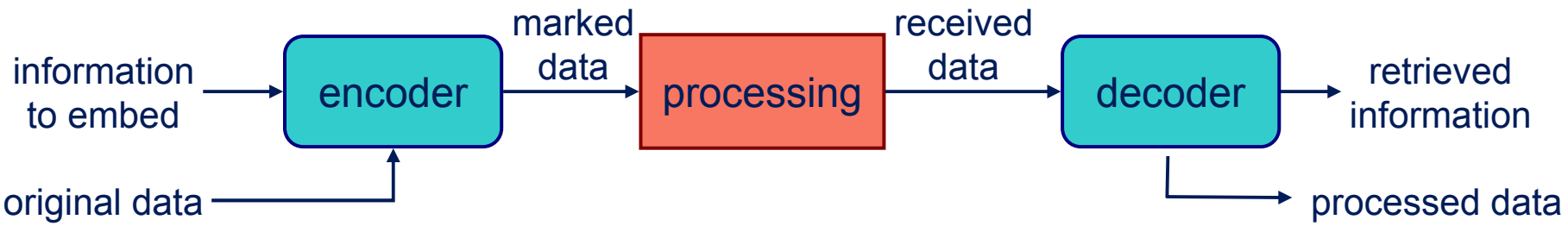
Information Hiding, Digital Watermarking and Steganography

An Introduction to Basic Concepts and Techniques

Nasir Memon
Polytechnic University, Brooklyn

Information Hiding

- Information Hiding: Communication of information by embedding it in and retrieving it from other digital data.
- Depending on application we may need process to be imperceptible, robust, secure. Etc.



Where can we hide?

- Media
  - » Video
  - » Audio
  - » Still Images
  - » Documents
- Software
- Hardware designs
- Etc.
- We focus on data hiding in media.
- We mainly use images but techniques and concepts can be suitably generalized to other media.

Why Hide?

- Because you want to protect it from malicious use
  - » Copy protection and deterrence - Digital Watermarks
- Because you do not want any one to even know about its existence
  - » Covert communication – Steganography
- Because it is ugly
  - » Media bridging,
  - » Meta Data embedding
- To get a free ride
  - » Hybrid digital analog communication, captioning.

Fundamental Issues

- **Fidelity**

  The degree of perceptual degradation due to embedding operation.

- **Robustness**

  The level of immunity against all forms of manipulation (intentional and non-intentional attacks).

- **Payload**

  The amount of message signal that can be reliably embedded and extracted (subject to perceptual constraints at the designated level of robustness).

- **Security**

  Perhaps the most misunderstood and ignored issue. Meaning of security depends on the application as we shall see later.

Classification Basis for Information Hiding Methods

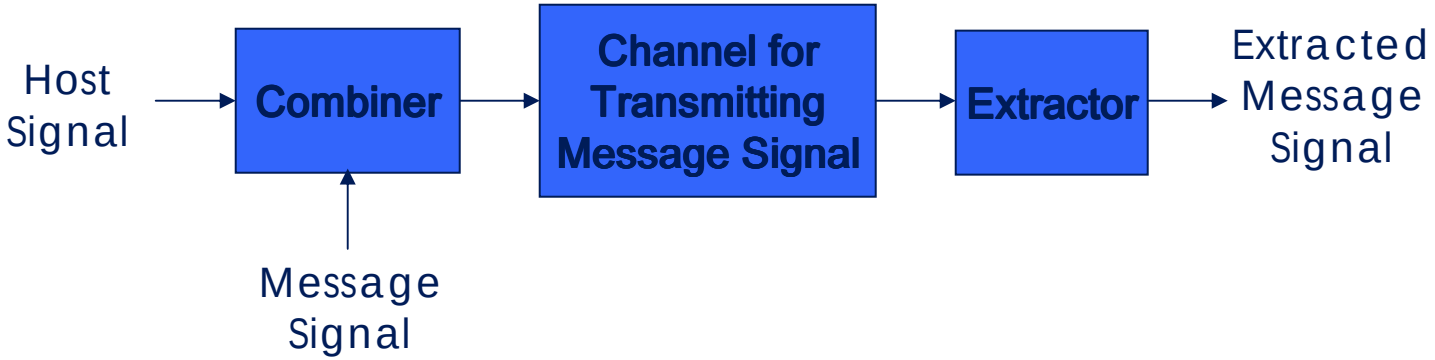
- The nature of *host signal, i.e.,* audio, video, image, text, programs, etc.
- Robust, fragile, semi-fragile.
- The need for *host signal* at message extraction–blind or non-blind (private or public).
- The type of communication–synchronous or asynchronous.
- The *threat model* – intentional (malicious) or non-intentional attacks.
  - » Digital Watermarking
  - » Steganography
  - » Data Hiding

Truly Interdisciplinary

- Information Theory and Communication
- Signal Processing and Transforms
- Game Theory
- Coding Theory
- Detection and Estimation Theory
- Cryptography and Protocol Design

A Communication Perspective

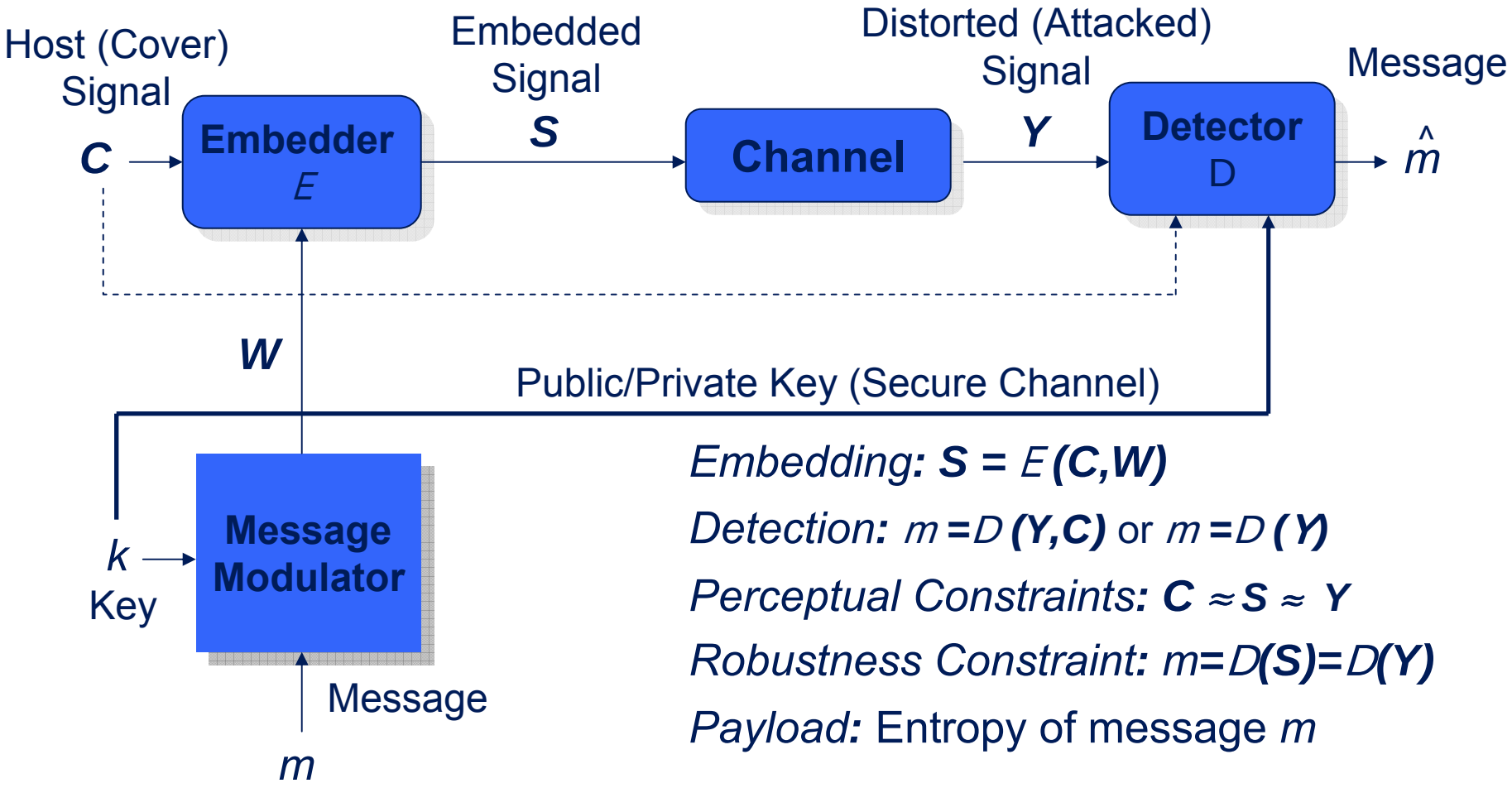
Communication model for Information Hiding



Information hiding differs from traditional communication systems in the operation of the combiner. *(Beyond modulation…)*

- i.e., in classical communications no "*similarity*" constraint is imposed on the carrier signal and its modulated version.

Generic Model & Terminology



Embedding: $\mathbf{S} = E(\mathbf{C},\mathbf{W})$

Detection: $m = D(\mathbf{Y},\mathbf{C})$ or $m = D(\mathbf{Y})$

Perceptual Constraints: $\mathbf{C} \approx \mathbf{S} \approx \mathbf{Y}$

Robustness Constraint: $m = D(\mathbf{S}) = D(\mathbf{Y})$

Payload: Entropy of message m

The Parallel Between Communication and Information Hiding Systems

Communication Framework	**Information Hiding Framework**
Encoder-Decoder	Embedder-Detector
Side information	Host signal
Channel noise	Attack
Power constraints	Perceptual distortion Limits
Signal to noise ratio (SNR)	Embedding distortion to attack distortion (WNR)

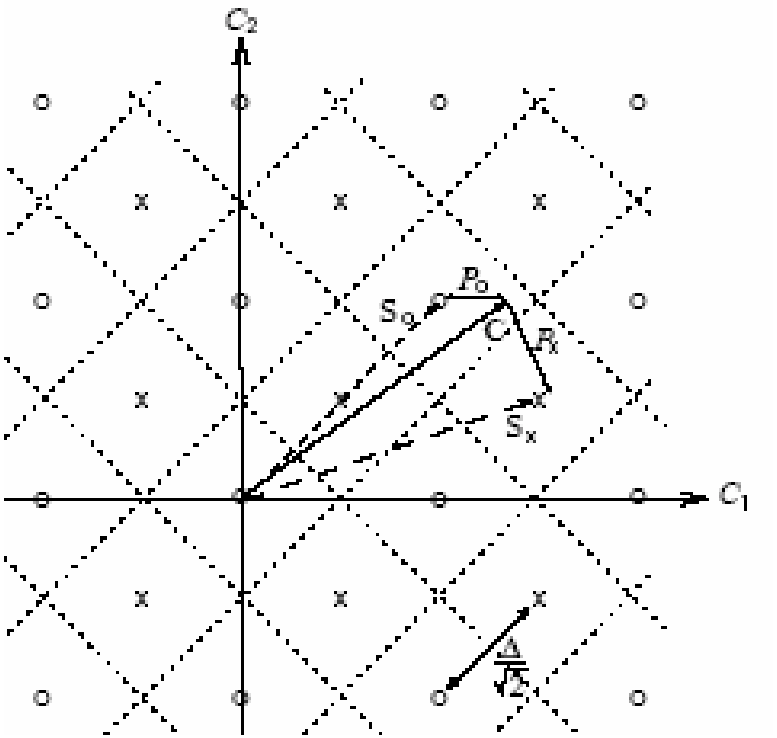
Goal: Designing codes for reliable communication between the two parties.

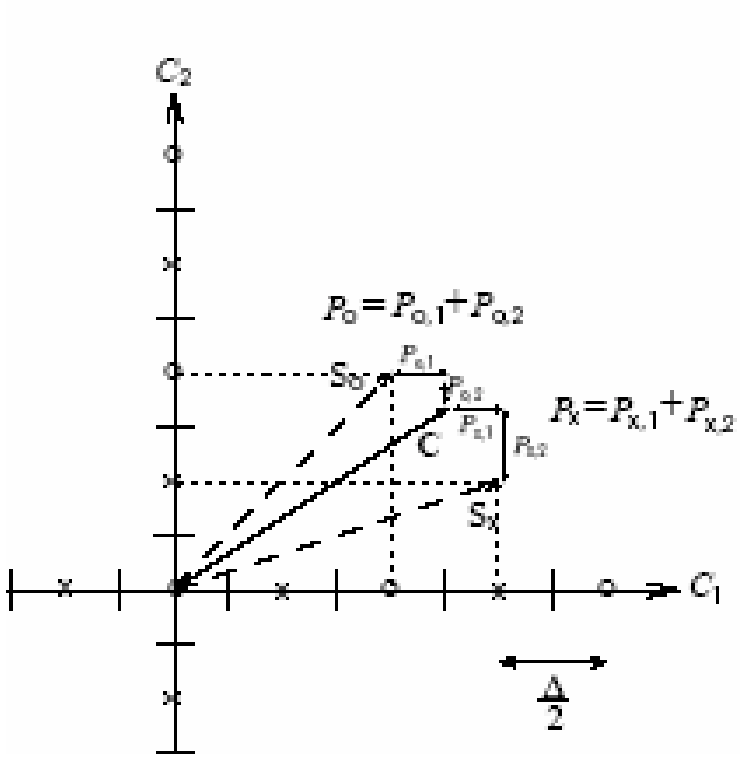
Lattice Quantization Case

- Example for $C \in \Re^2$
  - » Embedding a binary symbol by use of a two-dimensional lattice.
  - » Embedding two binary symbols by use of two unidimensional lattices.

o → m=0

× → m=1





Optimum Embedder/Detector Design

- Nested lattice codes provide an efficient algebraic structured binning scheme.
  - » A high dimensional *fine* lattice is partitioned into cosets of *coarse* lattices.
  - » Embedding is by quantizing *C* to the nearest lattice point in the coarse lattice.
  - » Detection is by quantizing *Y* to the nearest lattice point in the fine lattice.
  - » The embedding rate is designated by diluting the coset density in the *fine* lattice.

Optimum Embedder/Detector Design

- High dimensional constructions are not feasible.
- Lattices with simpler structures are employed.
  - » Cartesian products of low-dimensional lattices.
  - » Recursive quantization procedures.
    - – Trellis coded quantization
  - » Error correction codes.

Digital Watermarks

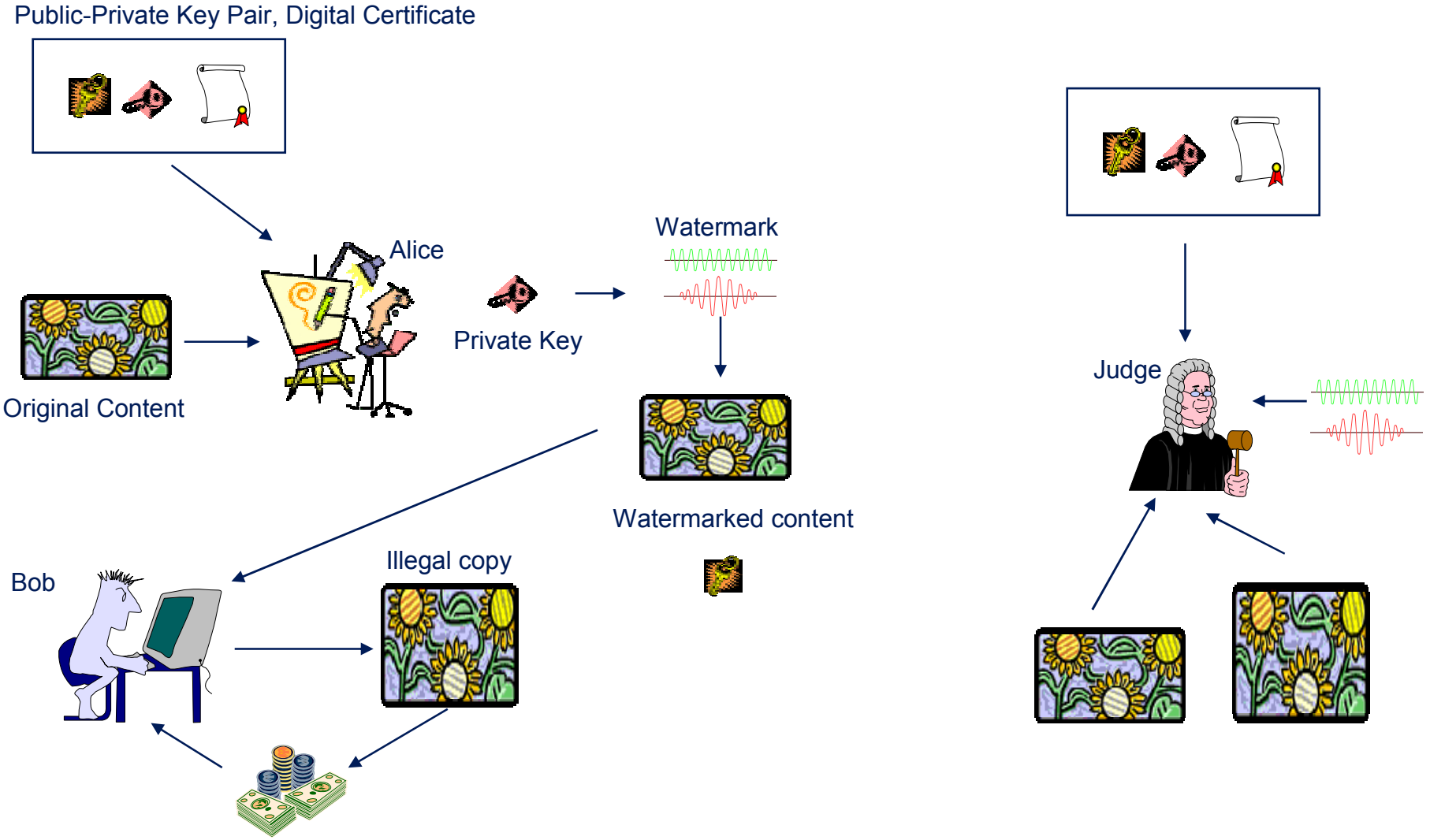
What is a Watermark?

- A watermark is a “secret message” that is embedded into a “cover message”.
- Usually, only the knowledge of a secret key allows us to extract the watermark.
- Has a mathematical property that allows us to argue that its presence is the result of deliberate actions.
- Effectiveness of a watermark is a function of its
  - » Stealth
  - » Resilience
  - » Capacity

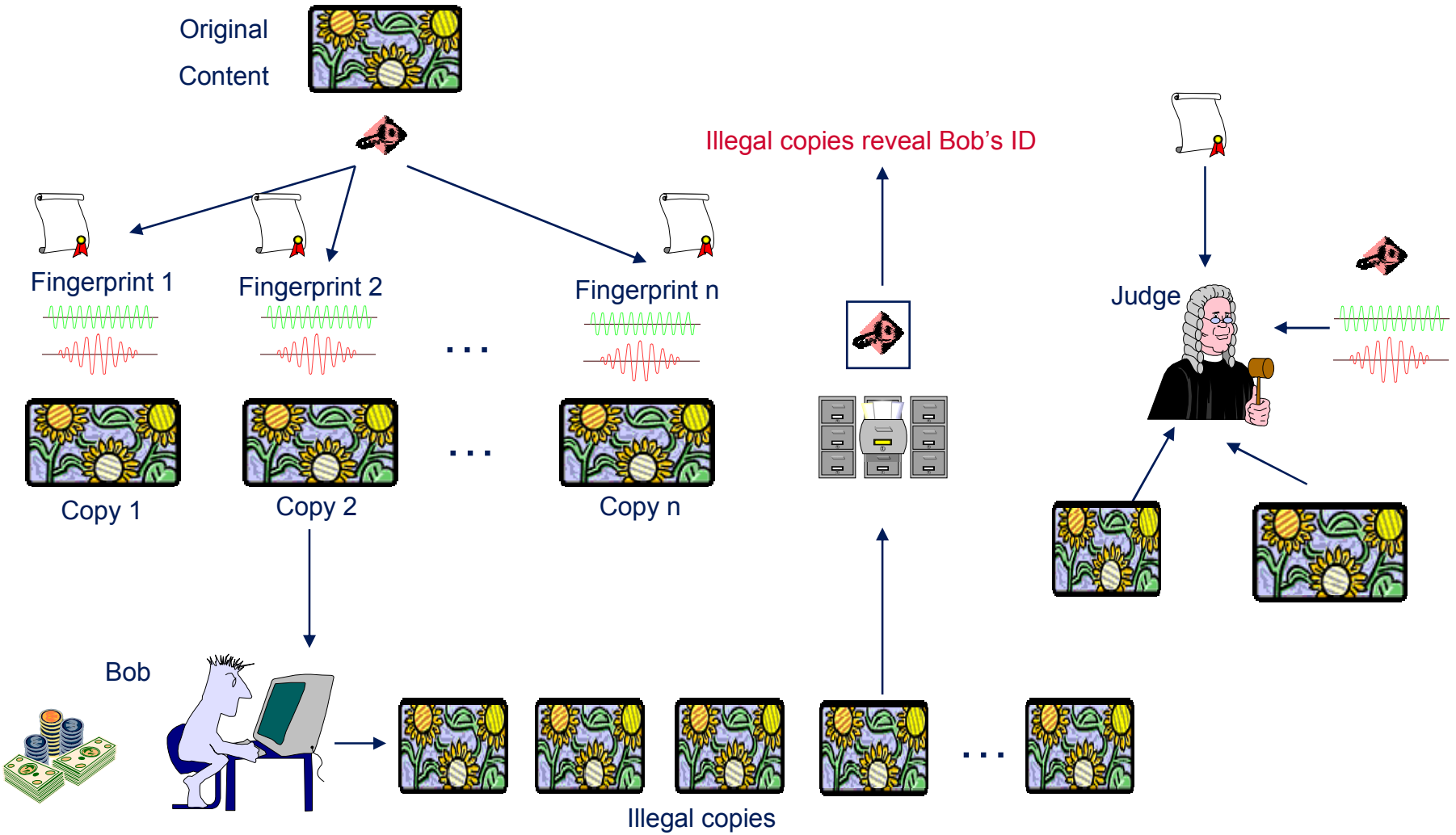
Why Watermark?

- Ownership assertion.
- Fingerprinting.
- Content labeling.
- Copy prevention or control.
- Content protection (visible watermarks).
- Authentication.
- Media Bridging
- Broadcast Monitoring
- Etc.

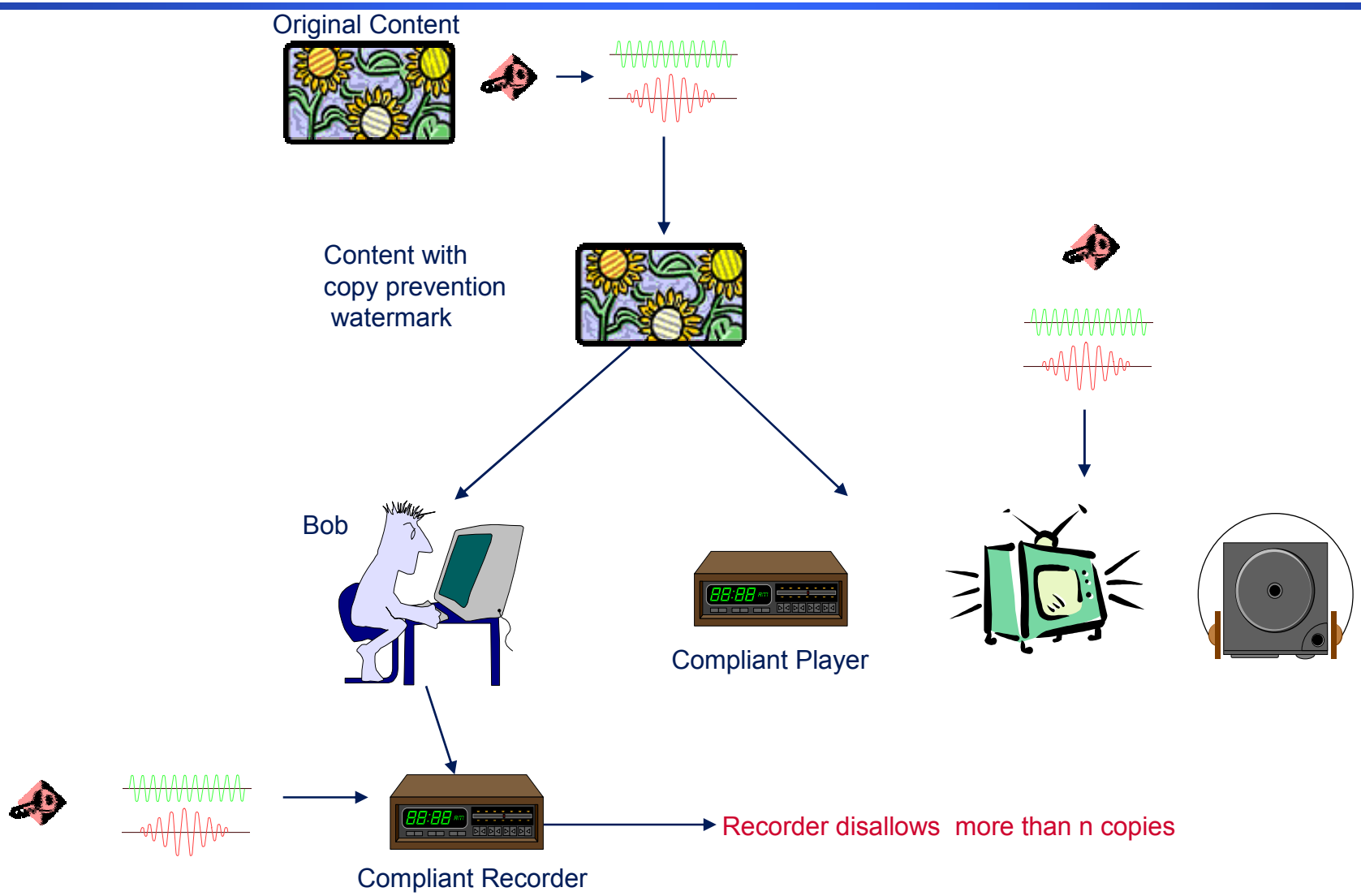
Ownership Assertion



Fingerprinting



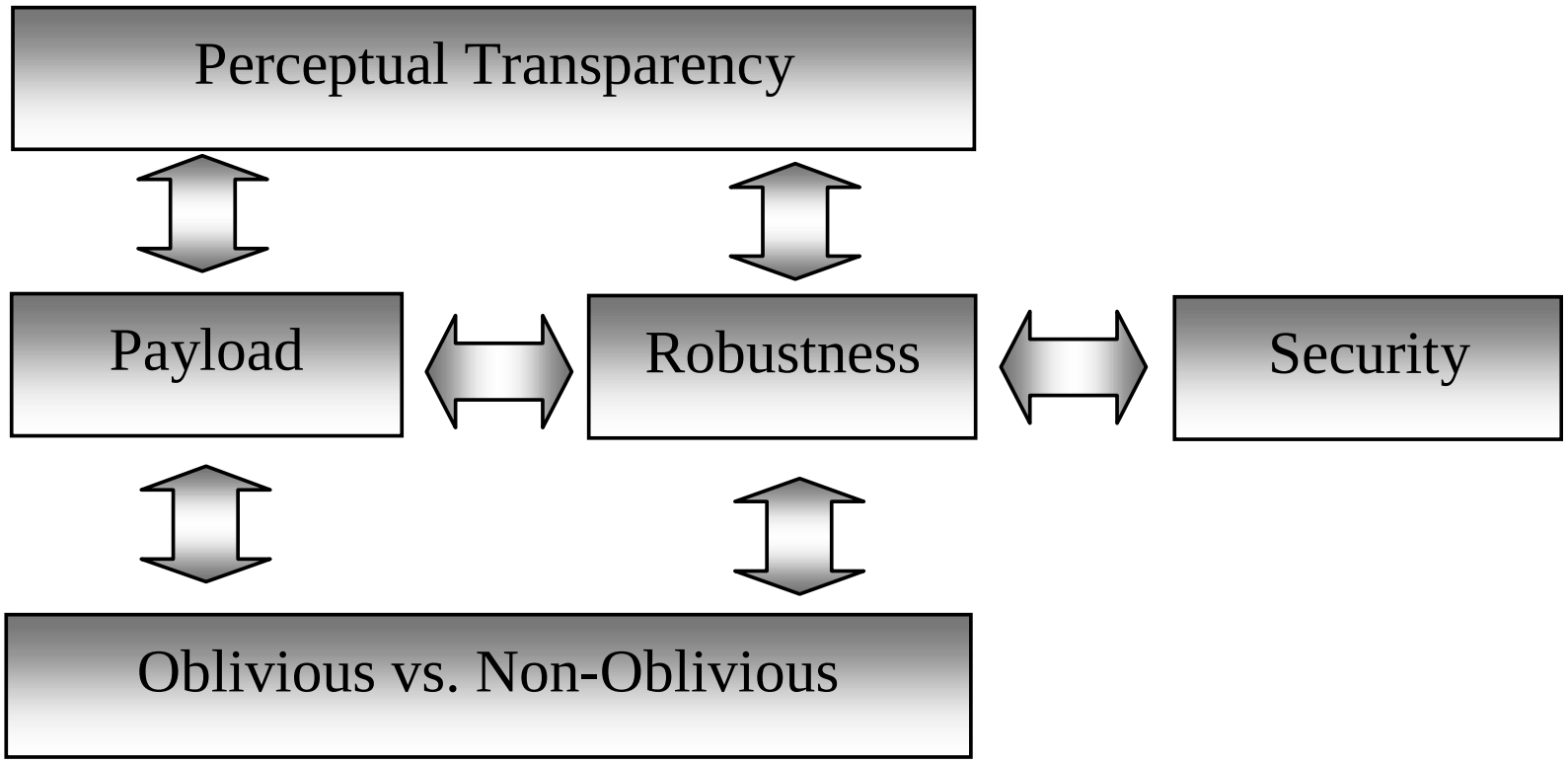
Copy Prevention and Control



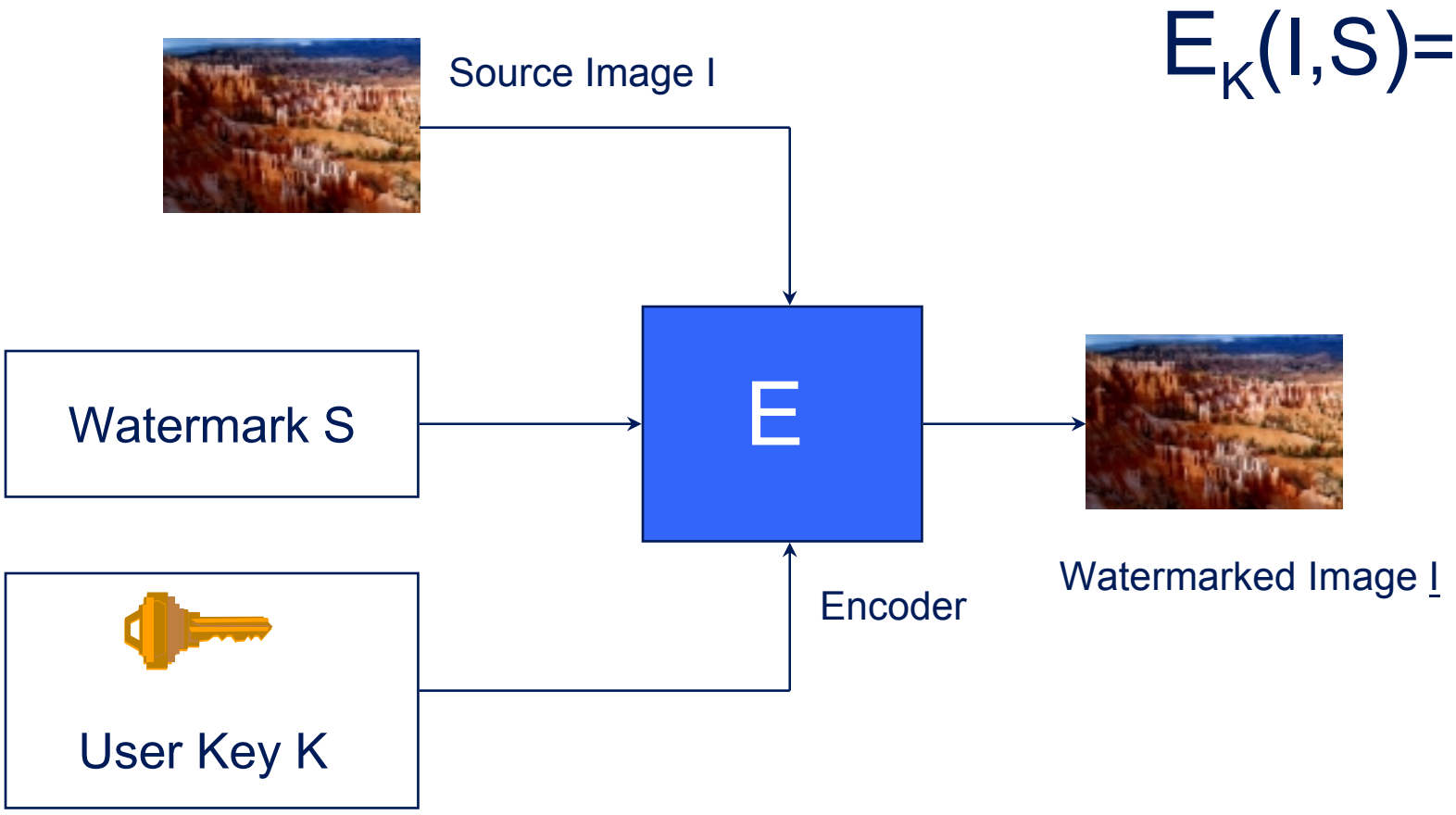
Requirements

- Requirements vary with application.
  - » Perceptually transparent - should 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.

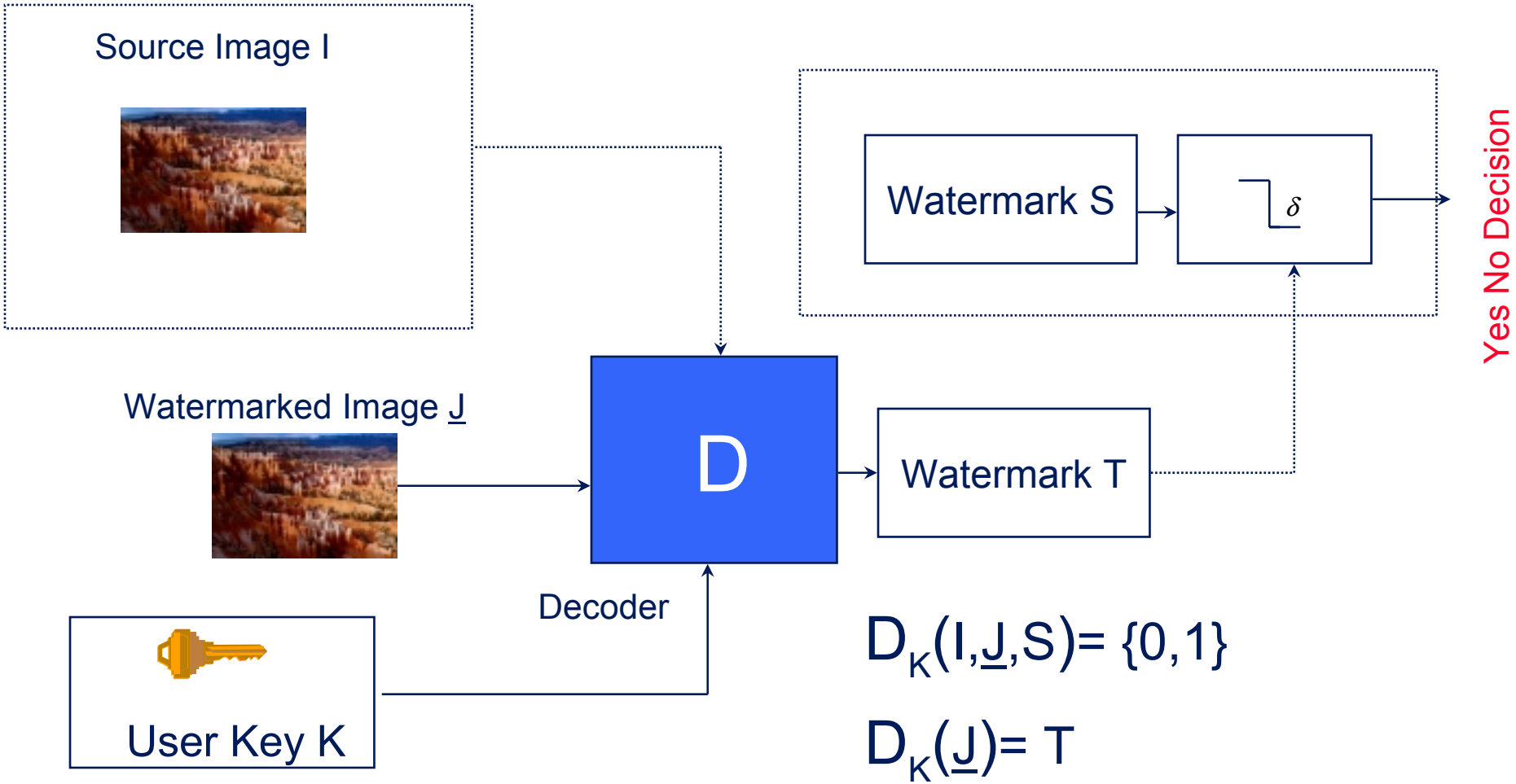
Requirements are Inter-related



Watermarking Encoding



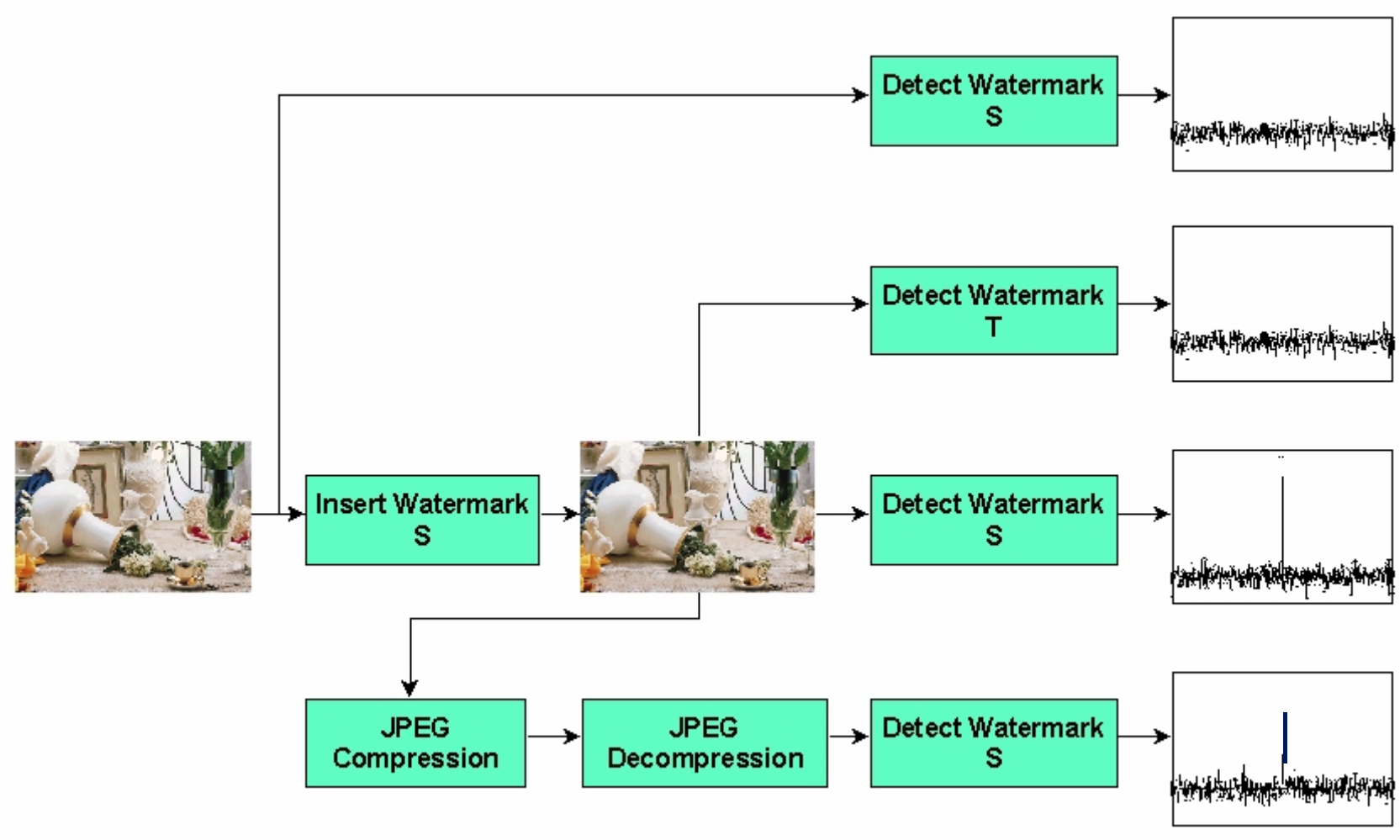
Watermarking Decoding



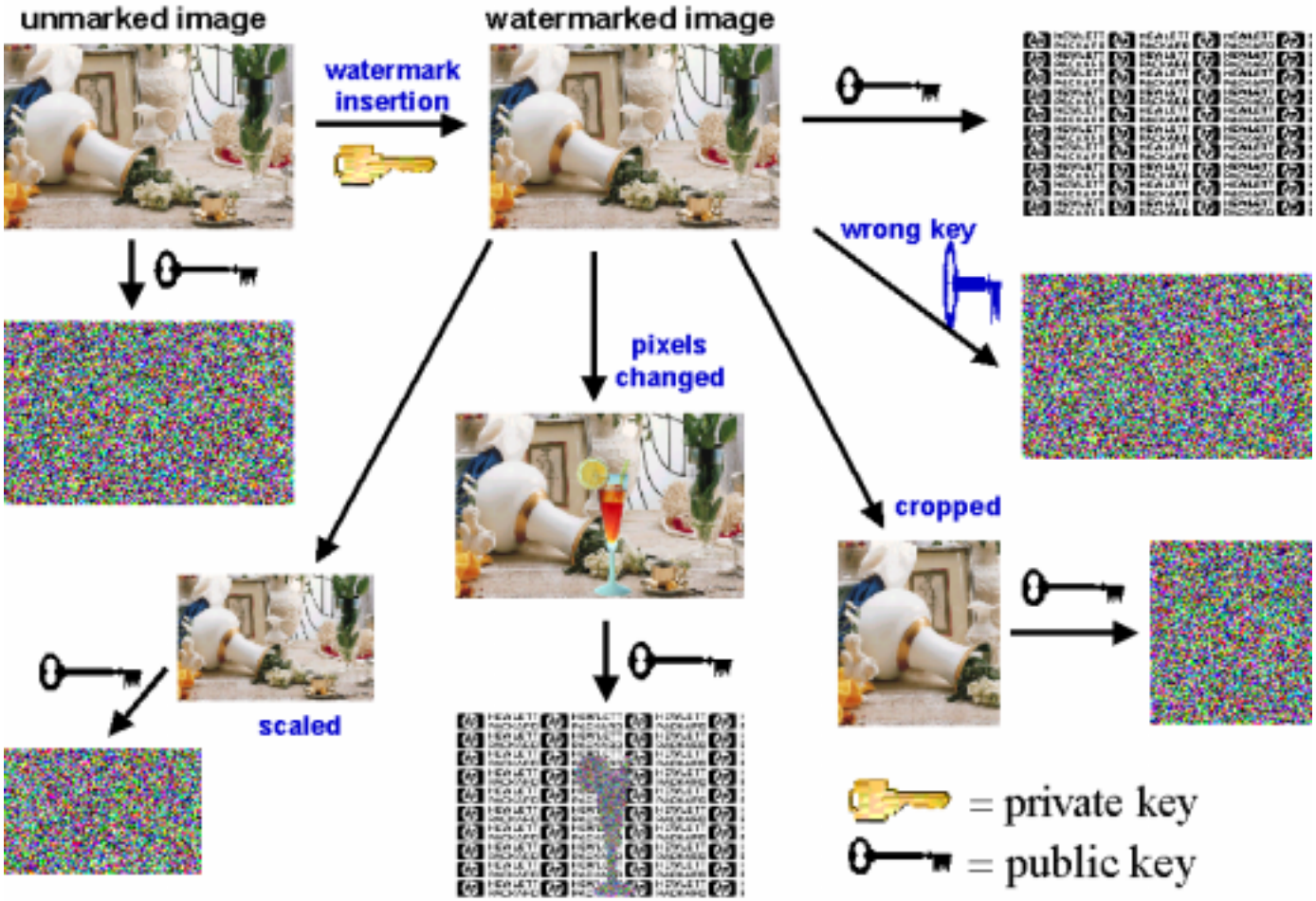
Classification

- According to method of insertion
  - » Additive
  - » Quantize and replace
- According to domain of insertion
  - » Transform domain
  - » Spatial domain
- According to method of detection
  - » Private - requires original image
  - » Public (or oblivious) - does not require original
- According to type
  - » Robust - survives image manipulation
  - » Fragile - detects manipulation (authentication)

Robust Watermarks



Fragile Watermarks



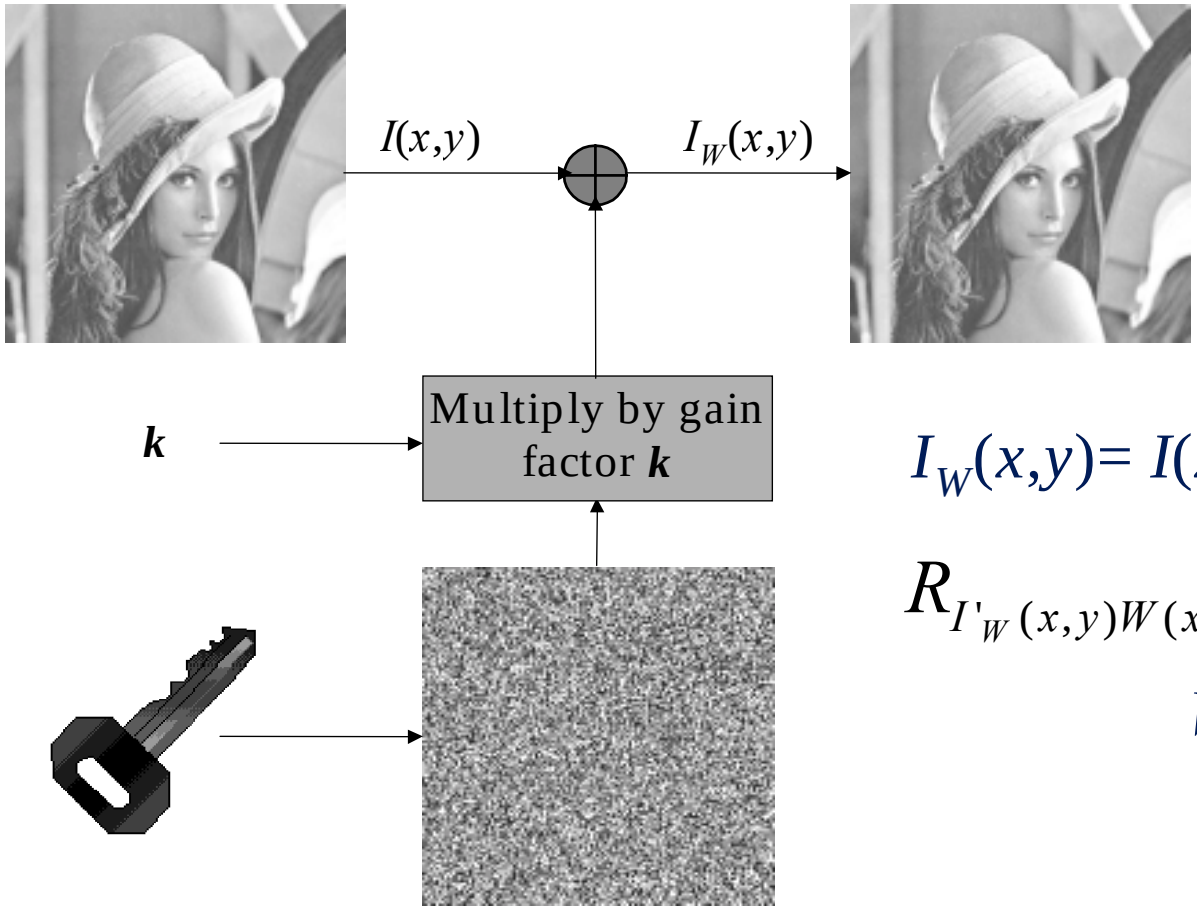
Detects and localizes any change to watermarked images.

Example of a Simple Spatial Domain Robust Technique

- Pseudo-randomly (based on secret key) select *n* pairs of pixels (a_i, b_i). The expected value of $\sum_{i=1}^{n} (a_i - b_i)$ is 0.
- Increase (a_i) by 1 and decrease (b_i) by 1. The expected value of $\sum_{i=1}^{n} (a_i - b_i)$ is 2n.
- To detect watermark, check $\sum_{i=1}^{n} (a_i - b_i)$

Example

Additive Watermarks



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

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

$$R_{I'_W(x,y)W(x,y)} > T \rightarrow$$

$W(x,y)$ detected

$$< T \rightarrow$$

No $W(x,y)$ detected

Additive Transform Domain Technique

- Embed sequence pseudo-randomly chosen iid Gaussian samples s_i into perceptually significant frequency components $f(I)$ of I (eg 1000 midband DCT coefficients).
- Example insertion formula. $f'_i = f_i + \alpha s_i$
- To detect s in J compute. $t_i = f_i(J) - f_i(I)$
- Confidence measure is. $$c = \frac{\sum_i t_i s_i}{\sqrt{\sum t_i^2 \sum s_i^2}}$$
- Watermark remarkably robust.

Example

Original

\+

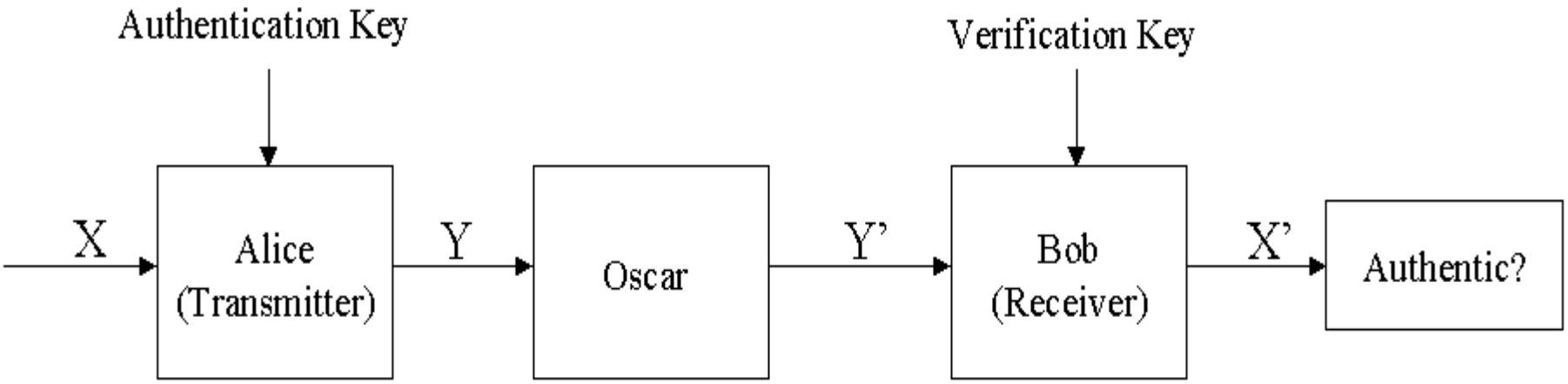
Watermark

=

Watermarked image

Multimedia Authentication

Authentication Codes



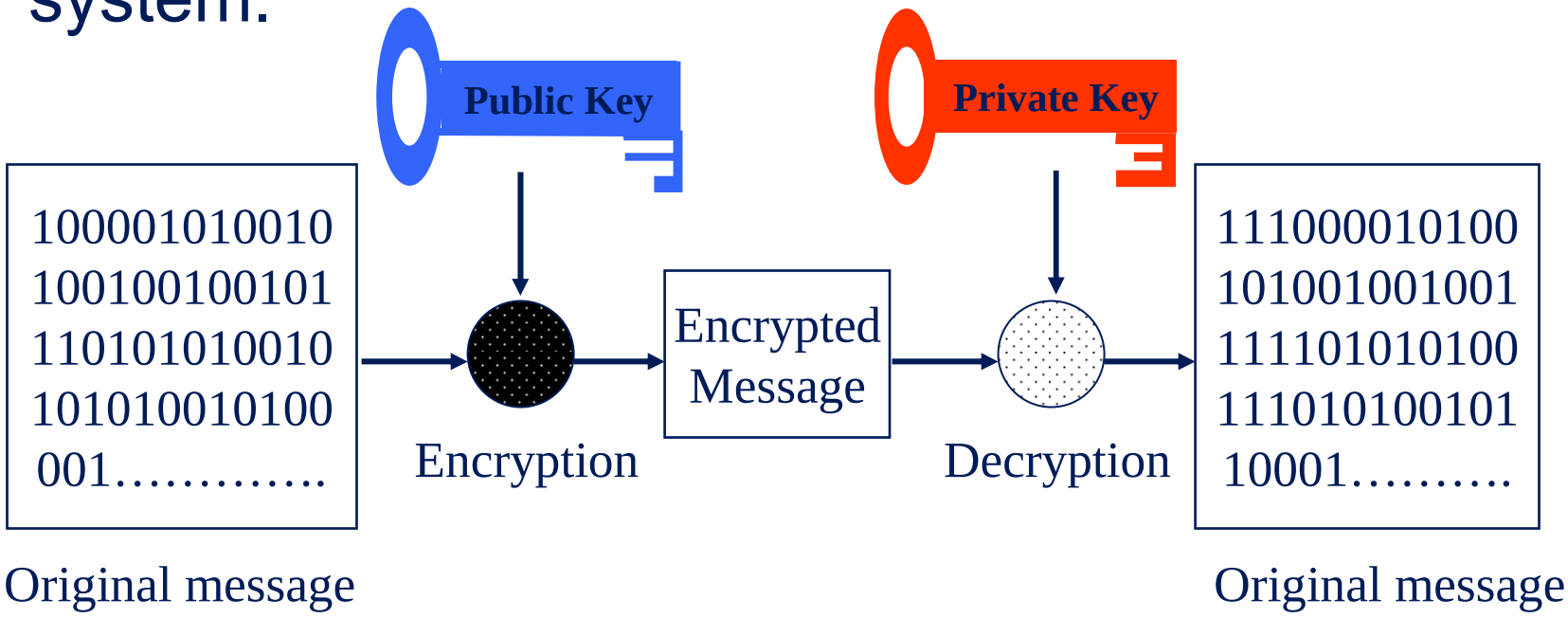
- Provides means for ensuring integrity of message
- Independent of secrecy - in fact sometimes secrecy may be undesirable!

Public-Key Cryptosystems

- **Public-key cryptography** was invented in 1976 by Diffie and Hellman in order to solve the key management problem. The system consists of two keys:
  - A **public key**, which is published and can be used to encrypt messages.
  - A **private key**, which is kept secret and is used to decrypt messages.
- Since the private key is never transmitted or shared, the problem of key management is greatly reduced.

Public-Key Cryptography

The most popular public-key encryption in use today is the **RSA** (Rivest-Shamir-Adleman) system.



Public-Key Cryptosystems for Authentication

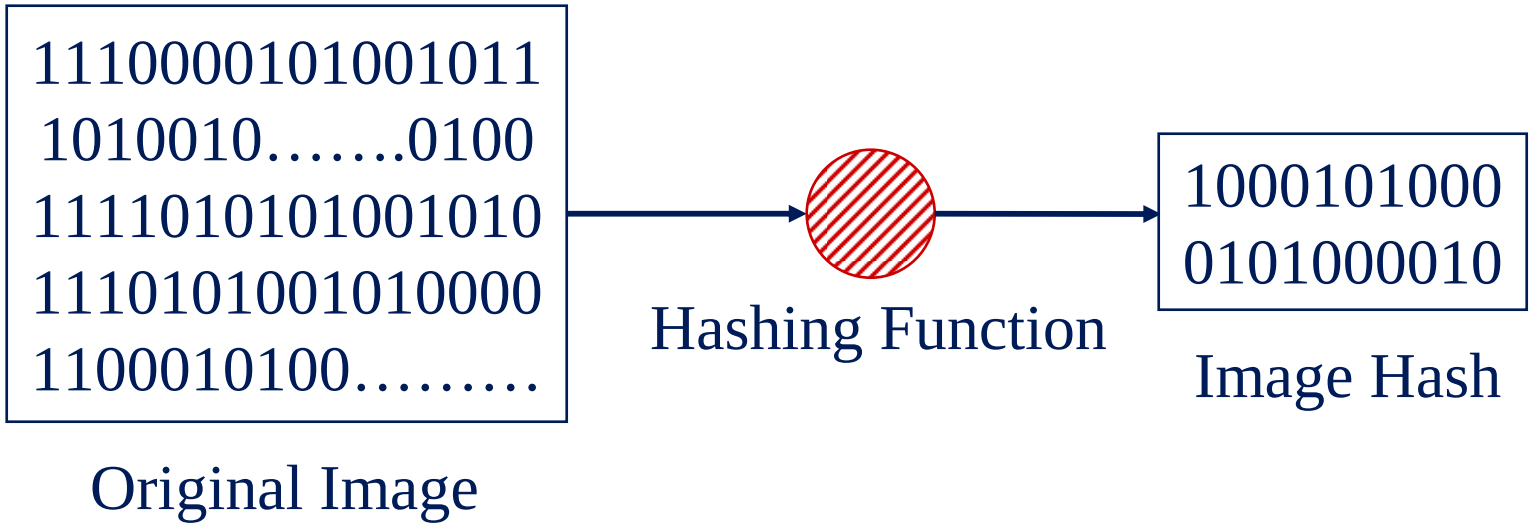
- Certain public-key cryptographic systems in which the roles of the public and private keys in encryption and decryption can be reversed, can also be used for authentication:
  - » Prior to sending a message, the sender encrypts the message with his/her private key.
  - » The message can be decrypted by the public using the public key of the signatory (no secrecy involved).
  - » Since it is computationally infeasible to find the private key from the public key and the known message, the decryption of the message into meaningful text constitutes its authentication.

One-Way Hash Functions

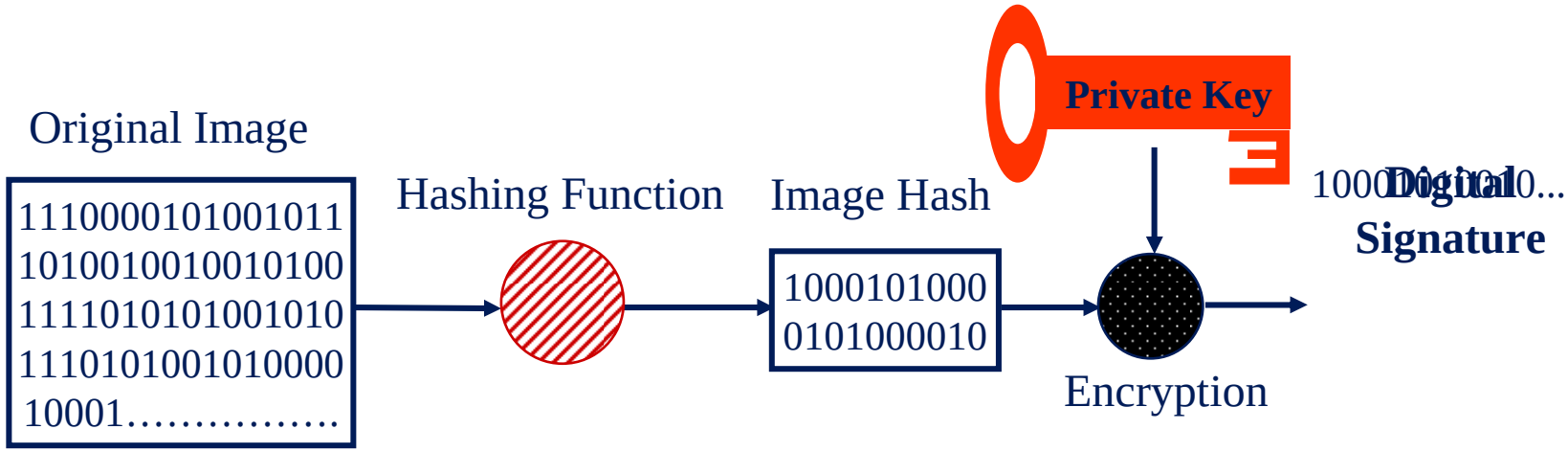
- **Hash function**: A computation that takes a variable-size input and returns a fixed-size digital string as output, called the **hash value**.
- **One-way hash function:** A hash function that is hard or impossible to invert, also called a **message digest function**.
- The one-way hash value can be thought of as the **digital fingerprint** of an image because:
  - » It is extremely unlikely for two different images to hash to the same value.
  - » It is computationally infeasible to find an image that hashes to a given value: precludes an attacker from replacing the original image with an altered image.

One-Way Hash Functions

- Examples of hash functions used for digital signatures are:
  - » 20-byte **secure hash algorithm** (SHA-1) that has been standardized for government applications.
  - » 16-byte **MD2**, **MD4**, or **MD5** developed by Rivest.

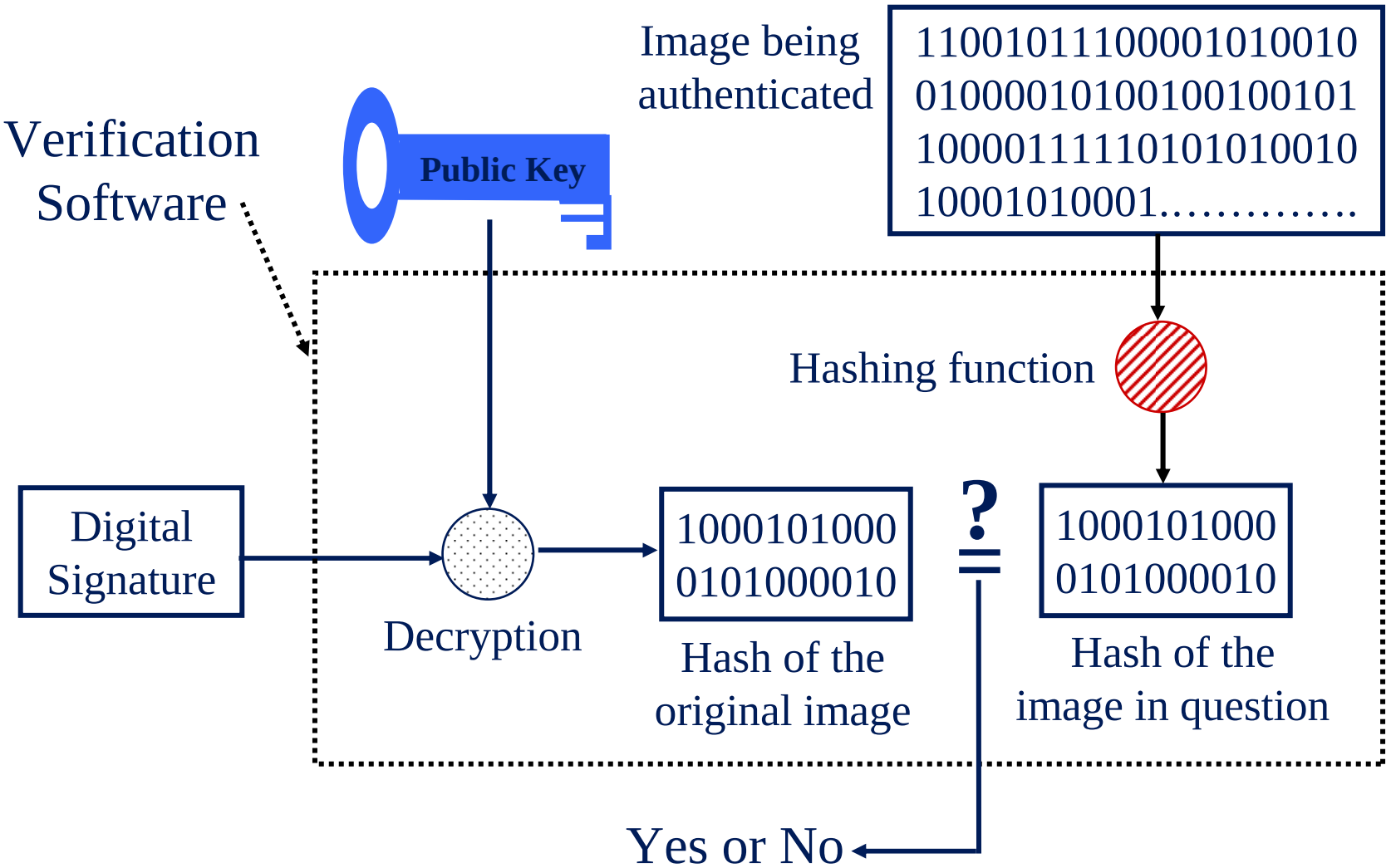


Digital Signature Generation



- A **digital signature** is created in two steps:
  - » A fingerprint of the image is created by using a one-way hash function;
  - » The hash value is encrypted with the private key of a public-key cryptosystem. Forging this signature without knowing the private key is computationally infeasible.

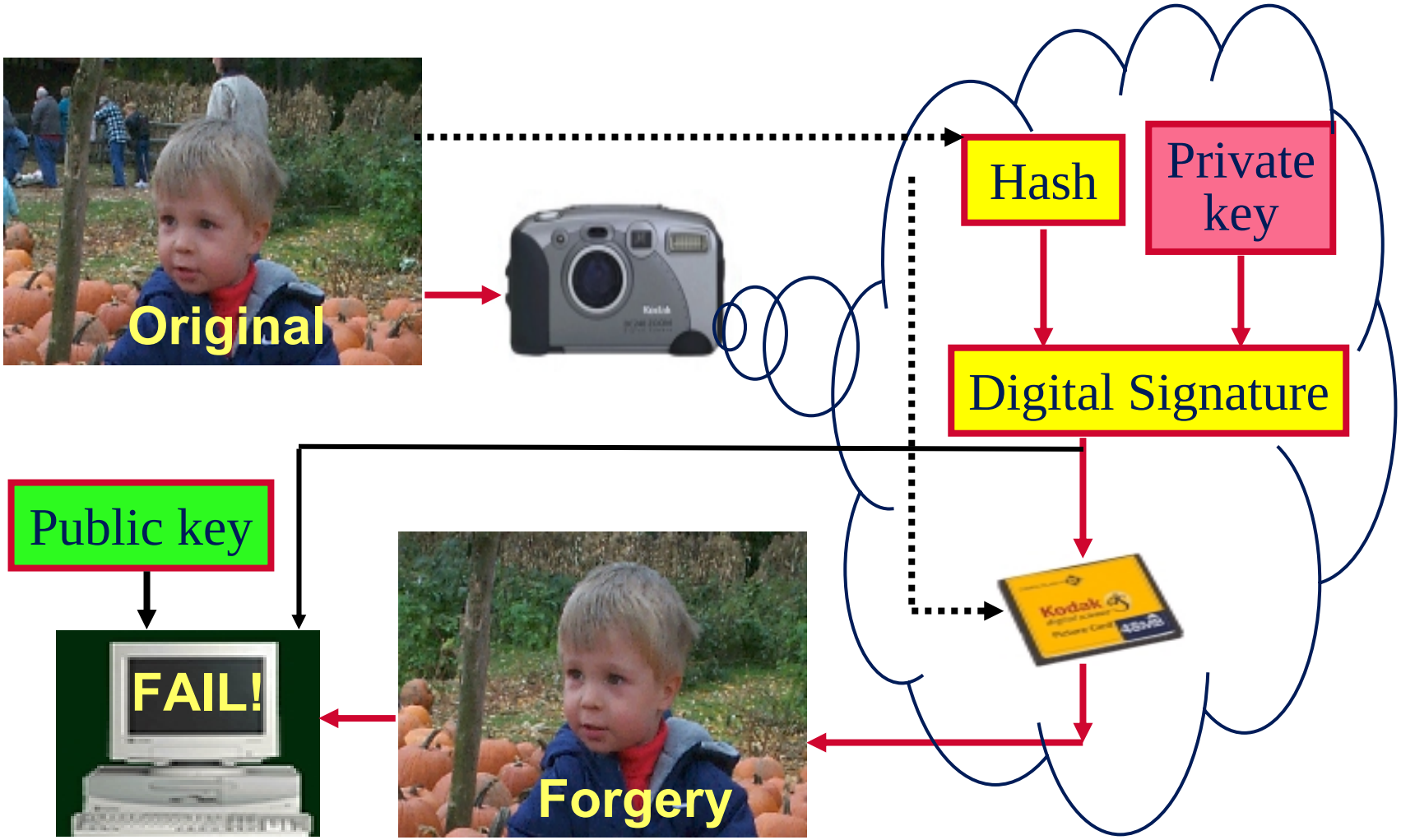
Digital Signature Verification



Techniques for Authentication

- Achieved by adding redundancy
  - » authenticator, tag, etc., or
  - » structure of message
- In some sense like Error Correcting Codes
- Private Key - Public Key <=> Authentication - Digital Signature
- Attacks
  - » Substitution
  - » Impersonation
  - » Choice of above

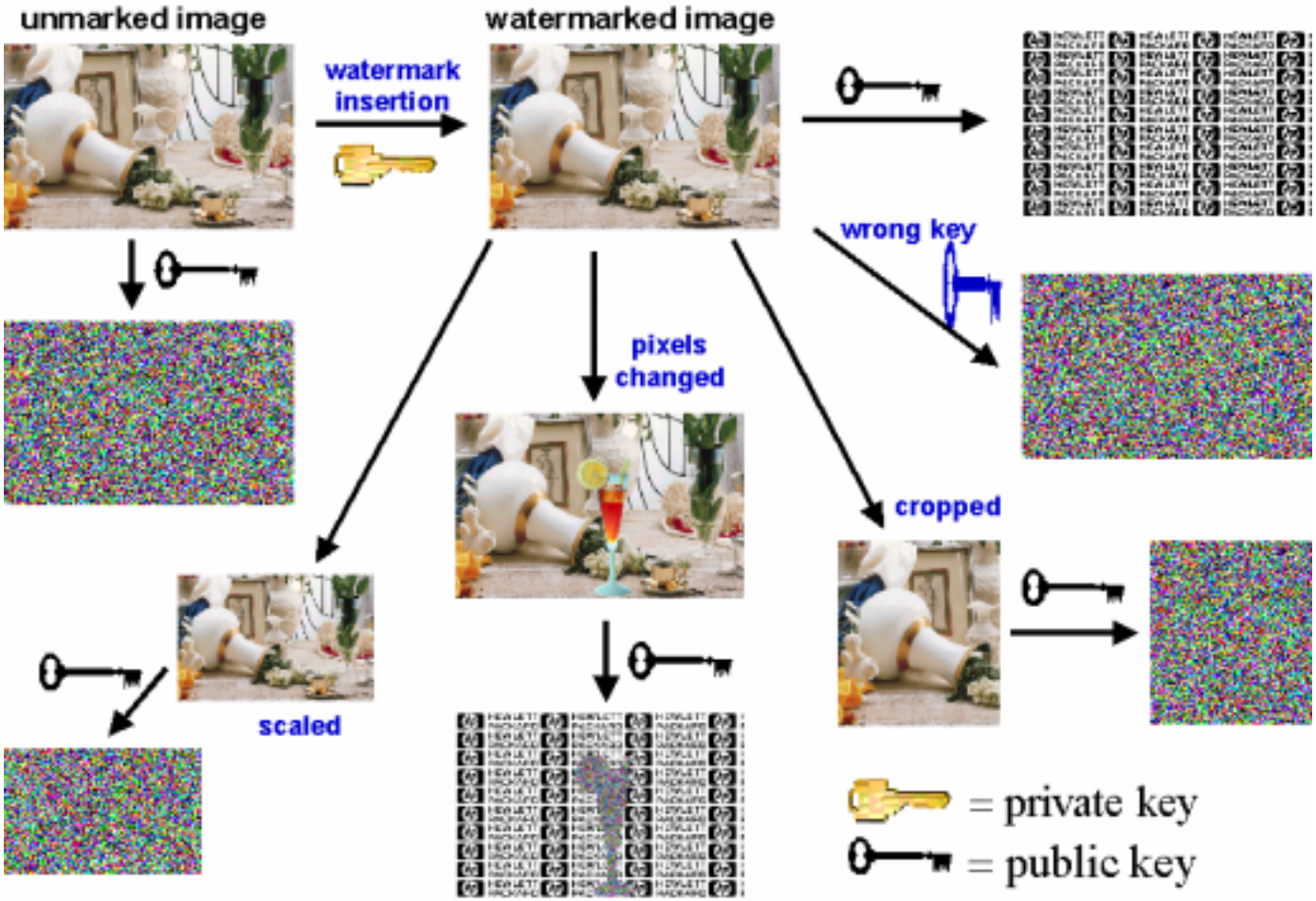
Digital Signature Authentication



Authentication of Multimedia - New Issues

- Authentication of content instead of specific representation -
  - » Example - JPEG or GIF image.
- Embedding of authenticator within content
  - » Survive transcoding
  - » Use existing formats
- Detect local changes
  - » Simple block based authentication could lead to substitution attacks
- Temporal relationship of multiple streams

Fragile Watermarks

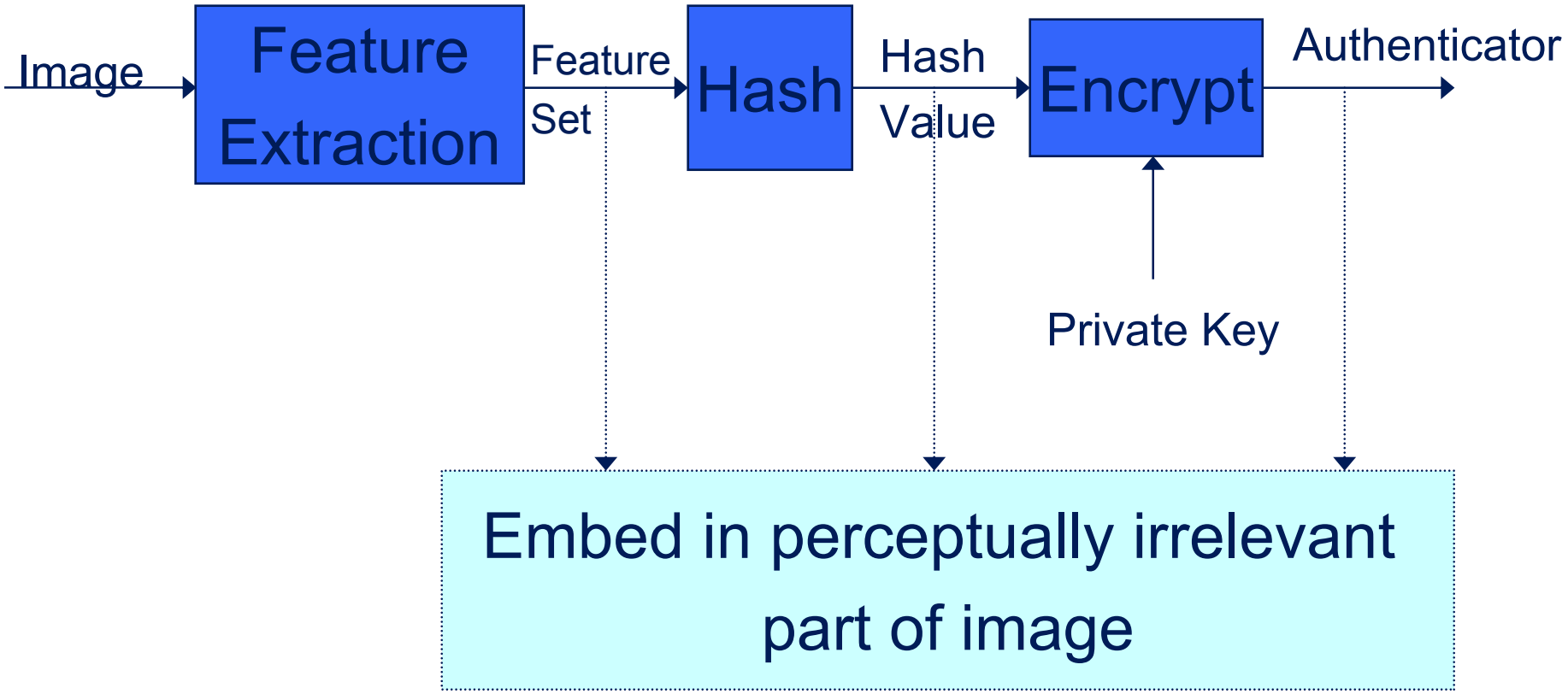


Detects and localizes any change to watermarked images.

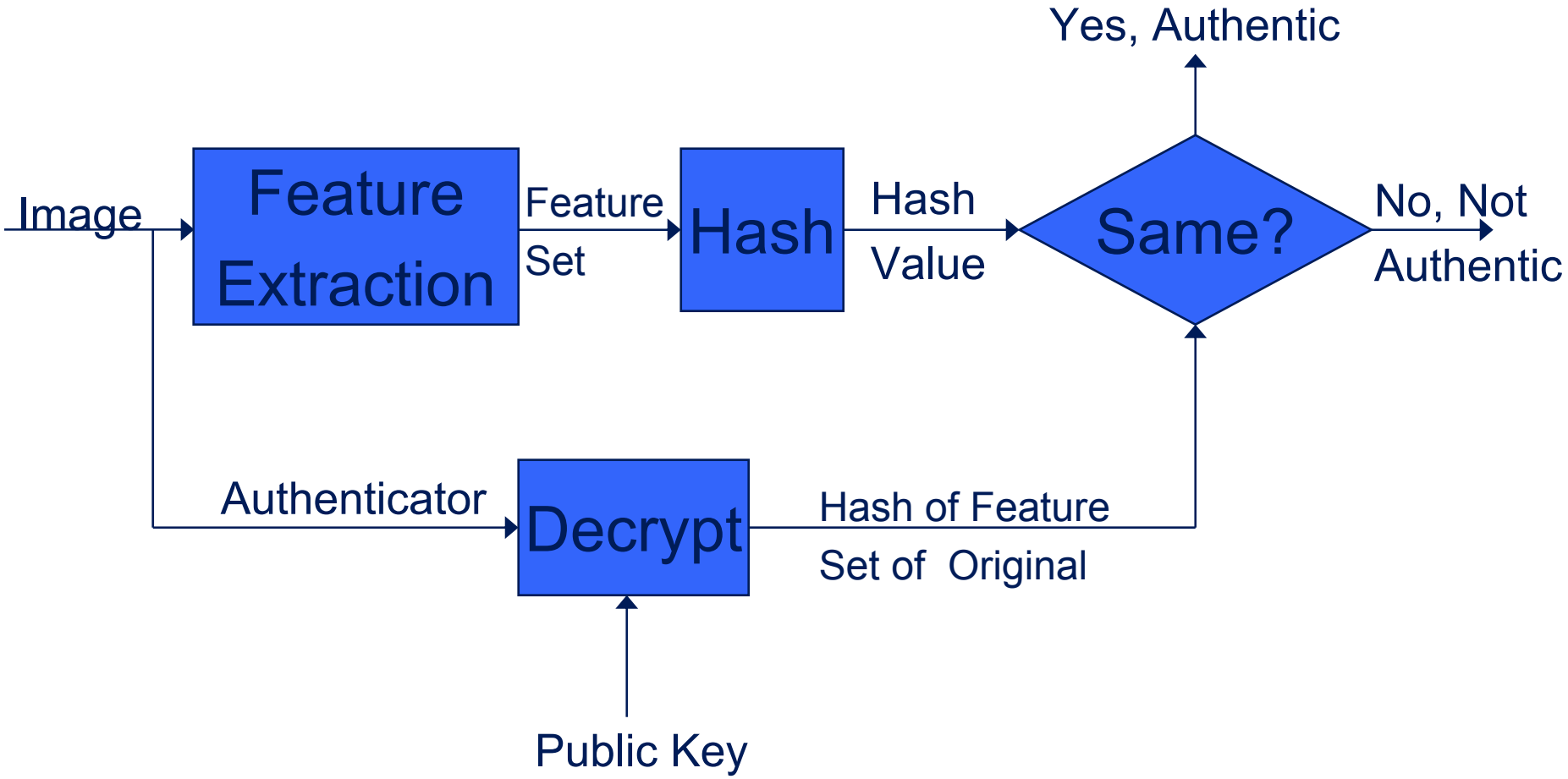
Limitations of Fragile Watermarks

- Essentially same as conventional authentication – authenticate representation and not “content”.
- The differences being –
  - » Embed authenticator in content instead of tag.
  - » Treat data stream as an object to be “viewed” by an human observer.
  - » Computationally efficient?

Feature Authentication



Feature Authentication (contd.)

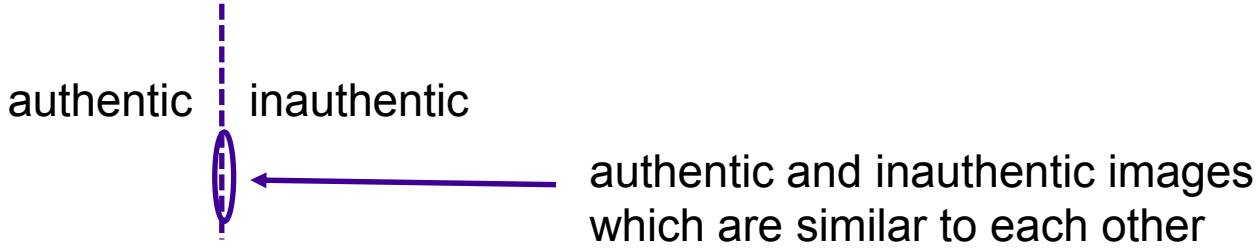


Limitations of feature authentication

- Difficult to identify a set of definitive features.
- Set of allowable changes has no meaningful structure – certain “small changes” may not be allowed but the same time “large” changes may be allowed in other situations.
- “Strong” features facilitate forgeries.
- “Weak” features cause too may false alarms.

Difficulties with content authentication of images

- Content is difficult to quantify.
- Malicious (benign) modifications are difficult to quantify.
- Images considered as points in continuous space means there is not a sharp boundary between authentic and inauthentic images.



Distortion Bounded Authentication

- Problem 1: allow flexibility in authentication to tolerate small changes
- Problem 2: to characterize and quantify the set of allowable changes
  - » Bound the errors
  - » Perceptual distortion or pixel value distortion
- Provide “guarantees” against substitution attacks.
- Approach – bounded tolerance authentication
  - » (semi-fragile)Watermarking techniques offer flexibility but most do not offer bounds

Distortion Bounded Authentication

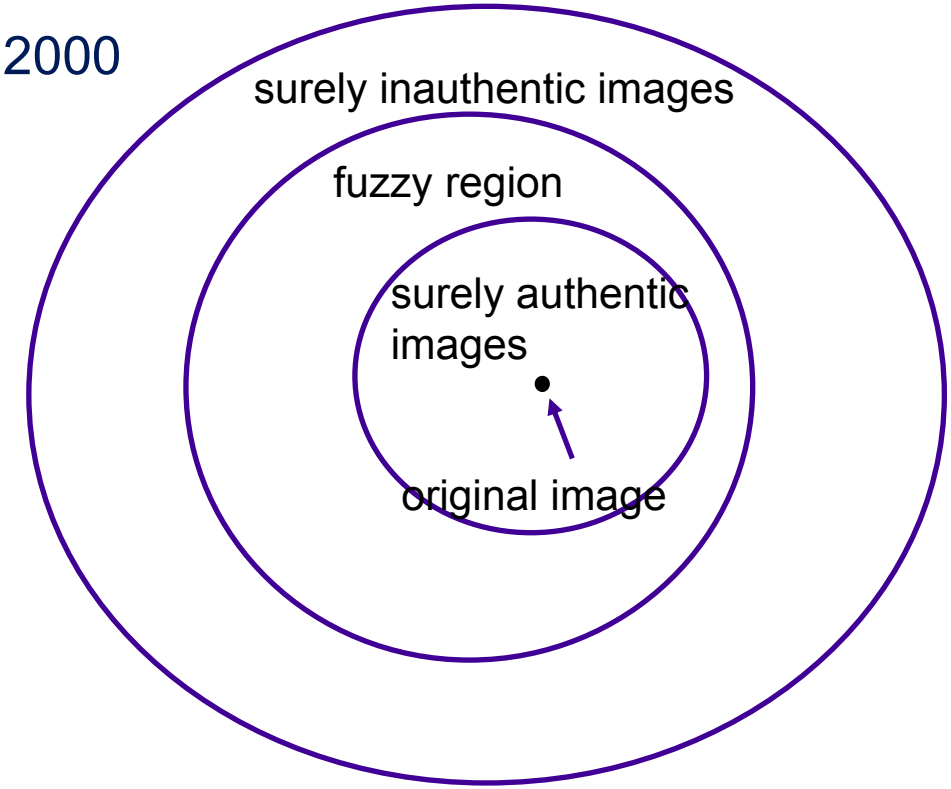
- Quantize image blocks or features prior to computing authenticator.
- Quantization also done prior to verifying authenticity of image.
- Enables distortion guarantees – image considered authentic as long as change made does not cause quantized version to change.
- Can be used in many different ways

Limitations

- Distortion added to “original” image.
- Similar problems as feature authentication, though to a lesser degree.
- Significant changes may indeed be possible within specified set of allowable changes.
- How to define set of allowable changes?

A Better Approach?

Chai Wah Wu - 2000



Fuzzy region: authenticity of image is uncertain.

Multimedia Fingerprinting

Definitions

- *A fingerprint* is a characteristic of an object that can be used to distinguish it from other similar objects.
  - » *E.g.,* human fingerprints, marks on a fired bullet
- *Fingerprinting* is the process of adding *fingerprints* to an object or of identifying the *fingerprint* of an object that is intrinsic to an object.
  - » Early examples: Table of logarithms with modified least significant digits, maps drawn with slight deliberate variations. Thatcher documents.
- The advent of digital objects and their unauthorized distribution has lead to the need for novel fingerprinting techniques.

Classification of Fingerprinting techniques (Wagner)

- Logical fingerprinting.
  - » Object is digital. The fingerprints are computer-generated and subject to computer processing.
- Physical fingerprinting.
  - » This is the opposite of logical fingerprinting. Here the fingerprints depend ***on*** physical characteristics of the object.

Classification of Fingerprinting techniques

- Perfect fingerprinting.
  - » Any alteration to the object that will make the fingerprinting unrecognizable must necessarily make the object unusable. Thus the distributor can always identify the recipient.
- Statistical fingerprinting.
  - » Given sufficiently many misused objects to examine, the distributor can gain any desired degree of confidence that he has correctly identified the compromised user. The identification is, however, never certain.
- Normal fingerprinting.
  - » This is a catch-all category for fingerprinting that does not belong to one of the first two categories.

Classification of Fingerprinting techniques

- Recognition.
  - » Recognize and record fingerprints that are already a part of the object.
- Deletion.
  - » The omission of some legitimate portion of the original object.
- Addition.
  - » Legitimate addition
  - » Modification.

Classification of Fingerprinting techniques

- Discrete fingerprint.
  - » An individual fingerprint with only a limited number of possible values.
    - – Binary fingerprint.
    - – N-ary fingerprint.
- Continuous fingerprint.
  - » Here a real quantity is involved and there is essentially no limit to the number of possible values.

Digital Fingerprints

- *A mark is a position in an object that can be in one of a fixed number of different states (Boneh and Shaw)*
  - » *I.e.,* a codeword comprised of a number of letters from a preset alphabet
- A *fingerprint* is a collection of marks
- Fingerprinting has two concerns
  - » How to mark an object
  - » How to use these marks to create a fingerprint
- Fingerprinting cannot prevent unauthorized distribution, but acts as a deterrence mechanism by helping trace illegal copies back to source
  - » **traitor**: authorized users who redistribute content in an unauthorized manner
  - » **traitor tracing**: identifying traitors based on redistributed content

Marking Assumption

The assumption states that a marking scheme designed to resist collusion and trace traitors with the following properties exist:

1. Colluding users may detect a specific mark only if the mark differs between their copies. Otherwise the mark cannot be detected.
   - If there is no collusion, fingerprint reduces to a serial number
2. Users cannot change the state of an undetected mark without rendering the object useless.

Basically, limits actions of colluding users

Boneh-Shaw Construction

- Targeted at generic data with “Marking assumptions” (1998)
  - » an abstraction of collusion model
  - » *E.g.,* assume a 6-bit content marked in the 2nd, 4th , and 5th positions and let m_1, m_2 and m_3 be the marked contents

m_1	=	0	1	1	1	0	1
m_2	=	0	1	1	0	1	1
m_3	=	0	0	1	1	1	1

  - » If m_1, m_2 and m_3 collude the positions of the marks are determined
  - » If $m_1$and m_2 collude only 4th and 5th marks can be identified

Boneh-Shaw Construction

- Focus on tracing one of the colluders
  - » *Totally c-secure* fingerprinting codes: Given a coalition of at most c traitors, an illegal copy can be traced back to at least once traitor in the coalition.
    - – Proved that for c>1 no such codes exist assuming colluder may leave marks in unreadable state
  - » Used randomization techniques to construct *ε-error c-secure* codes that are able to capture at least one colluder, out of a coalition of *c-colluders*, with a probability of *1-ε* for some small error rate of *ε.*

Collusion Secure Codes

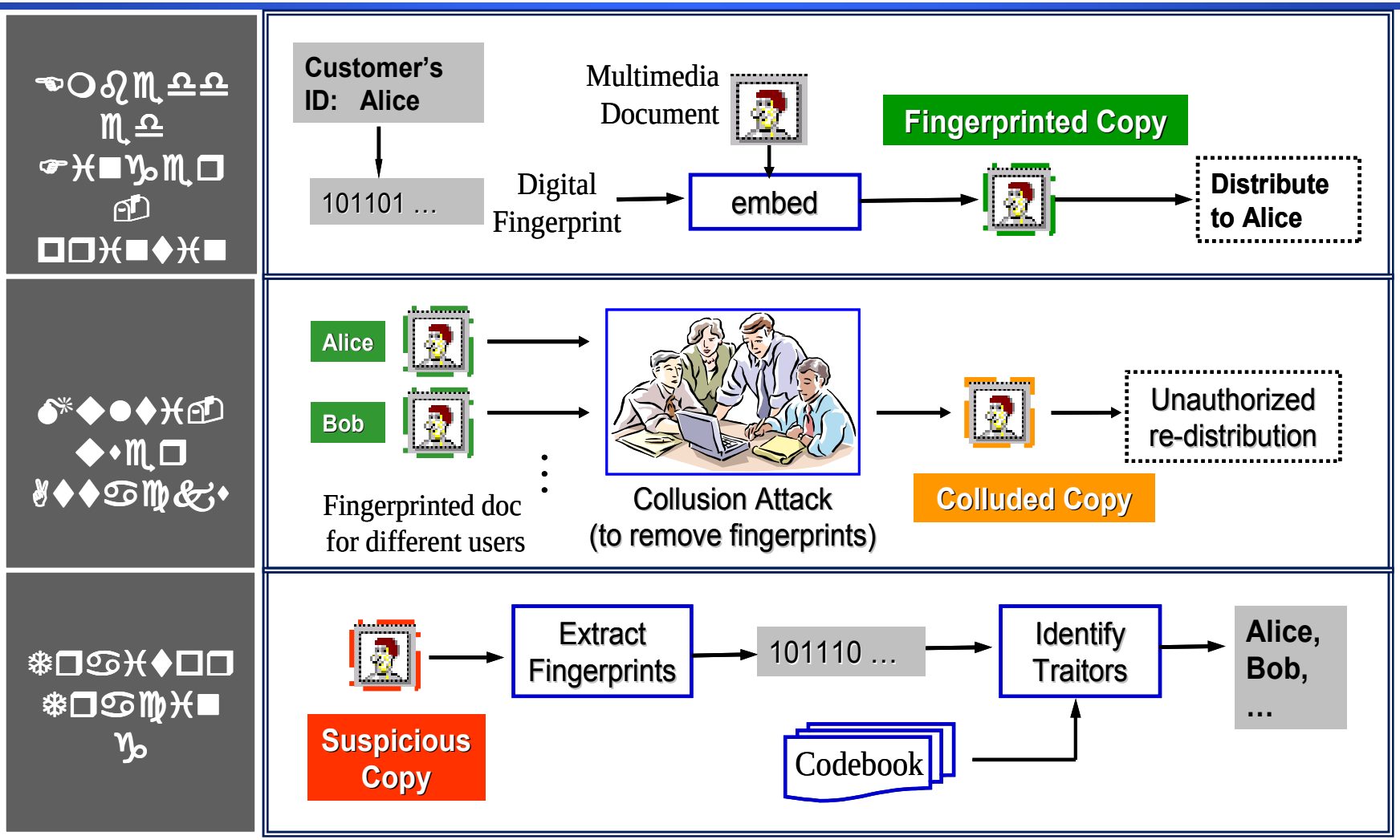
- Generate a code matrix whose rows are distinct fingerprints
- In the matrix, above the main diagonal is all ones and below is all zeros
  - » May look like stairs, and the stairs width determine the ε, *i.e.,*

    m_1 : 111111111111
    m_2 : 000111111111
    m_3 : 000000111111
    m_4 : 000000000111
  - » Prior to embedding each fingerprint is randomly permuted using a fixed permutation
- A collusion will *most likely* generate a codeword different than m_1, m_2 , m_3 and m_4.

Collusion Secure Codes (cont'd)

- Initially, fingerprints are far from each other (Hamming distance)
- The detector decodes the colluded fingerprint to *nearest* initial fingerprint in the code matrix
- Arbitrarily small ε yields very long codes
  - » Collusion resistance proportional to fourth root of content size (*i.e.,* to capture at least one of c-colluders code length must be of the order $O(c^4 \log c)$).
- Lot of follow up work in crypto literature that extends and improves Boneh-Shaw results.

Embedded Fingerprinting for Multimedia



What is Different?

- New issues with multimedia
  - » “Marking assumptions” do not directly carry over …
  - » Some code bits may become erroneously decoded due to strong noise and/or inappropriate embedding
  - » Can choose appropriate embedding to prevent colluders from arbitrarily changing the embedded fingerprint bits
- Want to trace as many colluders as possible
- Major Concerns
  - » How to embed/detect the fingerprint
    - – Deploy techniques from watermarking
  - » How to generate the fingerprint
    - – Utilize techniques from coding theory
  - » The type of attack the fingerprinted object undergoes

Marking Assumption for Multimedia Fingerprinting

Marking assumption considers a scheme with two specific requirements

- Fidelity requirement (Easy to satisfy)
  - » Marks are perceptually invisible and can be discovered only by comparison
  - » Unmarked object is not available
- Robustness requirement (Difficult to achieve)
  - » Undetected marks cannot be altered or removed

Spread-Spectrum Fingerprint Embedding / Detection

- Spread-spectrum embedding/detection
  - » Provide very good tradeoff on imperceptibility and robustness, esp. under non-blind detection
  - » Typical watermarking-to-noise (WNR) ratio: -20dB in blind detection, 0dB in non-blind detection
- Embedding: $\mathbf{X=S+\alpha W}$ where **S** is the original object, $\mathbf{W_i}$ is the fingerprint, and **α** is the embedding strength
- Detection: Analysis of the similarity between **Y** and $\mathbf{W_i}$, *i.e., correlation(**Y**,$\boldsymbol{W_i}$)* or *correlation(**Y-S**,$\boldsymbol{W_i}$)*

Fingerprinting Generation

- Choice of modulation schemes

Orthogonal modulation $\mathbf{w}_j = \mathbf{u}_j$

# of fingerprints = # of orthogonal bases

(Binary) coded modulation $\mathbf{w}_j = \sum_{i=1}^{B} b_{ij}\mathbf{u}_i$

for $b_{ij} \in \{0,1\}$ or $b_{ij} \in \{\pm 1\}$

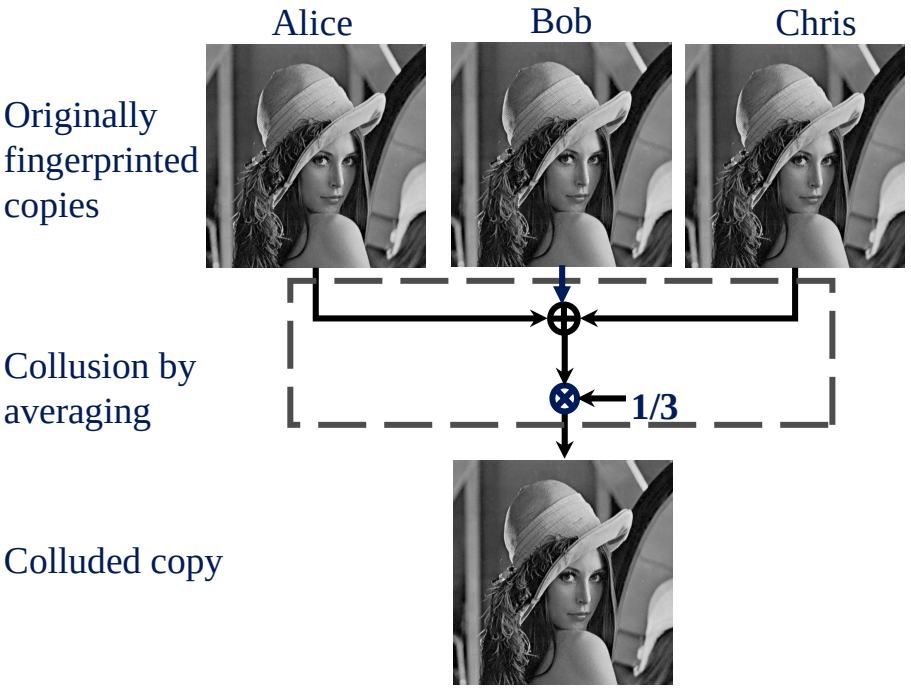
# of fingerprints >> # of orthogonal bases

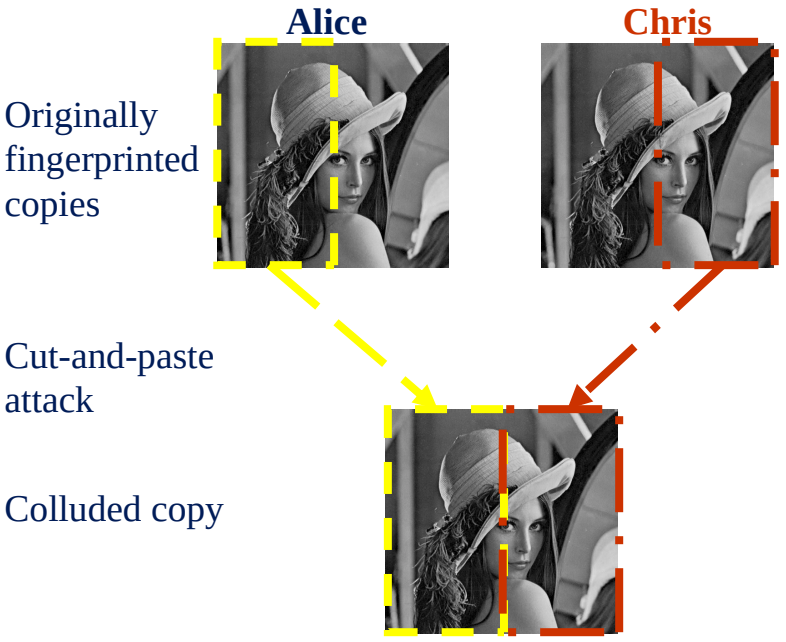
Attacks on Fingerprinting Systems

- Attacks on the marking system
  - » Exploiting the robustness of the fingerprinting embedding and detection scheme
  - » Collusion attacks. Collusion: $\mathbf{Y=g(X_1,X_2, \ldots X_k)}$ where g(.) is a function designating the nature of modification on collection of fingerprinted objects available.
    - – More effective
    - – May yield even better quality than the distributed object
- The traitor may have two types of goal
  1. Removal of the fingerprints from the fingerprinted-object
  2. Framing an innocent user
- *Design Goal: Improve collusion resistance w.r.t. type-1 while increasing robustness to type-2 attacks.*

Collusion Attacks by Multiple Users

- Interesting collusion attacks become possible
- Fairness: *Each colluder contributes equal share through averaging, interleaving, and nonlinear combining*





Linear vs. Nonlinear Collusion

- Linear collusion by averaging is simple and effective
- Colluders can output any value between the minimum and maximum values, and have high confidence that such spurious value is within the range of JND.
  - *Important to consider nonlinear collusion as well.*
- Order statistics based nonlinear collusions

$$V_j = g(y_j^{(k)})_{k \in S_C} = x_j + JND_j \cdot g(w_j^{(k)})_{k \in S_C}$$

$$V_j^{ave};\quad V_j^{\min};\quad V_j^{\max};\quad V_j^{median}$$

$$V_j^{\min\max} = average\left(V_j^{\min}, V_j^{\max}\right)$$

$$V_j^{\mathrm{mod}\,neg} = V_j^{\min} + V_j^{\max} - V_j^{med}$$

$$V_j^{randneg} = \begin{cases} V_j^{\min} & \text{w.p. } p \\ V_j^{\max} & \text{w.p. } 1-p \end{cases}$$

(Image) Steganography and Steganalysis

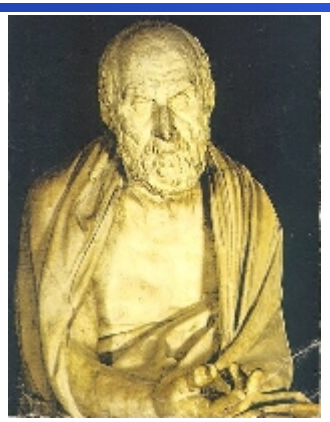
Steganography

- ***Steganography -*** "covered writing".
- For example (sent by a German spy during World War I),

  Apparently neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affects pretext for embargo on byproducts, ejecting suets and vegetable oils.

  Pershing sails from NY June I.

Ancient Steganography



Herodotus (485 – 525 BC) is the first Greek historian. His great work, The Histories, is the story of the war between the huge Persian empire and the much smaller Greek city-states.

Herodotus recounts the story of **Histaiaeus**, who wanted to encourage **Aristagoras of Miletus** to revolt against the Persian king. In order to securely convey his plan, Histaiaeus shaved the head of his messenger, wrote the message on his scalp, and then waited for the hair to regrow. The messenger, apparently carrying nothing contentious, could travel freely. Arriving at his destination, he shaved his head and pointed it at the recipient.

Ancient Steganography

Pliny the Elder explained how the milk of the thithymallus plant dried to transparency when applied to paper but darkened to brown when subsequently heated, thus recording one of the earliest recipes for invisible ink.

Pliny the Elder.
AD 23 - 79

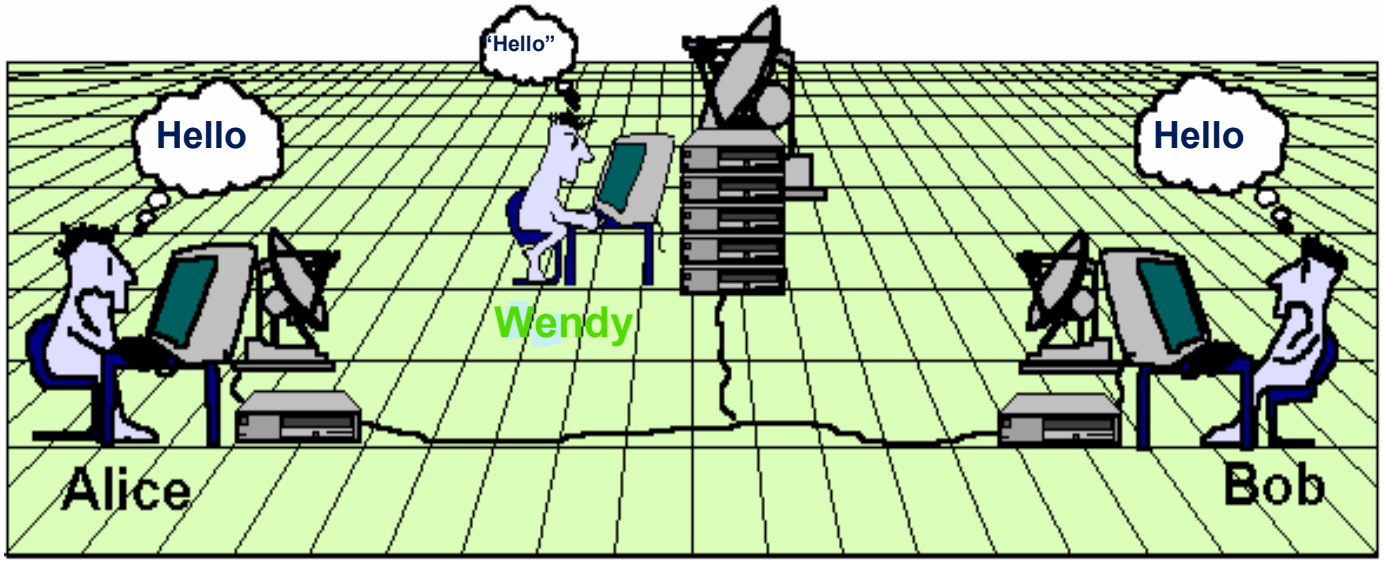
The **Ancient Chinese** wrote notes on small pieces of silk that they then wadded into little balls and coated in wax, to be swallowed by a messenger and retrieved at the messenger's gastrointestinal convenience.

Renaissance Steganography

Giovanni Battista Porta
(1535-1615)

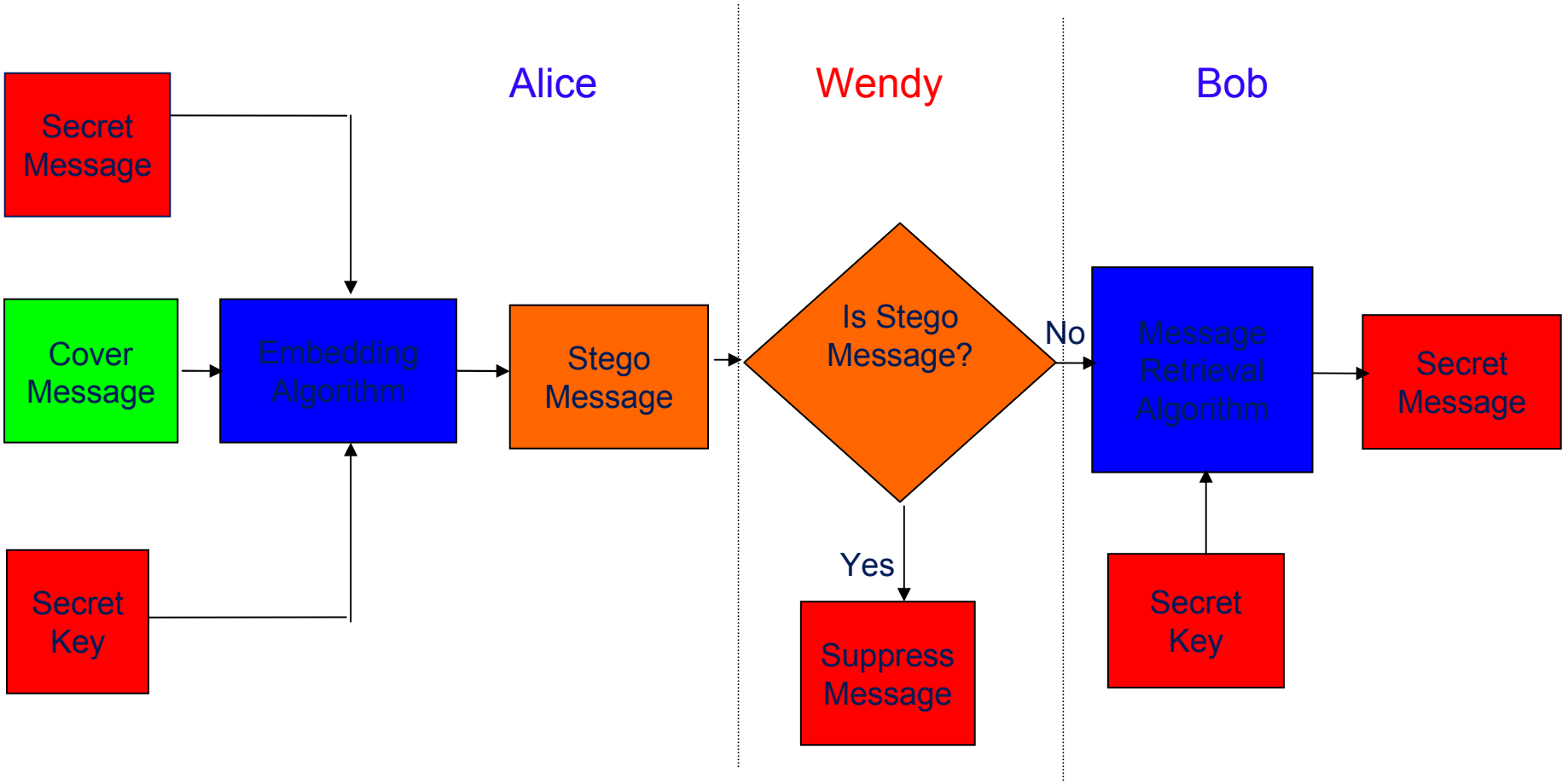
Giovanni Battista Porta described how to conceal a message within a hard-boiled egg by writing on the shell with a special ink made with an ounce of alum and a pint of vinegar. The solution penetrates the porous shell, leaving no visible trace, but the message is stained on the surface of the hardened egg albumen, so it can be read when the shell is removed.

Modern Steganography - The Prisoners' Problem



- Simmons – 1983
- Done in the context of USA – USSR nuclear non-proliferation treaty compliance checking.

Modern Terminology and (Simplified) Framework



Secret Key Based Steganography

- If system depends on secrecy of algorithm and there is no key involved – *pure steganography*
  - » Not desirable. Kerkhoff's principle.
- Secret Key based steganography
- Public/Private Key pair based steganography

Active and Passive Warden Steganography

- Wendy can be *passive*:
  - » Examines all messages between Alice and Bob.
  - » Does not change any message
  - » For Alice and Bob to communicate, Stego-object should be indistinguishable from cover-object.
- Wendy can be *active*:
  - » Deliberately modifies messages by a little to thwart any hidden communication.
  - » Steganography against active warden is difficult.
  - » Robust media watermarks provide a potential way for steganography in presence of active warden.

Steganalysis

- *Steganalysis* refers to the art and science of discrimination between stego-objects and cover-objects.
- Steganalysis needs to be done without any knowledge of secret key used for embedding and maybe even the embedding algorithm.
- However, message does not have to be gleaned. Just its presence detected.

Cover Media

- Many options in modern communication system:
  - » Text
  - » Slack space
  - » Alternative Data Streams
  - » TCP/IP headers
  - » Etc.
- Perhaps most attractive are multimedia objects -
  - – Images
  - – Audio
  - – Video
- We focus on Images as cover media. Though most ideas apply to video and audio as well.

Steganography, Data Hiding and Watermarking

- Steganography is a special case of data hiding.
  - » Data hiding in general need not be steganography. Example – Media Bridge.
- It is not the same as watermarking.
  - » Watermarking has a malicious adversary who may try to remove, invalidate, forge watermark.
- In Steganography, main goal is to escape detection from Wendy.

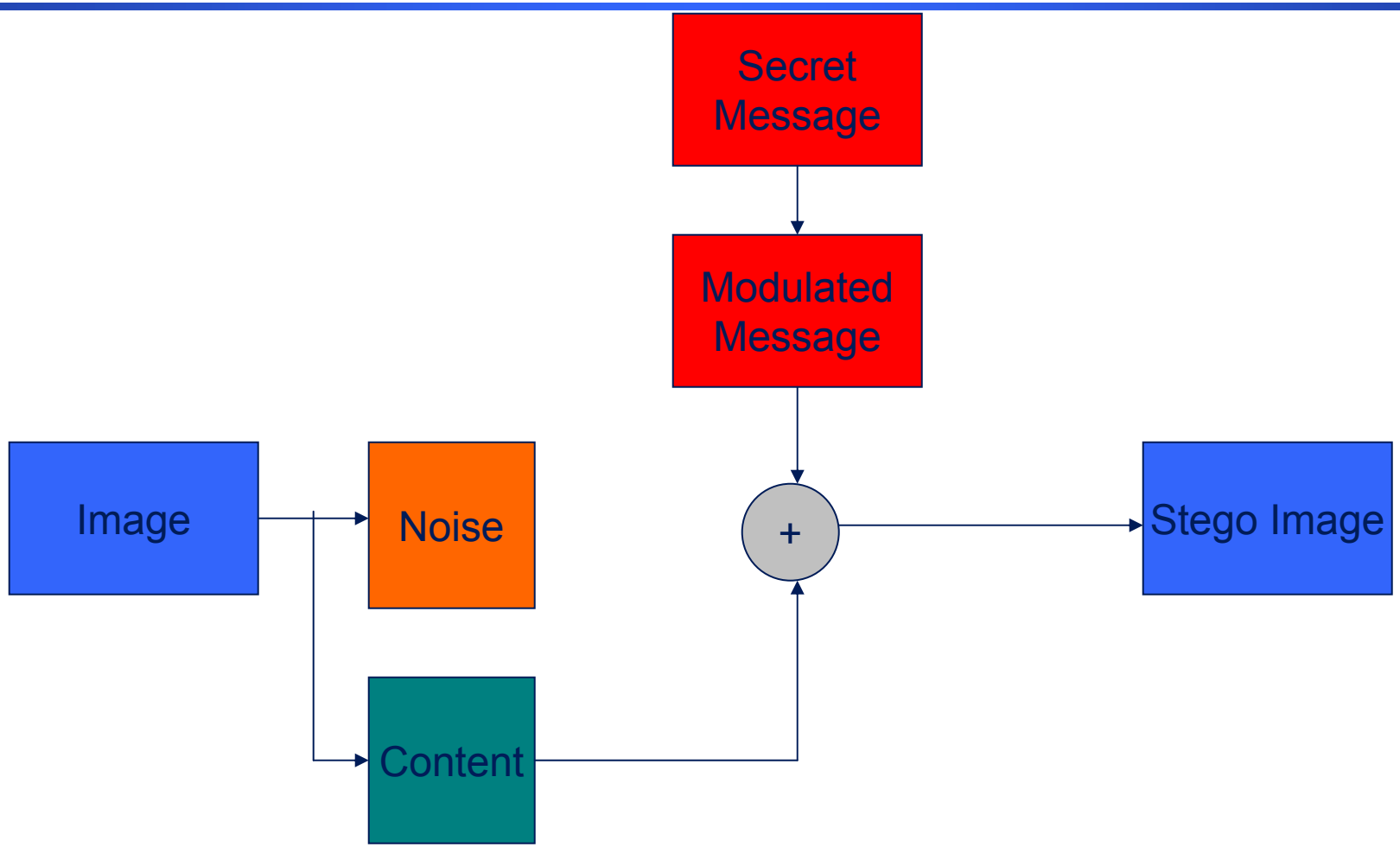
Information Theoretic Framework

- Cachin defines a Steganographic algorithm to be ε secure if the relative entropy between the cover object and the stego object pdf's is at most ε :

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

- Perfectly secure if $\varepsilon = 0$
- Example of a perfectly secure techniques known but not practical.

Steganography in Practice



Steganalysis in Practice

- Techniques designed for a specific steganography algorithm
  - » Good detection accuracy for the specific technique
  - » Useless for a new technique
- Universal Steganalysis techniques
  - » Less accurate in detection
  - » Usable on new embedding techniques

Simple LSB Embedding in Raw Images

- LSB embedding
  - » Least significant bit plane is changed. Assumes passive warden.
- Examples: Encyptic, Stegotif, Hide
- Different approaches
  - » Change LSB of pixels in a random walk
  - » Change LSB of subsets of pixels (i.e. around edges)
  - » Increment/decrement the pixel value instead of flipping the LSB

LSB Embedding

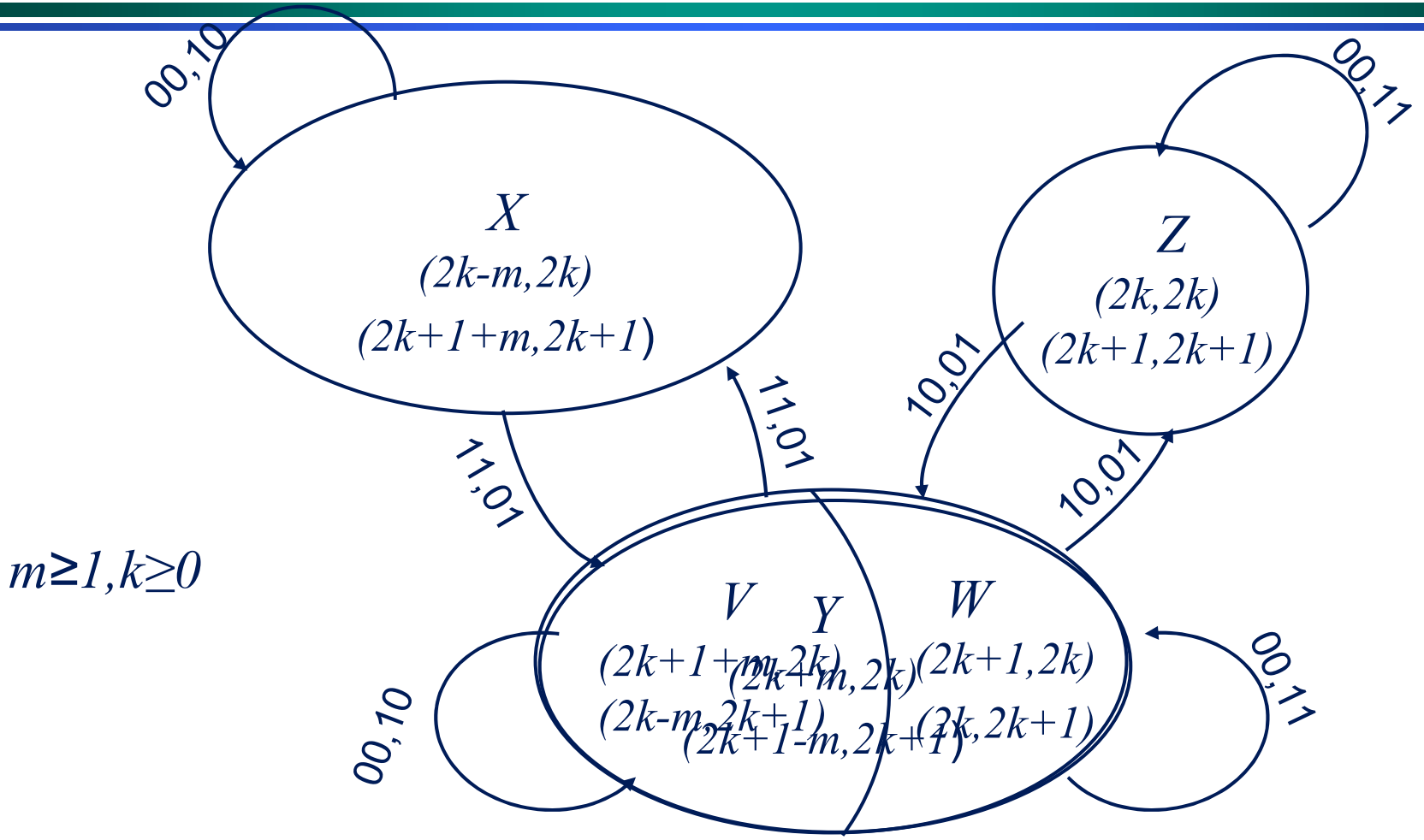
Steganalysis of LSB Embedding

- PoV steganalysis - Westfeld and Pfitzmann.
  - » Exploits fact that odd and even pairs from “closed set” under LSB flipping.
  - » Accurately detects when message length is comparable to size of bit plane.
- RS-Steganalysis - Fridrich et. al. [14]
  - » Very effective. Even detects around 2 to 4% of randomly flipped bits.

LSB steganalysis with Primary Sets

- Proposed by Dumitrescu, Wu, Memon
  - » Based on statistics of sets defined on neighboring pixel pairs.
  - » Some of these sets have equal expected cardinalities, if the pixel pairs are drawn from a continuous-tone image.
  - » Random LSB flipping causes transitions between the sets with given probabilities, and alters the statistical relations between their cardinalities.
  - » Analysis leads to a quadratic equation to estimate the embedded message length with high precision.

State Transition Diagram for LSB Flipping



X, *V*, *W*, and *Z*, which are called **primary sets**

Transition Probabilities

- If the message bits of LSB steganography are randomly scattered in the image, then

$$\text{i)}\ \ \rho(00) = \left(1 - \frac{p}{2}\right)^2,$$

$$\text{ii)}\ \ \rho(01) = \rho(10) = \frac{p}{2}\left(1 - \frac{p}{2}\right),$$

$$\text{iii)}\ \ \rho(11) = \left(\frac{p}{2}\right)^2.$$

- Let X, Y, V, W and Z denotes sets in original image and X', Y'. W' and Z' denote the same in stego image.

Message Length in Terms of Cardinalities of Primary Sets

- Cardinalities of primary sets in stego image can be computed in terms of the original

$$|X'| = |X|\left(1 - \frac{p}{2}\right) + |V|\frac{p}{2}$$

$$|V'| = |V|\left(1 - \frac{p}{2}\right) + |X|\frac{p}{2}$$

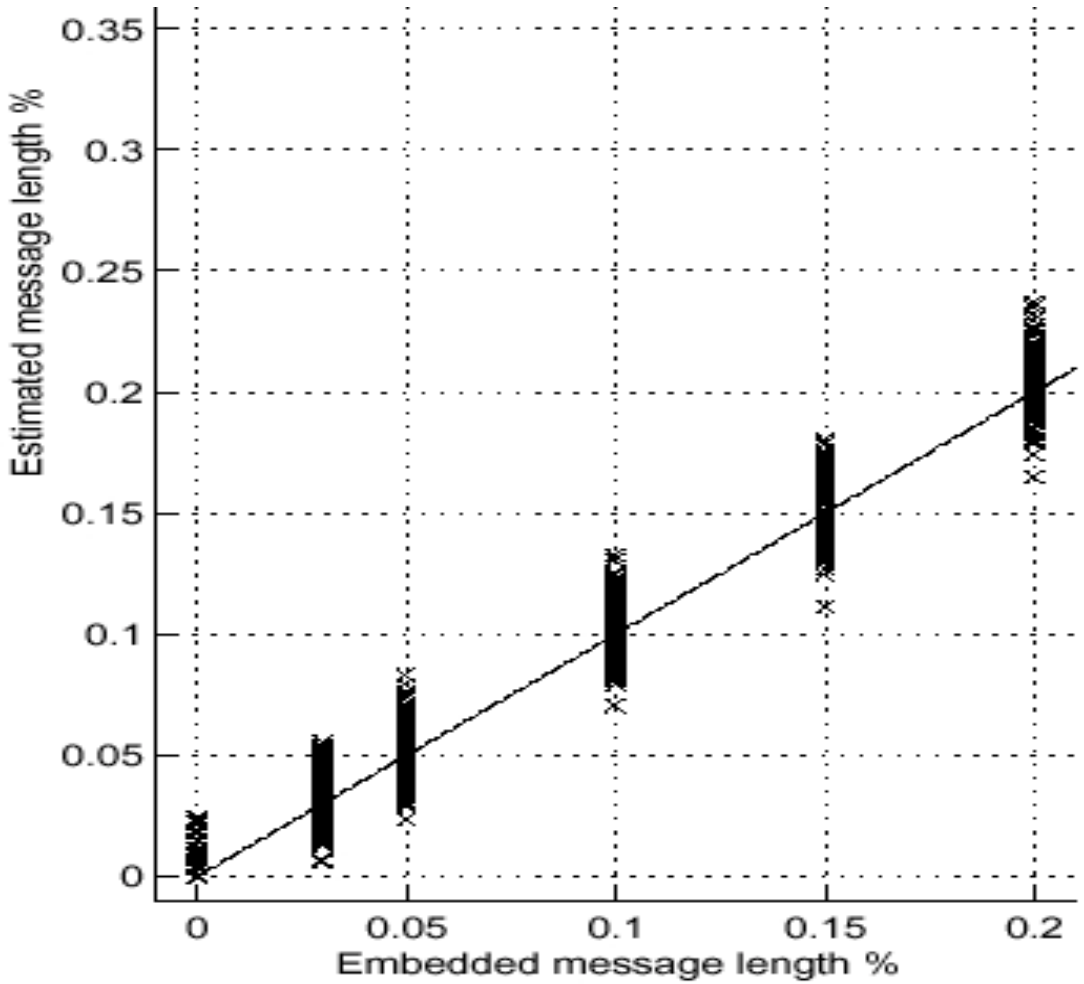
$$|W'| = |W|\left(1 - p + \frac{p^2}{2}\right) + |Z|p\left(1 - \frac{p}{2}\right)$$

- Assuming $E\{|X|\} = E\{|Y|\}$ and some algebra, we get:

$$0.5\gamma \cdot p^2 + \left(2|X'| - |P|\right)p + |Y'| - |X'| = 0$$

- Where $\gamma = |W' \cup Z'|. = |W \cup Z|.$

Simulation Results



Embedding in JPEG Images

- Embedding is done by altering the DCT coefficient in transform domain
- Examples: Jsteg, F5, Outguess
- Many different techniques for altering the DCT coefficients

F5

- F5 uses hash based embedding to minimize changes made for a given message length
- The modifications done, alter the histogram of DCT coefficients
- Given the original histogram, one is able to estimate the message length accurately
- The original histogram is estimated by cropping the jpeg image by 4 columns and then recompressing it
- The histogram of the recompressed image estimated the original histogram

F5 plot

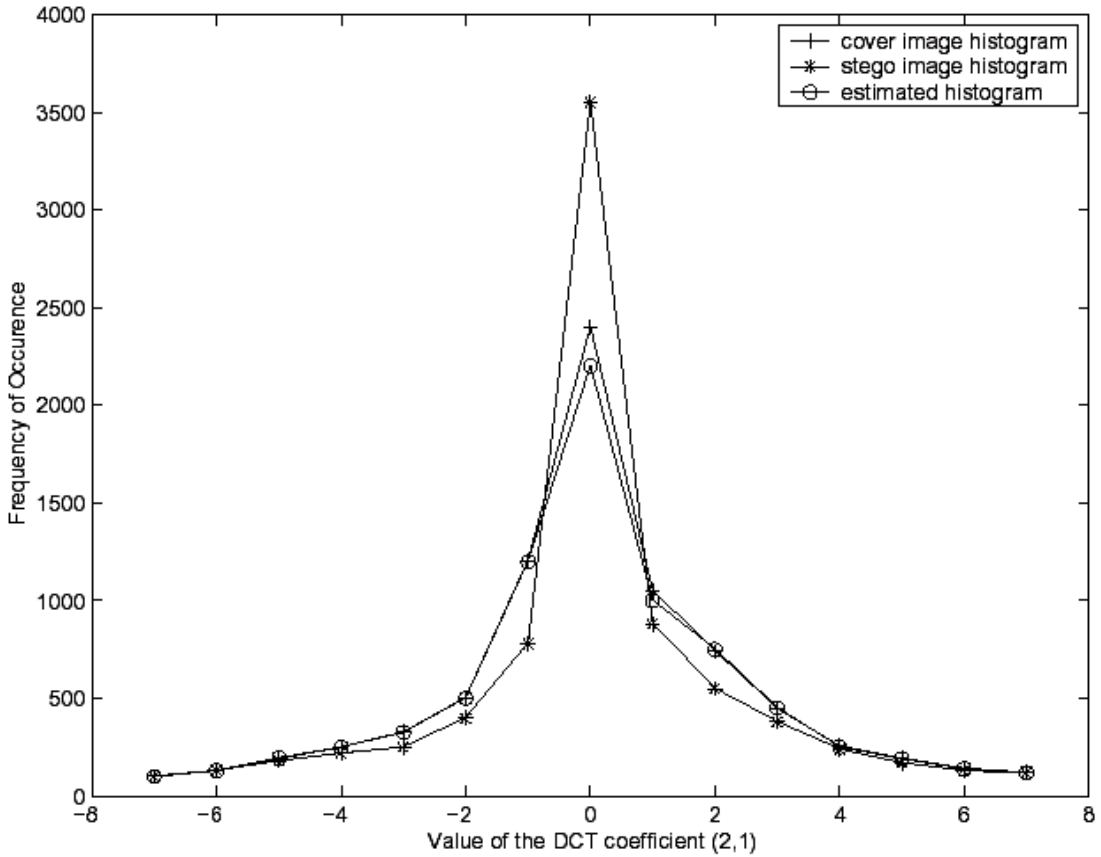


Fig. 5. The effect of F5 embedding on the histogram of the DCT coefficient (2,1).

Outguess

- Embeds messages by changing the LSB of DCT coefficients on a random walk
- Only half of the coefficients are used at first
- The remaining coefficients are adjusted so that the histogram of DCT coefficient would remain unchanged
- Since the Histogram is not altered the steganalysis technique proposed for F5 will be useless

Outguess

- Researchers proposed “blockiness” attack
- Noise is introduced in DCT coefficients after embedding
- Spatial discontinuities along 8x8 jpeg blocks is increases
- Embedding a second time does not introduce as much noise, since there are cancellations
- Increase or lack of increase indicates if the image is clean or stego

Universal Steganalysis Techniques

- Techniques which are independent of the embedding technique
- One approach – identify certain image features that reflect hidden message presence.
- Two problems
  - » Calculate features which are sensitive to the embedding process
  - » Finding strong classification algorithms which are able to classify the images using the calculated features

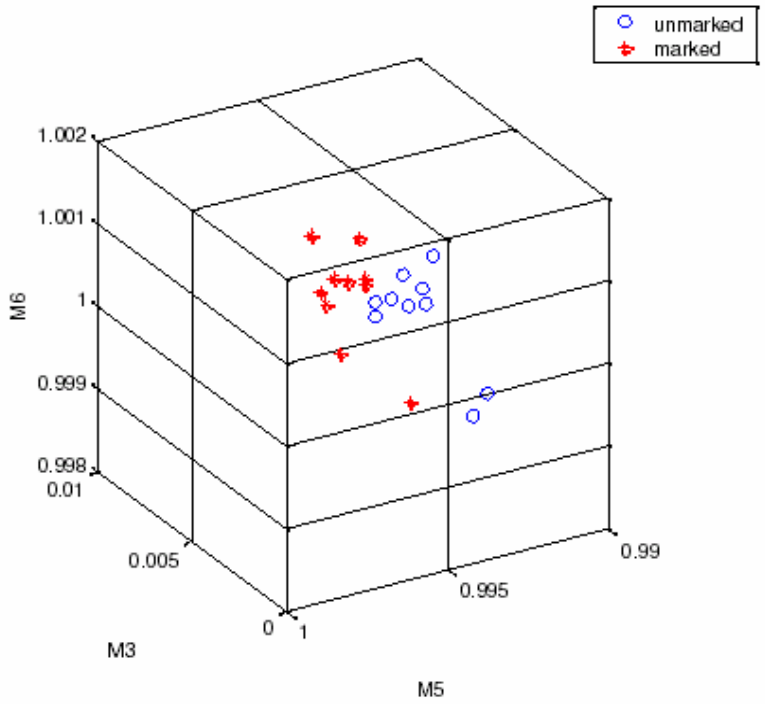
What makes a Feature "good"

- A good feature should be:
  - » Accurate
    - – Detect stego images with high accuracy and low error
  - » Consistent
    - – The accuracy results should be consistent for a set of large images, i.e. features should be independent of image type or texture
  - » Monotonic
    - – Features should be monotonic in their relationship with respect to the message size

IQM

- IQM's can be used as features
- From a set of 26 IQM measures a subset with most discriminative power was chosen
- ANOVA is used to select those metrics that respond best to image distortions due to embedding

IQM



Scatter plot of 3 image quality measures showing separation of marked and unmarked images.

Classifiers

- Different types of classifier used by different authors.
  - » MMSE linear predictor
  - » Fisher linear discriminates as well as a SVM classifier
- SVM classifiers seem to do much better in classification
- All the authors show good results in their experiments, but direct comparison is hard since the setups are very much different.

So What Can Alice (Bob) Do?

- Limit message length so that detector does not trigger
- Use model based embedding.
  - » Stochastic Modulation
- Adaptive embedding
  - » Embed in locations where it is hard to detect.
- Active embedding
  - » Add noise after embedding to mask presence.
  - » Outguess