Information hiding and steganography

Information Hiding

¨ Information Hiding is a branch of computer science that deals with concealing the existence of a message

¨ It is related to cryptography whose intent is to render messages unreadable except by the intended recipients ¨ It employs technologies from numerous science disciplines:

– Digital Signal Processing (Images, Audio, Video)

– Cryptography

– Information Theory\Coding Theory

– Data Compression

– Human Visual/Auditory perception

¨ There are four primary sub-disciplines of Information Hiding

– Steganography

– Watermarking

– Covert Channels

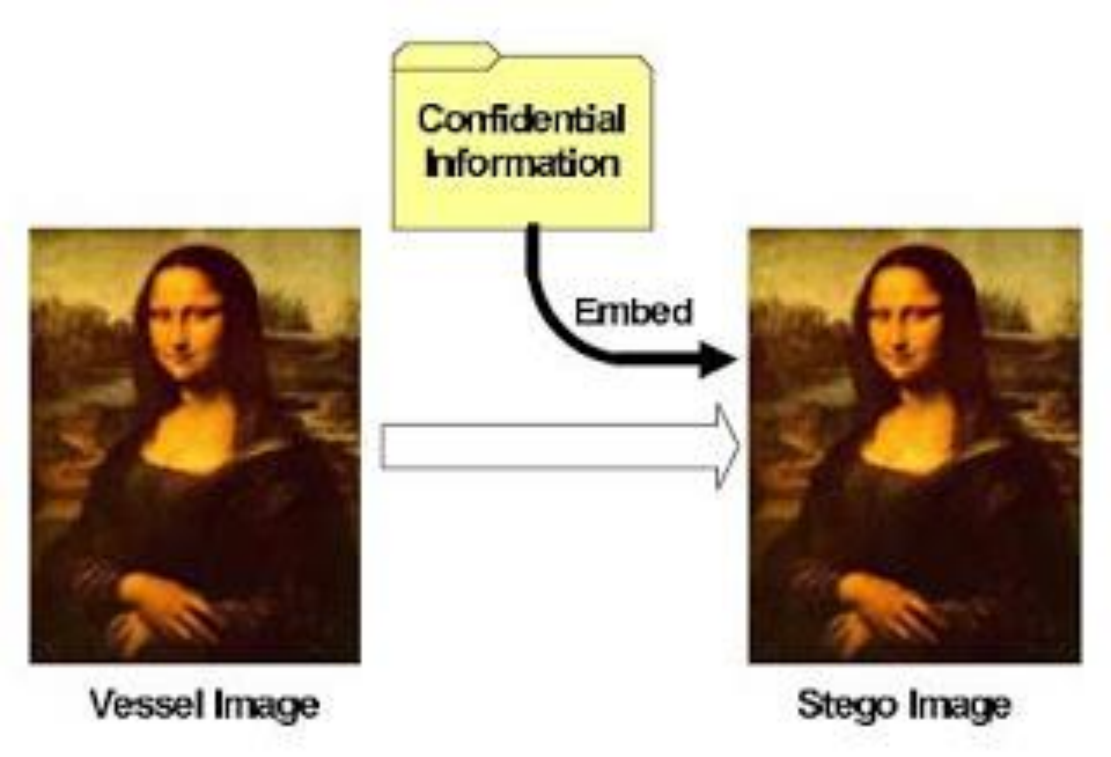
– Anonymity

- Encryption and data hiding are two technologies that play major roles in information security and assurance

Digital Steganography

Steganography

- The art of hiding data in a file so that only the sender and intended recipient suspect the presence of hidden data
  - A form of security through obscurity
- Very easy to accomplish
- Harder to detect and decrypt
- BMP, JPG, TXT, HTML/XML, PDF, PNG, GIF, AU, WAV, MP3, AVI, TIF, TGA, DLL, EXE



- **Steganography** is the art and science of writing hidden messages in such a way that no one apart from the intended recipient knows of the existence of the message.
- This can be achieve by concealing the existence of **information** within seemingly harmless **carriers or cover**
- **Carrier:** text, image, video, audio, etc.

Information Hiding – Watermarking, Steganography and Cryptography

There are two major branches of information hiding, **Steganography** and **Watermarking**

Watermarking:

- Communication in watermarking is the host signal, with the embedded data providing copyright protection.
- The existence of a watermark is often declared openly.
- Any attempt to remove or invalidate the embedded content renders the host useless.

Cryptography:

- Doesn’t conceal the communication.
- Scrambles the data to prevent eavesdroppers understanding the content.
- Cryptography involves various methods and implementations.
- May be considered complementary and orthogonal (unrelated).
- Once the presence of hidden information is revealed or even suspected, the purpose of steganography is defeated.

Information Hiding – Watermarking, Steganography and Cryptography

- As cryptanalysis is the counterpart of cryptography, steganalysis is the counterpart of steganography.
- A *steganalyst* tries to determine the existence of a covert communication channel between two parties and either break or alter their communication.
- While cryptology states that a cipher is broken when the attacker is able to gain information on the content of the payload, a steganography technique is considered broken when its mere existence is proven.

Information Hiding – Watermarking, Steganography and Cryptography

- The main purpose of Steganography, which means ‘writing in hiding’ is to hide data in a cover media so that others will not be able to notice it.
- While cryptography is about protecting the content of messages, steganography is about concealing their very existence

Steganography

- Steganography equation is ‘Stego-medium = Cover medium + Secret message + Stego key’.
- The general model of data hiding can be described as follows. The embedded data is the message that one wishes to send secretly. It is usually hidden in an innocuous message referred to as a cover-text or cover-image or cover-audio as appropriate, producing the stego-text or other stego-object.
- A stego-key is used to control the hiding process so as to restrict detection and /or recovery of the embedded data to parties who know it.

Steganography

- While steganography can be achieved using any cover media, we are concerned with hiding data in digital images.
- The features expected of a stego-medium are imperceptibility and robustness, so that the secret message is known only to the intended receiver and also the stego-medium being able to withstand attacks from intruders.
- The amount of secret message embedded should be such that it doesn’t reduce the quality of the stego image.
- The goal of steganography is to embed secret data into a cover in such a way that no one apart from the sender and intended recipients even realizes there is secret data.

Steganography - Applications

- The applications of information hiding systems mainly range over a broad area from military, intelligence agencies, online elections, internet banking, medical-imaging and so on.
- In the current situation digital images are the most popular carrier/cover files that can be used to transmit secret information.

Steganography - Applications

- To have secure secret communications where cryptographic encryption methods are not available.
- To have secure secret communication where strong cryptography is impossible.
- In some cases, for example in military applications, even the knowledge that two parties communicate can be of large importance.
- The health care, and especially medical imaging systems, may very much benefit from information hiding techniques.

Classification of Steganographic methods

- 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.
- Public key steganography where a public key and a private key is used for secure communication.

CLASSIFICATION OF STEGANOGRAPHY TECHNIQUES

1. Spatial Domain Techniques
2. Transform Domain Techniques
3. Spread Spectrum
4. Statistical Method
5. Distortion Technique

Spatial Domain

- These techniques use the pixel gray levels and their color values directly for encoding the message bits. These techniques are some of the simplest schemes in terms of embedding and extraction complexity.
- The major drawback of these methods is amount of additive noise that creeps in the image which directly affects the Peak Signal to Noise Ratio and the statistical properties of the image.
- Moreover these embedding algorithms are applicable mainly to lossless image compression schemes like TIFF images. For lossy compression scheme like JPEG, some of message bits get lost during compression step.

Transform Domain

- These techniques try to encode message bits in the transform domain coefficients of the image. Data embedding performed in transform domain is widely used for robust watermarking.
- Similar techniques can also realize large capacity embedding for Steganography. Candidate transforms include discrete cosine Transform (DCT), discrete wavelet transform (DWT), and discrete Fourier transform (DFT).
- By being embedded in the transform domain, the hidden data resides in more robust areas, spread across the entire image, and provides better resistance against signal processing. Eg: we can perform a block DCT and, depending on pay-load and robustness requirements, choose one or more components in each block to form a new data group that, in turn, is pseudo randomly scrambled and undergoes a second-layer transformation.

- Modification is then carried out on double transform domain coefficients using various schemes.
- These techniques have high embedding and extraction complexity. Because of robustness properties of transform domain embedding, these techniques are more applicable to “Watermarking” aspect of data hiding.

Spread Spectrum

- Spread spectrum transmission in radio communication transmit message below level of noise frequency.
- In Steganography it deals either with cover image as noise or tries to add as pseudo-noise in the cover image.
- Cover Image As Noise: A system that treats the cover image as noise can add a single value to that cover image. This value must be transmitted below that noise level. This means that the channel capacity of the image changes significantly. Thus, while this value can be a real number, in practice, the difficulty in recovering a real number decreases the value to a single bit.

- To permit the transmission of more than one bit, the cover image has to be broken into sub images. When these sub cover images are tiles, the technique is referred to as direct-sequence spread spectrum Steganography.
- When the sub cover images consist of separate points distributed over the cover image, the technique is referred to as frequency-hopping spread-spectrum Steganography. These techniques require searching the image for the carrier in order to then retrieve the data.
- These techniques are robust against gentle JPEG compression and can be made more robust through the pre-distortion of the carrier.

STATISTICAL METHODS

- Also known as model-based techniques, these techniques tend to modulate or modify the statistical properties of an image in addition to preserving them in the embedding process.
- This modification is typically small, and it is thereby able to take advantage of the human weakness in detecting luminance variation.
- Statistical Steganography techniques exploit the existence of a “1-bit”, where nearly a bit of data is embedded in a digital carrier.

STATISTICAL METHODS

- This process is done by simply modifying the cover image to make a sort of significant change in the statistical characteristics if “1” is transmitted, otherwise it is left unchanged.
- To send multiple bits, an image is broken into sub-images, each corresponding to a single bit of the message.

Distortion Technique

- It require original cover image during decoding process where decoder functions to check for differences between original cover image and distorted cover image in order to restore secret message.
- Encoder, adds a sequence of changes to cover image. So, information is described as being stored by signal distortion .
- Using this technique, a stego-object is created by applying sequence of modifications to cover image. This sequence of modifications is selected to match secret message required to transmit. Message is encoded at pseudo-randomly chosen pixels.
- If stego-image is different from cover image at given message pixel, then message bit is a “1.” Otherwise, message bit is a “0.”
- Encoder can modify “1” value pixels in such manner that statistical properties of image are not affected (which is different from many LSB methods).

Steganography Properties

- A few key properties that must be considered when creating a digital data hiding system are
  - Imperceptibility: Imperceptibility is the property in which a person should be unable to distinguish the original and the stego-image.
  - Embedding Capacity: Refers to the amount of secret information that can be embedded without degradation of the quality of the image.
  - Robustness: Refers to the degree of difficulty required to destroy embedded information without destroying the cover image.

Goals of Information Hiding - Security

It is secure if it cannot be removed even with full knowledge of the embedding algorithm without knowledge of the secret key

Can it be detected by human perception? (Invisibility)

– See distortion/noise in an image

– Hear distortion/noise in speech or music?

Can it be detected by statistical analysis? (Undetectability)

Does it leave easily detectable signatures?

Levels of Failure:

– Detection - Proof of existence of message

– Extraction – removing without destroying the cover

– Destruction – destroying the message without destroying the cover

Goals of Information Hiding - Capacity

- ¨ How much data can a cover image hold?
- – There is a physical limit (unless the cover file size is increased)
- – There is a limit as to when the data will be noticeable
- ¨ Typically, as more capacity is used, the lower the security and robustness

Goals of Information Hiding – Robustness

- ¨ How well does the data maintain integrity in the face of modifications?
- ¨ The modifications we are concerned with are quite common
- – Images: blurring, sharpening, scaling, cropping, contrast, gamma, brightness, rotation, skewing, recoloring, printing/copying/scanning, etc.
- – Audio: filtering (think bass/treble), volume adjustment, stereo to mono, etc.
- – Video: any image/audio modification, add/delete frames, temporal adjustments, frame swapping, frame averaging
- – Also: lossy compression, A/D and D/A conversion, and sophisticated attacks
- ¨ Robustness is achieved through redundant encoding of the message which reduces the capacity