Lecture 15. Capacity of IH systems

1. Capacity of stegosystems

The notion of channel capacity is well known for communication due to the pioneering Shannon's work[64] at 1948. The most essential significance of this notion(also due to Shannon) is the *following theorem*:

If the code rate R is lower than capacity C then there exist coding and decoding algorithms providing a decreasing to zero error probability after decoding(error correction) P_e as the code block length approaches to infinity. If $R > C$ then these conditions do not hold.

Shannon formula for channel coding with discrete time Gaussian channel is

$$C = \frac{1}{2} log_2 \left(1 + \frac{P_s}{P_n}\right) \tag{1}$$

where P_s - signal power, P_n- noise power.

Definition of capacity of stego channel is different:

This is the maximum code rate R for which also $P_e \to 0$ as the code block length $n \to \infty$ but with additional condition that relative entropy D is given.

In the paper[65] has been proved the capacity for Gaussian CO, that is irrealistic for practice.

Let us prove that relation for SG capacity when there is Gaussian noisy stego channel (See Lecture 5). Then C does not depend on statistic of CO.

We assume that informed legal decoder takes a decision about the embedding of the *i*-th codeword by making

$$i' = \arg\max_i \sum_{l=1}^{n} (C'_w(l) - C(l))(-1)^{b_{il}} \cdot \pi(l), \qquad (2)$$

where $C'_w(l) = C_w(l) + \varepsilon(l)$, i-the i –th code word, l – the l-th bit of the code word ;

ε is a zero-mean Gaussian i.i.d. sequence with variance σ_ε^2. In Lecture 5 was presented the formula for D:

$$D = 0.77n[ln(1+\eta)^{-1} - (1+\eta)^{-1}], \qquad (3)$$

where $\eta = \frac{\sigma_\varepsilon^2}{\sigma_n^2}$ is the *noise-to-watermark-ratio(NWR).*

D by (3) may be approximated for typically large NWR as

$$D = 0.36n\eta^{-2} \qquad (4)$$

That allows to present NWR as

$$\eta = 0.36\sqrt{\frac{n}{D}} \qquad (5)$$

On the other hand, using additive bound for P_e[66]:

$$P_e \le (2^k - 1)Q(\sqrt{\frac{d}{\eta}}),\ Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}}dt \qquad (6)$$

where *k* – is the number of information symbols in *n*-bit code block, *d* – the minimal code distance for (*n*,*k*) – code.

After simple transforms and taking into account that for any (*n,k*) a block-code $d \leq n$, we get from (5) and (6):

$$P_e \leq \exp\left(R \cdot n \cdot ln2 - \frac{n\sqrt{D}}{1.2\sqrt{n}}\right) = \exp(R \cdot n \cdot ln2 - \frac{\sqrt{nD}}{1.2}) \qquad (7)$$

We can see from (7) that $P_e \to 0$ as $n \to \infty$ for any security level $D > 0$ only if

$$R \leq 1.2\sqrt{\frac{D}{n}} \qquad (8)$$

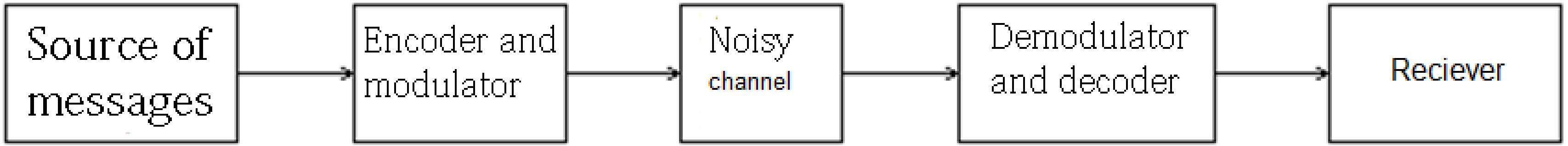
The inequality (8) means that the capacity of SG for a noisy attack channel and for any given security level D is ZERO. This result is quite reasonable: the greater is the block length the greater is the information that can be used by an attacker in order to distinguish the CO and SG.

We can find also from (7) that the code rate obey to square root law of SG[65] and more early[15].

With “the point of view of additive bound” any SG for which in order to keep constant the security level with increasing of the block length, it is required to decrease the WNR, and hence it has zero capacity.(See also [67])

But even in the case C=0 it is still possible to embed into CO some amount of secure and reliable bits.

2. Capacity of watermark systems.



Scheme of ordinary noisy channel

Shannon’s formula(1948)for channels with discrete time

$$C = \frac{1}{2}\log_2\left(1+\frac{P_s}{P_n}\right)\text{(bit/sample)}$$

where P_s – signal power
P_n – noise power (1)

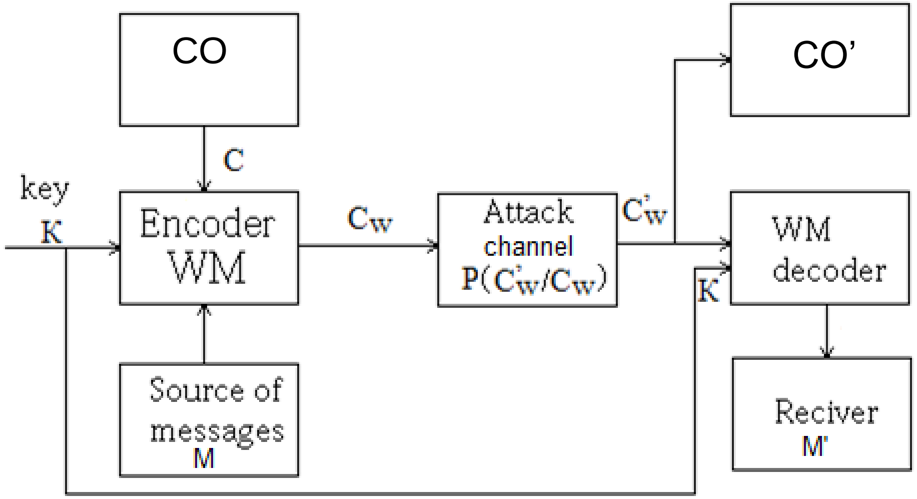
$C \to \infty$, if $P_s \to \infty$ or $P_n \to 0$,
$C \to 0$, if $P_s \to 0$ or $P_n \to \infty$

C = $\frac{1}{2}$ bit/sample, if P_s=P_n

Remark. Result is non-constructive because it gives only a potential opportunities of message transmission : it is impossible to transmit messages reliably with the rate more than C and possible with the rate C-ε, where ε>0. However it was an open problem how has to be design coder and decoder. These problem is solved in “Coding theory”[43].

Scheme of WM system.[19]

The main differences among communication systems and WM systems :



1. The main interference (CO) is own at the WM encoder .

2. Attacks should not make significantly worse the quality of CO.

$P(C'_W/C_W)$

3.Attack channel is not necessary equivalent to an additive noise .There is some conditional probability distribution

4. Generally speaking , between a designer of WM system and attacker there exists some game –situation :

Designer wants to provide a maximum embedding rate given small error probability whereas an attacker try to minimize the rate or (and) in an increasing of the error probability. (Similar situation appears in a communication system under the condition of signal jamming)

Constrains under WM embedding and attack

$$\sum_{(C,W,K)} \left(P(C,W,K)\, D\left(C,C_w\right)\right) \le D_1 \qquad (2)$$

where C - CO, W - WM, К – stegokey, D(C,C_w) – distortion measure of CO after embedding

Typical distortion measure (mean-square error-MSE):

$$D(C,C_w) = E\left[\left(C - C_w\right)^2\right] \qquad (3)$$

$$\sum_{\left(C_w,C'_w\right)} P\left(C'_w / C_w\right) D\left(C_w,C'_w\right) \le D_2 \qquad (4)$$

where C_w – CO after WM embedding, C'_w- attacked C_w, $D(C_w,C'_w)$ - distortion measure after attack .

MSE is :

$$D(C_w,C'_w) = E\left[\left(C_w - C'_w\right)^2\right] \qquad (5)$$

It is worth to noting that more naturally is to take a restriction based on distortions between CO and attacked CO :

$$\sum_{\left(C,C'_w\right)} P\left(C,C_w\right) D\left(C,C'_w\right) \le \widetilde{D_2} \qquad (6)$$

Extra definitions

Let us define some encoding function f_N(…) and decoding function Y_N(…) for WM system:

$C_w^N = f_N\left(C^N, K^N, M\right)$, $M' = Y_N\left(C_w'^N, K^N\right)$, where N – is the block length

Then let us define the probability of incorrect WM message block decoding as follows:

$$P_{e,N} = P\left(Y_N(C_w'^N, K^N) \neq M'\right) \tag{6}$$

The WM rate is: $$R = \frac{\log |M|}{N} \tag{7}$$

where |M| - is the total number of messages which can be embedded into N samples. If k bits is embedded into N samples , then

$$R = \frac{k}{N} \text{ (bits/samples)} \tag{8}$$

Definition of WM system capacity

We assume that an attacker knows an encoder function $f_N(\ldots)$, whereas legal decoder knows both encoder function $f_N(\ldots)$, and attack channel $P(C'_w/C_w)$
(This is not very strong restriction because these information can be get observing a statistic of WM embedding).
Capacity of WM system is maximum embedding rate under the condition that $P_{e,N} \rightarrow 0$ as $N\rightarrow\infty$, after trial of all encoding and decoding functions and attack channels ($f_N(\ldots)$,$Y_N(\ldots)$, $P(C'_w/C_w)$), which satisfy the restrictions D_1 and D_2 (see (2) and (4)).
It was proved in [19], that

$$\underline{C} = \max_{P(C_w/C)} \min_{P(C'_w/C_w)} I(C_w;C'_w/C) \qquad (9)$$

where $I(C_w;C'_w/C)$ - amount of conditional Shannon’s information,
and $P(C_w/C)\,P(C'_w/C_w)$ satisfy the restrictionsD_1 and D_2.

If informed decoder is used (CO is known at the decoder) then we have

$$\tilde{\underline{C}} = \min_{P(C'_w/C_w)} \max_{P(C_w/C)} I(C_w;C'_w/C) = C \qquad \text{(see (9))} \qquad (10)$$

General properties of WM system capacity [19]:

1. **<u>C</u>**=C(D_1,D_2) increases monotonically with D_1 and decreases monotonically with D_2.

2. **<u>C</u>**(D_1,D_2) – is convex on D_2.

3. **<u>C</u>**(D_1,D_2) ≤ log|C|, where |C| - is amount of all possible CO
(This means that it is more easy to embed WM into CO of more size).

4. If D_1=0, then **C**(0, D_2)=0 for all D_2.
(This means that it is impossible to embed anything without some corruption of CO. This statement by the way is untrue for so called “reversible WM” (see Lecture 13)).

5. There exists always such D'_2 : **C**(D_1,D_2)=0, if D_2≥D'_2 for any D_1(This means that if distortions are not limited then nothing can be embedded (“No methods against crowbar»).

6. Estimation attack is very important.
So if an attacker was able to find such C'_W that $H\left(C'_W / C\right) \le \varepsilon$ then $C < \varepsilon$

Calculation of C for particular models of CO

1. *Binary i.i.d samples with distortion function based on Hamming distance.*

It was proved in [19], that in this case

$$C=h(D_1*D_2)-h(D_2), \qquad (10)$$

where $D_1*D_2=D_1(1-D_2)+D_2(1-D_1)$, $h(x)= -(x \log_2 x+(1-x) \log_2(1-x))$,

That is achieved with embedding by the rule

$$C_w(n) := C(n) \oplus Z(n) \quad n := 1,2 \ldots N \qquad (11)$$

where $Z(n) \epsilon (0,1)$, (i.i.d)$P(Z(n)=1)=D_1$, $P(Z(n)=0)=1-D_1$, while the message M is encoded into Z(n), n=1,2 … N,

and for optimal attack:

$$C'_w(n) := C(n) \oplus \varepsilon(n) \qquad (12)$$

where $\varepsilon(n) \epsilon (0,1)$,(i.i.d), $P(\varepsilon(n)=1)=D_2$, $P(\varepsilon(n)=0)=1-D_2$

For $D_1 \geq {}^1/_2$ and $D_2 < {}^1/_2$, $C=1-H(D_2)$

For $D_2 > {}^1/_2$, C=0(evidently).

If $P(C(n)=1)=P(C(n)=0)={}^1/_2$, then WM system becomes *ideal stegosystem .*

Remark. The model considered above is not very interesting for practice because real CO (even presented in binary form) is not (i.i.d).

2. CO is Gaussian $N(0,\sigma^2)$ i.i.d sequence with SME (Euclidean) as distortion function.

It was proved in [19], that in this case we get for informed decoder

$$C = \begin{cases} \frac{1}{2}\log\left(1 + \frac{D_1}{\beta D_2}\right), & \text{if } D_2 < \sigma^2 + D_1 \\ 0, & \text{if } D_2 \geq \sigma^2 + D_1 \end{cases} \qquad (13)$$

where $\beta = \left(1 - \frac{D_2}{\sigma^2 + D_1}\right)^{-1}$,

Capacity above is achieved if the embedding follows the rule

$$C_w(n)=C(n)+Z(n), \ n=1,2 \ldots,N \qquad (14)$$

where $Z(n) \in N(0,D_1)$, $Z(n)$ – i.i.d

The best attack (with attacker’s point of view) is

$$P\left[C'_w(n), C_w(n)\right] = \begin{cases} N\left(\beta^{-1} C_w(n), \beta^{-1} D_2\right), & \text{if } D_2 < \sigma^2 + D_1 \\ \delta\left(C'_w(n)\right), & \text{if } D_2 \geq \sigma^2 + D_1 \end{cases} \qquad (15)$$

where $\delta(y) = \begin{cases} 1, & \text{if } y = 0 \\ 0, & \text{if } y \neq 0 \end{cases}$

Specification of the condition C=0 for $D_2 \geq \sigma^2 + D_1$

In fact , in this case an attacker can simply let $C'_w(n) \equiv 0$, independently on $C_w(n)$ because we get then

$$D_2 = E\left[\left(C'_w(n) - C_w(n)\right)^2\right] = E\left[C_w^2(n)\right] \leq \sigma^2 + D_1 \leq D_2$$

The main properties of WM system capacity for Gaussian CO.

1. C depends on CO, namely from its variance σ^2 through the parameter β. However, in a particular important case when $D_1 << \sigma^2$, $D_2 << \sigma^2$, we get by (13) that β≈1 and then

$$C = \frac{1}{2}\log\left(1 + \frac{D_1}{D_2}\right) \tag{16}$$

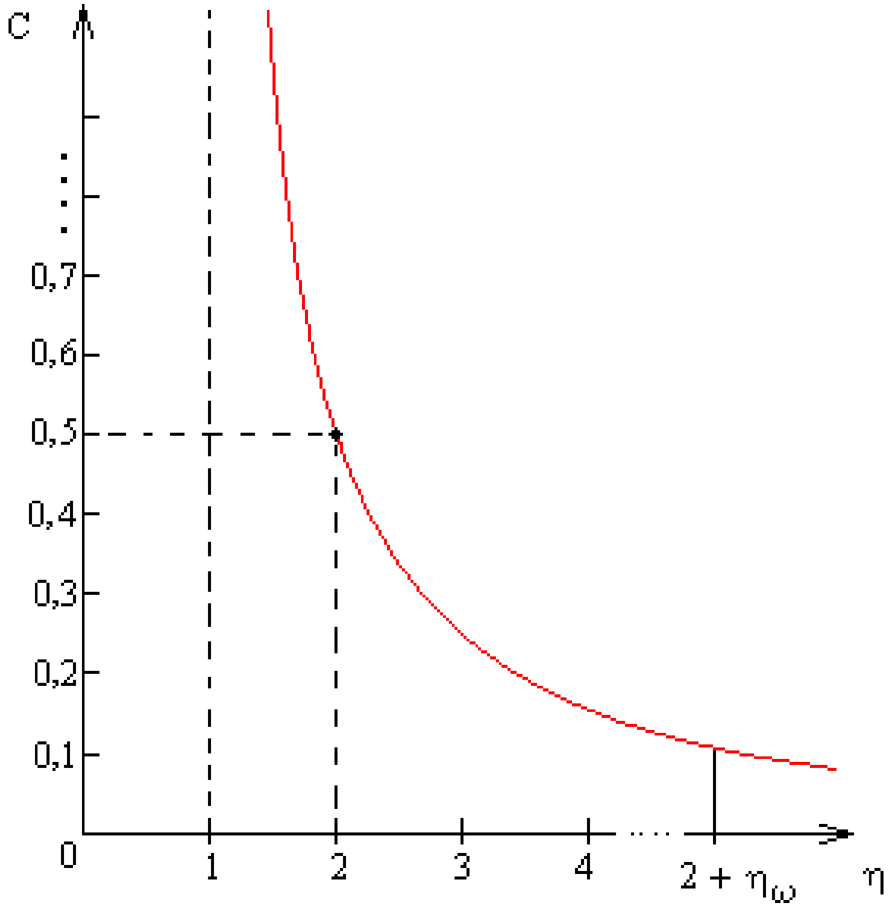
We can see from (16) that «C» does not depend on CO C(n).

2. Expressing(16) in terms of signal /noise ratio after embedding and after attack ($\eta_a = \sigma^2/(D_1+D_2)$), we get

$$C = \frac{1}{2}\log\left(\frac{\eta}{\eta - 1}\right) \tag{17}$$

where $\eta = {}^{\eta_w}/_{\eta_a}$

Graph of C versus η by(17)



If $\eta=1$, e.g. $\eta_w=\eta_a \Rightarrow D_2=0$ (no attack), then $C\to\infty$

If $\eta=2$, e.g. $\eta_w/\eta_a=2$, then $C=1/2$ (embedding of one bit into two samples)

If $\eta\to\infty \Rightarrow D_2\to\infty$, then $C\to0$

(In reality C=0 already under the condition that

$D_2\geq\sigma^2+D_1 \Leftrightarrow \eta>2+n_w$

Remark. The value of WM capacity is $C\neq\infty$ for any digital CO.

How it is important (or not) the notion of WM capacity?

1. It was proved in [19] that if CO C(n) is not Gaussian i.i.d sequence then under the very high requirements to CO quality (e.g. D1, D2→0) capacity C achieves C by (16) asymptotically, with optimal embedding by (14), and optimal attack by (15).

   However , there may be other attacks more effective for practical application. So, “estimation attack” or “geometric attacks” can be preferential than additive noise attack because they require more complex (and may be unknown) embedding and extraction algorithms.

2. It was proved in [19], that with the use of “blind” decoder capacity holds true by the same relation (13), as with the use of informed decoder.

   However in a real design of WM system with “blind “ decoder the embedding rate can be much less than with informed one.