Multimedia Security: Open Problems and Solutions

S. Voloshynovskiy [a,1], O. Koval [a] F. Deguillaume [a] and T. Pun [a]
[a] *CUI-University of Geneva, 24, rue General Duour, 1211, Geneva, Switzerland*

Abstract. In this paper we introduce and develop a framework for visual data-hiding technologies that aim at resolving emerging problems of modern multimedia networking. First, we present the main open issues of multimedia security and secure communications. Secondly, we formulate multimedia data-hiding as communications with side information and advocate an appropriate information-theoretic framework for the analysis of different data-hiding methods in various applications. Finally, we discuss data-hiding-based solutions to some multimedia security related problems.

Keywords. Data-Hiding, Robust Watermarking, Copyright Protection, Tamper Proofing, Secure Communications, Steganography

1. Introduction

The mass diffusion of digital media and the explosive growth of telecommunication are reshaping the lifestyles of ordinary people, research and industry. Over the last decade, the rise of digital telecommunication technologies has fundamentally altered how people work, think, communicate, and socialize.

Despite the obvious progress of multimedia communications, these developments carry with them a number of risks such as copyright violation, prohibited usage and distribution of digital media, secret communications, and network security. Therefore, security, scalability and manageability amongst others become issues of serious concern, as current solutions do not satisfy anymore the growing demands of multimedia communications.

In the scope of this paper, we will focus on a possible solution for multimedia security in order to prevent unauthorized data exchange and to ensure secure communications. Two main objectives will be addressed: the first one is to introduce and to overview a novel approach to multimedia security based on data-hiding technologies. We will consider theoretical fundamentals of digital data-hiding technologies and will demonstrate the relevance of data-hiding problems to digital communications. We will show the advantages of data-hiding based multimedia security protocols over the traditional general means of security based on encryption, scrambling and firewall systems. The second objective of the paper is to demonstrate some of the main achievements in the field of digital data-hiding technologies for multimedia security.

[1] Correspondence to: S. Voloshynovskiy, CUI-University of Geneva, 24, rue General Duour, 1211, Geneva, Switzerland. Tel.: +41 22 379 7637; Fax: +41 22 379 7780; E-mail:svolos@cui.unige.ch.

The paper is organized as follows: Section 2 formulates the main requirements to the multimedia security systems. Section 3 introduce digital data-hiding as a mean for multimedia security and secure communications. Section 4 considers authentication and tamper proofing. Section 5 presents secure communications and Section 6 concludes the paper.

$\mathbf{Notation}$. We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables $\mathbf{X}$, corresponding small letters x and $\mathbf{x}$ to denote the realizations of scalar and vector random variables, respectively. The superscript N is used to denote length-N vectors $\mathbf{x} = x^N = \{x[1], x[2], ..., x[N]\}$ with ith element $x[i]$. We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to indicate that a random variable X is distributed according to $p_X(x)$. Calligraphic fonts $\mathcal{X}$ denote sets $X \in \mathcal{X}$ and $|\mathcal{X}|$ denotes a cardinality of set.

2. Multimedia security: main requirements

Multimedia content security has a number of specific requirements that should allow to answer to the following questions: *Who has issued the multimedia content? Who is the content owner? When was the content issued? Who has access right to the content? Is the content modified? Where was the content modified? What was the original content before modification?*

The list of the related problems (like esteblishing secure and undetectable communications) is very broad and from a traditional point of view there does not seem to exist any common means of satisfying all these requirements. However, there are some common aspects of secure and reliable communications that could be addressed by novel technologies based on digital data-hiding.

3. Multimedia data-hiding

Multimedia data-hiding represents a reliable mean for secure communications. It provides a “virtual” channel of digital communications through the embedding of some secret unperceived information directly into the multimedia content.

It should garantee: perceptually invisible data embedding; reliable extraction of embedded information; security provided by a proper key management and undetectability of the hidden data presence by the existing detection tools.

We consider multimedia data-hiding with respect to three main applications that should address the open issues presented in Section 2: **robust watermarking; authentication and tamper proofing; secure communications**.

3.1. Robust watermarking

Robust watermarking should provide the reliable communication of a message m in the body of a multimedia content under a broad list of various intentional and unintentional attacks constituting watermarking channel (Figure 1).

The goal of the information embedder consists in the invisible “integration” of a specifically preprocessed message m into the original content $\mathbf{x}$ based on some secret key K. We assume that the message M, uniformly distributed over the message set $\mathcal{M}$ of cardinality $|\mathcal{M}|$, is encoded based on a secret key into some watermark $\mathbf{w}$, $w[i] = f_i(m, x^i)$, and embedded into a host data $\mathbf{x}$, producing the stego data $\mathbf{y}'$, $y'[i] = x[i] + w[i]$. The message m typically has a 64-bit length, i.e., $|\mathcal{M}| = 2^{64}$, and is content independent. In 1-bit watermarking, only a binary decision about the watermark presence/absence can be required. As another example, the printing industry only requires 16 bits for document tracking aiming at identifying the distribution channels.

The admissible distortion for watermark embedding is D_1:

$$E[d_1^N(\mathbf{X}, \mathbf{Y}')] \leq D_1, \tag{1}$$

where $d_1^N(\mathbf{X}, \mathbf{Y}') = \frac{1}{N}\sum_{i=1}^{N} d_1(x[i], y'[i])$ denotes $N-$vector distortion between $\mathbf{X}$ and $\mathbf{Y}'$ and $d_1(x[i], y'[i])$ is the element-wise distortion between $x[i]$ and $y'[i]$.

The channel is characterized by a transition probability $p(y|w, x)$, and can be quite general. In the particular case of intentional attacks, the attacker aims at removing the watermark $\mathbf{w}$ from $\mathbf{y}'$ producing the attacked data $\mathbf{y}$. The admissible attacker distortion is D_2 that is defined in the same way as (1) between vectors $\mathbf{y}'$ and $\mathbf{y}$: $E[d_2^N(\mathbf{Y}', \mathbf{Y})] \leq D_2$. One should also note another possibility to define the attacker distortion between the original data $\mathbf{x}$ and the attacked data $\mathbf{y}$. The decoder produces the estimate of $\hat{M}$ based on $\mathbf{y}$ using:

$$\hat{m} = g(y^N), \tag{2}$$

where $g(.)$ denotes the decoding rule and $\mathbf{y} = y^N$ is the distorted stego data. The decoding error occurs when $\hat{M} \neq M$. A particular case of generalized decoding rule $g(.)$ is the maximum a posteriori (MAP) decoding rule, which minimizes the probability of error, i.e. $\hat{m} = argmax_{m\in\mathcal{M}} p(m|y^N)$.

If $\mathbf{x}$ is not known at encoder and decoder it acts as an interference. In the case of watermarking, the host data $\mathbf{x}$ is available at the encoder. Therefore, this case can be considered as communication with side information available at the encoder (Figure 2) that was considered by Gel'fand and Pinsker in 1980 in non-watermarking applications. The capacity of this scheme was found as [11]:

$$C = max_{p(u,w|x)}\left[I(U;Y) - I(U;X)\right], \tag{3}$$

where U is an auxiliary random variable.

Costa (1983) has considered the above problem in the Gaussian context and found that, if $U = W + \alpha X$ and $\alpha = \frac{\sigma_w^2}{\sigma_w^2+\sigma_z^2}$, then $C = \frac{1}{2}\log_2\left(1 + \frac{\sigma_w^2}{\sigma_z^2}\right)$.

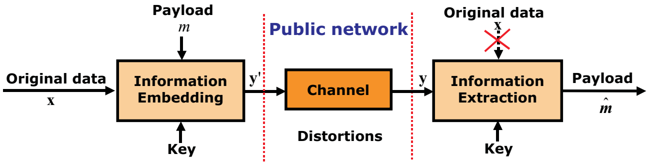

Figure 1. Generalized diagram of robust watermarking.

Having considered theoretical fundamentals of robust watermarking, we concentrate on the practical data-hiding schemes. They can be classified depending whether or not the side information about the host data is used at the encoder.

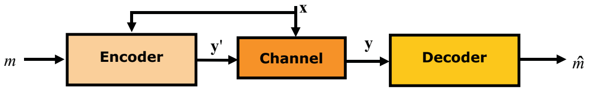

Figure 2. Robust watermarking as communications with side information at encoder.

Spread spectrum (SS) data-hiding does not directly use information about the host image for the watermark generation:

$$y'[i] = x[i] + w[i]. \tag{4}$$

In the most practical SS robust watermarking schemes proper spreading is applied for security, redundancy and geometrical attacks resistance reasons. This spreading is performed over the host data using a key-dependent spreading sequence $s[j] \in \{\pm 1\}$ such that $w[j] = c[k]s[j]$, $j \in S_k$, and where $\mathbf{c}$ is the codeword of length L_c that is mapped to 2-PAM, i.e., $\mathbf{c} \in \{\pm 1\}^{L_c}$ and S_k are non-overlapping subsites that are used for the allocation of each bit of codeword $\mathbf{c}$. Additionally, the watermark can be embedded exploiting particularities of the human visual system (HVS) and the details of perceptually adapted watermarking can be found in [6,17,19,21].

Three main variations of practical host interference free data-hiding are: Least Significant Bit Modulation (LSBM), Quantization Index Modulation (QIM) [3], Scalar Costa Scheme (SCS) [7].

The LSBM encoder embeds the data according to the next rule:

$$y'[i] = Q(x[i]) + d[i] = x[i] + d[i] + (Q(x[i]) - x[i]) = x[i] + w[i]. \tag{5}$$

The image is first precoded based on an uniform quantizer $Q(\mathbf{x})$ with a step Δ and then the M-PAM watermark $\mathbf{d}$ is added to this image (meaning that $Q(x)$ output is requantized to M levels). The LSBM decoder performs the direct estimation of the message:

$$\hat{d}[i] = y[i] - Q(y[i]). \tag{6}$$

The binary QIM encoder performs host image quantization using two sets of quantizers $Q_{-1}(.)$ and $Q_{+1}(.)$ that are shifted by Δ with respect to each other:

$$y'[i] = Q_d(x[i]) = x[i] + (Q_d(x[i]) - x[i]) = x[i] + w[i], \tag{7}$$

where $Q_d(.)$ denotes the quantizer for $d = -1$ and $d = +1$. The QIM decoder performs the ML-estimation:

$$\hat{d} = argmin_{d \in \{\pm 1\}} |y[i] - Q_d(y[i])|^2. \tag{8}$$

Contrarily to the LSBM and the QIM, which do not use any prior information about the attacking channel state, the SCS exploits the knowledge of the AWGN channel statistics at the encoder. The SCS encoding rule is:

$$y'[i] = x[i] + \alpha(Q_d(x[i]) - x[i]) = x[i] + \alpha w[i]. \tag{9}$$

Decoding in the binary SCS is performed according to (8).

Summarizing the above discussion, we can point out the main requirements to the robust watermarking. It requires the embedding of a 64-bit content independent message into the original image in an invisible manner specified by a proper distortion criteria. Strong robustness to all intentional and unintentional attacks is also required including both signal processing and geometrical transformations. The security requirement calls for a proper resistance against message removal that would be based on the knowledge of the algorithm.

4. Integrity Control and Tamper proofing

The goal of integrity control and tamper proofing consists in the verification of content integrity, in the detection of local modifications in multimedia data, in the recovering of the original content based on the available copy of modified/tampered content. The generalized integrity control and verification system (Figure 3) consists of three main parts. Information embedding part performs D_1-distortion-constrained embedding of the payload b into the original data $\mathbf{x}$. Contrarily to the robust watermarking, b is content dependent and related to the original data by some mapping $p(b|x)$ that might represent some hashing, features or compressed version of the original content and has a higher rate (about 5-10 Kbits depending on the size of the original data).

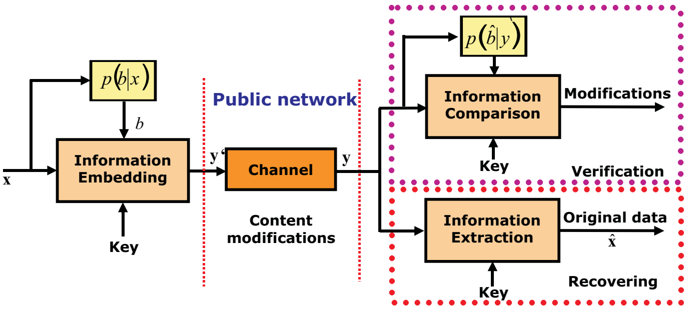

Figure 3. Generalized diagram of integrity control, tamper proofing and self-recovering systems.

The behavior of the channel $p(y|y')$ also differs significantly from the corresponding robust watermarking channel. Contrarily to the latter case, where the attacker is interested in D_2-constrained impairing the reliable watermark detection/decoding, the protocol attacker in the former case targets modifying or counterfeiting the visual appearance of the original content. In this case the document is either partially modified or a fraction of the document is copied into another document. Therefore, the global introduced distortion D_2 is of secondary importance for the evaluation of the degree of document modification for this application (Figures 4). The goal of the decoder of a tamper proofing system is thus to reliably detect the intentional or unintentional modifications, and to point out the modified areas or preferably reconstruct the original content.

Therefore, from the attacker perspective the integrity of the document should be preserved in such a way that the authentication watermark will not be capable to detect the introduced modifications. Recently, a lot of attention in the watermarking community was drawn to the investigation of new protocol attacks against tamper proofing systems [5]: such attacks are mostly advanced substitution attacks including the *cut-and-*

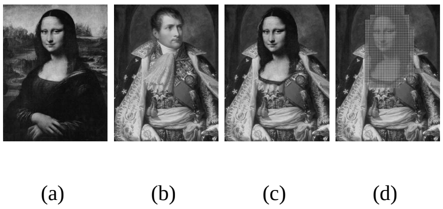

Figure 4. Tamper proofing example: (a) and (b) original images, (c) result of collage between (a) and (b), (d) highlighted regions indicate the content modifications.

paste attack [1], the *vector-quantization* (VQ) or *Holliman-Memon* attack [16], image compositions and the *collage attack,* as well as cryptographic attacks targeting the used hashing function.

To withstand the above attacks one should properly design a data-hiding scheme that should resolve two related problems: the first one is the *detection of modifications*; the second one is the *recovering of the original data* $\mathbf{x}$ after content modifications. Leaving the latter issue outside of the scope of this paper, *authentication and tamper proofing* watermarking will be discussed in the remaining part of this section.

Authentication aims at checking the authenticity of a document and of its source, while tamper proofing detects unauthorized modifications. Early authentication watermarks are the Yeung-Mintzer scheme [23] which authenticates each pixel with respect to a binary logo, and [4,22], which divide the image in blocks and attach cryptographic hash-codes or signatures within blocks.

However, most of schemes based on block-wise independent hashing are vulnerable to substitution attacks which exploit databases of images all protected with the same key. The cut-and-paste attack takes parts of several protected images and pastes them together (preserving the watermark synchronization) to form a new image. The collage attack is a cut-and-paste attack which uses rather large parts: in that case these parts are individually validated by the decoder and only the boundaries between them are indicated as tampered. Even more powerful VQ attack [15] allows the construction of completely arbitrary good quality images, which are wrongly authenticated by the decoder, by pasting blocks from already watermarked images. Moreover, regarding robust watermarking, most existing schemes are vulnerable to the copy attack [18]. This is a potential problem for many practical applications: if the watermark can be copied, how to be sure that the document actually holds the decoded copyright?

Various methods of blocks or hash-codes/signatures chaining, undeterministic signatures, etc. have been proposed for authentication watermarks against substitution attacks [1]. Fridrich [9] proposed to embed unique identifiers (ID) or “time-stamps” within individual images or even within individual blocks, a method which efficiently and conveniently defeats collage attacks. To make robust watermarks resistant against the copy attack, one possibility is to include host related data into the watermark by joining robust and authentication watermarks in a *hybrid* scheme. Therefore the hybrid scheme can resolve problems related to copyright, authenticity and integrity in an integrated framework, and furthermore it is also able to defeat both protocol attacks above: the copy attack is made impossible since local signatures mismatch if the watermark is copied from one image to another; and regarding the collage attack, the robust part of the hybrid watermark can help us to identify the areas coming from different sources since they hold different robust watermarks [5,9,20].

Most of proposed authentication (or *fragile*) watermarks are strictly sensitive to any change: even a single pixel modification is detected. Thus, they are not suitable for compression nor for digital/ analogue conversion. Media-conversion compatible schemes called *semi-fragile* watermarks have then been proposed, based on *robust visual hashing* as well as on an embedding approach which resists against a certain level of “acceptable and non-malicious” distortions [10,14].

5. Secure Communications

The goal of secure communications is to securely deliver some content via the public networks. Among the existing possibilities for secure communications is a visual “encryption” or scrambling that should provide additional error resilience in the case of lossy transmission. The second possibility is steganography that ensures secure content delivery by hiding it into the covert media in the undetectable manner.

5.1. Visual scrambling

The goal of visual scrambling (Figure 5) consists in the enciphering of visual content in a way suitable for reliable communications over public networks. The content that should be securely communicated is scrambled at the encoder based on the private key in such a way that it cannot be anymore visually recognized. Contrarily to traditional data enciphering, it is required here to ensure additionally to encription the error resilience as well as to avoid any redundancy in headers, meta data and attachments and to provide format independence. Finally, the decoder should provide reliable descrambling of the content even if bits, blocks or packets have been corrupted during transmission. One of possible solutions to this problem based on phase encryption was proposed in [12].

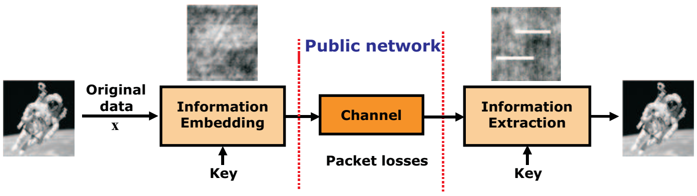

Figure 5. Generalized diagram of visual scrambling.

5.2. Steganography

Steganography (Figure 6), originally designed for hidden communications, should provide a certain level of security for public communications. The encoding/ decoding part of steganographic systems are similar to robust watermarking. However, it has reduced robustness requirements allowing a higher embedding rate. It should withstand unintentional attacks such as format conversion, slight lossy compression and in some special cases analog to digital conversion. While most existing steganographic tools can provide perceptually invisible data-hiding, the stochastic visibility of hidden data still remains a challenging task. Therefore, to be secure, the steganographic system should satisfy a set of requirements. The main one consists in providing the statistical indistinguishability between the cover data and the host data in terms of, for instance, a relative entropy [2].

The basic steganography protocol requires high-rate communications. Thus, the host interference cancellation issue should be resolved. The QIM and SCS-based embedding for steganographic purposes [8,13], have proven that the SCS-based steganography is secure according to the ϵ-security criterion [2].

However, this is a global criterion that does not reflect the local content modifications. This means that the content can be modified locally in such a way that the attacker can detect it either visually or using some specially designed statistical tests, while the relative entropy can be very low. Thus, in order to achieve undetectability when local data analysis is performed, new more accurate design criteria of steganographic systems should be exploited.

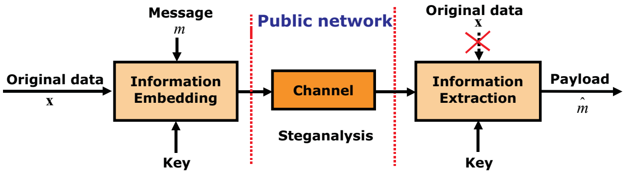

Figure 6. Generalized diagram of steganography-based secure communications.

6. Conclusion

In this paper we considered the problem of multimedia security from the data-hiding perspeectives. We formulated the main open issues of multimedia security and discussed watemarking-based solution to these issues. We presented digital communications with side informations as a unified theoretical framework for the analysis of digital data-hiding and considered its main applications. The main requirements, design principles, generalizations are underlined in the paper.

Acknowledgment

This paper was partially supported by SNF Professorship grant No PP002-68653/1, the European Commission through the IST Programme under contract IST-2002-507932-ECRYPT and IM2 projects. The authors are thankful to Y. Rytsar, J. Vila, E. Topak and R. Villan for many helpful and interesting discussions.

References

[1] P. S. L. M. Barreto, H. Y. Kim, and V. Rijmen. Toward a secure public-key blockwise fragile authentication watermarking. In *ICIP 2001*, pages 494–497, Thessaloniki, Greece, October 2001.

[2] C. Cachin. An information-theoretic model for steganography. In *IHW'98*, Portland, Oregon, USA, April 1998.

[3] B. Chen and G. W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. on Information Theory.*, 47:1423–1443, May 2001.

[4] D. Coppersmith, F. Mintzer, C. Tresser, C. W. Wu, and M. M. Yeung. Fragile imperceptible digital watermark with privacy control. In *Proceedings of SPIE 1999*, San Jose, CA, USA, January 1999.

[5] F. Deguillaume, S. Voloshynovskiy, and T. Pun. Secure hybrid robust watermarking resistant against tampering and copy attack. *Signal Processing*, 83(10):2133–2170, 2003.

[6] J. F. Delaigle, C. De Vleeschouwer, and B. Macq. Watermarking algorithm based on a human visual model. *Signal Processing*, 66:319–335, 1998.

[7] J. Eggers, J. Su, and B. Girod. A blind watermarking scheme based on structured codebooks. In *Secure images and image authentication, IEE Colloquium*, pages 4/1–4/6, London, UK, April 2000.

[8] J.J. Eggers, R. Bäuml, and B. Girod. A communications approach to image steganography. In *Proceedings of SPIE: Electronic Imaging 2002, Security and Watermarking of Multimedia Contents IV*, volume 4675, pages 26–37, San Jose, CA, USA, January 2002.

[9] J. Fridrich. A hybrid watermark for tamper detection in digital images. In *ISSPA'99 Conference*, Brisbane, Australia, August 1999.

[10] J. Fridrich. Visual hash for oblivious watermarking. In *IS&T/SPIE Proceedings*, volume 3971, San Jose, California, USA, January 2000.

[11] S.I. Gel'fand and M.S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[12] Z. Grytskiv, S. Voloshynovskiy, and Y. Rytsar. Cryptography and steganography of video information in modern communications. In *TELSIKS'97*, volume 1, pages 164–167, Nis, Yugoslavia, October 1997.

[13] P. Guillon, T. Furon, and P. Duhamel. Applied public-key steganography. In *Proceedings of SPIE 2002*, volume 4675, San Jose, CA, USA, January 2002.

[14] H. Hel-Or, Y. Yitzhaki, and Y. Hel-Or. Geometric hashing techniques for watermarking. In *ICIP 2001*, 2001.

[15] M. Holliman and N. Memon. Counterfeiting attacks on linear watermarking systems. In *Proc. IEEE Multimedia Systems 98, Workshop on Security Issues in Multimedia Systems*, Austin, Texas, June 1998.

[16] M. Holliman and N. Memon. Couterfeting attacks on oblivious block-wise independant invisible watermarking schemes. In *IEEE Trans. on Image Processing*, volume 9, pages 432–441, March 2000.

[17] M. S. Kankanhalli and R. K. R. Ramakrishnan. Content based watermarking of images. In *Multimedia and Security Workshop at ACM Multimedia'98, Bristol, U.K.*, September 1998.

[18] M. Kutter, S. Voloshynovskiy, and A. Herrigel. Watermark copy attack. In *IS&T/SPIE's 12th Annual Symposium, Electronic Imaging 2000: Security and Watermarking of Multimedia Content II*, volume 3971, San Jose, California USA, 23–28 jan 2000.

[19] J.-F. Delaigle M. Bertran and B. Macq. Some improvements to HVS models for fingerprinting in perceptual decompressors. In *ICIP 2001*, pages 1039–1042, Thessaloniki, Greece, October 2001.

[20] University of Geneva Stochastic Image Processing (SIP) Group. SIP Watermarking Technology. http://watermark.unige.ch/wmg_technology.html.

[21] S. Voloshynovskiy, A. Herrigel, N. Baumgaertner, and T. Pun. A stochastic approach to content adaptive digital image watermarking. In *IHW'99*, pages 212–236, September 29 - October 1st 1999.

[22] P. W. Wong. A public key watermark for image verification and authentication. In *ICIP 1998*, volume 1, 1998. MA11.07.

[23] M. M. Yeung and F. C. Mintzer. An invisible watermarking technique for image verification. In *ICIP 1997)*, volume 2, pages 680–683, Washington, DC, USA, October 26-29 1997.