Witsenhausen's counterexample and its links with multimedia security problems

Pedro Comesaña[1,2], Fernando Pérez-González[1,2], and Chaouki T. Abdallah[1⋆]

[1] Electrical and Computer Engineering Department, University of New Mexico, Albuquerque, NM, USA,
pcomesan@unm.edu
[2] Signal Theory and Communications Department, University of Vigo, 36310, Vigo, Spain

Abstract. Witsenhausen's counterexample was proposed more than four decades ago in order to show that affine control strategies are not optimal for systems with non-classical information patterns. Finding the optimal solution to Witsenhausen's problem however remains an open problem. Recently, the stochastic control community has re-discovered Costa's Dirty Paper result as a potential solution to Witsenhausen's problem. In this paper the similarities and differences between Witsenhausen's scenario and multimedia security problems are reviewed, and the historical evolution of the solutions to Witsenhausen's problem compared with those proposed for watermarking detection.

Keywords: Control Theory, Dirty Paper Coding, Multimedia Security, Quantization-Based Techniques, Watermark Detection, Witsenhausen's counterexample

1 Introduction

Control theory is a multidisciplinary field of research where Engineering, Mathematics and Physics interplay. The goal of control design is to modify the input of a dynamical system (which may be thought of as a physical plant) in order to make the system's output follow a reference value. The combined system is composed of a physical plant, and a controller (usually implemented in software), which is the part of the system in charge of modifying the plant's input. Whenever randomness is involved in the system input or dynamics, stochastic control is typically used. Please see [12] for a detailed discussion of such concepts.

⋆ Research supported by the European Union under project REWIND (Grant Agreement Number 268478), the European Regional Development Fund (ERDF) and the Spanish Government under projects DYNACS (TEC2010-21245-C02-02/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010), the Galician Regional Government under projects "Consolidation of Research Units" 2009/62, 2010/85 and SCALLOPS (10PXIB322231PR), and the Iberdrola Foundation through the Prince of Asturias Endowed Chair in Information Science and Related Technologies.

In real applications, several different controlled systems (with the corresponding controllers) could co-exist. If those controllers are allowed to make decisions without communicating between them or communicating with a centralized controlling entity, the resulting problem is usually denoted Decentralized Control. One of the most interesting cases in that scenario is that where the different controllers observe different input signals. It is within this particular framework that Witsenhausen proved more than four decades ago that the optimal control strategy, even when linear systems with quadratic performance objective and Gaussian noise are considered, can not be an affine function of the state [36].

Although the non-optimality of affine strategies has been formally proven, Witsenhausen's work does not establish what the optimal solution is. Indeed, this problem has received great attention, and even today the optimal solution for Witsenhausen's problem remains elusive. Interestingly, in the last years Grover *et al.* have pointed out links between distributed control, in particular one of the most promising approaches to Witsenhausen's problem, and Costa's dirty paper coding [21, 22].

The objective of our work is to present Witsenhausen's problem from a media security perspective. A review of the solutions proposed for this problem are introduced and compared for the first time with those proposed for different multimedia security applications; similarities and differences between Witsenhausen's counterexample and multimedia security problems are pointed out.

The remaining of this paper is organized as follows: after briefly introducing our notation in Sect. 1.1, Witsenhausen's problem is formally presented in Sect. 2. Digital watermarking classical problems are reviewed in Sect. 3, and the solutions proposed to Witsenhausen's problem are summarized and compared with watermarking detection strategies in Sect. 4. Then, Witsenhausen's problem is compared with two classical problems in multimedia security: authentication (in Sect. 5) and reversible watermarking (in Sect. 6). Finally, the conclusions of this work are presented in Sect. 7.

1.1 Notation

We denote scalar random variables with capital letters (e.g. X) and their outcomes with lowercase letters (e.g. x). The same notation criterion applies to L-dimensional random vectors and their outcomes, denoted in this case by bold letters (e.g. $\mathbf{X}$, $\mathbf{x}$). The ith component of a vector $\mathbf{X}$ is denoted as X_i. The power of signal $\mathbf{X}$ is denoted by $\sigma_X^2 \triangleq \mathrm{E}\{X_i^2\}$, being valid for any i, as the components of the considered vectors are assumed to be i.i.d..

2 Witsenhausen's counterexample

The general objective in stochastic control is to minimize the expected value of a target function, for given noise distributions [12]. The most basic scenario is based on what is referred to as the *classical information pattern*, which assumes that all actions performed by the controllers are based on the same data, and

that any data available at a given time will be also available at all later times. In that framework, and if linear systems, quadratic objective criteria and Gaussian noise are considered, the optimal solution was proven to be an affine function of the system's state.

In 1968 Hans S. Witsenhausen [36] showed by means of a counterexample, that affine control functions are no longer optimal solutions to problems where the information pattern is not classical. This counterexample is based on the modification by a first controller of a variable x (which in most of his paper is assumed to follow a Gaussian distribution) by adding a variable w. The resulting variable $y = x + w$ passes through a Gaussian channel; we denote by n the realization of that Gaussian channel, and by $z = y + n$ the output variable, which in turn feeds into a second controller that aims to provide an estimate $\hat{y}$ of y based only on z with an estimation error $q = y - \hat{y}$. This framework does not follow a classical information pattern, since x is known only to the first controller, but not to the second. The proposed target function is

$$k^2 \mathrm{E}\{W^2\} + \mathrm{E}\{Q^2\},$$

i.e., the sum of a weighted version of the variance of the signal introduced by the first controller (that is, $k^2\mathrm{E}\{W^2\}$) and the variance of the estimation error at the second controller $\mathrm{E}\{Q^2\}$ (assuming that both W and Q are zero-mean).

Witsenhausen derived the optimal *affine* solution, and compared its achieved target function value with that obtained by using $w = \sigma_X \mathrm{sign}(x) - x$, and $q = y - \sigma_X \tanh(\sigma_X z)$. He was then able to show that the value of the target function achieved by this strategy is strictly smaller than that obtained by the optimal affine solution thus proving that affine solutions for control problems with non-classical information patterns are no longer optimal.

Nevertheless, as Witsenhausen established in his paper, the solution proposed as counterexample is itself far from being optimal. Since the publication of the original paper, a large number of papers in the stochastic control field have been published in an attempt to derive the optimal controlling strategies for Witsenhausen's counterexample. An optimal solution for a general framework has not yet been found, but the proposed schemes have considerably reduced the value of the target function resulting from Witsenhausen's original strategy. Some of these proposals are reviewed in the following sections, then compared with different strategies designed for dealing with multimedia security problems.

2.1 Vector Witsenhausen's problem

Although Witsenhausen's original counterexample used scalar signals, Grover and Sahai have introduced the vector version of that problem [20], where the signal entering the first controller $\mathbf{x}$ is modified by the addition of a signal $\mathbf{w}$, yielding $\mathbf{y}$. The observation noise $\mathbf{n}$ is added to $\mathbf{y}$, so that the second controller must provide an estimate of $\mathbf{y}$, which we denote by $\hat{\mathbf{y}}$, based just on $\mathbf{z} = \mathbf{y} + \mathbf{n}$. Similarly to the scalar case, the target function in the vector scenario is given by

$$k^2 \mathrm{E}\{||\mathbf{W}||^2\} + \mathrm{E}\{||\mathbf{Q}||^2\}.$$

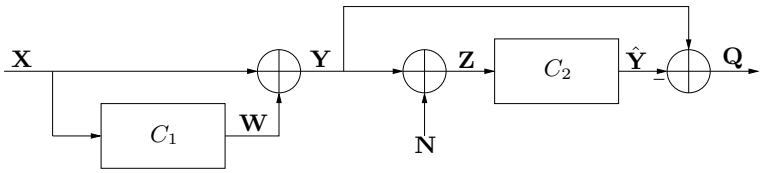

Fig. 1. Block-diagram of Witsenhausen's multidimensional problem.

As we discuss later, this extension to the multidimensional case allows for the use of more complex coding strategies that lead to target function reduction.

A block-diagram of Witsenhausen multidimensional problem is found in Fig. 1.

3 Classical problems in digital watermarking

In digital watermarking, two major problems are typically distinguished: one-bit watermarking (a.k.a. *zero-bit watermarking*, or *watermark detection problem*) and *multibit watermarking* (a.k.a. *watermark decoding problem*).[3] In the first problem, the receiver side tries to determine whether a given watermark, which is *a priori* known, is present or not at the available signal. This then is a binary hypothesis problem. A block-diagram of the watermark detection problem is shown in Fig. 2. On the other hand, in the multi-bit watermarking problem the presence of a watermark is assumed, and the objective is to determine which message, from a finite set of possibilities, has been embedded at the transmitter side; consequently, it is a multiple hypothesis problem.

Due to their different natures, the measures used for quantifying the goodness of one-bit and multi-bit watermarking are also different. For one-bit watermarking the probability of false-positive (or false-alarm) and the probability of false-negative (or missed-detection) are used, while in the multi-bit watermarking, the probability of decoding error (or some related measure, as the Bit Error Rate) is typically used. Considering these measures, and the imperceptibility constraints the watermark detection and decoding problems are formalized as

$$\min_{\mathbf{W}:\mathrm{E}\{||\mathbf{W}||^2\}\leq D_e, P_{fp}\leq P_{fp}^{target}} P_{fn}, \text{ and}$$
$$\min_{\mathbf{W}:\mathrm{E}\{||\mathbf{W}||^2\}\leq D_e} P_e,$$

respectively, where P_{fp} stands for the false positive probability, P_{fn} for the false negative probability, P_e for the probability of decoding error, and D_e for the maximum allowed mean embedding distortion. Note that the watermark detection problem may be defined from its dual counterpart, i.e. by fixing a target false negative probability and minimizing the false positive probability.

[3] Other watermarking problems, such as those of steganography, authentication, or reversible watermarking, may be regarded as subclasses where additional constraints or modified versions of the target function are considered.

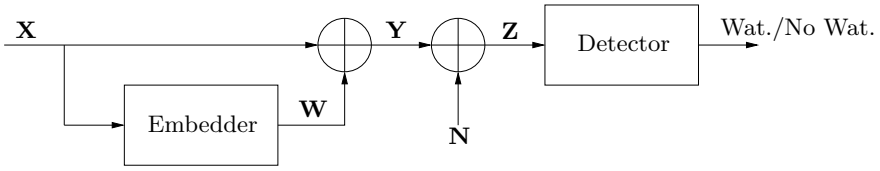

Fig. 2. Block-diagram of watermark detection problem.

Since the birth of digital watermarking in the late 1990's, a large number of different strategies have been proposed for dealing with both problems. The significant advances achieved so far have dramatically improved the performance of early watermarking schemes, to the point that a well-grounded theory is now available.

Although the block-diagrams for Witsenhausen's counterexample and the watermark detection problem are rather similar, one wonders how deep this similarity actually is, and what are the main differences between both problems. The target of the next section is to explore these similarities and differences.

3.1 Similarities and differences of Witsenhausen's counterexample with watermark detection

Although at first, the two problems may seem completely different, the fact is that the problems share some common traits:

- Both schemes have a non-classical information pattern. Specifically, in watermarking detection, the embedder observes the original host signal $\mathbf{x}$, while the detector must make its decision based solely on the received attacked signal, $\mathbf{z}$. Similarly, in Witsenhausen's counterexample, the first controller observes $\mathbf{x}$, while the second controller estimates its output based only on the observation of $\mathbf{z}$.
- Another similarity stems from the constraint on the watermark variance; this constraint is explicit in the watermark detection problem, and implicit in Witsenhausen's counterexample. Indeed, in the latter case the watermark variance (i.e., the variance of the signal introduced by the first controller) can not be arbitrarily large, as this would increase the score of the target function yielding a *de facto* non-feasible point. In other words, it is the target function itself that constrains the considered embedding functions to use a reduced watermark variance. This constraint on the watermark variance will strongly influence which codes are good for transmitting the desired information from the embedder (or the first controller) to the detector (or the second controller).
- In both cases, the transmitted signals are sent through an additive white Gaussian channel. As the channel noise distribution is the same in both cases, the shape of the used codes will share some geometrical characteristics that make them suitable for coping with such noise distribution.

- Both scenarios deal with zero-rate problems, meaning that no additional information, besides that used for estimating the signal produced by the first controller or the presence of the watermark, is transmitted. This the main reason for focusing our comparison of multimedia security and Witsenhausen's counterexample on watermark detection techniques (and in the next sections, on other zero-rate multimedia security problems). A remarkable exception to the zero-rate nature of Witsenhausen's counterexample can be found in [21], where Grover and Sahai consider the use of Costa-based schemes for conveying additional information.
- In both cases, the host signal may be interpreted as an interfering factor. This is now obvious and widely recognized within the watermarking research community. In fact, one of the major drawbacks of early embedding schemes (both for detection and decoding watermarking) was the interference due to the host signal, which made the reliable transmission of information very hard. One of the first mechanisms devised for dealing with this problem was to introduce the watermark in reduced-dimensionality domains, in an attempt to improve the signal-to-noise (SNR) ratio achieved in the original domain, but at the expense of a reduced available payload.

  Similarly, in Witsenhausen's counterexample, the larger the variance of the host signal $\mathbf{x}$, the more difficult it is for the second controller to estimate the transmitted signal $\mathbf{y}$. As a degenerate case illustrating this fact, one may think of a zero-variance host signal; in such case both controllers could use trivial strategies (e.g., $\mathbf{w} = 0$ and $\hat{\mathbf{y}} = 0$) that provide a null-score of the non-negative target function, and, consequently, an optimal solution. As the variance of the host signal is increased, the design of suitable strategies become harder and the target function increases.

Concerning the differences between both problems, probably the most obvious is the fact that in Witsenhausen's counterexample, an estimate of the watermarked signal must be provided. Nevertheless, if one considers this estimate as a two-step procedure (as is done in state-of-the-art schemes), where the watermarked signal estimate is itself based on a preliminary estimate of the codeword used at the first controller, this difference is not so distinctive as one could initially think. Indeed, this first stage of the estimate in Witsenhausen's counterexample may be seen as a characteristic shared with the watermark detection problem, as in both cases one is looking for the codeword used at the embedder. In Witsenhausen's problem however, there is only one set (codebook) of possible codewords as opposed to the watermark decoding problem, where there is a codebook for each possible transmitted message.

Consequently, although both problems seem to be rather different at first sight, the fact is that strong links between them exist. These connections constitute the reason why techniques used to solve each problem are quite similar. Since the watermark detection problem is the one in multimedia security that shares more characteristics with Witsenhausen's scenario, in the next section we review some of the strategies proposed in the literature for dealing with

Witsenhausen's counterexample, and analyze their relationship with well-known watermark detection methods.

4 A review of existing techniques for Witsenhausen's problem

The first strategies proposed for minimizing Witsenhausen's target function date back to Witsenhausen's original paper [36]. The first of them, is the non-optimal affine solution. In that case, $y = \lambda x$, with λ an appropriate real constant. The multidimensional version of this embedding strategy has been used for a long time in watermarking, both for detection and decoding problems, where it is known as *Multiplicative Spread-Spectrum*. For example, Barni *et al.* analyze in [4] a watermarking scheme with embedding function $y_i = x_i(1 + \lambda b_i)$, both for watermark detection and decoding, where b_i the message to be hidden. This was not the first time that such a strategy was used in a watermarking context; Cox *et al.* [10] had employed it several years before, although just for multibit watermarking.

In the counterexample given by Witsenhausen, the signal produced by the first controller is constructed as $w = \sigma_X \text{sign}(x) - x$, so $y = \sigma_X \text{sign}(x)$. The result resembles a sign-quantization strategy. This strategy may be interpreted as an example of the well-known Quantization Index Modulation (QIM) proposed by Chen and Wornell [8], where only one index is used and the quantizer has only two possible levels that are symmetrical (antipodal) about the origin. Although this is clearly not the most general configuration of QIM, this embedding strategy also fits within the QIM framework. Additionally, and although this strategy was sufficient in Witsenhausen's scenario for improving the results achieved by the optimal affine strategy, from the watermarking perspective it has the serious drawback of the large embedding distortion that it requires (as the watermarked signal is binary antipodal). To the best of our knowledge, this embedding strategy has never been used in the multimedia security field.

A non-affine strategy similar to that proposed by Witsenhausen, but with improved performance, was put forward by Bansal and Basar [3]. In this case $w = \sigma_X \sqrt{2/\pi} \text{sign}(x) - x$, but a binary antipodal quantizer is used. Furthermore, this work proves that affine solutions may still be optimal even in non-classical information patterns scenarios, if the target function does not depend on the product of the control variables. Finally, they also considered a generalized control strategy to exploit the benefits from both linear and non-linear strategies, proposing the use of $w = \epsilon \text{sign}(x) + \lambda x - x$.

An interesting result illustrating the hardness of the search for the optimal solution to Witsenhausen's counterexample is due to Papadimitriou and Tsitsiklis [33]. The authors proved that the discretized version of Witsenhausen's problem is fundamentally intractable, as it is an NP-complete problem therefore ustifying the lack of progress in the search for the optimal solution. Additionally, they relate the complexity of the discrete problem with that of the continuous one, proving the nonexistence of realistic algorithms for the continuous case.

In [11] Deng and Ho used ordinal optimization to study Witsenhausen's counterexample. Specifically, they implicitly deal with the problem of the large embedding distortion induced by the use of binary antipodal quantizers. By introducing multilevel quantizers, the authors were able to reduce the quantization error, or in other words, the variance of the signal introduced by the first controller. As long as the quantization levels are far enough for ensuring their correct estimation at the detector, an increased number of quantization levels reduces Witsenhausen's cost.

This idea is further explored at [25] by Lee *et al.*. In that work the authors study the effect of the number of quantization levels (if the quantization levels are broken down, the first stage of Witsenhausen's target function is reduced, but the difficulty of estimating y at the second controller increases), the quantization boundary values, and the quantized values. Even more interestingly, by using numerical methods, they prove that by considering piecewise linear functions, instead of pure step functions, the target function may be decreased. This last result clearly links to another well-known concept in watermarking: *Distortion Compensation* (DC). The basic idea behind Distortion Compensation is that by adding back a part of the quantization error to the quantized signal, the quantization error variance (i.e., the watermark variance in the watermarking application, or the variance of the signal introduced by the first controller in Witsenhausen's scenario) is reduced, allowing to inflate the used quantizer, and consequently providing increased robustness against channel attacks. This idea, which was proposed for the first time by Costa [9] in a purely information theoretic scenario, has been exploited in the watermarking field by the DC-QIM [8] and Scalar Costa Scheme (SCS) [13] schemes. Remarkably, this link was overlooked in the stochastic control literature at [25] (where DC quantization schemes were implicitly proposed for the first time in that scenario), and later discovered by Grover and Sahai (see, for example, [21]), whose works we will discuss below. Finally, a hierarchical search numerical method is proposed in [25] for the computation of the parameters of the proposed scheme (number of quantization levels, quantization boundaries, and value of the quantization levels).

Similar results showing the convenience of using sloped step functions (or similarly distortion-compensated quantization strategies) have also been obtained by Baglietto *et al.* [2] and Li *et al.* [26]. Baglietto *et al.*'s methodology is based on constraining the control functions to have some parameterized fixed structure, denoted by nonlinear approximation networks; stochastic approximation is used for solving the resulting nonlinear programming problem. On the other hand, the approach followed by Li *et al.* is based on discretizing the problem and formulating it as a potential game; the authors solve it by using the learning algorithm known as *Fading Memory Joint Strategy Fictitious Play with Inertia* [30].

In recent years, Grover and Sahai have published a series of interesting articles on Witsenhausen's counterexample, where they establish connections between this problem and open problems in communications, such as those of the cognitive radio channel, the multiple access channel with partial state in-

formation at the encoder, state masking [32] or Dirty Paper Coding (DPC) [9] (although surprisingly they do not mention the links with existing data hiding methods, as the aforementioned DC-QIM and SCS). Indeed, in [21] the authors interpret Witsenhausen's problem as a particular case of the wireless communication problem which they name *Assisted Interference Suppression*. Furthermore, in [21] the authors also propose several solutions to Witsenhausen's vector problem, where the most suitable alternative is chosen depending on the working-point (defined by σ_X^2 and k^2):

- $\mathbf{w} = -\mathbf{x}$, so $\mathbf{y} = \mathbf{0}$. In this case, the estimation at the second controller is trivial, and the variance of the estimation error is null. Therefore, only the first stage of Witsenhausen's cost will be non-null (specifically, $k^2\sigma_X^2$). This strategy makes sense for small values of k^2 and small values of σ_X^2.
- $\mathbf{w} = \mathbf{0}$, so $\mathbf{y} = \mathbf{x}$. This may be considered to be the counterpart of the previous scheme, as the first stage of Witsenhausen's target function is zero. The estimation at the second controller will be an MMSE estimation (i.e., Wiener's filter) of $\mathbf{y}$ given $\mathbf{z}$, yielding a Witsenhausen's cost of $\sigma_X^2\sigma_N^2/(\sigma_X^2 + \sigma_N^2)$. This strategy makes sense for large values of k^2.
- A randomized nonlinear controller based on the quantization of $\mathbf{x}$ by using a random codebook of square radius per dimension equal to $\sigma_X^2 - \sigma_W^2$. This is a *pure* quantizer, in the sense that distortion compensation is not considered. Due to its connections with the rate-distortion function, and noticing that the cardinality of the quantizer is chosen in order to avoid decoding mistakes, the authors name this strategy *Joint Source-Channel Coding* (JSCC). This approach makes sense for small values of k^2 and large values of σ_X^2.
- Dirty Paper Coding based strategy [9]. Similarly to the previous scheme, a random codebook is used, although in this case the square radius per dimension is equal to $\alpha^2\sigma_X^2 + \sigma_W^2$, where α stands for the distortion compensation parameter. Interestingly, the authors consider both the case where DPC is used for conveying additional information to that about $\mathbf{y}$ (related to the previously explained multi-bit watermarking problem), and the case where the quantization aims at aiding the estimation of $\mathbf{y}$. Therefore, no additional information is transmitted, yielding a Costa's zero-rate problem, related to watermark detection problem. The estimation stage is based on first quantizing the received signal $\mathbf{z}$ with the considered codebook (as done in conventional DPC schemes), followed by a second stage where an MMSE estimator is applied to estimate the self-noise at the first controller. The quantization error at the first controller scaled by $(1 - \alpha)$, given the total quantization noise at the second controller (which assuming that the codeword used at the first controller is correctly determined, is the sum of the self-noise and the channel AWGN). This two-step procedure reveals one of the similarities between Witsenhausen's counterexample and watermark detection problems as outlined in Sect. 3.1.
- Marginal improvements are also achieved by implementing the first controller as a two-stage process in which $\mathbf{X}$ is first scaled down to reduce its power, and then DPC is applied.

Additionally, it is important to point out that in [21] the authors show that the performance that can be achieved by using Witsenhausen's multidimensional version problem, is strictly better than that achieved by using its scalar counterpart, due to the advantages brought about by multidimensional codes.

Grover *et al.* also proposed in [19] the use of lattice-based quantization strategies for performing the quantization. Their analysis is based on considering the packing and covering radii, following an approach similar to that in [14].

Although lattice-based quantization strategies to multimedia security have been typically used for mult-ibit watermarking, where their optimality was proven by Erez and Zamir [15], there are also examples where they have been used for detection purposes. One of the most relevant contributions in this direction is due to Liu and Moulin [27], where the authors derive the error exponents of watermarking detection both for Additive Spread Spectrum and QIM. They assume that lattices are used in the QIM case and that the distortion compensation parameter takes the value proposed by Costa [9], being the detection region a hypershpere. Other work where quantization-based schemes were suggested for dealing with one-bit watermarking is [34], where Pérez-Freire *et al.* proposed to quantize the correlation of a series of pseudorandom-sequences and the original host signal (i.e., the projection of $\mathbf{x}$ onto a pseudorandomly generated subspace) without considering distortion compensation. Furthermore, they propose several detection regions in order to determine if the received signal $\mathbf{z}$ is watermarked or not.

Finally, we would like to mention that strategies which are not based on quantizing the host signal were proposed in the watermarking literature for reducing the host signal interference. As an example, in [6] Cannons and Moulin's effectively reduce the host signal interference thanks to the exploitation of a hash of the original signal available at the detector. Given that the hash provides information about the original host signal belonging to a given subset of the signal space, it allows to condition the probability density function (pdf) considered by the detector.

Another strategy for reducing host signal interference not based on quantization, is the Improved Spread Spectrum technique due to Malvar and Florencio [29], which was initially intended for decoding watermarking scenarios. In contrast, no control strategies similar to [6] or [29] have been suggested for dealing with Witsenhausen's counterexample.

5 Links between Witsenhausen's counterexample and authentication

Multimedia authentication methods may be basically divided into two main categories: those that complement the digital content under analysis with an additional authentication code used for checking the authenticity (e.g. Image Messages Authentication Codes [37], Approximate Message Authentication Codes [18], or Noise Tolerant Message Authentication Code [5]), and those that embed the information required for checking whether the considered contents is a

fake or not, in the content itself by using watermarking techniques. As we are interested in the comparison with Witsenhausen's counterexample, we focus in this work on the second category. For those methods, and similarly to zero-bit watermarking, the authentication problem is a binary hypothesis test, where the binary decision tries to determine if a given content was modified or not. In any case, as long as the authenticating methods are based on estimating the codeword used at the embedder, most of the similarities pointed to above between Witsenhausen's counterexample and watermark detection still hold.

These similarities are even more pronounced for some particular authentication methods, such as the one due to Martinian *et al.* [31]. In that work, the authentication process is split into two steps: first, the codeword used at the embedder is estimated, and then a reconstruction of the original host signal is produced. This estimate of the original host signal $\mathbf{x}$ (instead of $\mathbf{y}$, as in Witsenhausen's counterexample) is required to be free from the effects of any modifications by the editor, so it will be effectively defined by the encoder. Furthermore, the set of possible modifications on the watermarked signal $\mathbf{y}$ is constrained to verify a so-called reference channel model. For the case of Gaussian host and channel and quadratic distortion measures, Martinian *et al.* use a Gaussian random codebook for quantizing the original host signal, including distortion compensation.

A scheme conceptually similar to that in [31], but without the final estimation stage, was proposed by Fei *et al.* [16]. There, the authors define the *admissible set*, a deterministic set that characterizes the legitimate modifications that an authenticated signal may be subjected to in order to be considered authentic, similarly to the reference channel in [31]. Additionally, Fei *et al.*'s scheme is also based on the quantization of the original host signal, although in this case the use of distortion compensation is not considered, and lattice-based quantization instead of random quantization is used. This difference raises a security problem since a periodic structure, such that of the lattice, should not be used for determining the authenticity of a content, since as soon as an attacker has access to any authenticated object he/she could produce as many falsely authenticated contents as he/she wishes. This of course is not the case for random-coding based schemes, where the observation of a centroid does not supply any information about the rest of the codewords in the considered . On the other hand, this lack of structure renders random-coding-based schemes difficult to implement in practice. The solution proposed by Fei *et al.* relies on two nested lattices, where the coarse lattice is actually used for quantizing the original host signal, and in a second stage a point of the fine lattice within the Voronoi region of the chosen coarse-lattice centroid is pseudo-randomly selected depending on a secret key and the coarse-lattice centroid itself. In doing so, the security of the resulting scheme can be increased, in the sense that an attacker observing other watermarked contents but who is not aware of the secret key, could not produce a falsely authenticated content.

Interestingly, both embedding authentication methods just reviewed reveal a subtle similarity and a inherent difference between Witsenhausen's scenario

and authentication applications. On one hand, Witsenhausen's solutions can be seen to include a reference channel. This is the case for example, for DPC-based techniques where the distortion compensation parameter value is determined by considering a certain channel distribution (AWGN) and a given noise variance. If the actual channel at the output of the first controller were different, then estimation errors could arise; for example, if the variance of the AWGN were underestimated when computing the distortion compensation parameter, the total noise variance might be larger than the maximum allowed one for guaranteeing an error-free decoding. On the other hand, one does not need to take into account the security constraints in Witsenhausen's counterexample, as an attacker does not exist in that framework.

6 Links between Witsenhausen's counterexample and reversible watermarking

Since reversible watermarking also aims to estimate a signal which is not known at the receiver side, one may think that this is a multimedia security problems that is more similar to Witsenhausen's problem. Nevertheless, several characteristics of reversible watermarking, and especially of the scenarios where its use was proposed, give rise to some important differences:

- All reversible watermarking schemes we surveyed are mult-ibit. This probably owes to the applications reversible watermarking was designed for, e.g. medical or military images [17], in contrast to copyright applications, where watermark detection is typically used.
- Similarly to [31], in reversible watermarking the decoder tries to estimate the original host signal $\mathbf{x}$, instead of the watermarked signal $\mathbf{y}$ (as is the case in Witsenhausen's problem).
- All studied reversible watermarking schemes consider a discrete host alphabet, in contrast to Witsenhausen's Gaussian-distributed x. This is not a trivial difference, as it will strongly determine the feasible solutions in each case.
- Most of the studied reversible watermarking schemes do not consider that the watermarked content could be further corrupted by noise, i.e. they assume that $\mathbf{n} = \mathbf{0}$. Again, this assumption may make sense in most reversible watermarking applications. One exception to this lack of channel noise consideration is found in [17], although it is studied just from a empirical point of view.
- Most of the studied schemes are constrained to provide perfect estimates of the host signal, i.e. $\hat{\mathbf{X}} = \mathbf{X}$. Remarkably, this is not the case in [35], where a non-null distortion between the host signal and its estimate is considered.
- Most of the studied schemes are based on performing a lossless coding of the Least Significative Bits (or of some kind of prediction error) and use the remaining room for sending the additional data (e.g. [17], [1], [24]). Nevertheless, this strategy was proven to be non-optimal [23].

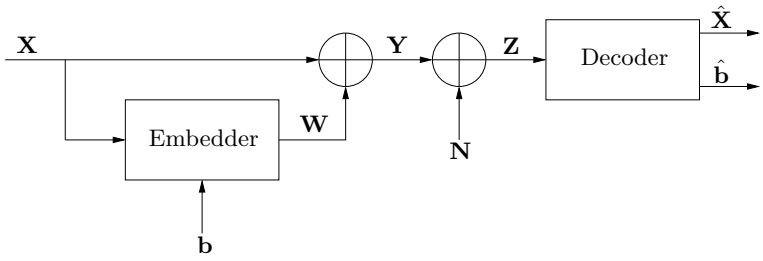

Fig. 3. Typical block-diagram of reversible watermarking.

Concerning the similarities, some reversible watermarking schemes using quantization methods exist (e.g. [7]), although this is not the most general approach to this problem. Additionally, embedding distortion constraints are typically considered. Concerning this last point, although at first sight it could seem to be obvious, note that given that the target is just the estimation of the original host signal at the decoder, the watermarked signal may be arbitrarily far from the original host. Indeed, methods where the distortion embedding constraint is not considered due to this fact have been proposed [28].

A typical block-diagram of reversible watermarking illustrating some of the characteristics mentioned above can be found in Fig. 3.

7 Conclusions

Similarities and differences between multimedia security problems and Witsenhausen's counterexample have been revealed. Although the definitions and application domains of these problems are intrinsically different, the similarities we have spotted (especially for authentication and watermark detection) explain why similar solutions had been independently proposed for both problems.

One of the main conclusions one can derive from this comparison is the fact that dirty paper coding is suitable for reducing the host interference (or state interference in control) in a range of scenarios much wider than the one initially proposed by Costa. This fact seems to encourage the use of dirty paper based techniques for multimedia security applications where it has not been used so far, as for example reversible watermarking, robust hashing or active forensics. Although it could well be the case that DPC were not optimal in those scenarios, as it is the case in Witsenhausen's counterexample, if the gain with respect to conventional approaches were as large as for Witsenhausen's problem, DPC would be worth considering.

References

1. Alattar, A.M.: Reversible watermark using the difference expansion of a generalized integer transform. IEEE Transactions on Image Processing 13(8), 1142–1156 (August 2004)

2. Baglietto, M., Parisini, T., Zoppoli, R.: Numerical solutions to the witsenhausen counterexample by approximating networks. IEEE Transactions on Automatic Control 46(9), 1471–1477 (September 2001)
3. Bansal, R., Basar, T.: Stochastic teams with nonclassical information revisited: When is an affine law optimal. IEEE Transactions on Automatic Control 32(6), 554–559 (June 1987)
4. Barni, M., Bartolini, F., De Rosa, A., Piva, A.: Optimum decoding and detection of multiplicative watermarks. IEEE Transactions on Signal Processing 51(4), 1118–1123 (April 2003)
5. Boncelet, C.G.: The ntmac for authentication of noisy messages. IEEE Transactions on Information Forensics and Security 1(1), 35–42 (March 2006)
6. Cannons, J., Moulin, P.: Design and statistical analysis of a hash-aided image watermarking system. IEEE Transactions on Image Processing 13(10), 1393–1408 (October 2004)
7. Celik, M.U., Sharma, G., Tekalp, A.M., Saber, E.: Reversible data hiding. In: Proceeding of the IEEE International Conference on Image Processing. vol. 2, pp. 157–160 (December 2002)
8. Chen, B., Wornell, G.W.: Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. IEEE Transactions on Information Theory 47(4), 1423–1443 (May 2001)
9. Costa, M.H.M.: Writing on dirty paper. IEEE Transactions on Information Theory 29(3), 439–441 (May 1983)
10. Cox, I.J., Kilian, J., Leighton, F.T., Shamoon, T.: Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing 6(12), 1673–1687 (December 1997)
11. Deng, M., Ho, Y.C.: An ordinal optimization approach to optimal control problems. Automatica 35(2), 331–338 (1999)
12. Dorato, P., Abdallah, C., Cerone, V.: Linear Quadratic Control: An Introduction. Krieger Publishing Company, Malabar, FL (2000)
13. Eggers, J.J., Bäuml, R., Tzschoppe, R., Girod, B.: Scalar Costa Scheme for information embedding. IEEE Transactions on Signal Processing 51(4), 1003–1019 (April 2003)
14. Erez, U., Listsyn, S., Zamir, R.: Lattices which are good for (almost) everything. IEEE Transactions on Information Theory 51(10), 3401–3416 (October 2005)
15. Erez, U., Zamir, R.: Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding. IEEE Transactions on Information Theory 50(10), 2293–2314 (October 2004)
16. Fei, C., Kundur, D., Kwong, R.H.: Analysis and design of secure watermark-based authentication systems. IEEE Transactions on Information Forensics and Security 1(1), 43–55 (March 2006)
17. Fridrich, J., Goljan, M., Du, R.: Lossless data embedding – new paradigm in digital watermarking. EURASIP Journal on Applied Signal Processing (2), 185–196 (2002)
18. Ge, R., Arce, G.R., Di Crescenzo, G.: Approximate message authentication codes for n-ary alphabets. IEEE Transactions on Information Forensics and Security 1(1), 56–67 (March 2006)
19. Grover, P., Park, S.Y., Sahai, A.: The finite-dimensional witsenhausen counterexample. IEEE Transactions on Automatic Control Submitted
20. Grover, P., Sahai, A.: A vector version of witsenhausen's counterexample: A convergence of control, communication and computation. In: Proceedings of the 47th IEEE Conference on Decision and Control (CDC). pp. 1636–1641 (December 2008)

21. Grover, P., Sahai, A.: Witsenhausen's counterexample as assisted interference supression. International Journal on Systems, Control and Communications 2(1/2/3), 197–237 (2010)
22. Grover, P., Wagner, A., Sahai, A.: Information embedding meets distributed control. IEEE Transactions on Information Theory (2010), submitted
23. Kalker, T., Willems, F.: Capacity bounds and constructions for reversible datahiding. In: Proceedings of the IEEE International Conference on Digital Signal Processing. vol. 1, pp. 71–76 (April 2002)
24. Kamstra, L., Heijmans, H.J.A.M.: Reversible data embedding into images using wavelet techniques and sorting. IEEE Transactions on Image Processing 14(12), 2082–2090 (December 2005)
25. Lee, J.T., Lau, E., Ho, Y.C.: The witsenhausen counterexample: A hierarchical search approach for nonconvex optimization problems. IEEE Transactions on Automatic Control 46(3), 382–397 (March 2001)
26. Li, N., Marden, J.R., Shamma, J.S.: Learning approaches to the witsenhausen counterexample from a view of potential games. In: Proceedings of the 48th IEEE Conference on Decision and Control. pp. 157–162 (December 2009)
27. Liu, T., Moulin, P.: Error exponents for one-bit watermarking. In: IEEE International Conference on Acoustics, Speech, and Signal Processing. vol. 3, pp. 65–68 (April 2003)
28. Macq, B.: Lossless multiresolution transform for image authenticating watermarking. In: Proceedings of EUSIPCO. pp. 533–536 (September 2002)
29. Malvar, H.S., Florencio, D.A.F.: Improved spread spectrum: A new modulation technique for robust watermarking. IEEE Transactions on Signal Processing 51(4), 898–905 (April 2003)
30. Marden, J.R., Arslan, G., Shamma, J.: Joint strategy fictitious play with inertia for potential games. IEEE Transactions on Automatic Control 54(2), 208–220 (February 2009)
31. Martinian, E., Wornell, G.W., Chen, B.: Authentication with distortion criteria. IEEE Transactions on Information Theory 51(7), 2523–2542 (July 2005)
32. Merhav, N., Shamai, S.: Information rates subject to state masking. IEEE Transactions on Information Theory 53(6), 2254–2261 (June 2007)
33. Papadimitriou, C.H., Tsitsiklis, J.: Intractable problems in control theory. SIAM J. Control and Optimization 24(2), 639–654 (July 1986)
34. Pérez-Freire, L., Comesaña, P., Pérez-González, F.: Detection in quantization-based watermarking: Performance and security issues. In: Delp III, E.J., Wong, P.W. (eds.) Proceedings of SPIE. Security, Steganography, and Watermarking of Multimedia Contents VII, vol. 5681, pp. 721–733. SPIE, San Jose, CA, USA (January 2005)
35. Willems, F., Kalker, T.: Reversible embedding methods. In: Proceedings of the 40thAllerton Conference on Communications, Control and Computing. pp. 1462–1471 (2002)
36. Witsenhausen, H.S.: A counterexample in stochastic optimum control. SIAM J. Control 6(1), 131–147 (1968)
37. Xie, L., Arce, G.R., Graveman, R.F.: Approximate image message authentication codes. IEEE Transactions on Multimedia 3(2), 242–252 (June 2001)