding error for DM under desynchronization attack with offsets modelled as a random walk. HWR = 20 dB, WNR = 10 dB, $\nu = 0.04$, $N = 600$.

6. CONCLUSIONS

We have presented a novel PLL-based approach to synchronization recovery for DM watermarked signals. The ability to track different fine desynchronization attacks has been demonstrated. Some issues which need further investigation include the effect of a mismatch between interpolation filters used in the attack channel and by the receiver on the performance of the proposed scheme. Also assessing the performance of the proposed synchronizer relative to perceptually acceptable limits of fine desynchronization attacks on real signals requires further investigation. We are currently examining the inclusion of distortion compensation and the extension of the M&M TED to two dimensional settings.

7. REFERENCES

- [1] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Attacks on copyright marking systems," in *Second International Workshop on Information Hiding*, D. Aucsmith, Ed., Portland, Oregon, U.S.A., April 1998, pp. 219–239, Springer-Verlag.
- [2] R. Bäuml, J.J. Eggers, and J. Huber, "A channel model for watermarks subject to desynchronization attacks," in *Procs. of SPIE: Security and Watermarking of Multimedia Contents IV*, San José, USA, Jan 2002, vol. 4675.
- [3] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.
- [4] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing*, Wiley Series in Telecommunications and Signal Processing. John Wiley and Sons, Inc., 1998.
- [5] K.H. Mueller and M. Müller, "Timing recovery in digital synchronous data receivers," *IEEE Trans. on Communications*, vol. 24, pp. 516–531, May 1976.