6
Esquema de Polarização do Detetor para Técnicas Frágeis, Semi-frágeis e Híbridas

No capítulo anterior foi verificado, na ótica da metodologia proposta, que o código turbo pode ser empregado diretamente para projeto de sistemas de marcação robustos. Neste capítulo, é apresentado um esquema de polarização do detetor, em uma implementação com codificação turbo, que permite a operação em modo frágil, semi-frágil ou robusto, marcando o hospedeiro com apenas uma marca d'água. Uma simulação computacional é implementada objetivando a obtenção de resultados que respaldem a metodologia e o esquema propostos.

6.1
Polarização do Detetor em Técnicas de Marcação D'Água Digital

Observando novamente o desempenho do código turbo típico, exemplificado na Figura 6.1 (repetição da figura 5.2), evidenciamos que a curva de desempenho possui duas regiões: uma região de alta derivada (região com condição de fragilidade), para $WNR_N < 3.75dB$; e uma região de baixa derivada (região com condição de robutez), para $WNR_N > 3.75dB$. As duas regiões são bem delimitadas por um "joelho", correspondendo a uma razão marca - ruído normalizada que denominamos WNR_N^p, que no exemplo é $3.75dB$. Na ausência de ataque (ruído), $WNR_N \rightarrow \infty$. À medida que o ataque é incrementado, o valor WNR_N é reduzido, e o sistema transita de uma região de baixas derivadas para outra de altas derivada na curva de desempenho. Objetivando a operação em modo frágil, propõe-se adicionar um ruído bem determinado na deteção, de modo a trazer diretamente a operação do esquema, no receptor, à região com condição de fragilidade, para $WNR_N < WNR_N^p$. Assim, é bastante apropriado referir-se a tal ruído determinado de *ruído de polarização*, e à WNR_N^p, como *razão de polarização marca d'água - ruído* . Assim, o sistema pode ser polarizado através da simples adição ao sinal $\mathbf{y}$, na entrada do receptor, de um ruído de polarização

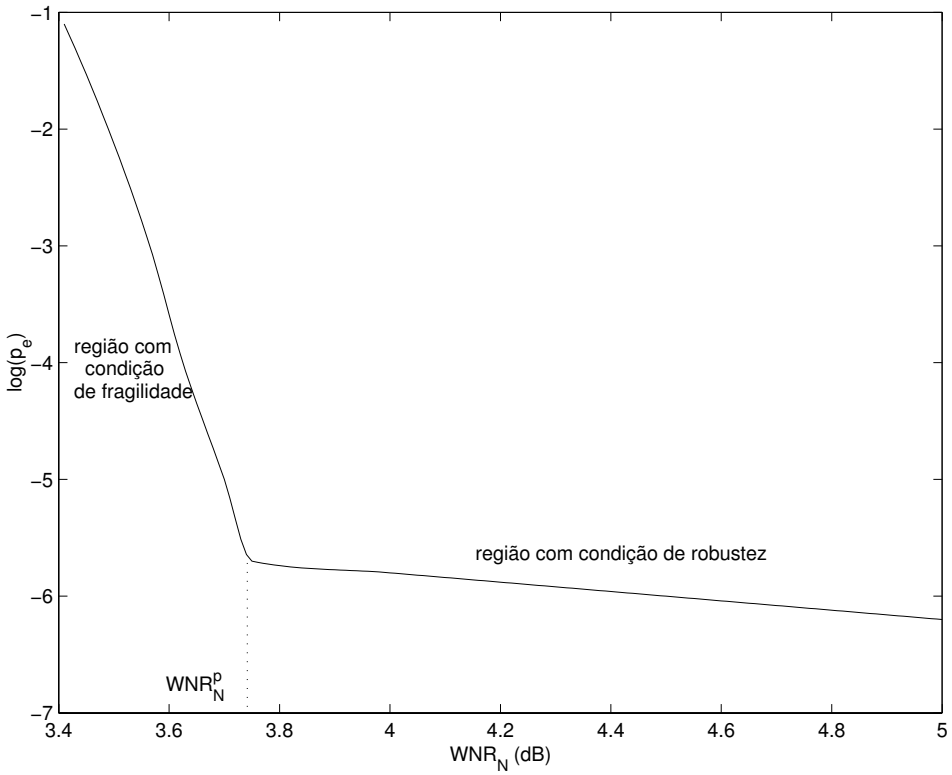

Figura 6.1: Desempenho de um codificador turbo típico com taxa 1/2.

gaussiano, com média zero e variança $\sigma^2_{n_p}$, com potência

$$\sigma^2_{n_p} = \frac{\sigma^2_{wm}}{WNR^p_N} = \frac{\rho D^M}{WNR^p_N}. \tag{6-1}$$

Assim, mediante a introdução do ruído de polarização no receptor, a deteção da marca d'água será processada diretamente na região com condição de fragilidade, i.e., tal que qualquer incremento no ruído de ataque ocorrerá severa degradação em p_e.

É bastante interessante notar que apesar de apenas uma marca d'água modular o hospedeiro, o esquema, na recepção, pode ser ajustado de forma a selecionar a operaração no modo fágil ou no robusto. A operação no modo robusto é alcançada diretamente, conforme discutido no capítulo anterior, sem adição de qualquer ruído de polarização. Entretanto, se o objetivo é a operação no modo frágil, basta adicionar o ruído de polarização no receptor, conforme recém proposto. A figura 6.2 ilustra esta idéia de seleção do modo de operação. Por exemplo, para verificações de direitos autorais

em relação ao hospedeiro, o modo robusto de deteção é selecionado, enquanto para verificação de autenticidade do hospedeiro, o modo polarizado (frágil) é selecionado. Este tipo de técnica é denominado na literatura [18] como técnica híbrida. Vale observar que o esquema da figura 6.2 pode ser facilmente redesenhado, com deteção paralela (com e sem ruído de polarização), de forma a possibilitar o processamento simultaneo em modo frágil e robusto.

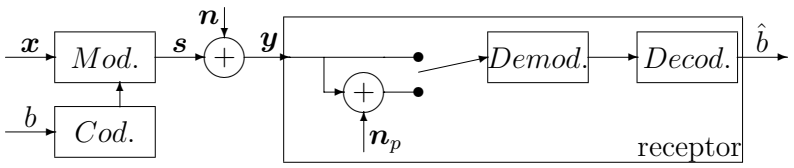

Figura 6.2: Esquema de marcação d'água digital, com polarização do detetor, com opção de seleção de operação robusta ou frágil.

É oportuno comentar que em outras aplicações, a proposta polarização pode ser efetuada no modulador, no lado da transmissão. Por exemplo, se a "marca d'água" a ser utilizada é uma informação sensível (*covert communication*), a polarização no modulador garantirá que esta informação sensível, que é a própria marca d'água, seja perdida ("auto-destruída") se for sujeita a qualquer ameaça (ataque) externa. De um modo geral, para qualquer aplicação em que a distorção do hospedeiro não seja um requisito restritivo, durante sua distribuição/transmissão, permite que a polarização seja conduzida na modulação. Contudo, neste caso, cabe evidenciar que, se a polarização efetuada no transmissor não for reversível no receptor, somente a deteção no modo frágil será possível.

6.2 Implementações de Técnicas de Marcação D'Água Codificadas Semi-Frágeis

O objetivo desta seção é a investigação de como mensurar a intensidade do ataque aplicado ao hospedeiro. Uma técnica que permite conduzir tal mensuração, aproveitando o recém proposto esquema de polarização, vai ser discutida a seguir.

Tal tipo de técnica, que é interessante empregar em aplicações de verificação de autenticidade e integridade do hospedeiro, é classificada na

literatura [18] como um tipo de técnica de deteção de adulteração *semi-frágil*, e também pode ser entendida como uma técnica de avaliação "soft".

Descrevemos a seguir o esquema da técnica proposta. Considerando o código turbo típico da Figura 6.1, vamos supor que o sistema está polarizado no joelho da curva de desempenho, representado pelo ponto ($WNR_N^p = 3.75$, $p_e^p = 10^{-5.5}$). Caso nenhum ataque seja aplicado, a medição da taxa de erros fornecerá um valor BER^p, aproximadamente igual a p_e^p. Assim na análise que segue vamos empregar o valor p_e^p, em vez de BER^p, eliminando a necessidade desta medição.

Por outro lado, quando um desconhecido ataque **n** faz-se presente, ocorre o deslocamento do ponto de operação na curva de desempenho de (WNR_N^p, p_e^p) para (WNR_N^*, p_e^*). O valor p_e^* pode ser estimado através da medição do correspondente BER^*. A medição do BER pode ser realizada de duas formas. Uma forma, é disponibilizar o bloco b de bits da marca d'água digital original para comparação com a estimação do bloco $\hat{b}$ resultante da deteção no modo frágil. Alternativamente, podemos também utilizar a estimação $\hat{b}$ resultante da deteção no modo robusto para fins de comparação com a estimação anteriormente efetuada com o processamento no modo frágil. Esta alternativa torna o esquema mais prático e seguro, pois a marca d'água original não é mais necessária no receptor.

Agora estamos em condições de propor que uma avaliação 'soft' da intensidade do ataque seja definida como $WNR_N^p - WNR_N^*$. Esta variação da razão marca-ruído pode ser estimada através da medição da variação $BER^* - p_e^p$ observada. Notar que esta estimativa só faz sentido em um esquema com polarização, uma vez que quando não há polarização, $WNR_N^p \rightarrow \infty$. A idéia desta proposta é ilustrada na figura 6.3. Deve ser observado neste exemplo que, um incremento do ataque menor do que $0.2dB$, correspondendo a variação de $WNR_N^p = 3.75$ para $WNR_N^* = 3.55$, implica em uma redução do BER por um fator maior do que 300.

A polarização do esquema, forçando a excursão da BER na região com condição de fragilidade, onde ataques de pequena intensidade provocam uma grande degradação na BER, aumenta significativamente a precisão em que a intensidade do ataque pode ser estimada. O mesmo esquema poderia ser polarizado na região com condição de robustez, mas a precisão da estimativa ficaria bastante reduzida. Algumas aplicações podem requerer

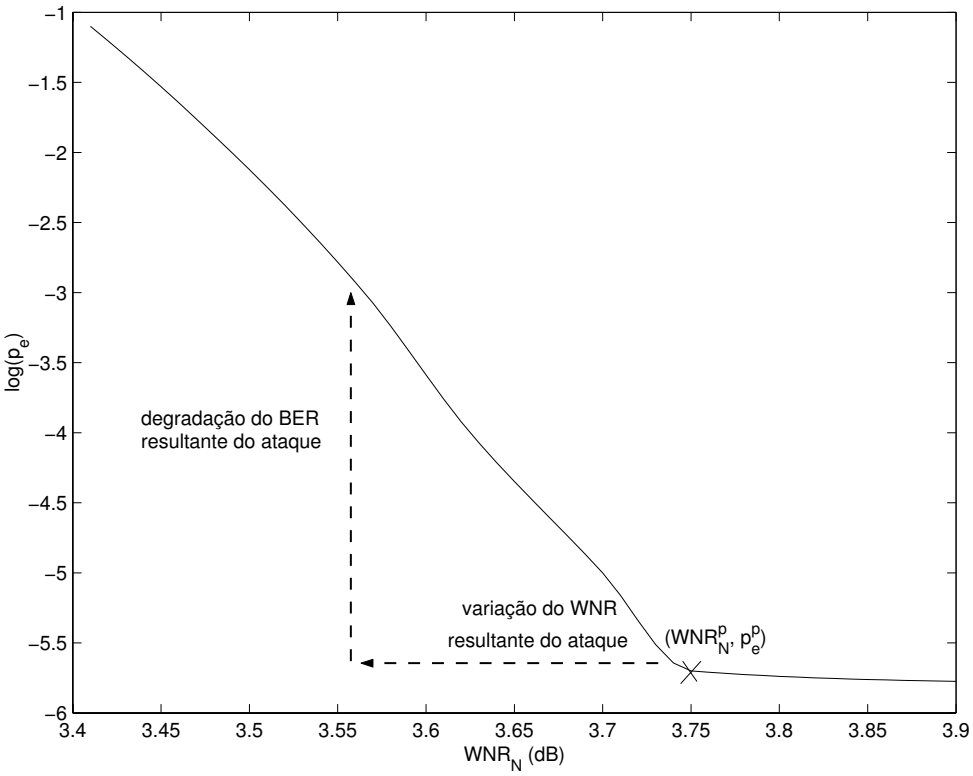

Figura 6.3: Variação da taxa de erro resultante de ataque para um esquema polarizado codifocado com código turbo.

uma fragilidade com sensitividade reduzida, i.e., que a inclinação da curva de desempenho não seja tão alta. Esta redução da sensitividade permite ampliar a faixa dinâmica de estimação do ataque. Em particular para as implementações com codificadores turbo, a referida redução de sensitividade, da região com condição de fragilidade, pode ser simplesmente obtida diminuindo o número de iterações do algoritmo de decodificação (veja o apêndice A).

Uma importante observação é que a viabilidade e praticidade do equema proposto são obtidas quando, por meio da polarização, a excursão da BER fica contida em um domínio onde sua medição é viável. No exemplo considerado para a codifacação turbo, $BER^* \in [10^{-5.5}, 10^{-1}]$)). Assim, não basta polarizar o esquema em uma região com condição de fragilidade, mas o correspondente domínio de excursão da BER nesta região deve possibilitar sua medição em termos práticos. Desta forma, podemos concluir que a possibilide da implementação do esquema proposto para técnicas semi-frágeis, resulta não só da polarização do detetor, mas também da

adequada caracterísitca da curva de desempenho do código turbo.

Finalizando esta seção, a discussão proposta de deteção semi-frágil não considerou a possibilidade de localizar o ataque no hospedeiro, e nem a possibilidade em distinguir ataques maliciosos (intencionais) de ataques inocentes, como, por exemplo, uma compressão de imagem. Entretanto, a flexibilidade do esquema proposto indica a potencialidade do mesmo para tais aplicações, e a investigação futura, neste sentido, é de interesse. Desde já, podemos registrar que a possibilidade em utilizar o esquema proposto para conduzir o processamento simultâneo, em modo robusto e frágil, permite efetuar-se a comparação das marcas d'água estimadas na saída de cada tipo de processamento (robusto e frágil), evidenciando a facilidade de empregar o esquema em aplicações de localização do ataque.

6.3 Resultados de Simulações Experimentais

Uma simulação computacional foi implementada objetivando a obtenção de resultados experimentais que validem o esquema de polarização proposto, e por conseqüência a própria metodologia de modelamento da fragilidade e da robustez introduzida no Capítulo 5. Nesta simulação, implementou-se uma técnica da modulação/demodulação tradicional, em uma arquitetura codificada com o mesmo código turbo (TC) utilizado na obtenção do desempenho representado na figura 6.1, com taxa 1/2.

O sinal hospedeiro utilizado na simulação foi uma imagem Lena (256 x 256 pixels) em escala de cinza. A marca d'água é um bloco de 8196 bits. Cada bit da marca d'água modula (SS) um bloco de 16 pixels ($N = 16$) da imagem hospedeira para produzir a imagem marcada. Em seguida, um ataque de ruído gaussiano é adicionado à imagem marcada, e finalmente, na deteção, um ruído gaussiano de polarização é introduzido, antes de processar a demodulação SS e a decodificação turbo. Vale lembrar que o comprimento da marca d'água deve ser grande o suficiente para que as medições do BER não possuam significativos desvios do valor estatísitco p_e.

Nosso maior interesse na simulação é observar o comportamento da fragilidade do esquema com polarização. Assim, um ruído de polarização, tal que $WNR_N^p = 3.3\ dB$ é adicionado no receptor, antes do processamento

da deteção, correspondendo a uma polarização próxima ao joelho da curva de desempenho. Desta forma, a demodulação e a decodificação é conduzida considerando a adição do ataque (ruído), correspondendo ao deslocamento da operação do esquema, na curva de desempenho, de $WNR_N^p = 3.3\ dB$ para um novo valor $WNR_N < WNR_N^p$. Após processada a deteção e a correspondente estimação da marca d'água, o *número de bits errados* (NEB) é determinado. A fim de obter uma estimativa da média e do desvio padrão de NEB, para cada intensidade de ataque considerado (correspondendo a valores distintos de WNR_N), a simulação implementada foi executada 50 (cinquenta) vezes, considerando 50 ruídos gaussianos de ataque $\mathbf{n}$, gerados aleatoriamente, para cada intensidade de ataque considerado (correspondendo a um dado valor de WNR_N). Os resultados obtidos são consolidados e apresentados na Figura 6.4, onde as curvas são parametrizadas em função do número de iterações ($iter$) do algoritmo TC de decodificação. Nesta figura, tendo em vista que nosso interesse é observar o comportamento do NEB em função do ataque, o ruído de polarização não está adicionado ao ruído de ataque para fins de cálculo da razão marca-ruído (WNR_N).

Antes de prosseguir com a análise dos resultados, vamos evidenciar que, para todos os ataques considerados em nossa simulação, se a polarização não for introduzida, a decodificação corrige todos os eventuais bits errados ($NEB = 0$). Isto é muito importante para permitir uma implementação segura do esquema de polarização, i.e., o detetor não necessita do conhecimento a priori da marca d'água original para determinar o NEB. Assim, para determinar o NEB, a simulação primeiramente deve ser executada sem considerar qualquer polarização, a fim de obter uma ótima estimação dos bits da marca d'água através deteção no modo robusto.

Primeiramente, deve ser observado na figura que, com o esquema de polarização, ataques de intensidades extremamente baixas podem ser detetados, simplesmente pela verificação de erros na recepção. Em nossa simulação, ataques correspondendo a $WNR_N = 97dB$, ou a $Eb/N_0 = 94dB$, podem ser identificados. Assim, conforme nossa expectativa, o esquema já revela sua propriedade de elevada sensitividade e fragilidade, permitindo na prática a deteção de quase qualquer tentativa de adulteração significativa do sinal hospedeiro.

Outra conclusão importante da análise dos resultados é a possibilidade de classificar a intensidade do ataque, em função do NEB medido,

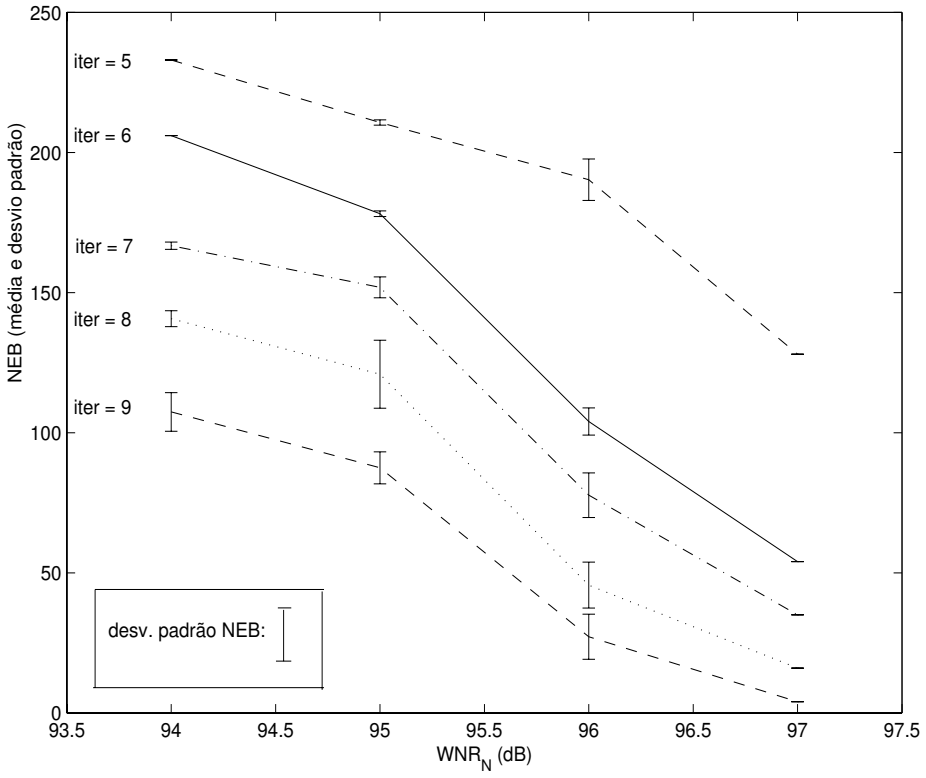

Figura 6.4: Média e desvio padrão de NEB para valores de WNR_N que correspondem ataques distintos.

confirmando a adequação do esquema de polarização para aplicações que requerem sistemas semi-frágeis. Como exemplo, se a deteção é conduzida com $iter = 7$, e como resultado é medido um $NEB = 72$, então no contexto das curvas obtidas, e correspondentes variançcas, podemos classificar ou estimar o ataque como tendo resultado uma razão marca-ruído em uma vizinhança de $WNR_N = 96dB$. Observar que, para um dado número de iterações ($iter$), e para os ataques considerados em nossa simulação, embora exista uma variança para cada intesidade de ataque, o maior (menor) NEB medido, para uma dada intesidade de ataque, foi sempre menor (maior) do que todos os $NEBs$ medidos quando aplicado uma intensidade de ataque superior (inferior), considerando todos os 50 ataques aleatórios para cada intesidade de ataque. Este resultado demonstra a precisão e confiabilidade do esquema proposto para classificação do ataque.

Podemos também extrair dos resultados apresentados na figura 6.4 alguma evidência do comportamento da variança (ou do desvio padrão) em função de $iter$ e WNR_N. Por exemplo, para baixos ataques o desvio é pequeno. Em nossa simulação, para $WNR_N = 97dB$, o desvio foi nulo para

todos os valores do parâmetro *iter* considerados. Podemos explicar tal comportamento observando que neste caso o especificado ruído de polarização é muito superior aos ataques aleatórios, e assim o mesmo é dominante no processo de decodificação, reduzindo (até eliminando) a influência do ataque no mesmo. Ainda, aparentemente, para uma dada intensidade de ataque, o desvio de NEB inicialmente aumenta com *iter*, mas à medida que o algoritmo de decodificação começa a convergir, para maiores valores de *iter*, este desvio começa a decrescer. Este comportamento pode ser explicado argumentando que à medida que o parâmetro *iter* é elevado, o algoritmo de decodificação começa a limpar os erros provocados pelos ataques aleatórios, e quando estes erros são todos corrigidos, os erros provocados pelo ruído de polarização tornam-se dominantes, fazendo o desvio de NEB diminuir. O comportamento de NEB em função de WNR_N pode ser interpretado com argumentos similares, e uma investigação mais aprofundada destes comportamentos é de interesse.

É interessante observar que a estimativa da intensidade do ataque, efetuada utilizando um determinado valor do parâmetro *iter*, pode ser checada, ou confirmada, observando os NEBs resultantes quando empregado outros valores deste parâmetro. No exemplo de nossa simulação, onde estimou-se o ataque utilizando $iter = 7$, pode-se confirmar a estimativa desse ataque verificando o NEB para o valor $iter = 5$. Neste caso, em função dos resultados da simulação efetuada, devemos esperar um valor $NEB = 190 \pm 8$. De uma maneira mais genérica, pode-se conceber um método para estimação do ataque considerando um tipo de correlação entre os NEBs medidos, para diversos valores de *iter*, e o padrão de valores previstos conforme os resultados da simulação. Entedemos que este tipo de método pode aumentar a confiabilidade e precisão da estimativa, e sua investigação é de interesse.

Como qualquer dispositivo de medida, o esquema proposto possui uma faixa dinâmica para estimativa do ataque. Uma forma de ampliar esta faixa, e possibilitar a estimativa de valores maiores da intensidade do ataque, é adaptar o proposto esquema de polarização. Podemos ampliar a faixa dinâmica introduzindo um valor diferente (reduzido) para o ruído de polarização, tal que WNR_N^p fique distante (à direita) do joelho da curva de desempenho. Se o ruído de polarização é tal que, na presença do ataque, WNR_N assuma novamente valores na vizinhança do joelho da curva de desempenho, então o esquema irá operar novamente na região ótima de

alta sensitividade. Assim, nesta nova polarização do esquema, o valor do ataque que traga novamente o valor de WNR_N para o joelho da curva de desempenho é conhecido à priori. Nesta situação, se o NEB medido for desprezível, podemos classificar o ataque como inferior àquele conhecido a priori. Quando uma medição significativa para NEB for obtida, podemos classificar o ataque como superior àquele conhecido a priori, e ainda estimar o mesmo, considerando agora que o sistema opera na mesma situação do esquema original. Assim, concluimos que este artifício de sintonia de polarização permite ampliar a faixa dinâmica de identificação e estimação do ataque. Na prática, podemos realizar uma primeira estimativa, de um desconhecido ataque, somente através da sintonia do ruído de polarização. Quando o esquema, na presença do ataque, estiver sintonizado no joelho da curva (correspondendo ao início da medição significativa de erros - NEB), o ataque pode ser estimado considerando que a sintonia no joelho da curva corresponde à composição do conhecido ruído de sintonia adicionado ao ataque. Como o ponto do joelho da curva de desempenho é bem conhecido, no caso do código turbo, podemos facilmente estimar a intensidade do ataque. A Figura 6.5 ilustra o descrito conceito de sintonia da polarização

Conforme já indicamos, implementou-se na simulação a tradicional técnica SS de modulação/demodulação. A fim de eliminar a influência da interferência do hospedeiro na deteção, e nos resultados da simulação, o sinal hospedeiro foi removido na demodulação, como em uma técnica não-cega. Outra forma de minimizar a interferência do hospedeiro na deteção seria implementar técnicas de modulação/demodulação cegas do tipo ISS, QIM ou SCS. Assim, entendemos que os resultados de nossa simulação, considerando a implementação da técnica não-cega, correspondem aos de uma implementação de uma técnica cega ideal. Na prática, existe uma pequena interferência do hospedeiro na deteção, quando do emprego de técnicas cegas, o que não impede o emprego do esquema proposto, uma vez que o artifício da sintonia da polarização, conforme explicado no parágrafo anterior, pode ser empregado considerando também a interferência do hospedeiro. Inclusive, podemos utilizar o esquema de sintonia de polarização até para estimar qual é a intensidade da interferência do hospedeiro na deteção. Para tanto, no esquema de polarização na ausência de ataque, basta considerar a interferência do hospedeiro como se um tipo de ataque fosse, e proceder a já descrita sintonia para estimá-lo. Desta forma, também pode-se empregar a técnica de sintonia de polarização para avaliar a eficiência da técnica de modulação na eliminação da interferência

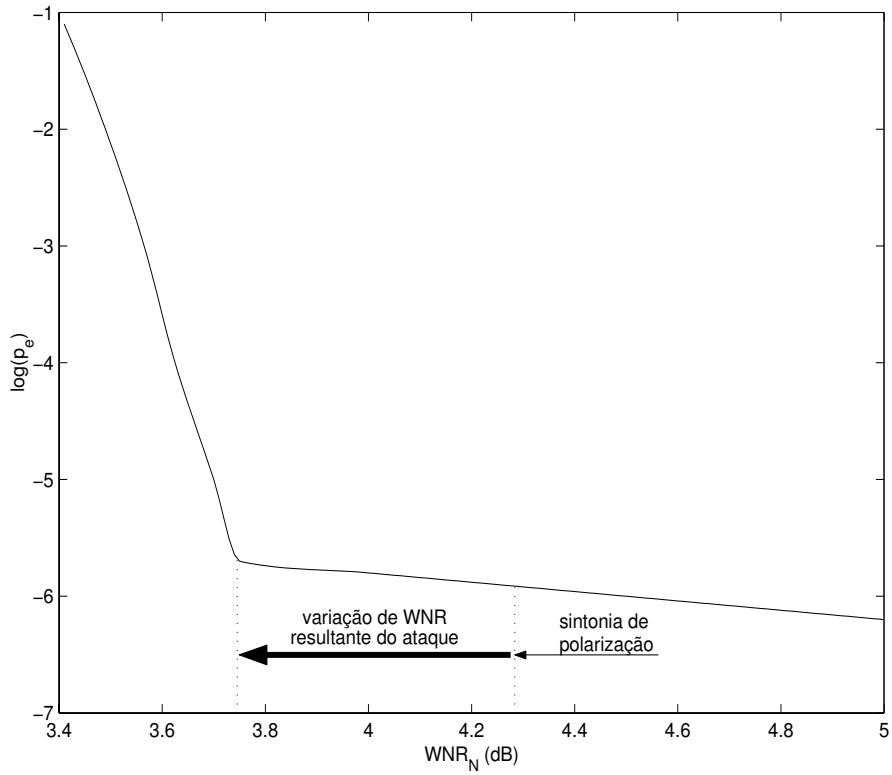

Figura 6.5: Sintonia da polarização para ampliação da faixa dinâmica da estimativa da intensidade do ataque.

do hospedeiro na deteção.

Finalmente, cabe mencionar que o código turbo apresenta um melhor desempenho quando empregado em blocos de bits maiores. Assim, se utilizarmos uma marca d'água de maior comprimento, o proposto esquema de polarização apresentará resultados ainda mais atrativos para aplicações práticas. Como, em muitas aplicações, o domínio para marcação do hospedeiro é limitado, um eventual incremento do comprimento da marca d'água, pode ser compensado pela redução do tamanho N do bloco do hospedeiro modulado por cada bit, considerando ainda o limite de distorção admissível no hospedeiro. Ainda, a marca d'água de maior comprimento fornece uma medida de NEB mais precisa, assim como a correspondente estimativa da intensidade do ataque.

6.4 Comentários

É bastante interessante observar que, como o modelo de polarização proposto é fundamentado na teoria da informação e em sistemas de comunicação, outros cenários de emprego do mesmo podem ser sugeridos. Por exemplo, o esquema pode ser diretamente empregado para caracterização de canais de comunicação, onde o hospedeiro é a portadora, a marca d'água corresponde aos próprios bits de informação, e o ruído do canal corresponde ao ataque que deseja-se estimar. Da mesma forma, o descrito esquema de polarização no detetor, pode indicar se uma determinada transmissão foi sujeita a interferência (*jamming*). E ainda, em comunicações "camufladas" (*covert communications*), a polarização na transmissão, garantirá que a informação secreta será perdida (deteção com baixa confiabilidade), caso haja algum tipo de interferência (ataque) até que o hospedeiro, portador da informação, alcance seu destinatário (receptor).

Outro cenário de emprego do esquema proposto reside no campo dos sensores. Já é largamente conhecida a capacidade de utilizar canais de propagação como transdutores de grandezas físicas, i.e., variações na grandeza que deseja-se monitorar provocam alterações nas condições de propagação do canal, o que resulta na possibilidade de detectar correspondentes alterações de fase e/ou intensidade da onda propagante. Aqui sugerimos que variações destas grandezas físicas também podem ser monitoradas observando o NEB de uma seqüência de bits que se propague pelo canal, e que o esquema de polarização proposto, empregado em uma arquitetura com codificação turbo, proporciona alta sensitividade na correspondente medição de NEB, o que é uma caracterísica essencial quando deseja-se um sensor de alta precisão.