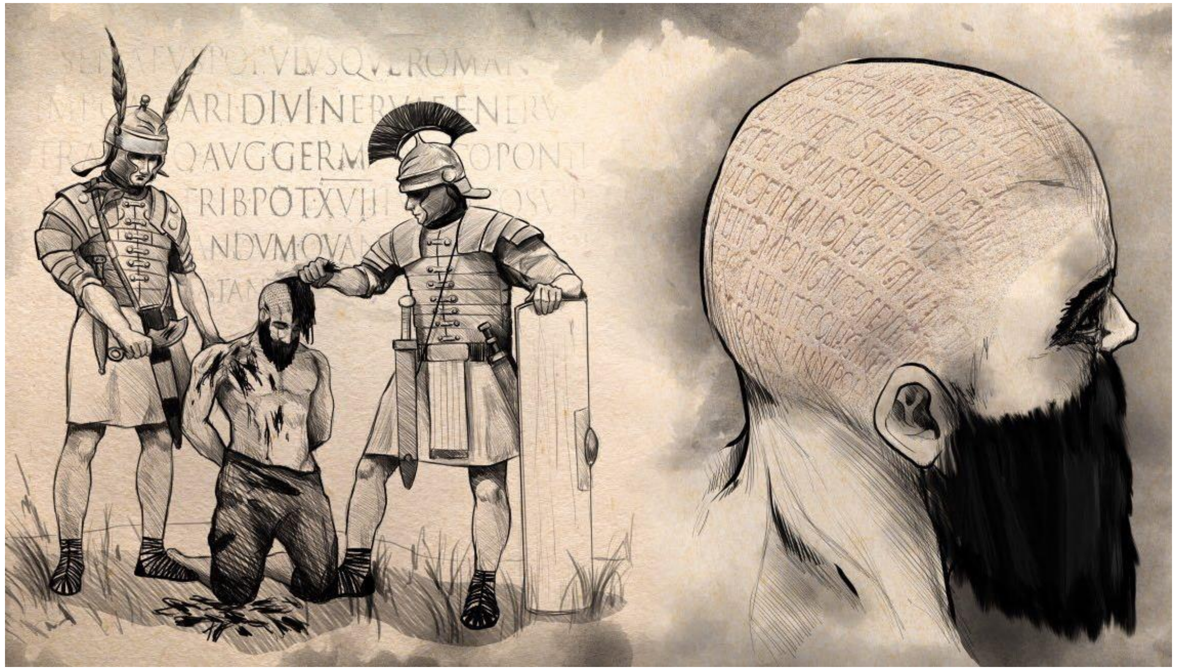
Unit 2 : Steganography and Steganalysis

Basics

- The word steganography comes from the Greek ‘steganos’, meaning covered or secret, and ‘graphy’, meaning writing or drawing. Therefore, steganography literally means covered writing.
- Steganography simply takes one piece of information and hides it within another.
  - Computer files (images, sounds recordings, even disks) contain unused or insignificant areas of data.
  - Steganography takes advantage of these areas, replacing them with information (encrypted mail, for instance).
  - The files can then be exchanged without anyone knowing what really lies inside of them.
  - An image of the space shuttle landing might contain a private letter to a friend.
  - Rumor has it that terrorists used steganography to transmit messages to one another.

History

In the history of the Persian Wars, Herodotus tells of a messenger who shaved his head and allowed a secret message to be tattooed on his scalp. He waited until his hair grew back. Then he journeyed to where the recipient awaited him and shaved his head again. The message was revealed. It was history's first use of steganography.



Invisible Ink

Ancient Romans used to write between lines using invisible ink based on various natural substances such as fruit juices, urine, and milk. Their experience was not forgotten: even nowadays children play spies and write secret messages that appear only when heated.

Invisible Ink

During the World War II the Germans developed the microdot. A secret message was photographically reduced to the size of a period, and affixed as the dot for the letter 'i' or other punctuation on a paper containing a written message. Microdots permitted the transmission of large amounts of printed data, including technical drawings, and the fact of the transmission was effectively hidden.

Steganography vs Cryptography

Steganography	Cryptography
Steganography refers to Cover Writing.	Cryptography refers to Secret Writing.
Used from ancient time till modern era	Used in modern era
Structure of data remains same.	Structure of data can be altered.
Attack in Steganography is termed as Steganalysis.	Attack in Cryptography is termed as Cryptanalysis.
Steganography supports Confidentiality and Authentication.	Cryptography supports Confidentiality, Authentication, Data integrity and Nonrepudiation.
Steganography requires a parameter like key.	Cryptography may not need any key.
Hide the message	Does not hide message
Only sender and receiver knows the existence of message	Everybody knows the existence of message
End result is stego media	End result is cipher text

Classification of Steganography

Steganography is classified into 3 categories,

- **Pure steganography** where there is no stego key. It is based on the assumption that no other party is aware of the communication.
- **Secret key steganography** where the stego key is exchanged prior to communication. This is most susceptible to interception.
- **Public key steganography** where a public key and a private key is used for secure communication.

What it does

- Hides information needing to be kept secret in unused, redundant parts of data.

- Data most often hidden in :
  - Text/ Word Documents
  - Audio files
  - Digital Images
  - Disk Space
  - Software and Circuitry
  - Network Packets
  - Strands of Human DNA (Genome Seg.)

Principles

Computer Steganography is based on two principles.

- The first one is that the files containing digitized images or sound can be altered to a certain extent without loosing their functionality.
- The other principle deals with the human inability to distinguish minor changes in image color or sound quality, which is especially easy to make use of in objects that contain redundant information, be it 16-bit sound, 8-bit or even better 24-bit image. The value of the least significant bit of the pixel color won’t result in any perceivable change of that color.

Techniques

Three common techniques used in Steganography :

- **Substitution** : **LSB Method -** replaces the last bit in a byte.
  - **Advantages:** Simplest approach to hide data in an image file
  - **Disadvantages :** does not take well with file changing.
- **Injection** : embedding the message directly into the carrier object
  - **Disadvantage** : Makes the file size much larger
- **Generation of a new file :** Start from scratch
  - **Advantage :** there is never an original file to compare to

Text Steganography

- Text steganography can be applied in the digital makeup format such as PDF, digital watermark or information hiding.
- It is more difficult to realize the information hiding based on text. The simplest method of information hiding is to select the cover first, adopt given rules to add the phraseological or spelling mistakes, or replace with synonymous words.
- Ex: TextHide hides the information in the manner of text overwriting and words’ selection.

TEXT STEGANOGRAPHY

Since Everyone Can Read, Encoding Text
In Neutral Sentences Is Doubtfully Effective

SECRET INSIDE

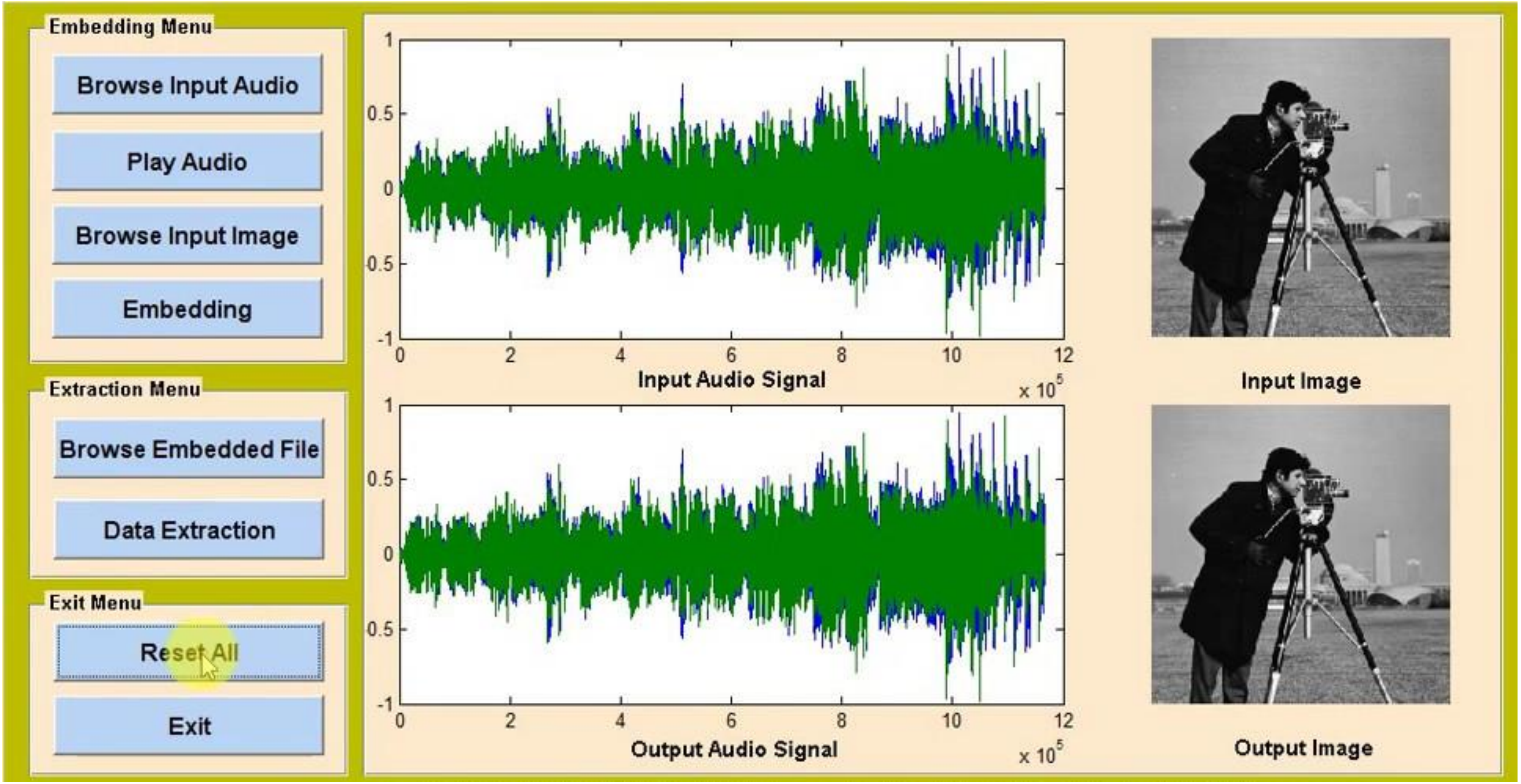
Text Steganography Methods

- Very challenging – small amounts of repetitive data
- **Line-shift encoding**: actually shifting the line of text up or down.
  - Value depends on whether the line was up or down from a stationary line.
- **Word-shift encoding**: uses horizontal space between words, while maintaining the natural spacing appearance, to obtain a value for the hidden message.
- **Feature specific encoding**: encoding a secret message by changing certain attributes such as the length of letters.

- Text Steganography in Markup Languages (HTML)
- Text Steganography in Specific characters in words
- Open Spaces
- Semantic Methods

Audio Steganography

- Embedding secret messages into digital sound is known as audio Steganography.
- Audio Steganography methods can embed messages in WAV, AU, and even MP3 sound files.
- The properties of the human auditory system (HAS) are exploited in the process of audio Steganography


Embedding Menu
Browse Input Audio
Play Audio
Browse Input Image
Embedding
Extraction Menu
Browse Embedded File
Data Extraction
Exit Menu
Reset All
Exit
1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
Input Audio Signal
x 10^5
Input Image
1
0.5
0
-0.5
-1
0
2
4
6
8
10
12
Output Audio Signal
x 10^5
Output Image

Audio Steganography Methods

The most challenging

- **Sample Quantization**: 16-bit linear sampling architecture used for WAV and AIFF formats.
- **Temporal Sampling Rate**: uses selectable frequencies; the higher the sample rate the more usable data will become.
- **Perceptual Sampling**: usually an MP3 file, changes statistics of audio by only encoding parts listener can perceive.
  - Perceived sound maintained but actual signal is changed.

Image Steganography

- Using image files as hosts for steganographic messages takes advantage of the limited capabilities of the human visual system.
- Some of the more common method for embedding messages in image files can be categorized into two main groups, image domain methods and transform domain methods.
- Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover image

1 0 10100010
01000101 0 01010
0 1 0 1010001010 010
01010 0101010 101010
01 00 1 0 1010001010 0
1010001010 0101010 10
1010 0101010 101010
0001010 0101010 1010
1010 101010
0001010 01010

Digital Images

- Anything that can be encoded in a bit-stream can be hidden in digital image
- Cover Images can be:
  - **8-bit**:
    - 256 color choices
    - Small file size
  - **24-bit**:
    - 16 million color choices
    - Large file size
  - **Gray-scale**:
    - Shades slightly from byte to byte

Image Compression

Image compression offers a solution to large image files. Two kinds of image compression are lossless and lossy compression. Both methods save storage space but have differing effects on any uncompressed hidden data in the image.

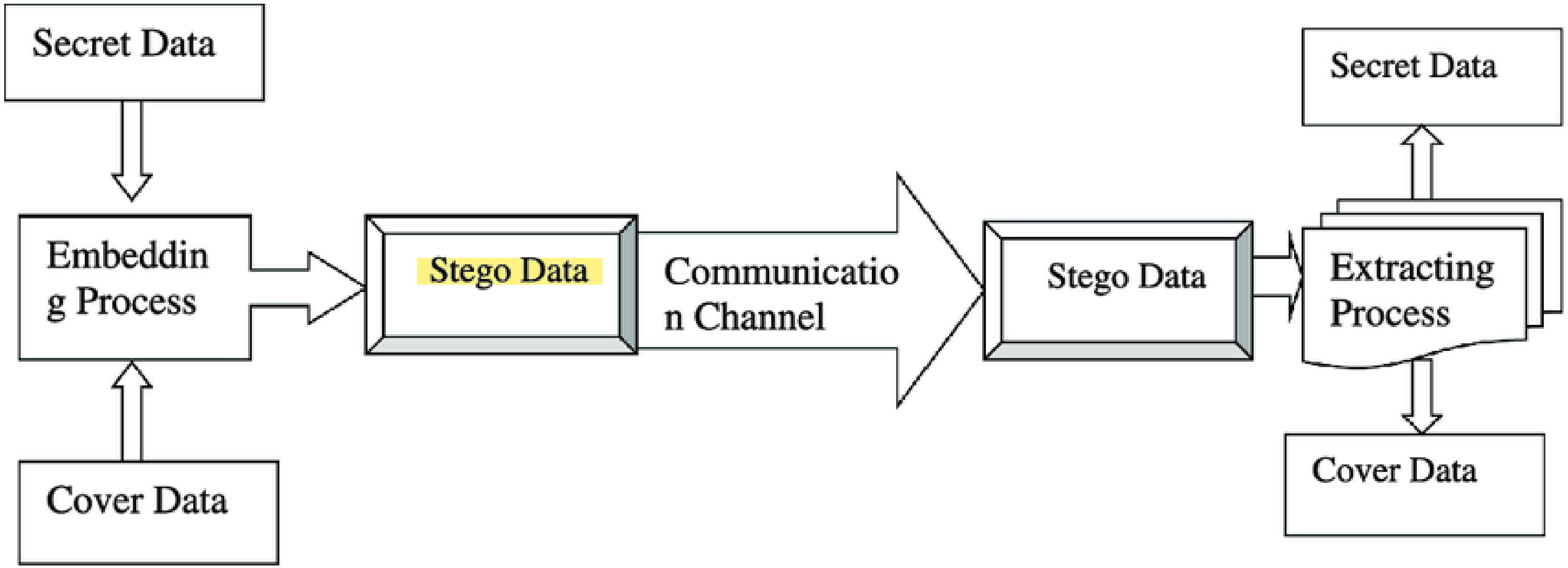
- “**Lossy**” JPEG(Joint Photographic Experts Group) format files, offers high compression, but may not maintain the original image's integrity. Hence it is called “lossy”.

- “**Lossless**” compression maintains the original image data exactly, It is thus more favored by steganographic techniques. Eg: (BMP),(GIF) Formats.

Steganography Process

- The data to be concealed is compressed and hidden within another file.
- The first step is to find a file which will be used to hide the message (also called a carrier or a container.)
- The next step is to embed the message one wants to hide within the carrier using a steganographic technique.
- Two different techniques commonly used for embedding are:
  - Replace the least significant bit of each byte in the [carrier] with a single bit for the hidden message.
  - Select certain bytes in which to embed the message using a random number generator; resampling the bytes to pixel mapping to preserve color scheme, in the case of an image; hiding information in the coefficients of the discrete cosine, fractal or wavelet transform of an image; and applying mimic functions that adapt bit pattern to a given statistical distribution."

Block Diagram of Steganography



Steganography Terms

- **Carrier or Cover File** - A Original message or a file in which hidden information will be stored inside of it .
- **Stego-Medium** - The medium in which the information is hidden.
- **Stego** – The resulted message after embedding the secret message into cover.
- **Stego Image** – Image with the hidden information.
- **Non Stego Image** – Natural image with no hidden information
- **Embedded or Payload** - The information which is to be hidden or concealed.
- **True Positive** – while testing, if a test image is correctly detected as a stego image.
- **True Negative** – while testing, if a test image is correctly identified as a non stego image.
- **False Positive** – while testing, if a test image is incorrectly detected as a stego image.
- **False Negative** – while testing, if a test image is incorrectly identified as a non stego image.

Possible uses of Steganography

- Combine explanatory information within an image.
- Protection of data alteration
- Confidential communications and secret data storing
- Access control system for digital content distribution
- Media Database Systems
- Usage in modern printers
- Maintaining anonymity
- Alleged use by intelligence services

Drawbacks

- Could degrade or render an image
- Could counteract and be counterproductive with original image
- Password leakage may occur and it leads to the unauthorized access of the data
- Third parties with steganography detection and cracking tools can view the message
- Steganographic software can’t protect the watermark
- Easier to open free Web-based e-mail

Legitimate Usage

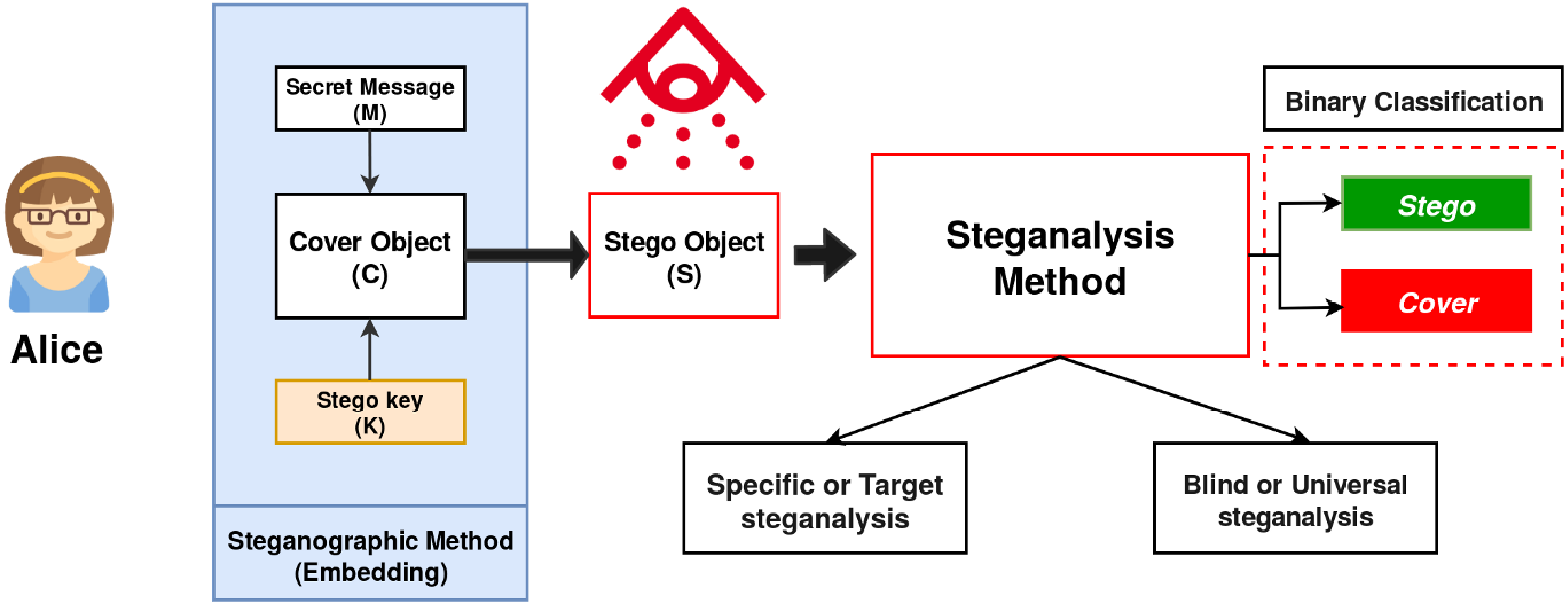
- **Digital Watermarking**
  - Prevent illegal modification, copying, distribution
    - Eg. DVD recorders detect copy protection on DVDs that contain embedded authorizations
  - Identify in Ownership disputes, content authentication
- **Provide explanatory information with an image** (like doctor's notes accompanying an X-ray)
- **Printers**
  - Tiny yellow dots, barely visible, contains date and time-stamps, encoded serial numbers
- **Used to hide the existence of sensitive files on storage media**

Illegitimate Usage

- **Corporate Espionage**
  - Theft of trade secrets
- **Terrorism**
  - USA Today article by Jack Kelly – “ Terror groups hide behind web encryption” (Feb 5, 2001)
  - Hiding secrets in websites like e-bay, amazon, porn websites, transmission via chat rooms, P2P sharing networks, etc.
  - However, no official proof or record has been produced
- **Child Pornography**

Steganalysis

- Steganalysis is the practice of attacking steganography methods for the detection, extraction, destruction and manipulation of the hidden data in a stego object.
- Attacks can be of several types for example, some attacks merely detect the presence of hidden data, some try to detect and extract the hidden data, some just try to destroy the hidden data by finding the existence without trying to extract hidden data and some try to replace hidden data with other data by finding the exact location where the data is hidden.


Alice
Secret Message (M)
Cover Object (C)
Stego key (K)
Steganographic Method (Embedding)
Stego Object (S)
Steganalysis Method
Binary Classification
Stego
Cover
Specific or Target steganalysis
Blind or Universal steganalysis

Classification of attacks

- **Stego only attack:** only stego object is available for analysis.
- **Known cover attack**: both cover and stego are known.
- **Known message attack**: in some cases message is known and analyzing the stego object pattern for this embedded message may help to attack similar systems.
- **Chosen stego attack**: steganographic algorithm and stego object are known.
- **Chosen message attack**: here steganalyst creates some sample stego objects from many steganographic tools for a chosen message and analyses these stego objects with the suspected one and tries to find the algorithm used.
- **Known stego attack**: cover object and the steganographic tool used are known.

Approaches of Steganalysis

Visual attacks: By analyzing the images visually, like considering the bit images and try to find the difference visually in these single bit images.

Structural attacks: The format of data file often changes as the data to be hidden is embedded, identifying these characteristic structural changes can detect the existence of image, for example in palette based steganography the palette of image is changed before embedding data to reduce the number of colors so that the adjacent pixel color difference should be very less. This shows that groups of pixels in a palette have the same color which is not the case in normal images.

Statistical attacks: In these type of attacks the statistical analyses of the images by some mathematical formulas is done and the detection of hidden data is done based on these statistical results. Generally, the hidden message is more random than the original data of the image thus finding the formulae to know the randomness reveals the existence of data.

Steganalysis Techniques

1. **Signature Steganalysis :** Steganography methods hide secret information and manipulate the images and other digital media in ways as to remain imperceptible to human eye. Steganography alters the media properties due to the insertion of message bits in the form of degradation or repeated patterns, which act as signatures that convey the existence of embedded message. For detecting the existence of hidden message in a suspicious image is to look for these repetitive patterns signatures of a steganography tool .These particular signatures automatically exploit the tool used in embedding the messages. Such methods looks at palette tables in GIF images and any anomalies caused there by common stego tools. When the message is embedded sequentially such attacks give promising results but, are hard to automatize and their reliability is highly doubtful.

2. **Statistical steganalysis :** The statistics of an image undergo alterations due to information hiding. Statistical steganalysis analyses the underlying statistics of an image to detect the secret embedded information. Statistical steganalysis is more commanding than signature steganalysis, because mathematical techniques are more responsive than visual perception

2.1 **Specific statistical steganalysis**: These types of techniques are established by analyzing the embedding operation and determining certain image statistics. Such techniques need a detailed knowledge of embedding process. These techniques capitulate very accurate results when used against a target steganography technique. Specific statistical steganalytic tools is used for finding secret message from stego-images embedded by LSB embedding, LSB matching, spread spectrum, JPEG compression and other transform domain.

2.2 **Universal statistical steganalysis**: Universal statistical steganalysis comprise the statistical steganalysis method that is not tailored for a specific steganography embedding method. It requires less or even no priori information of the under attack steganographic methods for detection of secret message. It uses a learning based strategy which involves training based on cover and stego-images. Neural network, clustering algorithms and other soft computing tools are used to construct the detection model from the experimental data. These techniques do not depend on the behavior of embedding algorithms.