

Influência da Codificação da Marca D'Água Digital em suas características de Robustez ou Fragilidade

Marcos de Castro Pacitti* e Weiler Alves Finamore**

Resumo—Neste artigo é proposta uma nova metodologia para análise e desenvolvimento de técnicas de marcação d'água digital, considerando principalmente suas características de robustez ou fragilidade requeridas em função da aplicação pretendida. A utilização da metodologia introduzida revelou uma importante conclusão: as propriedades dos códigos empregados nas técnicas de marcação d'água digital são determinantes para o estabelecimento do grau de robustez ou fragilidade destas técnicas. Também foi constatado neste artigo as excelentes propriedades do código turbo, principalmente na obtenção da característica de fragilidade do sistema, e é proposta a calibração das técnicas de marcação digital a fim de aproveitar tais propriedades.

Palavras-Chave—Marca d'água digital, fragilidade, robustez, códigos turbo, ganho de codificação.

Abstract—This paper introduces a new methodology for digital watermarking techniques analysis and development mainly considering the required robustness and fragility characteristics dependent on the desired application. The use of the introduced methodology revealed an important conclusion: The codes properties used on digital watermarking techniques are determinants for establishing the robustness or fragility degree of those techniques. This paper also noticed the turbo code excellent properties mainly in obtaining the system fragility characteristic and we propose a calibration on the digital watermarking techniques to benefit from those properties.

Keywords—Digital watermarking, information hiding, information security, fragility, robustness, turbo codes, coding gain.

I. INTRODUÇÃO

A utilização de marca d'água, onde permeia-se a informação (a marca) em um documento hospedeiro de forma a não ser imediatamente percebida e a dificultar a sua reprodução, já vem sendo bastante empregada, objetivando reprimir a falsificação. Estas técnicas já são utilizadas em documentos e dinheiro há séculos. Atualmente, com a ampla utilização da representação digital de documentos, imagens, áudio, e outros sinais, a proteção de direitos de reprodução e autorais utilizando técnicas de marcação d'água digital tornou-se uma área muito ativa de pesquisa (veja extensa bibliografia em [1]). Naturalmente, muitas outras aplicações surgiram nesta nova perspectiva digital do assunto [2], incluindo as aplicações de segurança nacional, como as para verificação de integridade/autenticidade da mídia, para “comunicação camouflada”

Este trabalho foi parcialmente financiado pela Comissão de Implantação do Sistema de Controle do Espaço Aéreo Brasileiro (CISCEA)- Comando da Aeronáutica - Ministério da Defesa.

*Marcos de Castro Pacitti, CISCEA, Rio de Janeiro, Brasil, E-mail: pacitti@cc.sivam.gov.br.

**Weiler Alves Finamore, CETUC / PUC - RJ, Rio de Janeiro, Brasil, E-mail: weiler@cetuc.puc-rio.br

(*covert communication*) da informação, e para “rastreamento de traidor” (*traitor tracing*).

A marcação d'água neste novo contexto é um problema complexo, com questões que envolvem não somente a marcação digital e fatores subjetivos da percepção humana, mas também o projeto sistêmico, segurança, e uma série de aspectos econômicos e legais. Em nosso artigo, tratamos apenas de parte do problema: O projeto da técnica de marcação digital objetivando revestí-la das características específicas de robustez ou fragilidade.

As propriedades de robustez ou fragilidade da técnica de marcação digital correspondem a capacidade da mesma em recuperar ou não, respectivamente, a marca d'água sob determinadas condições de ataque (ruído) na mídia hospedeira.

Num projeto, estas propriedades são consideradas em função da aplicação pretendida. Normalmente, a robustez é requisito para as aplicações envolvendo garantia de direitos sobre a mídia hospedeira, e a fragilidade, nas aplicações relacionadas a verificação da integridade e autenticidade da mesma.

Visando orientar o projeto de técnicas de marcação digital, neste artigo propomos uma nova abordagem metodológica no entendimento e análise das características de robustez e de fragilidade, possibilitando o desenvolvimento de um modelo analítico do problema.

A utilização do novo modelo, possibilitou verificarmos que a obtenção das características de robustez e de fragilidade da técnica de marcação digital são fortemente determinadas pela codificação aplicada na marca d'água, indicando que a utilização de códigos visa não apenas proporcionar ao sistema o conhecido ganho de codificação [4]. Assim, o desenvolvimento de códigos para marcação digital deve considerar os compromissos entre os requisitos do ganho de codificação e de robustez (ou fragilidade).

Na seção II, apresentamos o modelo do problema de marcação d'água digital, incluindo a descrição dos parâmetros envolvidos na análise e no projeto das técnicas de marcação. Na seção III, apresentamos nossa proposta de modelamento das características de robustez e fragilidade da técnica de marcação digital. Na seção IV identificamos a forte influência da codificação da marca d'água nestas características, e constatamos as excelentes propriedade dos códigos turbo para emprego objetivando a obtenção de técnicas frágeis, e na seção V propomos a calibração destas técnicas para beneficiarem-se desta influência.

II. MODELO DO PROBLEMA

O problema de modulação ou marcação d'água binária é genericamente descrito na fig. 1. Nesta figura o sinal hospedeiro é representado por um vetor $\mathbf{x} \in \mathbb{R}^N$.¹ O bit de informação b é codificado, e o sinal hospedeiro \mathbf{x} é modulado pelos bits codificados gerando o vetor $\mathbf{s}(b, \mathbf{x}) \in \mathbb{R}^N$ na saída do modulador.² A taxa de marcação é de $1/N$ bits de informação por dimensão (amostra do sinal hospedeiro), e o vetor \mathbf{s} normalmente possui uma restrição de distorção.

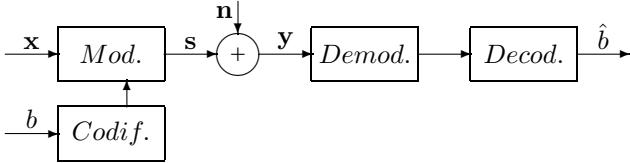


Fig. 1. Modelo do problema de marcação/estimação do bit de informação b no hospedeiro \mathbf{x} .

A distorção (D_b) do sinal \mathbf{s} , em relação a \mathbf{x} , provocada pela modulação com o bit b , é definida por

$$D_b = \frac{1}{N} \sum_{i=1}^N (s_i - x_i)^2 \quad (1)$$

assim, podemos considerar que a energia do bit (E_b) necessária para provocar a correspondente distorção é

$$E_b = ND_b \quad (2)$$

O sinal marcado \mathbf{s} está sujeito a diversas manipulações comuns de processamento de sinal e de tentativas voluntárias de remover a informação marcada. Estas manipulações são denominadas de ruído ou ataque, e também estão restritas pela distorção provocada no sinal. Para efeitos de análise, normalmente considera-se que o ataque é resumido a um ruído gaussiano aditivo $\mathbf{n} \in \mathbb{R}^N$ cujas componentes são descorrelacionadas, possuem média nula e variância $\sigma_n^2 = N_0/2$. Assim, na entrada do demodulador temos o sinal $\mathbf{y} = \mathbf{s} + \mathbf{n}$. Supondo o ruído \mathbf{n} estatisticamente independente em relação às demais variáveis, temos que a distorção no sinal \mathbf{y} em relação a \mathbf{x} é

$$D_y = D_b + \sigma_n^2. \quad (3)$$

O demodulador e o decodificador processam o sinal \mathbf{y} em função da técnica de modulação e de codificação utilizadas, e o decodificador fornece como saída uma estimativa \hat{b} do bit de informação b . É interessante notar que o modelo apresentado é análogo a um sistema de comunicação onde há um sinal interferente \mathbf{x} e um ruído aditivo \mathbf{n} .

Na Tabela I introduz-se algumas notações a fim de padronizar os parâmetros utilizados na análise. Nesta tabela $\sigma_x^2 = E[x_i^2]$, e foi considerado que $E[x_i] = 0$. Note que

¹O vetor \mathbf{x} é qualquer representação conveniente de todo ou parte do sinal hospedeiro. No caso do sinal hospedeiro ser uma imagem, o vetor \mathbf{x} pode possuir componentes correspondentes a valores selecionados de pixels ou coeficientes de uma transformada (DCT ou DWT, por exemplo).

²A modulação do vetor \mathbf{x} corresponde a alterações nos valores de suas componentes. Algumas técnicas conhecidas de modulação são a SS [5], QIM[6], STDM[6] e QP[7].

TABELA I

NOTAÇÃO DOS PARÂMETROS DE ANÁLISE

Parâmetro	Definição
Razão marca-ruído	$WNR = \frac{D_b}{\sigma_n^2}$
Razão marca-ruído normalizada	$WNR_N = \frac{E_b}{\sigma_n^2}$

$WNR_N = 2E_b/N_0$, ou equivalentemente, WNR_N (dB) = E_b/N_0 (dB) + 3 dB

Sugere-se ao leitor a consulta da referência [4] para maiores esclarecimentos quanto aos aspectos e limitantes envolvidos no projeto de um sistema de marcação d'água digital.

III. MODELAMENTO DA ROBUSTEZ E FRAGILIDADE EM UM SISTEMA DE MARCAÇÃO D'ÁGUA DIGITAL

A robustez representa a capacidade de um sistema de marcação digital suportar ataques (ruído) sem que a marca d'água seja afetada, isto é, ainda ser possível detectá-la (recuperá-la) mesmo na presença do ataque. Desta forma, um requisito de projeto é estabelecer um limiar onde ainda faz sentido (valor perceptivo humano da mídia) recuperar a marca d'água. Desta forma, a fim de parametrizar nosso projeto, definimos D^T como a **distorção total** que o sinal hospedeiro original (\mathbf{x}) pode sofrer, sem que o mesmo perca o valor perceptivo. Neste caso, o sistema de marcação é projetado de maneira que a marca d'água “resista” ao ataque desde que $D_y < D^T$. Esta resistência é definida pela **probabilidade de erro admissível** (p_e^a) na recuperação da marca, isto é, a maior BER que admite-se na recuperação (detecção) da marca d'água. Assim, temos o segundo parâmetro de projeto (p_e^a), significando que, para a condição de robustez, enquanto o ataque (σ_n^2) não provocar uma distorção total superior a D^T , deverá ser possível recuperar a marca d'água com $BER < p_e^a$.

Já a fragilidade de um sistema de marcação, representa a capacidade do mesmo “perder” a capacidade de detecção ($BER > p_e^a$) da marca d'água antes da distorção provocada pelo ataque (D_y) ultrapassar a distorção total D^T suportável pelo sinal hospedeiro. Isto significa que a BER na recuperação da marca deve ultrapassar o mínimo admissível (p_e^a), ainda com $D_y < D^T$, isto é, **sem** que o sinal hospedeiro esteja criticamente corrompido.

Para conduzir uma análise mais detalhada do problema, é adequado expressar os condicionantes do mesmo. Primeiramente, deve-se considerar a curva de desempenho do sistema empregado, que é representada por uma função ($p_e = f(WNR_N)$) decrescente. Na referência [4] pode-se verificar o comportamento deste desempenho para algumas técnicas de marcação lá consideradas. Em seguida, considerando os parâmetros de projeto definidos, estabelece-se a condição de máxima distorção,

$$D_y = D_b + \sigma_n^2 < D^T, \quad (4)$$

e a condição da probabilidade de erro admissível na recuperação (detecção) da marca d'água,

$$BER = p_e < p_e^a . \quad (5)$$

Definindo o **fator de robustez** $\rho = D_b/D^T$, da equação 4 obtem-se

$$WNR_N > N\rho/(1 - \rho) = WNR_N^{D^T} , \quad (6)$$

onde $WNR_N^{D^T}$ é a razão marca-ruído em que ocorre a distorção total suportável ($D_y = D^T$). Já da equação 5 obtem-se

$$WNR_N > f^{-1}(p_e^a) = WNR_N^a , \quad (7)$$

onde WNR_N^a é a razão marca-ruído que representa o limiar de perda de detecção da marca d'água ($BER = p_e^a$). Assim, as equações 6 e 7 representam os condicionantes para o projeto.

Na situação inicial ideal (ausência de ruído), a razão marca-ruído é infinita. A medida em que o ataque (ruído) é aplicado, esta razão marca-ruído é reduzida. A dinâmica como WNR_N^a e $WNR_N^{D^T}$ são alcançados, em virtude do incremento do ataque, estabelece as condições de robustez ou fragilidade do sistema de marcação.

Para facilitar a análise que segue, na fig. 2 representamos uma típica curva de desempenho de um sistema de marcação digital. No caso, foi considerado a marcação não codificada ideal discutida em [4]. Nesta figura também foi arbitrado, para fins ilustrativos, um ponto, $(WNR_N = WNR_N^a, p_e = p_e^a)$, que representa a condição de limiar de “perda” de detecção da marca d'água.

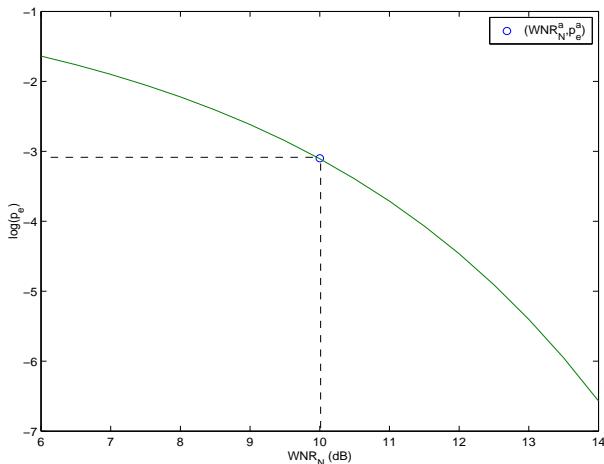


Fig. 2. Curva de desempenho para a marcação não codificada ideal [4], com representação de ponto arbitrado (WNR_N^a, p_e^a) .

A escolha do fator de robustez ($0 < \rho < 1$) e do número de amostras N utilizadas na modulação determinarão o valor de $WNR_N^{D^T}$ do sistema de marcação. Temos duas situações a serem consideradas:

- $WNR_N^{D^T} > WNR_N^a$: Neste caso, é possível incrementar o ataque, provocando a distorção do sinal hospedeiro, além da situação crítica ($D_y > D^T$), e ainda garantir a recuperação (detecção) da marca d'água, até $WNR_N =$

WNR_N^a , quando o sinal hospedeiro já encontra-se severamente distorcido. Esta situação corresponde ao sistema de marcação robusto, e sua principal aplicação visa garantir os direitos sobre a mídia hospedeira.

- $WNR_N^{D^T} < WNR_N^a$: Neste caso, o incremento do ataque provoca a perda da capacidade de detecção da marca d'água ($p_e < p_e^a$) antes do sinal hospedeiro ser considerado criticamente distorcido ($D_y < D^T$). Esta situação corresponde ao sistema de marcação frágil, que objetiva provocar a perda da marca d'água sob qualquer tentativa de adulteração do sinal hospedeiro, sendo que sua principal aplicação visa a garantia da autenticidade e integridade da mídia hospedeira.

Assim, o parâmetro $\Delta WNR_N = WNR_N^{D^T} (dB) - WNR_N^a (dB)$ define o tipo do sistema de marcação. Para $\Delta WNR_N > 0$, tem-se o sistema robusto, e para $\Delta WNR_N < 0$, o sistema frágil. E o valor $|\Delta WNR_N|$ representa um grau de robustez ou fragilidade do sistema. Outro grau de robustez ou fragilidade de uma técnica de marcação d'água digital a ser considerado, é a sensitividade das variações na BER percebida na detecção como resultado das variações incrementais do ataque, isto é, da “rapidez” em que a p_e^a é alcançada em virtude do incremento do ataque (σ_n^2). Esta dinâmica será abordada na próxima seção.

IV. INFLUÊNCIA DA CODIFICAÇÃO NAS CARACTERÍSTICAS DE ROBUSTEZ OU FRAGILIDADE DA TÉCNICA DE MARCAÇÃO D'ÁGUA DIGITAL

Naturalmente, quando empregam-se códigos correctores de erros em um sistema de marcação digital, objetiva-se reduzir a razão marca-ruído (WNR_N ou E_b/N_o), para uma pretendida probabilidade de erro de operação. Esta redução é denominada de ganho de codificação [4]. Contudo, como veremos, a característica da curva de desempenho do código será determinante na obtenção dos requisitos de fragilidade ou robustez, indicando que o em emprego de códigos destina-se não somente a promover o citado ganho de codificação.

Um grau de robustez de um sistema de marcação é entendido como a capacidade do mesmo em impedir que uma variação incremental do ataque (ruído) provoque grande variação (redução) na BER (p_e) na detecção da marca d'água. Assim, neste caso, deseja-se que a curva de desempenho $p_e = f(WNR_N)$ possua baixa inclinação para $WNR_N > WNR_N^a$, indicando que a perda da capacidade de detecção da marca d'água ($BER > p_e^a$) será provocada somente por um grande “esforço” de ataque.

Da mesma maneira, um grau de fragilidade de um sistema de marcação digital também pode ser entendido como a capacidade do mesmo em garantir que pequenos incrementos no ataque provoquem uma grande degradação da BER. Assim, nesta situação, deseja-se que a curva de desempenho $p_e = f(WNR_N)$ possua alta inclinação para $WNR_N > WNR_N^a$, indicando que a perda da capacidade de detecção da marca d'água ($BER > p_e^a$) será provocada por um pequeno incremento do ataque. Note que sistemas assim projetados também fornecem uma excelente ferramenta de graduação (medição) da intensidade do ataque em função das variações da BER percebidas na detecção.

Assim, as característica de robustez ou fragilidade em um sistema de marcação d'água digital são obtidos diretamente das propriedades da curva de desempenho $p_e = f(WNR_N)$. Estas propriedades são basicamente definidas pela técnica de marcação não codificada utilizada e, principalmente, pela codificação empregada. Desta forma, concluímos que, nestas circunstâncias, o desenvolvimento ou a escolha do código a ser empregado deve considerar não apenas o ganho de codificação, mas também a forte influência do mesmo na obtenção dos requisitos de robustez ou fragilidade desejados para a operação do sistema de marcação.

Particularmente, o código turbo [3] apresenta propriedades que permitem projetar um sistema de marcação robusto ou frágil, pois possui dois intervalos com características bem definidas. Um intervalo de alta inclinação para baixos WNR_N , e outro de baixa inclinação para médios e altos WNR_N , permitindo operação tanto frágil quanto robusta, respectivamente. A fig. 3 ilustra uma típica curva de desempenho de um código turbo.

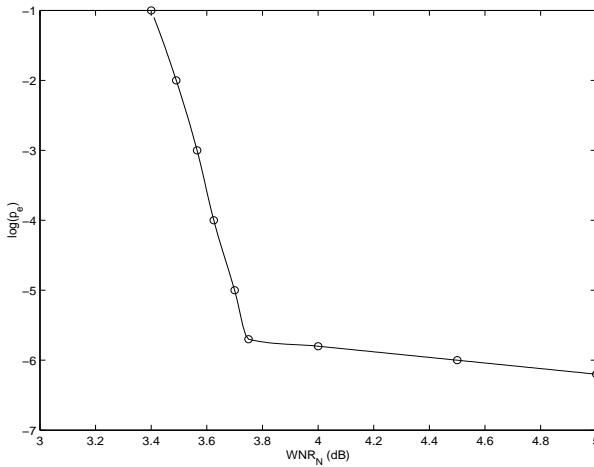


Fig. 3. Curva de desempenho típica de um código turbo com taxa 1/2.

É interessante constatar que, para a operação na região de alta inclinação (sistemas frágeis), uma variação menor do que 3% em WNR_N (ou E_b/N_0) provoca uma redução de mais de 10 vezes na BER, correspondendo a uma sensibilidade ($\Delta p_e\%/\Delta WNR_N\%$) superior a 300. Já a sensibilidade utilizando a técnica não codificada da fig. 2 é de aproximadamente 20. Desta forma, o código turbo pode aumentar em aproximadamente 15 vezes a sensibilidade (fragilidade) do sistema de marcação digital.

Na condição de robustez, objetiva-se operação na região de baixa sensibilidade, e da mesma forma, é possível verificar que a marcação não codificada possui a sensibilidade 2 vezes maior do que a sensibilidade quando empregado o código turbo na região de baixa inclinação, indicando que o código turbo também é vantajoso para aplicação quando deseja-se a robustez do sistema. Contudo, é possível que, para o caso robusto, a investigação de adaptações no projeto dos códigos turbo, ou de outro códigos, fornecerá uma melhor condição de robustez, isto é, uma menor inclinação na curva de desempenho do código, resultando na redução da sensibilidade.

Assim, as considerações de projeto de códigos turbo (características dos códigos RSC - "Recursive Systematic Convolutional" - componentes, estrutura do entrelaçador e número de iterações na decodificação), e de outros códigos, podem agora ser revistas tendo em vista a nova finalidade de aplicação proposta neste artigo.

Vale comentar que a presente proposta, de graduar a intensidade do ataque em função de variações da BER, também aplica-se às técnicas semi-frágeis, em que um critério de avaliação "softer" é desejado, e às técnicas híbridas, onde duas marcas, uma frágil e outra robusta, são empregadas [8]. Ainda, vale lembrar que a análise realizada limitou-se ao modelo de ataque gaussiano, assim, seu emprego, para aplicações onde outros tipos de ataque são considerados, deverá ser investigado quanto à possibilidade de extensão da metodologia apresentada na presente proposta. Finalmente, observamos, conforme já mencionado, que o processo de marcação d'água digital é complexo e envolve várias etapas. Uma importante etapa, não considerada neste artigo, é revestir o sistema com o nível de segurança adequado para a aplicação desejada, onde o emprego de técnicas criptográficas e a necessidade de gerência e geração de chaves são tratados.

V. CALIBRAÇÃO PARA SISTEMAS DE MARCAÇÃO D'ÁGUA DIGITAL FRÁGEIS

Considerando o emprego do código turbo cujo desempenho é representado na fig. 3 para obtenção de sistemas de marcação frágeis, surge o problema de como garantir a operação imediata na região de alta sensibilidade. Para este caso, propomos a calibração no modulador, que corresponde a adição de um *ruído de polarização* ao sinal hospedeiro (x) de forma a polarizar a marca d'água com $WNR_N = WNR_N^p$, onde WNR_N^p é a razão marca-ruído polarizada na "região frágil" do código. Assim, a intensidade do *ruído de polarização*, $\sigma_{n_p}^2$, é dada por

$$\sigma_{n_p}^2 = \rho D^T / WNR_N^p . \quad (8)$$

Por exemplo, no código turbo da fig. 3 sob análise, a polarização de WNR_N em 3.7 dB seria bastante adequada. Outros códigos podem ser investigados (ou mesmo outros tipos de códigos turbo) a fim de que a região de alta inclinação ocorra em valores de WNR_N maiores, possibilitando que o *ruído de polarização* seja menor, e provocando menor distorção no sinal hospedeiro.

Caso deseje-se apenas verificar (medir) a intensidade do ataque (adulteração do sinal hospedeiro) através da variação percebida na BER, basta adicionarmos o *ruído de polarização* somente na detecção (ao sinal y). Assim, a distorção provocada pelo *ruído de polarização* no sinal hospedeiro x é eliminada.

Desta forma, concluímos que a calibração para operação na região frágil do código pode ser conduzida no modulador ou no demodulador, dependendo da aplicação.

Como sugestão de pesquisas futuras para outras aplicações dos códigos correctores de erros, vale mencionar que a alta sensibilidade em ($\Delta p_e\%/\Delta(E_b/N_0)\%$) dos códigos turbo permite-nos propor sua aplicação em sensores tendo canais

de transmissão como transdutores (sensores a fibra óptica, por exemplo), onde a variação da grandeza física a ser monitorada provoque alteração no sinal que se propaga no canal, e consequentemente no valor de E_b/N_o , resultando em significativa variação da BER na detecção. Assim, variações de grandezas físicas podem ser monitoradas por variações da BER na detecção.

VI. CONCLUSÕES

Neste artigo, propusemos uma nova metodologia para balizar o desenvolvimento e análise de técnicas de marcação d'água digital codificada, considerando principalmente as necessárias características de robustez ou fragilidade destas técnicas. A utilização desta metodologia possibilitou derivarmos um importante resultado: A dominante influência das propriedades dos códigos corretores de erro na obtenção das características de robustez ou fragilidade de um sistema de marcação digital. Este é um resultado singular tendo em vista que normalmente os códigos são empregados apenas objetivando o ganho de codificação (operação próxima a capacidade do sistema), e sugere a revisão de projeto dos códigos existentes, ou a investigação de novos códigos, que objetivem a otimização das características de robustez ou fragilidade das técnicas de marcação d'água digital. Em particular, foi verificado que o código turbo proporciona uma excelente característica de fragilidade para estas técnicas, e foi proposta a calibração das mesmas para aproveitar tal característica.

REFERÊNCIAS

- [1] R. J. Anderson and F. A. Petitcolas, "Information hiding: an annotated bibliography" [online]. Available: www.petitcolas.net/fabien/steganography/bibliography/, 1999.
- [2] M. D. Swanson, M. Kobayashi, and A. H. Tewfik, "Multimedia data-embedding and watermarking technologies" *Proc. IEEE*, v. 86, p. 1064-1087, June 1998.
- [3] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes" in *ICC'93*, Geneva, Switzerland, p. 1064-1070, May 1993.
- [4] M. C. Pacitti, W. A. Finomore, "Limitante Inferior da Marcação D'Água Digital não Codificada" *submetido ao SBT'04*, Belém, Brasil, Set. 2004.
- [5] I. J. Cox, J. Killian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia" *IEEE Trans. Image Processing*, v. 6, p. 1673-1687, Dec 1997.
- [6] B. Chen and G. W. Wornell, "Quantization Index Modulation: A class of provably good methods for digital watermarking and information embedding" *IEEE Trans. Inform. Theory*, v. 47, No. 4, p. 1423-1443, May 2001.
- [7] F. Perez-González and F. Balado, "Quantized projection data hiding" *In Proc. of the IEEE International Conference on Image Processing (ICIP)*, Rochester (NY), USA, September 2002.
- [8] J. Fridrich, "Methods for Tamper Detection in Digital Images," *ACM Workshop on Multimedia and Security*, Orlando, FL, October 30-31, 1999, pp. 19-23.