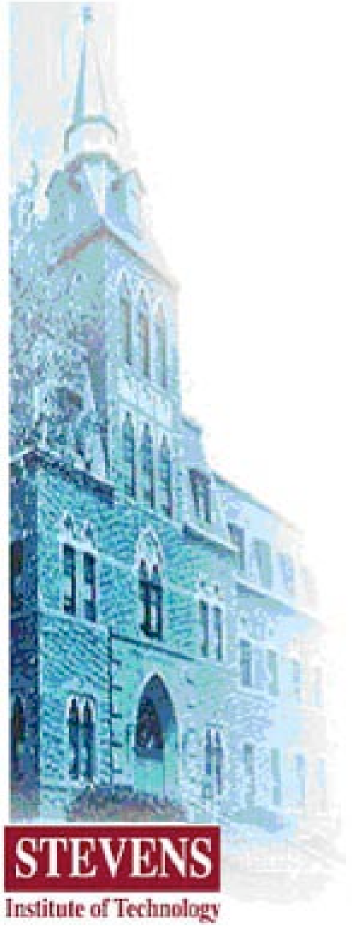


Introduction to Watermarking and steganography

Outline

- Watermarking
  - Definition and basics
  - Main applications
  - Attack mechanisms
  - Examples of some specific attacks
- Steganography & Steganalysis
  - What are they and How they work
  - Problem Model
  - Steganalysis category

What is a Watermark?

- A watermark is a “secret message” that is embedded into a “cover (original or host) message”.

- Only the knowledge of a secret key allows us to extract the watermark from the cover message.

- Effectiveness of a watermarking algorithm is a function of its
  - Resilience to attacks.
  - Capacity.
  - Stealth.

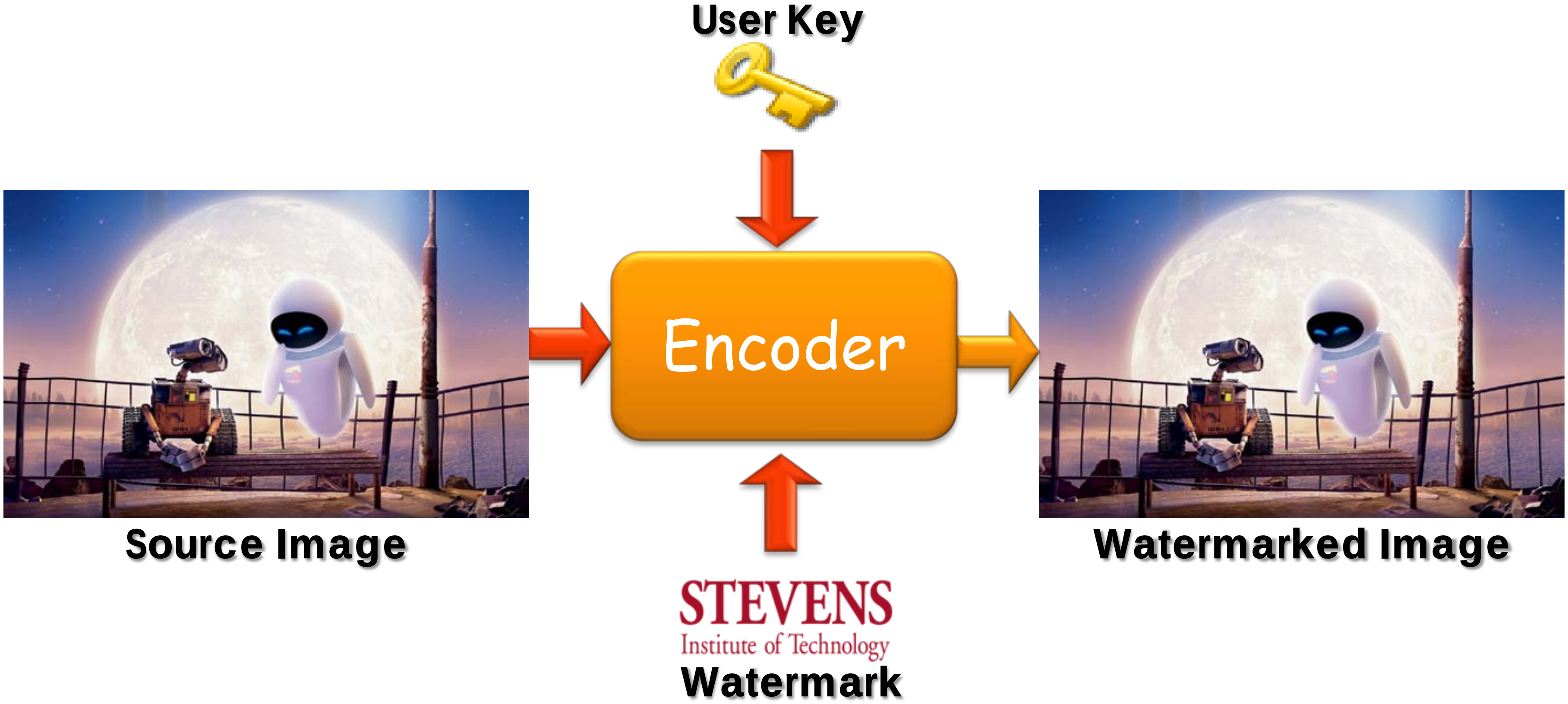
Media elements

- Multimedia data.
  - Video.
  - Audio.
  - Still Images.
  - Documents.
- Software.
- Hardware designs.

Multimedia Watermarks

- A digital watermark is a “secret key dependent” signal “inserted” into digital multimedia data.
- Watermark can be later detected/extracted in order to make an assertion about the data.
- A digital watermark can be.
  - Visible (perceptible).
  - Invisible (imperceptible).

Watermarking encoding process



User Key

Encoder

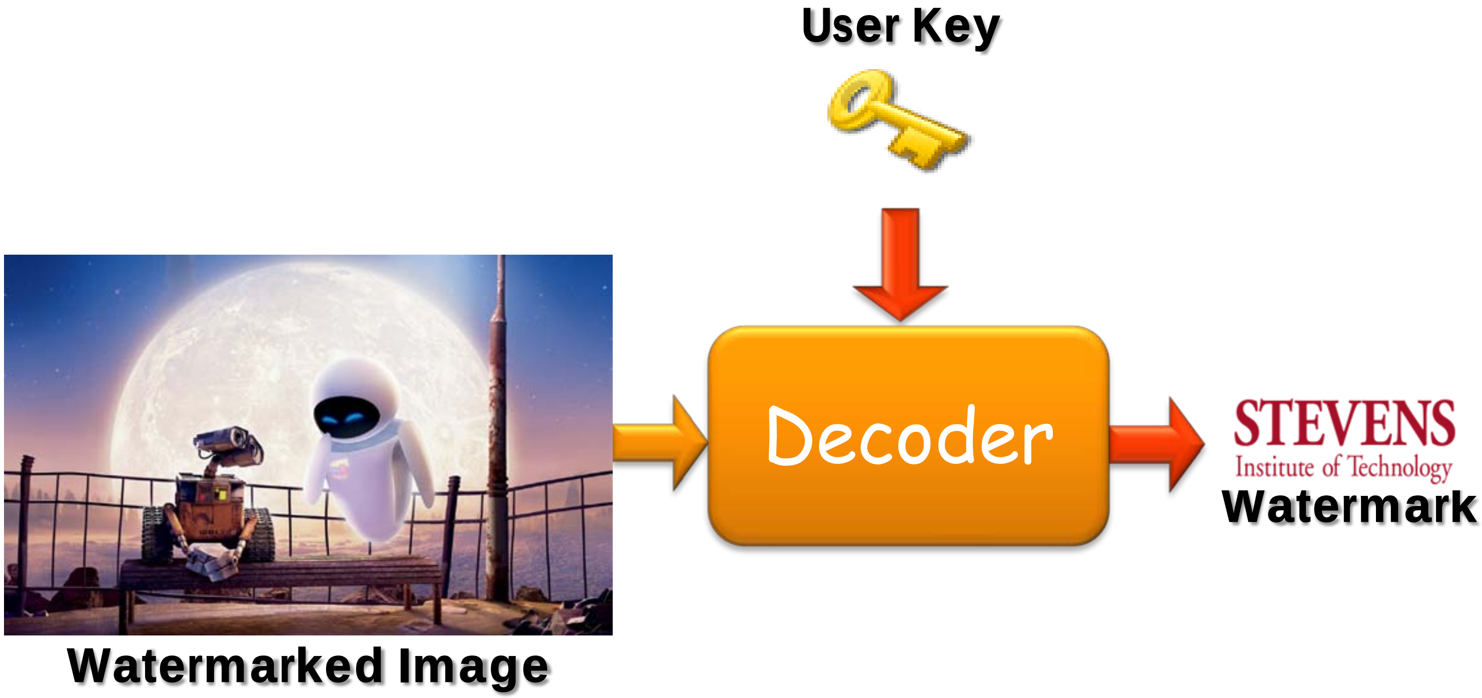
Source Image

Watermarked Image

STEVENS Institute of Technology

Watermark

Watermarking decoding process



User Key

Decoder

Watermark

Watermarked Image

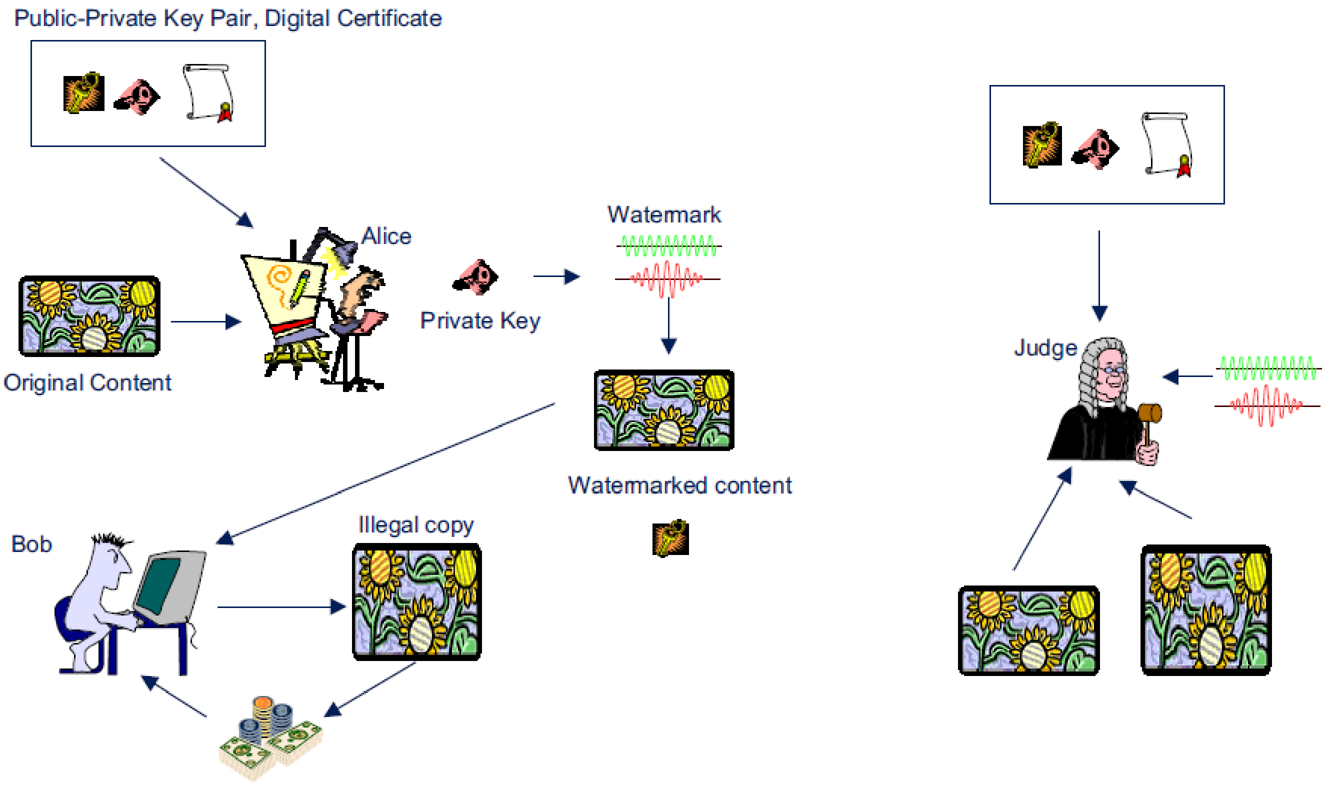
Watermarking applications

- Authentication.
  - Detect if image/video has been altered.
  - Digital cameras.
- Media Bridging.
  - Bridge media such as magazines and the Internet.
  - Digimarc.
- Broadcast Monitoring.
  - Keep track of when and where an advertisement is played.
  - ConfirMedia from Verance.

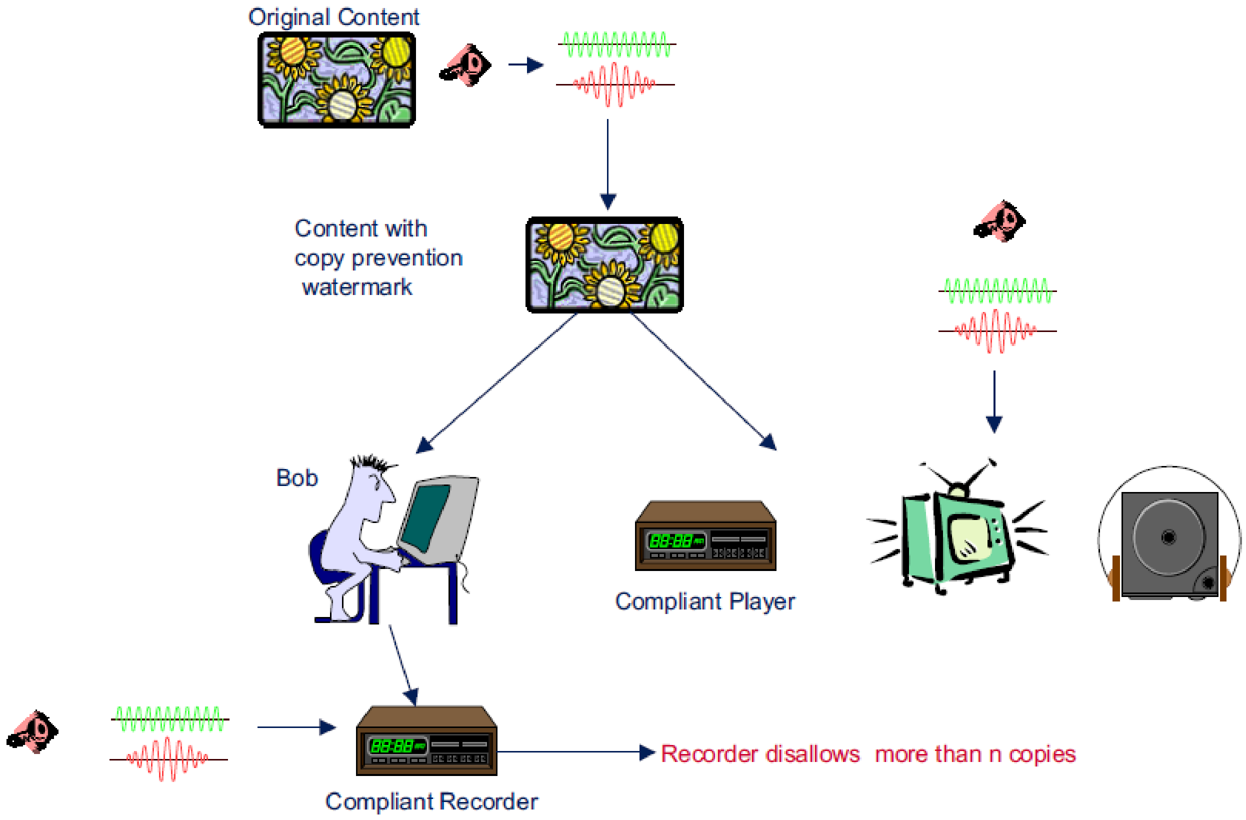
Watermarking applications

- Fingerprinting.
  - Identify the source of an illegal copy.
  - Unique watermark embedded in each copy.
  - DiVX, a modified version of DVD.
- Secret Communications.
  - Hide information such that general public do not know its presence.

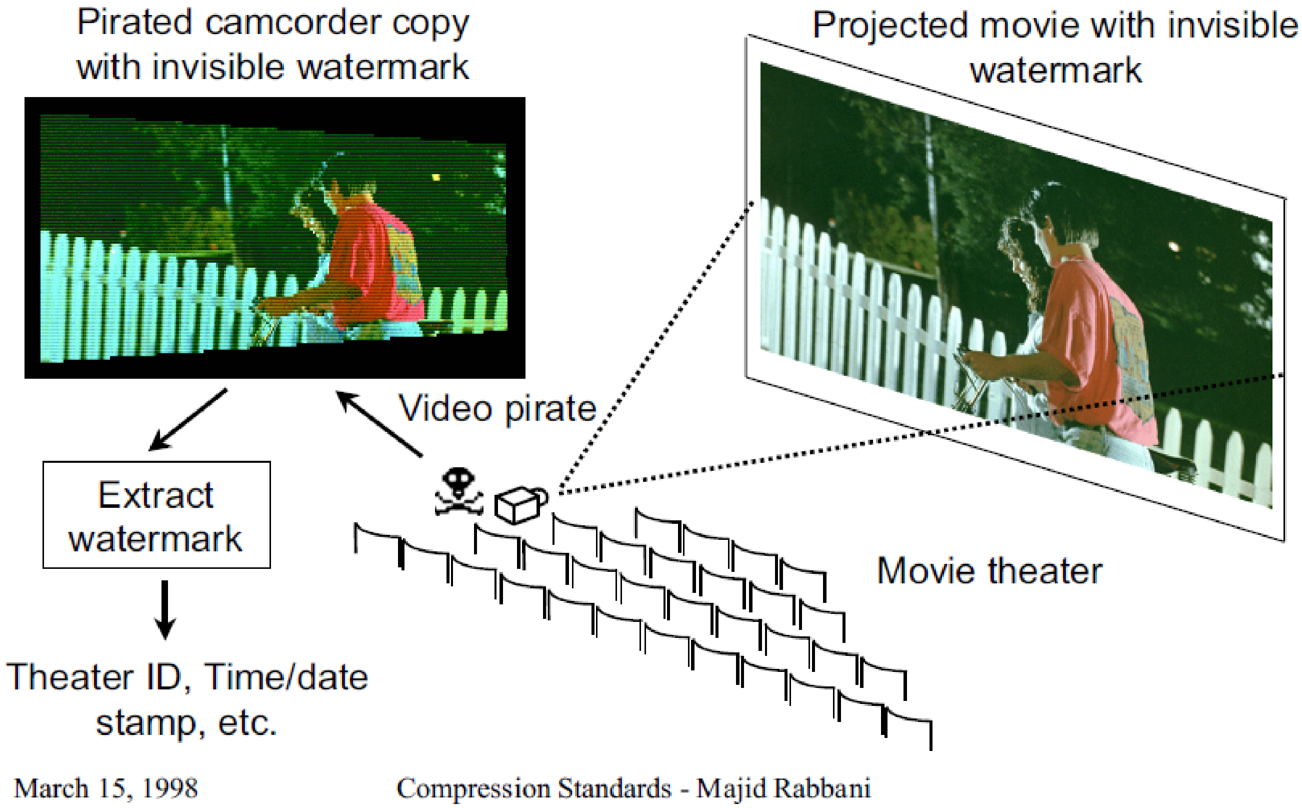
Ownership assertion



Fingerprinting for copy deterrence


Original Content
Content with copy prevention watermark
Bob
Compliant Player
Compliant Recorder
Recorder disallows more than n copies

Fingerprinting for digital cinema



Pirated camcorder copy with invisible watermark

Projected movie with invisible watermark

Video pirate

Extract watermark

Theater ID, Time/date stamp, etc.

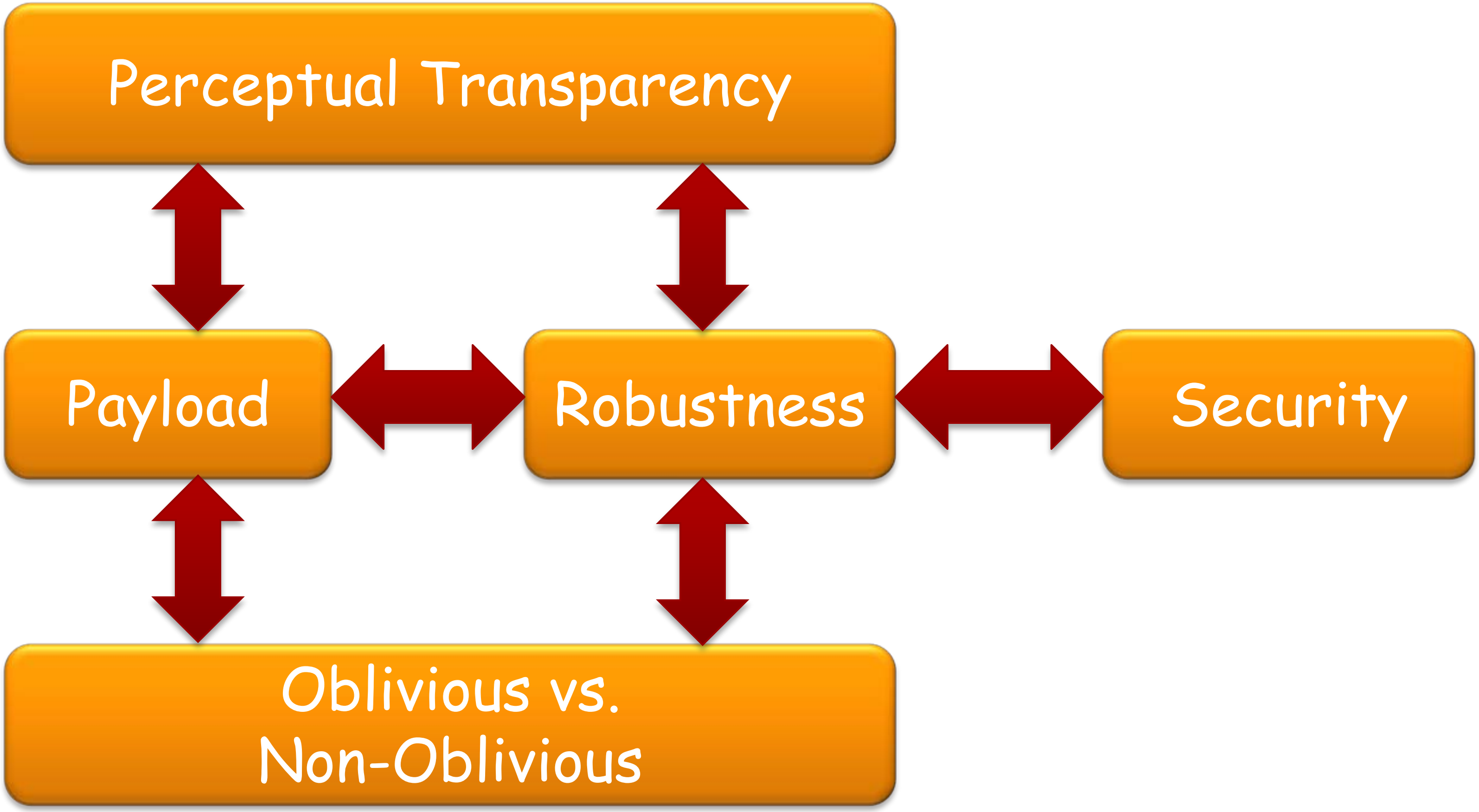
Movie theater

March 15, 1998 Compression Standards - Majid Rabbani

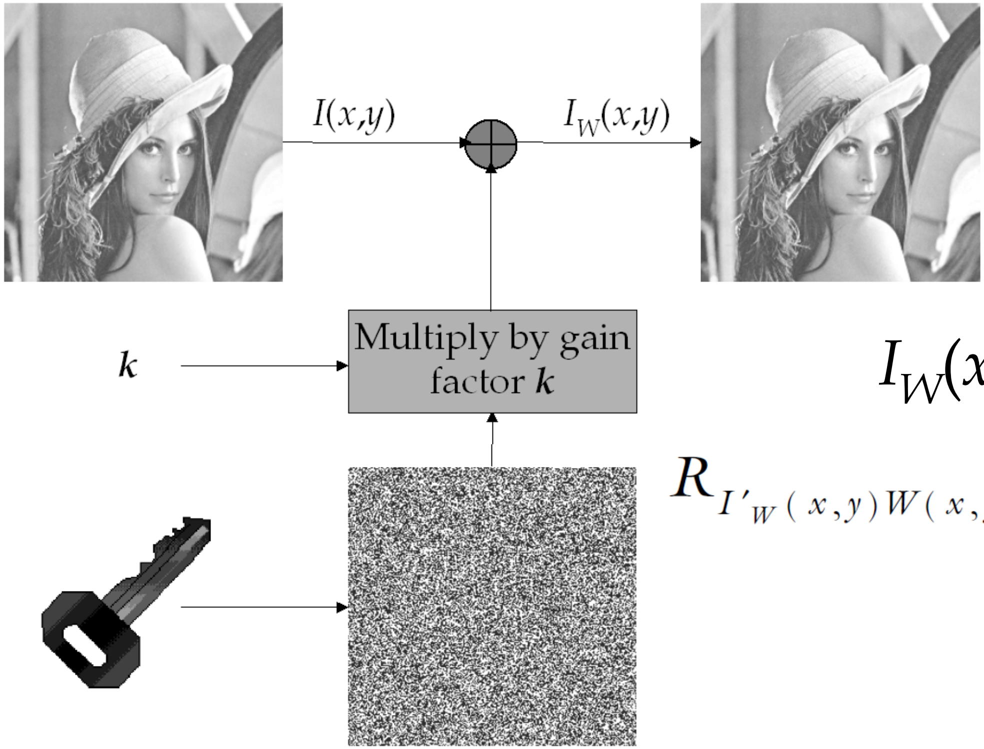
Watermarking requirement

- Requirements vary with application. For example:
  - Perceptually transparent - must not perceptually degrade original content.
  - Robust - survive accidental or malicious attempts at removal.
  - Oblivious or Non-oblivious - Recoverable with or without access to original.
  - Capacity – Number of watermark bits embedded.
  - Efficient encoding and/or decoding.

Contradicting Requirements



Example: Additive Watermarks

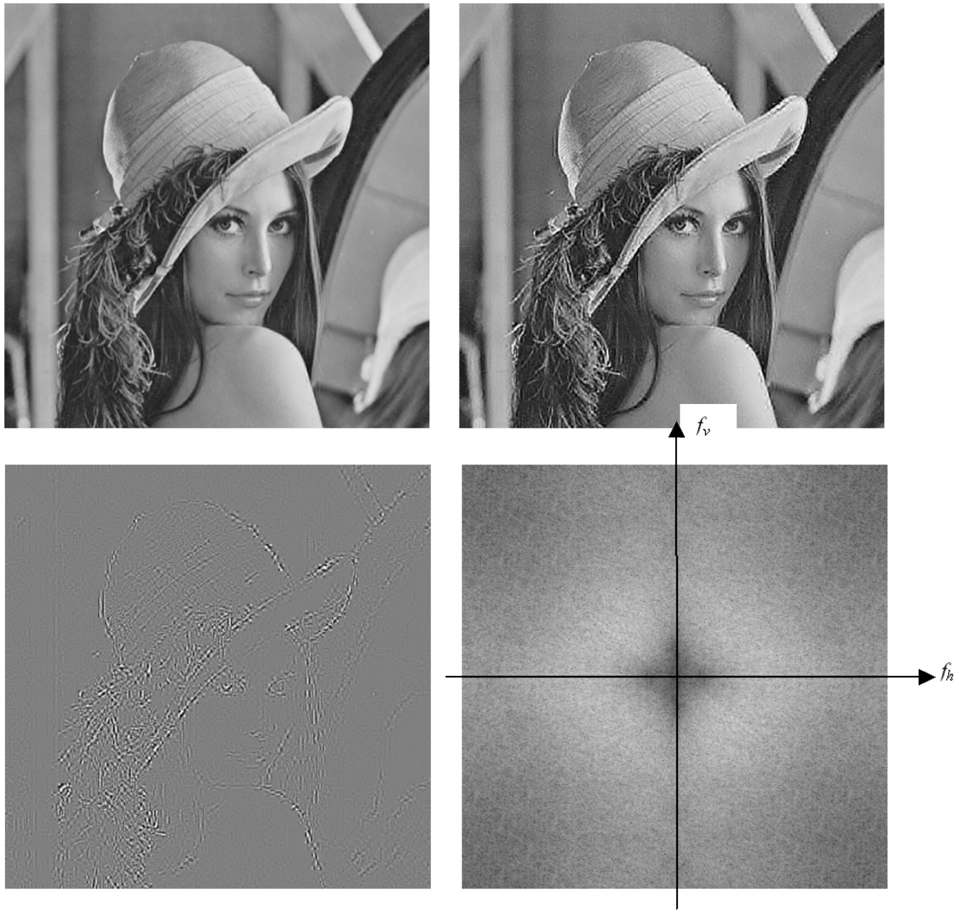


$W(x,y)$: Pseudo Random Pattern {-1,0,1}

$$I_W(x,y) = I(x,y) + k \cdot W(x,y)$$

$$\begin{array}{llll} R_{I'_W(x,y)W(x,y)} & > T & \rightarrow & W(x,y) \text{ detected} \\ & < T & \rightarrow & \text{No } W(x,y) \text{ detected} \end{array}$$

Additive watermarks



Watermark Attacks

- Active Attacks.
  - Hacker attempts to remove or destroy the watermark.
  - Watermark detector unable to detect watermark.
  - Key issue in proof of ownership, fingerprinting, copy control.
  - Not serious for authentication or covert communication.

Watermark Attacks

- Passive Attacks.
  - Hacker tries to find if a watermark is present.
  - Removal of watermark is not an aim.
  - Serious for covert communications.
- Collusion Attacks.
  - Hacker uses several copies of watermarked data (images, video etc.) to construct a copy with no watermark.
  - Uses several copies to find the watermark.
  - Serious for fingerprinting applications.

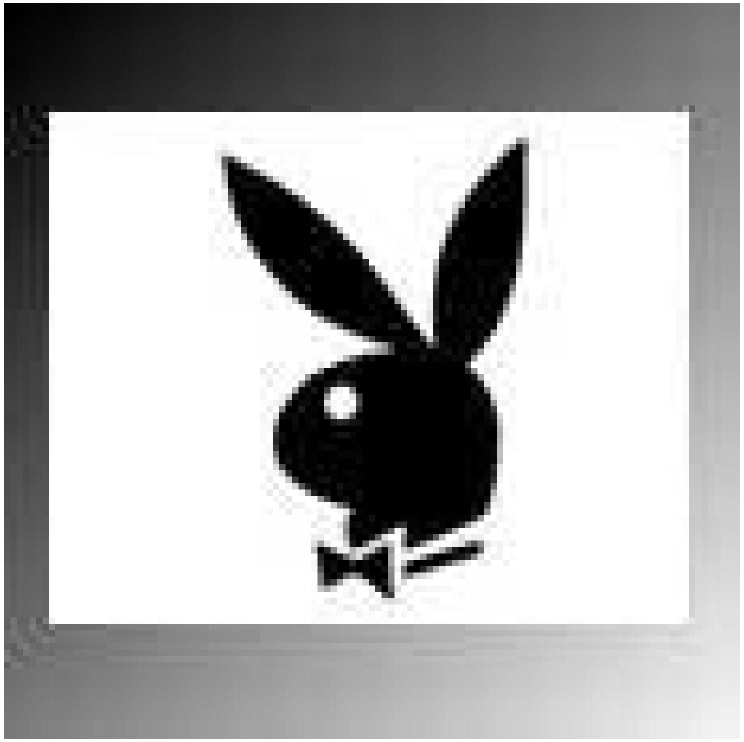
Watermark Attacks

- Forgery Attacks.
  - Hacker tries to embed a valid watermark.
  - Serious in authentication.
  - If hacker embeds a valid authentication watermark, watermark detector can accept bogus or modified media.

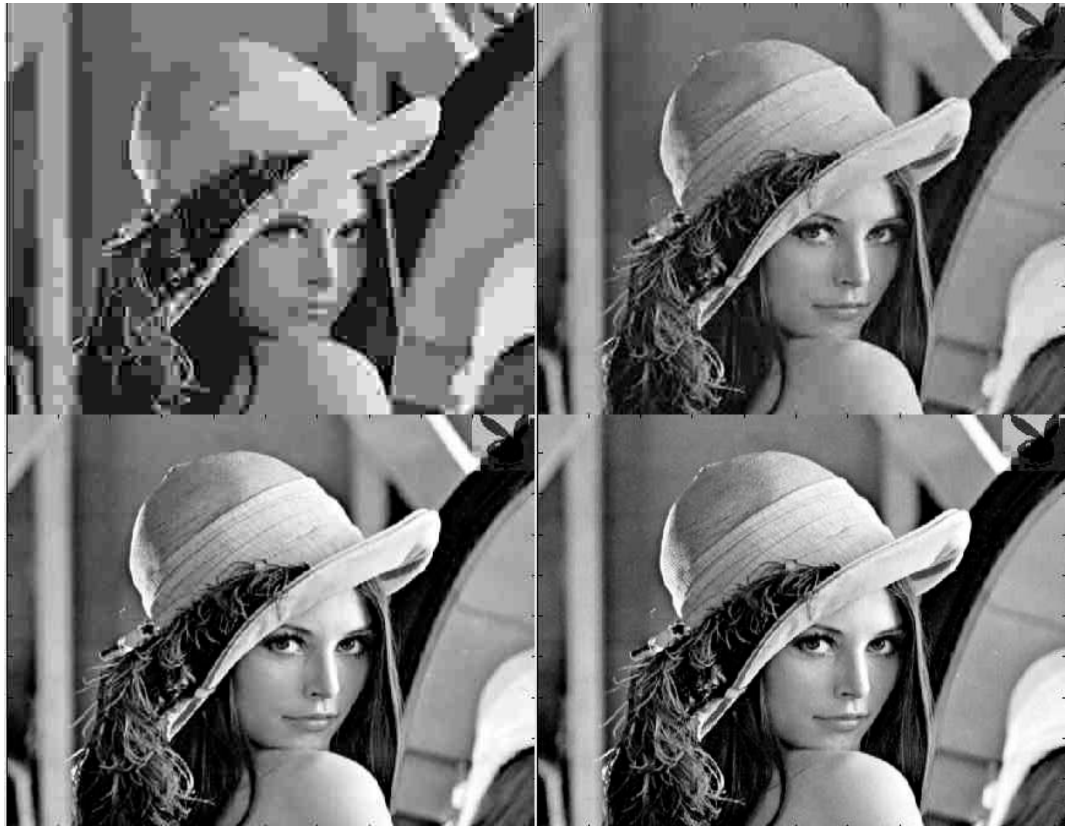
Content-based Watermarking

Original image

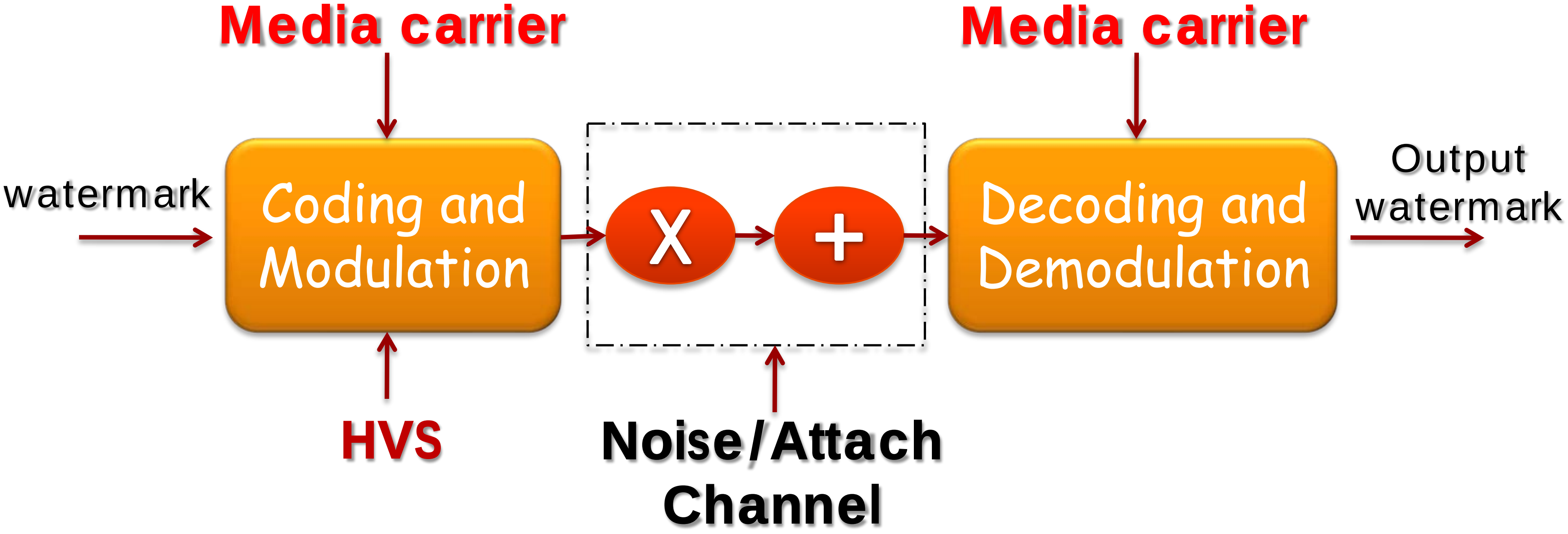
Watermark



JPEG Compression Attack



Watermarking as a communication system



Steganography and steganalysis

- Steganography:
  - the practice of hiding private or sensitive information within something that appears to be nothing out of the usual
- Steganalysis:
  - detect and/or estimate potentially hidden information from observed data with little or no knowledge about the steganography algorithm and/or its parameters

Steganography Terms

- Carrier File: a file which has hidden information inside of it.
- Stego-Medium: the medium in which the information is hidden.
- Redundant Bits: pieces of information inside a file which can be overwritten or altered without damaging the file.
- Payload: the information which is the be concealed.

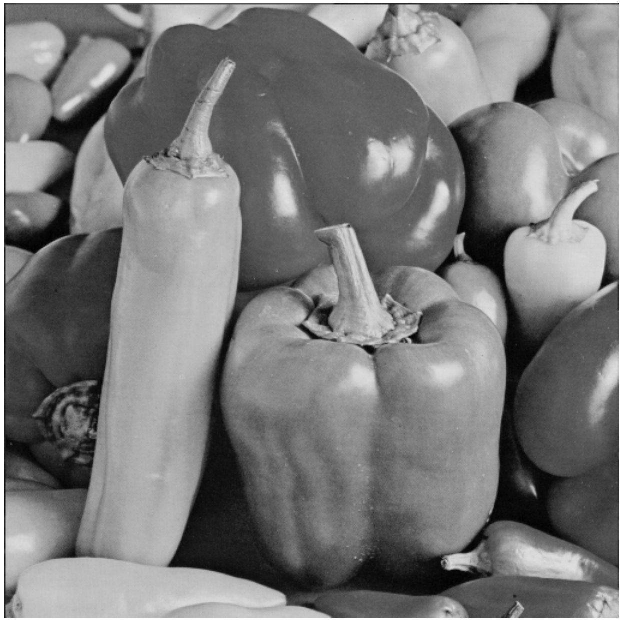
Steganography Methods

- BMP image based methods:
  - LSB
  - QIM
  - Spread spectrum
- JPEG image based methods:
  - Outguess
  - Steghide
  - F5
  - JP hide & seek
  - Perturbed quantization

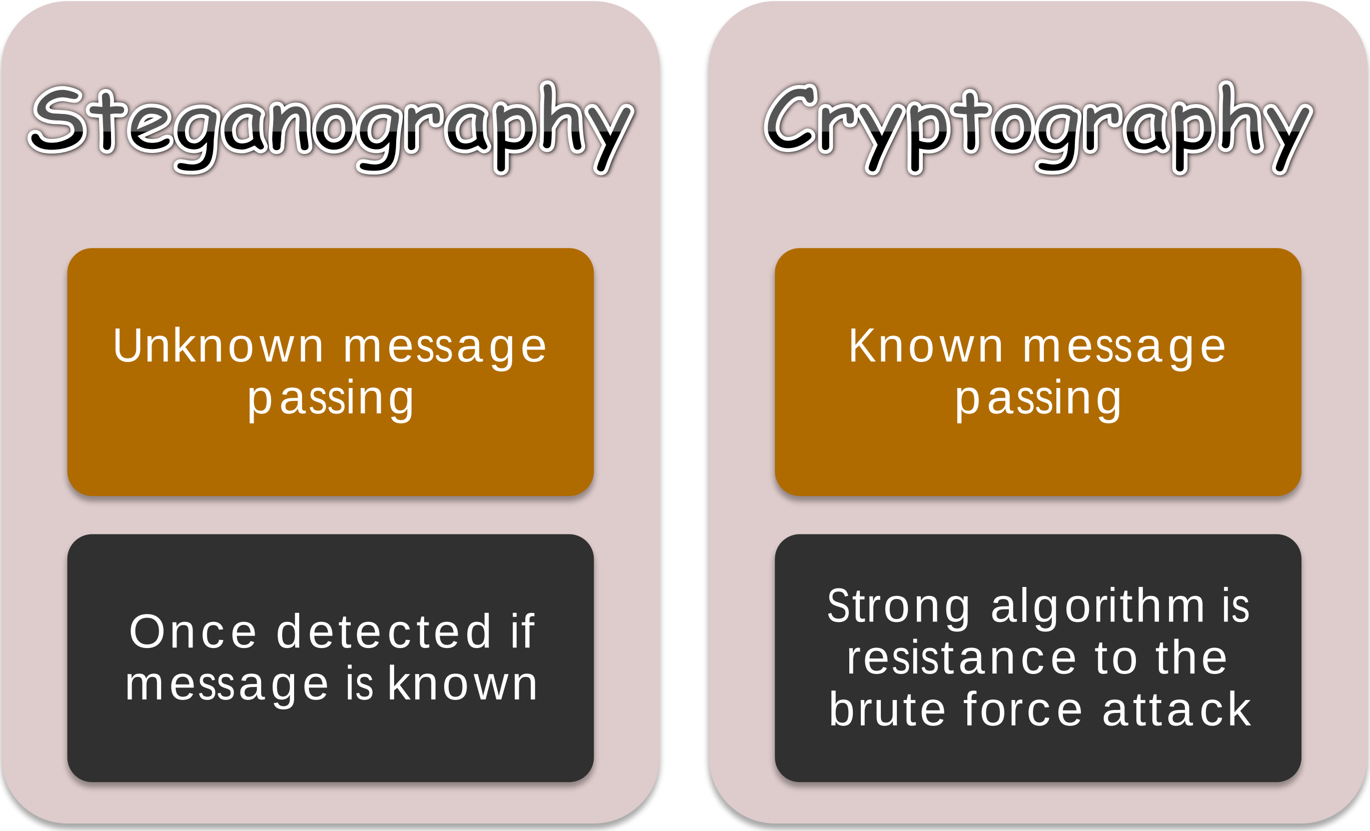
Example: LSB Encoding

Original image

Stego image



Steganography vs. cryptography



Steganography

Unknown message passing

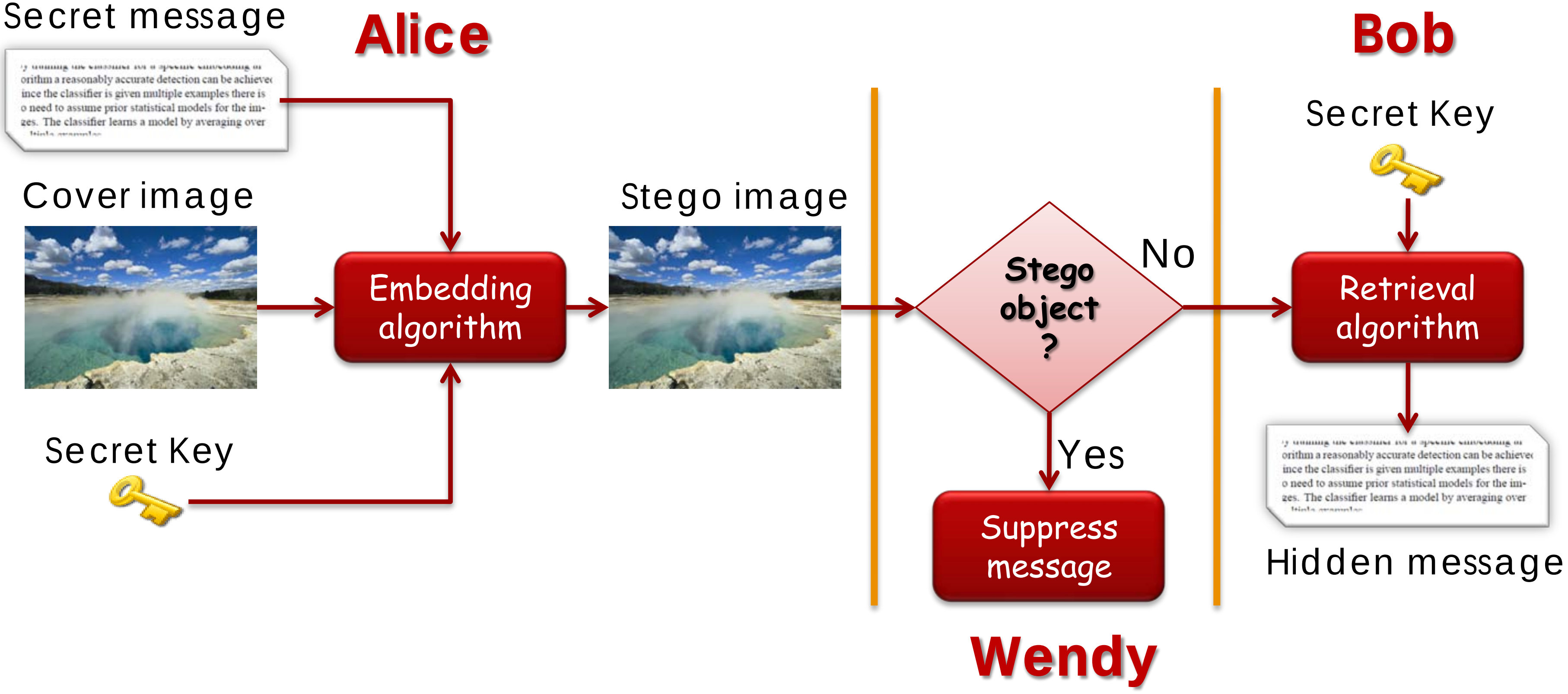
Once detected if message is known

Cryptography

Known message passing

Strong algorithm is resistance to the brute force attack

Framework for Secret Key Passive Warden Steganoraphy


Secret message
orithm a reasonably accurate detection can be achieve
ince the classifier is given multiple examples there is
o need to assume prior statistical models for the im-
ges. The classifier learns a model by averaging over
Alice
Cover image
Embedding algorithm
Stego image
Secret Key
Stego object ?
No
Yes
Suppress message
Wendy
Bob
Secret Key
Retrieval algorithm
orithm a reasonably accurate detection can be achieve
ince the classifier is given multiple examples there is
o need to assume prior statistical models for the im-
ges. The classifier learns a model by averaging over
Hidden message

Steganographic Security

- A steganographic system is considered to be insecure if the warden Wendy is able to prove the existence of a secret message.
- Cachin's security criteron: let P_C *denote* the probability distribution of cover-objects and P_S *denote the probability distribution* of stego-objects.

$$D(P_C||P_S) = \int P_C \cdot log\frac{P_C}{P_S} \leq \epsilon$$

Steganalysis classification based on the outcome

- *Passive steganalysis: Detect the presence or absence of a secret message in an observed message.*

- *Active steganalysis: Extract a (possibly approximate) version of the secret message from a stego* message.

Steganalysis classification based on information types

- Spatial diversity information based steganalysis:
  - Steganography methods could spread informationin the spatial domain and this information repeats itself in various forms in different spatial locations
- *Temporal diversity information based steganalysis:*
  - *Steganography information that appears repeatedly* over time can also aid steganalysis.

Steganalysis classification based on techniques

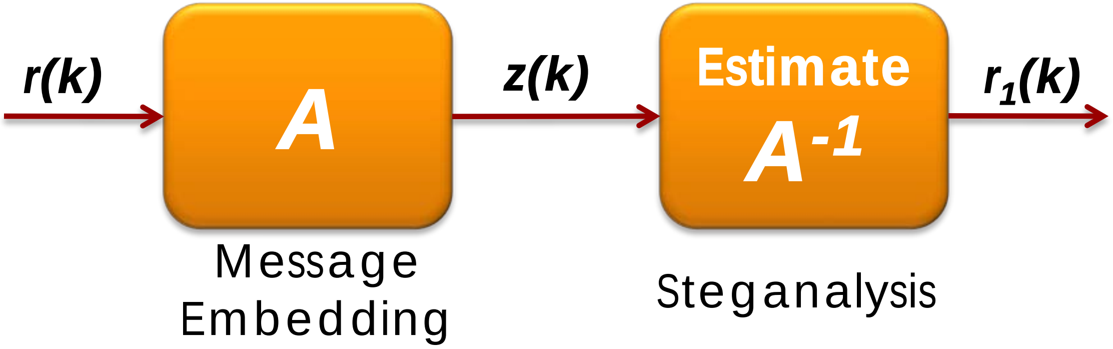
- Supervised learning based steganalysis
- Blind identification based steganalysis
- Parametric statistical steganalysis
- Hybrid techniques

Supervised learning based steganalysis

- Supervised learning methods construct a classifier to differentiate between stego and non-stego images using training examples.
- Some image features are first extracted and given as training inputs to a learning machine. These examples include both stego as well as non-stego messages
- The learning classifier iteratively updates its classification rule based on its prediction and the ground truth. Upon convergence the final stego classifier is obtained.

Blind identification based steganalysis

- Let ***z(k)*** *denote a random stego message vector observed* by the steganalyst, ***A*** be a representation of the embedding algorithm in matrix form, and ***r*** is the vector with the cover message and the secret message as its components. The steganalyst is now faced with the problem of inferring $\boldsymbol{A^{-1}}$ from ***z(k)****.*



Parametric statistical steganalysis

- Completely known statistics
- Partially known statistics
- Completely unknown statistics

Steganalysis measurement

- Accuracy
- Consistency
- Minimize false-positives