Introduction to Watermarking II

Chang Ee-Chien
School Of Computing
changec@comp.nus.edu.sg

Illustrate the robustness and security requirements of watermarking.

- Robustness: Random Noise
  Scaler Costa Scheme
- Security: Smart Attacker
  Zero Knowledge Detector.

Remark on Image Representation

image transformation,
e.g wavelet,DCT,Fourier

s1,s2,s3,.......

watermarking encoder

s1',s2',s3',....

inverse image transformation,

Thus, we assume the image are represented by $s_1, s_2, \ldots, s_n$. Assume each s_i is drawn from i.i.d normal distribution.

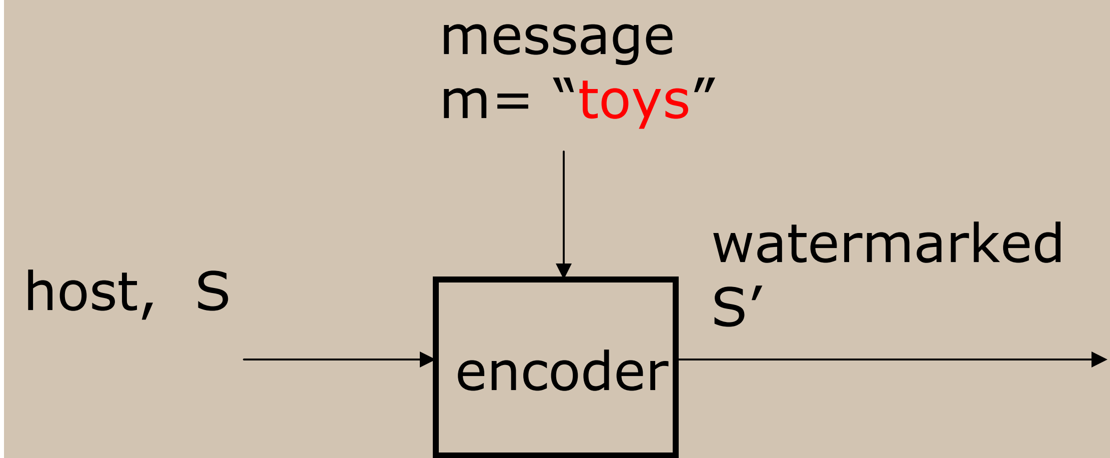
Watermarking as communication with side info.

message
m= “toys”

host, S
(cover Works,
original image,
work)

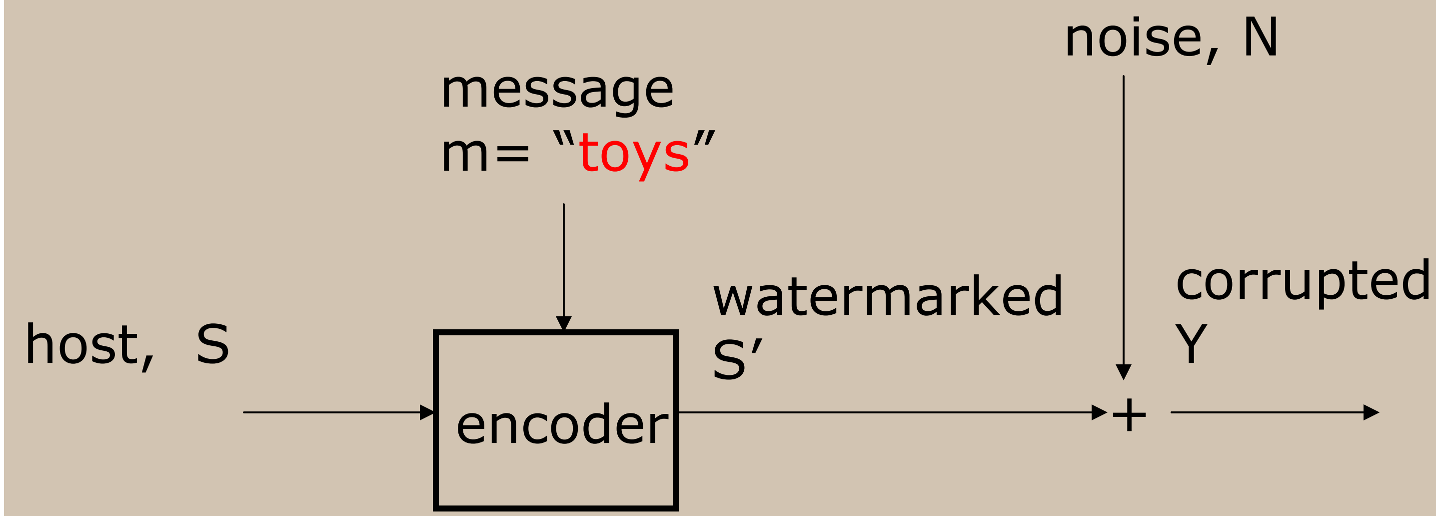
encoder

Watermarking as communication with side info.



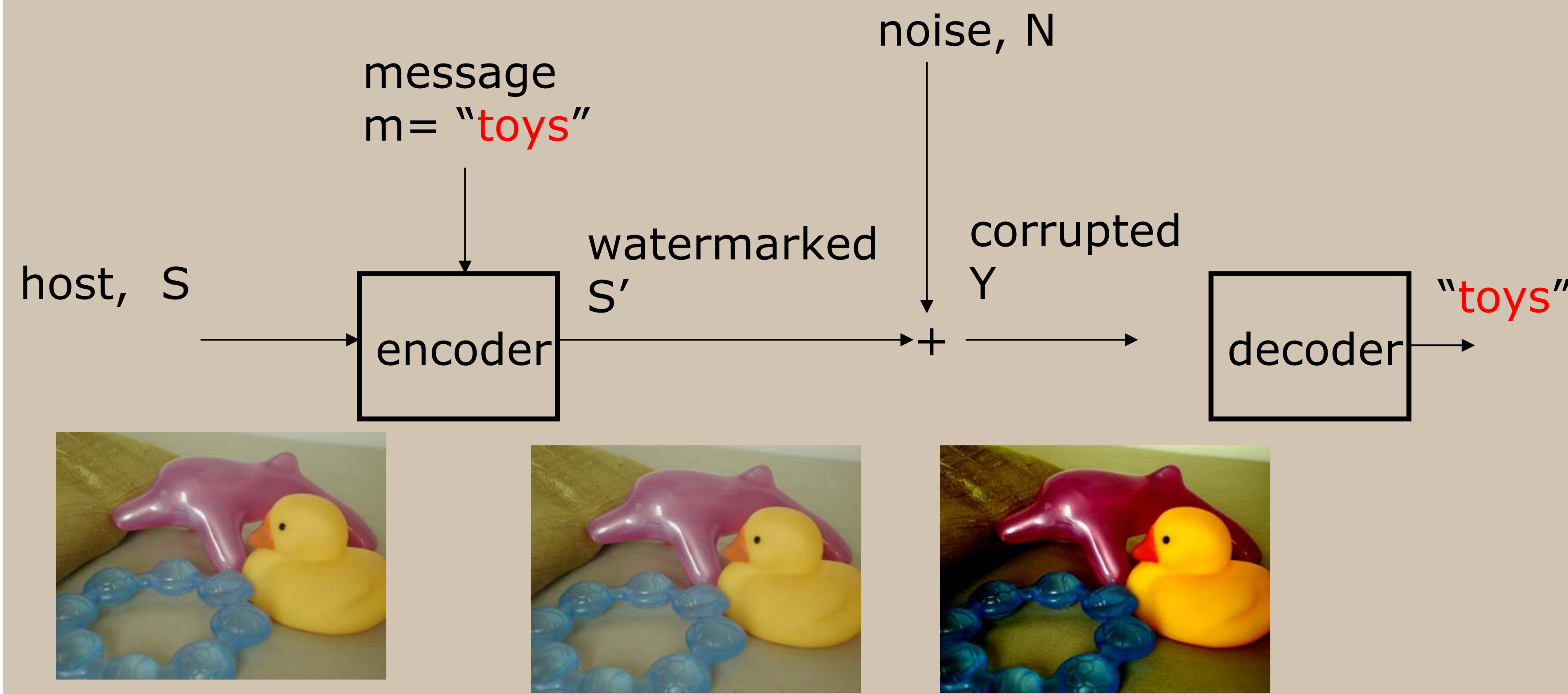
$$|| S' - S ||^2 < P$$

Watermarking as communication with side info.



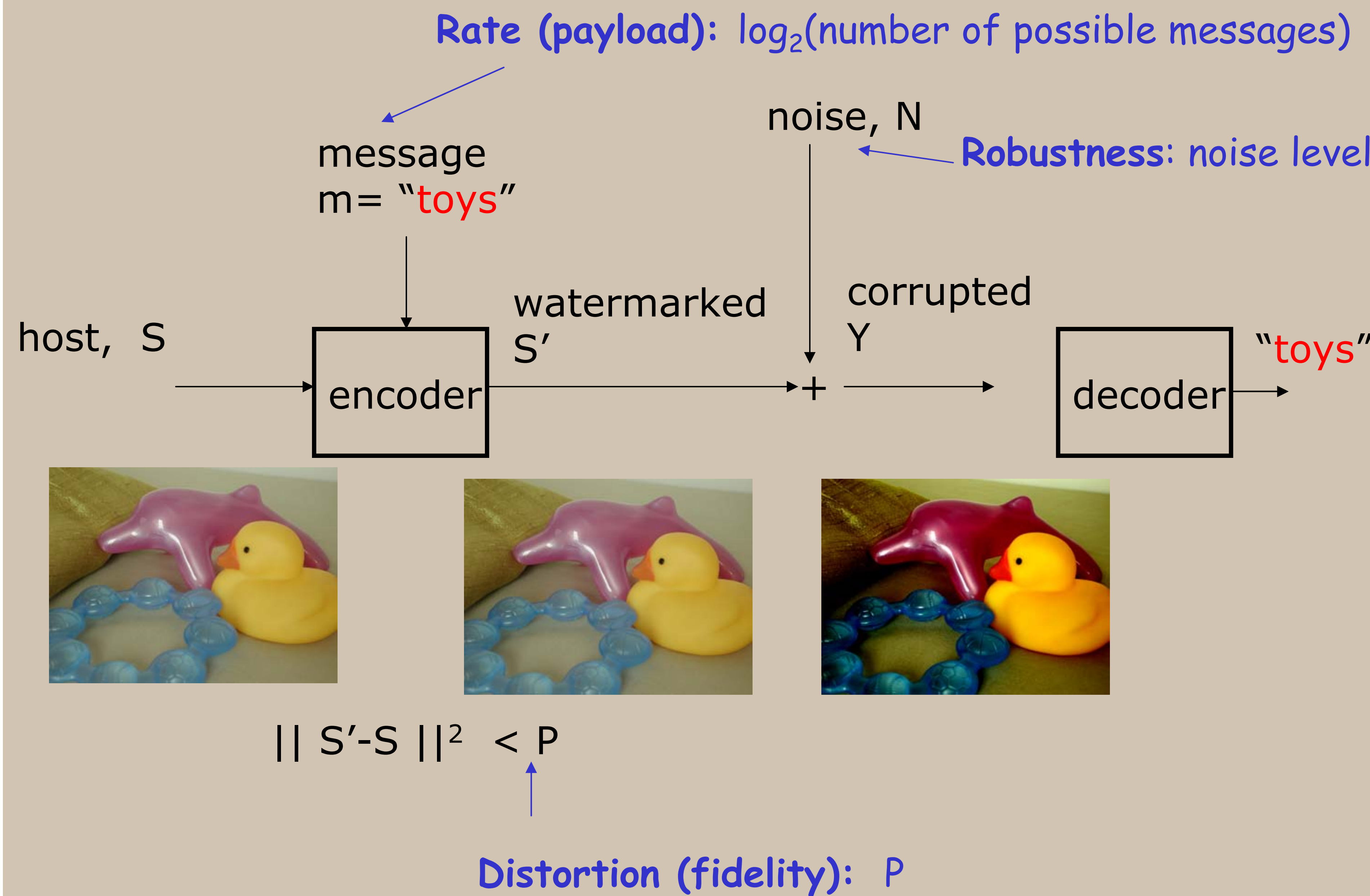
$$|| S'-S ||^2 < P$$

Watermarking as communication with side info.

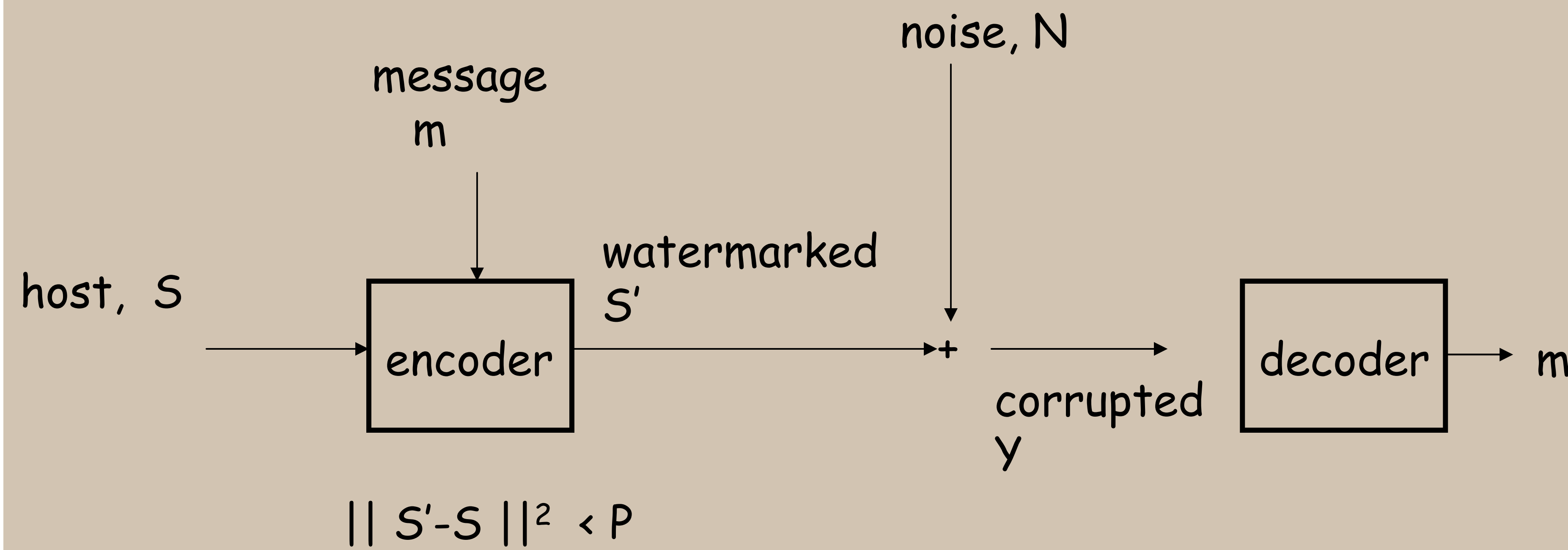


$$|| S' - S ||^2 < P$$

Performance Measure



Assumption



1. Host S is Gaussian,
i.e. $S = (S_1, S_2, S_3, \ldots, S_n)$, where S_i are n i.i.d Normal distributions.
2. Using 2-norm to measure distortion.
3. Noise is AWGN (additive White Gaussian Noise).

A Simple Scheme: Spread Spectrum method.

Set of messages $m \in \{ 0, 1 \}$
Let w0 and w1 be two pre-selected sequences,
and k a predefined constant. We can select w0 by randomly
choosing n coefficients from $\mathcal{N}(0,1)$

Encoding : Given the host $S=(s_1,s_2,s_3,\ldots,s_n)$, m, output the watermarked S'

$S' = S + k\,w0$ if m=0

$S' = S + k\,w1$ if m=1

Decoding : Given an image Y, output the embedded message

if (Y.w0) > (Y.w1) then output 0, else output 1.

Why is it robust to AGWN?

If $Y = (S + k\,w0) + \text{noise}$

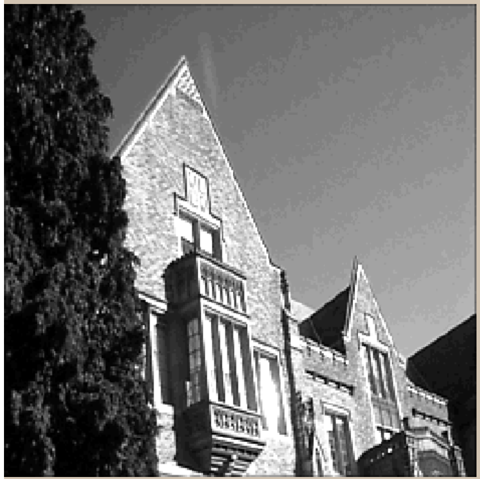
Then

$Y.w0 = S.w0 + k\,w0.w0 + \text{noise}.w0$ = small + large + small

$Y.w1 = S.w1 + k\,w0.w1 + \text{noise}.w1$ = small + small + small

Applying spread spectrum method in the pixel domain...

S



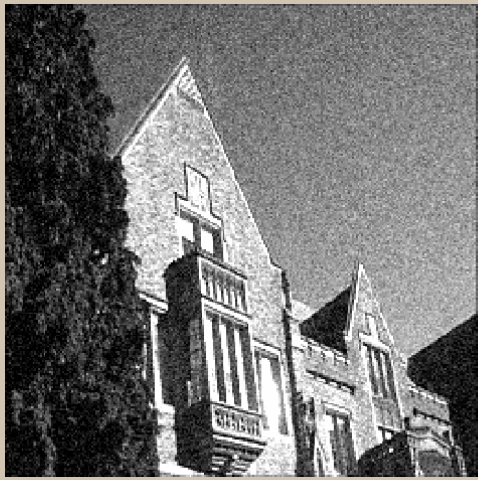
+ 0.1 *

k

w0

=

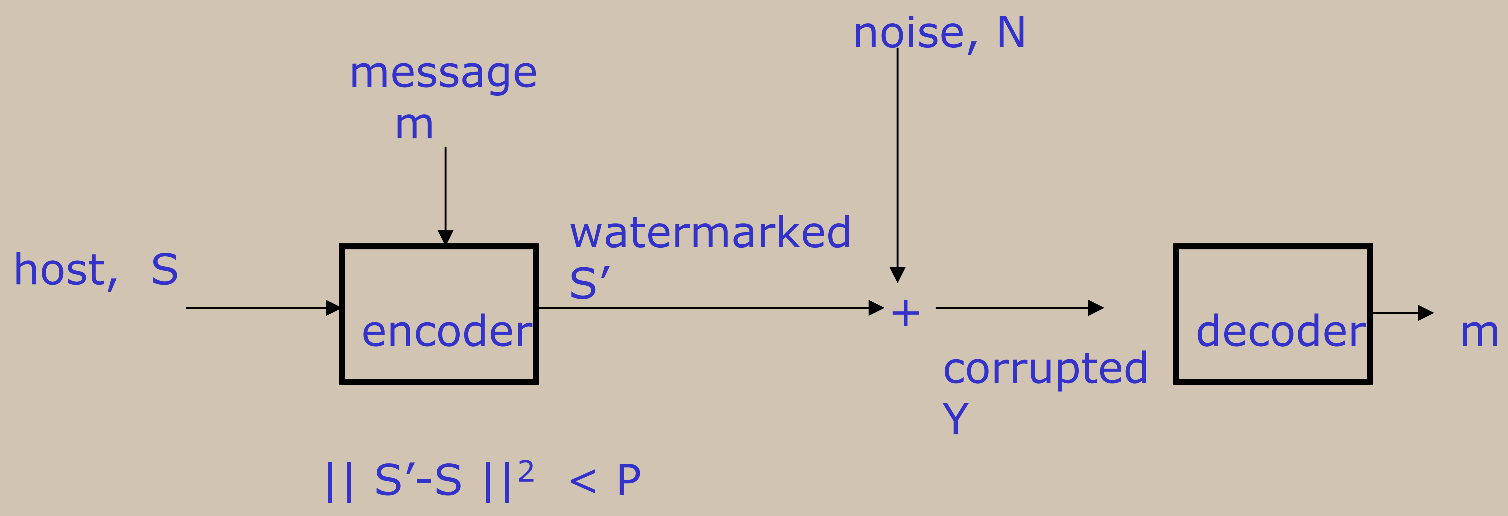
S'



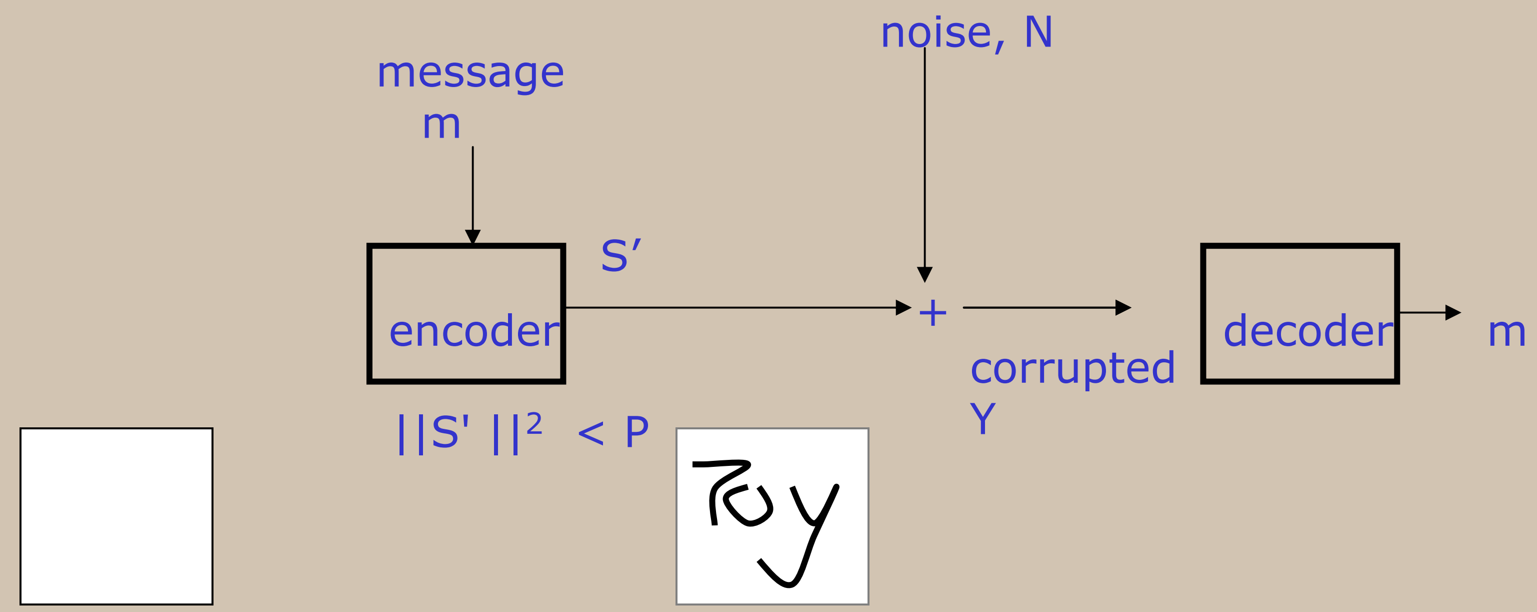
How much info can we hide?
A surprising result…

Capacity: Given a fixed ***distortion*** and ***noise level***, what is the max ***rate*** we can achieve?

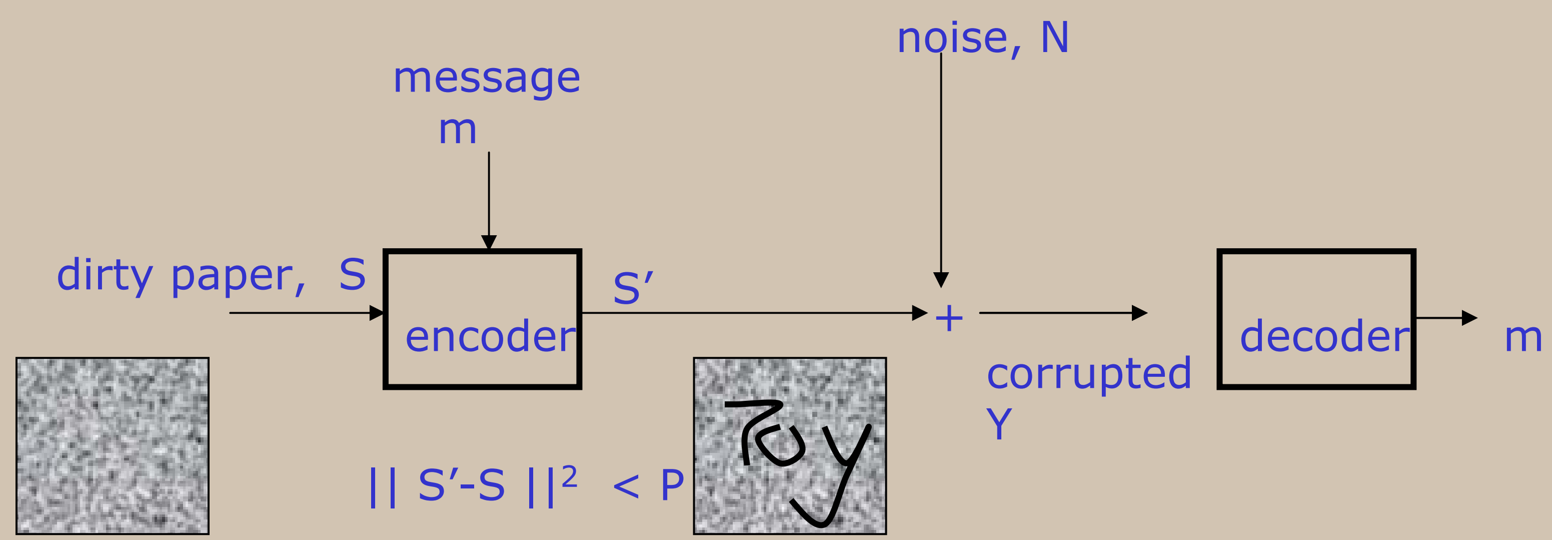
[1] **M.Costa**, Writing on Dirty Paper, *IEEE Trans. on Information theory,* 1983



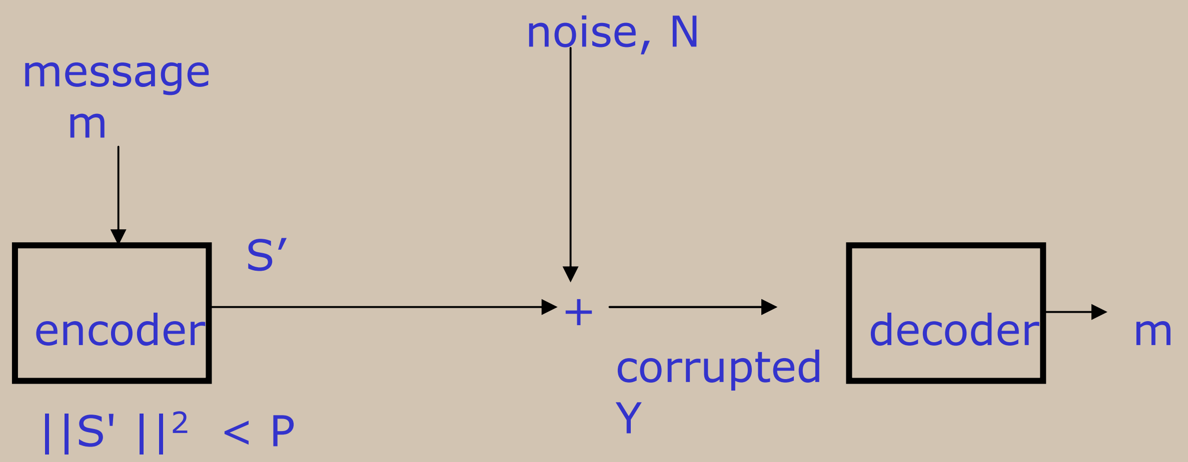
Writing on Clean Paper:



Writing on Dirty Paper:

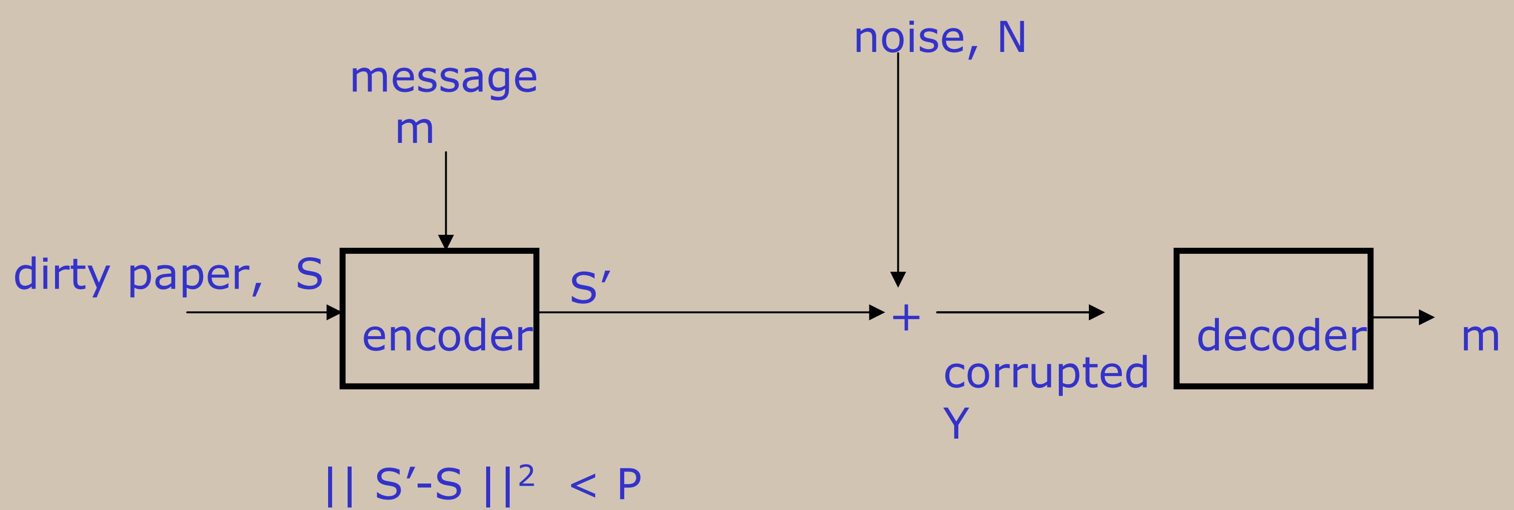


Writing on Clean Paper:



Shannon's 2nd Theorem: capacity = $(n/2)$ log $(1+ P/N)$.

Writing on Dirty Paper:



capacity = $(n/2)$ log $(1+ P/N)$.

Costa constructive proof is based on random code. Not practical.

Effort to realize "dirty code":

[2] **J. Chou, S.S. Pradhan and K. Ramchandran**.
On the Duality Between Distributed Source Coding and Data Hiding
33 Ailomar Conference on Signal, Sys. & Comp. 1999

[3] **B. Chen and G.W. Wornell**.
Quantization Index Modulation: A class of provable good methods for digitial watermarking and information embedding
IEEE Trans. Information Theory, 2001

[4] **M. Satring, J. Oostveen and T. Kalker**
Optimal Distortion Compensation for Quantization Watermarking
IEEE ICIP 2003.

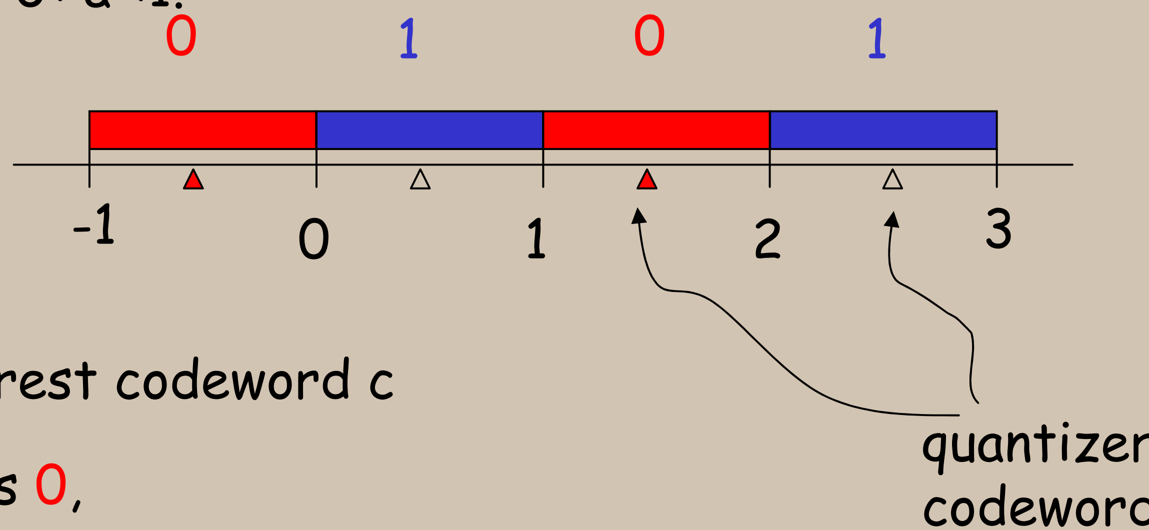
Scaler Costa Scheme: illustration of dirtycode

The scheme requires a error correcting code for binary data, and a scaler quantizer. The codewords in the quantizer is alternative labeled as 0 and 1.

A parameter is a constant $0 < \alpha < 1$.

Decoding

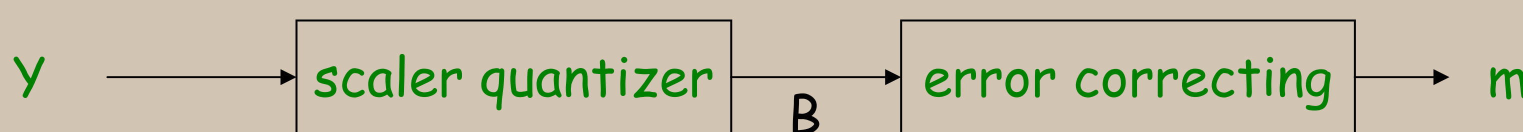
Given $Y = (y_1, y_2, \ldots, y_n)$.



For each y_i, find the nearest codeword c

let b_i = 0 if c is labeled as 0,
1 otherwise

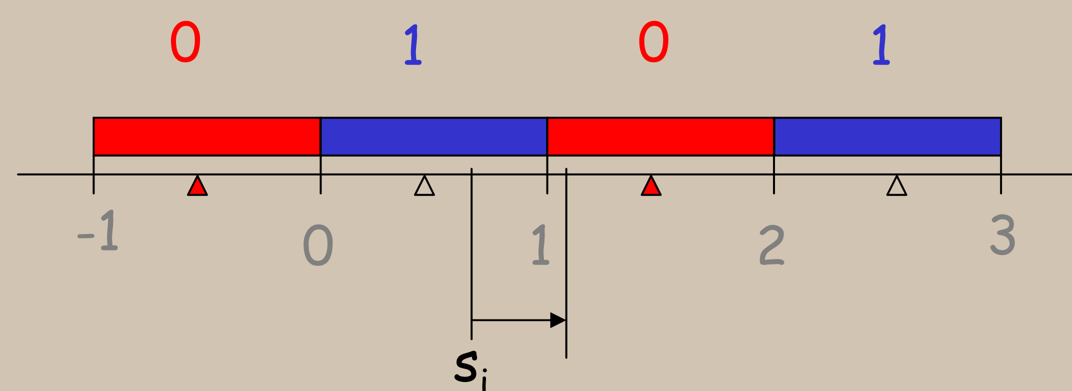
Let $B = (b_1, b_2, \ldots, b_n)$, Using the error correcting code, determine the message encoded in B.



Encoding: Given a message *m* and a host *S*.

1. Encode *m* using the error correcting code. WLOG, let the encoded bits be a sequence of *n* zeros. 0 0 0 0 … 0

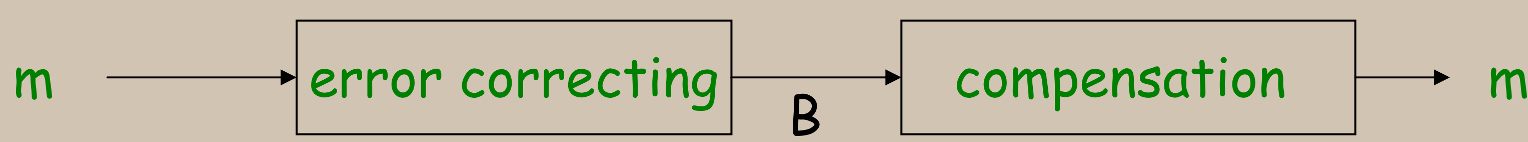
2. For each s_i, find the nearest codeword c_0 labeled as 0.



Let $s'_i = s_i + \alpha (c_0 - si)$

where $0 < \alpha < 1$ is the predefined constant (*distortion compensation*)

if α is chosen as 1, then the watermarked s'_i is just the codeword c_0



Why is the distortion small? Why is it robust? Why high rate?

Let conduct a simple Matlab experiment.

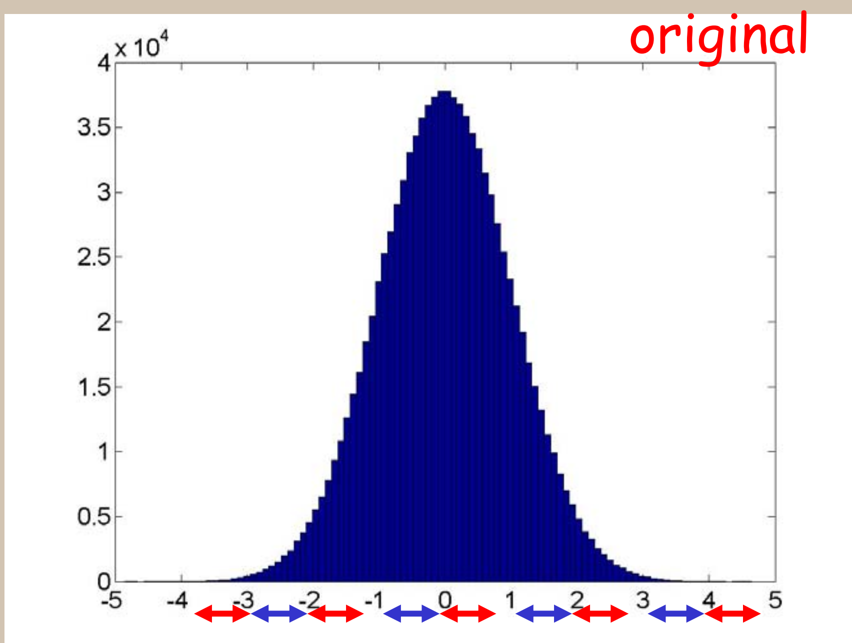
$n = 1\,000\,000$, $\alpha = 0.2$, $s_i \sim \mathcal{N}(0,1)$

Let's assume that the error correcting code can withstand probability of error $p = 0.495$ (i.e. a particular bit will "flip" with probability 0.495).

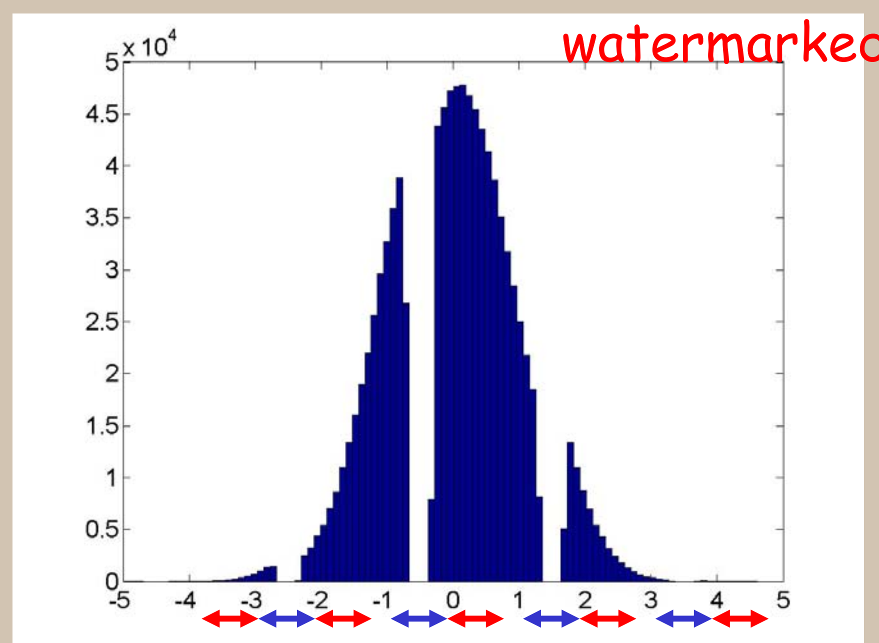
Thus, theoretically, it can encode

$$n\,(1 - \mathcal{H}(p)) \approx 72 \text{ bits of messages}$$

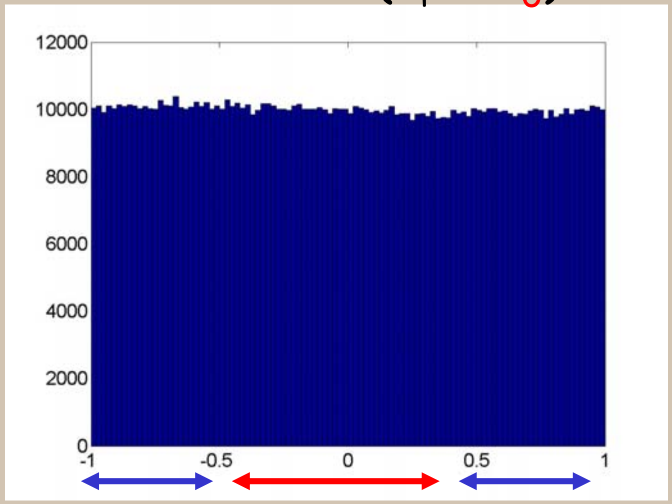
Distribution of s_i



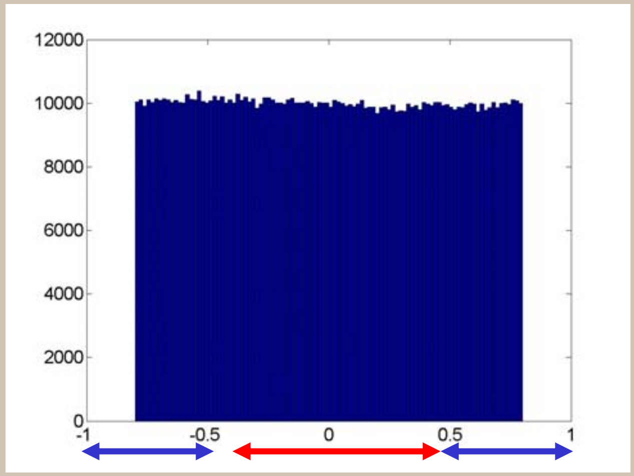
Distribution of s'_i



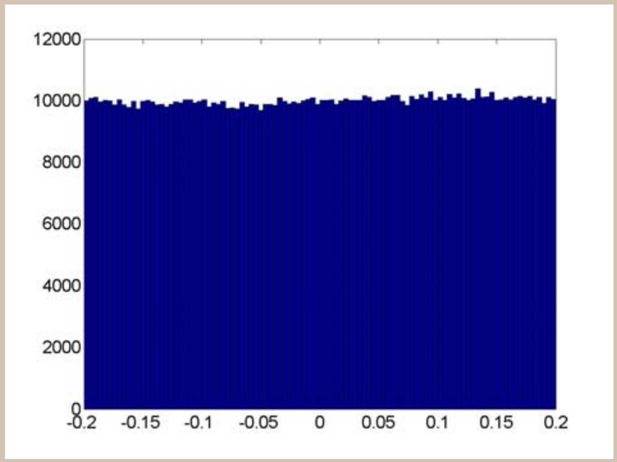
Distribution of $(s_i - c_0)$



Distribution of $(s'_i - c_0)$



Distribution of $(s'_i - s_i)$



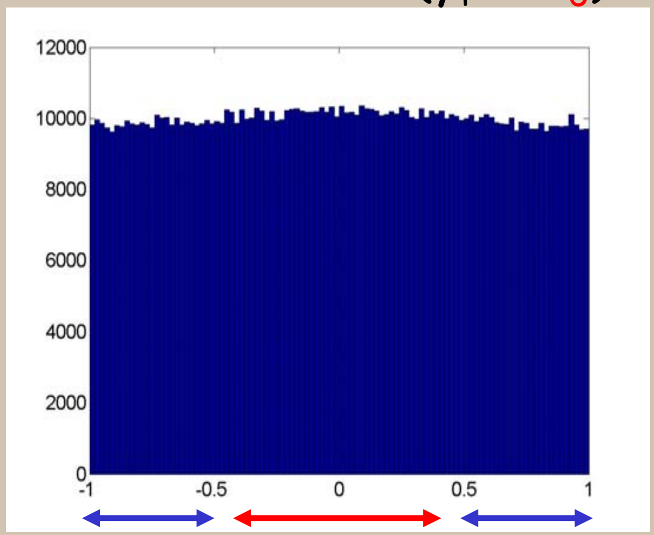
average distortion $P = || S - S' ||^2 \approx n\,(0.013)$

Now, add noise $N=(z_1,z_2,\ldots,z_n)$ to the watermarked S'.

$Y= S' + N$, where each $z_i \sim \mathcal{N}(0, 0.8)$. Thus the noise level is $0.8n$

The distribution of $(y_i - c_0)$ {recall c_0 is the nearest codeword labeled as 0}, can be approximated by the distribution of $(s_i - c)$ convolves with the noise p.d.f.

Distribution of $(y_i - c_0)$



Let's perform the watermark decoding on Y. Let B be the binary sequence obtained.
With high probability, the number of $b_i = 1$ is < 49.5%. Since the error correction code can withstand error of 0.495, thus, the message *m* can be recovered.

Note that the theoretical capacity is
$n/2 \log (1+P/N) > 10000$

Relationship with the Dirtycode

Sketch of the construction by Costa:

Preprocessing:

- Randomly choose many codewords.
- Each codeword is randomly labeled as a message (Because the number of codewords > number of messages, multiple codeword will be labeled as a same message).
- Decide a constant $0<\alpha<1$.

Encoding of *m* into S

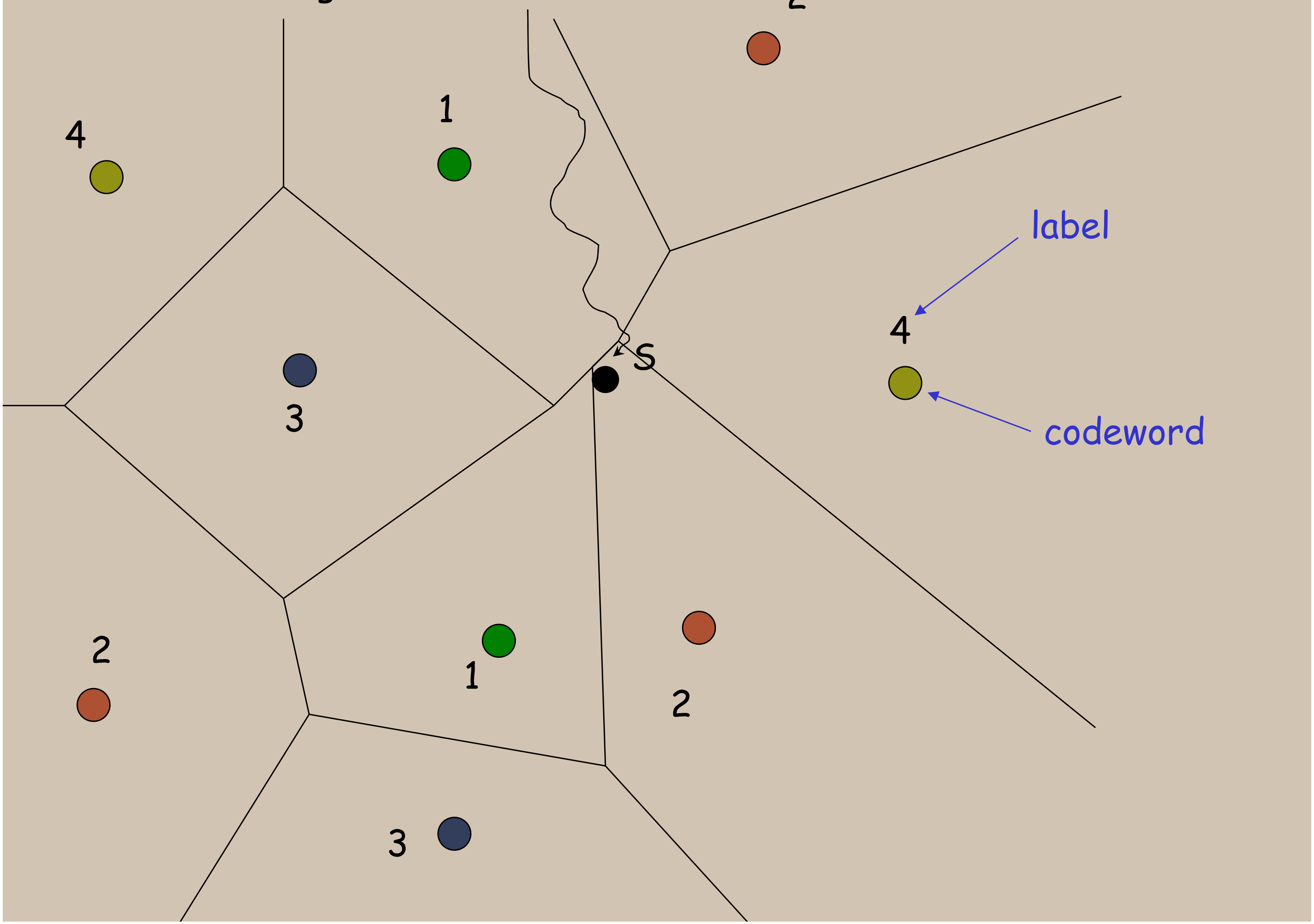
- Find the "nearest" codeword c labeled as *m*.
- Compute the watermarked $S' = S + \alpha\,(S-c)$.

Decoding

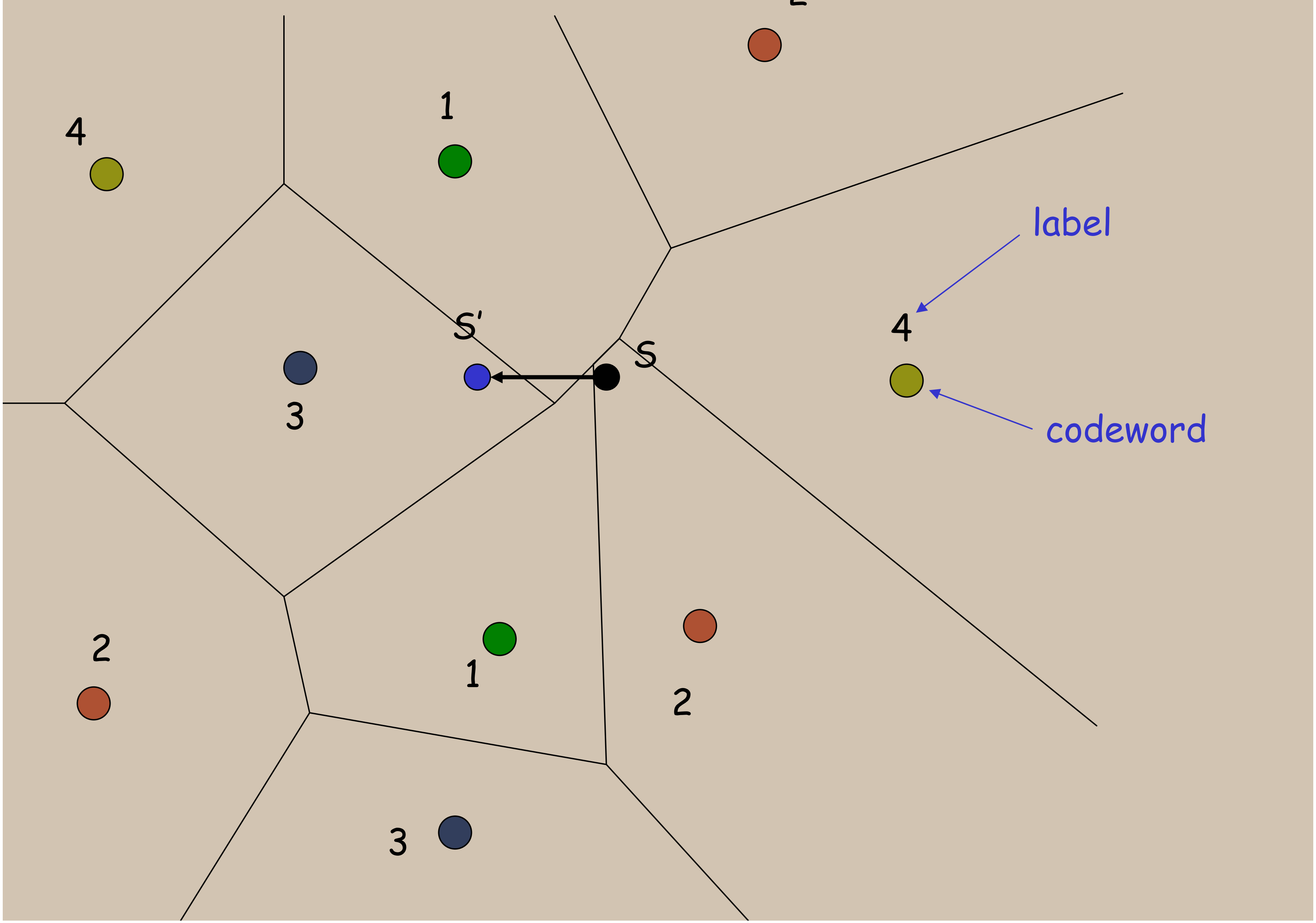
- Find the nearest codeword c. Output its label.

The random code book can be viewed as a 2-layer quantizer. In the scaler Costa scheme, the combination of scaler quantizerand error correcting code can be view as a high dimensional quantizer.

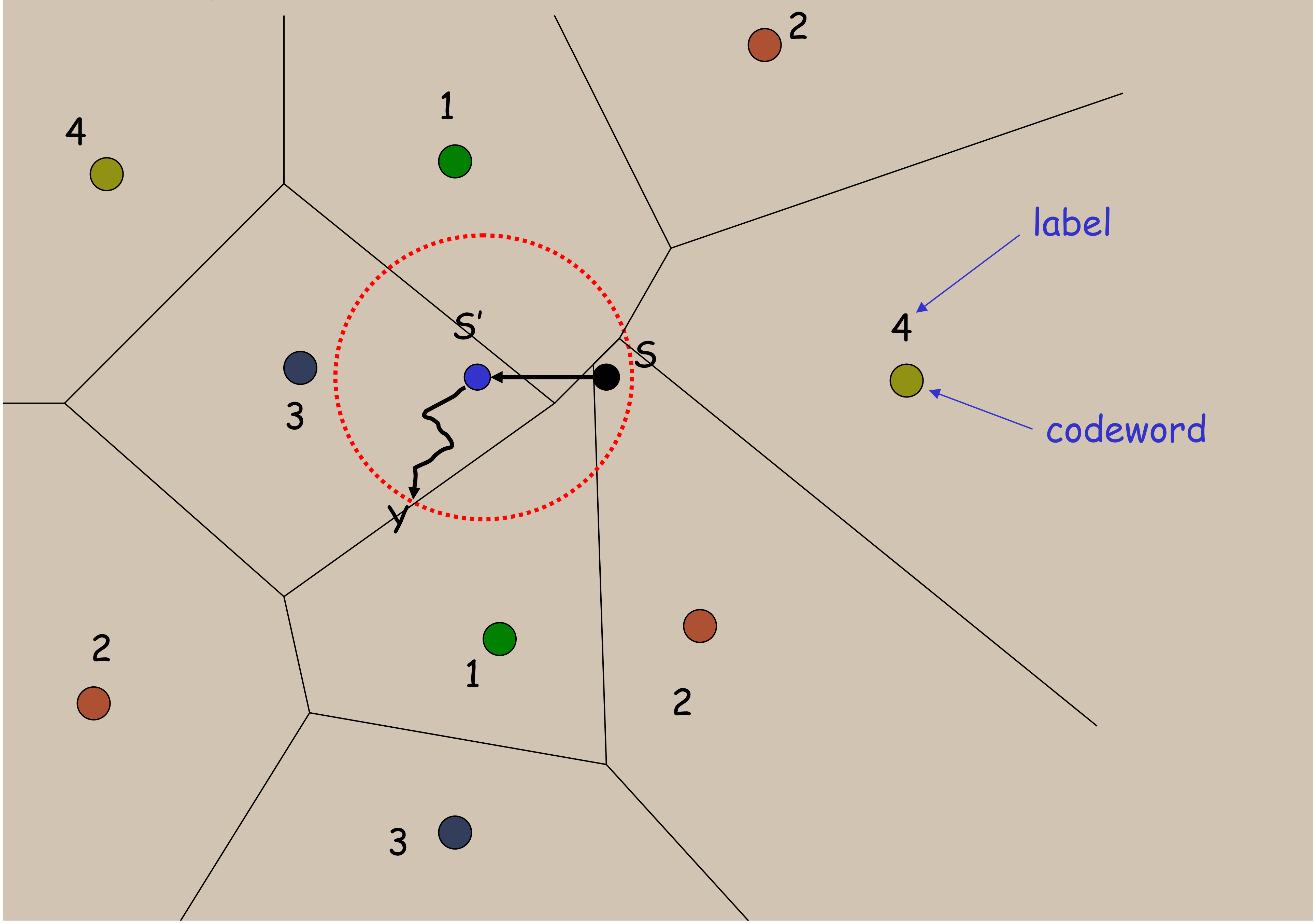
To embed message “3” into S



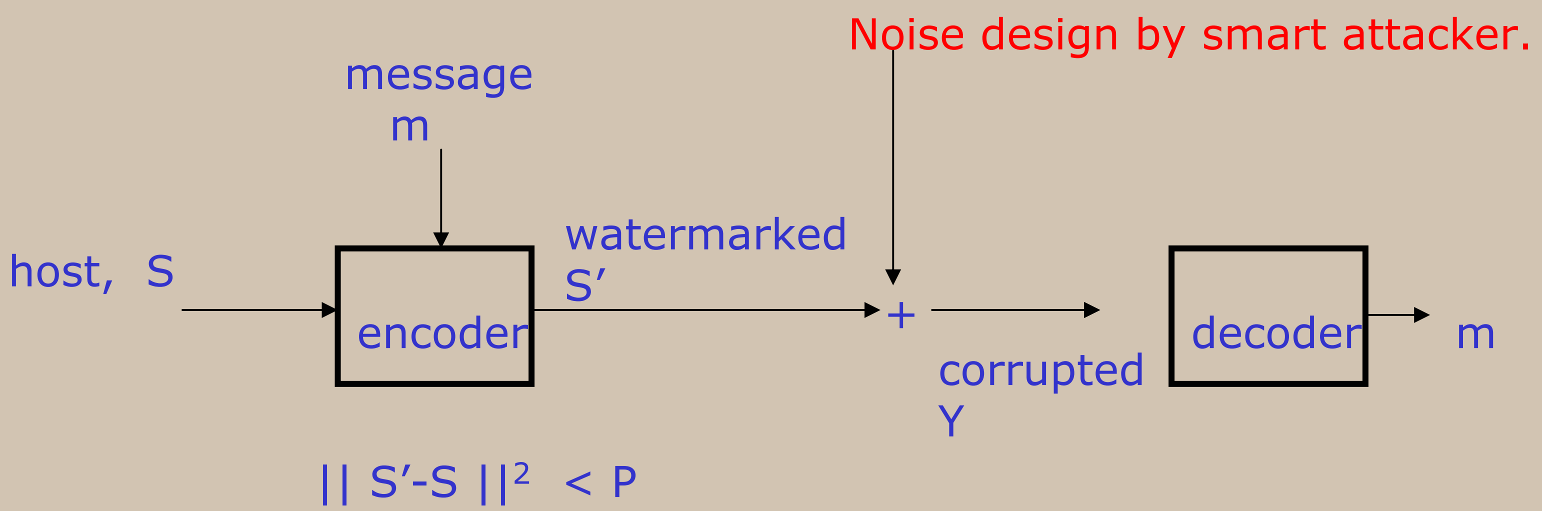
The watermarked S'



The corrupted Y. With high chances, Y is decoded as "3"



Security: smart attacker

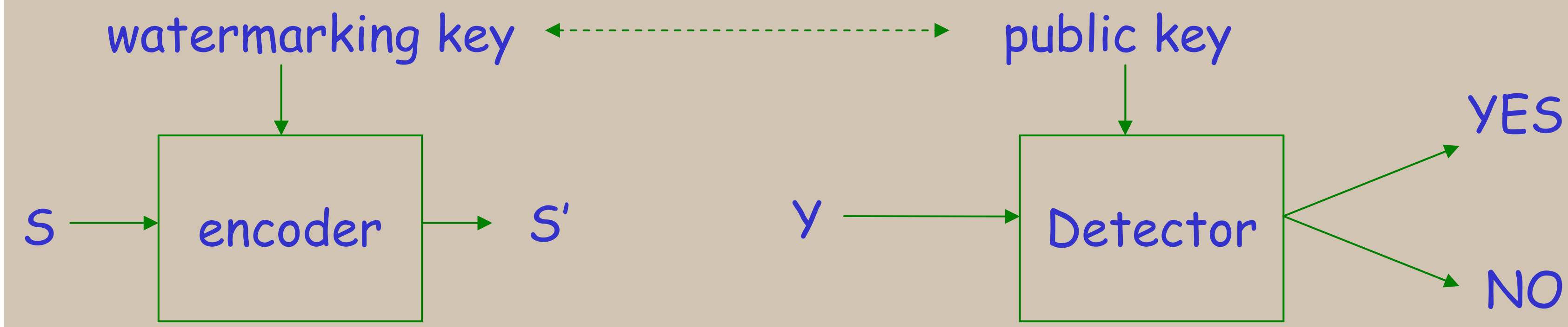


Here, we look at one aspect of security: Public watermarking scheme.

Public watermarking for copyright protection

Models of public watermarking

1) Public key watermarking:
The detector algorithm is known by everyone.



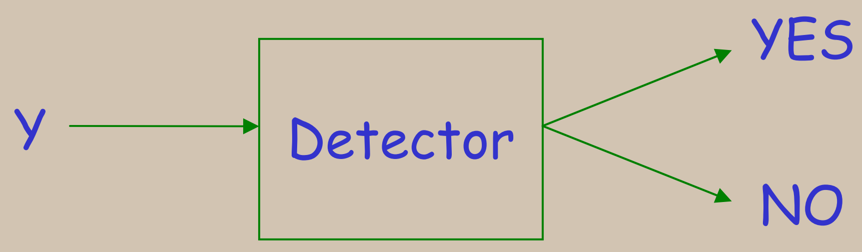
An attacker, given a watermarked Y (i.e. Detector(Y) gives YES), want to find a X, s.t. X is not watermarked, and X is close to Y.

A scheme is secure if any attacker, given Y, will not able to find such X, even if the attacker know the public key.

Main question: *does such a secure scheme exist?*
Insofar, there is not scheme that is satisfactory secure.
Example of attempt: asymmetric watermarking scheme.

2) Detector as black box:

The detector is a black box. To check whether an image Y is watermarked, the verifier sends it to the black box. The black box will return YES or NO. The verifier always trust the black box.



An attacker, given a watermarked Y, want to find a non-watermarked X, s.t. X is closes to Y, using as little calls to the black box as possible. If he succeed, he has "inverted" the watermarking process.

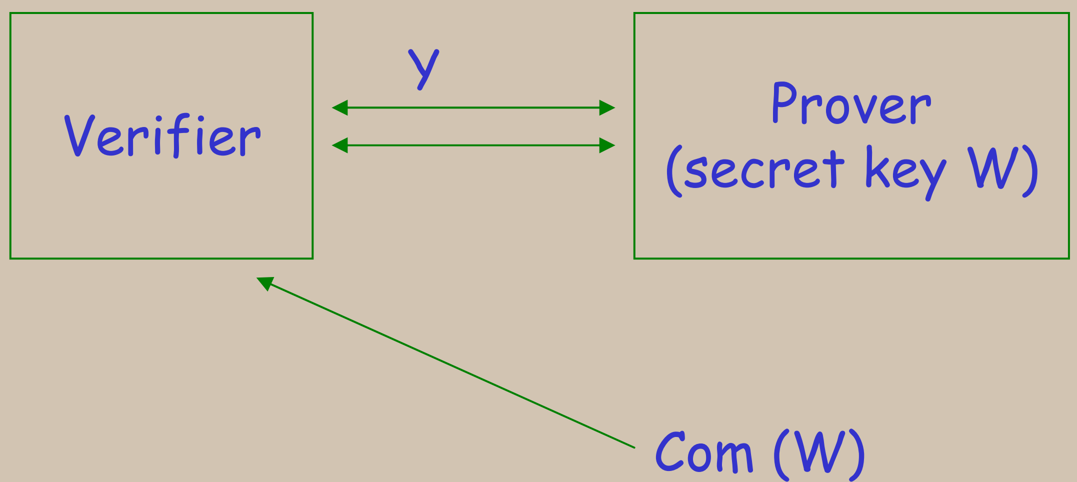
Main question: *Does a non-invertible scheme exist?*

Probably the answer is no.

3) Zero knowledge detector.

Same as the setting of black box. However, the verifier don't trust the black box. The issue here is, whether the *Prover* can convince the verifier that the image Y is indeed watermarked, and yet the verifier gain no additional information, except the fact that Y is watermarked.

Initially, the Prover publishes a commitment com(W) of the secret W. We can viewed com(W) as a encrypted W.



A Zero-knowledge Detector

Let's use the spread spectrum method as the underlying watermarking scheme...

Secret key: $W = (w_1, w_2, \ldots, w_n)$, The com(W) is published.

Encoding: Given S, output the watermarked $S' = S + W$

Detector: Given Y, Output YES if $Y \cdot W > T$,
where T is a predefined constant.

The job of the zero knowledge detector is to convince the verifier that, indeed the inner product of $Y \cdot w$ is $> T$, and the published value is a commitment of w.

[5] **A. Adelsback and A. Sadeghi.**
Zero-knowledge watermark detection
4th Int. Workshop on Info. Hiding, 2000

Note that inner product on vectors can be carried out in a series of scaler addition and multiplication. There is a known commitment scheme that

1) given C_1, C_2, which is a commitment of a, b respectively, $C_1 * C_2$ (mod N) is a commitment of **a+b (mod N)**.

Furthermore, it allows a prover to show the following using zero knowledge interactive proof.

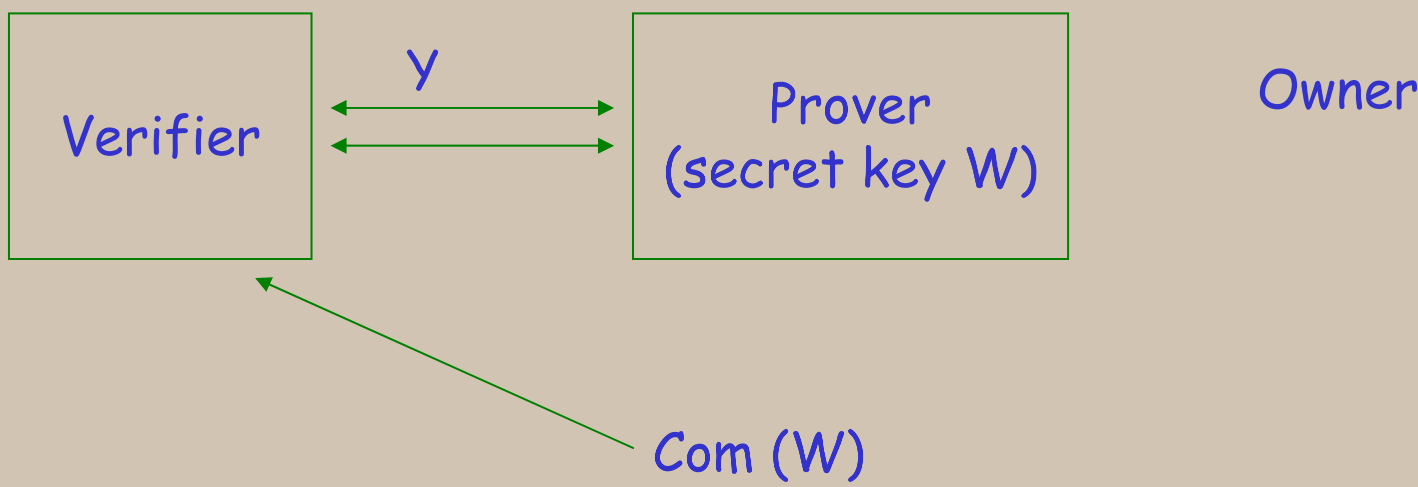
2) given C_1, C_2, C_3, prove that C_1, C_2, C_3 are commitment of some a,b,c respectively, s.t. **a = b * c (mod N)**

3) given C_1, T, prove that C_1 is a commitment of some value a, s.t **a>T**.

With the above 3 tools, it is easy to construct a zero-knowledge detector.

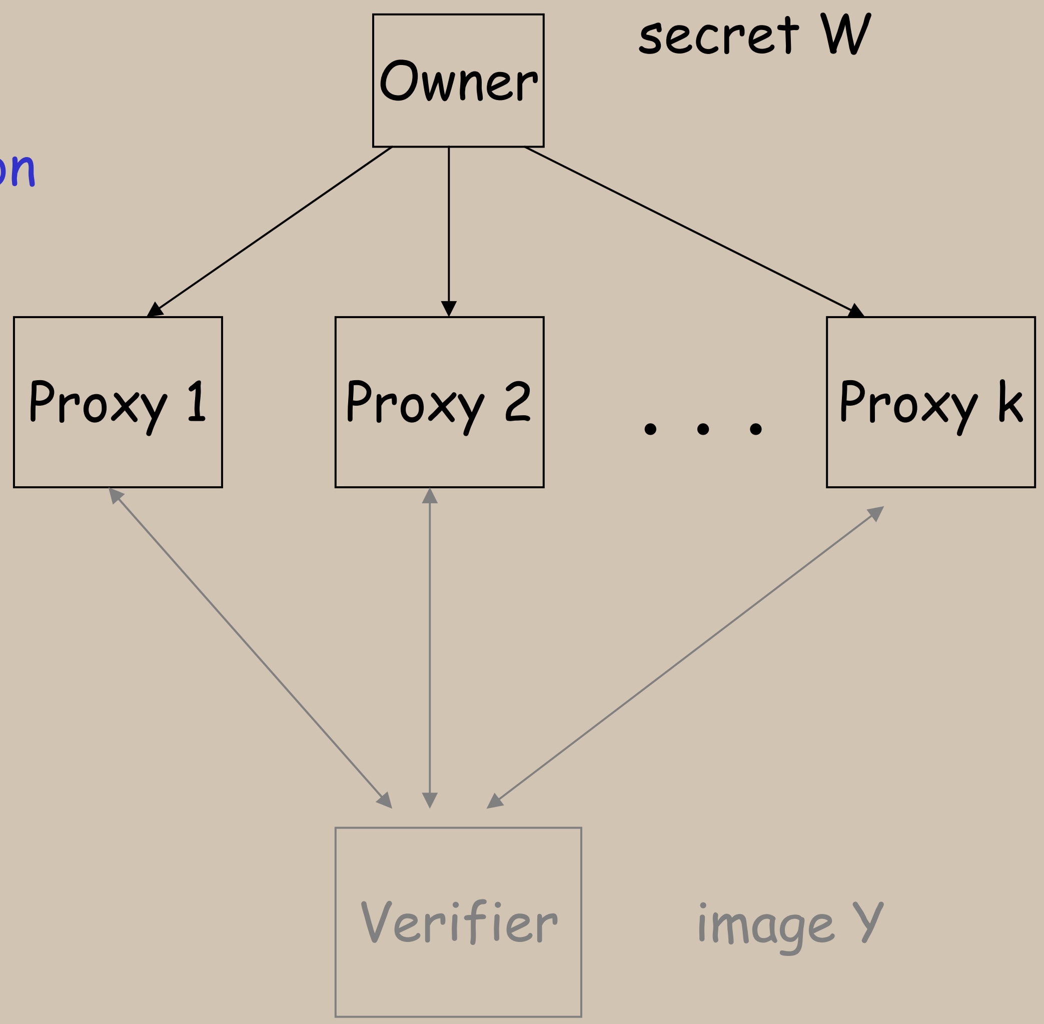
4) Detection by proxies

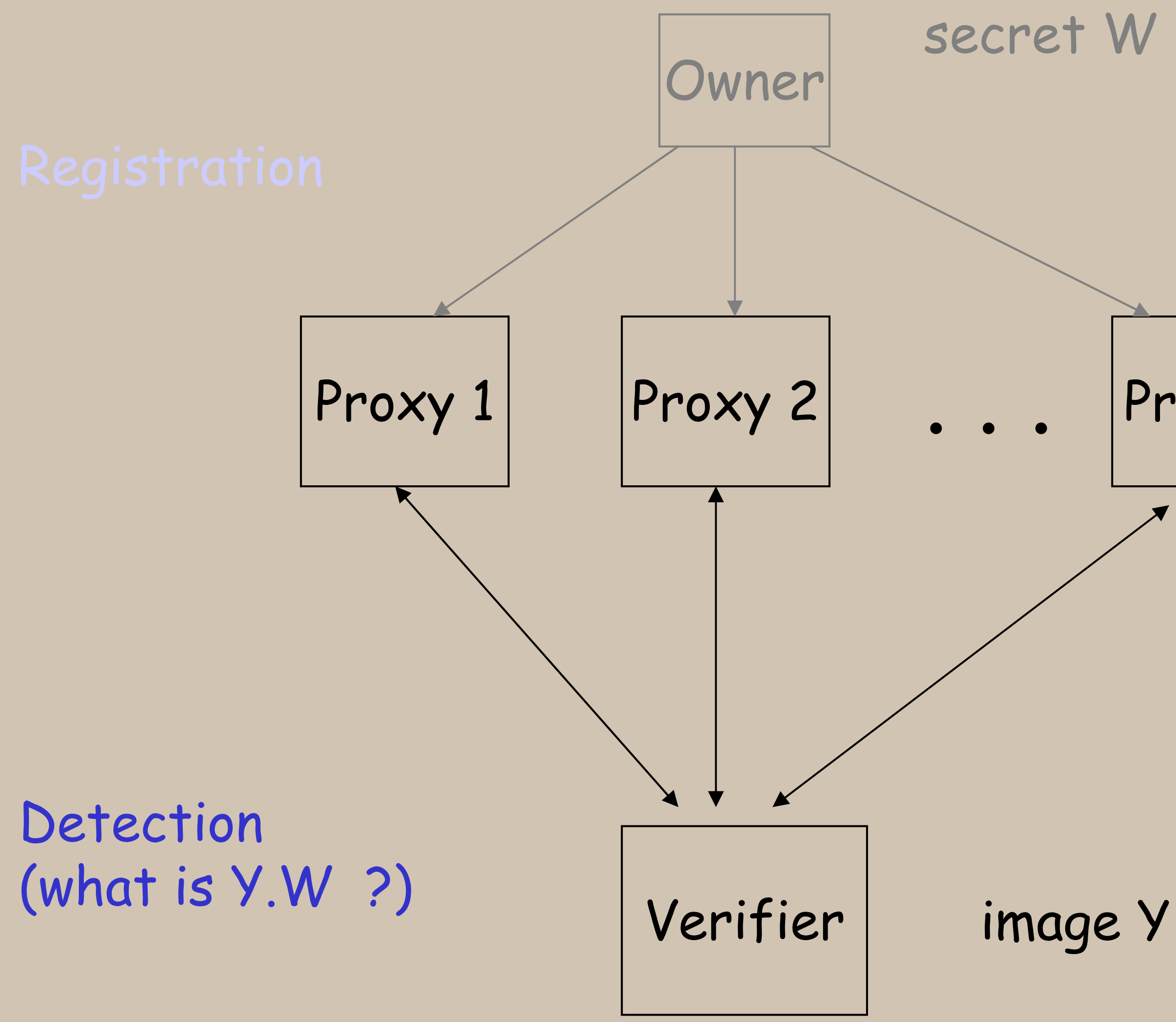
Under the zero-knowledge detector model, the owner has to give his secret key W to the prover. So he has to trust the prover. Furthermore, the Verifier has to get the com(W) from a yellow page, so, he has to trust the yellow page.



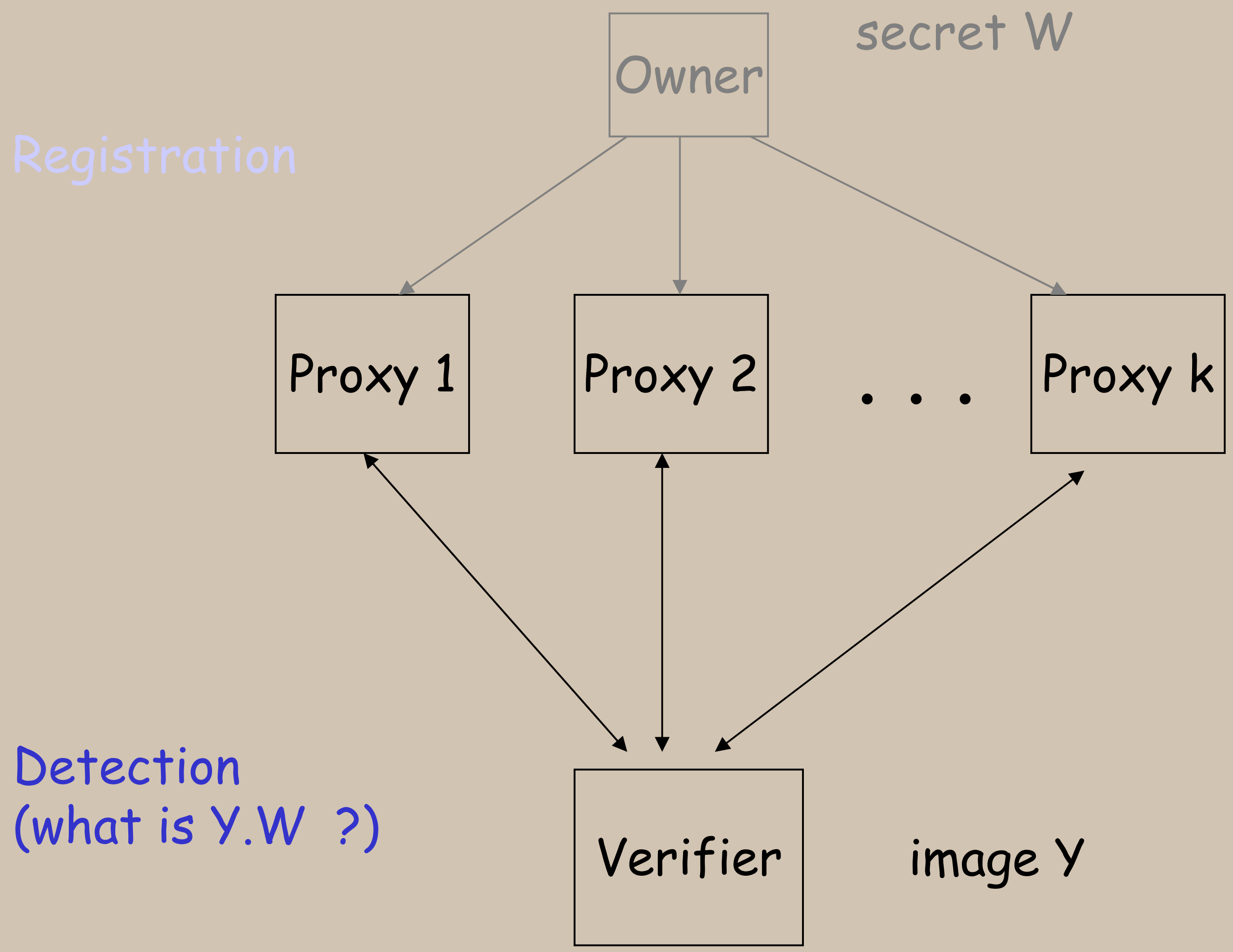
Using multiple proxies, we can have a scheme where no individual can be trusted. The security is achieved if majority is honest.

Registration





secret W
Owner
Registration
Proxy 1
Proxy 2
. . .
Proxy k
Detection
(what is Y.W ?)
Verifier
image Y



[6] **Q.Li and E.C. Chang,**
Public Watermark Detection using Multiple Proxies and Secret Sharing
2nd Int. Workshop on Digital Watermarking, 2003

References:

Books:

[7] **I.J.Cox, M.L.Miller and J.Bloom**, *Digital Watermarking,* Morgan Kaufmann 2002

[8] **S.Katzenbeisser and F.A.P. Petitcolas**, editors *Information Hiding: techniques for steganography and digital watermarking,* 2000

Webs:

[9] http://www.watermarkingworld.org/