An Efficient Method for Content Reconstruction with Self-Embedding

Paweł Korus, *Student Member, IEEE,* Andrzej Dziech

***Abstract*—This paper presents a new model of the content reconstruction problem in self-embedding systems, based on an erasure communication channel. We explain why such a model is a good fit for this problem, and how it can be practically implemented with the use of digital fountain codes. The proposed method is based on an alternative approach to spreading the reference information over the whole image, which has recently been shown to be of critical importance in the application at hand. Our study presents a theoretical analysis of the inherent restoration trade-offs. We analytically derive formulas for the reconstruction success bounds, and validate them experimentally with Monte Carlo simulations and a reference image authentication system. We perform an exhaustive reconstruction quality assessment, where the presented reference scheme is compared to 5 state-of-the-art alternatives in a common evaluation scenario. Our study leads to important insights on how self-embedding schemes should be constructed to achieve optimal performance. The reference authentication system designed according to the presented principles allows for high-quality reconstruction, regardless of the amount of the tampered content. The average reconstruction quality, measured on 10,000 natural images is 37 dB, and is achievable even when 50% of the image area becomes tampered.**

***Index Terms*—Image & Video Processing for Watermarking and Security**

I. Introduction

The capability of reconstructing the original content is one of the most compelling features of digital image authentication schemes. In addition to the content hashes for authentication purposes, an encoder embeds in the image a *reconstruction reference*, which describes the content, and which can be used by a dedicated decoder to restore the tampered image fragments. Hence the term *self-embedding*, coined in the original publication [1].

In the most common approach, the reference information is a reduced-quality version of the original image. Alternatives also exist, where the reconstruction reference no longer has a direct interpretation of an image. In [2], it is constructed from the redundancy provided by traditional error correction codes. In [3], content reconstruction is modeled as an irregular sampling problem, and projections onto convex sets are used for restoration. The reference data is obtained by logical exclusive disjunction on coefficient polarity information and pseudo-random bit sequences. A formulation of the content reconstruction problem in terms of compressive sensing has recently been proposed in [4].

A number of content reconstruction schemes have been proposed so far with various applications in mind, ranging from content authentication [1], [3], [5], [6], through error concealment [7], [8], [9], to privacy protection [10], [11]. Regardless of the application at hand, the reference information needs to be communicated to the decoder through an unreliable channel, i.e., the tampered digital image.

In many schemes, the reference information regarding a particular image block (i^{th}) is simply embedded into a different block (j^{th}), often chosen pseudo-randomly. As a result, the i^{th} block can be recovered only if j^{th} is still authentic. This problem, referred to as the *reconstruction dependency*, cripples even recent schemes, e.g., [12] where the reference information is duplicated and the decoder has simply two chances for successful reconstruction.

In fact, the decoder needs not to recover the whole reference stream, but only the necessary fragments corresponding to the tampered image regions. Since it is not possible to determine *a priori* which regions will be tampered, the remaining portion of the reference stream eventually turns out as unnecessary and contributes to the waste of the watermark's capacity. This problem is referred to as the *reference waste*.

Both problems can be solved with proper scheme design. As recently shown, the reference waste can be mitigated by reusing authentic image content, and the reconstruction dependencies can be eliminated by distributing the reference information over the image [13], [4]. In these schemes, the image content is randomly divided into groups, and a random linear projection of each individual group is performed to obtain the reference data. The resulting bits are then pseudo-randomly scattered over the image. The reconstruction is possible if the number of tampered elements within each group is below a certain threshold, determined with the use of the binomial distribution. When the threshold is exceeded, approximation techniques may be employed. Such schemes, where the reconstruction fidelity deteriorates with the tampering rate, are referred to as *flexible*.

Despite the variety of available research, there is no general model of the content reconstruction problem. If an existing scheme needs to be adapted to different requirements, the available experimental results no longer apply, and it is diffi-

P. Korus, A. Dziech are with the Department of Telecommunications of the AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland, fax: +48 12 6342372. Telephone: +48 12 6173805 (P. Korus), +48 12 6172616 (A. Dziech), E-mail: pkorus@agh.edu.pl, dziech@kt.agh.edu.pl

Supplementary multimedia materials, including example reconstruction results and Matlab source code are available at http://ieeexplore.ieee.org/.

The research leading to these results has received funding from the INDECT project funded by European Community's Seventh Framework Programme (FP7 / 2007-2013) under grant agreement no.[218086].

cult to estimate the impact of the prospective modifications. It is also not clear, how efficiently it is possible to trade-off various system parameters. This phenomenon is clearly visible by considering the trade-off between the image quality and the restoration conditions. The former is expressed by means of objectively measured distortion of the original image, e.g., using the peak signal to noise ratio (PSNR). The latter is usually the maximum tampering rate, i.e., area of modifications for which the reconstruction is still possible.

On the one hand, the scheme from [5] allows for lossless reconstruction with bit-wise accuracy. It is achieved by combining a full-quality reconstruction reference with difference expansion for its embedding. While the resulting stego-image is significantly distorted, it can be used to perfectly restore the original content, both in the watermarked and the tampered regions. The cost for this superior quality is a limited tampering rate of 3.2%. On the other hand, [14] proposes a scheme which withstands a tampering rate of 59%, but provides only low-quality restoration. The behavior of this trade-off has not received sufficient attention yet. The question of how efficiently it is possible to exchange the tampering rate for the reconstruction quality still remains unanswered.

In this study, we present a new model of the content reconstruction problem, which allows for formal analysis of the inherent restoration trade-offs. It involves an alternative way of spreading the reference information over the whole image. Instead of using a combined approach with locally-applied random projection and globally-applied scattering, we adopt a spreading mechanism which uses the whole image in the process. Unlike in scattering, each image block uniformly contributes to the entire embedded watermark. As a result, every pixel carries information about the whole image. With the same rate of reference information per image block, the proposed approach allows for working with higher tampering rates than other self-embedding schemes.

We theoretically analyze the problem based on the foundations from communication theory. We use the erasure communication channel as a starting point, and adapt it to take into account individual properties of the problem at hand. Due to the intrinsic availability of the tampering localization features in image authentication systems, the erasure channel is a natural model of the self-embedding problem. We derive formulas for the reconstruction success bounds, which allow for analytical assessment of the restoration performance. The presented theoretical results are verified experimentally by Monte Carlo simulations using a reference image authentication scheme. Our analysis leads to general guidelines for the construction of efficient self-embedding systems.

We also present an exhaustive performance evaluation on a representative test set of natural images. Reconstruction quality assessment reported in existing studies is often based on a small set of commonly known images, without considering the influence of the tampering pattern. For the sake of fair comparison with other schemes, we perform the assessment in a common evaluation framework. The presented reference scheme delivers superior performance compared to state-of-the-art alternatives. Our scheme is the first to allow for high quality reconstruction under extensive tampering. With up to 50% of the tampered image area, it allows for reconstruction with the average PSNR of 37 dB.

The remaining part of this paper is organized as follows. Section II summarizes the most recent research addressing the quality-related trade-offs. A formal problem statement and a derivation of the reconstruction success bounds are presented in Section III. The experimental evaluation scenario and the details of the utilized reference image authentication system are described in Section IV. We conclude in Section V. A detailed comparison with the alternative spreading approach from [13] follows in Appendix A.

II. Related Work

The most recent research on self-embedding focuses on flexible and adaptive schemes. In the flexible system from [15], 5 most significant bits of two randomly selected pixels are combined by exclusive disjunction, then grouped with other pixel pairs and embedded into 3 least significant bit-planes of randomly selected image blocks. Depending on the authenticity of individual pixels, the combined 5 bits of a pixel pair can be either fully or partially recovered. The remaining uncertainty is resolved by exploiting local pixel correlations.

In [16], the authors propose to use a dithered binary version of the image as a reference, and perform the reconstruction by inverse halftoning of the recovered watermark. As a result of image modifications, the watermark gradually becomes a random noise and the restoration result quickly ceases to be discernible. Feasible tampering rates are not reported.

The scheme proposed in [4] groups scattered image blocks, and performs linear projection of their discrete cosine transform (DCT) coefficients using a pseudo-random Gaussian matrix. Hence, it spreads the reference information within a single block group. The complete reference bit-stream is then scattered over the whole image. During content reconstruction, the necessary coefficients within each individual group are recovered either using compositive reconstruction or by compressive sensing, depending on whether the resulting problem is over- or under-determined.

In [13], the authors describe two schemes, with constant-fidelity and with flexible restoration. In the former, the reference information is obtained by randomly ordered 5 most significant bit-planes. The latter uses a pyramidal decomposition of the image blocks, and defines a three-part scalable reference stream. In both schemes, the stream is further divided into subsets, which are locally spread using pseudo-random binary matrices. Concatenated reference information from all of the subsets is then scattered over the whole image. In the flexible variant, the reconstruction quality exhibits three possible levels, depending on the extractable parts of the reference stream.

The reconstruction adaptivity presented in [6] stems from an additional quality descriptor which defines several fidelity levels. By controlling the reconstruction quality for each image block individually, it is possible to bias the scheme either towards better quality or better tampering rate. The resulting varying-length reference is then encoded into a constant-length payload in order to exploit the maximum capacity of the watermarking scheme. The utilized LT code [17] spreads

the information about each image fragment over the image. This scheme, however, fails to exploit the remaining authentic content and is limited by traditional bounds of a general erasure channel.

Content adaptivity has also been demonstrated in [18], [19]. An individual image block is represented with one of multiple defined reference rates, depending on the amount of texture. The primary objective is to improve the reconstruction quality with simultaneous reduction of the reference payload. In [18], one of the defined rates is null, and its corresponding blocks are recovered with the use of inpainting.

III. Formal Problem Statement

This section describes the considered content reconstruction problem in terms of an erasure communication channel. First, we introduce a generic reconstruction framework and explain why this communication model is a good fit for the problem at hand. We emphasize the differences between the communication process in a general and in the considered scenario, and derive formulas for the reconstruction success bounds. The resulting model is referred to as the *self-recovery* model.

The operation of many image authentication and reconstruction systems can be summarized in terms of a generic framework, regardless of the assumed formulation of the content reconstruction problem. There are three fundamental properties which differentiate such systems: the reference generation and reconstruction method, the payload encoding method and the data embedding scheme. As it will be demonstrated later, all of the three aspects need to be designed properly.

Let I denote the original, unprotected image and $I_i : i = 1, \ldots, N$ the i^{th} image block in the raster scan order. Let $g_b(I_i)$ denote a reconstruction reference generation function for a single image block which generates exactly b bits of reference information. In general, $b = b(i)$, but in this study we consider $b = const$. Thus, $g_b(\cdot)$ generates a complete $Nb-$bit reconstruction reference:

$$\mathbf{r} = r_1, \ldots, r_N = g_b(I_1), \ldots, g_b(I_N).$$

An inverse function $g_b^{-1}(\cdot)$ restores the image block from the reference bit-stream. We denote the restored image block as $I'_i = g_b^{-1}(r_i)$.

Let $h(\cdot)$ denote a hashing function, which generates a cryptographic hash from the image block content I_i, the block payload Y_i, the block number i and a secret key k. The number of hash bits produced is denoted as $|h|$. For brevity, let us also denote the resulting hash for i^{th} image block as:

$$h_i = h(I_i, Y_i, i, k).$$

Let also

$$f(I_i, Y_i, h_i) \rightarrow \hat{I}_i$$

denote the embedding function, and

$$f^{-1}(\hat{I}_i) \rightarrow Y_i, h_i$$

the blind watermark recovery function. Both the watermark embedding and recovery functions are key controlled, e.g., by pseudo-random order of the selection channel. We omit this obvious dependency for the sake of notation simplicity. The capacity of the watermarking scheme is $B+|h|$ bits per block.

Algorithm 1 A generic image authentication and reconstruction scheme.

Require: I
Require: $h(), f(), g(), f^{-1}(), g^{-1}()$
Require: $b, B : b \leq B$
  for $i = 1 \rightarrow N$ **do**
    $r_i \leftarrow g_b(I_i)$
  end for
  $\mathbf{r} \leftarrow [r_1, \ldots, r_N]$
  Encode $\mathbf{r}$ *to obtain watermark payload* $Y_i : i = 1, \ldots, N$
  for $i = 1 \rightarrow N$ **do**
    $h_i \leftarrow h(I_i, Y_i, i, k)$
    $\hat{I}_i \leftarrow f(I_i, Y_i + h_i)$
  end for
  Tamper selected image blocks : $\hat{I}_i \rightarrow \tilde{I}_i$
  for $i = 1 \rightarrow N$ **do**
    $\tilde{Y}_i, \tilde{h}_i \leftarrow f^{-1}(\tilde{I}_i)$
    $\hat{h}_i \leftarrow h(\tilde{I}_i, \tilde{Y}_i, i, k)$ Generate the tampering map :
    $e_i \leftarrow \tilde{h}_i = \hat{h}_i$
  end for
  Discard $\tilde{Y}_i : e_i \neq 1$
  Regenerate $r_i : e_i = 1$
  Remove the dependencies on $r_i : e_i = 1$ *from* $\tilde{Y}_i$
  Recover $\mathbf{r}$ from remaining $\tilde{Y}_i = Y_i$ *// Only* $r_i : e_i \neq 1$
  Reconstruct $I'_i : e_i \neq 1$

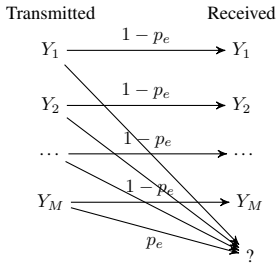


Fig. 1: M-ary erasure channel with probability of erasure p_e.

The operation of a generic image authentication scheme is described in Algorithm 1. In earlier schemes, certain operations from this framework are either skipped completely or trivialized. For instance, often the payload encoding phase consists of simple permutations of the reference stream blocks, leading to reconstruction dependencies.

Communication of the reconstruction reference to the decoder clearly resembles the erasure channel (Fig. 1). Each image block carries a single symbol of the watermark payload. Since the erasure (tampering) localization information is intrinsically available after content authentication, the decoder sees the transmitted symbols either as correctly transmitted or erased, for authentic and tampered blocks, respectively. Since the payload is authenticated along with the image content, the transition probabilities between symbols are negligible.

The best known family of codes for the erasure channel are the digital fountain codes [20]. Given a stream of input symbols, they produce a potentially limitless stream of same-length output symbols. The output symbols are computed by bit-wise exclusive disjunction on randomly selected input symbols. Hence, each output symbol carries information about the whole image. The transmitted message can be successfully decoded from arbitrary portions of the resulting stream, pro-

vided that sufficient amount of correct symbols is collected. In this study, we use the random linear fountain (RLF) code.

In the proposed approach, the reconstruction reference $\mathbf{r}$ consists of N b-bit *reference blocks*. The stream is then divided into constant length B-bit *reference symbols* $X_k : k = 1, \ldots, K$ which are then encoded with the RLF to produce N *embedding symbols* Y_i. In case an image block is tampered, the decoder marks the corresponding embedding symbol as erased and continues with reference data decoding. The code rate, denoted as $\lambda = \frac{K}{N}$, reflects the rate of the effective payload with respect to the available watermark capacity. In this study, we focus on the case of $K \leq N$ ($\lambda \leq 1$).

The key insight to distinguish the cases of general communication and self-recovery communication over the erasure channel is the success criterion. For the reconstruction process to succeed, the decoder only needs selected fragments of the transmitted message which correspond to the factually needed reference blocks. The remaining portion of the reference stream should be recoverable from the authentic image fragments. This leads to a concise definition of the communication process.

Definition 1. *The self-recovery communication problem is a special case of an erasure channel which:*

1) *uses the media content as a communication channel,*
2) *carries the message which describes the media content itself, i.e., the reconstruction reference,*
3) *aims to recover only the reference data of the altered (erased) media fragments,*
4) *allows to recover an identical reconstruction reference from image fragments with embedded payload, provided that they are authentic.*

The latter condition means that the reconstruction reference generation algorithm must be invariant to the utilized embedding scheme, i.e., $g_b(I_i) = g_b(\hat{I}_i)$. Upon computing the tampering map, the decoder can recalculate the reference blocks for authentic image areas and remove these dependencies from the correctly recovered symbols. Hence, the decoder can make the reconstruction reference forget what it knows about the appearance of authentic blocks and reduce the problem to tampered blocks only. The selective decoding capability implies the necessity for random access to the reference stream. In this study, this property is guaranteed by using $b = const$. An alternative approach would be to use a reconstruction quality descriptor [6].

Fig. 2 shows the operation of an image authentication system based on the proposed approach. In addition to the information processed in the successive steps of the algorithm it also demonstrates the impact of the misalignment between the reference blocks and symbols. A single erased reference block might invalidate two reference symbols and, thus, limit the prospective reduction of the decoding problem. This effect can be minimized or even eliminated by proper choice of b and B. This issue will be discussed in detail later on.

The reduction of the reference decoding problem can also be explained by examining a matrix representation of digital fountain codes. The generator matrix:

$$\mathbf{G}_{N\times K} = [G_{i,k}] : i = 1, \ldots N; k = 1, \ldots, K; G_{i,k} \in \{0, 1\}$$

fully represents the code by assigning value 1 to elements i, k iff the k^{th} reference symbol X_k is included in the i^{th} embedding symbol Y_i. An example random code for $K = 7$ and $N = 9$ is shown in Fig. 3a. The rank of the matrix is 7 and all of the reference symbols can be decoded from this set of embedding symbols.

Consider now a tampering pattern from Fig. 2 where Y_1 and Y_6 are erased. The resulting code is represented by a matrix (Fig. 3b) of rank 6 and it is no longer possible to successfully decode all X_j symbols. We can, however, regenerate the reconstruction reference of known blocks, restore the reference symbols $\{X_2, X_3, X_6, X_7\}$ and remove these dependencies from the original Y_i to produce Y_i' (Fig. 3c). The rank of the resulting matrix is 3 and it is possible to recover the remaining 3 unknown reference symbols $\{X_1, X_4, X_5\}$ (Fig. 3d).

The bound on the allowed tampering rate can be calculated analytically. Let M denote the number of authentic image blocks, and automatically the embedding symbols Y_i. Then $\gamma = \frac{M}{N}$ denotes the block survival rate and $\tilde{\gamma} = 1 - \gamma$ the tampering rate. Given the probability of decoding error δ, the bound on the reconstruction success for a general erasure channel and a random linear fountain code is [20]:

$$\begin{aligned} M &\geq K + \epsilon(\delta), \\ \gamma &\geq \lambda + \frac{\epsilon(\delta)}{N}, \end{aligned} \tag{1}$$

where $\epsilon(\delta)$ represents the overhead or imperfectness of the code. For an ideal code, the decoder would always be capable of successful decoding if the number of received symbols is equal to the number of input symbols, i.e., $\epsilon(\delta) \equiv 0$. For the RLF, the overhead is bounded by [20], [21]:

$$\epsilon(\delta) \leq \log_2 \frac{1}{\delta}.$$

In practice, for long messages (large images) $\frac{\epsilon}{N} \approx 0$. Hence, in the succeeding derivations this term will be disregarded. The equality in (1) defines the bound on the maximal allowed tampering rate, referred to as the γ_1 bound.

By exploiting the described properties of the self-recovery communication problem, the number of necessary reference symbols becomes reduced:

$$\begin{aligned} \gamma &\geq \lambda\rho(\lambda, \gamma), \\ \rho(\lambda, \gamma) &: \mathbf{R}^+ \times [0, 1] \rightarrow [0, 1]. \end{aligned} \tag{2}$$

$\rho(\lambda, \gamma)$ is the *reconstruction demand*, i.e., the expected value of the fraction of reference symbols X_i which need to be decoded from the remaining embedding symbols $\{\hat{Y}_i : e_i = 1\}$ for a given tampering rate $\tilde{\gamma}$.

A. Upper Bound on Maximal Tampering Rate

The best possible reconstruction capability is achieved when the tampering in a new reference block r_i yields minimal impact on the reference symbols X_j. Such a situation occurs when the reference blocks and symbols are perfectly aligned

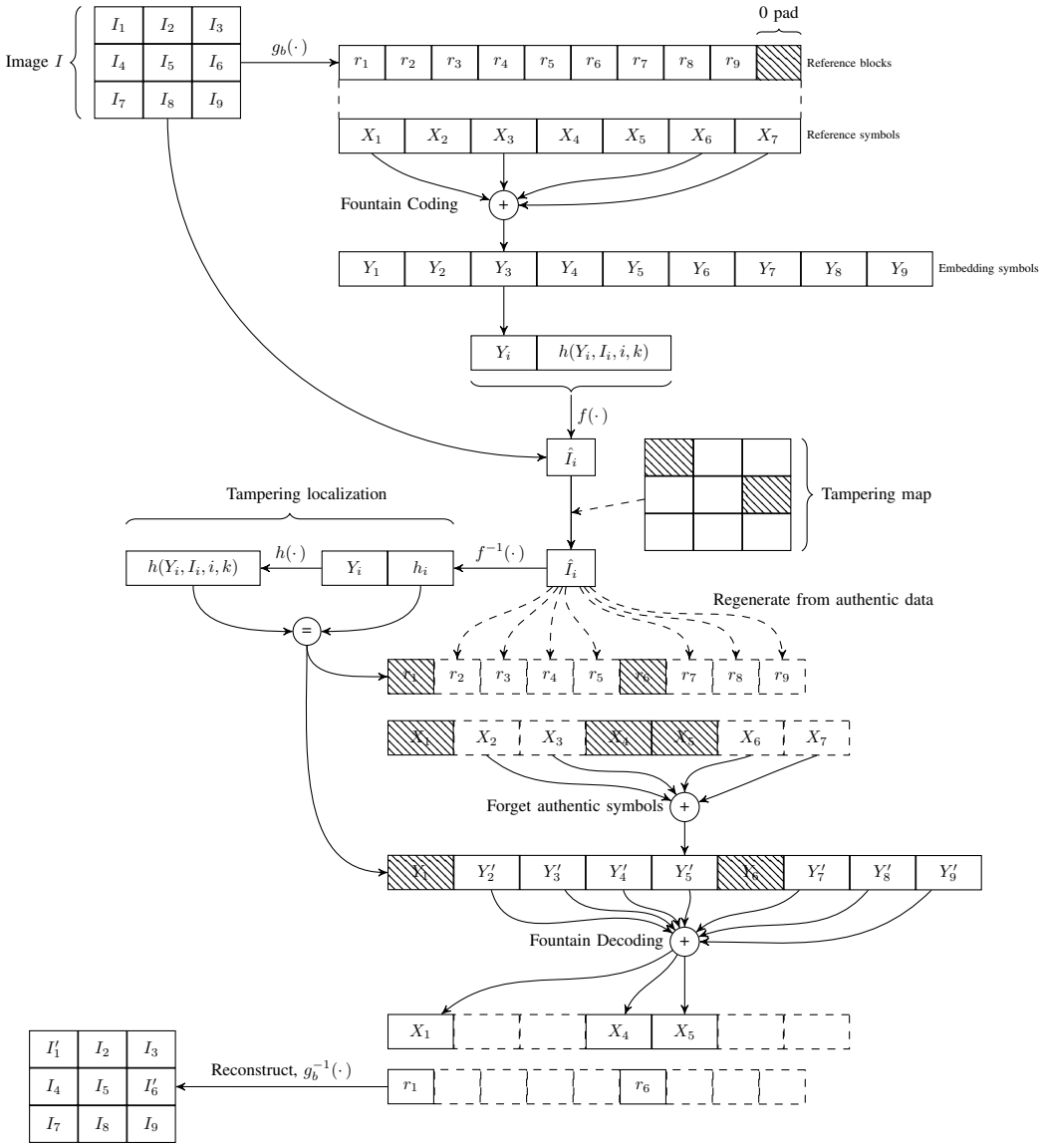


Fig. 2: Operation of the reference image authentication system. It illustrates the reduction of the decoding problem and the impact of prospective misalignment between the division of the reconstruction reference into reference blocks and symbols.

or when the erasure pattern is continuous over r_i, e.g., when image blocks $i = 1, \ldots, N - M$ are tampered. Then:

$$\rho(\lambda, \gamma) = (1 - \frac{M}{N}) = (1 - \gamma) = \tilde{\gamma}.$$

Hence, from (2):

$$\gamma \geq \lambda(1 - \frac{M}{N}) = \lambda(1 - \gamma),$$

$$\gamma \geq \frac{\lambda}{\lambda + 1}. \tag{3}$$

The equality in (3) defines the optimistic reconstruction success bound, referred to as the γ_2 bound. What is important, is that this optimistic bound is relevant not only for continuous erasure patterns. It can also be reached for $\lambda \in \mathbf{N}$. This phenomenon will be explained in more detail later on.

B. Typical Bound on the Maximal Tampering Rate

In general, linear growth of the reconstruction demand with the tampering rate cannot be safely assumed. The tampering locations can be distributed over the whole image and the misalignment between the boundaries of reference blocks and symbols will cause a more precipitous increase of the reconstruction demand. An example such situation has already been shown in Fig. 2 where the damage of one image block I_6 implicates the necessity to decode two reference symbols X_4 and X_5.

In this section, we derive the reconstruction success bound for a random tampering pattern, where the image blocks for modification are chosen randomly. The main factor that influences the reconstruction demand is the misalignment between the reference blocks and symbols. Three fundamental cases can be distinguished:

1) $hcf(b, B) = b \Leftrightarrow \frac{1}{\lambda} \in \mathbf{N}$,

(a) Full reconstruction problem

rank 7	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Y_1	1	0	0	1	0	0	1
Y_2	0	1	1	1	0	0	1
Y_3	1	0	1	0	0	1	1
Y_4	1	1	0	0	1	0	1
Y_5	0	0	0	1	0	1	1
Y_6	1	1	1	0	0	0	1
Y_7	0	0	0	1	1	0	1
Y_8	0	1	0	1	1	0	1
Y_9	1	0	1	0	0	1	1

(b) Erased Y_1 and Y_6

rank 6	$\overline{X_1}$	X_2	X_3	$\overline{X_4}$	$\overline{X_5}$	X_6	X_7
Y_2	0	1	1	1	0	0	1
Y_3	1	0	1	0	0	1	1
Y_4	1	1	0	0	1	0	1
Y_5	0	0	0	1	0	1	1
Y_7	0	0	0	1	1	0	1
Y_8	0	1	0	1	1	0	1
Y_9	1	0	1	0	0	1	1

(c) Removed dependencies on known X_i

rank 3	$\overline{X_1}$	$\overline{X_4}$	$\overline{X_5}$
$Y_2' = Y_2 \oplus X_2 \oplus X_3 \oplus X_7$	0	1	0
$Y_3' = Y_3 \oplus X_3 \oplus X_6 \oplus X_7$	1	0	0
$Y_4' = Y_4 \oplus X_2 \oplus X_7$	1	0	1
$Y_5' = Y_5 \oplus X_6 \oplus X_7$	0	1	0
$Y_7' = Y_7 \oplus X_7$	0	1	1
$Y_8' = Y_8 \oplus X_2 \oplus X_7$	0	1	1
$Y_9' = Y_9 \oplus X_3 \oplus X_6 \oplus X_7$	1	0	0

(d) Solved reconstruction problem

rank 3	$\overline{X_1}$	$\overline{X_4}$	$\overline{X_5}$
$Y_3' = Y_3 \oplus X_3 \oplus X_6 \oplus X_7$	1	0	0
$Y_2' = Y_2 \oplus X_2 \oplus X_3 \oplus X_7$	0	1	0
$Y_2' \oplus Y_7' = Y_7 \oplus X_7 \oplus Y_2 \oplus X_2 \oplus X_3$	0	0	1

Fig. 3: Decoding problem reduction in the matrix representation of the random linear fountain code: the full code generator matrix $\mathbf{G}$ allows to recover all reconstruction symbols (a), after erasing two symbols, the whole message cannot be decoded (b) but after eliminating the dependencies on known symbols (c), the remaining necessary symbols can be easily decoded (d).

2) $1 < hcf(b, B) < b$,
3) $hcf(b, B) = 1$.

$hcf(\cdot)$ is the highest common factor. In the first of the identified cases, B is a multiple of b and the probability of invalidating a reference symbol can be calculated in a straightforward manner. Since each reference symbols is overlapped by exactly $\frac{1}{\lambda}$ reference blocks, it will need to be decoded provided that any of the corresponding reference blocks is required. Hence:

$$\rho(\gamma, \lambda) = 1 - \gamma^{\frac{1}{\lambda}}.$$

We immediately see that if $\lambda = 1$ ($b = B$) the reconstruction bound is identical to the previously calculated optimistic bound γ_2.

We do not distinguish between the cases when $\frac{1}{\lambda} \notin \mathbf{N}$. In general, when $hcf(b, B) < b$, the misalignment between the reference blocks and symbols makes it possible for a missing block to invalidate multiple symbols. We aim to express the reconstruction demand in a convenient form of:

$$\rho(\gamma, \lambda) = 1 - \gamma^{\alpha(\lambda)},$$

with a case dependent function $\alpha(\lambda)$. The necessary theoretical estimate can be obtained by analyzing the overlap between the reference blocks and symbols.

Proposition 1. *The average number of reference blocks overlapping a reference symbol is bounded by* $\frac{1}{\lambda} + 1$.

Proof: Let $b, B \in \mathbf{N} : 0 < b \leq B \ \wedge \ hcf(b, B) = 1$. Division of a bit-stream into b-bit blocks and B-bit symbols is shown in Fig. 4. The misalignment between the starting points of blocks and symbols is denoted as Δ. Depending on the location of the starting point of the next block within a symbol, the number of blocks that overlap a single symbol can assume two values: $\lceil \frac{1}{\lambda} \rceil$ or $\lceil \frac{1}{\lambda} \rceil + 1$. The average number of overlapping reference blocks per reference symbols is a weighted average:

$$(1 - \beta)\lceil \frac{1}{\lambda} \rceil + \beta(\lceil \frac{1}{\lambda} \rceil + 1),$$

stemming from the proportion between the two types of symbols. For any $n \in \mathbf{N}$:

$$Bn \bmod b = 0, 1, \ldots, b - 1,$$

with a uniform distribution of $\{0, 1, \ldots, b-1\}$. If $Bn \bmod b = 0$ or $Bn \bmod b \geq B \bmod b$ there is no room for the extra block and the number of overlapping blocks is $\lceil \frac{1}{\lambda} \rceil$. If $0 < Bn \bmod b < B \bmod b$, the additional block will result in a total of $\lceil \frac{1}{\lambda} \rceil + 1$. Thus, the average number of reference blocks overlapping a reference symbol is:

$$(1 - \frac{(B \bmod b) - 1}{b})\lceil \frac{1}{\lambda} \rceil + \frac{(B \bmod b) - 1}{b}(\lceil \frac{1}{\lambda} \rceil + 1).$$

By treating the misalignment of $\Delta = 0$ as if it produced the additional overlapping symbol, we obtain an upper bound of:

$$(1 - \frac{B \bmod b}{b})\lceil \frac{1}{\lambda} \rceil + \frac{B \bmod b}{b}(\lceil \frac{1}{\lambda} \rceil + 1).$$

After simple algebraic expansion:

$$(1 - \frac{B \bmod b}{b})\lceil \frac{1}{\lambda} \rceil + \frac{B \bmod b}{b}(\lceil \frac{1}{\lambda} \rceil + 1) = \frac{B \bmod b}{b} + \lceil \frac{1}{\lambda} \rceil$$

For $x \in \mathbf{R}$, $x = \lfloor x \rfloor + \{x\}$ where $\{\cdot\}$ denotes the fractional part. $\frac{B \bmod b}{b} = \{\frac{1}{\lambda}\}$, and hence:

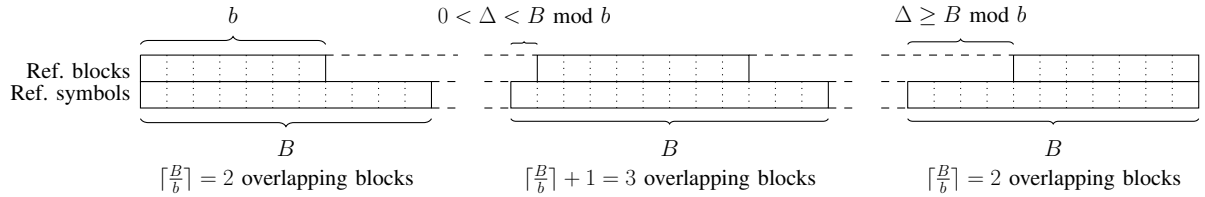


Fig. 4: Impact of the misalignment between the reference blocks and symbols on the number of overlapping blocks per single symbol; Example stream division with $b = 7, B = 11$.

$$\frac{B \bmod b}{b} + \lceil \frac{1}{\lambda} \rceil = \{\frac{1}{\lambda}\} + \lfloor \frac{1}{\lambda} \rfloor + 1 = \frac{1}{\lambda} + 1.$$

□

We will derive a common worst case estimate based on the case of $hcf(b, B) = 1$. We will show that $1 - \gamma^{\frac{1}{\lambda}+1}$ is not only an upper bound on the reconstruction demand but also serves as its good approximation.

The reconstruction demand is also bounded by two weaker bounds:

$$1 - \gamma^{\lceil \frac{1}{\lambda} \rceil} \leq \rho(\gamma, \lambda) \leq 1 - \gamma^{\lceil \frac{1}{\lambda} \rceil + 1},$$

and peaks towards the lower bound whenever $hcf(b, B) > 1$.

Proposition 2. *Given the fraction of necessary reference blocks $\tilde{\gamma}$, the expected fraction of reference symbols that need to be decoded is approximately $1 - \gamma^{\frac{1}{\lambda}+1}$.*

Proof: In the proof of proposition 2 we have shown that the proportion between the reference symbols overlapped by $\lceil \frac{1}{\lambda} \rceil + 1$ and by $\lceil \frac{1}{\lambda} \rceil$ reference blocks is approximately $\{\lambda^{-1}\} : 1 - \{\lambda^{-1}\}$. Thus, the expected value of the fraction of authentic reference symbols:

$$1 - \rho(\gamma, \lambda) \approx \{\lambda^{-1}\}\gamma^{\lceil \lambda^{-1} \rceil + 1} + (1 - \{\lambda^{-1}\})\gamma^{\lceil \lambda^{-1} \rceil}. \quad (4)$$

We now estimate the right hand side as follows:

$$\{\lambda^{-1}\}\gamma^{\lceil \lambda^{-1} \rceil + 1} + (1 - \{\lambda^{-1}\})\gamma^{\lceil \lambda^{-1} \rceil} \approx \gamma^{\frac{1}{\lambda}+1},$$

$$\{\lambda^{-1}\}\gamma\gamma^{\lceil \lambda^{-1} \rceil} + \gamma^{\lceil \lambda^{-1} \rceil} - \{\lambda^{-1}\}\gamma^{\lceil \lambda^{-1} \rceil} \approx \gamma^{\{\lambda^{-1}\} + \lceil \lambda^{-1} \rceil},$$

$$\{\lambda^{-1}\}\gamma^{\lceil \lambda^{-1} \rceil}(\gamma - 1) + \gamma^{\lceil \lambda^{-1} \rceil} \approx \gamma^{\{\lambda^{-1}\}}\gamma^{\lceil \lambda^{-1} \rceil},$$

$$\{\lambda^{-1}\}(\gamma - 1) + 1 \approx \gamma^{\{\lambda^{-1}\}}.$$

Since the only dependency on λ is on $\{\lambda^{-1}\}$, the approximation will be fulfilled for all $\lambda^{-1} \in \mathbf{R}$ if it is fulfilled for $\lambda^{-1} \in [0, 1]$. Hence:

$$\lambda^{-1}(\gamma - 1) \approx \gamma^{\lambda^{-1}} - 1.$$

This approximation stems from the Taylor series expansion of the right hand side around $\gamma = 1$:

$$\lambda^{-1}(\gamma - 1) - \frac{(\gamma - 1)^2(\lambda - 1)}{2\lambda^2} + O((\gamma - 1)^3).$$

□

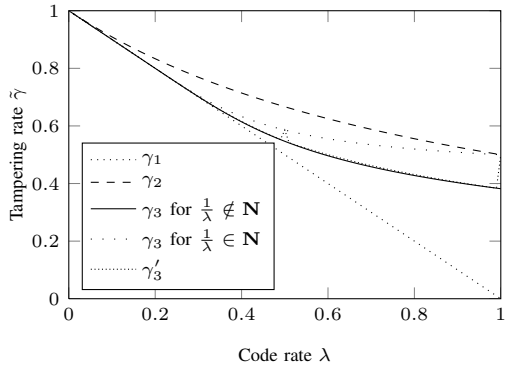


Fig. 5: Derived reconstruction success bounds.

Proposition 3. *The behavior of the reconstruction demand is well represented by the following function:*

$$\rho(\gamma, \lambda) = \begin{cases} 1 - \gamma^{\frac{1}{\lambda}}, & \text{if } \frac{1}{\lambda} \in \mathbf{N}, \\ 1 - \gamma^{\frac{1}{\lambda}+1}, & \text{otherwise}. \end{cases} \quad (5)$$

For $\frac{1}{\lambda} \in \mathbf{N}$ the function (5) is an exact formula for the reconstruction demand. In the remaining case, it serves as a close pessimistic approximation. By substituting (5) to (2) we obtain the definition of the last of the considered reconstruction success bounds, the γ_3 bound:

$$\begin{cases} \gamma \geq \lambda\ (1 - \gamma^{\frac{1}{\lambda}}), & \text{if } \frac{1}{\lambda} \in \mathbf{N}, \\ \gamma \geq \lambda\ (1 - \gamma^{\frac{1}{\lambda}+1}), & \text{otherwise}. \end{cases} \quad (6)$$

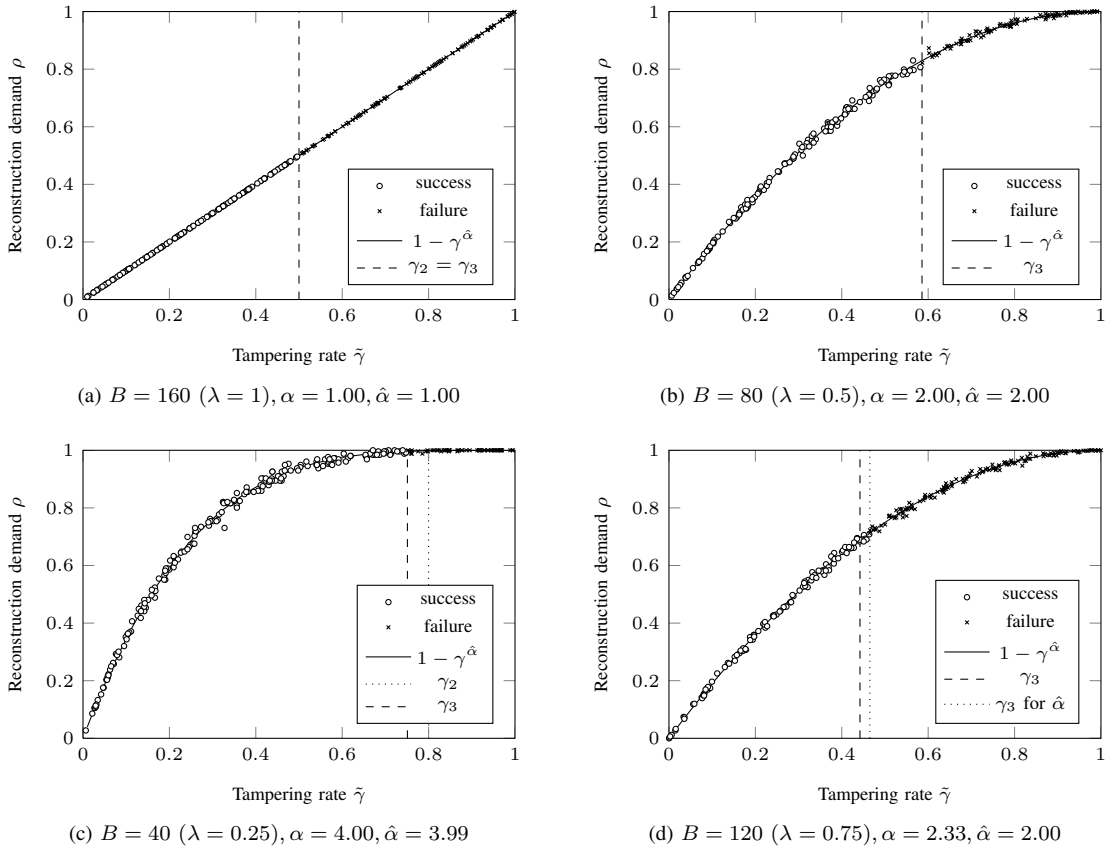
It is not possible to analytically derive a formula for γ_3 and (6) needs to be solved numerically. Fig. 5 shows all of the defined reconstruction success bounds γ_1, γ_2 and γ_3. For reference, the figure also shows a tampering rate bound derived using a more accurate form of $\rho(\gamma, \lambda)$ from (4):

$$\gamma \geq \lambda\ (1 - \{\lambda^{-1}\}\gamma^{\lceil \lambda^{-1} \rceil + 1} - (1 - \{\lambda^{-1}\})\gamma^{\lceil \lambda^{-1} \rceil}).$$

This bound is denoted as the γ_3' bound.

IV. Experimental Evaluation

Experimental validation of the proposed content reconstruction model is divided into two main parts. Firstly, we assess the accuracy of the assumed reconstruction demand estimate (5), and validate the theoretical reconstruction success bounds via Monte Carlo simulations. Secondly, we perform an exhaustive evaluation of the reconstruction quality and compare the results with state-of-the-art self-embedding schemes. The experiments are performed using a reference image authentication system.



(a) $B = 160$ $(\lambda = 1), \alpha = 1.00, \hat{\alpha} = 1.00$

(b) $B = 80$ $(\lambda = 0.5), \alpha = 2.00, \hat{\alpha} = 2.00$

(c) $B = 40$ $(\lambda = 0.25), \alpha = 4.00, \hat{\alpha} = 3.99$

(d) $B = 120$ $(\lambda = 0.75), \alpha = 2.33, \hat{\alpha} = 2.00$

Fig. 6: Experimental and theoretical dependency between the reconstruction demand and the tampering rate for selected reference payloads λ. The theoretical reconstruction success bounds γ_1, γ_2 and γ_3 are shown by means of vertical helper lines.

A. Reference Image Authentication System

The operation of the reference image authentication scheme follows the general algorithm and principles described in Section III. This section merely describes the implementation of the h, f and g_b functions.

The image is divided into non-overlapping 8×8 px blocks. We use bit substitution for data embedding in the 3 least significant bit-planes. Such embedding strategy is used by most of alternative schemes, which facilitates fair comparison of the reconstruction performance. The remaining 5 bit-planes are considered as visually important and are transformed into DCT domain for generating the reconstruction reference. Such construction ensures that the embedded watermark does not interfere with the reconstruction reference generation basis.

The resulting 192 bits of watermark capacity are divided into two parts. 32 bits are used for embedding the h_i hashes obtained by shortening the MD5 hashes by exclusive disjunction on neighboring bit pairs. The remaining $B = 160$ bits are used for embedding the reconstruction reference. Variations of λ are obtained by controlling b.

The reconstruction reference generation function performs quantization of the DCT coefficients. Let

$$S_i(x, y) : x, y \in \{0, \ldots, 7\},$$

denote the coefficient with coordinates (x, y) from the i^{th} image block. A group of coefficients is defined by a constant sum of their coordinates, $x + y = \text{const}$. Hence, there exist 15 groups: $x + y = 0, 1, \ldots, 14$; each quantized with a dedicated code-book. The DC coefficient (group 0) is quantized uniformly, and the remaining groups are quantized with a Lloyd-Max code-book [22]. The precision of the used code-books can be represented by a 15-D allocation vector. During the quality assessment experiment, we use the following allocation vector:

$$[8, 6, 5, 4, 4, 3, 3, 3, 2, 2, 0, 0, 0, 0, 0],$$

which results in a total of $b = 160$ reference bits.

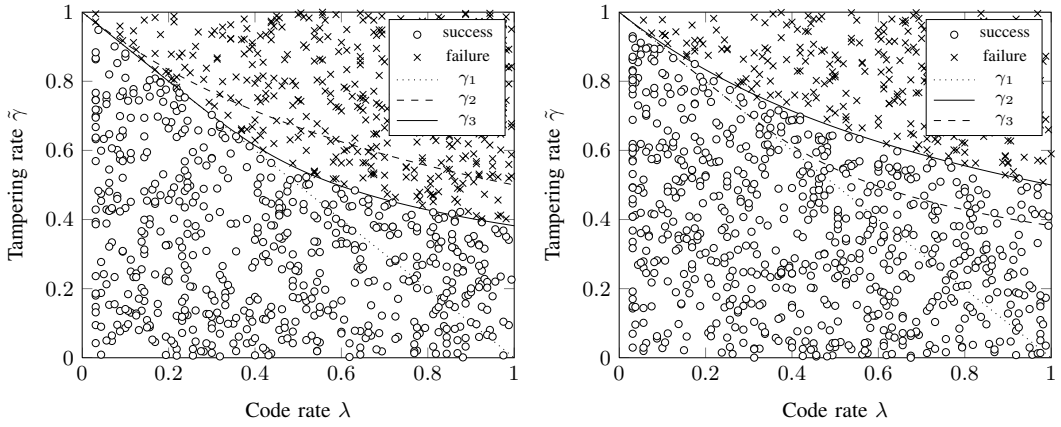
B. Reconstruction Success Bounds

For the purpose of validating the assumed reconstruction demand estimate, we perform a full *content protection* $\rightarrow$ *tampering* $\rightarrow$ *reconstruction* cycle. The process is carried out for $b = 8, 9, \ldots, 160$ with randomly selected tampering rates $\tilde{\gamma}$, drawn from a uniform distribution $U(0, 1)$. During the reconstruction process, the decoder records the observed reconstruction demand value ρ.

This experiment is repeated 250 times for each of the considered values of b, and for each 250-point result set, we perform a fit to the $\phi(\gamma)$ function:

$$\phi(\gamma) = 1 - \gamma^{\alpha}. \tag{7}$$

The shape parameter α is estimated by solving a non-linear least squares problem. Fig. 7 shows the obtained estimates $\hat{\alpha}$ and the corresponding theoretical results. Just as expected,



(a) Random tampering pattern

(b) Continuous tampering pattern

Fig. 8: Experimental validation of the reconstruction success bounds using Monte Carlo simulations. Each sample represents a single reconstruction attempt. The applicable theoretical bound between the successes and failures is shown with a solid line.

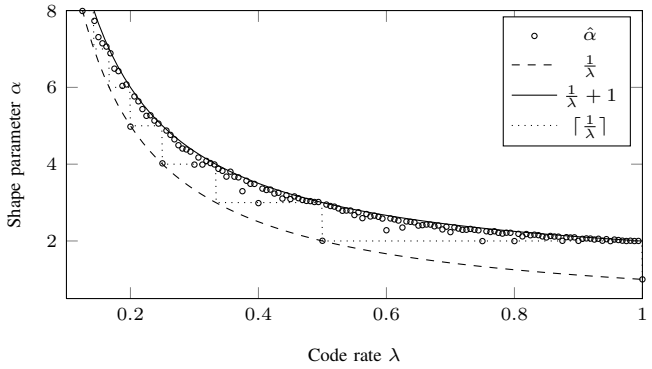


Fig. 7: Comparison of empirical vs. theoretical estimates of the shape parameter α.

the estimates are well approximated by $\frac{1}{\lambda}+1$ with occasional peaks towards the lower bound of $\lceil\frac{1}{\lambda}\rceil$ whenever $hcf(b,B) > 1$. The experiment was performed on 256×256 px images.

Fig. 6 shows scatter-plots of $\rho \times \tilde{\gamma}$ for $\lambda = 1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}$. The plots show the obtained samples, the relevant function fits and vertical helper lines representing the theoretical success bounds. The success-failure distinction is provided by different marks of the samples.

This experiment clearly demonstrates that the derived theoretical results closely approximate the behavior of the reconstruction demand. If $1 < hcf(b,B) < b$, the assumed estimate is potentially least accurate. However, the differences in the tampering rate bounds are marginal, and in favor of the practically achievable bounds. An example case of $\lambda = \frac{3}{4}$, where the difference between the theoretical value of the shape parameter α and its estimate $\hat{\alpha}$ is well visible, is shown in Fig. 6d. For reference, the plot shows also a γ_3 bound obtained from $\hat{\alpha}$. The practically achievable tampering rate bound is 2.3% better compared to the theory.

The obtained results also show the validity of the theoretical reconstruction success bounds. The empirical bound between the reconstruction success and failure coincides with the γ_3 bound. We also observe that for the $\lambda = 1$ case (in general for $\lambda \in \mathbf{N}$), the reconstruction success bound reaches the optimistic bound of γ_2 from (3).

The purpose of the next experiment was to validate all of the introduced bounds: γ_1, γ_2 and γ_3. In each iteration, we encoded the image with a random λ setting, tampered a random fraction $\tilde{\gamma}$ of the available image blocks, and attempted to perform content reconstruction. The evaluation was done on 256×256 px images, and 5,000 iterations were made with different seeds for the pseudo-random number generator.

We consider two tampering patterns: with successive and random blocks tampered. The obtained results are shown in Fig. 8. The success and failure cases are marked with circles and crosses, respectively. The applicable theoretical bound is shown with a solid line. For the sake of presentation clarity, we show here only 1,000 samples and only the results for γ_2 and γ_3. Complete results are provided as supplementary materials.

C. Reconstruction Quality Evaluation

In this experiment, we perform exhaustive evaluation of the reconstruction quality. In addition to the presented reference algorithm, we consider 5 of state-of-the-art self-embedding schemes, both with constant [14], [13] and with flexible reconstruction quality [13], [15], [4]. In order to facilitate fair comparison of the reconstruction performance, we have reimplemented the schemes in a common evaluation framework[1]. The least significant bits are set to 100_2 in all of the schemes. This step is used in [13] to boost the PSNR scores. The watermarking-inflicted distortion is identical for all of the schemes, and equals 37.9 dB.

[1]Due to reference value saturation, stemming from large DC coefficients, the original scheme from [4] exhibits prohibitively poor performance for dark and bright images. The issue has been fixed in our evaluation by adjusting the quantization procedure in (8); $2f_t$ is used instead of f_t. While this operation limits the maximal reconstruction quality to approximately 38 dB, it allows for correct operation on dark and bright images. A detailed description of the problem, and example reconstruction results for both the original and the corrected versions are available in the supplementary materials.

(a) 6882.png (b) 4767.png (c) 9011.png

(d) 131.png (e) 7683.png (f) 4749.png

Fig. 9: Example test images from the BOWS2 data-set.

TABLE I: Mean reconstruction quality on 48 natural images.

Scheme	Reconstruction PSNR [dB] for various $\tilde{\gamma}$						$\tilde{\gamma}_{\max}$
	0.050	0.100	0.200	0.300	0.400	0.475	
[4]	37.2	34.8	31.7	29.1	28.1	27.0	0.60
[15]	37.3	35.6	33.3	31.6	30.2	29.2	0.54
[13]-B	31.8	31.7	31.7	28.7	28.7	25.8	0.66
[14]	28.5	28.4	28.4	28.4	28.4	28.4	0.59
[13]-A	40.7	40.7	40.7	-	-	-	0.24
Proposed	36.4	36.4	36.4	36.4	36.4	36.4	0.50

Our evaluation is performed on a set of 48 images of size 512×512 px, selected from 10,000 gray-scale natural images from the BOWS2 data-set [23]. The selected specimens span the space of possible image characteristics, i.e., include dark, medium and bright content with various amount of details, measured as an average standard deviation of individual 8×8 px blocks. Example images are shown in Fig. 9. The complete evaluation results, together with reconstruction examples are provided as supplementary materials.

The images are encoded with the considered schemes, and then randomly modified with the tampering rates from 0.025 to 0.6 with a 0.025 step. The tampering pattern is the same for all of the schemes. The experiment is repeated 30 times, with different seeds for the pseudo-random number generator.

The average PSNR scores for selected tampering rates are collected in Table I. The results are averaged over 48 images and 30 independent replications of the experiment. The presented reference scheme is capable of high-quality reconstruction, regardless of the tampering rate. While for low tampering rates [13]-A and two flexible schemes [4], [15] can deliver better performance, in case of extensive tampering the proposed algorithm is clearly superior. The threshold tampering rate is usually between $\tilde{\gamma} = 0.05$ and $\tilde{\gamma} = 0.2$, depending on the image itself.

The proposed algorithm is also beneficial when compared to constant-fidelity schemes. It delivers significantly better quality than [14], with only 9% worse tampering rate. The [13]-A scheme, with reference rate of 320 bits per block, allows for perfect recovery of 5 most significant bit-planes (expected PSNR is 40.7 dB). The maximum tampering rate

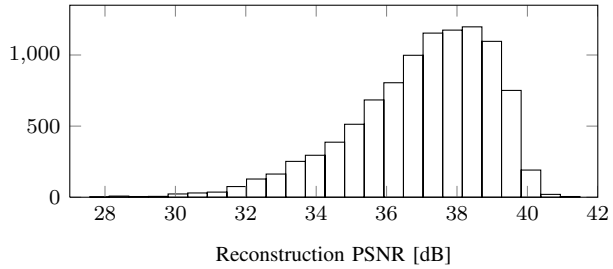


Fig. 10: Histogram of reconstruction PSNR on 10,000 images from the BOWS2 data-set.

for this scheme is 0.24. With $b = 320$ ($\lambda = 2$), the success bound of the proposed method is $\tilde{\gamma}_2 = \tilde{\gamma}_3 = 0.33$. A detailed analysis of this performance gap follows in Appendix A.

Fig. 10 shows a histogram of the PSNR scores of the proposed scheme for the complete reconstruction reference. The results were collected on the entire BOWS2 data-set, i.e., on 10,000 natural images. The average and the median PSNR are 36.99 dB and 37.34, respectively. 99% of the obtained scores fall in the $[31.3; 41.5]$ dB range.

Individual reconstruction fidelity scores for example test images are shown in Fig. 11. The plots not only clearly show the threshold tampering rates, but also demonstrate the characteristic behavior of the systems. Two flexible schemes [4], [15] reveal systematic deterioration of the reconstruction fidelity. The scheme [4] is more susceptible to the distributions of details in the image. For images with large areas of solid low-detail blocks, the curves may not be monotonic. The scheme [15] operates directly on pixel intensities, and is not affected by the problem. The plots also demonstrate the expected three distinct quality levels for the [13]-B scheme.

A reconstruction example is shown in Fig. 12. For the sake of presentation clarity, the images are 256×256 px, and are shown clipped to the half of their height. The tampered area (marked in Fig. 12a) is rectangular, and the tampering rate is $\tilde{\gamma} = 0.265$. The proposed algorithm yielded the best quality. Large amount of high-energy DCT coefficients causes reference value saturation, most visible in [14] and [13]-B. The fidelity of [4] is limited by reconstruction artifacts, typical for this scheme when dealing with larger tampering rates.

D. Discussion and Limitations

The presented reference self-embedding scheme allows for high-quality reconstruction, even under extensive tampering. The reconstruction quality does not deteriorate with the growing extent of modifications, which can reach up to 50% of the image area. Interestingly, by decreasing the amount of information in the reconstruction reference, we can obtain only a limited improvement in the reconstruction performance. For instance, by using only 50% of the available capacity, we can increase the maximal tampering rate to merely 59%.

Compared to state-of-the-art schemes, our approach can deliver better reconstruction performance. The reference scheme can be easily adapted to different requirements, as the proposed reconstruction model allows for straightforward exchange of the reconstruction fidelity for the tampering rate.

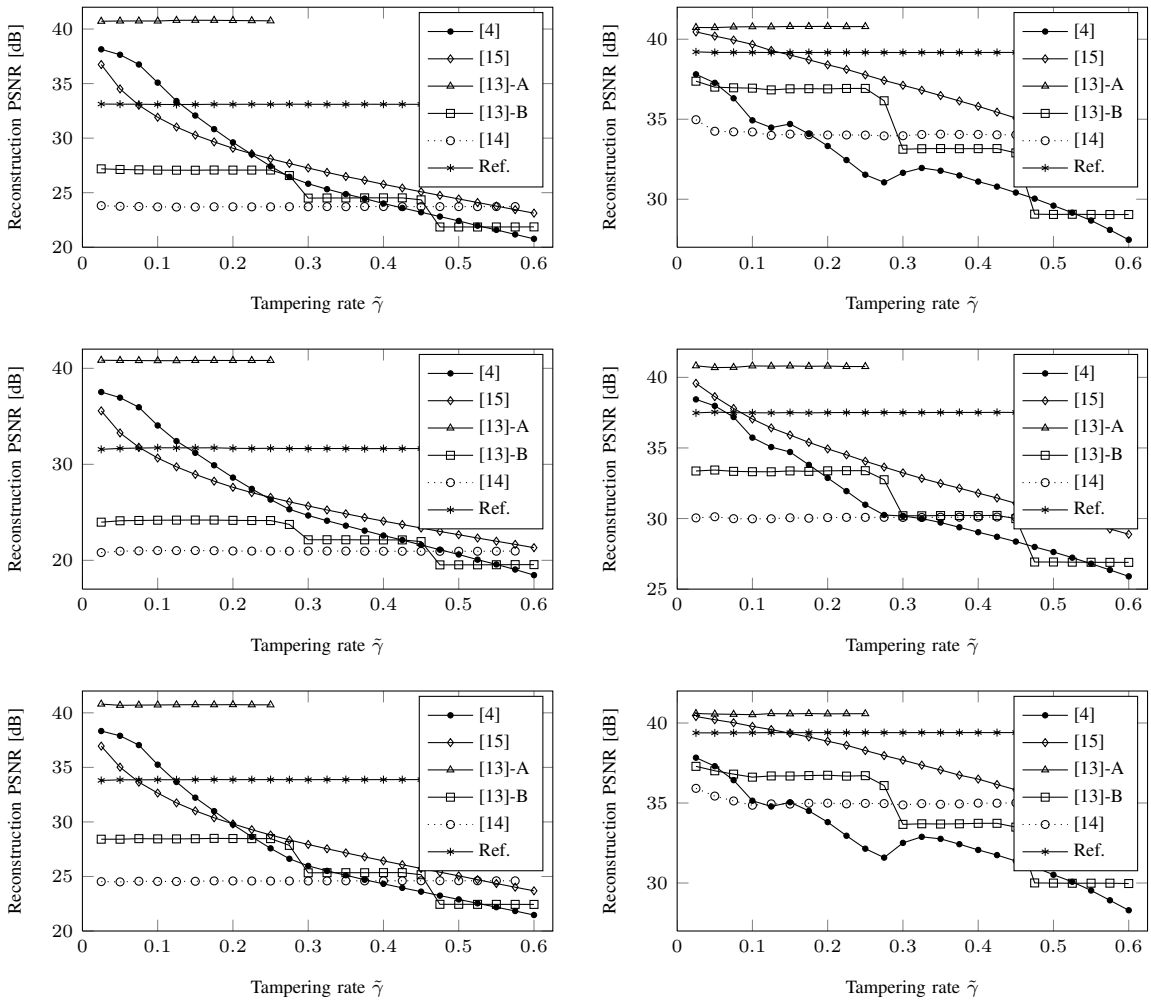


Fig. 11: Reconstruction PSNR scores for selected test images under varying tampering rates; from top left: 6882.png, 4767.png, 9011.png, 131.png, 7683.png, 4749.png.

Our analysis leads to clear guidelines for the construction of efficient content reconstruction systems:

1) Fountain codes can be used to generate a reference stream of arbitrary length while uniformly spreading the reference information over the whole image,
2) During distribution of the reference information over the image, each content fragment should uniformly contribute to the whole embedded watermark,
3) Exploitation of the authentic image content is essential to achieve optimal performance,
4) The watermark embedding function f should not interfere with the reference generation function g_b,
5) Random access to the reference stream is necessary to determine its actually needed fragments,
6) The fraction of the necessary reference stream fragments should be linearly proportional to the tampering rate, i.e., $\rho = \tilde{\gamma}$. This can be achieved by perfect alignment between the reference blocks and the reference symbols.

With the use of authentic content, it is possible to make the reference stream forget about the authentic fragments. Hence, the watermark capacity is not wasted, and it is possible to embed reference streams longer than the available capacity.

Perfect alignment between the reference blocks and symbols can also be achieved by embedding multiple shorter symbols in a single image block. Asymptotically, 1-bit symbols could be considered, and RLF would produce $160N$ watermark bits from bN reference bits. Then, the applicable success bound would always be the optimistic γ_2. However, due to high computational complexity of the RLF decoding, such an approach is not feasible. The typical solution of using a sparse generator matrix and belief propagation for the decoding is not applicable for the self-recovery problem, as low-degree symbols are likely to be quickly reduced to null useless symbols in the process of eliminating the knowledge of authentic image fragments. This issue could potentially be addressed by designing a dedicated degree distribution, but it is a separate research problem. With the use of M-ary symbols, the proposed approach can be efficiently implemented in practice, and with proper choice of (b, B), the optimistic success bounds can still be reached.

We believe that the adoption of the proposed model might enable the development of robust, easily customizable reconstruction systems. The presented reference scheme shows a practical implementation of the model for lossless digital

(a) Original image & tampering pattern, 6882.png (b) Proposed scheme, 30.82 dB (c) [4], 23.55 dB

(d) [13]-B, 24.73 dB (e) [15], 25.11 dB (f) [14], 21.40 dB

Fig. 12: Reconstruction results shown on clipped 256×256 px images; tampering pattern shown in (a); rate $\tilde{\gamma} = 0.265$.

images. While B needs to be constant, the fidelity of individual blocks can vary, i.e., $b = b(i)$. The derived analytical formulas, and the notion of the reconstruction demand can be directly used to calculate the reconstruction success bounds for such adaptive schemes. The problem will be studied in detail in our future work.

V. Conclusions

In conclusion, the presented analysis gives new insights into the inherent self-recovery trade-offs. We have shown that the erasure channel is a valid model of the content reconstruction problem. By redefining the success criterion, and exploiting the remaining authentic content it is possible to theoretically analyze the reconstruction performance with the use of communication theory. Based on these foundations, we have derived the formulas for the appropriate success bounds. They allow for straightforward assessment of the expected performance for an arbitrary configuration of the restoration process.

The presented self-recovery model provides a general framework of the problem. It does not impose any specific watermark embedding strategy, reference generation function, or hash calculation algorithm. Only a small set of necessary requirements is given, in order for the model to be applicable in practice. For the sake of fair comparison of the achievable reconstruction performance, we have experimentally evaluated a reference image authentication system, designed with the same embedding strategy and payload as most of existing alternative schemes.

Our work shows that high-quality reconstruction is still possible, even under extensive tampering.

Appendix A

In this appendix we compare the proposed reconstruction reference spreading technique with the one used in [13]. We will show that both techniques are asymptotically equivalent, and are essentially two different approaches of practical implementation of the same high-level paradigm. The fundamental concept involves spreading of the reference information over the whole image, and exploitation of the remaining authentic content to aid the restoration process.

We consider the scheme [13]-A, which features constant-quality reconstruction, and uses 320 most significant bits per 8×8 px image block as reference information. The bits are randomly permuted, and organized into L-bit subsets. A N' pixel image contains $5N'/L$ such subsets. To fit into the available watermark capacity, the subsets are then projected onto $L/2$-bit vectors with the use of a pseudo-random binary matrix of size $L/2 \times L$. After concatenation, the reconstruction reference is randomly permuted, and embedded into 3 least significant bit-planes.

Therefore, the method divides the reconstruction problem into smaller problems, where the spreading mechanism is applied locally to randomly selected image portions. The probability of successful reconstruction is:

$$P_S = P_{LI}^{5N'/L},$$

where P_{LI} is the probability of success within a single subset. It is calculated with the use of two binomial distributions, and a recursive formula for the probability that a random binary matrix is of sufficient rank. Asymptotically, we consider the case of $L = 5N'$, i.e., with a single subset covering the whole image. Such a configuration is equivalent to the proposed approach operating on 1-bit symbols (Section IV-D).

If there exists only a single subset, the amount of tampered, and extractable elements no longer has a stochastic character, and the problem resolves to solving a $(1-\tilde{\gamma})\ 5N'/2 \times \tilde{\gamma}\ 5N'$ linear system in GF(2) arithmetic. It becomes possible if the number of columns is at most equal to the number of rows, i.e.:

$$(1-\tilde{\gamma}_z)\ 5N'/2 = \tilde{\gamma}_z\ 5N' \Rightarrow \tilde{\gamma}_z = \frac{1}{3}. \tag{8}$$

The obtained asymptotic bound is identical to $\tilde{\gamma}_2$ in our approach (3). We will now analyze the reconstruction success

probability, and demonstrate how it converges to a threshold in $\tilde{\gamma}_z$. Since the original calculation procedure from [13] is not feasible for the considered problem size, we will use it in a different form. Let $E[\tilde{\gamma}]$ denote the probability of tampering a single image block, and also the expected tampering rate. The success probability for a single subset is:

$$P_{LI} = \sum_{i=0}^{L/2} \sum_{j=0}^{L} P_v(i) P_{nT}(j)(1 - q(i,j)),$$

where:

$$P_{nT}(j) = e^{\ln\binom{L}{j} + j \cdot \ln(E[\tilde{\gamma}]) + (L-j) \cdot \ln(1-E[\tilde{\gamma}])},$$
$$P_v(i) = e^{\ln\binom{L/2}{i} + i \cdot \ln(1-E[\tilde{\gamma}]) + (L/2-i) \cdot \ln(E[\tilde{\gamma}])}.$$

In such form, P_{nT} and P_v can be efficiently calculated by using the logarithmic gamma function to obtain the binomial coefficients. The term $1 - q(i,j)$ denotes the probability that a random binary matrix of size $i \times j$ has sufficient rank. Instead of the recursive formula from [13], we approximate it as:

$$q(i,j) \approx \begin{cases} 1, & \text{if } j > i, \\ 2^{-i}, & \text{if } j = 1, \\ 0.712, & \text{if } j = i, \\ 2^{j-i}, & \text{otherwise.} \end{cases}$$

The approximation is founded on boundary analysis [21]:

Lemma 1. *Let V be a vector space of dimension n over $GF(q)$ and let $m \geq n$. Then, the probability that m random vectors in V span the whole space V is:*

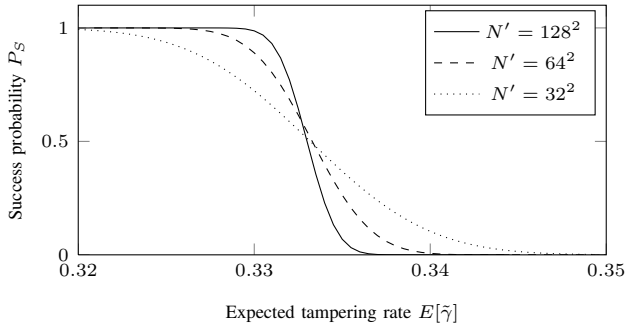
$$\prod_{i=1}^{n} (1 - \frac{1}{q^{m-n+i}}) \geq \begin{cases} 0.288, & \text{if } m = n \text{ and } q = 2, \\ 1 - \frac{1}{q^{m-n}(q-1)}, & \text{otherwise.} \end{cases}$$

Equivalently, this also bounds the probability that a random $m \times n$ matrix over $GF(q)$ has rank n.

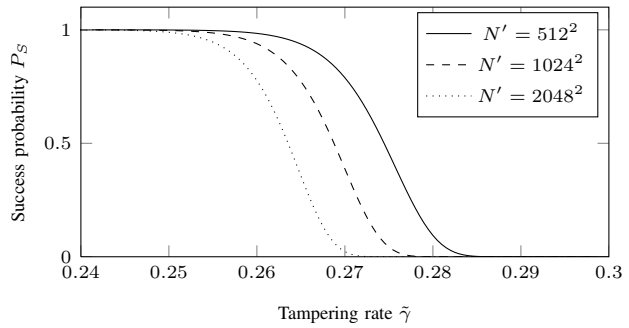
With a fixed $L = 5N'$, the success probability has only one degree of freedom, i.e., the size of the image. As N' increases, the reconstruction becomes more probable around $\tilde{\gamma}_z$. Fig. 13a shows P_S vs. $E[\tilde{\gamma}]$ for different images sizes. Just as expected, the slope becomes steeper, and the curve approaches a threshold in $\tilde{\gamma}_z$.

We can conclude that the proposed spreading technique, and the one from [13] are essentially two different methods of practical implementation of the same high-level concept. Instead of dividing the reconstruction problem into smaller fragments, our approach uses M-ary symbols for reference information processing in a single spreading process.

Our approach has three major benefits. Firstly, it does not suffer from an inherent performance penalty, and can still reach the optimistic success bound $\tilde{\gamma}_2$. Given the same rate of reference information, it allows for working with higher tampering rates, e.g., for 320 bits per block ($\lambda = 2$), the difference in the maximal tampering rate is 0.33 vs. 0.24. Secondly, the method from [13] is susceptible to image size, and the performance deteriorates as N' increases (Fig. 13b). Thirdly, the proposed approach, formulated in terms of digital fountain coding, can be conveniently analyzed with communication theoretic tools. The applicable success bounds are expressed in terms of well-defined formulas for an arbitrary configuration of the system.



(a) Asymptotic variant, $L = 5N'$



(b) Original variant, $L = 512$

Fig. 13: Probability of successful recovery of the spreading method from [13] for different image sizes.

APPENDIX B

Table II summarizes the achievable reconstruction performance, and the adopted approaches in state-of-the-art self-embedding schemes. The reconstruction quality results are presented in terms of fuzzy scores:

- *Low* - typical restoration PSNR $<$ 30 dB,
- *Medium* - typical restoration PSNR $<$ 35 dB,
- *High* - typical restoration PSNR $<$ 40 dB,
- *V. High* - typical restoration PSNR $>$ 40 dB,
- *Loss-less* - no distortion, PSNR = ∞ dB.

The *reference origin* column indicates the origin of the reconstruction reference: P - pixel domain, T - transform domain, and the number of most significant bit-planes used in the reference generation process.

REFERENCES

[1] J. Fridrich and M. Goljan, "Images with self-correcting capabilities," in *Proc. of IEEE International Conference on Image Processing*, 1999.

[2] J. Lee and C. S. Won, "Authentication and correction of digital watermarking images," *Electronics Letters*, vol. 35, no. 11, pp. 886–887, 1999.

[3] X. Zhu, A. Ho, and P. Marziliano, "A new semi fragile image watermarking with robust tampering restoration using irregular sampling," *Signal Processing : Image Communication*, vol. 22, no. 5, 2007.

[4] X. Zhang, Z. Qian, Y. Ren, and G. Feng, "Watermarking with flexible self-recovery quality based on compressive sensing and compositive reconstruction," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp. 1223–1232, 2011.

TABLE II: Reconstruction performance and approaches of state-of-the-art self-embedding schemes; notation: LSB - least significant bit substitution, DC - differential coding, DE - difference expansion, bpp - bits per pixel, bpb - bits per block.

Scheme	Embedding Method	PSNR	Restor. fidelity	Max. tamp. rate	Restoration approach	Reference origin	Reference rate	Reference spreading method	Reuses auth. data
Schemes with constant reconstruction quality									
Fridrich [1] - I	LSB	44 dB	Low	N/A	read permuted data	T / 8 MSB	64 bpb	N/A	N
Fridrich [1] - II	DC	33 dB	Low	N/A	read permuted data	P / 4 MSB	4 bpp	N/A	N
Zhang [5]	DE	29 dB	Loss-less	3.2	solve linear systems for random content groups	P / 8 MSB	8 bpp	local-spreading, global-scatter	Y
Zhang [14]	LSB	38 dB	Low	59	solve linear systems for random content groups	T / 5 MSB	54 bpb	local-spreading, global-scatter	Y
Zhang [24]	LSB	38 dB	V. High	6.6	brute-force MSB search until hash valid	P / 5 MSB	320 bpb	local-spreading, global-scatter	N
Zhu [3]	PST[1]	35 dB	Low	N/A	irregular sampling, projection on convex sets	T / 8 MSB[0]	N /A	N/A	N/A
Cheddad [16]	DWT[2]	42 dB	Low	N/A	inverse halftone of binary watermark	P / 1 bit dither	1 bpp	N/A	N
Zhang [13] - A	LSB	38 dB	V. High	24	solve linear systems for random content groups	P / 5 MSB	320 bpb	local-spreading, global-scatter	Y
Flexible or adaptive schemes									
Korus [6]	LSB	38 dB	Medium[3]	≤ 72[3]	general erasure communication, LT code	T / 5 MSB	variable	global-spreading	N
Qian [18]	LSB	51 dB	Medium	N/A	read permuted data, error correction coding	T / 7 MSB	variable	global-scatter	N
Qian [19]	LSB	38 dB	Medium	35	solve linear system for random content groups	T / 5 MSB	variable	local-spreading, global-scatter	N
Zhang [15]	LSB	38 dB	High[4]	54	*pixel xor* + bit estimation from local pixel correl.	P / 5 MSB	160 bpb	global-scatter	Y
Zhang [4]	LSB	38 dB	High[4]	60	compressive sensing / compositive reconstruction for random content groups	T / 5 MSB	161 bpb	local-spreading, global-scatter	Y
Zhang [13] - B	LSB	38 dB	Medium[5]	24-66	solve linear systems for random content groups	P / 5 MSB	20 + 45 + 102 bpb	local-spreading, global-scatter	Y
Lee [12]	LSB[6]	41 dB	High[4]	95	read permuted data, two chances	P / 5 MSB[7]	N/A	global-scatter	N
Different configurations of the proposed scheme									
$\lambda = 1$	LSB	38 dB	High	50	self-recovery, RLF code	T / 5 MSB	160 bpb	global spreading	Y
$\lambda = 2$	LSB	38 dB	V. High	33	self-recovery, RLF code	T / 5 MSB	320 bpb	global spreading	Y

[0] only DCT coefficient polarity
[1] quantization in Pinned Sine Transform (PST) domain
[2] bit substitution in DWT domain
[3] quality exchangeable for the tampering rate
[4] gradual drop of the restoration fidelity
[5] level-wise drop of the restoration fidelity
[6] modified version with additional smoothing
[7] down-scaled image used as reference

[5] X. Zhang and S. Wang, "Fragile watermarking with error free restoration capability," *IEEE Transactions on Multimedia*, vol. 10, no. 8, 2008.

[6] P. Korus and A. Dziech, "A novel approach to adaptive image authentication," in *Proc. of IEEE International Conference on Image Processing*, 2011.

[7] S. Yafei, Z. Li, W. Guowei, and L. Xinggang, "Reconstruction of missing blocks in image transmission by using self-embedding," in *Proc. Int Intelligent Multimedia, Video and Speech Processing Symp*, 2001, pp. 535–538.

[8] G. Gur, Y. Altug, E. Anarim, and F. Alagoz, "Image error concealment using watermarking with subbands for wireless channels," *IEEE Communications Letters*, vol. 11, no. 2, pp. 179–181, 2007.

[9] C. B. Adsumilli, M. C. Q. Farias, S. K. Mitra, and M. Carli, "A robust error concealment technique using data hiding for image and video transmission over lossy channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 11, pp. 1394–1406, 2005.

[10] G. Li, Y. Ito, X. Yu, N. Nitta, and N. Babaguchi, "Recoverable privacy protection for video content distribution," *EURASIP Journal on Information Security*, vol. 2009, January 2009.

[11] P. Korus, W. Szmuc, and A. Dziech, "A scheme for censorship of sensitive image content with high-quality reconstruction ability," in *Proc. of IEEE International Conference on Multimedia and Expo*, 2010.

[12] T.-Y. Lee and S. Lin, "Dual watermark for image tampering detection and recovery," *Pattern Recognition*, vol. 41, pp. 3497–3506, 2008.

[13] X. Zhang, S. Wang, Z. Qian, and G. Feng, "Reference sharing mechanism for watermark self-embedding," *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 485–495, 2011.

[14] X. Zhang, S. Wang, and G. Feng, "Fragile watermarking scheme with extensive content restoration capability," in *Proc. of International Workshop on Digital Watermarking*, 2009.

[15] X. Zhang, S. Wang, Z. Qian, and G. Feng, "Self-embedding watermark with flexible restoration quality," *Multimedia Tools and Applications*, vol. 54, pp. 385–395, 2011, 10.1007/s11042-010-0541-z.

[16] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "A secure and improved self-embedding algorithm to combat digital document forgery," *Signal Processing*, vol. 89, pp. 2324–2332, December 2009.

[17] M. Luby, "Lt codes," in *Proceedings of the 43rd Symposium on Foundations of Computer Science*, Washington, DC, USA, 2002, FOCS '02, pp. 271–280, IEEE Computer Society.

[18] Z. Qian and G. Feng, "Inpainting assisted self recovery with decreased embedding data," *IEEE Signal Processing Letters*, vol. 17, no. 11, pp. 929–932, Nov. 2010.

[19] Z. Qian, G. Feng, X. Zhang, and S. Wang, "Image self-embedding with high-quality restoration capability," *Digital Signal Processing*, vol. 21, no. 2, pp. 278–286, Mar. 2011.

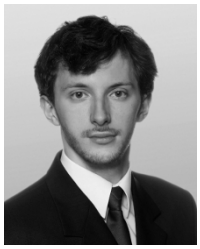
[20] D. J. MacKay, "Fountain codes," *IEE Proceedings Communication*, vol. 152, no. 6, 2005.

[21] R. P. Brent, S. Gao, and A. G. B. Lauder, "Random Krylov spaces over finite fields," *SIAM Journal on Discrete Mathematics*, vol. 16, no. 2, pp. 276–287, Feb. 2003.

[22] S. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129 – 137, mar 1982.

[23] "The dataset from the 2nd bows contest," http://bows2.ec-lille.fr/, 2007, Visited on 26 March 2012.

[24] X. Zhang and S. Wang, "Fragile watermarking scheme using a hierarchical mechanism," *Signal Processing*, vol. 89, pp. 675–679, 2009.



Paweł Korus received his M.Sc. degree in telecommunications from the AGH University of Science and Technology in 2008 (with honors). He is currently a Ph.D. candidate at the Department of Telecommunications of AGH University of Science and Technology.

Since 2008 he has actively participated in both national, and international research projects, e.g., BONE, INDECT, INSIGMA. His research interests include multimedia security, information hiding, digital watermarking, and multimedia communications.

Andrzej Dziech holds the position of a full professor at the Department of Telecommunications of AGH University of Science and Technology in Krakow, Poland.

He received his M.Sc. and Ph.D. degrees from the Institute of Electrical Engineering in Saint Petersburg in 1970 and 1973, respectively, and the D.Sc. from Technical University of Poznan in 1978. He is an author of 6 books, and nearly 180 publications. He was a supervisor of 18 Ph.D. students.

His fields of interest are related to digital communications, image and data processing, data compression, information and coding theory, random signals, computer communications networks and signal processing. He was awarded 4 times for research achievements by the Ministry of Education of Poland.

Professor Dziech actively participated in numerous international research projects, e.g., Tempus, Knixmas, Calibrate. Currently, he is coordinating a European Union FP7 integrated project INDECT.