Worst Case Attack on Quantization based Data Hiding

Ning Liu
ECE Department
Stevens Institute of Technology
Hoboken, NJ 07030
nliu@stevens.edu

K.P. Subbalakshmi
ECE Department
Stevens Institute of Technology
Hoboken, NJ 07030
ksubbala@stevens.edu

Abstract

Currently, most quantization based data hiding algorithms are built assuming specific distributions of attacks, such as additive white Gaussian noise (AWGN), uniform noise, and so on. In this paper, we prove that the worst case additive attack for quantization based data hiding is a 3-δ function. We derive the expression for the probability of error (P_e) in terms of distortion compensation factor, α, and the attack distribution. By maximizing P_e with respect to the attack distribution, we get the optimal placement of the 3-δ function. We then experimentally verify that the 3-δ function is indeed the worst case attack for quantization based data hiding.

1 Introduction

Digital data hiding refers to the process of hiding secondary data in host datafor various applications including covert communications, access control, ownership assertion and annotation. Existing data hiding algorithms operate in the spatial [15], or the frequency domain [9, 13]. The mechanism for embedding the hidden data includes spread spectrum based methods [2, 5, 8], quantization based methods [4, 6, 12, 11, 10] and others [17].

Most of the quantization based data hiding work focus on the tradeoff between embedding induced distortion, robustness to attacks, and capacity under power constraints for both the embedder and the attacker. For example, some researchers have proposed several embedding algorithms to optimize the tradeoff against specific attack distributions, such as AWGN and uniform noise. Optimal strategies for the attacker against these data hiding algorithms have also been proposed. For example, Goteti et al. [7] proposed "*QIM watermarking games*" and derived a solution of this game based on the Bhattacharyya bound on the probability of error (P_e). Note that this solution is approximate since it is based on a bound on the probability of error rather than an exact analytical expression. Tzschoppe et al. [14] proposed a complexity reduced watermarking game and derived the numerical solution (as opposed to an analytical solution) to this game by applying the Blahut-Arimoto algorithm [3, 1]. Vila-Forcen et al. [16] hypothesized (without proof) that the 3-δ function is the worst case additive attack against quantization based data hiding schemes. They also find a solution for the watermarking game in higher range of the watermarking noise ratio (WNR $\geq \frac{4}{3}$).

In this paper, *we theoretically prove that the worst case additive attack for quantization based data hiding is indeed a 3-δ function, and we then derive the optimal position of the 3-δ function for the entire range of WNR.* Using a set of images (Barbara, Airplane, Baboon, Lena, Sena, Apple) as the host, we test the 3-δ attack signal against the scalar Costa scheme (SCS) [6] and the dither modulation embedding with distortion compensation (DC-DM) scheme proposed by Chen et al. [4]. The experimental results show that the 3-δ attack results in more probability of error than AWGN and the uniform noise.

2 Problem Formulation

In this paper, we use the quantization based data hiding scheme as the embedding method. Note that throughout this paper, we use scalar uniform quantizer in the data hiding scheme and the mathematical analysis is based on binary embedding which can be easily extend to non-binary cases. For the analysis in this paper, we make the flat-host assumption [12] which means that the statistics of the host can be assumed to be uniform and with infinite variance (in comparison to that of the hidden data). Further, we assume that the additive noise, N, is zero mean: $(E(N) = 0)$ and

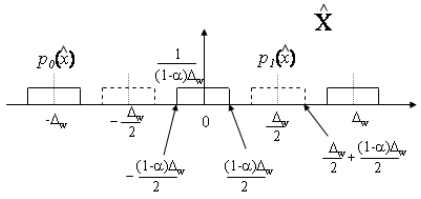

Figure 1. The *pdf* of the stego-image when a single bit is embedded. $p_k(\hat{X})$ represents the *pdf* of the stego-image when a bit $k \in \{0,1\}$ is embedded.

is power constrained: $(\sigma_N^2 \leq D_n)$, and the embedding power is constrained as well: $(\sigma_W^2 \leq D_w)$.

Let a bit $k \in \{0,1\}$ be embedded in the host signal, X, to produce $\hat{X}_k$. Let this signal be corrupted by the noise process N to give $Y_k = \hat{X}_k + N$. The *pdf* of Y_k, $f_Y^{(k)}$, is given by

$$f_Y^{(k)}(t) = f_{\hat{X}}^{(k)}(t) \otimes f_N(t), \tag{1}$$

Without lost of generality, we assume that the quantizer Q_k for embedding a bit k is given by

$$Q_k(X) = i\Delta_w + k\Delta_w/2,$$

for $(k-1)\Delta_w/2 + i\Delta_w \leq X < (k-1)\Delta_w/2 + (i+1)\Delta_w$, where i is an integer $\in [-M, M]$, $X \in [-M\Delta_w, M\Delta_w]$, and Δ_w is the step size of the quantizer. Then, the stego-signal after distortion compensation is:

$$\hat{X}_k = Q_k(X) + (1-\alpha)\left(X - Q_k(X)\right) \quad k = \{0,1\}. \tag{2}$$

where α is the distortion compensation factor.

2.1 The Probability of Error

With these assumptions, the *pdf* of the stego-signal for embedding a hidden bit is shown in Figure-1. Since the hidden data is uniformly distributed,

$$P_e = \sum_{i=-M}^{M} \int_{\frac{\Delta_w}{4}+i\Delta_w}^{\frac{3\Delta_w}{4}+i\Delta_w} f_Y^{(0)}(y)dy, \; i \in [-M, M]. \tag{3}$$

Substitute Eqn-1 into Eqn-3,

$$P_e = \int_{-\infty}^{\infty} f_N(x)\Big(\sum_{i=-M}^{M} \int_{\frac{\Delta_w}{4}+i\Delta_w}^{\frac{3\Delta_w}{4}+i\Delta_w} f_{\hat{X}}^{(0)}(t-x)dt\Big)dx$$

Let us denote $\sum_{i=-M}^{M} \int_{\frac{\Delta_w}{4}+i\Delta_w}^{\frac{3\Delta_w}{4}+i\Delta_w} f_{\hat{X}}^{(0)}(t-x)dt$ by $G(x)$, then P_e can be written in terms of $G(x)$ as:

$$P_e = \int_{-\infty}^{\infty} f_N(x)G(x)dx \tag{4}$$

In the rest of this paper, we will prove that the worst case attack is a 3-δ function and derive the optimal placement for the 3-δ function to achieve the worst case attack.

3 Worst Case Attack: 3-δ function

The worst case attack can be obtained by maximizing P_e over $f_N(x)$: $f_N^*(x) = \arg\max_{f_N(x)} P_e$, under the constraint: $(\int_{-\infty}^{\infty} x^2 f_N(x)dx = \sigma_n^2 \leq D_N)$. Since $G(x)$ is expressed differently in different ranges of α, we consider each range of α and show that the worst case attack is the 3-δ function in each case.

Case 1: $\alpha \geq \frac{1}{2}$

In this case, according to the definition of $G(x)$ in Eqn-4, $G(x)$ is given by,

$$G(x) = \begin{cases} 0, & |x| < \frac{(2\alpha-1)\Delta_w}{4} \\ \frac{(1-2\alpha)\Delta_w + |x|}{4(1-\alpha)\Delta_w}, & \frac{(2\alpha-1)\Delta_w}{4} \leq |x| < \frac{(3-2\alpha)\Delta_w}{4} \\ 1, & \frac{(3-2\alpha)\Delta_w}{4} \leq |x| \leq \frac{(1+2\alpha)\Delta_w}{4} \\ C(x) & |x| > \frac{(1+2\alpha)\Delta_w}{4}. \end{cases} \tag{5}$$

(where $0 \leq C(x) \leq 1$) and graphically described in Figure-2.

The goal of the worst case attack is to maximize P_e as follows,

$$P_e^* = \max_{f_N(x)} P_e, \text{ where } \int_{-\infty}^{\infty} x^2 f_N(x)dx \leq D_N \tag{6}$$

As seen from Figure-2, $G(x)$ is symmetric over "0", so $\int_{-\infty}^{\infty} f_N(x)G(x)dx = \int_{-\infty}^{\infty} f_N(x)G(-l)dx$. Hence, $P_e = \int_{-\infty}^{\infty} f_N(x)G(x)dx = \int_{-\infty}^{\infty} f_N(-x)G(x)dx$. So, $f_N(x)$ is also symmetric over zero. Let $f_N^*(x)$ represent the optimal $f_N(x)$ maximizing P_e. Because of the symmetric of $f_N(x)$, we only need to find $f_N^*(x)$, in the interval $x \in [0, \infty)$.

From Eqn-4 and Eqn-5, $f_N^*(x) = 0$ for $|x| \in (0, \frac{(2\alpha-1)\Delta_w}{2})$, since $P_e = 0$ for any attack power spent in this range. We now calculate P_e for other ranges of x.

* **When** $|x| \geq \frac{(3-2\alpha)\Delta_w}{4}$,

$$P_e = 2\int_{\frac{(3-2\alpha)\Delta_w}{4}}^{\infty} f_N(x)G(x)dx \leq 2\int_{\frac{(3-2\alpha)\Delta_w}{4}}^{\infty} f_N(x)dx,$$

with equality only if $f_N(x) = 0$ for $|x| \in (\frac{(2\alpha+1)\Delta_w}{4}, \infty)$, according to the definition of $G(x)$ in this case.

Let $P = 2\int_{\frac{(3-2\alpha)\Delta_w}{4}}^{\infty} f_N(x)dx$, then the attacking power spent in the range of $|x| \in \left[\frac{(3-2\alpha)\Delta_w}{4}, \infty\right)$ becomes

$$2\int_{\frac{(3-2\alpha)\Delta_w}{4}}^{\infty} x^2 f_N(x)dx \geq P\left(\frac{(3-2\alpha)\Delta_w}{4}\right)^2, \tag{7}$$

with equality only if $f_N(x) = P\delta\left(x - \frac{(3-2\alpha)\Delta_w}{4}\right)$. $f_N^*(x) = 0$ for $|x| \in \left(\frac{(3-2\alpha)\Delta_w}{4}, \infty\right)$, because when $f_N(x) = P\delta\left(x - \frac{(3-2\alpha)\Delta_w}{4}\right)$, the attacking signal results in the maximum error with the least attacking power in the range of $|x| \in \left[\frac{(3-2\alpha)\Delta_w}{4}, \infty\right)$.

* **When** $x \in \left[\frac{(2\alpha-1)\Delta_w}{4}, \frac{(3-2\alpha)\Delta_w}{4}\right]$**,** Let's denote $B = \int_{\frac{(2\alpha-1)\Delta_w}{4}}^{\frac{(3-2\alpha)\Delta_w}{4}} f_N(x)dx$.

Theorem 1 ***The pdf of the optimal attack noise,*** $f_N^*(x)$***, is an impulse function in the range*** $|x| \in \left[\frac{(2\alpha-1)\Delta_w}{4}, \frac{(3-2\alpha)\Delta_w}{4}\right]$***, and is given by,*** $f_N^*(x) = B\delta(x-l)$, $\frac{(2\alpha-1)\Delta_w}{4} \leq l \leq \frac{(3-2\alpha)\Delta_w}{4}$**,**

Proof: Let $f_N(x)$ be any noise probability density function. Let $\xi_N(x)$ be a specific example given by

$$\xi_N(x) = B\delta(x-l), \frac{(2\alpha-1)\Delta_w}{4} \leq l \leq \frac{(3-2\alpha)\Delta_w}{4}$$

We show that in order to achieve the same value of P_e, $\xi_N(x)$ needs to spend less power than any other $f_N(x)$.

Let P_e be the probability of error achieved by $f_N(x)$, then

$$P_e = \int_{\frac{(2\alpha-1)\Delta_w}{4}}^{\frac{(3-2\alpha)\Delta_w}{4}} f_N(x)G(x)dx \tag{8}$$

If $\xi_N(x)$ achieves the same P_e,

$$\int_{\frac{(2\alpha-1)\Delta_w}{4}}^{\frac{(3-2\alpha)\Delta_w}{4}} \xi_N(x)G(x)dx = \int_{\frac{(2\alpha-1)\Delta_w}{4}}^{\frac{(3-2\alpha)\Delta_w}{4}} f_N(x)G(x)dx$$

From the definition of $G(x)$ (Eqn-5),we get

$$\int_{\frac{(2\alpha-1)\Delta_w}{4}}^{\frac{(3-2\alpha)\Delta_w}{4}} f_N(x)\frac{(1-2\alpha)\Delta_w + x}{4(1-\alpha)\Delta_w}dx$$

$$= \int_{\frac{\Delta_w}{4} - \frac{(1-\alpha)\Delta_w}{2}}^{\frac{\Delta_w}{4} + \frac{(1-\alpha)\Delta_w}{2}} B\delta(x-l)\frac{(1-2\alpha)\Delta_w + x}{4(1-\alpha)\Delta_w}dx$$

$$\Longrightarrow \int_{\frac{(2\alpha-1)\Delta_w}{4}}^{\frac{(3-2\alpha)\Delta_w}{4}} x f_N(x)dx = Bl \tag{9}$$

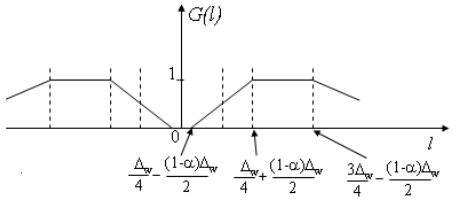

Figure 2. The function of $G(x)$ **for** $\alpha \geq \frac{1}{2}$

For $x \in \left[\frac{(2\alpha-1)\Delta_w}{4}, \frac{(3-2\alpha)\Delta_w}{4}\right]$, the powers associated with attack distribution of $\xi_N(x)$ and $f_N(x)$ are given by:

$$\text{Power}\left(\xi_N(x)\right) = \int_{\frac{(2\alpha-1)\Delta_w}{4}}^{\frac{(3-2\alpha)\Delta_w}{4}} x^2\xi_N(x)dx,$$

$$\text{Power}\left(f_N(x)\right) = \int_{\frac{(2\alpha-1)\Delta_w}{4}}^{\frac{(3-2\alpha)\Delta_w}{4}} x^2 f_N(x)dx,$$

respectively. By simple mathematical manipulation (The details is omitted because of the page limitation), it can be shown that

$$\text{Power}(f_N(x)) - \text{Power}(\xi_N(x)) \geq 0$$

with equality only if $f_N(x) = \xi_N(x)$. So Theorem-1 is proved.■

Since $f_N(x)$ is symmetric over zero, for $x \in [-\frac{(3-2\alpha)\Delta_w}{2}, -\frac{(2\alpha-1)\Delta_w}{2}]$, $f_N^*(x) = B\delta(x+l)$.

Case 2: $\alpha < \frac{1}{2}$**,**

$G(x)$ is now given by

$$G(x) = \begin{cases} \frac{1-2\alpha}{2(1-\alpha)}, & 0 \leq |x| < \frac{(1-2\alpha)\Delta_w}{4} \\ \frac{(1-2\alpha)\Delta_w + |x|}{4(1-\alpha)\Delta_w}, & \frac{(1-2\alpha)\Delta_w}{4} \leq |x| < \frac{(1+2\alpha)\Delta_w}{4} \\ C(x), & |x| \geq \frac{(1+2\alpha)\Delta_w}{4} \end{cases} \tag{10}$$

where $0 < C(x) \leq \frac{1}{2(1-\alpha)}$ with equality only if $|x| = \frac{(1+2\alpha)\Delta_w}{4}$. Following the same logic as in the case of $\alpha \geq \frac{1}{2}$, $f_N^*(x) = 0$, for $|x| \in (0, \frac{(1-2\alpha)\Delta_w}{4})$ and $(\frac{(1+2\alpha)\Delta_w}{4}, \infty)$

Theorem 2 ***The optimal attack noise pdf*** $f_N^*(x)$ ***is an impulse function in the range of*** $[\frac{(1-2\alpha)\Delta_w}{4}, \frac{(1+2\alpha)\Delta_w}{4}]$**, and is given by** $f_N^*(x) = B\delta(x-l)$, $\frac{(1-2\alpha)\Delta_w}{4} \leq l \leq \frac{(1+2\alpha)\Delta_w}{4}$**, where** $B = \int_{\frac{(1-2\alpha)\Delta_w}{4}}^{\frac{(1+2\alpha)\Delta_w}{4}} f_N(x)dx$**.**

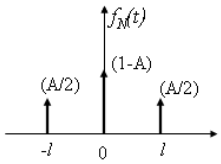

Figure 3. The *pdf* of the worst case attack: 3-δ function

The proof is omitted, since it follows the same logic as the proof for Theorem-1. Also, in this case, for $x \in \left[-\frac{(1+2\alpha)\Delta_w}{4}, -\frac{(1-2\alpha)\Delta_w}{4}\right]$, $f_N^*(x) = B\delta(x+l)$.

Consolidating the above two cases, we can conclude that the worst case attack signal is a 3-δ function as shown in Figure-3. Let's define

$$l_{lower} = \begin{cases} \frac{(1-\alpha)\Delta_w}{2} - \frac{\Delta_w}{4}, & \alpha < \frac{1}{2} \\ \frac{\Delta_w}{4} - \frac{(1-\alpha)\Delta_w}{2}, & \alpha \geq \frac{1}{2} \end{cases} \quad (11)$$

$$l_{upper} = \begin{cases} \frac{3\Delta_w}{4} - \frac{(1-\alpha)\Delta_w}{2}, & \alpha < \frac{1}{2} \\ \frac{(1-\alpha)\Delta_w}{2} + \frac{\Delta_w}{4}, & \alpha \geq \frac{1}{2} \end{cases} \quad (12)$$

$$f_N^*(x) = \frac{A}{2}\delta(x+l) + (1-A)\delta(x) + \frac{A}{2}\delta(x-l), \quad (13)$$

where $l \in [l_{lower}, l_{upper}]$ and $A \in [0,1]$ is a scalar factor. Since the attack power is constrained by D_N,

$$\int_{-\infty}^{\infty} x^2 f_N(x)dx \leq D_N \quad (14)$$

Substituting Eqn-13 for $f_N(x)$ from the above equation, we get

$$A = \frac{\sigma_n^2}{l^2} \quad (15)$$

Since $0 \leq A \leq 1$, we get $l \geq \sigma_n$. If $\sigma_n > l_{upper}$,

$$f_N^*(x) = \frac{1}{2}\delta(x+l) + \frac{1}{2}\delta(x-l), \quad (16)$$

where $l = l_{upper}$. Otherwise, if $\sigma_n \leq l_{upper}$,

$$f_N^*(x) = \frac{A}{2}\delta(x+l) + (1-A)\delta(x) + \frac{A}{2}\delta(x-l), \quad (17)$$

where $l \in [\sigma_n, l_{upper}]$ Using Equations-(4), (5), (10), (11), (16), and (17), we express P_e as a function of l and the watermark to noise ratio (WNR $= \frac{\sigma_w^2}{\sigma_n^2}$)

If WNR $< \frac{1}{12}$,

$$P_e = \begin{cases} \frac{2\alpha l\sigma_n^2 + \sqrt{3}(1-2\alpha)\sigma_w(2l^2-\sigma_n^2)}{4\sqrt{3}(1-\alpha)\sigma_w l^2}, & \alpha \leq \frac{\sqrt{3\text{WNR}}}{2(1-\sqrt{3\text{WNR}})} \\ \frac{1}{2(1-\alpha)}, & \frac{\sqrt{3\text{WNR}}}{2(1-\sqrt{3\text{WNR}})} < \alpha < \frac{1}{2} \\ 1, & \alpha \geq \frac{1}{2} \end{cases} \quad (18)$$

If WNR $\geq \frac{1}{12}$,

$$P_e = \begin{cases} \frac{\sigma_n^2}{l^2}\frac{(1-2\alpha)\sqrt{3}\sigma_w + 2\alpha l}{4(1-\alpha)\sqrt{3}\sigma_w}, & \frac{1}{2} \leq \alpha < \frac{3\sqrt{3\text{WNR}}}{2(1+\sqrt{3\text{WNR}})}, \\ \frac{2\alpha l\sigma_n^2 + \sqrt{3}(1-2\alpha)\sigma_w(2l^2-\sigma_n^2)}{4\sqrt{3}(1-\alpha)\sigma_w l^2}, & \alpha < \frac{1}{2}, \\ 1, & \alpha \geq \frac{3\sqrt{3\text{WNR}}}{2(1+\sqrt{3\text{WNR}})}, \end{cases} \quad (19)$$

4 Optimal placement of the 3-δ function

In the previous subsection, we showed that the optimal strategy for an attacker is a 3-δ function under attack power constraint $\int_{-\infty}^{\infty} x^2 f_N(x)dx \leq D_N$. We also showed the optimal placement of the 3-δ function for some special cases of α and WNR:

$$l^* = \begin{cases} \frac{(1+2\alpha)\Delta_w}{4}, & \frac{\sqrt{3\text{WNR}}}{2(1-\sqrt{3\text{WNR}})} < \alpha < \frac{1}{2}, \ \text{WNR} < \frac{1}{12} \\ \frac{(3-2\alpha)\Delta_w}{4}, & \alpha \geq \frac{3\sqrt{3\text{WNR}}}{2(1+\sqrt{3\text{WNR}})}, \ \text{WNR} \geq \frac{1}{12} \\ \frac{(3-2\alpha)\Delta_w}{4}, & \alpha \geq \frac{1}{2}, \ \text{WNR} < \frac{1}{12} \end{cases}$$

The corresponding probability of error (P_e) are

$$P_e = \begin{cases} \frac{1}{2(1-\alpha)}, & \frac{\sqrt{3\text{WNR}}}{2(1-\sqrt{3\text{WNR}})} < \alpha < \frac{1}{2}, \ \text{WNR} < \frac{1}{12} \\ 1, & \alpha \geq \frac{3\sqrt{3\text{WNR}}}{2(1+\sqrt{3\text{WNR}})}, \ \text{WNR} \geq \frac{1}{12} \\ 1, & \alpha \geq \frac{1}{2}, \ \text{WNR} < \frac{1}{12} \end{cases}$$

In the rest of this section, we will find the optimal placement of the 3-δ function for the other ranges of α and WNR.

Case 1: $\frac{1}{2} \leq \alpha < \frac{3\sqrt{3\text{WNR}}}{2(1+\sqrt{3\text{WNR}})}$ and WNR $> \frac{1}{12}$

We now define two terms: globally optimal probability of error, $P_e^{\text{glo}}(l^*)$, which refers to the maximum value of P_e obtained *anywhere* along l, and locally optimal probability of error, $P_e^{\text{loc}}(l^*)$, which refers to the maximum value of P_e within the valid range of l. We are interested in determining only $P_e^{\text{loc}}(l^*)$ because we have a power constraint which determines the valid range for l. In some cases, the local and global optima coincide (and therefore lie inside the valid range). Under this condition, we can use the partial differential method to determine the optimal P_e, $P_e(l^*)$. When the locally optimal point is different from the globally optimal solution, we need to use special methods.

Let $l_{\text{glo}}^* = \arg\ \max_l P_e(l)$ denote the global optima. and the maximum power be $P_e(l_{\text{glo}}^*)$. To maximize $P_e(l)$ with l, we take the partial differential of P_e with respect to l,

$$\frac{\partial P_e}{\partial l} = \frac{\sigma_n^2((6\alpha-3)\sigma_w - \sqrt{3}\alpha l)}{6l^3(1-\alpha)\sigma_w} \quad (20)$$

and equalize it to zero to get $l^*_{\text{glo}} = \frac{\sqrt{3}\sigma_w(2\alpha-1)}{\alpha}$. It is easily verified that l^*_{glo} indeed maximizes P_e.

If the local and global optima coincide, the global optimum, l^*_{glo}, lies in the valid range. Then, according to Eqn-17, $\sigma_n \le l^*_{\text{glo}} \le \frac{(3-2\alpha)\Delta_w}{4}$. So in this case, $\frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1} \le \alpha \le \frac{5}{6}$,

$$P_e(l^*) = \max_{f_N(x)} P_e(l) = \frac{\alpha^2}{12(1-\alpha)(2\alpha-1)\text{WNR}}, \quad (21)$$

and $l^* = \frac{\sqrt{3}\sigma_w(2\alpha-1)}{\alpha}$.

So now we deal with the case where the local and global optima do not coincide, which means l^*_{glo} does not lie in $[\sigma_n, \frac{(3-2\alpha)\Delta_w}{4}]$. We split the problem into 2 cases: $l^*_{\text{glo}} < \sigma_n$ and $l^*_{\text{glo}} > \frac{(3-2\alpha)\Delta_w}{4}$, which induce $\frac{1}{2} \le \alpha < \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1}$ and $\alpha > \frac{5}{6}$.

* **When $\frac{1}{2} \le \alpha < \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1}$:** Substituting σ_n for l in Eqn-20, $\frac{\partial P_e}{\partial l} < \frac{\sigma_n^2(\sqrt{3}\alpha\sigma_n - \sqrt{3}\alpha l)}{6l^3(1-\alpha)\sigma_w} \le 0$, for $l \in [\sigma_n, \frac{\Delta_w}{4} + \frac{(1-\alpha)\Delta_w}{2}]$. That is, P_e *is monotonically decreasing in the valid range* $[\sigma_n, \frac{\Delta_w}{4} + \frac{(1-\alpha)\Delta_w}{2}]$. Therefore the local optimal solution is at the left edge of the range, $l^* = \sigma_n$ and

$$P_e(l^*) = \max_{f_N(x)} P_e(l) = \frac{2\sqrt{3}\alpha - 3(2\alpha-1)\sqrt{\text{WNR}}}{12(1-\alpha)\sqrt{\text{WNR}}}.$$

* **When $\alpha > \frac{5}{6}$:** Using the same method as ($\frac{1}{2} \le \alpha < \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1}$), we can get $l^* = \frac{\sqrt{3}\sigma_w(3-2\alpha)}{2\alpha}$ and

$$P_e(l^*) = \max_{f_N(x)} P_e(l) = \frac{4\alpha^2}{3(3-2\alpha)^2\text{WNR}}, \quad (22)$$

So combining the solution for $P_e(l)$ for both cases of local optimal l^*,

$$l^* = \begin{cases} \frac{\sqrt{3}\sigma_w(3-2\alpha)}{2\alpha}, & \alpha > \frac{5}{6} \\ \frac{\sqrt{3}\sigma_w(2\alpha-1)}{\alpha}, & \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1} \le \alpha \le \frac{5}{6} \\ \sigma_n, & \frac{1}{2} \le \alpha < \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1} \end{cases}$$

The corresponding $P_e(l^*)$ is given by

$$P_e(l^*) = \begin{cases} \frac{4\alpha^2}{3(3-2*\alpha)^2\text{WNR}}, & \alpha > \frac{5}{6} \\ \frac{\alpha^2}{12(1-\alpha)(2\alpha-1)\text{WNR}}, & \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1} \le \alpha \le \frac{5}{6} \\ \frac{2\sqrt{3}\alpha - 3(2\alpha-1)\sqrt{\text{WNR}}}{12(1-\alpha)\sqrt{\text{WNR}}}, & \frac{1}{2} \le \alpha < \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1} \end{cases}$$

Following the same logic with Case 1, we can get the l^* and corresponding $P(l^*)$ for the rest two cases: Case 2, $\alpha < \frac{1}{2}$ and $\text{WNR} \ge \frac{1}{12}$; Case 3, $\alpha \le \frac{\sqrt{3\text{WNR}}}{2(1-\sqrt{3\text{WNR}})}$ and $\text{WNR} < \frac{1}{12}$.

Consolidating the solutions for all ranges, we get the optimal placement of the 3-δ function, l^* given by:

$$\begin{cases} \frac{(1+2\alpha)\sqrt{3}\sigma_w}{2\alpha}, & \alpha \le \frac{\sqrt{3\text{WNR}}}{2(1-\sqrt{3\text{WNR}})},\ \text{WNR} < \frac{1}{48} \\ \frac{(1+2\alpha)\sqrt{3}\sigma_w}{2\alpha}, & \alpha < \frac{1}{6},\ \text{WNR} \ge \frac{1}{48} \\ \frac{\sqrt{3}\sigma_w(1-2\alpha)}{\alpha}, & \frac{1}{6} \le \alpha < \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}+1},\ \text{WNR} \ge \frac{1}{48} \\ \sigma_n, & \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}+1} \le \alpha < \frac{\sqrt{3\text{WNR}}}{2(1-\sqrt{3\text{WNR}})},\ \frac{1}{48} \le \text{WNR} < \frac{1}{12} \\ \frac{(1+2\alpha)\sqrt{3}\sigma_w}{2\alpha}, & \frac{\sqrt{3\text{WNR}}}{2(1-\sqrt{3\text{WNR}})} \le \alpha \le \frac{1}{2},\ \text{WNR} < \frac{1}{12} \\ \frac{\sqrt{3}\sigma_w(3-2\alpha)}{2\alpha}, & \alpha > \frac{1}{2},\ \text{WNR} < \frac{1}{12} \\ \sigma_n, & \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}+1} \le \alpha < \frac{3\sqrt{3\text{WNR}}}{2(1+\sqrt{3\text{WNR}})},\ \frac{1}{12} \le \text{WNR} < \frac{25}{48} \\ \frac{\sqrt{3}\sigma_w(3-2\alpha)}{2\alpha}, & \frac{3\sqrt{3\text{WNR}}}{2(1+\sqrt{3\text{WNR}})} \le \alpha < 1,\ \frac{1}{12} \le \text{WNR} < \frac{4}{3} \\ \sigma_n, & \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}+1} \le \alpha < \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1},\ \text{WNR} \ge \frac{25}{48} \\ \frac{\sqrt{3}\sigma_w(2\alpha-1)}{\alpha}, & \frac{\sqrt{3\text{WNR}}}{2\sqrt{3\text{WNR}}-1} \le \alpha < \frac{5}{6},\ \text{WNR} \ge \frac{25}{48} \\ \frac{\sqrt{3}\sigma_w(3-2\alpha)}{2\alpha}, & \frac{5}{6} \le \alpha < \frac{3\sqrt{3\text{WNR}}}{2(1+\sqrt{3\text{WNR}})},\ \frac{25}{48} \le \text{WNR} < \frac{4}{3} \\ \frac{\sqrt{3}\sigma_w(3-2\alpha)}{2\alpha}, & \frac{5}{6} \le \alpha \le 1,\ \text{WNR} \ge \frac{4}{3} \end{cases} \quad (23)$$

The corresponding $P_e(l^*)$ for WNR from -20dB to 20dB is graphically described in Figure-4.

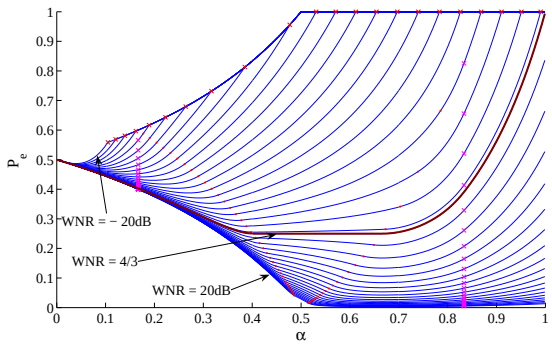

Figure 4. The optimal solution for the attacker, $P_e(l^*)$ (WNR=-20dB 20 dB)

5 Experimental Results and Discussion

Using a set of 512x512 gray images (Barbara, Airplane, Baboon, Lena, Sena, Apple) as the host, we embed one bit per pixel. The performance comparison between uniform noise, AWGN and the derived 3-δ attack against DC-DM and SCS are shown in Figure-5. As seen from this figure, the 3-δ attack results in more probability of error and is more attacking-efficient against DC-DM and SCS comparing to uniform noise and AWGN.

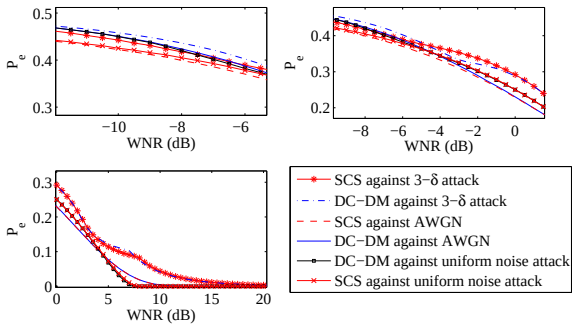

Figure 5. The P_e comparison between the 3-δ attack, AWGN and uniform noise attack against DC-DM and SCS.

As assumed by the definition of worst case additive attack, the 3-δ attack results in more P_e than any other attacks for quantization based data hiding. As seen from Figure-5, experimental results confirm the assumptions.

6 Conclusion

In this paper, we modeled the general quantization based data hiding scheme as a power constrained data hiding framework. Using the P_e as the cost function, we proved that the best strategy for the attacker is a 3-δ function and derived the mathematical expression of the optimal strategy for the attacker. Experimental results show that the best strategy for the attacker results in the maximum P_e compared to other attacks for any given quantization based scheme.

References

[1] S. Arimoto. An algorithm for calculating the capacity of an arbitrary discrete memoryless channel. *IEEE Transactions on Information Theory*, IT-18(1):14–20, 1972.

[2] W. Bender, D. Gruhl, and N. Morimoto. Techniques for data hiding. In *Proc. SPIE*, volume 2420, 1995.

[3] R. E. Blahut. Computation of channel capacity and rate distortion functions. *IEEE Transactions on Information Theory*, IT-18(4):460–473, 1972.

[4] B. Chen and G. W. Wornell. Quantization index modulation: A class of probably good methods for digital watermarking and information embedding. *IEEE Trans. on Information Theory*, 47(4):1423–1443, 2001.

[5] I. Cox, J. Kilian, T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6:1673–1687, Dec. 1997.

[6] J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod. Scalar costa scheme for information embedding. *IEEE Transactions on Signal Processing*, 51:1003 –1019, 2003.

[7] A. K. Goteti and P. Moulin. QIM watermarking games. In *Proc. IEEE International Conference on Image Processing (ICIP'04)*, Singapore, Oct. 2004.

[8] D. F. HS Malvar. Improved spread spectrum: a new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing*, 2003.

[9] C.-T. Hsu and J.-L. Wu. A DWT-DFT composite watermarking scheme robust to both affine transform and jpeg compression. *IEEE Transactions on Circuits System and Video Technology*, 2003.

[10] N. Liu and K. P. Subbalakshmi. Non-uniform quantizer design for image data hiding. In *IEEE International Conference on Image Processing*, Singapore, Oct. 2004.

[11] N. Liu and K. P. Subbalakshmi. Vector quantization based scheme for data hiding for images. In *Proc. SPIE International Conference on Electronic Images'04*, San Jose, CA, Jan. 2004.

[12] F. Perez-Gonzalez, F. Balado, and J. Martin. Performance analysis of existing and new methods for data hiding with known-host information in additive channels. *IEEE Transactions on Signal Processing*, 51(4):960 –980, 2003.

[13] K. Solanki, N. Jacobsen, U. Madhow, B. S. Manjunath, and S. Chandrasekaran. Robust image-adaptive data hiding using erasure and error correction. *IEEE Transactions on Image Processing*, 13(12):1627–1639, Dec. 2004.

[14] R. Tzschoppe, R. Bauml, R. Fischer, J. Huber, and A. Kaup. Additive non-gaussian noise attacks on the scalar costa scheme (SCS). In *Proc. SPIE International Conference on Electronic Images'05*, volume 5681, San Jose, CA, Jan. 2005.

[15] M. Utku-Celik, G. Sharma, E. Saber, and A. Murat-Tekalp. Hierarchical watermarking for secure image authentication with localization. *IEEE Transactions on Image Processing*, 11(6):585–595, June 2002.

[16] J. Vila-Forcen, S. Voloshynovskiy, O. Koval, F. Perez-Gonzalez, and T. Pun. Worst case additive attack against quantization-based data-hiding methods. In *Proc. SPIE International Conference on Electronic Images'05*, volume 5681, San Jose, CA, Jan. 2005.

[17] P. Wong and N. Memon. Secret and public key image watermarking schemes for image authentication and ownership verification. *IEEE Transactions on Image Processing*, 10(10):1593–1601, Oct. 2001.