C1

Método de Inserção de Marcas d'Água Robustas em Sinais Digitais

por
Dimitri do Brasil DeFigueiredo

Dissertação de Mestrado apresentada à Universidade Estadual de Campinas – Unicamp, Faculdade de Engenharia Elétrica e de Computação, como requisito parcial para a obtenção do grau de Mestre em Engenharia Elétrica

Orientador:
Prof. Dr. Max Henrique Machado Costa

Banca Examinadora:
Prof. Dr. Max Henrique Machado Costa – FEEC/Unicamp
Prof. Dr. Hae Yong Kim – EPUSP/USP
Prof. Dr. Jaime Portugheis – FEEC/Unicamp
Prof. Dr. Reginaldo Palazzo Jr. – FEEC/Unicamp

Campinas, São Paulo
Julho de 2001

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

D361m

DeFigueiredo , Dimitri do Brasil
Método de inserção de marcas d'água robustas em sinais digitais / Dimitri do Brasil DeFigueiredo. --Campinas, SP: [s.n.], 2001.

Orientador: Max Henrique Machado Costa.
Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Teoria da codificação. 2. Códigos de controle de erros (Teoria da codificação). 3. Criptografia. 4. Propriedade intelectual. 5. Direitos autorais. 6. Marcas d'água. I. Costa Max Henrique Machado. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Resumo

Este trabalho tem como objetivo o estudo de marcas d'água digitais robustas. Uma marca d'água digital robusta não pode ser facilmente retirada sem comprometer a qualidade do sinal no qual está inserida. A inserção de uma marca d'água robusta em um sinal digital pode ser utilizada para estabelecer uma ligação permanente entre o sinal digital e seu proprietário. Em um sistema de marca d'água digital assimétrico a capacidade de detectar uma marca não implica na capacidade de retirá-la. Estes sistemas têm, portanto, uma maior flexibilidade. Neste trabalho são estabelecidas algumas considerações importantes para a derivação de um sistema de marca d'água robusto e assimétrico. Em seguida, utilizando-se das mesmas e da distância de Hamming como medida de distorção, propõe-se um método assimétrico de inserção de marcas d'água. O método utiliza conhecimento do sinal para transmitir a maior quantidade de informação possível. São derivadas algumas condições que quando satisfeitas tornam o método seguro contra qualquer forma de ataque.

Abstract

A robust digital watermark cannot be removed from the marked signal without compromising its quality. Thus, these watermarks can be used to establish a permanent link between the watermarked signal and its owner. Asymmetric watermarking schemes have a wider range of applications than symmetric ones. In such systems, the ability to detect the watermark does not provide the means which would enable an attacker to also remove the watermark. In this thesis, a few rules-of-thumb for robust asymmetric watermarking schemes are established and a new scheme is proposed. The Hamming distance is chosen as the distortion measure throughout. The proposed scheme uses information about the signal to be watermarked to maximize the embedded information rate. Also, the necessary conditions for the scheme to become secure against an arbitrary distortion attack are established.

Agradecimentos

Gostaria de agradecer a todos que, direta ou indiretamente, contribuíram para a conclusão deste trabalho. Um reconhecimento especial à minha irmã, Danielle, pelo apoio incondicional em todos os momentos.

Agredeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro.

Índice

Lista de Figuras

Capítulo 1

Introdução

A idéia de propriedade já está sedimentada na sociedade moderna e faz parte do dia a dia, não é em geral questionada e parece bastante natural. Com esta idéia surge a noção de um bem, que até um passado relativamente recente era um conceito atribuído a um objeto do mundo físico: uma casa, um terreno, um casaco, etc. No fim do século XIX, houve uma ampliação do conceito de propriedade para incluir também o que é agora denominado propriedade intelectual. A propriedade intelectual estende o conceito de propriedade a entidades que não pertencem ao mundo físico: uma música, um texto, uma invenção, etc. Uma característica comum dos bens advindos da propriedade intelectual é que todos se baseiam em informação. No caso de um texto por exemplo, não são as páginas de papel que se constituem em objeto de propriedade intelectual mas sim a forma com que as palavras são utilizadas para comunicar idéias. Por exemplo, o texto só pode ser recitado em uma propaganda na televisão com a autorização da editora ou do autor, isto é, de seu proprietário.

Esta característica comum dos bens advindos da propriedade intelectual faz com que os mesmos sejam diferentes dos demais bens em vários aspectos importantes. Estes bens, que serão chamados de agora em diante de bens intelectuais, não possuem por exemplo um "espaço físico ocupado". Outra diferença entre os bens materiais e os bens intelectuais está ligada ao conceito de roubo. Para quase a totalidade dos bens materiais existentes, após a ocorrência de um roubo, o proprietário não pode mais usufruir do bem. Isto é radicalmente diferente do que ocorre com bens intelectuais, onde após a ocorrência do furto, que pode ser a simples duplicação de informação, o proprietário continua sendo capaz de utilizar o bem plenamente. Fazer uma cópia idêntica, a partir do nada, de um carro é considerado uma coisa muito boa, seria ótimo se pudéssemos fazê-lo; fazer uma cópia idêntica, a partir do nada, de um programa de computador é crime.

Contudo, com o atual desenvolvimento tecnológico ficou muito fácil duplicar informação que se encontre em formato digital, ou seja, informação que se encontra

armazenada de forma que possa ser utilizada por computadores ou outros equipamentos digitais. Aos bens intelectuais que podem ser colocados desta forma chamamos bens digitais. Como é muito fácil duplicar bens digitais e a duplicação é em geral considerada roubo, é fácil roubá-los. Logo, fazer com que a lei seja cumprida neste caso é uma tarefa árdua. A situação é agravada ainda mais por dois fatores: como os bens estão em forma digital a duplicação dos mesmos gera cópias idênticas e, além disto, as cópias podem ser enviadas para qualquer lugar através, *e.g.*, através da Internet. Há alguns anos, quando isto não era ainda verdade, o problema não era considerado tão sério. Como exemplo das mudanças ocorridas, pode-se citar o recente choque causado pela Napster que levou a indústria fonográfica a tomar medidas legais enérgicas para proteger seus bens.

Grande quantidade de recursos está sendo utilizada para a criação de bens digitais. Supondo que as atuais leis de propriedade intelectual promovam o bem estar comum através do incentivo para a criação de um número cada vez maior destes bens, é do interesse de todos que estas leis sejam cumpridas. Mecanismos capazes de fazer com que todos cumpram as leis são então de grande valia para o bem comum. A procura por tais mecanismos é a força motriz que impulsiona uma série de novas pesquisas relacionadas ao tratamento de informação.

No contexto de mecanismos que ajudem a se fazer cumprir as leis de propriedade intelectual é que as pesquisas na área de *marca d'água digital* tem se destacado. Uma marca d'água digital é uma pequena quantidade de informação inserida em um bem digital. Esta pequena quantidade de informação pode ter qualquer significado, mas o grande atrativo é fazer com que seja feita uma ligação entre o proprietário e o bem. Um exemplo simples são os avisos de direitos autorais que se encontram impressos nas contracapas de livros. Tais avisos estabelecem quem é o proprietário dos direitos autorais da obra. Outro exemplo são as pequenas marcas semi-transparentes que podem atualmente ser vistas em um dos cantos das imagens de televisão. Da mesma forma, é possível se inserir uma marca d'água em um programa de computador que especifique quem é o dono do programa. Quando o programa for copiado, a marca também o é e a ligação entre proprietário e bem não é perdida.

Para as leis de propriedade intelectual uma marca d'água digital, que é um mecanismo capaz de fazer uma ligação entre proprietário e bem digital, só é útil se tiver validade legal, isto é, só se puder ser utilizada como evidência perante um tribunal. Para

isto, é necessário que a marca d'água digital seja idônea, no sentido de não beneficiar em uma dada situação qualquer uma das partes envolvidas em uma disputa. Por exemplo, em um transação comercial é feita a compra de 1 Kg de ouro puro. Contudo, o comprador após receber a mercadoria percebe que foi enganado e que a quantidade entregue é inferior a 1 Kg. Quando a quantidade entregue é apresentada perante um tribunal, o juiz deve ser capaz de medir exatamente o peso do ouro e para isto deve se valer de uma balança "idônea". Se a única balança que puder ser utilizada estiver descalibrada e sempre indicar um peso inferior ao real, a pesagem na balança não será considerada confiável pela corte. Portanto, a utilização da balança perde sua utilidade se a balança não for idônea. Similarmente, a utilização de uma marca d'água digital para estabelecer o proprietário de um bem digital perde sua utilidade se a marca d'água não for idônea.

Além de ser idônea, uma marca d'água digital também deve ser difícil de se remover, para ser útil a um tribunal. No exemplo acima, a pesagem da quantidade de ouro vendida só é valida se, após o compra, a quantidade de ouro não foi adulterada. Assim como é preciso garantir que o comprador de ouro não tirou pedaços do material após a compra, adulterando a quantidade de ouro vendida, também é preciso garantir que uma marca d'água digital não possa ser adulterada posteriormente à sua inserção. A grande maioria dos sistema de marca d'água digital disponíveis atualmente satisfaz a esta condição escondendo a marca. Não é uma tarefa fácil se remover a marca d'água por que não se sabe exatamente onde a mesma se encontra. Todavia, isto implica que somente o proprietário do bem digital sabe onde está a marca, o que por sua vez restringe a habilidade de terceiros de detectar o roubo de um bem digital. Por exemplo, uma apresentação ao vivo da Terceira Sinfonia de Beethoven pode ser gravada e ter seus direitos de reprodução adquiridos[1]. Supõe-se que, em seguida, uma marca d'água digital é inserida e a música é colocada à venda em um CD. Como existem várias interpretações da Terceira Sinfonia, ao adquirir uma tal interpretação, um ouvinte não tem como saber se trata-se de material "autêntico" ou não. Mas se a música é marcada com uma marca d'água digital que possa ser lida, então o ouvinte pode facilmente saber se quem vendeu o CD é a mesma pessoa que detêm os direitos de reprodução da música. Se uma reprodução da Terceira Sinfonia de Beethoven

[1] Embora ninguém possua direitos autorais sobre a obra de Beethoven é possível obter direitos sobre uma interpretação desta obra.

com a marca da Orquestra Filarmônica de Berlim é encontrada na loja da Orquestra Real inglesa, algo deve estar errado.

Neste trabalho é proposto um método para a inserção e detecção de marcas d'água digitais que tem como intuito satisfazer às necessidades mencionadas. Ao contrário da maioria dos métodos encontrados na literatura, o método proposto aplica-se principalmente quando a informação é representada de forma discreta. Além disto, em alguns casos a segurança da marca inserida é garantida no sentido de que, para extrair a marca, é necessário deteriorar muito a qualidade do bem digital no qual ela está inserida. O trabalho está organizado como se segue:

No segundo capítulo é feito um breve histórico dos diversos tipos de marca d'água existentes. São consideradas as diversas aplicações para marcas d'água digitais assim como as diversas formas de classificação das mesmas. A notação e o modelo matemático a ser utilizado no restante do texto são introduzidos e atenção especial é dada ao tratamento de marcas d'água com detecção cega. O Capítulo 3 é focado na derivação do método de inserção e detecção de marcas d'água digitais apresentado neste trabalho. A motivação intuitiva é apresentada seguida do método propriamente dito. Considerações a respeito da aplicação do sistema proposto também são apresentadas. Em seu fechamento, o Capítulo investiga a possibilidade de possíveis ataques ao método e avalia o desempenho do mesmo. Finalmente, no Capítulo de Conclusões são feitas algumas considerações finais e sugestões para o prosseguimento dos estudos.

A bibliografia utilizada é listada no fim por ordem alfabética de autor. Interessantemente, grande parte do material utilizado encontra-se disponível na Internet.

Capítulo 2

Considerações Preliminares

O tratamento científico da inserção de marcas d'água digitais é uma disciplina muito recente que tem atraído grande atenção de pesquisadores no últimos anos. O número de trabalhos publicados na área mais que duplicou a cada ano entre 1993, onde somente 2 trabalhos foram publicados, e 1998, quando houve 103 publicações internacionais [25]. Também recentemente, em 1996, foi realizado o primeiro workshop internacional sobre o assunto[2]. Devido à sua breve história, esta área de estudo encontra-se ainda permeada por outras disciplinas. O estudo de marcas d'água digitais utiliza técnicas de Processamento Digital de Sinais e de Criptologia[3] em conjunto, havendo também um estreito relacionamento com Teoria de Informação e Codificação.

Devido ao seu estreito relacionamento com várias disciplinas correlatas, o estudo de marcas d'água digitais é muitas vezes confundido com a Criptologia ou com a Esteganografia. Contudo, estas três disciplinas vêm a satisfazer diferentes requisitos. A Criptologia pode ser divida em duas áreas: a Criptografia, que estuda formas de codificar mensagens para que só possam ser lidas por pessoas autorizadas, e a Criptanálise, que concentra-se em conseguir decifrar tais mensagens "secretas". Na Criptologia é muito comum a utilização do modelo de comunicação entre duas "pessoas" quaisquer A e B, normalmente apelidadas de "**A**lice" e "**B**ob". Em um cenário padrão, Alice quer enviar uma mensagem M a Bob, mas não quer que ninguém além de Bob possa ler a mensagem. Para enviar a mensagem de forma segura Alice utiliza-se de técnicas de Criptografia e codifica (cifra) a sua mensagem M obtendo uma mensagem cifrada $C(M)$. Por outro lado, uma pessoa não autorizada, **C**arlos, pode querer ler a mensagem M a partir da mensagem cifrada $C(M)$ e para isto usa as técnicas da Criptanálise. Todos os participantes neste modelo de comunicação sabem da existência da mensagem $C(M)$, e a segurança do sistema se baseia na dificuldade de Carlos obter a mensagem original M, a partir da mensagem cifrada $C(M)$.

[2] *Information Hiding: First International Workshop,* Berlin, Alemanha, Maio 1996.

Se para obter a mensagem *M* a partir da mensagem cifrada *C(M)*, Carlos tiver que fazer um enorme esforço computacional, então o sistema é computacionalmente seguro. Por outro lado, se for impossível para Carlos obter *M* a partir de *C(M)* independentemente do esforço computacional utilizado, então o sistema é incondicionalmente seguro. Sistemas incondicionalmente seguros são seguros contra atacantes que tem capacidade computacional ilimitada.

Um problema esteganográfico clássico é o problema dos prisioneiros apresentado por Simmons [27]. Neste problema Alice e Bob encontram-se presos em celas separadas de uma prisão e estão arquitetando um plano de fuga. Contudo, para traçar os detalhes do plano Alice e Bob precisam se comunicar e toda a sua comunicação passa pelo carcereiro, Carlos. Se Carlos descobrir uma mensagem, mesmo que cifrada em qualquer uma das mensagens trocadas entre Alice e Bob, Carlos irá frustrar os planos de fuga colocando Alice e Bob na solitária. É necessário que Alice e Bob consigam esconder mensagens sobre o seu plano de fuga em mensagens que não irão despertar a suspeita de Carlos. Supõe-se que o mecanismo utilizado para esconder as mensagens é conhecido pelo carcereiro. Assim a segurança do sistema deve depender exclusivamente de uma senha secreta que somente Alice e Bob conhecem. *Observa-se que a Criptografia visa manter secreto o conteúdo de uma mensagem, por outro lado, a esteganografia visa manter secreta a própria existência da mensagem.*

As técnicas para obter informação sobre uma mensagem tais como: o remetente, o destinatário, o tamanho da mensagem, quando foi enviada, etc; são normalmente vistas como sendo parte da disciplina de Análise de Tráfego [1]. Técnicas deste tipo são empregadas, por exemplo, para descobrir quais são as ligações de uma organização criminosa. Como a Esteganografia visa esconder a própria existência de uma mensagem, pode-se classificar a Esteganografia como um ramo da Análise de Tráfego [1]. Para poder esconder mensagens as técnicas esteganográficas visam colocar uma mensagem dentro de outra que seja inócua. Assim, se Alice desejar mandar uma mensagem com o número do seu cartão de crédito para Bob, ao invés de enviar apenas estes dados, Alice pode enviar a Bob uma coleção de fotografias digitais, dentro das quais se encontra inserido, secretamente, o número de seu cartão de crédito. Embora Carlos queira saber qual o

[3] Utiliza-se aqui a nomenclatura moderna *Criptologia* ao invés da forma mais usual *Criptografia*.

número do cartão de crédito e esteja monitorando toda a comunicação, a mensagem está segura pois, aparentemente, não é nada além de algumas fotografias.

Uma outra aplicação, os sistemas de marca d'água digital para a proteção de direitos de propriedade intelectual tem o intuito de proteger a reprodução não autorizada de um bem digital. E em um contexto mais amplo onde marcas d'água digitais são inseridas com outros propósitos, a marca é inserida em um *sinal* digital específico. O sinal constitui, sob a nossa óptica, nada mais que um certo número de bits que pode ser uma música, um vídeo, um software, ou qualquer outra informação de valor. Este sinal em sua forma original é designado neste trabalho como o *sinal hospedeiro*. Como existe um grande número de gravadoras, editoras, estúdios de TV e demais entidades que detêm direitos autorais, assim como várias outras aplicações para marcas d'água digitais, cada entidade deve ser capaz de colocar a sua *marca* em qualquer sinal hospedeiro que venha a produzir, gerando um *sinal marcado*. Neste trabalho, entidades detentoras de direitos de propriedade intelectual são denominadas *autor* ou *proprietário*. A *marca* nada mais é que um padrão de bits especificado pelo *proprietário*. É exigido que a marca seja independente do sinal hospedeiro no sentido que qualquer marca pode ser inserida em qualquer sinal hospedeiro e posteriormente recuperada.

Na Fig. 1 vê-se um sistema de marca d'água digital onde Alice insere a marca e vende o sinal marcado (uma música, vídeo, software, etc) para Bob que por sua vez desfruta deste sinal e também detecta a presença da marca.

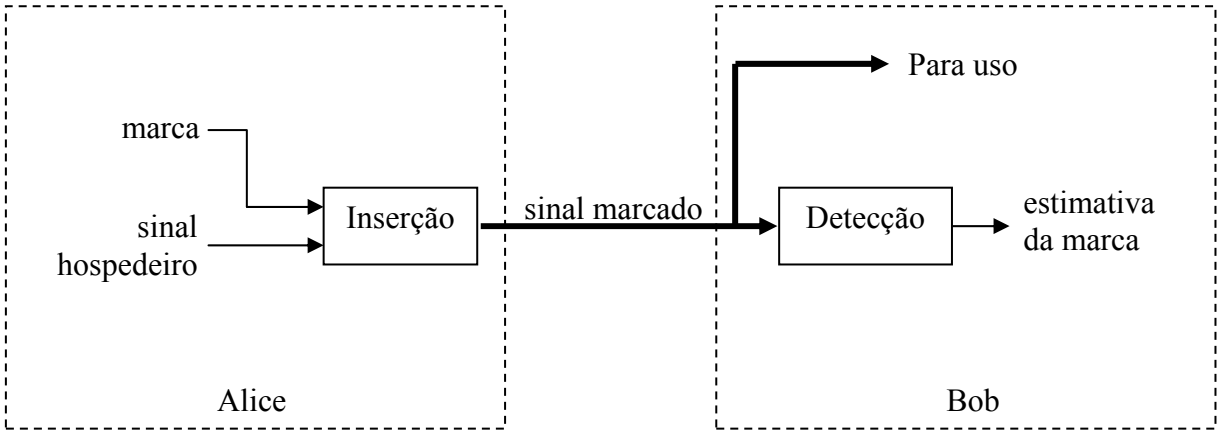


Figura 1 - Estrutura básica de um sistema de inserção de marca d'água digital.

Uma vez estabelecido o entendimento a respeito destes conceitos: a *marca d'água*, o *sinal hospedeiro* e o *sinal marcado* as demais necessidades impostas se devem ao modo como se deseja utilizar o sistema como um todo. Assim como na Esteganografia, uma marca d'água digital também insere uma mensagem, a *marca d'água*, em um mensagem inócua, o *sinal hospedeiro*[4]. Contudo, diferentemente da Esteganografia, não é sempre necessário que a existência da mensagem inserida (a marca d'água) seja secreta. O fundamental é o que ocorre com o sinal hospedeiro quando é feita uma tentativa indevida de retirar a marca, cujo conteúdo pode ou não ser secreto. No estudo de marcas d'água digitais o sinal hospedeiro tem valor próprio e isto deve ser considerado diretamente. Na Esteganografia o sinal hospedeiro somente deve ser preservado para que a mensagem secreta não seja detectada. Para enviar uma mensagem secreta, usando um sistema esteganográfico, o remetente da mensagem pode fazer com que um texto inócuo de Shakespeare seja substituído por um texto inócuo de Jorge Amado, contanto que tal substituição não cause suspeita. Tal substituição não faz sentido em um sistema de marca d'água digital pois viola um dos principais requisitos destes sistemas, a *transparência*. Como o sinal hospedeiro em um sistema de marca d'água tem valor próprio é importante que a inserção da marca não cause alterações que degradem significativamente a qualidade do sinal hospedeiro. Em outras palavras, a marca deve modificar o sinal hospedeiro o mínimo possível. Fazendo uma alusão ao caso de imagens, a marca d'água inserida tem que ser *transparente* para que não interfira na percepção propiciada pelo sinal hospedeiro. Por exemplo, ao se inserir uma marca d'água em uma música é importante que a qualidade da música continue adequada. Não há interesse em se escutar uma música distorcida e cheia de ruídos estranhos! Resumindo, em um sistema esteganográfico é necessário que seja impossível detectar uma mensagem inserida no sinal hospedeiro. E em um sistema de marca d'água digital é necessário que a informação inserida não comprometa a qualidade do sinal hospedeiro.

Por exemplo, as marcas d'água visíveis que atualmente aparecem em um dos quatro cantos de uma imagem de televisão, são um exemplo de marca d'água digital. Todos sabem que a marca está presente e portanto o sistema não é esteganográfico. Contudo a qualidade da imagem resultante não é comprometida. Embora semelhantes, os requisitos de sistemas

[4] Lembrando novamente, o bem digital é o sinal hospedeiro quando a marca d'água é utilizada para proteger a

esteganográficos e de marcas d'água não são os mesmos, se um requisito é satisfeito, isto *não* implica na satisfação do outro.

Outra técnica que é muito discutida na literatura e que pode ser facilmente confundida é a técnica de "impressão digital"[5]. Sistemas de "impressão digital" também têm o intuito de proteger a propriedade intelectual. Uma impressão digital pode ser vista como um número de série que é inserido em cada cópia legal de um bem digital. Se uma cópia ilegal do bem é achada é possível, através do "número de série", saber de qual cópia legal foi feita a cópia ilegal. A principal diferença entre um sistema de marca d'água digital e um sistema de "impressão digital" é que em um sistema de "impressão digital", a partir de um mesmo sinal hospedeiro, são conseguidos vários sinais marcados. Isto abre a possibilidade de que vários sinais marcados diferentes sejam usados em conjunto para retirar a marca. Este "conluio" não é possível no caso de marcas d'água, onde somente um sinal marcado está disponível.

2.1 Aplicações

Existe uma quantidade enorme de aplicações para sistemas de marca d'água digital. Cada aplicação requer uma marca com diferentes características. Seguem-se alguns exemplos ilustrativos.

Uma companhia que compra espaço publicitário pode usar uma marca d'água para monitorar os meios de comunicação e saber se suas propagandas estão sendo veiculadas como contratado, usando um sistema automático de detecção. Um marca d'água pode ser inserida na propaganda de televisão da companhia. Então toda vez que a propaganda for transmitida isto é registrado pelo detetor. Uma marca d'água digital desta forma deve ser detectada mesmo após pequenas modificações causadas pelos sistemas de transmissão e recepção, *i.e.* a marca deve ser *robusta* contra modificações típicas de sistemas de processamento de sinais. A marca também deve ser *difícil de falsificar*, para que a emissora de TV não possa enganar o cliente inserindo marcas falsas durante sua programação normal. Neste sistema o detetor pode se utilizar do sinal original para facilitar a detecção da marca d'água.

propriedade intelectual.

[5] Do inglês: *Digital fingerprinting*

Marcas d'água também foram sugeridas como uma forma segura de identificar radiografias para uso clínico. É importante que uma radiografia seja corretamente vinculada a um paciente devido aos sérios problemas que uma troca pode acarretar. Entretanto, radiografias são muito similares. Um sistema de marca d'água é capaz de formar um vínculo definitivo entre a imagem obtida da radiografia e o paciente. Para esta aplicação observa-se que a radiografia original não é guardada e portanto não está disponível para facilitar a detecção da marca.

O uso de fotografias como evidência legal está cada vez mais comprometido devido à facilidade com que tais fotografias podem ser adulteradas. Uma marca d'água pode servir para garantir a integridade de uma fotografia se alterações na fotografia fizerem com que haja uma modificação ou uma perda da marca. Marcas d'água digitais com esta característica são chamadas de *marcas frágeis* [6]. Sistemas que utilizam marcas frágeis normalmente são utilizados para detectar alterações e portanto é importante que uma marca frágil falsa não possa ser forjada. Em geral, se for fácil forjar uma marca frágil, a marca perde sua utilidade de detectar ou corrigir mudanças no sinal marcado. Basta modificar o sinal e posteriormente reinserir uma marca forjada para que a mudança não seja detectada.

Um exemplo final para a utilização de marcas d'água digitais é encontrado nos discos DVD. Um DVD que tem capacidade de armazenar 4.7Gbytes de informação, pode conter, assim como uma fita de videocassete, um filme gravado por um estúdio de cinema, um programa de televisão ou ainda um vídeo da última reunião de família. A perspectiva de que o DVD substitua o videocassete em um futuro próximo pressupõe que um gravador de DVDs esteja disponível para o consumidor a preço acessível. O consórcio que padroniza a utilização do DVD[7], preocupado com o potencial incentivo à pirataria causado pela disseminação de tais gravadores, incluiu um complexo sistema de controle de cópias nas especificações de funcionamento dos aparelhos DVD [3]. O sistema é projetado de forma a permitir a geração de um única cópia de programas de televisão e a cópia ilimitada de vídeos familiares. Além disto, o sistema não permite a cópia de filmes. Desta forma, não é possível gravar em um outro DVD o conteúdo de um DVD alugado na videolocadora. Contudo, é possível fazer uma única cópia do seu programa de televisão preferido para assisti-lo posteriormente em horário mais conveniente e também fazer cópias da filmagem

[6] Do inglês: *fragile watermarking*

de eventos familiares. Para que o sistema possa diferenciar entres estas situações é necessário inserir dois bits de informação no conteúdo de um DVD. Estes bits representam os seguintes estados:

Nunca copiar (*never_copy*) , para filmes cinematográficos.

Copiar uma vez (*copy_once*), para programas de televisão.

Não copiar mais (*copy_no_more*), para as cópias de programas de televisão.

Copiar livremente (*copy_freely*), para os vídeos familiares.

Através da leitura destes bits de informação um gravador de DVD é capaz de forçar o cumprimento das leis de propriedade intelectual. Todavia, todo aparelho DVD possui uma saída de vídeo analógica de excelente qualidade. Se ligado a esta saída, um gravador de DVDs pode inadvertidamente fazer inúmeras cópias ilegais de um disco marcado "nunca copiar**"**. Portanto, os bits de informação devem ser propagados através de conversões analógico-digitais e para isto um aparelho gravador deve ser capaz de identificar um sinal analógico como pertencente a um dos quatro estados acima. Para satisfazer a estas necessidades foi proposta a inserção de marcas d'água digitais. Estas marcas devem ser detectadas por qualquer aparelho gravador de DVDs e ao mesmo tempo devem ser muito difíceis de serem extraídas, sobrevivendo inclusive a um ataque que inclua conversões analógico-digitais. O sistema proposto neste trabalho atende a estes requisitos.

2.2 Classificação dos sistemas de marca d'água digital

Talvez o sistema de marca d'água digital mais simples consista em mudar os bits menos significativos de uma imagem de forma a inserir uma mensagem nos mesmos. Devido à grande quantidade de pixels, mensagens de tamanho razoável podem ser inseridas. Por exemplo, em uma imagem de 256×256 pixels o equivalente a 8 kbytes está disponível considerando-se apenas um bit em cada pixel. Infelizmente, esta forma de inserção é muito vulnerável a modificações para ser utilizada com marcas d'água robustas. É fácil remover a marca, basta alterar os bits menos significativos. De fato, qualquer processamento que modifique os bits menos significativos da imagem (tal como uma

[7] *The DVD Consortium*

filtragem) remove a marca. Por outro lado, um sistema que utilize os bits menos significativos de uma imagem pode ser utilizado para inserir uma marca d'água frágil. Para isto, é preciso fazer com que a marca d'água frágil inserida seja difícil de forjar. Marcas d'água frágeis tem o intuito de garantir a autenticidade. Elas podem ser removidas facilmente, mas dificilmente são forjadas.

Considerando marcas robustas, uma das principais falhas dos sistemas que inserem a marca nos bits menos significativos do sinal hospedeiro está justamente no fato dos bits menos significativos afetarem pouco a qualidade do sinal. Não há muita diferença entre uma amostra de áudio de CD de 16 bits com o valor de 10916 e outra com o valor de 10900. Isto faz com que marcas d'água inseridas desta forma sejam muito fáceis de serem removidas, já que é possível alterar à vontade os bits onde a marca encontra-se inserida sem causar grandes variações na qualidade do sinal hospedeiro. Um bom método para retirar uma marca robusta deste tipo é passar o sinal marcado por um sistema de compressão com perdas.

A observação de que a inserção de informação nos bits menos significativos de um sinal digital é uma forma muito frágil de marca d'água levou Cox *et al.* [8] a propor um sistema de marca d'água que insere a informação nas componentes perceptualmente significativas de um sinal através de técnicas de espalhamento espectral[8]. Apesar da utilização das componentes perceptualmente significativas dificultar a remoção de uma marca d'água, o que é desejável em sistemas de marca d'água robusta, também há uma conseqüência indesejável: A inserção da marca d'água causa necessariamente degradação do sinal hospedeiro. Segue que é importante atingir um equilíbrio entre a quantidade de degradação e a robustez da marca inserida.

Um outro método de inserção de marcas d'água foi proposto por Chen [4] e apresenta algumas características desejáveis. O método é denominado QIM (*Quantization Index Modulation*). Um dos principais atrativos deste método é que o mesmo atinge a máxima taxa de transferência de informação possível. Em outras palavras, uma marca inserida através deste método pode vir a transmitir tanta informação quanto possível[9]. Para inserir uma marca d'água utilizando QIM o sinal hospedeiro precisa ser quantizado gerando um sinal marcado. A escolha do quantizador a ser utilizado é que define a marca inserida.

[8] *Spread Spectrum*

Logo, no QIM, é necessário ter tantos quantizadores quantas diferentes marcas desejam-se inserir.

A marca é inserida em um sinal hospedeiro gerando um sinal marcado. Diz-se que uma marca d'água foi removida, se após alguma modificação no sinal marcado a mesma não é mais detectada. Podem-se classificar os diferentes sistemas de marca d'água digital de acordo com vários critérios. A seguir apresentam-se alguns deles:

- *Quanto à visibilidade*[10] *da marca:* Em imagens e vídeo, se uma marca d'água não pode ser percebida a olho nu, então é uma marca invisível. Caso contrário, é uma marca visível. Analogamente, em uma música pode-se ter uma marca audível ou não audível.
- *Quanto à facilidade de se remover a marca:* Se uma marca d'água digital pode ser removida com apenas pequenas alterações no sinal marcado a marca d'água é dita *frágil*. Caso contrário, se for necessário modificar muito o sinal marcado piorando a sua qualidade para retirar a marca, a mesma é dita *robusta*. Sistemas de marca d'água para proteção de bens digitais necessitam de marcas robustas.
- *Quanto ao uso do sinal hospedeiro para detectar a marca:* Se o sistema de detecção não se utiliza do sinal hospedeiro, isto é, o sinal original, para detectar a marca, então o sistema de marca d'água é um *sistema de detecção cega*[11]. Caso contrário o sistema é dito de *detecção não-cega*.
- *Quanto ao uso de senhas de segurança:* Se para detectar e inserir/remover uma marca d'água é utilizada uma mesma senha então o sistema é de *chave privada.* Neste caso, um pirata que possua a senha pode retirar a marca d'água. Se são utilizadas senhas diferentes, o sistema é de *chave pública*. Em sistemas com duas senhas um pirata pode possuir a chave de detecção mas não a chave de inserção/remoção de modo que é possível detectar a marca mas não extraí-la. Um sistema de duas senhas (de chave pública) onde uma senha é igual para todos os autores e todos os sinais hospedeiros é considerado por alguns autores um sistema *sem chave* [4]. Sistemas de chave pública são chamados sistemas *assimétricos*, enquanto sistemas de chave privada são chamados sistemas *simétricos*.

[9] Se o método for utilizado com compensação de distorção.
[10] É importante não confundir *visibilidade* com *transparência*. Neste texto todos os sistemas de marca d'água digital são transparentes (não comprometem a qualidade do sinal hospedeiro) mas em alguns casos a marca é *visível*, isto é, facilmente detectada por todos, e em outros não.

- *Quanto ao método de inserção da marca d'água:* Um sistema de marca d'água digital pode utilizar ou não o conhecimento a respeito do sinal hospedeiro na ocasião de inserção da marca para maximizar a quantidade de informação que o mesmo pode transmitir. Sob este aspecto o sinal hospedeiro é considerado uma interferência e a marca uma mensagem. Obviamente, isto é considerado somente quando o sinal hospedeiro não é conhecido no detetor (sistemas de detecção cega). Os sistemas de marca d'água de detecção cega podem ainda ser classificados como sistemas que rejeitam a interferência do sinal hospedeiro (e assim conseguem colocar mais informação em uma marca) e sistemas que não rejeitam a interferência do sinal hospedeiro[12] [4] .

2.3 Formulação do Problema

Neste texto, o estudo de marcas d'água digitais é considerado um problema de comunicação onde uma mensagem é adicionada ao sinal hospedeiro e transmitida juntamente com este através de um canal de comunicação hostil. O modelo considerado é mostrado na Fig. 2 . Tanto o receptor quanto o transmissor podem se utilizar de senhas criptográficas para detectar ou inserir a marca. Denota-se a chave de detecção por k_d e a chave de inserção por k_i . Para um sistema de chave privada $k_i = k_d$ e esta chave não é conhecida pelo atacante. O sinal hospedeiro s é de conhecimento do transmissor mas pode ou não (detecção cega) ser conhecido pelo receptor. O processo de inserção da marca também *pode* utilizar-se do conhecimento do sinal hospedeiro para gerar um sinal inserido[13] w que depende de s. O sinal marcado x é a soma do sinal hospedeiro s com o sinal inserido w . O canal de comunicação não é seguro e um atacante pode tentar remover a marca do sinal hospedeiro obtendo um sinal y de onde o detetor deve estimar a marca. Na Fig. 2 , o processo de inserção encontra-se cercado por um quadrado pontilhado e as setas tracejadas indicam informações que podem ou não estar disponíveis. As grandezas s, x, w e y também podem ser tratadas na forma vetorial, neste caso serão escritas em negrito $\mathbf{s}$, $\mathbf{x}$, $\mathbf{w}$ e $\mathbf{y}$.

[11] Do inglês: *blind watermarking scheme, non-blind watermarking scheme*.
[12] Do inglês: *host-interference rejecting/non-rejecting methods*.
[13] Observa-se que vários autores consideram w e não m a marca d'água.

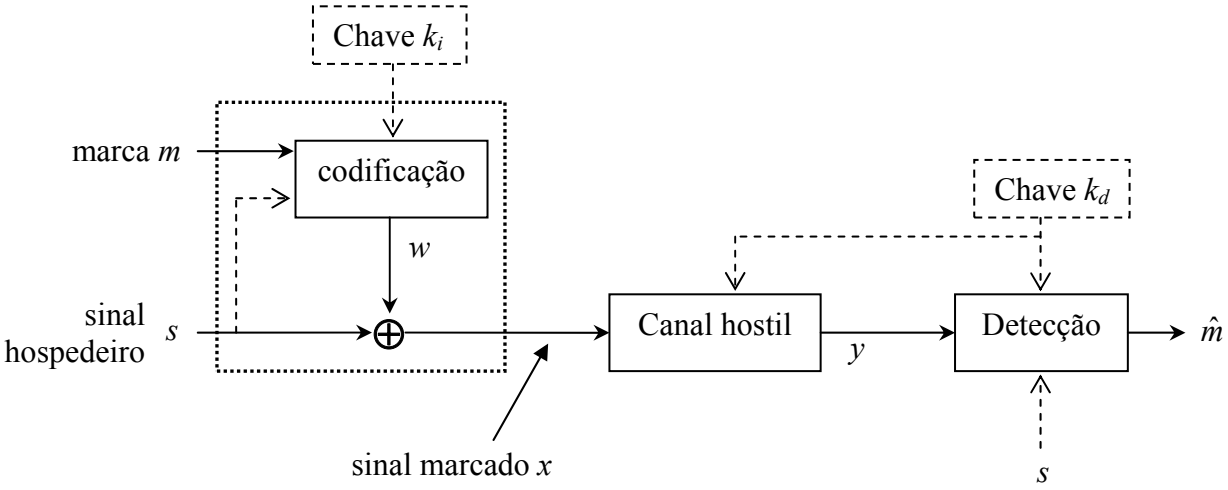


Figura 2 - Modelo de comunicação para sistema de marca d'água digital.

É importante relacionar este modelo de comunicação como o modelo apresentado na Fig. 1 onde Alice e Bob estavam presentes. Naquele modelo, Bob é o usuário do sinal vendido por Alice. Por exemplo, Alice pode produzir os CDs que Bob compra e ouve. Para adquirir o direito de ouvir os CDs de Alice basta Bob comprar os CDs. Naturalmente, no mundo real alguns compradores podem querer redistribuir o sinal. Por exemplo, Bob pode comprar um CD de Alice com o intuito de gravar várias cópias e vendê-las. *Portanto, não se pode supor que Bob seja confiável apesar de ser necessário que Bob possa utilizar o sinal marcado.* Esta situação leva à necessidade de prover Bob apenas com os meios necessários para que ele possa ouvir o CD comprado e nada mais. Bob não pode, por exemplo, obter a partir de algum processamento o sinal hospedeiro. Se isto ocorrer Bob pode retirar a marca d'água inserida. Logo, Alice não pode vender cópias do sinal hospedeiro, mas sim cópias do sinal marcado, as quais também devem ser suficientemente boas para que Bob possa usá-las. *No modelo de comunicação pode-se considerar que o usuário* (*Bob*) *é parte do canal hostil*. Nota-se que devido ao fato do usuário final não ser confiável *é necessário que a marca se propague*. Por exemplo, se um vídeo marcado, que está no formato MPEG, é gravado em uma fita VHS é interessante que a marca esteja presente na fita. Caso contrário, Bob pode gravar a fita e depois utilizá-la para fazer novamente uma codificação MPEG, retirando a marca.

Neste trabalho o conjunto de todos os sinais hospedeiros é designado por $\mathtt{S}$ e o conjunto de todos os sinais marcados por $\mathtt{X}$. Supõe-se que ambos $\mathtt{S}$ e $\mathtt{X}$ estão contidos em um espaço onde se pode definir uma medida de distância $d(\cdot,\cdot)$, que quantifique a distorção existente entre o sinal hospedeiro s e o sinal marcado x . Se para dois possíveis sinais marcados x_1 e x_2, $d(s,x_1) < d(s,x_2)$, então a qualidade de x_1 é melhor que a qualidade de x_2, *i.e.* x_1 pode ser usado para representar s de forma mais fidedigna que x_2 . Quanto mais fiel for a distância d em quantificar a distorção percebida pelo usuário, melhores e mais robustos serão os sistemas desenvolvidos a partir da mesma.

2.4 Métodos de Detecção Cega

Nos últimos anos foi proposta um grande número de métodos para a inserção de marcas d'água robustas. Contudo, a maior parte dos métodos propostos são simétricos, de detecção não-cega e obtém a sua robustez através do sigilo. Os métodos tornam-se seguros ou por que o atacante não sabe onde está a marca d'água ou por que ele não sabe exatamente qual é a marca. Em outras palavras, o atacante não é capaz de detectar a presença da marca e assim não pode removê-la. Felizmente, nem todos os métodos propostos são de detecção não-cega. Dentre os métodos com detecção cega propostos alguns merecem atenção especial e serão abordados nesta seção.

Um resultado surpreendente obtido por Costa [7] para o caso de sinais Gaussianos, que tem uma forma análoga para o caso discreto [18], mostra que a quantidade de informação que a marca d'água pode transmitir é independente do detetor conhecer ou não o sinal hospedeiro. Como o sinal hospedeiro é conhecido durante a inserção da marca d'água, a interferência causada por este pode ser completamente eliminada pelo codificador. Portanto, nestes casos, para um mesmo nível de distorção, um esquema de detecção cega consegue transmitir tanta informação quanto um esquema com detecção não-cega. Baseando-se nestas idéias, alguns métodos de inserção de marca d'água com detecção cega foram propostos. Estes métodos visam obter com detecção cega um desempenho tão bom quanto o obtido com detecção não-cega. De uma forma geral, os métodos atingem a capacidade de transmissão assintoticamente.

Segue-se uma breve descrição de alguns métodos de marca d'água. Os métodos aqui apresentados não são específicos para um determinado tipo de aplicação, podendo ser aplicados em uma variedade de modalidades de sinais hospedeiros. A escolha dos métodos aqui apresentados foi feita com o intuito de mostrar o funcionamento dos métodos mais freqüentemente encontrados na literatura, assim como métodos com características interessantes.

2.4.1 Espalhamento Espectral

O método de inserção de marcas d'água com detecção cega por espalhamento espectral proposto por Cox *et al.* [8] consiste em adicionar uma seqüência pseudo-aleatória $\alpha\mathbf{z}$, de pequena amplitude, às componentes do sinal hospedeiro **s**, gerando o sinal marcado $\mathbf{x} = \mathbf{s} + \alpha\mathbf{z}$.[14] O método supõe que o sinal hospedeiro **s** é ortogonal à seqüência pseudo-aleatória **z** utilizada. Após a inserção da marca o sinal marcado pode ser modificado gerando o sinal atacado $\mathbf{y} = \mathbf{x} + \mathbf{e}$. Para detectar a marca d'água é feita a correlação do sinal atacado **y** com a seqüência pseudo-aleatória **z** . Supondo que o sinal atacante **e** também é ortogonal a **z**, segue-se que

$$E[\mathbf{z}\,\mathbf{y}] = E[\mathbf{z}\,(\mathbf{s} + \alpha\mathbf{z} + \mathbf{e})] = E[\mathbf{z}\,\mathbf{s}] + \alpha E[\mathbf{z}^2] + \mathrm{E}[\mathbf{z}\,\mathbf{e}] = \alpha E[z^2] \qquad (2.1)$$

Para seqüências finitas o resultado é aproximado e é feito um teste de hipóteses na medida da correlação obtida. Podem-se observar dois pontos importantes. Em primeiro lugar, se a seqüência **z** é conhecida pelo atacante, a marca pode ser removida adicionando-se um sinal **e** apropriado (*e.g.* $\mathbf{e} = -\alpha\mathbf{z}$). Além disto, o método não utiliza qualquer conhecimento prévio que o codificador possa ter sobre o sinal hospedeiro **s** para ajudar na detecção. O sinal hospedeiro é tratado pelo detetor como uma interferência desconhecida. Este é um método de detecção cega que não rejeita a interferência do sinal hospedeiro.

2.4.2 QIM (*Quantization Index Modulation*)

Como já mencionado, o método QIM requer que o sinal hospedeiro *s*, a ser marcado, seja quantizado. O sinal quantizado constitui o sinal marcado *x* e a escolha do quantizador utilizado é que define a marca inserida. Seja $q_i(a, \Delta)$ a quantização do sinal *a*

por um quantizador de índice *i* que tem distância Δ entre os níveis de quantização. Para o QIM, pode-se escrever $x = q_m(s, \Delta)$. Neste método, o detetor da marca d'água quantiza novamente o sinal marcado com um quantizador de detecção q_d, que é a união de todos os possíveis quantizadores usados no processo de inserção:

$$q_d(\cdot, \Delta) = \bigcup_{\forall i} q_i(\cdot, \Delta) \tag{2.2}$$

Na união de quantizadores aqui representada pelo símbolo ∪, supõe-se que os níveis de quantização são únicos dentre todos os quantizadores e também que os intervalos de quantização do quantizador resultante são repartidos de modo a quantizar cada ponto com o nível de quantização mais próximo. Determinando-se o nível de quantização em que o sinal hospedeiro é quantizado se pode determinar qual quantizador foi utilizado e portanto qual a marca d'água inserida. Quanto maior a distância entre os níveis de diferentes quantizadores maior a robustez da marca inserida, já que, é necessário que haja mais distorção para confundir o detetor e retirar a marca. Isto pode ser conseguido aumentando-se a distância Δ entre os níveis de um mesmo quantizador. Todavia, aumentar Δ também aumenta a distorção gerada no processo de inserção da marca. Chen [4] propôs uma modificação do QIM que adapta a distância entre os níveis de reprodução, ao sinal hospedeiro. A idéia é retirar uma fração do erro gerado na quantização do sinal marcado. O processo de inserção torna-se:

$$x = q_m(s, \Delta/\alpha) + (1-\alpha)[s - q_m(s, \Delta/\alpha)] \tag{2.3}$$

O primeiro termo do lado direito da equação é o QIM padrão e o segundo termo cancela parte do erro introduzido pelo processo de quantização. O parâmetro α pode ser calculado de forma a maximizar a relação sinal ruído no detetor de marcas d'água[15]. Chen [4] batizou este método de QIM com compensação de distorção. O método atinge assintoticamente a capacidade de transmissão de informação para o caso Gaussiano considerado por Costa [7]. O método não é assimétrico já que ambos codificador e decodificador devem conhecer os quantizadores e também o parâmetro α.

[14] Para tornar o sistema mais robusto contra variações de escala nos sinais, Cox *et al.* [8] também consideram outras funções para inserir a marca, por exemplo, $\mathbf{x} = \mathbf{s}(1 + \alpha\mathbf{z})$. Mas a idéia principal é a mesma.
[15] Onde a relação é definida pela razão entre a distância mínima entre dois quantizadores ao quadrado e a interferência causada tanto pelo canal quanto pelo processo de inserção.

2.4.3 SCS (*Scalar Costa Scheme*)

Eggers *et al.* [10] propuseram um esquema de codificação derivado diretamente do método de codificação aleatória apresentado por Costa para atingir a capacidade do canal no caso Gaussiano. O método, denominado SCS, substitui o dicionário não estruturado do trabalho de Costa [7], por um dicionário estruturado como um reticulado, reduzindo a complexidade. O desempenho é muito bom e o método consegue obter altas taxas de transferências de informação na marca d'água[11] [12]. Novamente, o método proposto é de detecção cega mas não é um método assimétrico. Tanto o codificador quanto o decodificador (detetor de marcas) devem conhecer o dicionário utilizado. Para todos os métodos cegos de aplicação geral, este método apresenta o melhor desempenho.

2.4.4 Chou *et al.*

Uma abordagem de detecção cega apresentada por Chou *et al.*[5] baseia-se na dualidade existente entre dois problemas. O problema de inserção de marcas d'água e o problema de codificação distribuída de fonte. O problema de inserção de marcas d'água digitais pode ser considerado um problema de codificação de canal onde informação auxiliar está disponível no codificador. Neste enfoque a marca d'água é a mensagem a ser transmitida e o sinal hospedeiro é parte da interferência do canal de comunicação. Especificamente, a informação auxiliar disponível pode ser vista como o conhecimento que o codificador tem do sinal hospedeiro. Para um sistema de detecção cega, esta informação está disponível somente no codificador. Com o conhecimento da dualidade existente entre codificação de fonte com informação auxiliar (sobre a fonte) disponível no decodificador e codificação de canal com informação auxiliar (sobre o canal) disponível no codificador, Chou *et al.* [5] utilizaram como base um método de codificação distribuída de fonte [26][16] para inserir marcas d'água. Devido à estrutura genérica da analogia feita e ao foco no caso contínuo, Chou *et al.* não consideraram o caso discreto a não ser como motivação. Seguindo esta linha de raciocínio não fica claro qual método específico de codificação se deve utilizar. Isto levou a outro trabalho dos mesmos autores [6] onde uma solução

[16] O método proposto é discreto e é essencialmente o mesmo método proposto por Heegard [17] para a utilização de memórias com defeitos. Todavia, o trabalho [26] não faz referência ao trabalho de Heegard [17].

específica é proposta. Por outro lado, colocando o problema na forma contínua facilita-se muito a sua aplicação prática.

Como mencionado, o enfoque dos métodos de detecção cega citados é tentar tornar irrelevante o fato do detetor de marcas d'água não conhecer o sinal hospedeiro. O objetivo principal é, em geral, maximizar a capacidade de transmissão de informação através da marca. Normalmente, isto leva a métodos que são robustos contra tipos específicos de ataque. Por exemplo, Chou *et al.* [17] [6] supõe que tentativas de remoção da marca limitam-se à adição de ruído branco Gaussiano (AWGN) ou compressão de dados com padrão JPEG [17] [6]. Neste trabalho o objetivo principal é a robustez da marca inserida, contra ataques intencionais (arbitrários e irrestritos). A capacidade de transmissão de informação através da marca também é considerada mas não se constitui no objetivo principal.

2.5 Métodos Assimétricos

Os métodos de detecção cega requerem que o detetor não faça uso do sinal hospedeiro para a detecção da marca. Os métodos de chave pública ou assimétricos requerem ainda que a detecção da marca não torne o detetor capaz de removê-la. Logo, os métodos assimétricos são todos de detecção cega. Caso contrário, o detetor poderia passar adiante o sinal hospedeiro, que está disponível, sem marca, para a detecção. Em oposição à grande quantidade de métodos simétricos propostos, há poucos métodos de chave pública na literatura.

2.5.1 Espalhamento espectral com chave parcialmente conhecida

Em um método de chave pública vários detetores diferentes devem ser capazes de detectar a marca, mas nenhum deve ser capaz de removê-la. No método de espalhamento espectral proposto por Cox *et al.* [8] o detetor precisa conhecer a seqüência pseudo-aleatória **z** para detectar a marca. Contudo, conhecendo **z** o detetor pode também remover a marca d'água. Hartung e Girod [16] propuseram uma modificação do método de espalhamento espectral que pode ser vista como um método de chave pública. Ao invés de

prover cada detetor com a seqüência pseudo-aleatória **z** usada para inserir a marca, cada detetor recebe uma seqüência diferente $\mathbf{z}_i$ que é igual à seqüência **z** em apenas algumas posições. Às posições não coincidentes entre os dois sinais são atribuídos valores aleatórios. Escolhendo-se adequadamente as seqüências utilizadas e também aumentando-se o nível de distorção α (*cf.* eq. 2.1), ainda é possível detectar a marca d'água através da correlação descrita anteriormente. Embora um atacante possa fazer com que seja impossível a detecção da marca utilizando-se uma seqüência $\mathbf{z}_i$, um outro detetor ainda pode detectar a marca utilizando uma seqüência $\mathbf{z}_j$ diferente. Apesar de ser simples, o sistema tem sérias limitações com relação ao número de seqüências $\mathbf{z}_i$ disponíveis. Obviamente, há também a possibilidade de um "conluio" entre os detetores para retirar a marca.

2.5.2 Seqüências de Legendre e Autovetores

Outro método que tenta detectar a marca através do coeficiente de correlação é o método proposto por van Schyndel *et al.* [28], de seqüências de Legendre. Uma seqüência de Legendre **r** tem a interessante propriedade de que sua Transformada de Fourier é igual à própria seqüência, conjugada e multiplicada por uma ganho *h*. Representando a Transformada Discreta de Fourier através da multiplicação pela matriz $\boldsymbol{F}$ segue-se que $\boldsymbol{F}\,\mathbf{r} = h\,\mathbf{r}^*$, onde $\mathbf{r}^*$ denota a seqüência **r** conjugada e o ganho *h* pode ser complexo. Na detecção supõe-se que o sinal hospedeiro não é auto-correlacionado, desta forma a existência de um valor de autocorrelação maior que um certo limiar indica a presença da marca.

O método de seqüências de Legendre foi generalizado por Eggers *et al.* [13]. Ao invés de utilizar a Transformada de Fourier, uma transformação linear geral representada pela matriz **G** é utilizada. As diferentes marcas d'água que podem ser inseridas são os diferentes autovetores da matriz **G**. Para inserir uma marca d'água em uma seqüência de comprimento N a matriz **G** deve ser $N \times N$. Durante a inserção, um autovetor **v** é adicionado ao sinal hospedeiro $\mathbf{x} = \mathbf{s} + \mathbf{v}$. O sinal pode ser modificado por um atacante gerando o sinal atacado $\mathbf{y} = \mathbf{s} + \mathbf{v} + \mathbf{z}$. Para detectar a marca, a correlação entre o sinal recebido e sua transformação é calculada $c = \mathbf{y}^t\,\mathbf{G}\,\mathbf{y}\,/N$. Se o valor *c* for alto, então **y** deve conter um autovetor de **G**. A detecção funciona adequadamente se $(\mathbf{s} + \mathbf{z})^t\,\mathbf{G}\,(\mathbf{s} + \mathbf{z})$ for pequeno, *i.e.* se o sinal hospedeiro somado com as modificações introduzidas no ataque não for

significativamente correlacionado com a soma de suas transformações. No método de autovetores a matriz **G** serve de chave pública.

Embora o problema não seja tão grave no método de autovetores, podem ser retiradas as marcas de ambos os métodos através de uma busca exaustiva de todas as possibilidades para o sinal **v**. Outro problema encontrado em ambos sistemas é a interferência do sinal hospedeiro, que não é rejeitada. Além disto, para alguns casos, projetando o sinal marcado no espaço gerado pelos possíveis autovetores e subtraindo a projeção é possível retirar a marca e simultaneamente melhorar a qualidade do sinal. Este ataque foi descrito por Furon [9].

2.5.3 QIM, versão assimétrica

O método QIM sem compensação de distorção pode ser utilizado como um método assimétrico. Para remover uma marca inserida com o QIM é necessário quantizar novamente um sinal já quantizado e portanto aumentar a distorção. Isto foi proposto por Chen [4]. No entanto, a quantidade de distorção extra introduzida para um ataque bem sucedido é muito pequena, o que torna o método pouco robusto.

2.5.4 Função "one-way" de processamento de sinais

Outro método assimétrico para a inserção de marcas d'água foi proposto por Furon e Duhamel [14]. O método consiste em permutar o sinal hospederio de forma a obter uma função densidade espectral de potência (do sinal permutado) plana, $\widetilde{\Phi}_{ss}(\Omega) = P_s$. Em seguida, é adicionado ao sinal permutado $\widetilde{\mathbf{s}}$ um ruído colorido **z** independente do sinal hospedeiro **s**. Este ruído é obtido passando-se ruído branco de potência P_z por um filtro linear para obter $\Phi_{zz}(\Omega) = P_z|H(\Omega)|^2$. O sinal marcado permutado é dado por: $\widetilde{\mathbf{x}} = \widetilde{\mathbf{s}} + \mathbf{z}$. Como **z** é independente do sinal hospedeiro, a função densidade espectral de potência do sinal marcado permutado é dada por $\widetilde{\Phi}_{xx}(\Omega) = P_s + P_z|H(\Omega)|^2$. Esta função tem o mesmo formato que a função densidade espectral de potência da marca inserida. Através do teste de hipóteses sobre o formato da função densidade espectral de potência do sinal marcado, descrito em [14], é possível decidir se uma marca está presente ou não. Não é necessário

conhecer **z** para detectar a marca. Basta conhecer o formato da função densidade espectral de potência de **z**. Como, em geral, é impossível determinar **z** a partir de $\Phi_{zz}(\Omega)$ não é tão simples remover a marca quanto detectá-la. Contudo, uma filtragem que achate a função densidade espectral de potência do sinal marcado permutado consegue remover a marca. Além disto, escolhendo-se cuidadosamente o filtro, pode-se fazer com que as componentes mais afetadas sejam justamente as que foram mais distorcidas durante o processo de inserção. Através da filtragem é possível remover a marca, e até melhorar a qualidade do sinal obtido [9].

Capítulo 3

Sistema Proposto

Neste capítulo propõe-se um método para a inserção de marcas d'água digitais em sinais hospedeiros constituídos de uma seqüência de valores discretos. O método proposto não requer a utilização de chaves criptográficas e a detecção da marca é feita sem a utilização do sinal hospedeiro (detecção cega). A marca inserida é robusta e sua detecção pode ser feita por agentes não confiáveis.

Um sistema de marca d'água digital com detecção não-cega necessita que o detetor tenha disponível uma cópia do sinal hospedeiro. Isto limita o número de aplicações em que estes sistemas podem ser empregados. Do mesmo modo, em sistemas de chave privada, a utilização da mesma chave criptográfica tanto para a detecção quanto para a remoção da marca reduz a flexibilidade do sistema. Portanto, é desejável que se encontre um sistema que opere com detecção cega e chave pública. Deseja-se um sistema capaz de resistir a ataques intencionais. Seguindo o princípio de Kerckhoffs [1], supõe-se que o atacante tenha total conhecimento dos métodos de inserção/detecção e da chave pública. Contudo o atacante desconhece a chave privada e isto dificulta a remoção da marca robusta que é inserida. Além disso, supõe-se que o atacante também tenha capacidade computacional ilimitada. Portanto, procura-se um sistema incondicionalmente seguro. Embora deseje-se obter um sistema de marca d'água incondicionalmente seguro é necessário que seja computacionalmente fácil detectar a marca d'água. Em um sistema incondicionalmente seguro de chave pública, a chave pública pode servir para facilitar o esforço computacional necessário para calcular a marca d'água a partir do sinal marcado. Neste trabalho são considerados somente sistemas de chave pública que operem desta forma[17].

Supõe-se que tanto o sinal hospedeiro quanto o sinal marcado possam ser representados como vetores N dimensionais. Considera-se aqui o caso onde as componentes

[17] Em Criptologia, o sistema incondicionalmente seguro "*one-time-pad*" é simétrico. Como os requisitos são diferentes, não está claro se a mesma situação se aplica ao estudo de marcas d'água, ou seja, se podem haver sistemas de marca d'água incondicionalmente seguros e de chave pública.

do vetor sinal hospedeiro podem ser vistas como uma seqüência de variáveis aleatórias i.i.d. (independentes e identicamente distribuídas) com distribuição uniforme no alfabeto Z_q, onde Z_q é o anel de inteiros módulo q [18]. Logo, no caso binário, o sinal hospedeiro é modelado como uma seqüência i.i.d. Bernoulli(1/2). Vale lembrar que apesar da hipótese ser bastante restritiva, toda informação quando compactada tende a se aproximar desta estatística.

Como mencionado no capítulo anterior, é necessário considerar uma medida que quantifique o grau de distorção produzido pelo processo de inserção da marca d'água. Apesar de ser apenas uma aproximação, em geral é feita a hipótese que esta medida de distorção é uma métrica. Neste texto é utilizada a *distância de Hamming* para medir a distorção introduzida pelo sistema de marca d'água digital ou por possíveis atacantes. Neste trabalho considera-se a distância de Hamming entre dois vetores dada pela soma das distâncias de Hamming componente a componente. Portanto, se $\mathbf{x} = (x_1,x_2,...,x_n)$, $\mathbf{y} = (y_1,y_2,...,y_n)$ a distância de Hamming é:

$$\mathbf{d}(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} d(x_i, y_i) \tag{3.1a}$$

onde,

$$d(x,y) = \begin{cases} 1, & \text{se } x = y \\ 0, & \text{caso contrário} \end{cases} \tag{3.1b}$$

A distância de Hamming é muito usada para o caso binário, onde é equivalente à distância de Lee que em $Z_q = \{0,1,2,..., q\text{-}1\}$ é dada por:

$$d_{Lee}(x,y) = \begin{cases} |x-y| & , \text{se } |x - y| \le q/2 \\ q - |x-y| & , \text{caso contrário} \end{cases} \tag{3.2}$$

Em outros alfabetos finitos, quando mais símbolos estão disponíveis, talvez a distância de Lee seja mais usual. Nestes casos a distância de Lee se aproxima mais da idéia de distância usual de números reais, *i.e.* $d(x,y)=| x - y |$. Por exemplo, no anel de inteiros com 256 elementos, Z_{256}, os vetores $\mathbf{v}_1 = (0,2,6,1,3)$ e $\mathbf{v}_2 = (0,2,1,1,3)$ têm entre si uma distância de Hamming de 1 e uma distância de Lee de 5, contudo os vetores $\mathbf{v}_1 = (0,2,6,1,3)$ e $\mathbf{v}_3 = (0,2,245,1,3)$ que também têm uma distância de Hamming de 1, têm uma distancia de

[18] Neste trabalho q é considerado uma potência de um número primo, embora está condição não seja necessária para que Z_q seja um anel.

Lee de 17 e não de 245 – 6 = 239. Logo, a noção de distância de números reais também é perdida neste caso. Tomando a distância de Hamming como medida de distorção, efetivamente equivale a desprezar a influência da amplitude da variação de uma componente vetorial (como no caso acima, uma variação de 1 causa tanta distorção quanto uma variação de 239). À distância de Hamming interessa somente o número de componentes de valores diferentes.

Retornando ao modelo do sistema de marca d'água digital mostrado na Fig. 1, o usuário de um sinal digital quer que a qualidade do sinal usado seja tão boa quanto possível. Por exemplo, ao ouvir uma gravação de uma orquestra sinfônica, deseja-se a melhor qualidade da gravação possível. O sinal hospedeiro, a princípio, tem uma qualidade muito boa. Infelizmente, a inserção de uma marca d'água modifica o sinal hospedeiro e obtém-se, um sinal de qualidade inferior, o sinal marcado $\mathbf{x}$. Um possível atacante, por sua vez, modifica o sinal marcado. Espera-se que em um bom sistema de marca d'água esta modificação piore ainda mais a qualidade do sinal obtido. No entanto, degradações sucessivas não podem ser acumuladas indefinidamente. Existe um limite para o grau de distorção que um usuário pode tolerar. Neste trabalho o nível máximo tolerável de distorção $d_{máx}$ é medido a partir do sinal hospedeiro $\mathbf{s}$. Em um sistema em que não houve ataque, a distorção é dada pela distância entre o sinal hospedeiro $\mathbf{s}$ e o sinal marcado $\mathbf{x}$. Se, por outro lado, um atacante produzir um outro sinal $\mathbf{y}$ a partir do sinal marcado $\mathbf{x}$, a qualidade do sinal obtido, é naturalmente quantificada pela distância entre $\mathbf{y}$ e $\mathbf{s}$ e não entre $\mathbf{y}$ e $\mathbf{x}$.

Uma das estratégias de ataque que pode ser utilizada é inserir modificações tão pequenas quanto possível para que o ataque não deteriore muito mais a qualidade do sinal. Como a distância é uma métrica, vale a desigualdade triangular: $d(\mathbf{s},\mathbf{y}) \leq d(\mathbf{s},\mathbf{x}) + d(\mathbf{x},\mathbf{y})$. Por hipótese, a distância $d(\mathbf{s},\mathbf{x})$ esta dentro do limite aceitável, *i.e.* $d(\mathbf{s},\mathbf{x}) \leq d_{máx}$, já que o usuário final compra o sinal marcado $\mathbf{x}$. Também se pode definir ao redor de cada ponto $\mathbf{s}_i$ do espaço de sinais hospedeiros, uma esfera no interior da qual todos os pontos tem distorção aceitável. Diminuindo $d(\mathbf{x},\mathbf{y})$ espera-se que $d(\mathbf{s},\mathbf{y})$, que é limitada pela desigualdade acima, também seja pequena o suficiente para ser aceita pelo usuário. Em outras palavras, tenta-se garantir que $d(\mathbf{s},\mathbf{y}) < d_{máx}$ fazendo $d(\mathbf{x},\mathbf{y})$ tender a zero. É mais provável que este tipo de estratégia obtenha sucesso se $d(\mathbf{x},\mathbf{y}) = \Delta d$ puder ser pequeno e $\mathbf{x}$ encontrar-se no interior e não na casca da esfera. A Fig. 3 ilustra estas duas situações.

Como x_1 encontra-se no interior da região de distorção aceitável, *i.e.* dentro da esfera de raio $d_{máx}$, uma variação suficientemente pequena ainda resulta em um sinal de distorção aceitável. Por outro lado, no caso de x_0 situado na fronteira da região, mesmo fazendo distorções muito pequenas há a possibilidade do ataque produzir um sinal com distorção inaceitável.

Observa-se então que para tentar obter maior segurança contra ataques que fazem modificações pequenas, deve-se procurar levar o sinal marcado até o limiar de distorção aceitável. *Portanto, é interessante degradar o sinal marcado tanto quanto for tolerável*. Deste modo o atacante fica sem margem para modificar ainda mais o sinal marcado. Esta observação está de acordo com a prática de inserir marcas robustas nos componentes *mais* significativos do sinal.

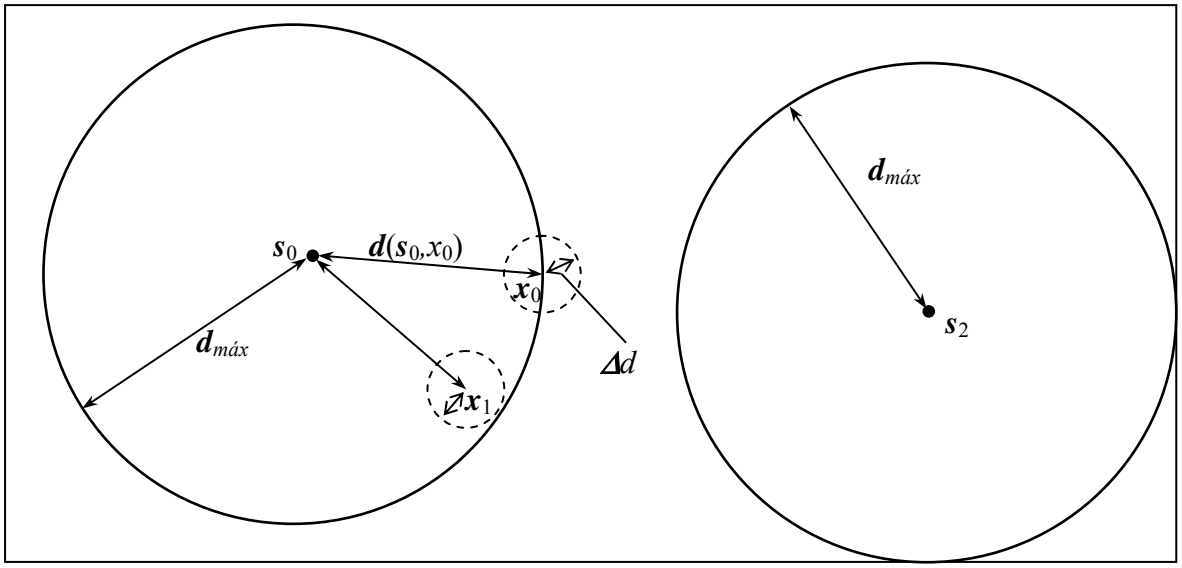


Figura 3 - Ataques com pequenas variações.

Uma consideração importante em sistemas de marca d'água digital de chave pública é que o usuário não é confiável. Todo usuário do sinal marcado é também um pirata em potencial. Isto implica que, uma marca d'água robusta deve se propagar junto com o sinal marcado. Caso contrário, a marca pode ser removida. Por exemplo: Bob, o pirata, comprou um CD de música com o intuito de regravá-lo. O CD foi feito por Alice que não sabe destas intenções escusas. Ao gravar uma música em um CD, são sempre colocados vários bits que especificam informações outras que não a música em si e Alice inadvertidamente colocou uma marca d'água nestes bits. Como os bits não fazem parte da música que é ouvida, a

marca é perdida assim que a música é enviada à saída analógica do tocador de CD. No entanto, a qualidade da música obtida na saída analógica é excelente. Bob pode obter um sinal não marcado a partir desta saída. Não adianta criptografar a saída analógica. Embora isto impossibilite a retirada da marca, também impossibilita a utilização do sinal marcado.

A necessidade de propagação da marca d'água tem conseqüências interessantes. Considera-se o caso de um sistema para marcar imagens, em dois estágios, funcionando da seguinte forma. Em um primeiro passo a marca d'água é inserida no sinal hospedeiro s que está no formato JPEG (".jpg") produzindo o sinal intermediário marcado x. Contudo x está em um novo formato que não pode ser usado diretamente. Em seguida, o sinal x é distribuído e vendido. Para usar x é necessário aplicar uma transformação $t(\cdot)$ que devolva o sinal ao formato JPEG e permita uma boa reprodução da imagem (muito próxima de s). Neste caso $v = t(x)$ deve conter toda a informação sobre a marca d'água que está contida em x, caso contrário, um pirata pode distribuir v como uma imagem própria. *Um sistema que marca um estágio intermediário não atende aos requisitos de um sistema robusto se a marca não se propagar.* Nos sistemas sem chave[19] há uma restrição ainda maior, já que o sistema é equivalente a um sistema com chaves públicas todas iguais. Logo, v deve conter não somente toda a informação da marca *que está em x* mas também toda a informação da marca de modo geral. Permitindo-se que informação sobre a marca d'água esteja presente também na chave pública e não somente no sinal marcado, pode haver o caso em que obtém-se do mesmo sinal marcado x, duas marcas d'água diferentes, utilizando-se duas chaves públicas diferentes. Como a chave pública não é usada pelo usuário a não ser para gerar a marca, a mesma pode ser modificada arbitrariamente. Isto não ocorre com o sinal marcado. Neste caso, o sistema fica vulnerável a um ataque que faça uma busca exaustiva pelo espaço de chaves públicas até encontrar uma chave que gere uma marca adequada aos propósitos do atacante (a marca do próprio atacante ou uma marca inválida). Em outras palavras, um atacante pode encontrar uma outra chave pública, ainda "sem dono", que quando utilizada para calcular a marca d'água gera a marca do atacante. Em seguida, o atacante torna esta chave, uma de suas chaves públicas (o atacante passaria a ter mais de uma chave pública). Observa-se que *qualquer* chave pública ainda não utilizada pode servir. No modelo utilizado neste trabalho, supõe-se que a capacidade computacional do

[19] Sistemas em que a chave pública é sempre a mesma para todos os autores e todos os sinais hospedeiros.

atacante é ilimitada, e portanto, é mais interessante considerar o caso alternativo em que *o sinal marcado especifica unicamente a marca*. Isto é, embora possa ser necessário um grande esforço computacional para calcular a marca a partir do sinal marcado x, a marca é inequivocamente determinada a partir do sinal marcado.

Do exemplo acima, pode-se notar outra conseqüência do fato do usuário não ser confiável. Para o sistema ser robusto contra ataques intencionais, o usuário jamais poderá ter uma cópia do sinal hospedeiro. De fato, é interesse do proprietário que o usuário, um pirata em potencial, compre um sinal tão ruim quanto ele possa tolerar. Este é o preço que usuários honestos pagam pela existência de usuários desonestos.

Como a marca d'água deve se propagar junto com o sinal marcado, *o sinal marcado deve ser a melhor estimativa do sinal hospedeiro que o usuário pode obter*. Um bom exemplo desta necessidade é encontrado se o sinal hospedeiro for codificado com um código corretor de erros antes da inserção da marca d'água. Se os erros causados pelo processo de inserção da marca puderem ser corrigidos através do código utilizado, então o usuário pode obter o sinal hospedeiro e o sistema não é robusto.

3.1 Motivação Intuitiva

Nesta seção são feitas algumas observações sobre o comportamento geral de sistemas de marca d'água digital com detecção cega e chave pública[20]. Estas observações fornecem a motivação para a forma do sistema que é proposto em seguida.

Deseja-se um sistema onde a segurança seja independente de qualquer esforço computacional. Ou seja, não importando a capacidade computacional de Bob, o sistema deve ser robusto. Primeiramente lembra-se que se Bob conseguir descobrir o sinal hospedeiro s, então o sistema não é seguro; por que Bob pode passar s adiante e s não é marcado. Considera-se então o caso hipotético de um sistema de marca d'água digital onde o sinal marcado x é uma função de duas variáveis s e m. Isto é, para inserir a marca d'água o codificador faz: $x = f(s,m)$. Deve também existir uma função h de forma que o decodificador calcule h para detectar a marca, $\hat{m} = h(y)$. Isto implica que a determinação

[20] Tratam-se de chaves públicas, como descrito anteriormente, que auxiliam no esforço computacional necessário para detectar a marca d'água.

de x também determina unicamente a marca inserida pois $m = h(x)$, conforme observado anteriormente esta condição é desejável também por outros motivos.

Para se ter um sistema em que a marca inserida m é pública e de conhecimento de todos é necessário que a função f não seja injetora. Caso contrário, conhecendo-se x e m é possível calcular a função inversa[21] de f e achar s. Portanto, para sistemas de chave pública a função f deve ser "muitos-para-um". Para cada sinal marcado, deve haver sinais hospedeiros suficientes no domínio de f para que a probabilidade de um atacante gerar um sinal s com distorção aceitável seja pequena. Da necessidade de x especificar unicamente m segue que, para qualquer x_i na imagem de f tal que $x_i = f(s_i, m_i)$, não deve existir nenhuma outra marca diferente m_j que também leve a este ponto. Isto pode ser observado no diagrama da Fig. 4, onde, para efeito de melhor visualização, S e X são mostrados como conjuntos diferentes.

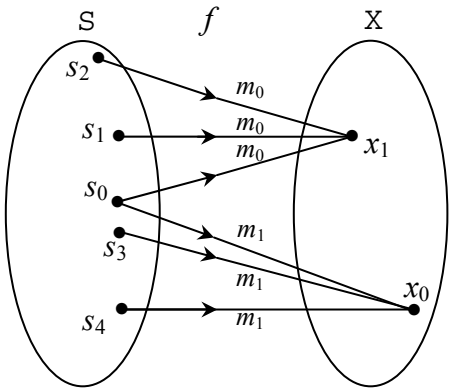


Figura 4 - Mapeamento desejável para a inserção de marca pública.

Como f é uma função, não pode ocorrer o que é mostrado na Fig. 5 onde o ponto s_0 é mapeado em dois pontos diferentes x_0 e x_1 para a mesma marca m_0 . Além disto a Fig. 5 também mostra outras situações indesejáveis. Uma delas é que caso o sinal marcado seja x_1 não há como saber se este foi obtido através de $x_1 = f(s_2, m_1)$, de $x_1 = f(s_1, m_0)$ ou ainda de $x_1 = f(s_0, m_0)$. Como há duas marcas diferentes haverá uma ambigüidade quando a função h for detectar a marca que pode ser tanto m_0 como m_1. Uma última característica inadequada é que se x_2 for o sinal marcado então o sinal hospedeiro pode ser determinado. Supondo-se novamente que o atacante tem capacidade computacional ilimitada, não importa se

computacionalmente é difícil calcular a inversa de *f*. Não importa se *f* é uma função injetora do tipo "one-way" ; neste modelo o atacante sempre consegue calcular a inversa se a função for injetora.

A mesma situação é apresentada de forma simplificada sob outro ponto de vista na Fig. 6. Nesta figura S e X são o mesmo conjunto e somente estão representadas as setas que partem de s_0 e s_4 na Fig. 5. Os círculos representam esferas onde a distorção com relação a x_i é menor ou igual a $d_{máx}$. Logo, somente os pontos *s* que se encontram dentro de uma esfera de raio $d_{máx}$ podem gerar o sinal marcado x_i . Para os demais, seria distorção além do limiar aceitável. Um atacante que quer descobrir qual *s* gerou um dado x_i pode restringir sua busca aos pontos no interior da esfera ao redor de x_i .

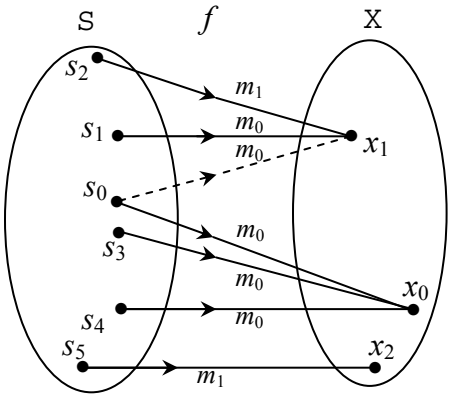


Figura 5 - Mapeamento indesejável para a inserção de marca pública.

A Fig. 6 mostra também que pode haver sobreposição das esferas. Pontos *s* que estão na região de sobreposição poderiam, segundo o critério de distorção, gerar os sinais marcados x_i de qualquer uma das esferas sobrepostas. Todavia, como o processo de inserção da marca é uma função, uma vez especificados *s* e *m* estará especificado o sinal marcado *x*. Isto leva na realidade a uma partição da região de sobreposição que é mostrada na Fig. 6 como uma linha pontilhada. Cada um dos pontos *s* da região de sobreposição será mapeado em um x_i ou em outro, pela função *f* . Pode ser então definida entorno de cada sinal marcado x_i uma região de sinais hospedeiros s_i que podem gerar aquele sinal marcado.

Quanto maior o número de sinais hospedeiros em cada região maior a incerteza sobre qual sinal hospedeiro gerou um dado sinal marcado. Para uma dada distorção, a

[21] A função *f* deve ser bijetora para ter inversa, mas se *f* for injetora pode-se definir uma função inversa modificando o contra-domínio de *f* para a sua imagem.

incerteza será sempre limitada pelo número de pontos no interior de uma esfera de raio $d_{máx}$. Ao se particionar a região de interseção das esferas diminui-se o número de pontos na região de incerteza e portanto diminui-se a própria incerteza. Para evitar esta redução, podem-se afastar as esferas umas das outras. Contudo, isto aumenta a distorção no processo de inserção da marca. Portanto, se o processo de inserção é uma função de *s* e *m* somente, para se aumentar a segurança é necessário aumentar a distorção. Isto pode não ser o comportamento adequado para o sistema procurado.

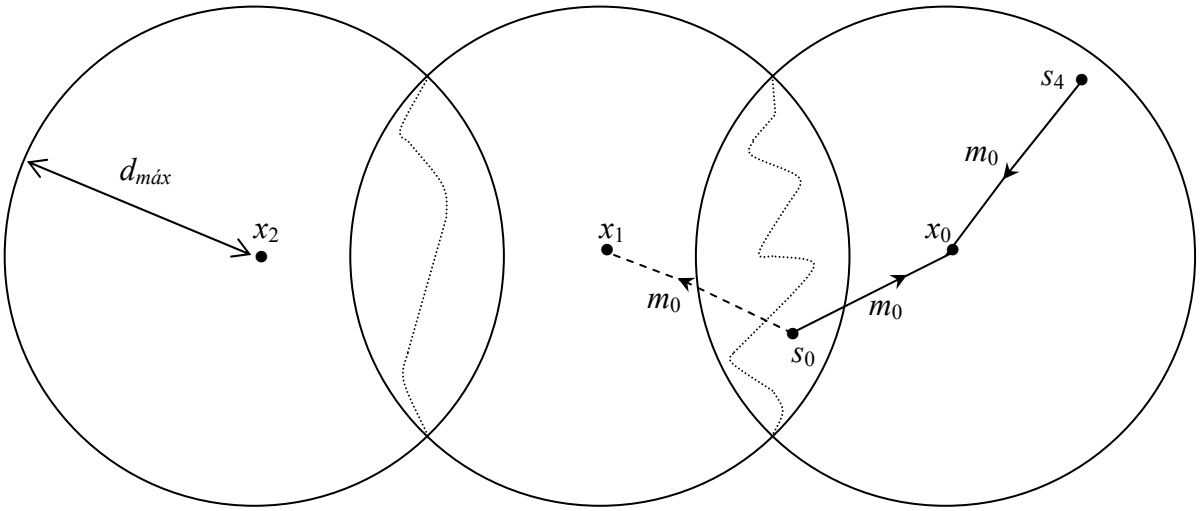


Figura 6 - Inserção de marca e partição do espaço.

Alternativamente, considera-se um sistema probabilístico com várias funções $f_k(s,m)$. Neste caso não é necessário que cada função f_k seja "muitos-para-1", já que, ao contrário do caso anterior, dados *x* e *m* onde $x = f_k(s,m)$, não se sabe qual função f_k inverter para obter s. Além disto, o diagrama mostrado na Fig. 6 onde o par (s_0,m_0) é mapeado tanto em x_0 quanto em x_1 é válido. Portanto, um possível atacante não consegue particionar a região de interseção de duas esferas, pois os pontos no interior desta região podem, a princípio, gerar ambos os pontos x_0 ou x_1 dependendo da função f_k que foi utilizada no processo de inserção da marca.

Para maximizar a incerteza é preciso fazer com que o maior número de pontos possível esteja dentro das esferas de distorção máxima e isto, neste caso, é limitado somente pelo raio $d_{máx}$. Não há restrição em trazer os pontos x_i mais para perto uns dos

outros e a região de sobreposição pode ser aumentada sem diminuir a incerteza. Como somente os pontos s que estão na região de sobreposição de duas esferas são compartilhados, é interessante notar que ao se aumentar o tamanho da região de sobreposição diminuindo-se a distorção, mais pontos são compartilhados e a segurança não é alterada. Logo, diminuindo-se a distorção não se afeta a segurança. Portanto, aparentemente, um sistema probabilístico deve apresentar um melhor desempenho para a inserção de uma marca d'água digital que o sistema determinístico apresentado anteriormente. Um sistema com chave pública onde uma chave privada especifica qual função f_k é utilizada para fazer $x = f_k(s,m)$, também deve ter um bom comportamento (embora a segurança neste caso seja apenas computacional se for possível calcular a chave privada a partir da pública). *Segundo esta análise, um sistema que não utiliza chaves deve ser preferencialmente probabilístico*.

Seguindo a idéia de que um sistema probabilístico se adapta melhor às necessidades de um sistema que insere marcas d'água, propõe-se o seguinte sistema ilustrativo que insere apenas um bit de informação de marca d'água. O sistema funciona assim:

- **Inserção**: A inserção da marca d'água é feita controlando-se a paridade do sinal hospedeiro a ser marcado. Paridade par significa que a marca d'água tem valor zero e paridade ímpar significa que tem valor um. Para controlar a paridade é *aleatoriamente* escolhido um bit no sinal hospedeiro e este bit é então modificado se necessário.
- **Detecção**: Para detectar a marca d'água simplesmente é calculada a paridade do sinal marcado, paridade par e ímpar representam as marcas 0 e 1 respectivamente.

Por exemplo, para colocar a marca "0" no sinal hospedeiro 01010111 aleatoriamente se escolhe uma posição qualquer. Neste caso a terceira foi escolhida. Como se deseja inserir a marca "0" é necessário que a paridade seja par. Calculando-se a paridade do sinal hospedeiro vê-se que ela é ímpar e portanto é necessário modificar o terceiro bit obtendo o sinal marcado 01110111. Para detectar a marca basta calcular a paridade do sinal marcado.

Este sistema não tem muita aplicação prática mas o mesmo possui uma propriedade interessante. Apesar de ser muito fácil detectar a marca (basta calcular a paridade do sinal) é muito difícil saber qual bit foi modificado no processo de inserção. Como a escolha do bit a ser modificado é feita aleatoriamente, um possível atacante deve ser capaz de estimar qual bit foi modificado somente através da inspeção do tipo de distorção causada. Por exemplo:

Se o sinal hospedeiro é uma imagem e os bits representam diferentes regiões desta imagem, para saber qual bit foi modificado deve-se observar qual parte da imagem está degradada. Todavia, nem sempre é fácil determinar qual a alteração feita a partir do tipo de distorção que esta causa. Se não for possível estimar quais bits foram modificados no processo de inserção da marca a partir das distorções existentes no sinal marcado, então o sistema ilustrativo proposto acima não dá informação adicional além da representada por x a um atacante que deseje obter o sinal hospedeiro.

Um problema básico do processo de inserção descrito acima é que a taxa de alarmes falsos é muito alta (o sistema não é idôneo). Ao se utilizar o detetor é detectada uma marca em todo sinal, marcado ou não. Logo, *todo sinal **não** marcado é considerado um sinal marcado*. Para amenizar este problema pode-se usar uma técnica simples que é ilustrada a seguir. Particiona-se o conjunto de marcas {0,1} em dois subconjuntos: o de marcas válidas $V = \{0\}$ e o de marcas inválidas $I = \{1\}$. Considera-se que se um sinal possui uma marca inválida, então, na realidade, ele não está marcado. De imediato, é possível notar que o número de marcas disponíveis diminui fazendo-se esta consideração. Contudo, a probabilidade de alarme falso também. Os sinais não marcados de paridade ímpar não serão erroneamente identificados com sinais marcados de marca d'água igual a "1". *Somente os sinais marcados de paridade par serão erroneamente considerados sinais marcados*. Neste caso simplificado a partição do espaço de marcas não resolve o problema. Apenas o ameniza, mas o exemplo é apenas ilustrativo.

Embora o sistema de paridade seja bastante seguro contra um atacante que tem o intuito de descobrir qual o sinal hospedeiro, isto não é suficiente. Para este sistema, um atacante é capaz de retirar a marca d'água sem saber qual o sinal hospedeiro, basta modificar um único bit. Por exemplo, um atacante pode modificar o último bit do sinal marcado 01110111, obtendo 01110110 e removendo a marca "1". É necessário que o sistema também seja robusto contra ataques que visam somente retirar a marca do sinal marcado. Para isto, o detetor deve ser capaz de detectar a marca em sinais que foram modificados durante um ataque. Isto implica em uma certa redundância onde pequenas alterações no sinal marcado não são suficientes para retirar a marca. Se o sistema de paridade inserisse mais informação, ao invés de um único bit, talvez fosse possível fazer com que a marca fosse resistente a ataques que visam modificar o sinal marcado. Como a

marca é a informação a ser transmitida, além de um ataque ao sinal marcado o processo de inserção também pode ser visto como parte de um canal de comunicação hostil, como mostrado na Fig. 7. Também de acordo com a figura, pode ser interessante codificar a marca para corrigir erros causados por este canal.

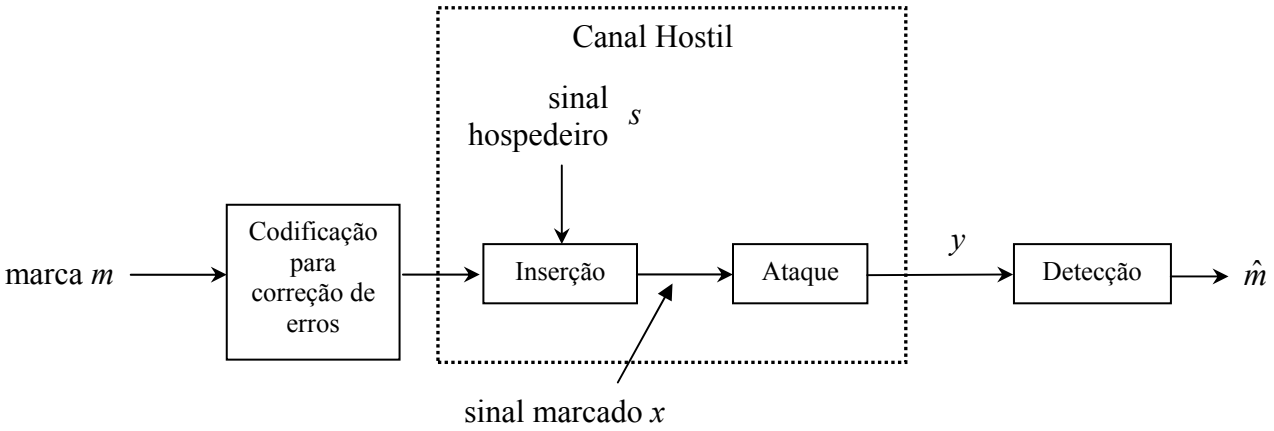


Figura 7 - Modelo de canal equivalente para a inserção da marca.

Para isto, pode ser utilizado um código corretor de erros. Considera-se o seguinte: Se o sistema inserisse 3 bits de informação, utilizando um código de repetição na marca é possível inserir redundância suficiente para detectar a marca mesmo após algumas modificações no sinal marcado. Alternativamente, um código corretor de erros melhor pode ser utilizado. É importante fazer a ressalva que embora o atacante modifique o sinal marcado x obtendo um sinal atacado y, *não é necessário corrigir erros em x mas sim em m*. Modificações que degradem x mas que não modifiquem m não são consideradas ataques bem sucedidos neste modelo. Não é necessário diferenciar entre modificações do sinal marcado que causem efeitos idênticos na marca. Estes últimos pontos devem ser melhor esclarecidos com um exemplo. Se um atacante modificar o sinal marcado $x = 01110111$ para $y = 11110111$ e isto não causar qualquer alteração na marca inserida, não há necessidade de correção. Por outro lado, se esta modificação causar uma mudança na estimativa da marca que é produzida pelo receptor, então deve haver um modo de melhorar *a estimativa*. Supondo que a marca inserida foi 000 e que uma modificação no quarto ou no quinto bit do sinal marcado faça com que a estimativa gerada se torne 001 então basta corrigir a estimativa de 001 para 000. Não é preciso saber qual bit do sinal marcado foi modificado.

No sistema de paridade, o detetor da marca simplesmente multiplica o sinal marcado por uma matriz de paridade de uma linha. Para inserir mais informação do que apenas um bit de paridade o esquema de codificação precisa ser modificado. Várias linhas precisam estar presentes na matriz de paridade. Como então o esquema pode ser modificado? Heegard [17] propõe um método de codificação para memórias com defeito que pode ser utilizado para a inserção de marcas d'água digitais. Segue-se uma descrição do método e de como utilizá-lo em sistemas de marca d'água.

3.2 Memórias com Defeitos e Marcas d'Água

Heegard estudou o problema de como utilizar memórias de computador com defeitos e propôs a utilização de códigos de bloco lineares de uma maneira engenhosa para maximizar a capacidade de armazenamento dessas memórias. O trabalho considera que existem dois tipos de defeitos em uma célula de memória. A célula de memória pode estar "travada"[22] em 0 ou 1 e não sair deste valor, o que é considerado um defeito de fabricação. Podem também ocorrer inversões ocasionais dos valores guardados de 0 para 1 ou vice-versa. Isto é considerado apenas uma falha aleatória. As células travadas não conseguem guardar nenhuma informação. O método proposto considera apenas o caso binário mas pode ser diretamente estendido para o caso apresentado aqui onde cada célula de memória pode estar em um de q estados diferentes. Considerando-se inicialmente apenas os defeitos de fabricação, somente o codificador, que escreve os dados na memória, conhece a localização das células travadas. O decodificador, que lê os dados, não sabe quais células estão travadas. Como trata-se de um método para codificação em bloco, considera-se uma n-upla de células de memória, cada célula com q estados possíveis.

Um ótimo exemplo é proposto por Chou [5]. Considera-se o caso de três células de memória binárias (*i.e* uma memória de três bits) onde uma das células encontra-se travada. Neste caso ainda é possível transmitir (ou guardar) 2 bits de informação mesmo se somente o codificador, que escreve na memória, souber qual a célula que está travada. Para isto, particiona-se o conjunto de palavras de 3 bits nos 4 conjuntos mostrados abaixo:

[22] Em inglês: *stuck-at*

A_0	A_1	A_2	A_3
000 111	001 110	010 101	011 100

Supõe-se que o defeito está na segunda célula que encontra-se travada em 1. Se deseja-se transmitir a informação binária 11_b, por exemplo, escolhe-se no conjunto A_3 uma palavra compatível com o padrão de erros. A palavra em questão é 011, a palavra que tem 1 na segunda posição. Escreve-se 011 na memória. Como a segunda célula já está travada em 1, ao ler a memória o decodificador obtém exatamente a palavra escrita. Como o mesmo conhece a tabela acima sabe que a palavra 011 faz parte de A_3 e portanto a informação transmitida é $11_b = 3$. Todos os conjuntos A_w acima são compatíveis com todos os tipos de defeito de uma só célula. Através deste esquema sempre é possível transmitir 2 bits de informação.

Devem-se escrever na memória somente palavras que a n-upla de células pode guardar sem erro. Por exemplo, no caso de 8 células binárias em que a primeira célula de memória está travada em 1, o codificador não pode escrever na memória nenhuma palavra que comece com 0 tal como **0**1110111, por que se assim o fizer, a palavra não será recuperada depois sem modificações. Para que o codificador sempre consiga escrever na memória somente palavras que podem ser guardadas sem erro, cada conjunto A_w deve conter uma palavra-código (palavra de n dígitos q-ários) compatível com cada tipo de defeito que pode ocorrer em uma n-upla de células de memória. O esquema de codificação é simples, o codificador simplesmente acha no conjunto A_w, onde w é a informação a ser gravada (escrita), uma palavra compatível com os defeitos da memória. Esta palavra pode ser escrita na memória sem modificações e portanto quando o decodificador ler a memória, pode identificar sem erro a qual conjunto a palavra pertence e então recuperar a informação transmitida w.

O esquema de codificação por blocos para memórias com defeito que se segue, assim como a descrição feita são de Heegard [17]. O esquema consiste em particionar as q^n n-uplas possíveis de serem escritas em uma memória de n células em $w = q^l$ conjuntos A_w onde:

$$A_w = \left\{ \mathbf{y} \,\middle|\, \mathbf{y}\mathbf{H}^t = \mathbf{w} \right\} \tag{3.3}$$

e **y** é uma *n*-upla dentre as q^n *n*-uplas de *n* células existentes (vista como um vetor linha). Cada conjunto A_w é o conjunto de todas as *n*-uplas que tem a síndrome **w** quando multiplicadas pela matriz **H**. Se um decodificador, que só lê a memória, conseguir distinguir sem erro entre os *w* conjuntos, *l* dígitos *q*-ários de informação podem ser guardados na memória. Supondo que **H** é a matriz de paridade de um código linear $C = [n, k, d]$ (*i.e.* de comprimento *n*, com *k* dígitos *q*-ários de informação e distância mínima de Hamming *d*) então o código tem matriz geradora **G** de tamanho $k \times n$, e **H** é de tamanho $(n\text{-}k) \times n$. Cabe ressaltar que *l* representa o número de dígitos *q*-ários de informação que o esquema deseja efetivamente guardar na memória, enquanto que *k* representa o número de dígitos *q*-ários de informação que o código linear a ser usado pelo esquema possui. Para obter a informação escrita o decodificador simplesmente calcula $\hat{\mathbf{w}} = \mathbf{y}\mathbf{H}^t$. Considera-se que a memória tem *u* defeitos de fabricação nas localizações $1 \le i_1 < i_2 < \ldots < i_u \le n$ e que estas células estão travadas nos valores $s_1, s_2, \ldots, s_u \in \{0,1,\ldots, q\text{-}1\}$. Seja **v** qualquer vetor que tem a síndrome **w** , *i.e.* um vetor tal que $\mathbf{w} = \mathbf{v}\mathbf{H}^t$. O codificador tenta resolver a equação matricial:

$$\mathbf{d}\mathbf{G}' = \left[s_1 - x_{i_1}, s_2 - x_{i_2}, \ldots, s_u - x_{i_u} \right] \tag{3.4}$$

para o vetor **d** de comprimento *k* onde

$$\mathbf{G}' = \left[\mathbf{g}_{i_1}, \mathbf{g}_{i_2}, \ldots, \mathbf{g}_{i_u} \right] \tag{3.5}$$

é a submatriz de **G** que tem como colunas as colunas onde estão os defeitos na memória. Se uma solução é achada então a palavra **x** dada por

$$\mathbf{x} = \mathbf{v} + \mathbf{d}\mathbf{G} \tag{3.6}$$

é gravada na memória. Observa-se que $x_{i_1} = s_1, x_{i_2} = s_2, \ldots, x_{i_u} = s_u$ e que $\mathbf{x}\mathbf{H}^t = \mathbf{w}$ de forma que **x** será corretamente decodificada.

Para que se possa escrever *l* dígitos *q*-ários de informação no esquema proposto por Heegard, o produto de uma palavra de *n* dígitos **y** pela matriz de paridade transposta $\mathbf{y}\mathbf{H}^t$, deve ser capaz de gerar q^l valores diferentes, de forma que existam q^l conjuntos A_w diferentes. A palavra de *n* dígitos *q*-ários **y** pertence a um subespaço de dimensão $1 \times n$. Além disso, **H** é de tamanho $(n\text{-}k) \times n$ e o produto $\mathbf{y}\mathbf{H}^t$ pertence a um subespaço de dimensão $1 \times (n\text{-}k)$. Portanto, para gerar q^l vetores diferentes é necessário que $(n\text{-}k) \ge l$. Resumindo, o esquema proposto por Heegard é capaz de transmitir *l* dígitos *q*-ários de

informação utilizando um código linear $C = [n, k, d]$ somente se valer a desigualdade $\boldsymbol{n - k \geq l}$.

No esquema de Heegard admite-se que os valores guardados nas células de memória travadas sejam arbitrários. Em outras palavras, o codificador (que grava os dados na memória) não controla estes valores. Quando utilizando o esquema de Heegard para inserir marcas d'água, a informação que deseja-se transmitir é a marca d'água e os valores guardados nas células de memória travadas são valores do sinal hospedeiro. O sinal marcado é obtido através da modificação das células de memória que não estão travadas. Supondo que muitas das células de memória estão travadas, apenas poucos dígitos do sinal hospedeiro podem ser modificados para gerar o sinal marcado. O sinal marcado é o valor realmente gravado (escrito) na memória e posteriormente lido pelo decodificador (o detetor da marca d'água).

Para inserir marcas d'água, tem-se uma limitação da discrepância entre o sinal hospedeiro e o sinal marcado. Como há uma distorção máxima tolerável, esta discrepância, que é quantificada pela *distância de Hamming*, é limitada superiormente. Supõe-se que a distorção máxima tolerável é de e dígitos q-ários. O número de dígitos que podem ser alterados, e, é definido pela aplicação. Por exemplo, no caso onde cada célula de memória pode estar em um de 3 estados, isto equivale a ter um sinal hospedeiro codificado de forma ternária. Se o usuário admite uma distorção máxima de $e = 2$, para o sinal hospedeiro 01221021 o sinal marcado 02221121 é admissível, mas o sinal 02221120 que modifica 3 posições, não o é. Podem-se modificar **quaisquer** e dígitos q-ários do sinal, contanto que no máximo e sejam alterados. Todas as outras (n-e) posições devem permanecer inalteradas. É como se o sinal hospedeiro tivesse sido guardado em uma memória que foi posteriormente danificada, deixando todas as (n-e) posições travadas. Por exemplo, se deseja-se marcar um sinal hospedeiro de 8 dígitos de comprimento e tolera-se uma discrepância entre o sinal marcado e o sinal hospedeiro de $e = 3$ dígitos, então pode-se supor que o sinal hospedeiro dado por 11010101_b foi guardado em uma memória, a qual foi parcialmente travada com o conteúdo deste sinal. Dada a limitação de distorção tolerada, somente 3 dígitos q-ários podem ser modificados e os outros 5 se encontram travados nos valores do sinal hospedeiro. Se as 5 primeiras células de memória encontram-se travadas,

então, o sinal marcado deve começar com a seqüência de dígitos (do sinal hospedeiro) 11010_b. Um possível sinal marcado seria **11010**000.

Para que um possível atacante não tenha como descobrir onde está a informação sobre a marca d'água convém escolher os *e* dígitos *q*-ários a serem modificados aleatoriamente. Todas as outras (*n*-*e*) posições devem permanecer inalteradas. O ideal é que através dos *e* dígitos *q*-ários que se podem modificar no sinal hospedeiro se possa transmitir também *e* dígitos *q*-ários de informação sobre a marca d'água. Seria uma pena, por exemplo, poder modificar 8 dígitos *q*-ários e conseguir transmitir com isto apenas 5 dígitos de informação. Logo, para este sistema ideal, *todo dígito q-ário que não pode ser mudado para transmitir informação da marca representa um dígito q-ário de memória que está travada (e, em conseqüência, do sinal hospedeiro).* Se *u* é o número de dígitos *q*-ários travados e *e* designa o número de dígitos *q*-ários que se podem modificar e que também se desejam transmitir sobre a marca, segue-se que $e + u = n$. Portanto, $u = n - e$. Como o esquema proposto por Heegard transmite *l* dígitos *q*-ários de informação, para usá-lo para transmitir *e* dígitos *q*-ários é necessário que $l \geq e$. Substituindo esta desigualdade na expressão acima obtém-se: $\boldsymbol{u \geq n - l}$.

Um resultado obtido por Heegard diz que para que cada conjunto A_w seja compatível com qualquer configuração de até *u* dígitos *q*-ários travados é *necessário e suficiente* que qualquer conjunto de *u* colunas da matriz geradora **G** tenha posto igual a *u*. Para que isto aconteça é necessário que qualquer conjunto de *u* colunas seja linearmente independente. **G** é a matriz geradora do código $C = [n, k, d]$ e também a matriz de paridade do código dual $C_d = [n_d, k_d, d_d]$. A condição da matriz de paridade de um código ter qualquer conjunto de *u* colunas linearmente independentes implica que a distância mínima entre as palavras do código é pelo menos igual a $u+1$ [22]. Logo, a distância mínima do código dual C_d é

$$\boldsymbol{d_d \geq u + 1} \tag{3.7}$$

Substituindo as desigualdades obtidas acima, obtém-se

$$d_d \geq u+1 \geq n - l+1 \geq n - (n - k)+1 \tag{3.8}$$

Logo,

$$d_d \geq n_d - k_d + 1 \tag{3.9}$$

pois (n-k) é o número de dígitos de informação no código dual. Pelo limitante de Singleton para qualquer código linear [n, k, d] tem-se que

$$d \leq n - k + 1 \tag{3.10}$$

Esta relação é satisfeita com igualdade somente quando o código for MDS[23]. Nota-se então que o código dual C_d é um código MDS, pois ambas as condições só podem ser satisfeitas simultaneamente com igualdade. Como o dual de um código MDS é também MDS [22], o código C que deve ser usado para codificar a marca d'água através do esquema de Heegard (utilizando todos os bits disponíveis) é um código MDS.

A recíproca também é verdadeira, isto é, se for utilizado um código MDS transmitindo k bits de informação no esquema de Heegard, sempre é possível achar uma palavra código em qualquer um dos conjuntos A_w que tenha as outras (n-k) posições fixadas arbitrariamente. Para um código MDS com k dígitos q-ários de informação, quaisquer ($n - k$) colunas da matriz de paridade **H** são linearmente independentes [22] e segundo Heegard esta condição é necessária e suficiente para que os conjuntos A_w sejam compatíveis com qualquer configuração de k dígitos travados.

A família de códigos MDS binários é limitada aos triviais códigos de repetição e paridade. Os códigos MDS não triviais conhecidos são os códigos de Reed-Solomon e os códigos estendidos derivados destes. Contudo, códigos de Reed-Solomon existem para outros corpos que não $GF(2)$ e portanto para usar estes códigos é necessário codificar cada componente do sinal hospedeiro em outro corpo finito $GF(q)$ que não o binário, como por exemplo $GF(256)$. Além disto, o comprimento do vetor que representa o sinal hospedeiro nesta forma deve ser igual a q-1. Uma vez estabelecidos estes parâmetros pode ser utilizado um código para qualquer nível de distorção desejado pois existe um código de Reed-Solomon para cada número de dígitos de informação de 1 a q-1.

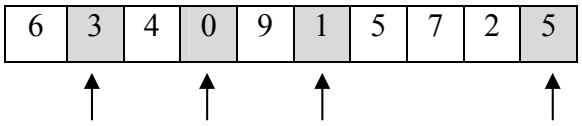
Segue-se um exemplo do método proposto. Supõe-se que após uma transformação de adaptação ao corpo finito, o sinal hospedeiro é representado por um vetor de componentes em $GF(11)$ de comprimento 10. Seja 4 a máxima distorção aceitável para o sinal assim representado, isto é, até 4 dígitos podem ser modificados ($d_H(s,x) \leq 4$) . O sinal hospedeiro é **s** = (6, 3, 4, 0, 9, 1, 5, 7, 2, 5). Deseja-se gerar o sinal marcado **x**. Como a distorção máxima aceitável é de 4 dígitos q-ários, os outros 6 dígitos são tidos como

[23] Do inglês: ***M**aximum **D**istance **S**eparable*

travados e utiliza-se um código de Reed-Solomon de parâmetros [10,6,5] , *i.e.* comprimento 10, 6 dígitos de informação e distância mínima 5. A matriz geradora e a matriz de paridade deste código na forma sistemática são as seguintes:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 3 & 5 & 8 \\ 0 & 1 & 0 & 0 & 0 & 0 & 3 & 10 & 7 & 7 \\ 0 & 0 & 1 & 0 & 0 & 0 & 4 & 4 & 8 & 6 \\ 0 & 0 & 0 & 1 & 0 & 0 & 5 & 8 & 7 & 4 \\ 0 & 0 & 0 & 0 & 1 & 0 & 7 & 4 & 10 & 8 \\ 0 & 0 & 0 & 0 & 0 & 1 & 3 & 5 & 8 & 1 \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} 10 & 8 & 7 & 6 & 4 & 8 & 1 & 0 & 0 & 0 \\ 8 & 1 & 7 & 3 & 7 & 6 & 0 & 1 & 0 & 0 \\ 6 & 4 & 3 & 4 & 1 & 3 & 0 & 0 & 1 & 0 \\ 3 & 4 & 5 & 7 & 3 & 10 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.11)$$

As síndromes que representam as marcas são vetores de 4 componentes $\mathbf{w} = (w_1,w_2,w_3,w_4)$, mas para controlar a probabilidade de falso alarme nem todas as marcas devem ser consideradas válidas. Sugerem-se utilizar somente 2 dígitos q-ários para as marcas válidas e deixar os outros 2 dígitos disponíveis para representar marcas inválidas. Arbitrariamente escolhem-se que as marcas válidas serão representadas por síndromes que começam com dois zeros, $(0,0,w_3,w_4)$. Como há 11 possibilidades para cada uma das duas componentes restantes, existe um total de $11^2 = 121$ marcas válidas diferentes. Supõe-se que deve ser inserida a marca 77 decimal que equivale a 70_{11}. Esta marca é então representada pela síndrome $\mathbf{w} = (0,0,7,0)$. Em seguida, são aleatoriamente escolhidas as 4 posições que serão modificadas no vetor sinal hospedeiro, como indicado no diagrama:



Posições selecionadas para alteração:
segunda, quarta, sexta e décima posições

Para achar no conjunto A_w uma palavra compatível com esta restrição utiliza-se o método proposto por Heegard. Como a matriz de paridade está na forma sistemática, para achar um vetor com síndrome igual a (0,0,7,0) basta acrescentar zeros à esquerda obtendo-se $\mathbf{v} = (0,0,0,0,0,0,0,0,7,0)$. Então resolve-se:

$$\mathbf{dG}' = \mathbf{s}' - \mathbf{v}', \quad (3.12)$$

isto é,

$$\mathbf{d}\begin{bmatrix} 1 & 0 & 0 & 1 & 3 & 5 \\ 0 & 0 & 0 & 3 & 10 & 7 \\ 0 & 1 & 0 & 4 & 4 & 8 \\ 0 & 0 & 0 & 5 & 8 & 7 \\ 0 & 0 & 1 & 7 & 4 & 10 \\ 0 & 0 & 0 & 3 & 5 & 8 \end{bmatrix} = \begin{bmatrix} 6-0 & 4-0 & 9-0 & 5-0 & 7-0 & 2-7 \end{bmatrix}. \qquad (3.13)$$

Como **G´** tem colunas linearmente independentes sua inversa existe e **d** pode ser calculado como **d** = (6,6,4,0,9,4). Portanto, o sinal marcado **x** é dado por **x** = **v** + **dG** = (6,6,4,0,9,4,5,7,2,3). Comparando **x** com **s** observa-se que somente as posições escolhidas foram modificadas.

s =	6	3	4	0	9	1	5	7	2	5	← sinal hospedeiro
x =	6	6	4	0	9	4	5	7	2	3	← sinal marcado

Finalmente, para detectar a marca basta multiplicar o sinal marcado **x** por $\mathbf{H}^t$ e obter:

$$\mathbf{w} = \begin{bmatrix} 6 & 6 & 4 & 0 & 9 & 4 & 5 & 7 & 2 & 3 \end{bmatrix} \left(\begin{bmatrix} 10 & 8 & 7 & 6 & 4 & 8 & 1 & 0 & 0 & 0 \\ 8 & 1 & 7 & 3 & 7 & 6 & 0 & 1 & 0 & 0 \\ 6 & 4 & 3 & 4 & 1 & 3 & 0 & 0 & 1 & 0 \\ 3 & 4 & 5 & 7 & 3 & 10 & 0 & 0 & 0 & 1 \end{bmatrix} \right)^t \qquad (3.14)$$

$$= \begin{bmatrix} 0 & 0 & 7 & 0 \end{bmatrix}$$

Como esta marca começa com dois zeros a mesma é válida e o sinal é considerado marcado, os dois dígitos finais (70_{11}) trazem a informação da marca.

Embora no exemplo acima seja difícil saber qual o sinal hospedeiro é muito fácil retirar a marca d'água. Se um atacante modificar apenas um dígito a marca é modificada. Se o atacante modificar somente o primeiro dígito do sinal marcado acima de 6 para 7 a marca obtida passa a ser **w** = (10,8,2,3). Obviamente, o sistema não é seguro contra este tipo de ataque. Para torná-lo mais seguro é necessário que modificações que ocorram na síndrome devido a pequenas modificações no sinal marcado possam ser corrigidas. Para isto, é necessário fazer a correção de erros na síndrome, o que é tratado na próxima seção.

3.3 Correção de Erros

Um atacante tenta remover a marca de um sinal fazendo modificações e através de grandes alterações no sinal marcado sempre é possível remover a marca inserida. Contudo, tais alterações também degradam excessivamente a qualidade do sinal. Não é preciso que o sistema de marca d'água digital seja capaz de recuperar a marca de um sinal marcado que sofreu várias degradações e está com péssima qualidade, é preciso apenas recuperar a marca naqueles sinais que apresentam qualidade boa. Tais sinais encontram-se a uma distância inferior a $d_{máx}$ do sinal hospedeiro original.

Em termos da distância de Hamming, um ataque causa tanta distorção quantos dígitos q-ários do sinal forem alterados. Todavia, como o ataque é feito no sinal marcado, que já é um sinal distorcido, se feito inteligentemente a distorção do ataque pode cancelar ou apenas modificar a distorção causada pelo processo de inserção da marca sem piorar (ou possivelmente melhorando) a qualidade do sinal.

Como o ataque a um sinal marcado consiste em fazer com que o detetor não mais detecte a presença da marca no sinal é interessante considerar o processo de detecção. Até agora o processo de detecção considerado é basicamente a multiplicação do sinal pela matriz de paridade de um código MDS, *i.e.* $\hat{\mathbf{w}} = \mathbf{y}\mathbf{H}^t$. Ainda utilizando a notação matricial e supondo que um atacante modifica a j-ésima componente do sinal marcado $\mathbf{x}$ de δ para obter o sinal atacado $\mathbf{y}$. Haverá uma modificação na estimativa de $\mathbf{w}$ obtida no detetor. Considerando a equação acima na forma transposta ($\mathbf{H}\,\mathbf{y}^t = \mathbf{w}^t$) pode-se escrever:

$$\begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,j} & \dots & h_{1,n} \\ h_{2,1} & h_{2,2} & \dots & h_{2,j} & \dots & h_{2,n} \\ h_{3,1} & h_{3,2} & \dots & h_{3,j} & \dots & h_{3,n} \\ \vdots & \vdots & & \vdots & & \vdots \\ h_{n-k,1} & h_{n-k,2} & \dots & h_{n-k,j} & \dots & h_{n-k,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j + \delta \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} w_1 + \delta h_{1,j} \\ w_2 + \delta h_{2,j} \\ \\ \vdots \\ \\ w_{n-k} + \delta h_{n-k,j} \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \\ \vdots \\ \\ w_{n-k} \end{bmatrix} + \delta \begin{bmatrix} h_{1,j} \\ h_{2,j} \\ \\ \vdots \\ \\ h_{n-k,j} \end{bmatrix} \quad (3.15)$$

Nota-se que para cada componente alterada em um ataque, o atacante adiciona à marca original um múltiplo de uma das colunas da matriz de paridade. A coluna adicionada depende de qual componente é alterado. Para muitas posições no sinal marcado,

modificando somente o valor da componente correspondente faz com que todos os dígitos da síndrome sejam modificados. Devido a esta grande capacidade de alteração da síndrome a partir de pequenas modificações no sinal marcado, a utilização de um código convencional para correção de erros pode parecer pouco interessante. No entanto esta abordagem pode ser bastante efetiva. O atacante também sofre algumas restrições no tipo de modificação que pode causar à síndrome se poucas posições puderem ser alteradas. Por exemplo, se a matriz **H** estiver na forma sistemática então modificando apenas um dígito do sinal marcado leva a apenas duas possibilidades para as mudanças que podem ocorrer na síndrome: ou todos os dígitos são alterados ou apenas um dígito é alterado. Não é possível alterar apenas parte (diferente de 1) dos dígitos.

Através da visualização do problema em forma matricial, não fica claro no entanto quais são as restrições impostas ao atacante se somente poucos dígitos do sinal marcado puderem ser alterados. Alternativamente, pode-se tentar utilizar a notação polinomial. Se um código MDS cíclico, tal como um código de Reed-Solomon for usado para inserir uma marca d'água, o sinal marcado resultante pode ser expresso como um polinômio em λ da seguinte forma:

$$x(\lambda) = p(\lambda)g(\lambda) + v(\lambda) \tag{3.16}$$

Esta equação é a forma polinomial da eq. 3.6 com **d** substituído por $p(\lambda)$. Na equação acima, $g(\lambda)$, de grau $(n\text{-}k)$, é o polinômio gerador do código $C = [n, k, d]$ utilizado para inserir a marca, $p(\lambda)$ é um polinômio qualquer de grau menor que k e $v(\lambda)$ é um polinômio que tem síndrome igual a $w(\lambda)$ (*i.e.* **w**). O processo de detecção pode ser implementado pela multiplicação pelo polinômio $h(\lambda)$ (onde $g(\lambda)\, h(\lambda) = \lambda^n\text{-}1$) seguida de redução módulo λ^n-1. Toda a informação sobre a marca encontra-se em $v(\lambda)$ pois:

$$x(\lambda)h(\lambda) \ \ mod\ (\lambda^n\text{-}1) = [p(\lambda)g(\lambda)h(\lambda) + v(\lambda)h(\lambda)]\ mod\ (\lambda^n\text{-}1) = v(\lambda)h(\lambda) \ \ mod\ (\lambda^n\text{-}1) \tag{3.17}$$

No caso de um ataque pode-se considerar que o sinal sofre a adição de um polinômio atacante $a(\lambda)$ que tem poucos coeficientes não nulos e portanto altera o sinal marcado em um pequeno número de posições. Expressando este ataque por meio de polinômios, tem-se

$$y(\lambda) = x(\lambda) + a(\lambda) = p(\lambda)g(\lambda) + v(\lambda) + a(\lambda) \tag{3.18}$$

Se não houver nenhuma informação adicional a respeito de $v(\lambda)$ não há como diferenciar $v(\lambda)$ de $a(\lambda)$ em $y(\lambda)$. Isto ocorre no caso da equação (3.18) acima onde não há código corretor de erros aplicado a $v(\lambda)$. Por outro lado, se houver uma estrutura ao polinômio $v(\lambda)$, pode ser possível remover a interferência de $a(\lambda)$. Fazendo $v(\lambda)$ ser múltiplo de um polinômio $f(\lambda)$, *i.e.* $v(\lambda) = i(\lambda)f(\lambda)$, pode ser possível diferenciar $v(\lambda)$ de $a(\lambda)$. Sugere-se para esta tarefa uma escolha simples, bastando fazer $f(\lambda)$ um polinômio que divide $g(\lambda)$. Desta forma tem-se

$$\begin{aligned} y(\lambda) &= x(\lambda) + a(\lambda) = p(\lambda)g(\lambda) + v(\lambda) + a(\lambda) \\ &= p(\lambda)g(\lambda) + i(\lambda)f(\lambda) + a(\lambda) \\ &= p'(\lambda)f(\lambda) + i(\lambda)f(\lambda) + a(\lambda) \\ &= [p'(\lambda) + i(\lambda)]f(\lambda) + a(\lambda) \end{aligned} \tag{3.19}$$

Do mesmo modo que em um código cíclico, onde $x(\lambda) = p(\lambda)g(\lambda)$, é possível corrigir erros considerando $y(\lambda)$ como uma palavra código alterada por $a(\lambda)$ de um código com polinômio gerador $f(\lambda)$. O sinal marcado torna-se neste caso uma palavra código de um código com polinômio gerador $f(\lambda)$, $x(\lambda) = [p'(\lambda) + i(\lambda)]\ f(\lambda)$. Cabe lembrar, no entanto, que esta estrutura é conseqüência da estrutura imposta a $v(\lambda)$. Ao se restringir o número de marcas possíveis, o sinal marcado torna-se uma palavra código. Não há como utilizar a estrutura imposta a $x(\lambda)$ para obter informação sobre o sinal hospedeiro, nem tampouco sobre a localização das modificações feitas no processo de inserção. Observa-se que $f(\lambda)$ é independente de qual $x(\lambda)$ está sendo considerado. O custo desta operação é que ela implica numa restrição do número de marcas disponíveis. Resumindo, para tornar o sistema robusto contra modificações no sinal marcado, ambos os processos de inserção e detecção da marca precisam ser modificados. O processo de inserção é modificado apenas fazendo com que as marcas inseridas sejam todas múltiplas de $f(\lambda)$ A detecção é modificada para corrigir pequenas modificações em $x(\lambda)$ antes de tentar extrair a marca inserida. Primeiro é feita a correção de erros em $x(\lambda)$ utilizando-se $f(\lambda)$ como polinômio gerador, para depois se tentar detectar a marca.

Em seguida é mostrado um exemplo para clarificar as sugestões feitas. Segue-se o exemplo anterior de um sinal hospedeiro de comprimento 10 em $GF(11)$. Como antes, considera-se $\mathbf{s} = (\ 6, 3, 4, 0, 9, 1, 5, 7, 2, 5\ )$ e que a máxima distorção tolerável é de 4

dígitos. Modificando-se o sistema para que seja capaz de recuperar a marca mesmo após o atacante modificar uma componente do sinal marcado. É necessário utilizar 2 dígitos q-ários dos 4 disponíveis para a inserção da marca d'água, para se fazer correção de erros. Os dois dígitos restantes são utilizados para as marcas válidas e inválidas. Utiliza-se um único dígito q-ário para as marcas inválidas. Desta forma, como só resta um dígito para as marcas válidas, há 11 marcas válidas disponíveis. Convenciona-se que as marcas válidas então tem dois dígitos, com o primeiro igual a zero (01, 02, etc). Desta vez supõe-se que a marca a ser inserida é a de número 7. O polinômio gerador $g(\lambda)$ do código $C = [n, k, d]$ usado para inserir a marca através das matrizes **G** e **H** é dado por:

$$\begin{aligned} g(\lambda) &= (\lambda - \alpha)(\lambda - \alpha^2)(\lambda - \alpha^3)(\lambda - \alpha^4) \\ &= (\lambda - 2)(\lambda - 4)(\lambda - 8)(\lambda - 5) \\ &= \lambda^4 + 3\lambda^3 + 5\lambda^2 + 8\lambda + 1 \end{aligned} \tag{3.20}$$

Pode-se então fazer, por exemplo,

$$f(\lambda) = (\lambda - \alpha)(\lambda - \alpha^2) = (\lambda - 2)(\lambda - 4) = \lambda^2 + 5\lambda + 3 \tag{3.21}$$

Para inserir a marca 7 o codificador simplesmente multiplica a marca válida $(0,7) = 0x + 7$ por $f(\lambda)$:

$$w(\lambda) = 7(\lambda^2 + 5\lambda + 3) = 7\lambda^2 + 2\lambda + 10 \tag{3.22}$$

Na forma matricial, tem-se

$$\begin{bmatrix} 0 & 7 & 2 & 10 \end{bmatrix} = \begin{bmatrix} 0 & 7 \end{bmatrix} \begin{bmatrix} 1 & 5 & 3 & 0 \\ 0 & 1 & 5 & 3 \end{bmatrix} \tag{3.23}$$

Após o cálculo da marca adequada, a inserção é feita como anteriormente. Como a matriz **H** está na forma sistemática basta estender esta síndrome colocando zeros à esquerda[24] para obter o vetor $\mathbf{v} = (0,0,0,0,0,0,0,7,2,10)$. As posições de inserção são escolhidas aleatoriamente. Supondo que escolhem-se as mesmas posições mostradas anteriormente calculando **d** em $\mathbf{dG}' = \mathbf{s}' - \mathbf{v}'$ e substituindo em $\mathbf{x} = \mathbf{v} + \mathbf{dG}$ o sinal marcado obtido é:

[24] O que equivale a multiplicar por uma matriz estendida com zeros à esquerda.

x = | 6 | 8 | 4 | 7 | 9 | 5 | 5 | 7 | 2 | 1 |

(para **d** = (6, 8, 4, 7, 9, 5))

Neste ponto, supõe-se que o atacante tenha modificado a terceira posição deste sinal para:

Y = | 6 | 8 | **9** | 7 | 9 | 5 | 5 | 7 | 2 | 1 |

Se for feita a multiplicação por **H** do sinal atacado, não se obtém a marca inserida originalmente (0,7,2,10), mas sim (2,9,6,2). Observa-se que todos os dígitos da marca foram alterados. Mas o detetor faz primeiro a correção de erros como em um código Reed-Solomon com polinômio gerador $f(\lambda) = \lambda^2 + 5\lambda + 3$ e só depois retira a marca. Supõe-se que o método utiliza as implementações mais práticas e adequadas de decodificação de códigos Reed-Solomon. Deste modo, obtém-se novamente o sinal marcado **x** = (6,8,4,7,9,5,5,7,2,1) e portanto a mesma marca (0,7,2,10) é obtida através do mecanismo de decodificação. Dividir pelo polinômio $f(x)$ é equivalente, na forma matricial, a multiplicar pela pseudo-inversa da matriz usada na codificação. Seja $\mathbf{A}^+$ a pseudo-inversa da matriz **A**. Então

$$\begin{bmatrix} 1 & 5 & 3 & 0 \\ 0 & 1 & 5 & 3 \end{bmatrix}^{+} = \begin{bmatrix} 5 & 9 & 2 & 1 \\ 10 & 1 & 6 & 1 \end{bmatrix}^{t} \tag{3.24}$$

Portanto,

$$\begin{aligned} \mathbf{m} &= [síndrome] \begin{bmatrix} 5 & 9 & 2 & 1 \\ 10 & 1 & 6 & 1 \end{bmatrix}^{t} \\ &= \mathbf{y}\ \mathbf{H}^{t} \quad \begin{bmatrix} 5 & 9 & 2 & 1 \\ 10 & 1 & 6 & 1 \end{bmatrix}^{t} \\ &= \mathbf{y} \begin{bmatrix} 5 & 6 & 10 & 6 & 0 & 0 & 5 & 9 & 2 & 1 \\ 4 & 10 & 1 & 6 & 1 & 4 & 10 & 1 & 6 & 1 \end{bmatrix}^{t} \end{aligned} \tag{3.25}$$

Finalmente, supondo que **y** foi corrigido, segue-se que

$$\begin{aligned} \mathbf{m} &= \begin{bmatrix} 6 & 8 & 4 & 7 & 9 & 5 & 5 & 7 & 2 & 1 \end{bmatrix} \begin{bmatrix} 5 & 6 & 10 & 6 & 0 & 0 & 5 & 9 & 2 & 1 \\ 4 & 10 & 1 & 6 & 1 & 4 & 10 & 1 & 6 & 1 \end{bmatrix}^{t} \\ &= \begin{bmatrix} 0 & 7 \end{bmatrix} \end{aligned} \tag{3.26}$$

O exemplo acima mostra que através da correção de erros na marca é possível aumentar a robustez do sistema de marca d'água. Contudo, para facilitar o entendimento os parâmetros utilizados foram pequenos. Um sistema mais plausível pode ser construído através de um código de Reed–Solomon de maior comprimento. Um código muito usado é o código Reed-Solomon [255,223,33] em $GF(2^8)$ que tem 32 bytes de paridade. Ao invés de lidar-se com bits, cada um dos 256 símbolos é equivalente a um byte de informação. Supondo que o limite tolerável de distorção é de 32 bytes, sugerem-se utilizar 22 dígitos para a correção de erros, 5 para marcas inválidas e 5 para marcas válidas. Desta forma é possível inserir um total de 2^{40} marcas distintas e corrigir 11 erros.

3.4 Análise de Desempenho

Como observado anteriormente é necessário que a marca d'água digital seja idônea. Naturalmente, uma marca idônea não é tendenciosa. Uma marca idônea não tende a beneficiar nenhuma das partes envolvidas em uma disputa. Isto leva a três requisitos, quais sejam:

1. É desejável que seja pouco provável que acidentalmente se possa detectar uma marca inexistente, *i.e.* a probabilidade de ocorrerem alarmes falsos deve ser baixa.
2. É também desejável que a marca seja detectada com alta probabilidade quando presente.
3. Deve ser muito difícil trocar uma marca por outra.

O sistema proposto apresenta melhor desempenho para códigos de bloco mais compridos (maior n). Para mostrar o desempenho do sistema proposto é considerado o caso com o código de Reed-Solomon com parâmetros [255,223,33] mencionado no fim da última seção. Para se utilizar este código supõe-se que a máxima distorção tolerável é de 32 componentes do sinal hospedeiro. Novamente, utiliza-se estes 32 dígitos q-ários (bytes) disponíveis, da seguinte forma:

- 22 dígitos utilizados para fazer correção de erros
- 5 dígitos q-ários utilizados para marcas inválidas
- 5 dígitos q-ários utilizados para marcas válidas

A probabilidade de falso alarme pode ser calculada supondo que os valores de cada componente do sinal hospedeiro tem distribuição uniforme sobre o alfabeto de $2^8 = 256$ símbolos. Supondo, como anteriormente, que todas as marcas válidas tem os seus 5 primeiros dígitos fixos em um valor específico (e.g. 0,0,0,0,0), então a probabilidade de uma marca d'água ser detectada em um sinal não marcado é igual à probabilidade de ocorrência desta seqüência. Calcula-se a seguir a distribuição de probabilidade das componentes da síndrome. Cada componente da síndrome é obtida através da multiplicação de uma linha da matriz de paridade pelo vetor sinal marcado. Cada componente do sinal marcado é independente das demais e tem uma função de massa de probabilidade uniforme. Ao se multiplicar cada componente do vetor por um valor não nulo na matriz de paridade ocorre somente uma permutação na função de massa de probabilidade. Por outro lado, se a entrada na matriz for nula então a função de massa de probabilidade atribui probabilidade 1 ao zero e é nula em todos os outros valores. Após a multiplicação, como os valores são independentes, a soma representa a convolução circular[25] das funções de massa de probabilidade. Desta operação novamente resulta uma distribuição uniforme. Assim a probabilidade de falso alarme é simplesmente dada por:

$$Pr(\textit{falso alarme}) = (1/256)^5 = 2^{-40} = 1.09 \times 10^{-12}$$

Bloom *et al.* [3] determinaram que para um sistema de marca d'água de vídeo a probabilidade de falso alarme deve ser da ordem de 10^{-12} alarmes por frame. Observa-se que este sistema esta próximo da ordem de grandeza necessária.

Considera-se agora o segundo requisito. Para um sinal marcado que não foi alterado, o sistema proposto sempre detecta a presença da marca d'água inserida sem erros. Em sinais alterados existe uma margem de tolerância até onde o sistema ainda detecta a presença da marca. Para estes casos é interessante determinar o que um atacante é capaz de fazer.

Para remover uma marca um atacante pode tentar fazer pequenas modificações no sinal marcado, caracterizando um ataque com pequena distorção. Como mencionado, para repelir tais ataques é interessante distorcer o sinal marcado até o limite tolerável e também tornar difícil que se façam pequenas alterações no sinal. Considera-se uma distribuição uniforme para as componentes do sinal hospedeiro. No sistema proposto acima é muito

[25] A convolução é circular por que a soma é módulo q.

provável que todas as posições escolhidas para a inserção da marca sejam alteradas. A probabilidade de que uma componente de uma posição escolhida para a inserção da marca permaneça inalterada é de apenas 1/256. Logo, tem-se uma distribuição binomial no número de posições alteradas n_a com parâmetros $p = 255/256$ e $n = 32$. Especificamente,

$$\Pr(n_a = k) = \binom{32}{k} 0{,}9961^k (1 - 0{,}9961)^{32-k} . \tag{3.27}$$

A probabilidade que todas as posições escolhidas sejam alteradas é de $(255/256)^{32} \approx 0{,}88$. A probabilidade que todas posições menos uma sejam alteradas é 0,11 e a probabilidade de que todas posições menos duas sejam alteradas é 0,0067. A ocorrência de todas as outras possibilidades tem probabilidade inferior a 3×10^{-4} . Nota-se que o sinal marcado está de uma forma geral muito próximo da fronteira da região de distorção aceitável. Além disto, devido à correção de erros (que no caso citado corrige até 11 erros) o atacante precisa fazer grandes alterações no sinal marcado. Caso contrário, a marca não é removida.

Uma outra linha de ataque ao sistema é tentar remover a marca d'água a partir da observação das distorções causadas pelo processo de inserção. Ao inserir a marca d'água, modificando o sinal hospedeiro, o sistema causa um certo nível de degradação do sinal. Pode ser que cada tipo de marca inserida cause uma degradação diferente. Se assim for, observando a degradação causada, é possível obter informações para retirar a marca. Este tipo de ataque pressupõe que o atacante tem uma "noção" de qual deve ser o sinal original para que a degradação causada pelo sistema possa ser percebida. Qualquer sistema de marca d'água que cause distorções está sujeito a este tipo de ataque. Como sistemas robustos devem necessariamente degradar o sinal para evitar ataques com pequenas variações, os mesmos são sempre sujeitos a estes ataques. Para minimizar a probabilidade de sucesso de um ataque que utilize esta tática é interessante que seja muito difícil distinguir entre distorções causadas por diferentes modificações. Em outras palavras, deve ser difícil saber qual modificação causou qual distorção. Em particular, deve ser difícil descobrir a posição das componentes modificadas. Um tipo especial deste ataque é obtido quando provadores humanos são utilizados para comparar a qualidade do sinal marcado com a de um sinal que teve somente uma componente modificada sucessivas vezes. Se o sinal ligeiramente modificado for preferido pelos provadores, isto pode indicar que aquela componente foi modificada no processo de inserção da marca. Contudo, não há garantias de

que isto seja verdade. Como várias componentes foram modificadas, pode se tratar de um "máximo local". Por exemplo, se o valor original da componente modificada na primeira comparação era de 51 no sinal hospedeiro e é de 62 no sinal marcado (ou seja, a componente foi realmente modificada no processo de inserção). Pode ocorrer que após uma série de comparações o valor preferido pelos provadores para esta componente seja 23. Não somente esta componente, mas também outras foram modificadas no processo de inserção e portanto o valor 51 poderá não ser o valor obtido. Pode ocorrer também o caso em que os provadores prefiram um sinal em que uma componente que não foi modificada tenha outro valor que não o original. Observa-se também que este tipo de ataque pode ser combatido com a escolha adequada dos parâmetros do método.

No processo proposto para a inserção da marca d'água a posição dos dígitos modificados é escolhida aleatoriamente. Com esta escolha e como todos os dígitos modificados são arbitrários, não há nenhum "padrão" no sinal marcado que forneça informação sobre quais foram as posições alteradas.

Uma última estratégia de ataque para retirar uma marca a ser considerada é a força bruta. Partindo do sinal marcado, no sistema com 22 dígitos para correção de erro, é necessário modificar um mínimo de 12 posições para remover uma marca d'água. Um atacante pode escolher aleatoriamente 12 posições e modificá-las de modo a retirar a marca. Na escolha de cada uma das 12 posições a serem alteradas no ataque, o atacante tem uma chance de acertar justamente uma posição que também foi escolhida no processo de inserção da marca. Neste caso, sem informação adicional é pouco provável que o atacante consiga retirar a distorção inserida. Sempre que o atacante tiver sorte de selecionar uma posição já modificada, a contribuição para a distorção proveniente daquela componente do sinal pode permanecer constante. Se, por outro lado, o atacante selecionar uma posição que não foi modificada anteriormente, então o atacante causa um aumento na degradação do sinal. Felizmente, a probabilidade do atacante escolher justamente posições que foram modificadas no processo de inserção é baixa. Como somente trinta e duas dentre duzentas e cinqüenta e cinco posições foram modificadas a probabilidade de uma posição já modificada ser escolhida é de aproximadamente 1/8. A probabilidade de que todas as 12 posições modificadas no processo de inserção sejam alteradas pelo atacante é aproximadamente $(1/8)^{12} = 1{,}45 \times 10^{-11}$. Esta é a probabilidade de que o ataque retire a

marca sem aumentar a distorção do sinal. Como exemplo adicional de robustez do processo, considera-se a situação em que o atacante altera 12 posições, das quais 7 são posições já alteradas na inserção da marca. Desta forma, o incremento na distorção do sinal seria igual a 5 posições. A probabilidade desta combinação de eventos é $\binom{12}{7}(1/8)^7(7/8)^5 = 1{,}9\times10^{-4}$ aproximadamente. O mais provável é que o atacante consiga remover a marca aumentando a distorção em pelo menos 9 posições. Isto ocorre com probabilidade 0,94. Como se observa do exemplo onde 7 posições previamente alteradas são escolhidas, é muito improvável que um atacante consiga remover a marca adicionando pouca distorção ao sinal. Existe uma margem entre a máxima distorção tolerável e a distorção mínima necessária para um ataque bem sucedido. No caso acima, a distorção máxima permitida no processo de inserção é de 32 dígitos, mesmo se um sinal com distorção igual a ou menor que 37 (32+5) ainda for considerado aceitável por um atacante, um sinal sem marca com esta qualidade só será obtido uma vez a cada 4700.

O sistema que utiliza um código em $GF(256)$ para a inserção de marcas d'água tem a vantagem de trabalhar diretamente com bytes como unidade de informação. Contudo, caso os parâmetros deste sistema ainda não sejam adequados, utilizando-se a próxima unidade de informação usual, *i.e.* 16 bits, e trabalhando em $GF(2^{16})$ as exigências possivelmente serão satisfeitas. Isto leva a sistemas onde a marca é inserida em seqüências que tem um comprimento muito grande quando comparados aos sistema em $GF(256)$. Sistemas em $GF(2^{16})$ tem comprimento aproximado de $64.000 \times 2 = 128$ kbytes. Contudo, 128 kbytes constituem apenas 8 segundos de música no formato padrão MP3 com taxa de 128 kbits/s[26]. Tal quantidade de informação ainda pode ser considerada suficientemente pequena para muitas aplicações. Um ajuste de parâmetros do método proposto pode tornar inviáveis, ataques promissores.

O sistema proposto não pode ser usado exatamente na forma apresentada para a inserção de *impressões digitais* uma vez que a disponibilidade de dois sinais com marcas diferentes ou com diferentes posições alteradas na inserção das marcas torna as posições

[26] A taxa de 128 kbits/s é muito comum na Internet e a mais encontrada em serviços de troca de arquivos de música (como o Napster).

modificadas imediatamente visíveis. Uma alternativa para esta aplicação é sugerida no capítulo de conclusão.

Capítulo 4

Conclusões

Neste trabalho propõe-se um sistema para a inserção de marcas d'água robustas. O sistema é assimétrico e permite que agentes não-confiáveis leiam a marca, pois a leitura da marca não torna os mesmos agentes aptos a retirá-la. O sistema não usa qualquer tipo de chave criptográfica. A detecção das marcas é feita por um detetor padrão cujos detalhes e forma de operação podem ser de conhecimento público. Não há a necessidade de dispositivos à prova de ataques intencionais (*tamper-proof*). O procedimento de inserção da marca é probabilístico e pode ser feito por qualquer agente que conheça o sinal hospedeiro.

O sistema baseia-se no fato de que o sinal hospedeiro pode ser representado como uma seqüência de variáveis aleatórias discretas i.i.d. e considera a distância de Hamming como medida de distorção. A utilização desta medida de distorção implica que o sinal hospedeiro está representado em um espaço onde valem as seguintes condições:

a) O número de componentes modificadas define a distorção introduzida no sinal hospedeiro, e não a amplitude da variação introduzida em cada componente modificada.
b) Modificações em diferentes componentes causam níveis equivalentes de degradação no sinal hospedeiro. Não há uma componente mais significativa. Por exemplo, a modificação da primeira componente do sinal gera um nível de distorção equivalente ao da modificação da última componente.

Além disto, para a análise de desempenho do sistema é considerado que as componentes do sinal hospedeiro tem distribuição uniforme. Desta forma, é possível calcular as probabilidades de interesse e não é possível para um atacante determinar onde a marca d'água foi inserida a partir da inspeção do sinal marcado. O relaxamento desta última condição segue como uma sugestão para trabalhos futuros.

O sinal hospedeiro deve estar codificado de forma a permitir degradação somente até um limiar máximo aceitável, a partir do qual modificações adicionais causam degradações intoleráveis. Valendo as condições anteriores, o sistema é vulnerável a apenas

um ataque, como descrito a seguir. Embora diferentes modificações causem o mesmo *grau* de distorção, as mudanças ocorridas quando da modificação desta ou daquela componente não precisam ser as mesmas. Uma dada modificação pode distorcer o sinal de uma forma diferente das outras. O ataque em questão consiste em determinar onde a marca d'água foi inserida através da determinação do tipo de distorção presente no sinal marcado. Todos os sistemas de marca d'água robusta estão sujeitos a este ataque e portanto deve ser difícil determinar as modificações introduzidas no sinal hospedeiro a partir da "apreciação" do sinal marcado.

A capacidade de resistir a ataques intencionais só está presente em um único método assimétrico de inserção de marcas d'água, o QIM. Para os demais métodos não há nenhuma garantia de segurança[27]. A existência de métodos capazes de resistir a tais ataques é inclusive questionada dependendo das hipóteses consideradas. Pode-se citar por exemplo [25]: "A existência de uma marca d'água digital para a qual um dado ataque de distorção bem sucedido não possa ser encontrado é uma questão ainda aberta.", ou ainda [21], "Na realidade, Alice, no caso a proprietária do conteúdo, pretende vender informação para um cliente não-confiável, Bob, sem permitir que Bob passe adiante a informação. Aparentemente, não há uma solução criptográfica ou baseada em Teoria da Informação para este problema." Contudo, dentro do modelo utilizado o método desenvolvido satisfaz às exigências de robustez impostas. A probabilidade de falso alarme, *i.e.* a probabilidade do sistema achar uma marca onde não existe nenhuma, pode ser controlada dividindo-se apropriadamente o espaço de marcas d'água. A probabilidade de não-detecção de uma marca d'água em um sinal não atacado é zero. E a probabilidade de um atacante implementar um ataque bem sucedido, ou seja, um ataque que retira a marca mantendo a qualidade do sinal hospedeiro, é baixa. Em todo este trabalho foram sempre considerados ataques intencionais.

O sistema proposto utiliza codificação de Reed-Solomon em duas camadas. A primeira faz correção de erros na síndrome e a segunda espalha a informação da marca sobre todos os dígitos do sinal hospedeiro. A combinação de ambas etapas dá uma estrutura ao sinal marcado. Contudo esta estrutura não pode ser utilizada para a remoção da marca.

[27] Métodos muito bons foram encontrados para aplicações específicas, por exemplo, a inserção de ecos (*echo hiding*). Mas também não há para estes casos garantia de que um novo ataque não possa ser arquitetado.

Alterando-se o código de Reed-Solomon utilizado para a inserção da marca d'água podem ser conseguidos parâmetros de robustez e de taxa de transferência de informação adequados a diferentes aplicações. Códigos com comprimento maior tendem a oferecer melhor desempenho. Embora os códigos de Reed-Solomon tenham comprimentos restritos a potências de primos, *i.e.* $n = p^m - 1$, onde p é primo, existem códigos com o número de dígitos de informação variando de 1 ao comprimento do código. Portanto, diferentes níveis de distorção podem ser contemplados.

Para o desenvolvimento do sistema proposto, foram consideradas algumas diretivas baseadas nas características dos sistemas de marca d'água desejados. Uma consideração importante a fazer em todo sistema de marca d'água robusta é supor que o usuário do sinal não é confiável. Neste trabalho observa-se também que a marca d'água robusta deve se propagar junto com o sinal.

Uma contribuição do trabalho são as características que devem ser satisfeitas por sistemas assimétricos de marcas d'água robustas a ataques intencionais:

- É interessante degradar o sinal marcado tanto quanto for tolerável.
- O sinal marcado deve ser a melhor estimativa do sinal hospedeiro que o usuário pode obter.

A primeira consideração deve ser feita para que o sistema seja robusto contra ataques que façam pequenas modificações no sinal marcado. A segunda decorre da necessidade do usuário não poder conhecer o sinal hospedeiro.

Por fim, observa-se que os sistemas probabilísticos parecem ser os melhores candidatos para sistemas sem chave, robustos a ataques intencionais.

O trabalho apresentado abre novas possibilidades e requer o desenvolvimento de técnicas que tornem possível a implementação prática da ferramenta proposta. Como sugestões para trabalhos futuros, considera-se a necessidade de adaptar sinais hospedeiros em geral (vídeos, músicas, etc) a um espaço adequado para a utilização do sistema proposto. A utilização de outras medidas de distorção, por exemplo, a distância de Lee, podem se constituir em objeto de métodos alternativos.

Como já observado, o método proposto não pode ser utilizado na forma apresentada para a inserção de impressões digitais. Esta limitação pode ser contornada se a inserção da marca não for feita isoladamente em um único bloco de dados. O sinal hospedeiro pode ser

dividido em vários blocos, nos quais marcas diferentes são inseridas, com isto, o sistema pode oferecer uma base segura para a implementação de sistemas de "impressão digital" utilizando a técnica proposta por Boneh e Shaw [2]. Para isto é necessário que mais de uma marca esteja disponível para cada autor.

Uma possível extensão do presente trabalho é o desenvolvimento de técnicas capazes de codificar sinais hospedeiros encontrados no mundo real em uma representação que apresente um limiar crítico de distorção aceitável. O sistema de inserção de marcas d'água aqui descrito distorceria o sinal, durante o processo de inserção da marca, até este limiar aceitável. A partir deste ponto, modificações adicionais produziriam quedas abruptas de qualidade de modo a provocar distorções intoleráveis no sinal. Esta adaptação de sinais é necessária para a implementação prática do método de inserção de marcas d'água aqui apresentado.

Bibliografia

[1] R. J. Anderson e F. A. P. Petitcolas, “On The Limits of Steganography”, *IEEE Journal on Selected Areas in Communications*, vol 16 n 4, pp 474-481, maio 1998.

[2] D. Boneh e J. Shaw, “Collusion-Secure Fingerprinting for Digital Data”, *Advances in Cryptology - CRYPTO 95*, pp 452--465, 1995.

[3] J. A. Bloom, I. J. Cox, T. Kalker, J. M. G. Linnartz, L. Miller e C. B. S. Traw, “Copy Protection for DVD Video”, *Proceedings of the IEEE*, vol 87 n 7, pp 1267-1276, julho 1999.

[4] B. Chen, “Design and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems”, tese de doutorado MIT, junho 2000.

[5] J. Chou, S. Sandeep Pradhan e K. Ramchandran, “On the duality between distributed source coding and data hiding”, *33rd Asilomar conference on Signals, System and Computers*, pp. 1503-1507, 1999.

[6] J. Chou, S. Sandeep Pradhan, L. El Ghaoui e K. Ramchandran, “A robust optimization solution to the data hiding problem using distributed source coding principles”, *Proc. SPIE conference on Electronic Imaging: Image and Video Communications and Processing*, janeiro 2000.

[7] M. H. M. Costa, “Writing on Dirty Paper”, *IEEE Transactions on Information Theory*, vol 29 n 3, pp 439-441, maio 1983.

[8] I. J. Cox, J. Kilian, F. T. Leighton e T. Shamoon, “Secure Spread Spectrum Watermarking for Multimedia”, *IEEE Transactions on Image Processing*, vol 6 n 12, PP 1673-1686, dezembro 1997.

[9] J. J. Eggers, J. K. Su e B. Girod, “Asymmetric Watermarking Schemes”, *Tagungsband des GI Workshops "Sicherheit in Mediendaten,"*, Berlin, Alemanha, setembro 2000.

[10] J. J. Eggers, J. K. Su e B. Girod, “A Blind Watermarking Scheme Based on Structured Codebooks,” *IEE Colloquium: Secure Images and Image Authentication*, Londres, Inglaterra, abril 2000.

[11] J. J. Eggers, J. K. Su e B. Girod, “Performance of a Practical Blind Watermarking Scheme”, *Electronic Imaging 2001*, San Jose, CA, USA, janeiro 2001.

[12] J. J. Eggers, J. K. Su e B. Girod, “Robustness of a Blind Image Watermarking Scheme”, *Proceedings of IEEE International Conference on Image Processing (ICIP 2000)*, Vancouver, Canada, setembro 2000.

[13] J. J. Eggers, J. K. Su e B. Girod, "Public Key Watermarking By Eigenvectors of Linear Transforms," *European Signal Processing Conference (EUSIPCO2000)*, Tampere, Finlandia, setembro 2000.

[14] T. Furon e P. Duhamel, “An Asymmetric Public Detection Watermarking Technique” *Workshop on Information Hiding*, Dresden, Alemanha, outubro 1999.

[15] T. Furon e P. Duhamel, "Robustness of Asymmetric Watermarking Technique", *IEEE International Conference on Image Processing (ICIP)*, Vancouver, Canada, setembro 2000.

[16] F. Hartung e B. Girod, "Fast public-key watermarking of compressed video", *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Santa Barbara EUA, outubro 1997.

[17] C. Heegard, "Partitioned Linear Block Codes for Computer Memory with "Stuck-at" Defects", *IEEE Trans. Information Theory* vol 29 n 6, pp 831-842, novembro 1983.

[18] C. Heegard e A. El Gammal, "On the Capacity of Computer Memory with Defects", *IEEE Trans. Information Theory* vol 29 n 5, pp 731-739, setembro 1983.

[19] M. E. Hellman, "An Extension of the Shannon Theory Approach to Cryptography", *IEEE Trans. Information Theory* vol 23 n 3, pp 289-294, maio 1977.

[20] L. R. Matheson, S. G. Mitchell, T. G. Shamoon, R. E. Tarjan, F. Zane, "Robustness and Security of Digital Watermarks", *Proceedings of Financial Cryptography*, Anguilla, BWI, pp 23-25, fevereiro 1998.

[21] M. Maes, T. Kalker, J-P. M. G. Linnartz, J. Talstra, G. F. G. Depovere e J. Haitsma, "Digital Watermarking for DVD Copy Protection", *IEEE Signal Processing Magazine*, pp 47-57, setembro 2000.

[22] F. J. McWiliams e N. J. A. Sloane, "The Theory of Error-Correcting Codes", North-Holland 1977.

[23] P. Moulin e J. A. O'Sullivan, "Information–Theoretic Analysis of Information Hiding" (*preprint*), outubro 1999.

[24] P. Moulin e J. A. O'Sullivan, "Information–Theoretic Analysis of Watermarking", *Proceedings ICASSP'00* , Istanbul, Turquia, junho 2000.

[25] F. A. P. Petitcolas, R. J. Anderson e M. G. Kuhn, "Information Hiding – A Survey", *Proceedings of the IEEE*, vol 87 n 7, pp 1062-1078, julho 1999.

[26] S. Sandeep Pradhan e K. Ramchandran, "Distributed source coding using syndromes (DISCUS) : Design and construction", *Proc. IEEE Data Compression Conference*, Snowbird EUA, março 1999.

[27] G. J. Simmons, "The prisoners´ problem and the subliminal channel", *Proceedings of CRYPTO'83*, pp 51-67, 1983.

[28] R. G. van Schyndel, A. Z. Tirkel e I. D. Svalbe, "Key independent watermark detection", *Proceedings of the IEEE International Conference on Multimedia Computing and Systems (ICMCS'99)*, vol 1, Florença, Itália, junho 1999.