Video Streaming Authentication over Satellite Networks

Gabriele Oligeri

Wireless Networks Laboratory
ISTI – CNR
Pisa, Italy

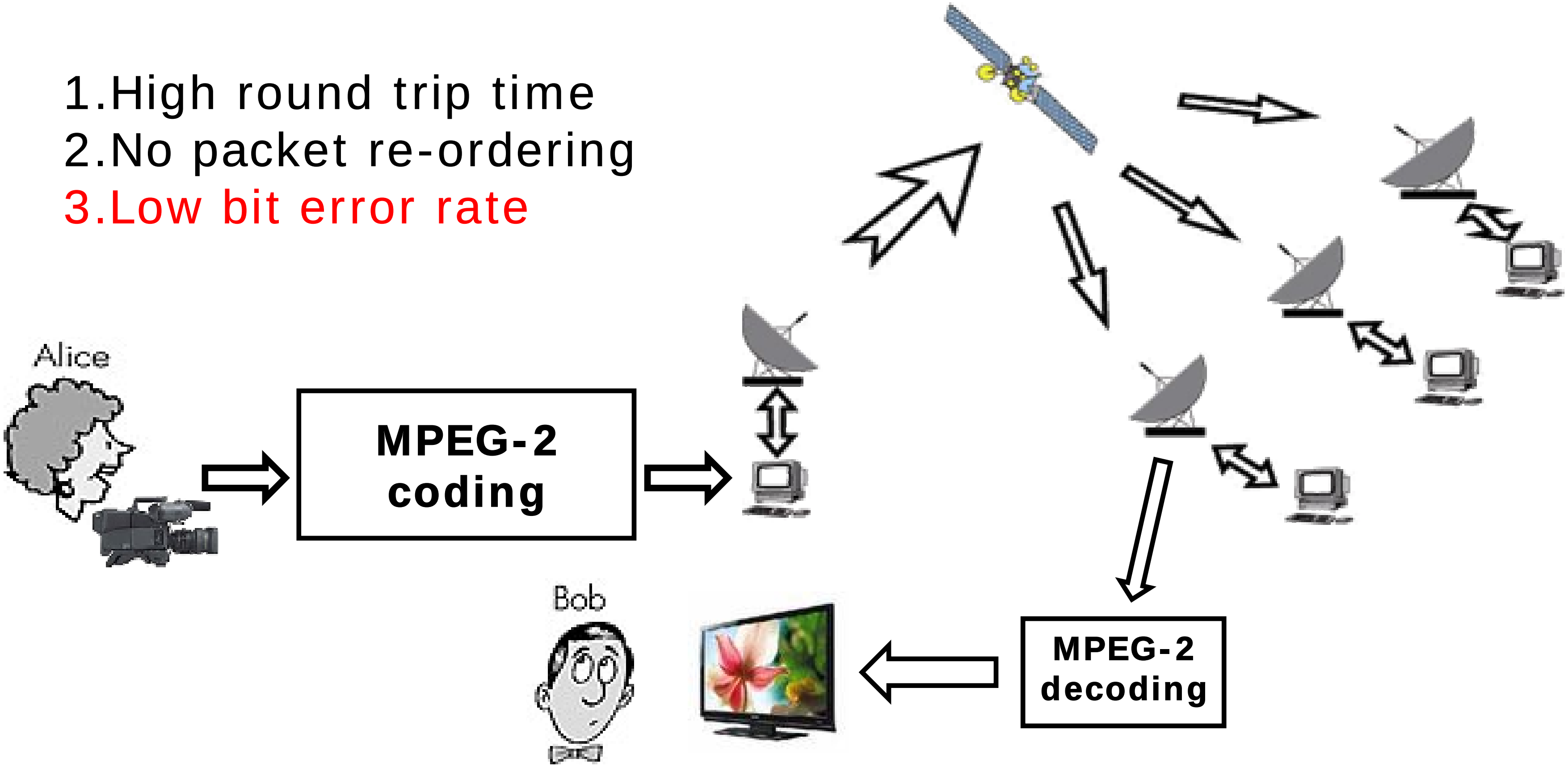
gabriele.oligeri@isti.cnr.it

ISTITUTO DI SCIENZA E TECNOLOGIE
DELL'INFORMAZIONE "A. FAEDO"

System model

Peculiar characteristics of multicast satellite networks :

1. High round trip time
2. No packet re-ordering
3. Low bit error rate



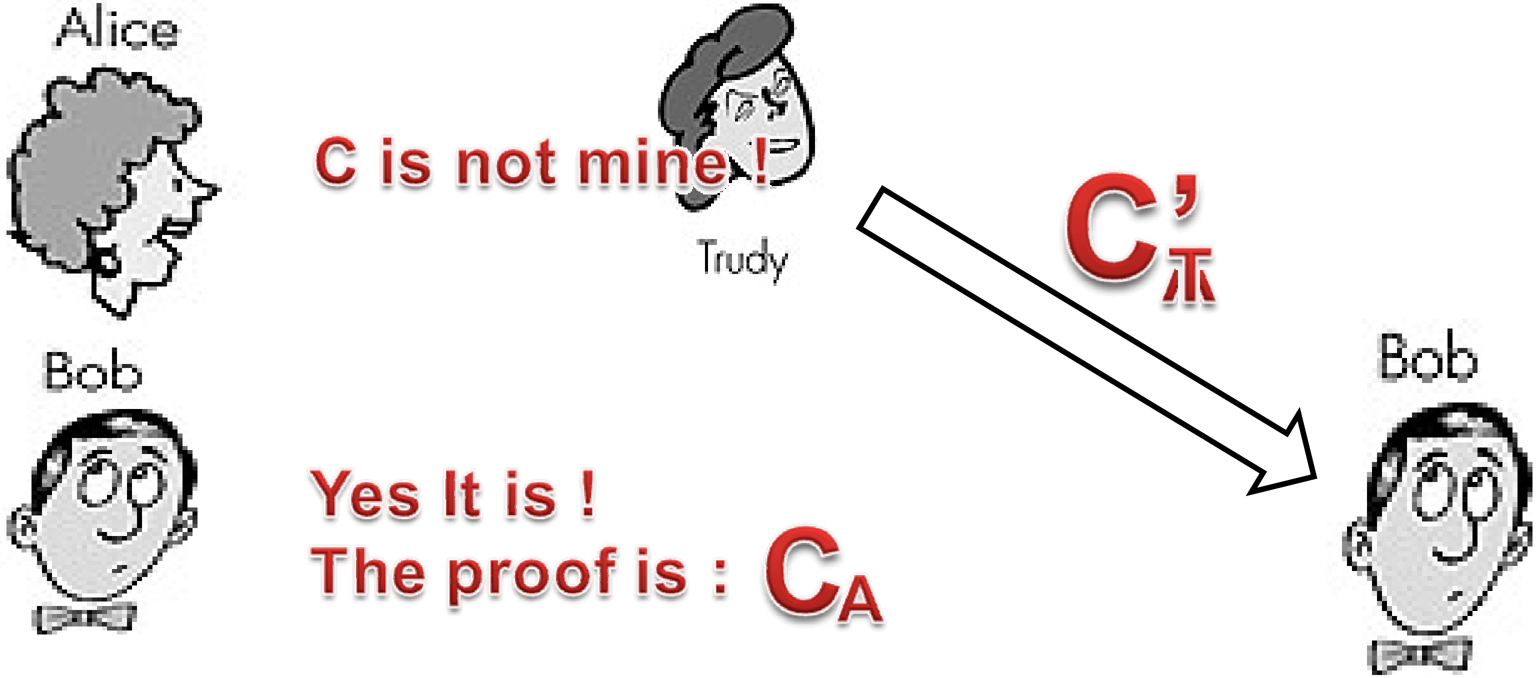
MPEG-2 coding

MPEG-2 decoding

Authentication properties

Different authentication features / levels:

1. Change of content (Integrity verification)
2. Change of ownership (Source verification)
2. Non repudiation



C is not mine !

Yes It is !
The proof is : C_A

Digital signature

Digital signature

=

RSA is an algorithm for public-key cryptography.

+

HASH is a transformation that takes an input and returns a fixed-size string

HASH

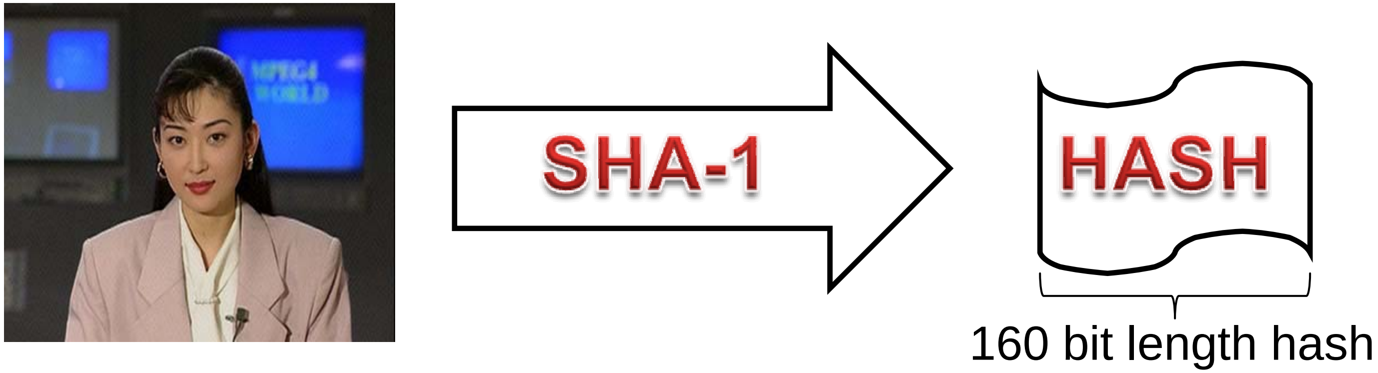
h : X → Y is a hash function if :

1. H can be applied to a block of data at any size
2. H produces a fixed length output

Important properties

1. **Preimage resistant** : given h it should be hard to find any m such that h = hash(m) (one way function)

2. **Collision-resistant** : it should be hard to find two different messages m_1 and m_2 such that $hash(m_1) = hash(m_2)$.

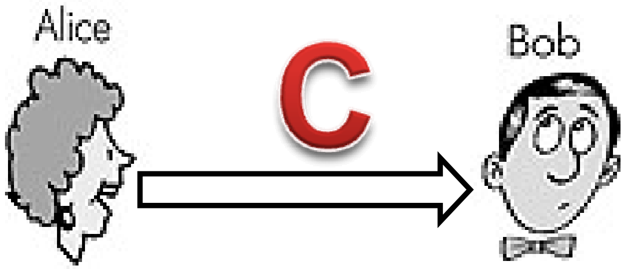


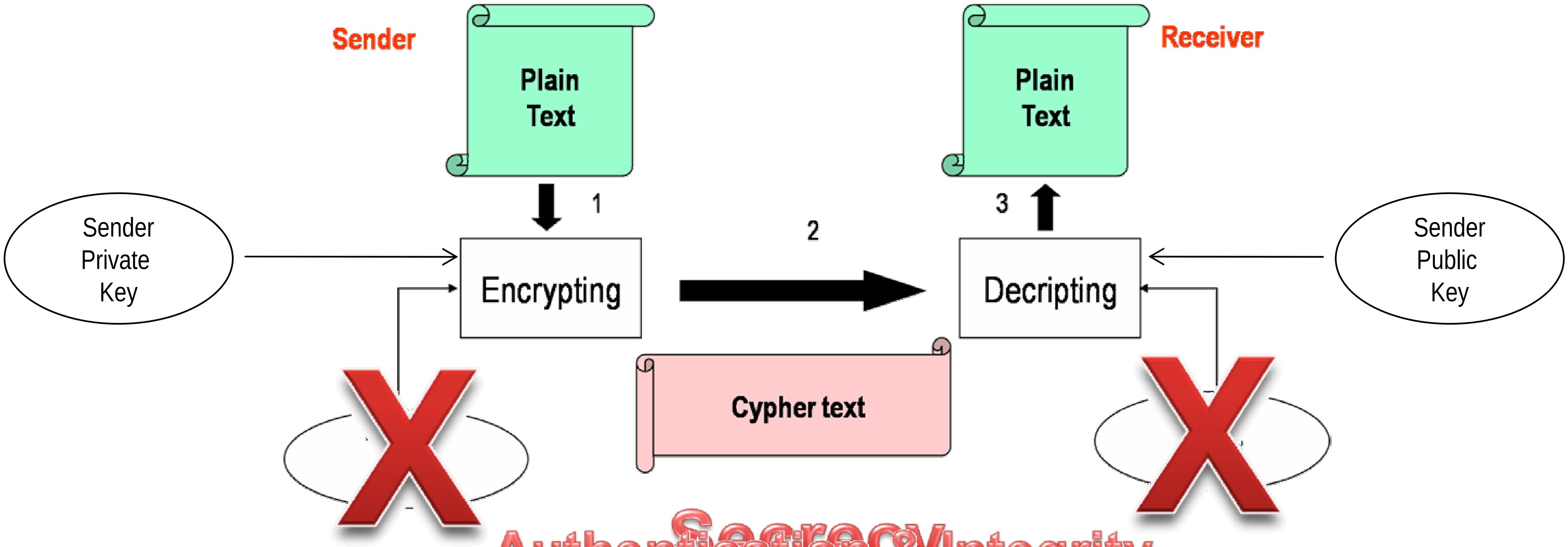
RSA (1 / 2)

- Two keys : One for the encryption, and one for the decription
- Each user has two keys :

Private key : Secret information
Public key : Information to be distributed

C = PRI (PUB (C))
C = PUB (PRI (C))

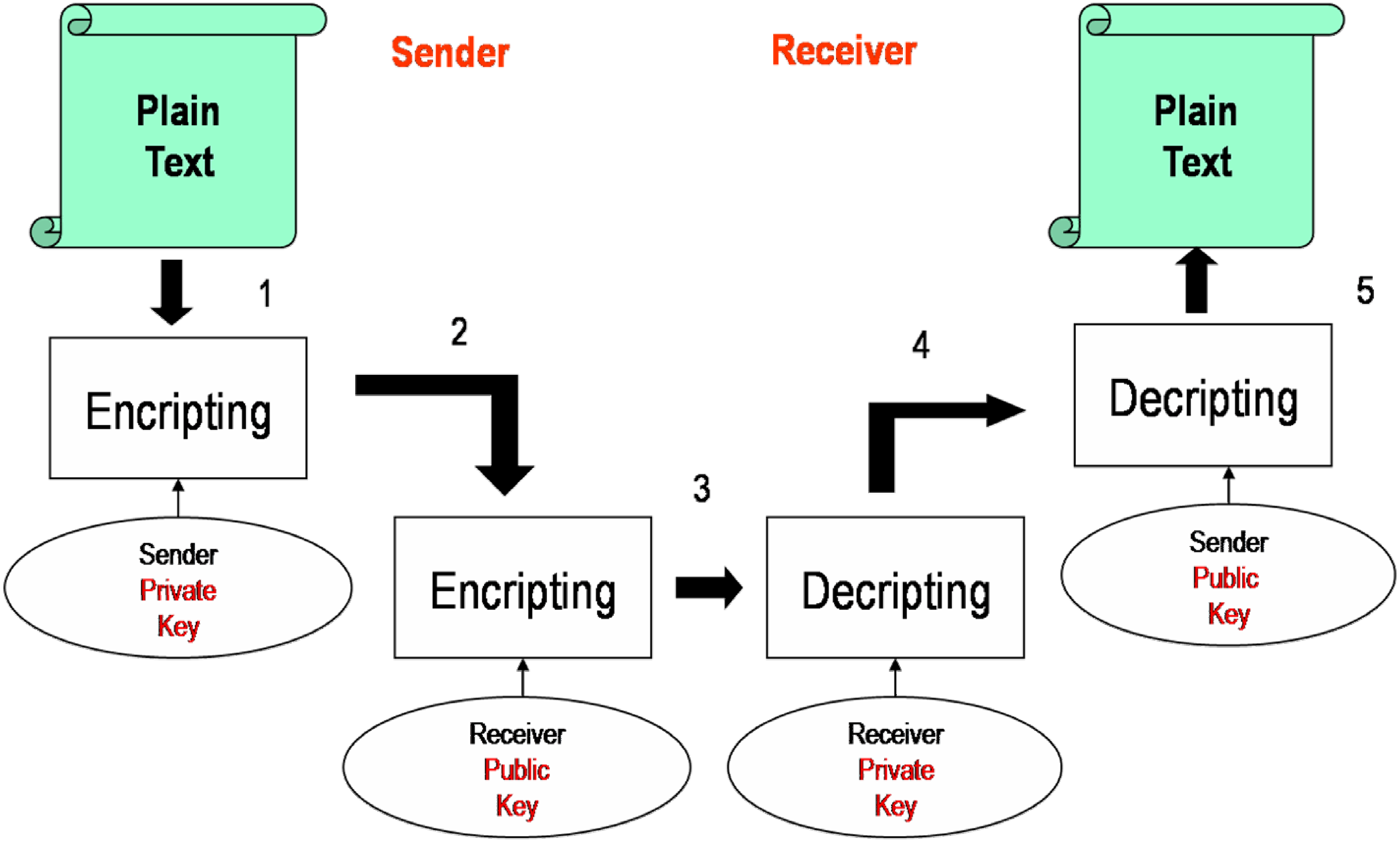



Sender
Plain Text
Receiver
Plain Text
1
2
3
Sender Private Key
Encrypting
Decripting
Sender Public Key
Cypher text

Secrecy
Authentication&Integrity

RSA (2 / 2)

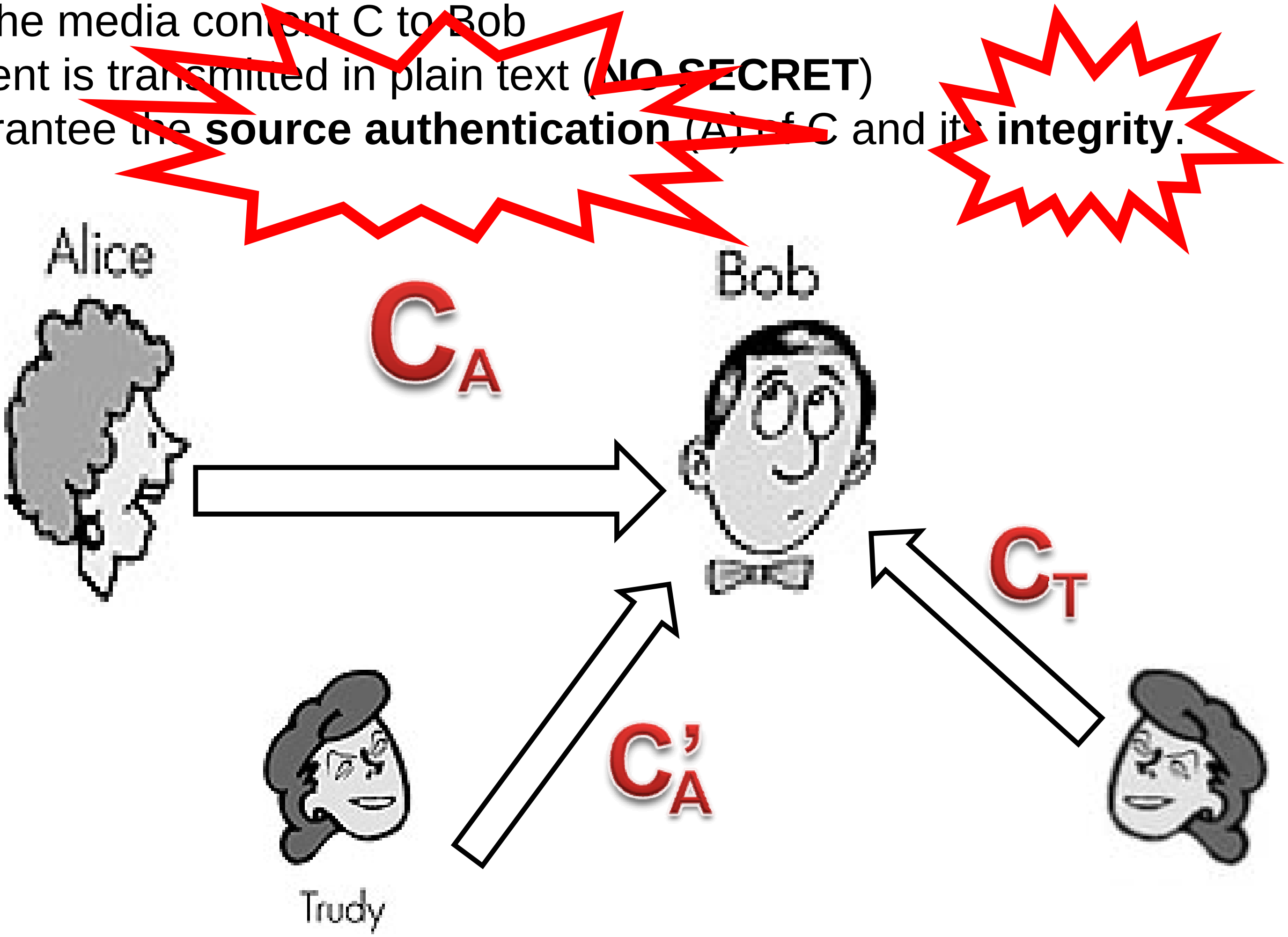
Put everything together : Secrecy + Authentication & Integrity



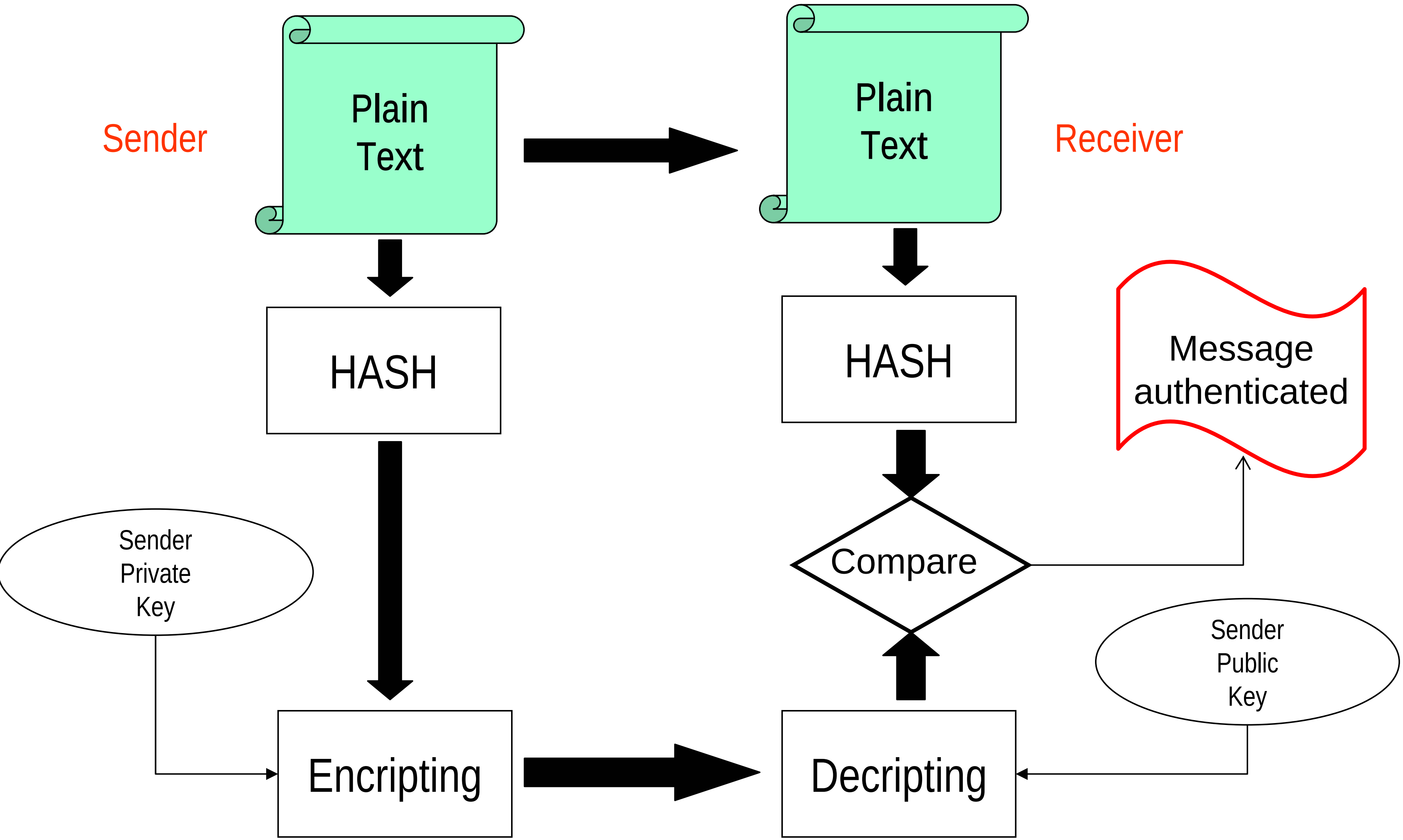
Digital signature (1/2)

The issue is **Source Authentication & Integrity**

- Alice transmits the media content C to Bob
- The media content is transmitted in plain text (**NO SECRET**)
- We want to guarantee the **source authentication** (A) of C and its **integrity**.



Digital Signature (2/2)



Sender

Plain Text

HASH

Sender Private Key

Encripting

Receiver

Plain Text

HASH

Message authenticated

Compare

Sender Public Key

Decripting

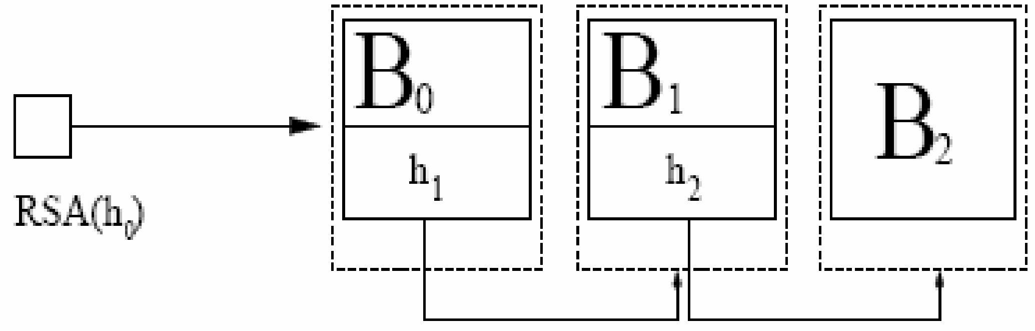
Authentication – The simple scheme

We have three entitities to authenticate : $B_0 \quad B_1 \quad B_2$

The first approach could be : $DS(B_0), DS(B_1), DS(B_2)$

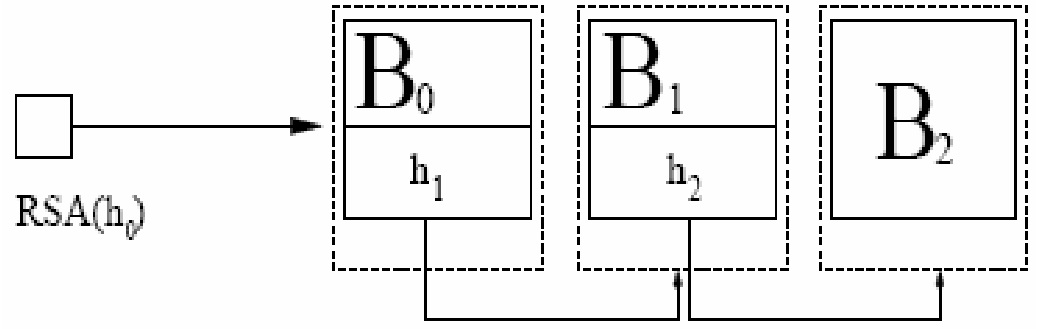
Three DS calculations : Too much computational overhead !!!

Hash chain : how to ammortize RSA computational overhead and propagate the source and content authentication



Authentication – The double chain

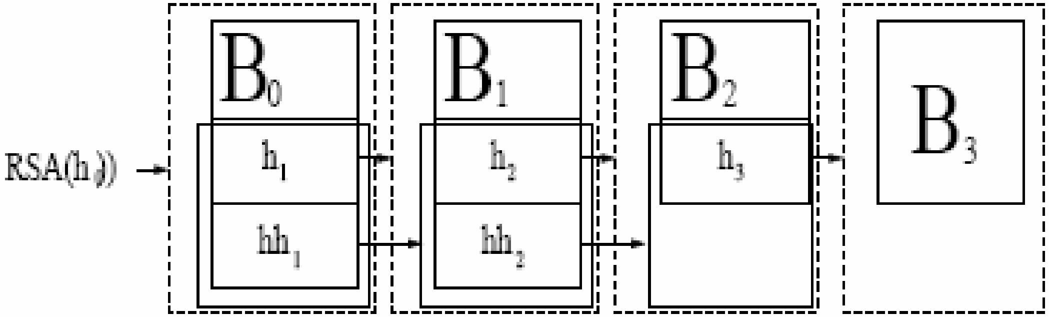
If there are changes in one only block, all subsequent blocks cannot be authenticated.



Dealing with bit error rate :

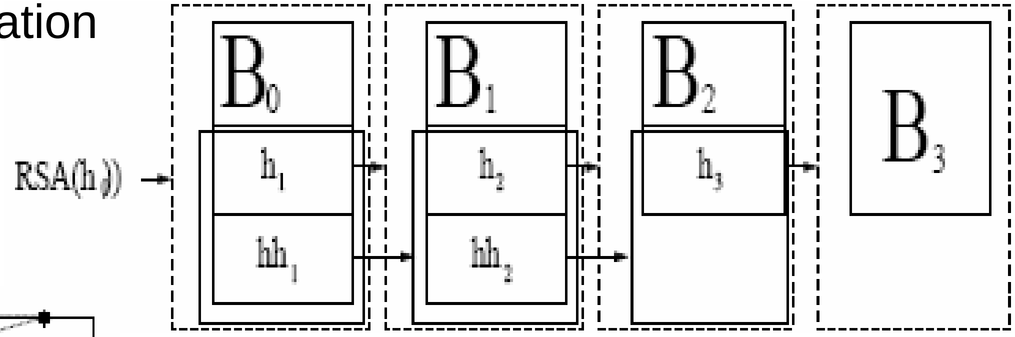
Two hash chains :

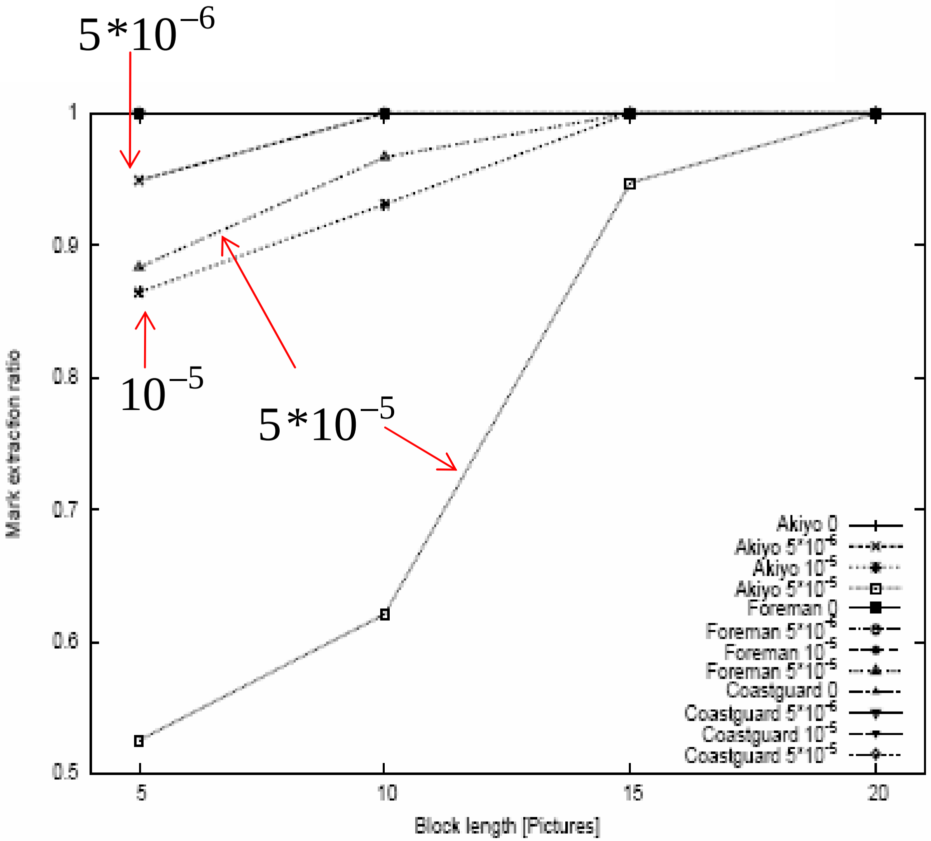
- ➢ one for the authentication of the subsequent block
- ➢ one for the authentication of the subsequent hashes



Authentication – Dealing with b.e.r.

- ➢ The block is a sequence of pictures ≥ 1
- ➢ We are able to introduce information (hash) into blocks, by means of a watermarking technique





$5*10^{-6}$

10^{-5}

$5*10^{-5}$

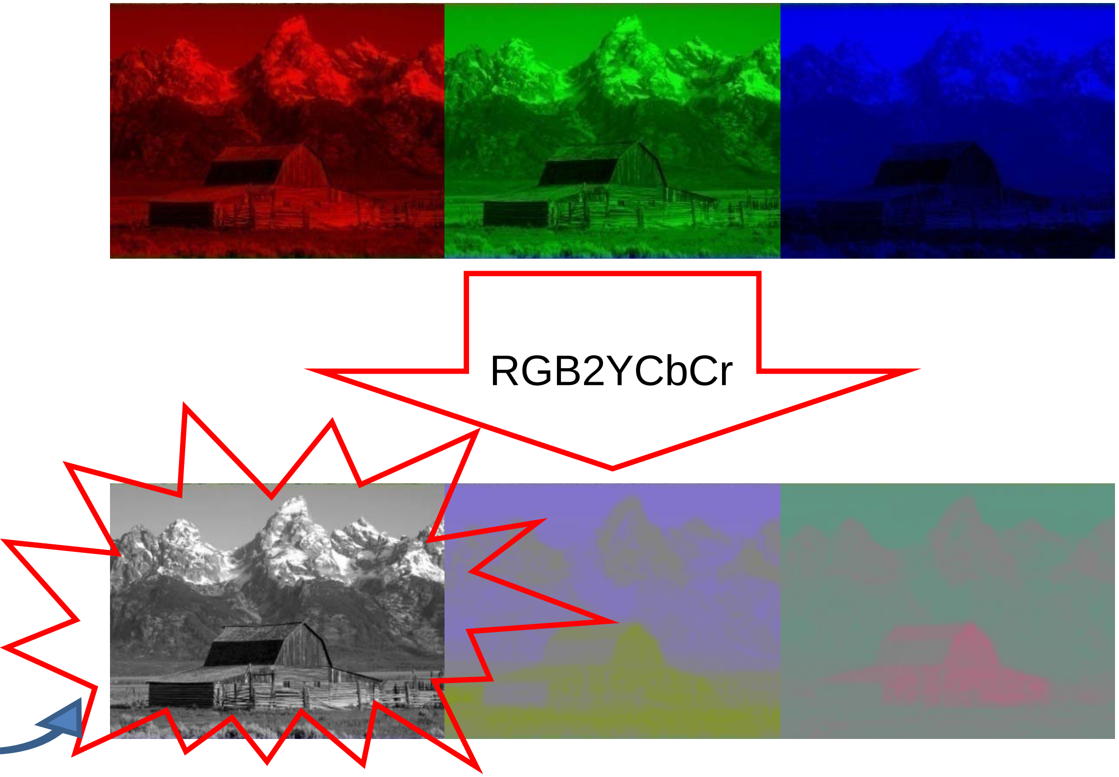
- ➢ Mark extraction ratio depends on number of mark replicas
- ➢ Bit error rate can be fight using longer block length

Color models

The **RGB color model** is an additive color model in which **red**, **green**, and **blue** light are added together in various ways to reproduce a broad array of colors.

YCbCr is a family of color spaces used in video and digital photography systems. **Y** is the luma component and **Cb** and **Cr** are the blue and red chroma components.

- It is a way of encoding RGB information.
- Highlights redundancies.

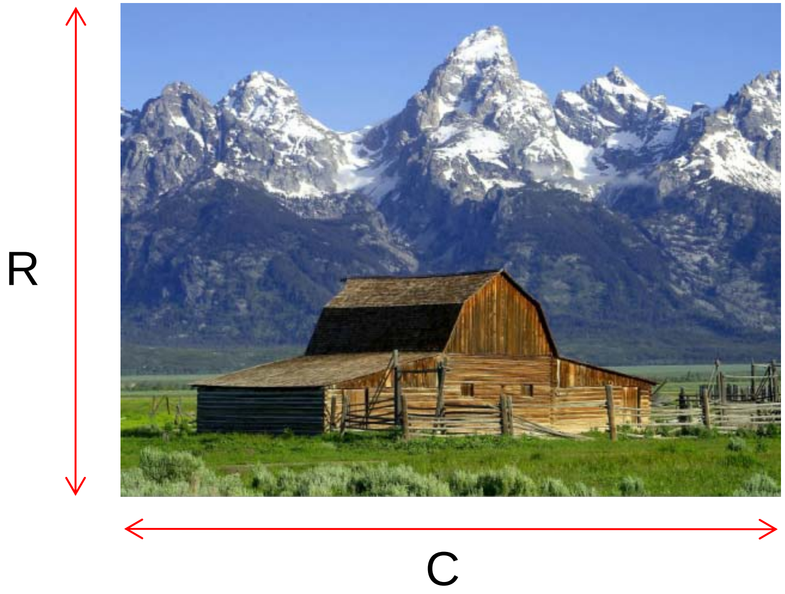


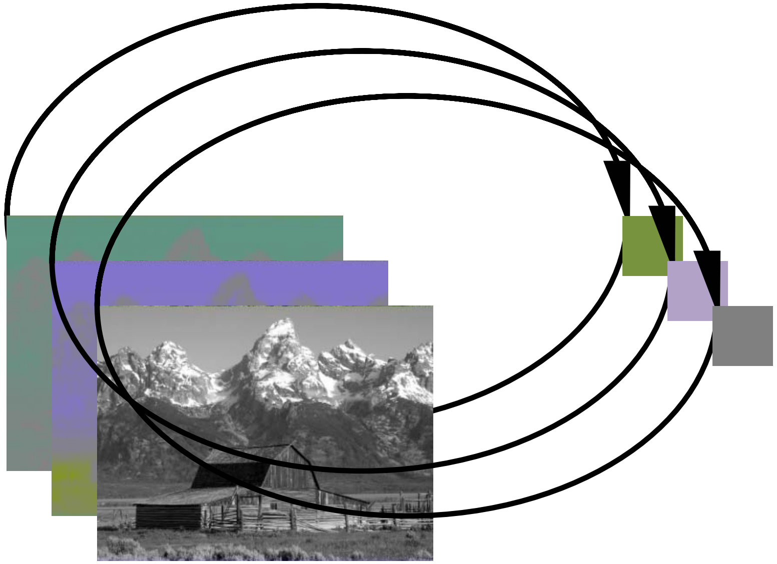
All the details are in the luminance component.

Media content representation (1/2)

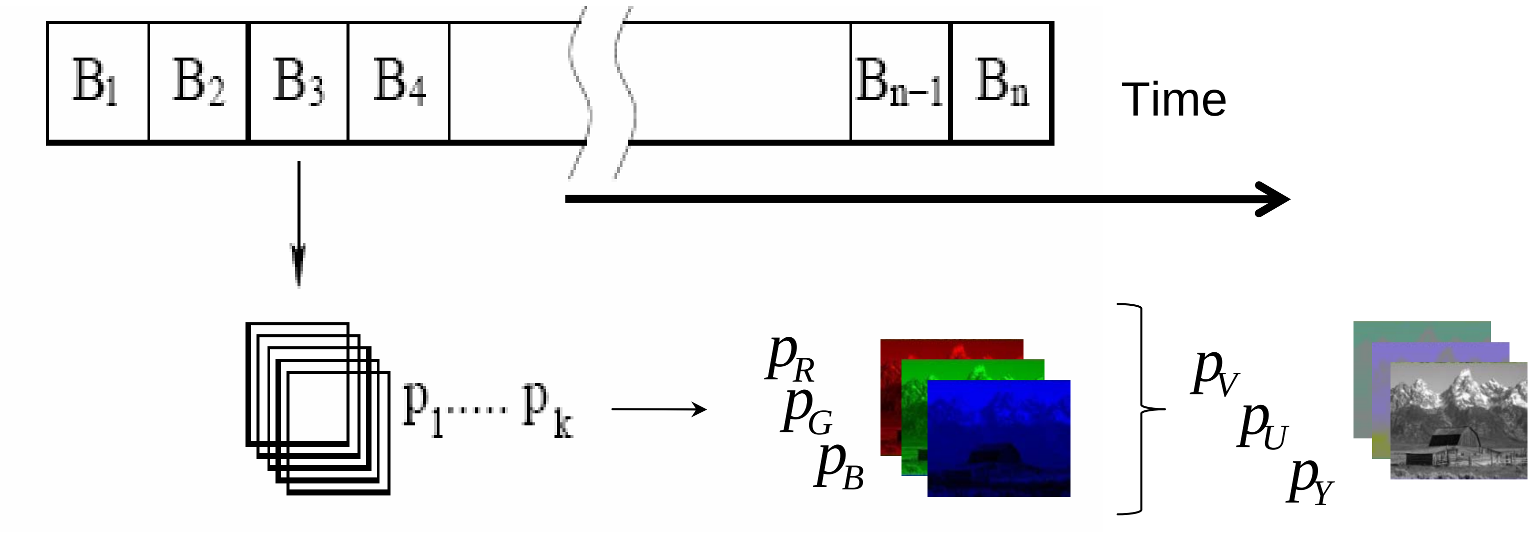
- Each picture can be considered as the composition of 3 matrix of R x C pixels.
- Each pixel value has 8 bit length and it is in the range [0..255]

- The mark will be embedded in the luminance component.
  - Chroma components are erased by MPEG-2 coding algorithm more than luma component.
  - The mark is embedded in the edges of the luminace component.





Media content representation (2/2)



1. A video stream can be considered as a three dimensional signal whose components are a bi-dimensional matrix (the picture) and the time.

2. Each block is composed by k picture (k>0)

3. Each picture is considered in the YCbCr space

Watermarking – Why ? How ?

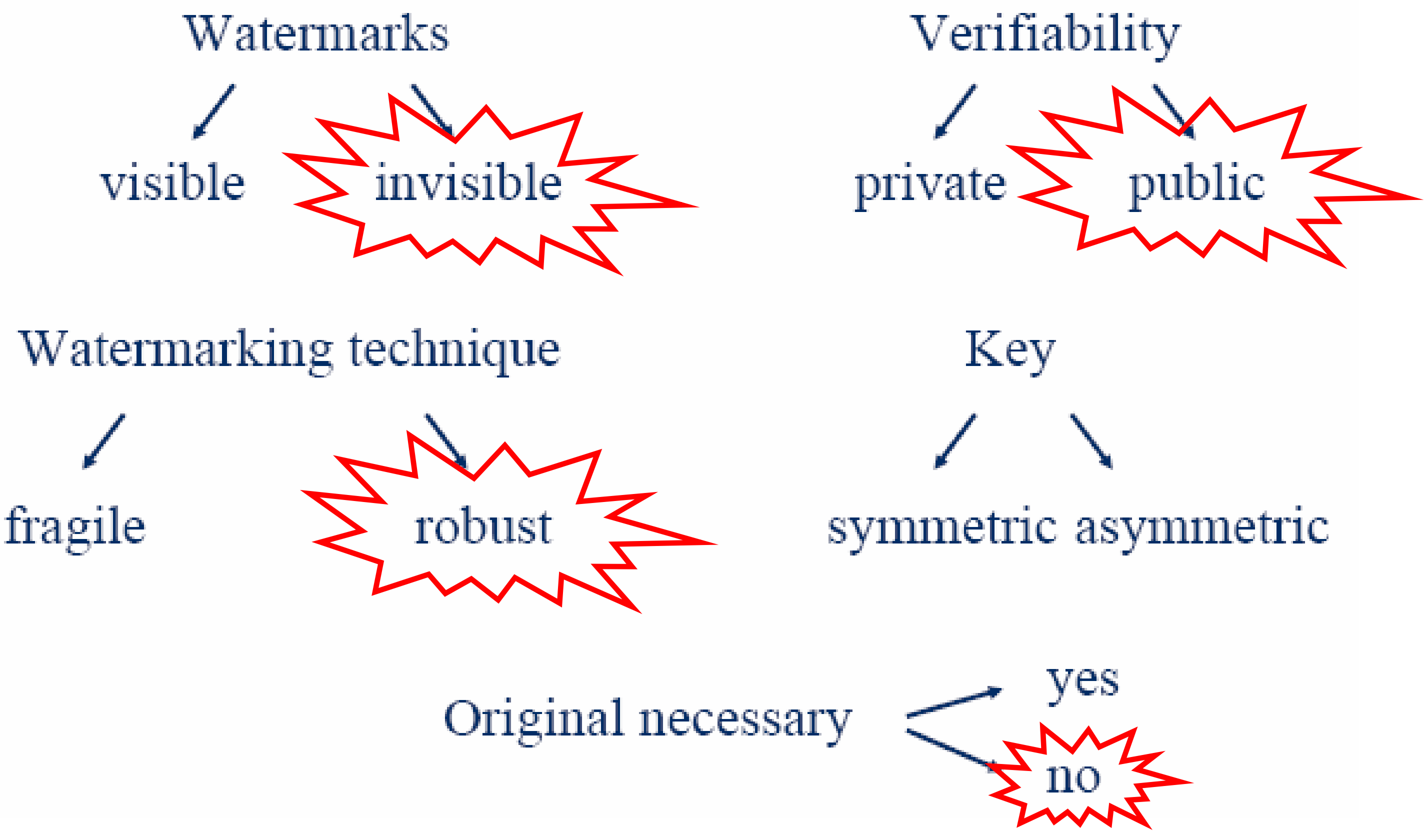
Why introducing data into video pictures ?

- ➢ No bandwidth overhead

How to introduce data into video pictures ?

- ➢ Watermarking

Watermarking concepts



Watermarks: visible / invisible

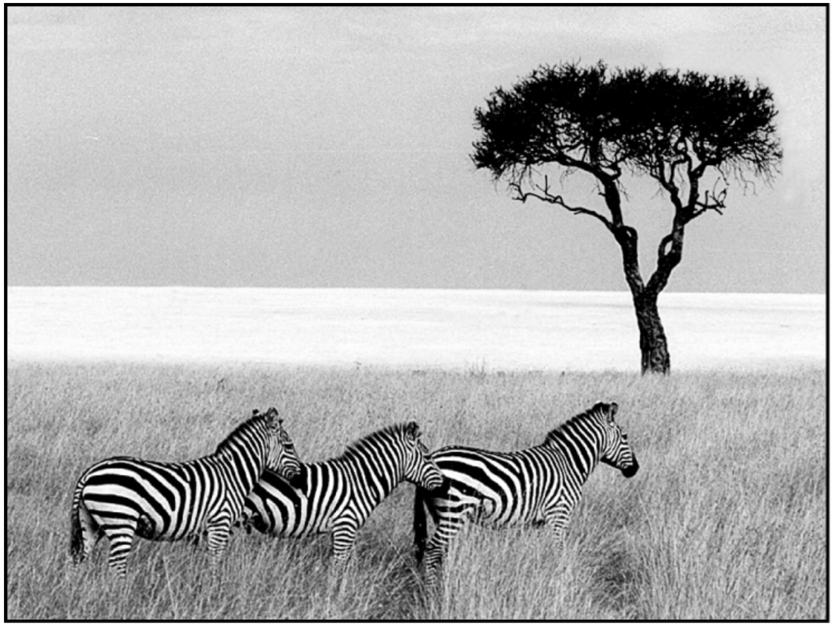
Verifiability: private / public

Watermarking technique: fragile / robust

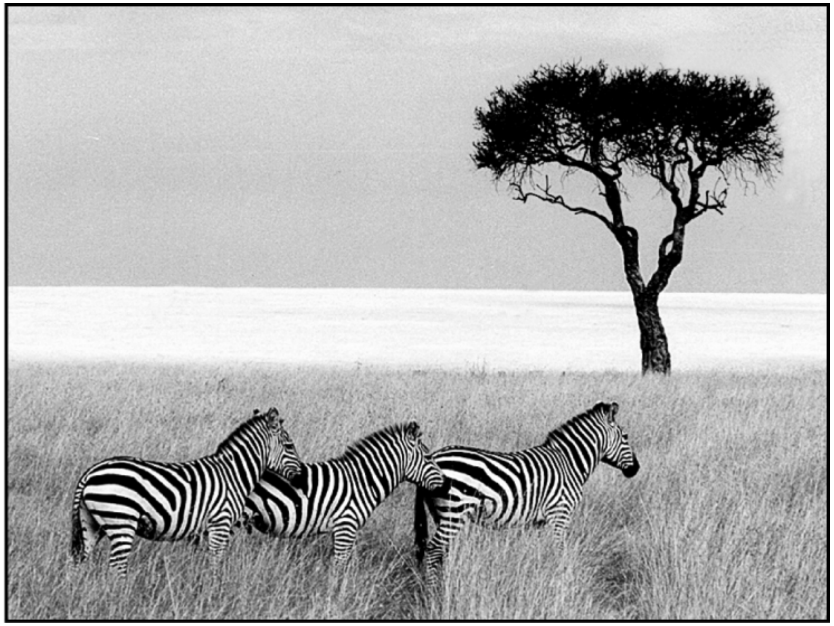
Key: symmetric asymmetric

Original necessary: yes / no

Steganography



Three zebras and a tree.

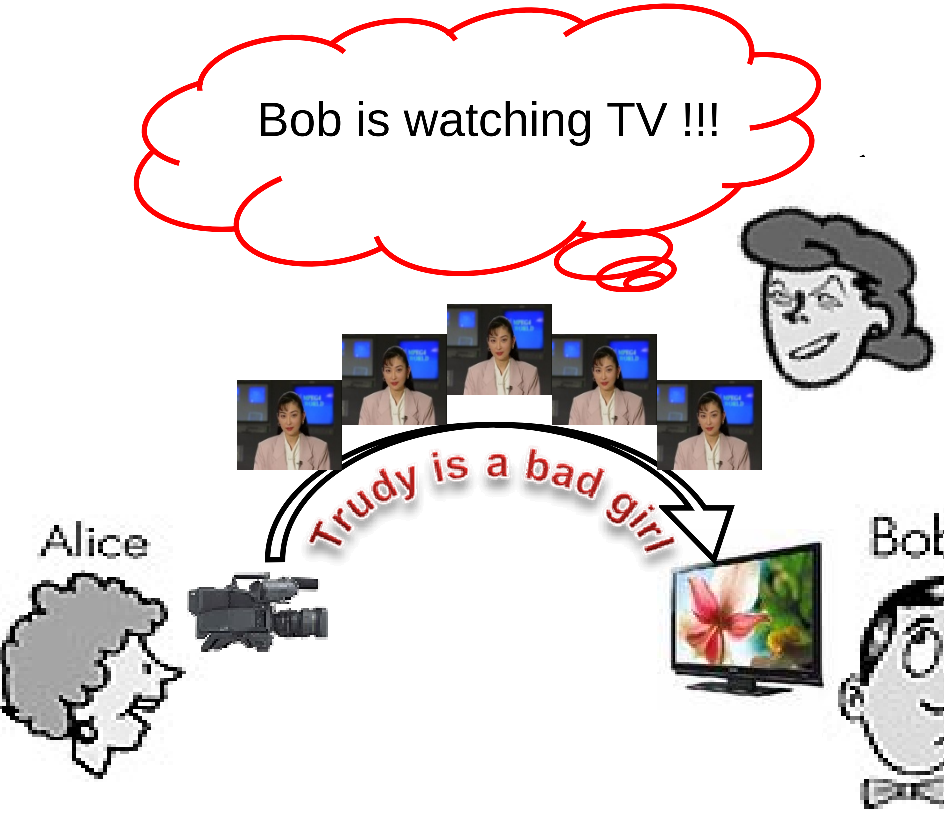


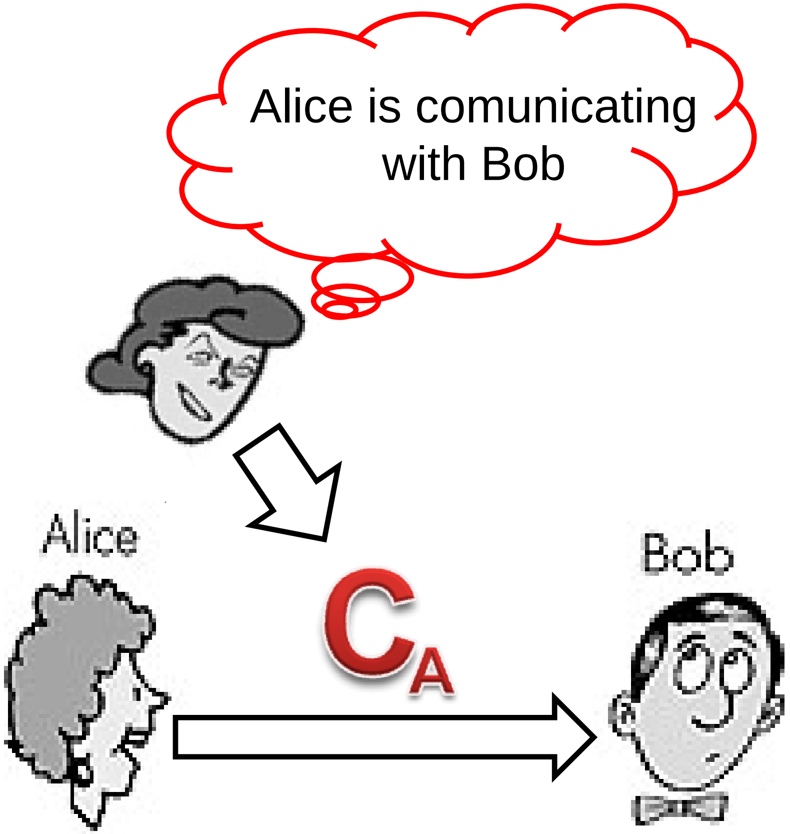
Three zebras, a tree, and the complete text of five plays by William Shakespeare.

Steganography / Cryptography

Cryptography obscures the meaning of a message, but it does not conceal the fact that there is a message.

Steganography is the art and science of writing hidden messages in such a way that no one apart from the sender and intended recipient even realizes there is a hidden message.





Watermarking overview (1)

How can watermarking be possible ?

- The visual system has very strong “error correction”

Requirements :

- Robust against tampering
  - Compression
  - Rotation
  - Resampling
  - Cropping
- Non perceptible

The mark embedding procedure

- **must have a negligible impact into the video stream quality**
- **should be robust to commonly used tampering techniques.**

The mark should be uniformly distributed over all the picture

Watermarking overview (2)

Original video

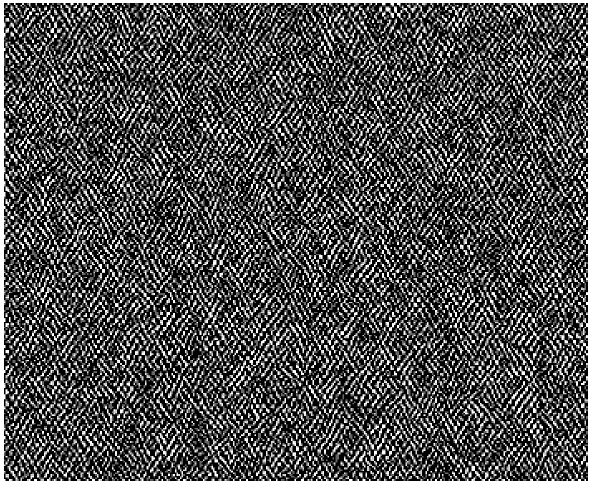
The mark is uniformly distributed over all the picture : robustness to **rotation, resampling, cropping**

-

Marked video

=

Amplified difference



Watermarking techniques

Least Significant Bit Modification

- Survive transformations such as cropping
- Any addition of noise or lossy compression is likely to defeat the watermark.
- An even better attack would be to simply set the LSB bits of each pixel to one fully defeating the watermark with negligible impact on the cover object.

Frequency domain

- This procedure is based on modifications of the image frequency domain coefficients; it thus has a minimal impact on the whole picture in the spatial domain.

- Bi-dimensional Discrete Cosine Transform (DCT2) approach.

Watermarking – DCT approach (1/2)

The Discrete Cosine Transform (DCT2) is closely related to the discrete Fourier transform.

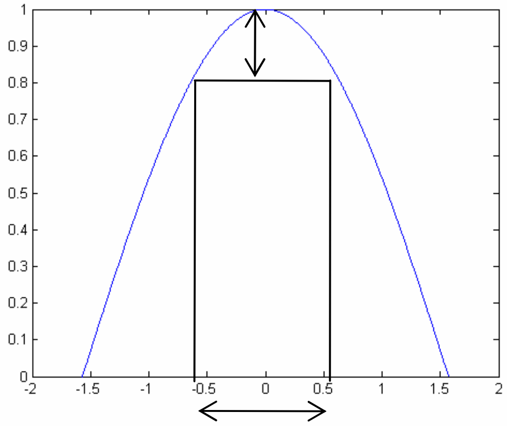
Suppose to have a matrix A with M rows and N columns, the transformed component B is :

Spatial domain → Frequancy domain

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2M} \quad , \quad \begin{array}{l} 0 \le p \le M-1 \\ 0 \le q \le M-1 \end{array}$$

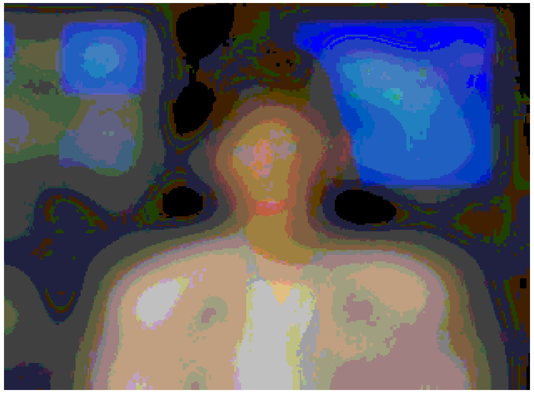
$$\alpha_p = \begin{cases} 1/\sqrt{M}, p = 0 \\ \sqrt{2/M}, 1 \le p \le M-1 \end{cases} \qquad \alpha_q = \begin{cases} 1/\sqrt{N}, q = 0 \\ \sqrt{2/N}, 1 \le q \le N-1 \end{cases}$$

Note that : The DCT tends to concentrate information, making it useful for image compression applications.

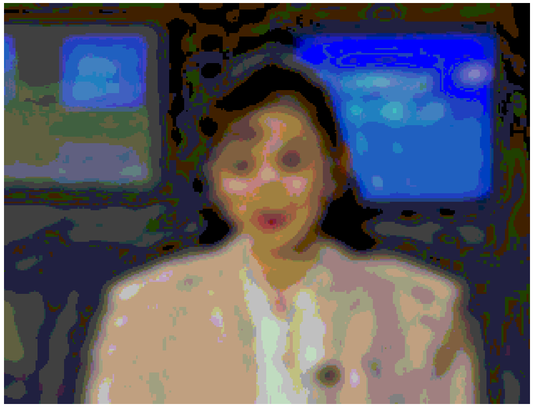


Watermarking DCT approach (2)

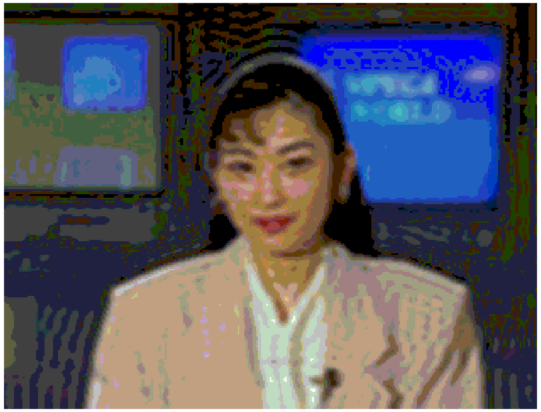
L = 10



L = 30



L = 70



DCT2 luminance components

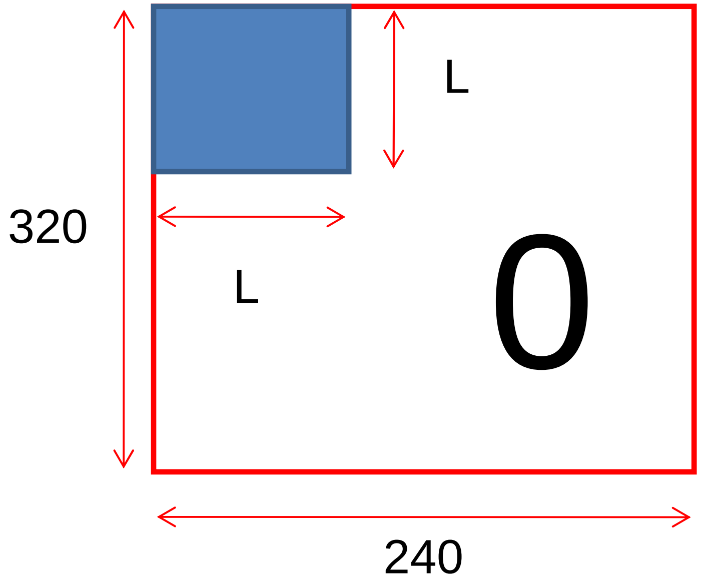
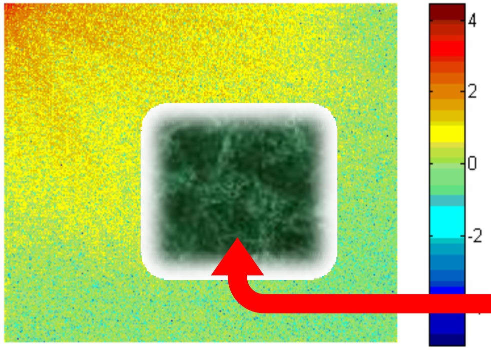


Image processing – DCT domain

Original Picture

Luminance Component

Hash Calculation



DCT2

Mark Embedding

Watermarking the picture

We want to embed $h = [-1, 1]$ in P

Input data :

$$P = \begin{matrix} -0.8808 & -0.8065 & 0.3192 & -0.0924 & -0.6532 \\ 0.3639 & 0.6363 & 0.0372 & -0.1352 & -0.2181 \\ -0.9151 & 0.6351 & 0.9459 & 0.6506 & 0.6628 \\ -0.8571 & 0.4449 & 0.2980 & -0.8331 & 0.6067 \\ 0.0433 & -0.7003 & 0.6007 & -0.7337 & -0.8791 \end{matrix}$$

$$P_i = \mathrm{DCT}(p_i(Y))$$

$$\overline{P}_i = P_i + |P_i| * \alpha * n_i * H$$

$$\overline{P}_i * n_i = P_i * n_i + |P_i| * \alpha * H$$

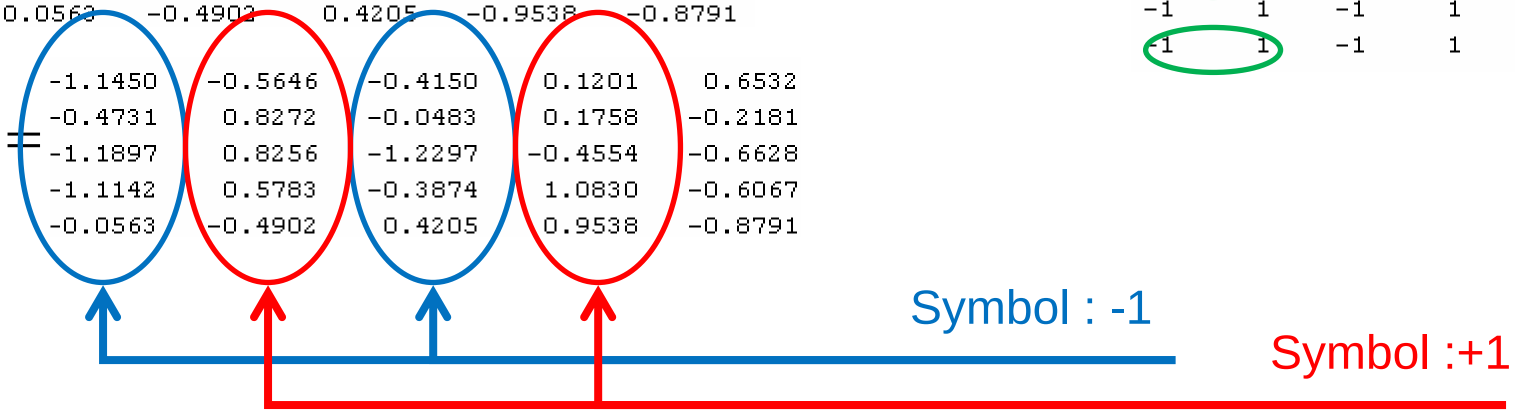
Amplification factor → $\alpha = 0.2$

Replication factor

$$n = \begin{matrix} 1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ -1 & 1 & 1 & -1 & 1 \end{matrix}$$

$$H = \begin{matrix} -1 & 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 1 & 0 \end{matrix}$$

$$\overline{P}_i = \begin{matrix} -1.1450 & -0.5646 & 0.4150 & -0.1201 & -0.6532 \\ 0.4731 & 0.8272 & 0.0483 & -0.1758 & -0.2181 \\ -1.1897 & 0.8256 & 1.2297 & 0.4554 & 0.6628 \\ -1.1142 & 0.5783 & 0.3874 & -1.0830 & 0.6067 \\ 0.0563 & -0.4902 & 0.4205 & -0.9538 & -0.8791 \end{matrix}$$



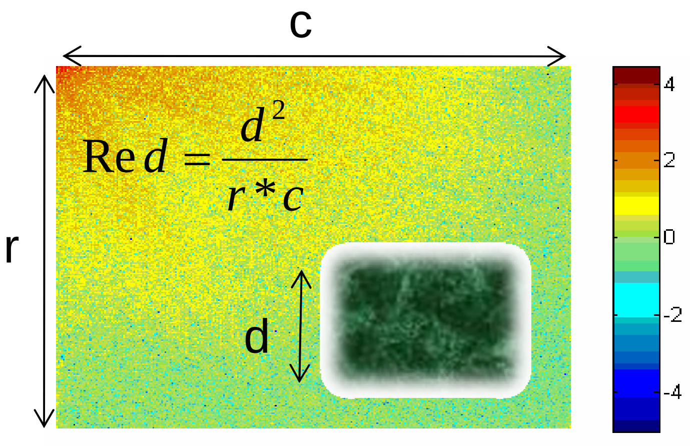
$$\overline{P}_i * n_i = \begin{matrix} -1.1450 & -0.5646 & -0.4150 & 0.1201 & 0.6532 \\ -0.4731 & 0.8272 & -0.0483 & 0.1758 & -0.2181 \\ -1.1897 & 0.8256 & -1.2297 & -0.4554 & -0.6628 \\ -1.1142 & 0.5783 & -0.3874 & 1.0830 & -0.6067 \\ -0.0563 & -0.4902 & 0.4205 & 0.9538 & -0.8791 \end{matrix}$$

Symbol : -1

Symbol :+1

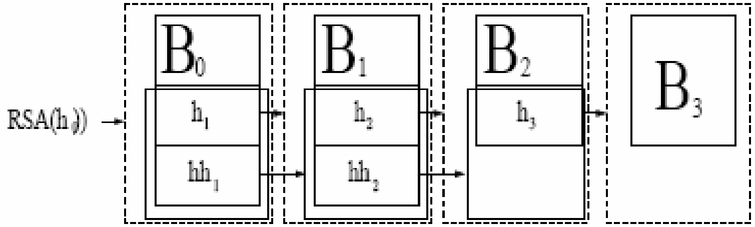
Watermarking and Authentication Issues

1. Mark extraction ratio
2. Authentication ratio
3. Coding robustness



$$\mathrm{Re}\,d = \frac{d^2}{r * c}$$

Video quality



Trade off between:
1.Mark embedding / extraction
2.Video quality

Mark carry authentication information:
If mark extraction fails, authentication cannot be possible

Configuration parameters
1.Message amplification factor
2.Message area
3.Block length

Simulation results

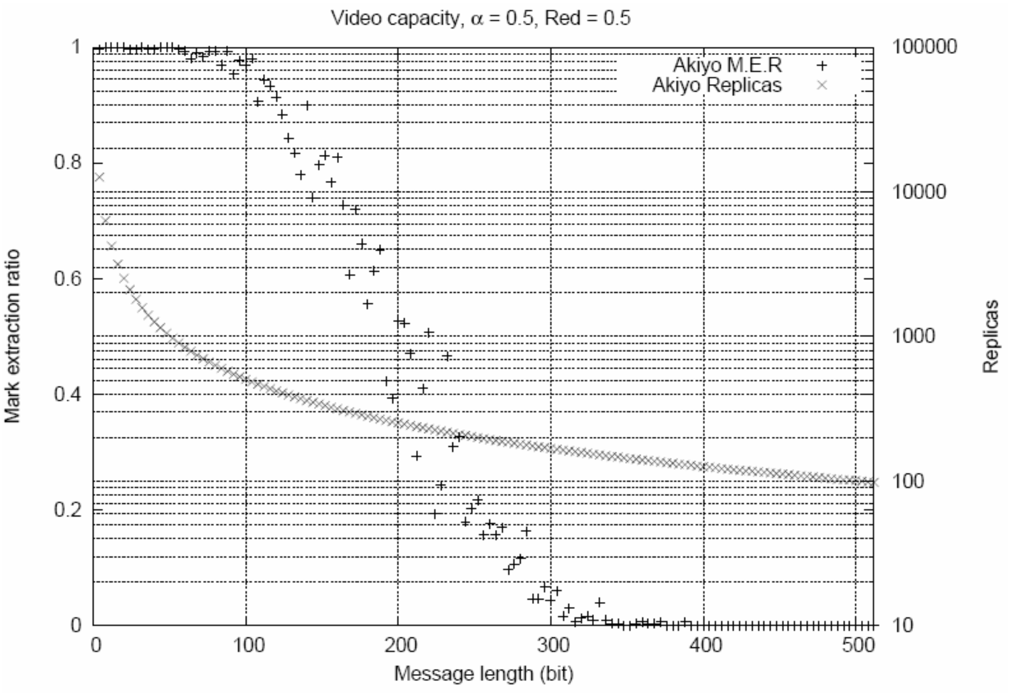
- Message length estimation
    - Fixing embedding parameters
    - Fixing the quality degradation (PSNR)
    - Mark extraction ratio VS message length

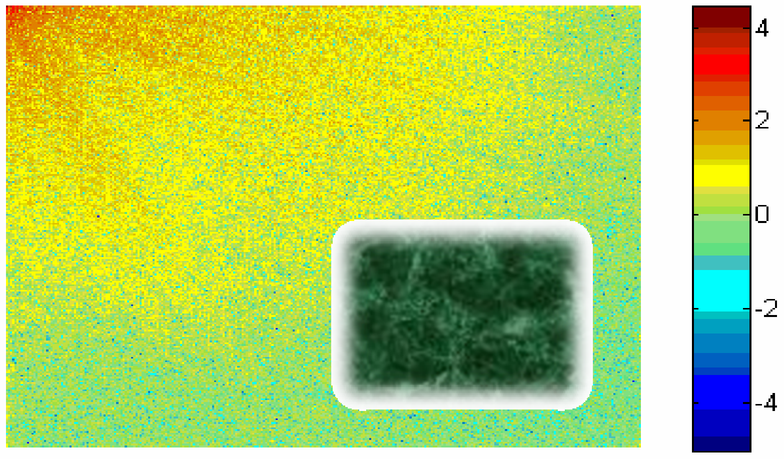
- Estimating optimal embedding parameters
    - Mark extraction ratio VS Quality degradation (PSNR)

- Estimating mark survival / robustness to MPEG-2
    - Mark extraction ratio VS Quality degradation (PSNR)

The video as a channel

Estimating “channel” bandwidth





1. Increasing the message length decreases the number of message replicas fixing the message area.
2. The mark extraction ratio decreases decreasing the number of replicas
3. Using these configuration parameters the message length threshold is 100 bit (the mark has been repeated 500 times).

Quality estimation - PSNR

Peak to Signal Noise Ratio (PSNR) is the ratio between the power of a signal and the power of corrupting noise that affects the fidelity of its representation.

$$\mathrm{PSNR} = 10 \log \frac{(2^n - 1)^2}{\sqrt{\mathrm{MSE(Y)}}}$$

$$\mathrm{MSE(Y)} = \frac{1}{3wh} * \sum_{i=1}^{N} (p_i^o(\mathrm{Y}) - p_i^w(\mathrm{Y}))^2$$

- ➢ b = 8 , color depth
- ➢ Only luma component is considered in the PSNR evaluation

Performance evaluation :

Mark extraction rate

VS

PSNR

- ➢ Amplification factor impacts on the video quality more than replication factor.
- ➢ C.P. valid for data emebedding are 24..27 and 30..34

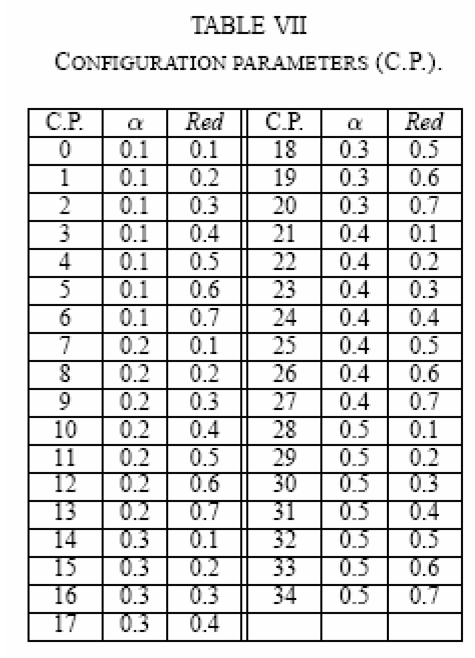
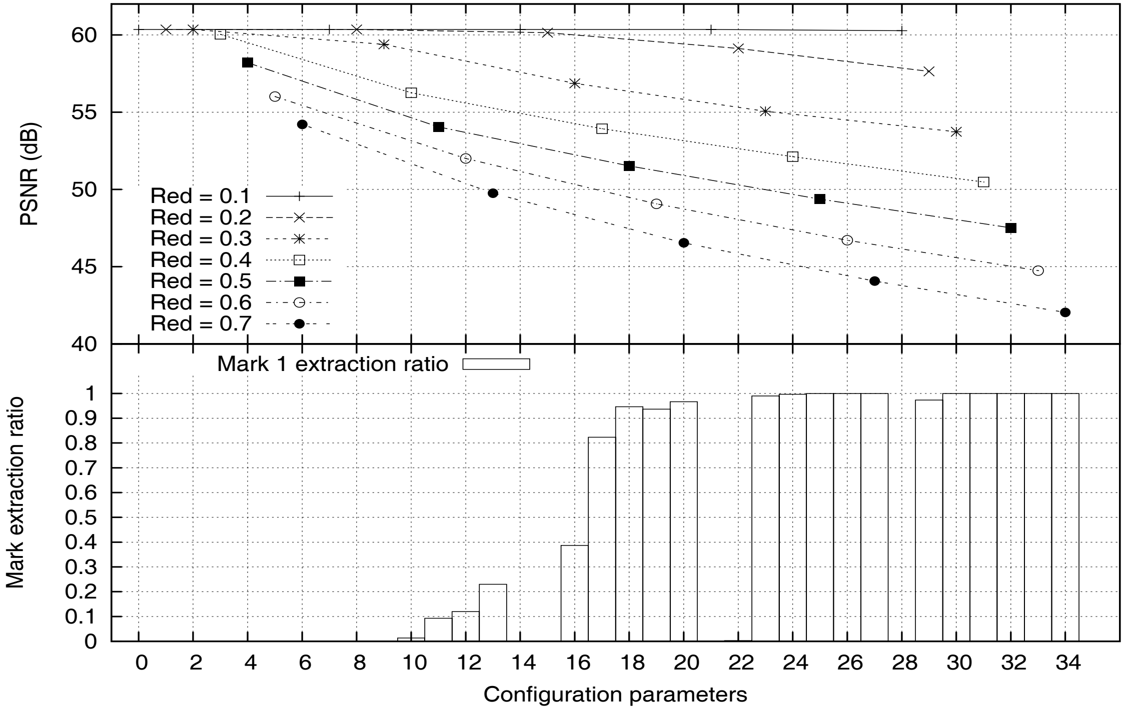


TABLE VII

CONFIGURATION PARAMETERS (C.P.).

C.P.	α	*Red*	C.P.	α	*Red*
0	0.1	0.1	18	0.3	0.5
1	0.1	0.2	19	0.3	0.6
2	0.1	0.3	20	0.3	0.7
3	0.1	0.4	21	0.4	0.1
4	0.1	0.5	22	0.4	0.2
5	0.1	0.6	23	0.4	0.3
6	0.1	0.7	24	0.4	0.4
7	0.2	0.1	25	0.4	0.5
8	0.2	0.2	26	0.4	0.6
9	0.2	0.3	27	0.4	0.7
10	0.2	0.4	28	0.5	0.1
11	0.2	0.5	29	0.5	0.2
12	0.2	0.6	30	0.5	0.3
13	0.2	0.7	31	0.5	0.4
14	0.3	0.1	32	0.5	0.5
15	0.3	0.2	33	0.5	0.6
16	0.3	0.3	34	0.5	0.7
17	0.3	0.4			

Raw Video



MPEG-2 Overview

Why multimedia contents should be compressed ?

- Multimedia contents are generally **redoundant** for the Human Visual System.
- Saving channel **bandwidth.**

How to compress data:

- Removal of temporal redundancy: inter-frame compression
- Removal of spatial redundancy (DCT): intra-frame compression
  - Quantisation of DCT coefficients

Compression – Time domain (1/2)

Temporal redundancy

Removal of temporal redundancy by means of inter-frame compression :

- **Motion vectors**

Compression – Time domain (2/2)

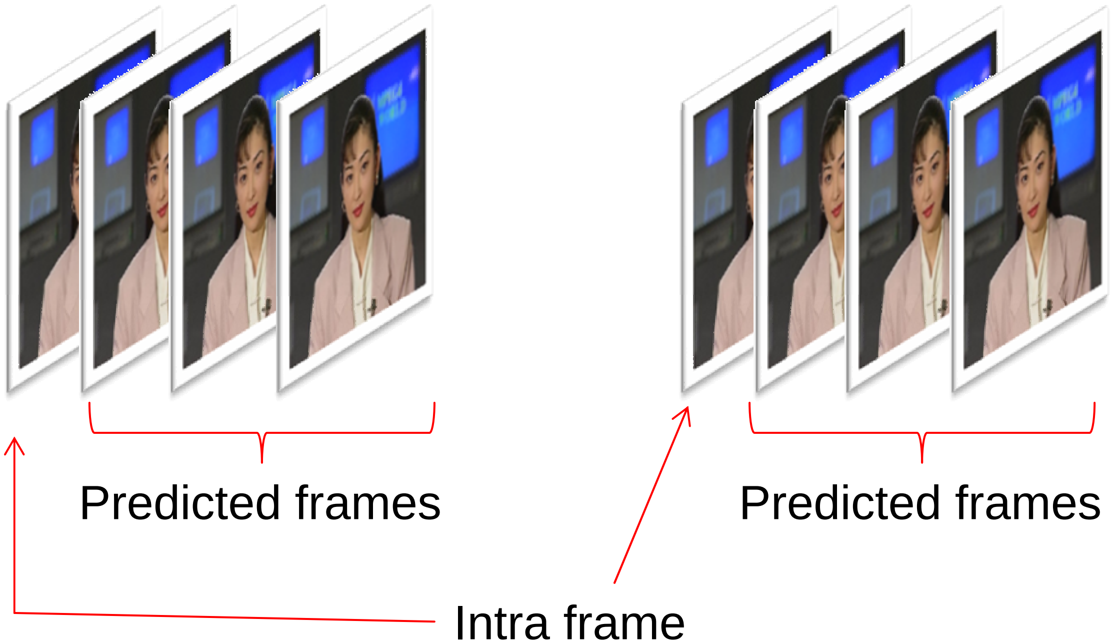
Three classes of video frame:

I (Intra) **frames**, make no reference to other frames

P (Predicted) **frames**, predicted from earlier I- or P-frames

B (Bi-directionally predicted) **frames**, predicted from both past and future frames

Typical sequence : I B B P B B P B B P I ......



Predicted frames

Predicted frames

Intra frame

Use **motion estimation** to predict the next frame.

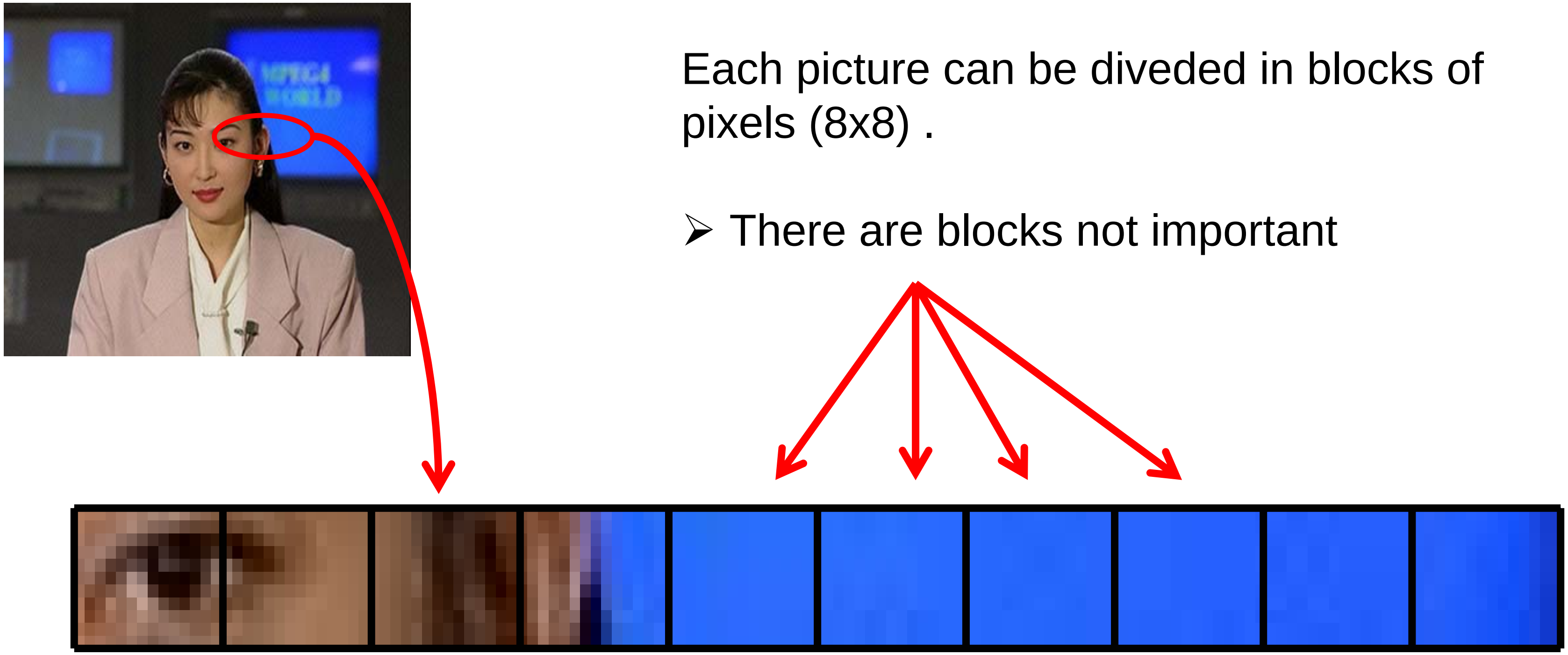
Use DCT to encode the difference between predicted and actual.

Compression – Space domain (1/3)

Spatial redundancy

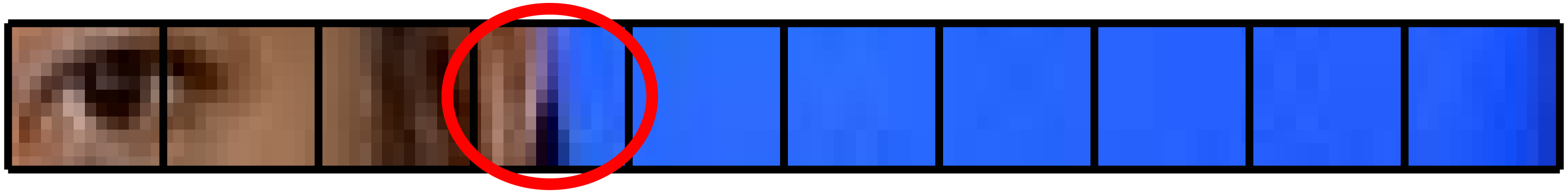
Each picture can be diveded in blocks of pixels (8x8) .

- ➢ There are blocks not important



Removal of spatial redundancy (DCT): intra-frame compression

Compression – Space domain (2/3)



Pixel values for a block

176	176	176	176	176	176	176	176
171	171	171	171	171	171	171	171
185	185	185	185	185	185	185	185
203	203	203	203	203	203	203	203
206	206	206	206	206	206	206	206
203	203	203	203	203	203	203	203
193	193	193	193	193	193	193	193
178	178	178	178	178	178	178	178

DCT2

Increasing horizontal frequency

Increasing vertical frequency

1106	12	-22	12	4	6	2	0
145	-15	-16	10	3	7	1	0
98	-4	-20	4	5	1	1	-1
52	-15	-8	1	-1	2	-2	0
18	-10	-1	-1	-1	1	-2	0
9	-4	-3	-2	1	-1	0	0
-4	2	-4	1	-3	2	1	0
-13	1	0	0	-1	1	1	2

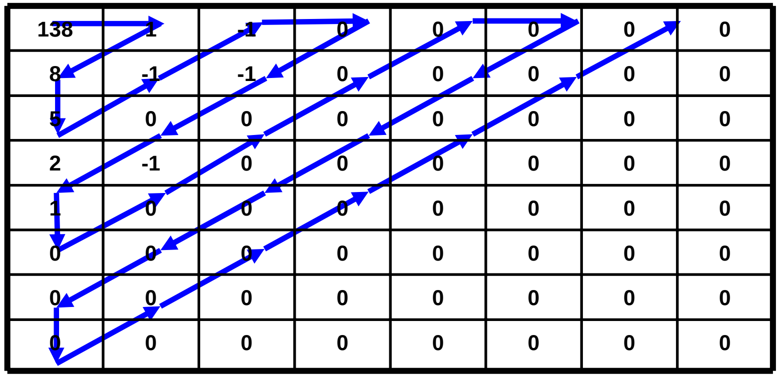
- Operates on blocks of 8x8 pixels.

- Discrete Cosine Transform (DCT) converts spatial elements to frequency domain (lossless).

Compression – Space domain (3/3)

138	1	-1	0	0	0	0	0
8	-1	-1	0	0	0	0	0
5	0	0	0	0	0	0	0
2	-1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

- ➢ DCT values after quantisation
- ➢ This is a **lossy step** in the algorithm



138	1	-1	0	0	0	0	0
8	-1	-1	0	0	0	0	0
5	0	0	0	0	0	0	0
2	-1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

- ➢ Conversion to serial data by zig-zag scanning
- ➢ Run length coding removes long strings of zeros
- ➢ Variable length coding replaces common values with shorter symbols

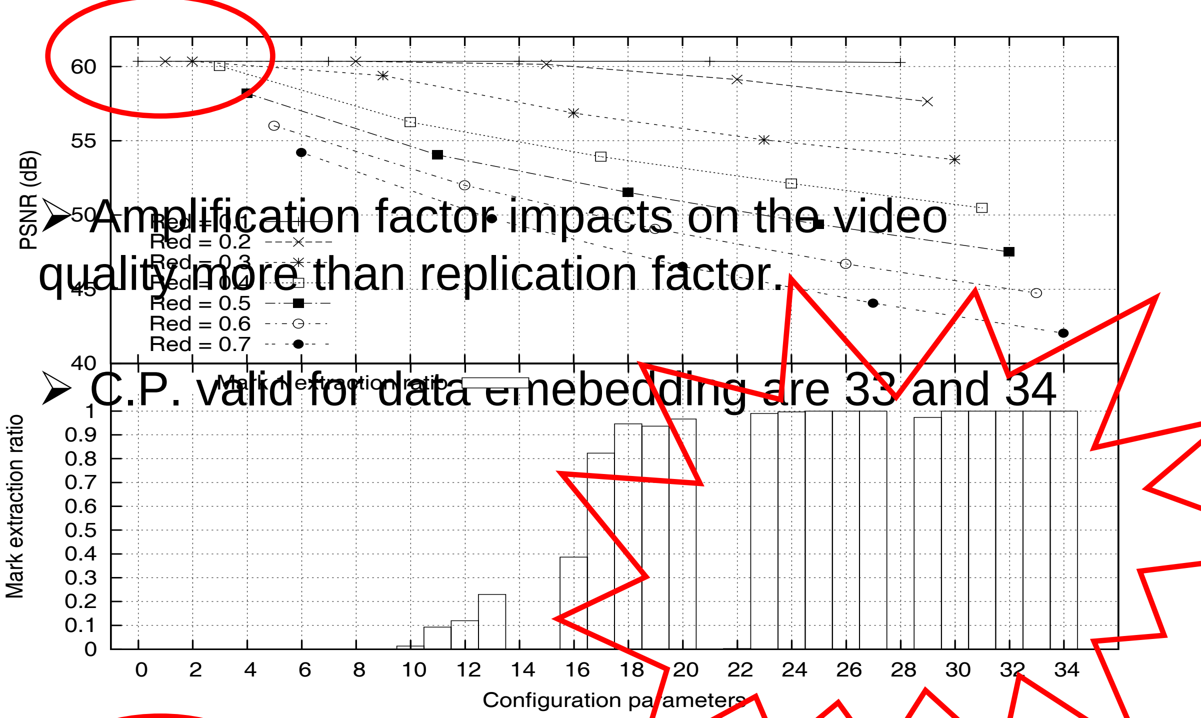
Performance evaluation :

Mark extraction rate

VS

PSNR

Raw Video



- ➢Amplification factor impacts on the video quality more than replication factor.
- ➢ C.P. valid for data emebedding are 33 and 34

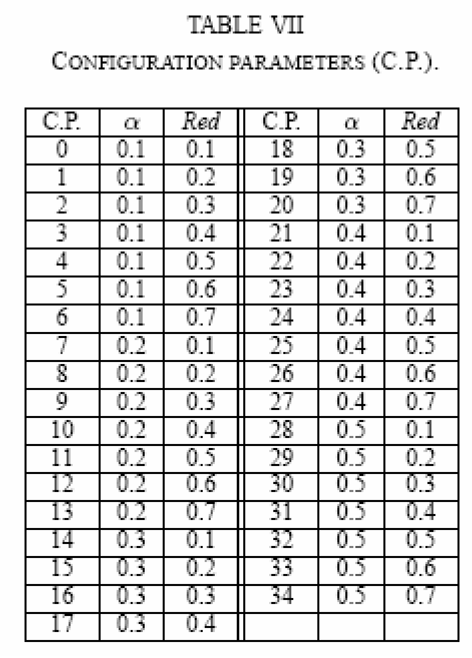
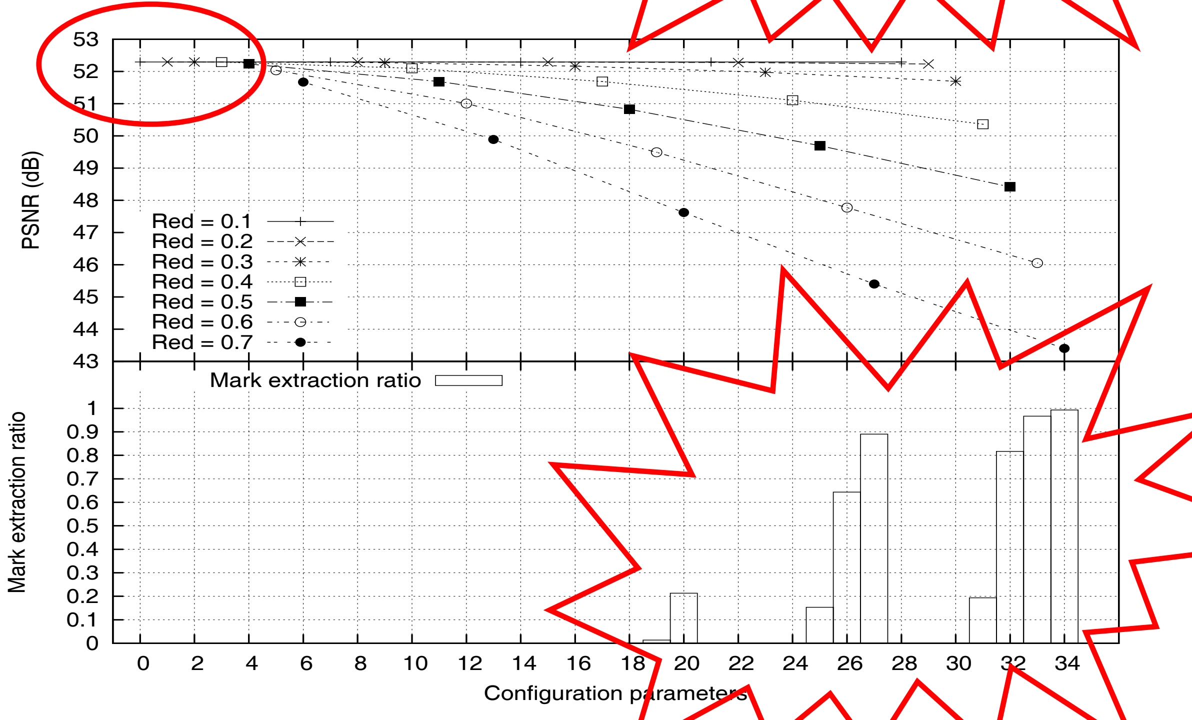


TABLE VII

CONFIGURATION PARAMETERS (C.P.).

C.P.	α	*Red*	C.P.	α	*Red*
0	0.1	0.1	18	0.3	0.5
1	0.1	0.2	19	0.3	0.6
2	0.1	0.3	20	0.3	0.7
3	0.1	0.4	21	0.4	0.1
4	0.1	0.5	22	0.4	0.2
5	0.1	0.6	23	0.4	0.3
6	0.1	0.7	24	0.4	0.4
7	0.2	0.1	25	0.4	0.5
8	0.2	0.2	26	0.4	0.6
9	0.2	0.3	27	0.4	0.7
10	0.2	0.4	28	0.5	0.1
11	0.2	0.5	29	0.5	0.2
12	0.2	0.6	30	0.5	0.3
13	0.2	0.7	31	0.5	0.4
14	0.3	0.1	32	0.5	0.5
15	0.3	0.2	33	0.5	0.6
16	0.3	0.3	34	0.5	0.7
17	0.3	0.4			

Mpeg2 Coded Decoded video



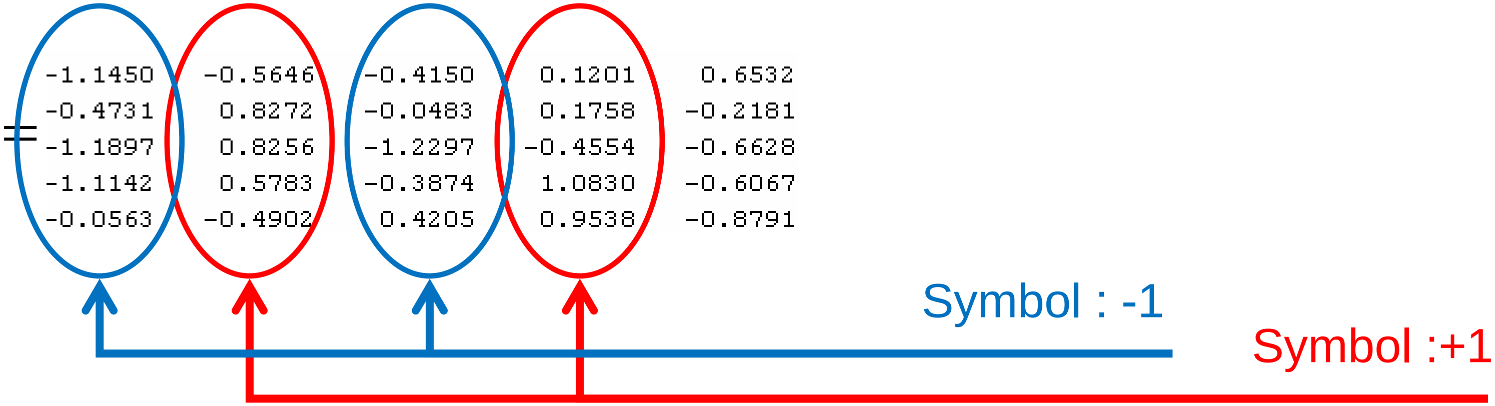
Mark extraction ratio – Theoretical Approach - (1 / 4)

$$P_i = \mathrm{DCT}(p_i(\mathrm{Y}))$$

$$\overline{P}_i = P_i + |P_i| * \alpha * n_i * H$$

$$\overline{P}_i * n_i = P_i * n_i + |P_i| * \alpha * H$$

- ➢ We consider all together the symbol values, and we evaluate the Probability Distribution Function.
- ➢ We derive a mathematical model for the symbol distributions
- ➢ We show a theoretical approach to the mark extraction ratio.



$$\overline{P}_i * n_i = \begin{matrix} -1.1450 & -0.5646 & -0.4150 & 0.1201 & 0.6532 \\ -0.4731 & 0.8272 & -0.0483 & 0.1758 & -0.2181 \\ -1.1897 & 0.8256 & -1.2297 & -0.4554 & -0.6628 \\ -1.1142 & 0.5783 & -0.3874 & 1.0830 & -0.6067 \\ -0.0563 & -0.4902 & 0.4205 & 0.9538 & -0.8791 \end{matrix}$$

Symbol : -1

Symbol :+1

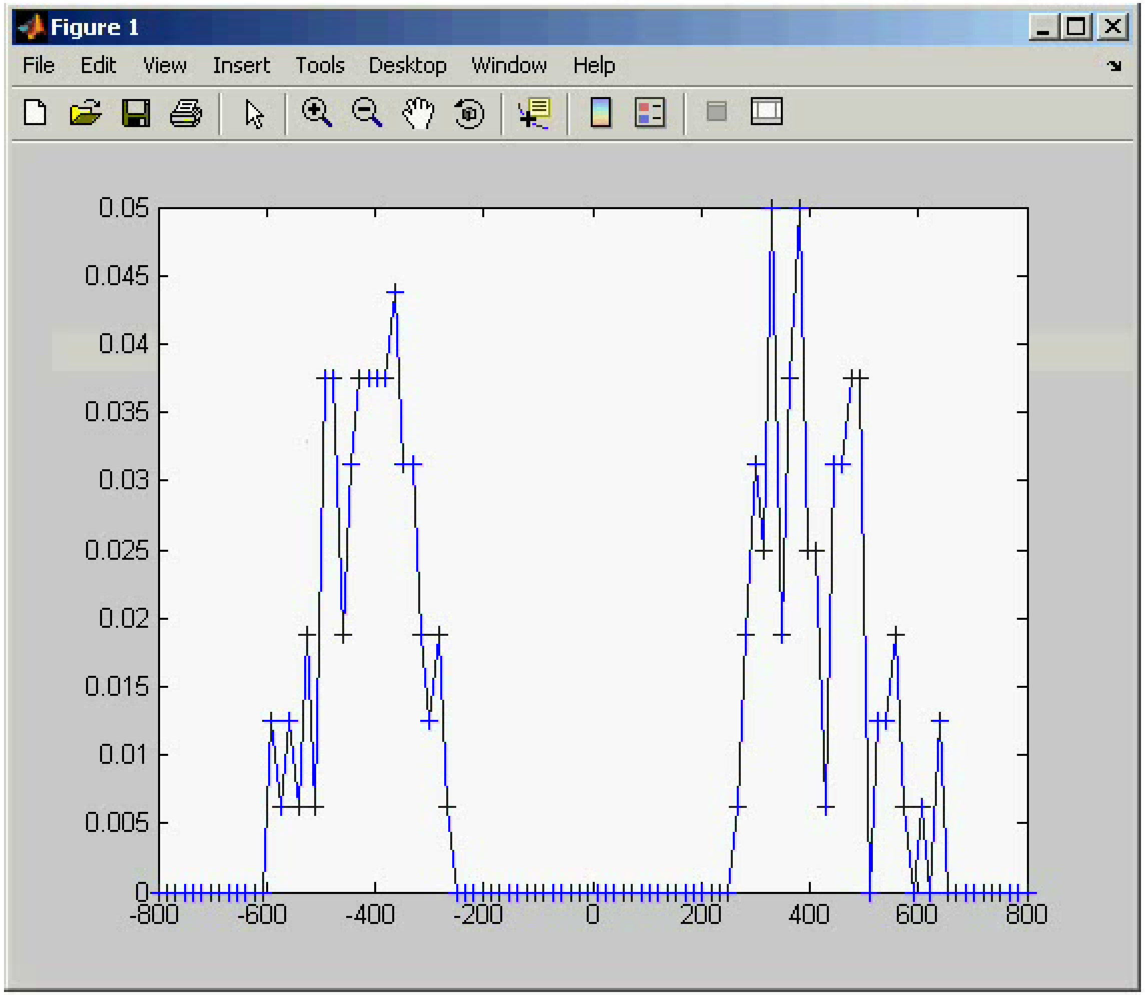
Mark extraction ratio – Theoretical Approach - (2 / 4)

- Akiyo video with configuration parameter **alpha = 0.5, red =0.5**
- One PDF for each picture , 300 pictures

- Symbols distributions are far from zero.

- Embedding parameters are good.

- The signum function works great.



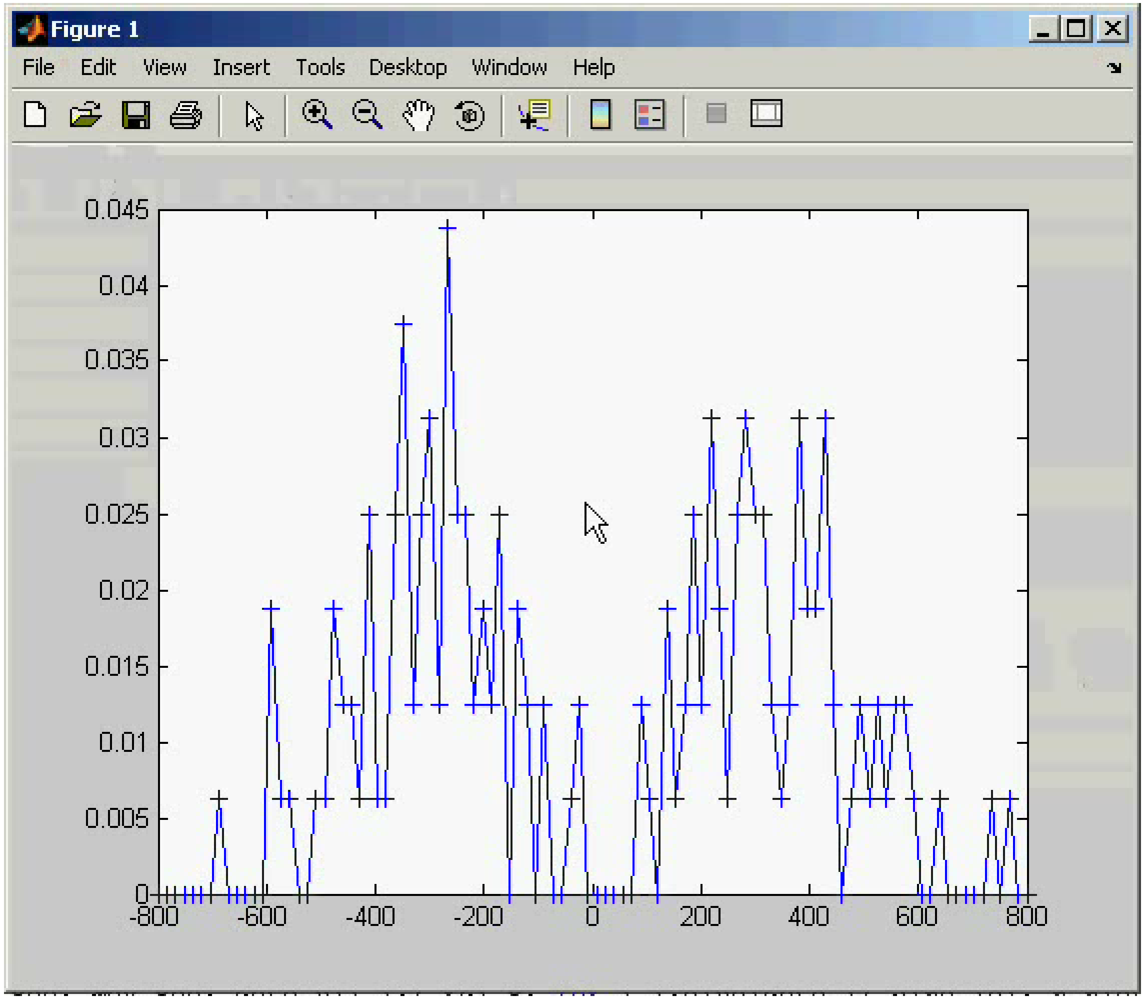
Mark extraction ratio – Theoretical Approach - (2 / 4)

- Akiyo video with configuration parameter **alpha = 0.2, red =0.7**
- One PDF for each picture , 300 pictures

- Symbols distributions are near zero.

- Embedding parameters are **not good**.

- There are symbols that will not be retrieval



Mark extraction ratio – Theoretical Approach - (3 / 4)

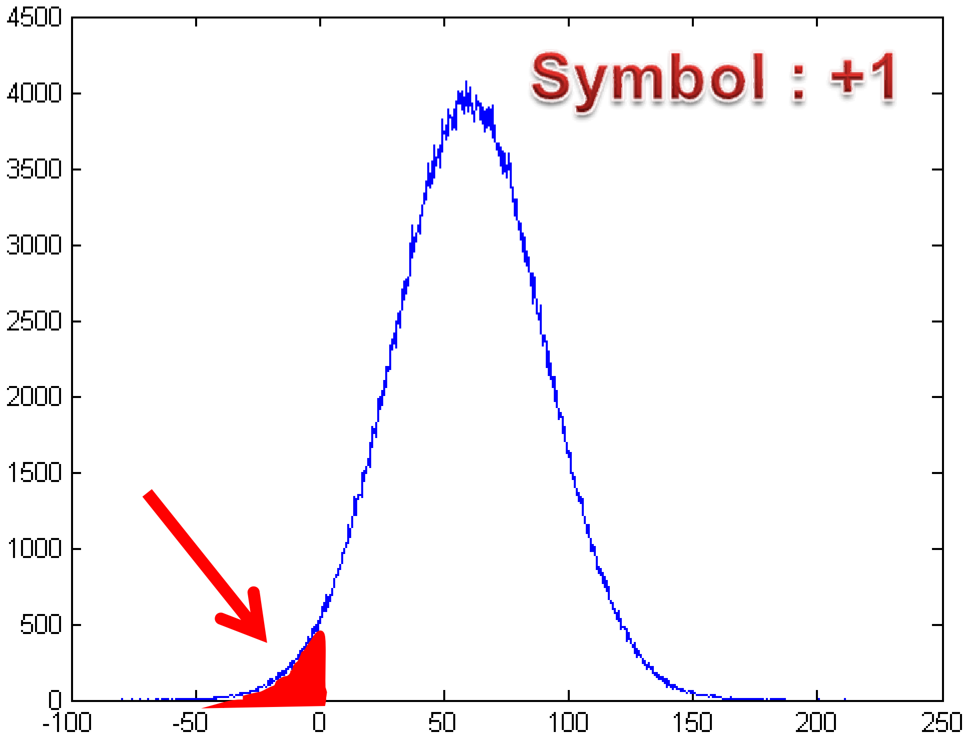
- The mark extraction probability will be :

$$P_t = (1 - P_e)^L$$

where L is the mark length and Pe is the probability of wrong retrieval.

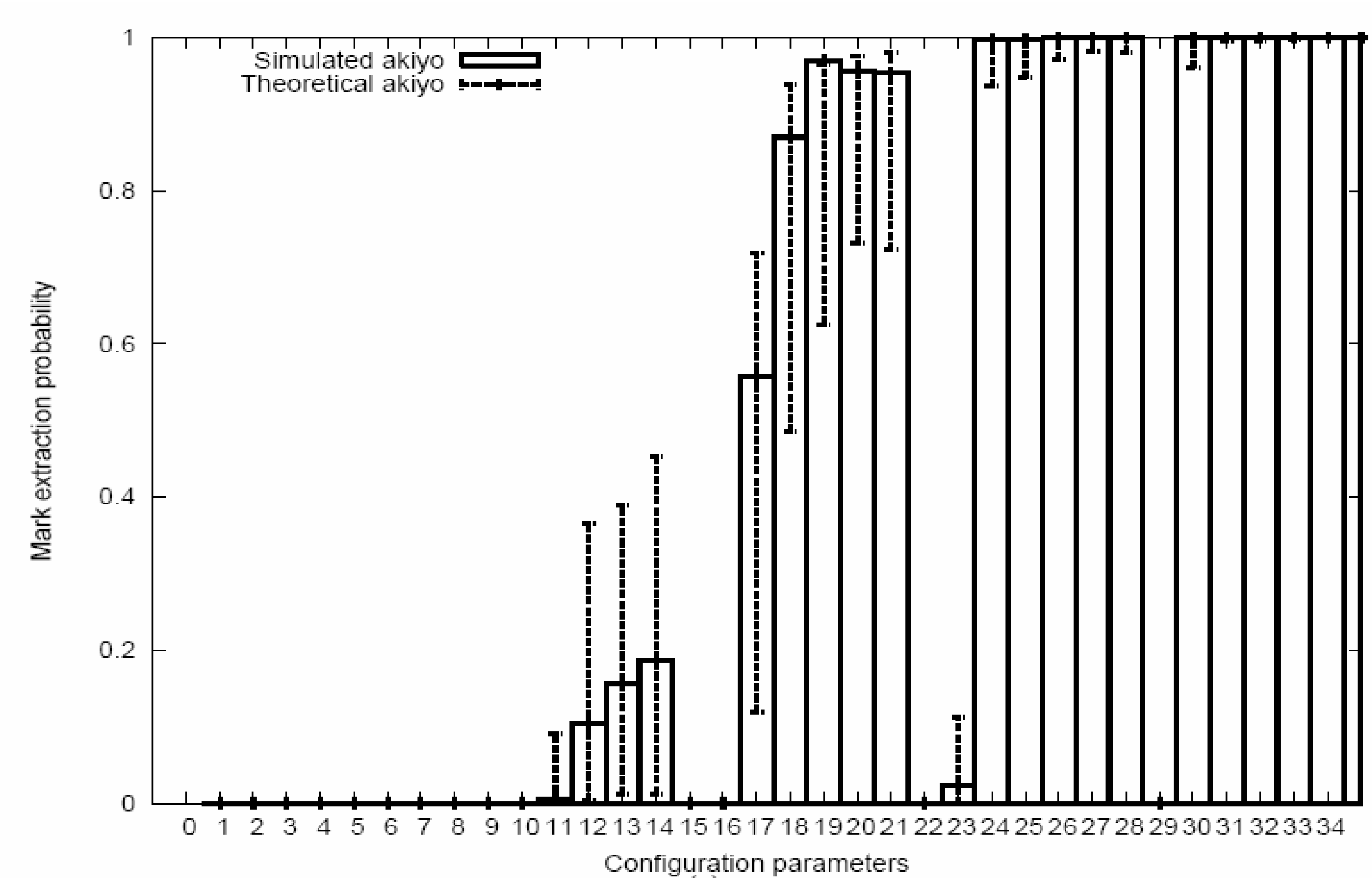
- The wrong symbol extraction probability P_e can be computed evaluating the area in the figure:

- Symbol distribution can be considered as normally distributed.



Mark extraction ratio – Theoretical Approach - (4/4)

- Akiyo video
- Theoretical and measured mark extraction ratio



Future works

- Optimizing mark embedding extraction
  - Embedding the mark directly in the MPEG-2 domain
  - Mark embedding enhancements
- New authentication scheme based on Tesla and BiBa
- Calculating the mark over features extracted from the pictures
- Using watermarking technique as a communication channel

The End

Thank you for your attention

Simple questions, please !