13. STEGANOGRAPHY and WATERMARKING

One of the most important property of (digital) information is that it is in principle very easy to produce and distribute unlimited number of its copies.

This might undermine the music, film, book and software industries and therefore it brings a variety of important problems concerning the protection of the intellectual and production rights that badly need to be solved.

The fact that an unlimited number of perfect copies of text, audio and video data can be illegally produced and distributed requires to study ways of embedding copyright information and serial numbers in audio and video data.

Steganography and watermarking bring a variety of very important techniques how to hide important information in an undetectable and/or irremovable way, especially in audio and video data.

Steganography and watermarking are main parts of the fast developing area of **information hiding**.

INFORMATION HIDING SUB-DISCIPLINES

Covert channels , especially in operating systems and networks. They are communication paths that were neither designed nor intended to transfer information at all.

These channels are typically used by untrustworthy programs to leak information to their owner while performing service for another program.

Anonymity is finding ways to hide meta content of the message (for example the sender and/or the recipients of the message). Anonymity is need when making on-line voting or to hide access to some web pages.

Steganography – covered writing - from Greek $\sigma\tau\varepsilon\gamma\alpha\nu\acute{o}$-$\varsigma$, $\gamma\rho\alpha\phi$-$\varepsilon\iota\nu$.

Watermarking - *visible digital watermarks* and also *imperceptible (invisible, transparent,....) watermarks* .

STEGANOGRAPHY versus WATERMARKING

Differences between steganography and watermarking are both subtle and essential.

The main goal of **steganography** is **to hide** a message m in some audio or video (cover)data d, to obtain new data d', practically indistinguishable from d, by people, in such a way that an eavesdropper **cannot detect** the presence of m in d'.

The main goal of **watermarking** is **to hide** a message m in some audio or video (cover)data d, to obtain new data d', practically indistinguishable from d, by people, in such a way that an eavesdropper **cannot remove or replace** m in d'.

It is also often said that the goal of steganography is to hide a message in **one-to-one communications** and the goal of watermarking is to hide message in **one-to-many communications**.

Shortly, one can say that cryptography is about **protecting** the content of messages, steganography is about **concealing** its very existence.

Steganography methods usually do not need to provide strong security against removing or modification of the hidden message. Watermarking methods need to to be very robust to attempts to remove or modify a hidden message.

APPLICATIONS of STEGANOGRAPHY

- 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.

APPLICATIONS of WATERMARKING

A popular application of watermarking techniques is to provide a proof of ownership of digital data by embedding copyright statements into video or image digital products.

Other applications include:

- Automatic monitoring and tracking of copy-write material on WEB. (For example, a robot searches the Web for marked material and thereby identifies potential illegal issues.
- Automatic audit of radio transmissions: (A robot can “listen” to a radio station and look for marks, which indicate that a particular piece of music, or advertisement , has been broadcast.
- Data augmentation – to add information for the benefit of the public.
- Fingerprinting applications (in order to distinguish distributed data)

Actually, watermarking has recently emerged as the leading technology to solve the above very important problems.

All kind of data can be watermarked: audio, images, video, formatted text, 3D models, model animation parameters, . . .

STEGANOGRAPHY/WATERMARKING

versus CRYPTOGRAPHY

The purpose of both is to provide secret communication.

Cryptography hides the contents of the message from an attacker, but not the existence of the message.

Steganography/watermarking even hide the very existence of the message in the communicating data.

Consequently, **the concept of breaking the system** is different for **cryptosystems** and **stegosystems (watermarking systems)**.

- A cryptographic system is broken when the attacker can read the secrete message.
- Breaking of a steganographic/watermarking system has two stages:
  - The attacker can detect that steganography/watermarking has been used;
  - The attacker is able to read, modify or remove the hidden message

A steganography/watermarking system is considered as insecure already if the detection of steganography/watermarking is possible.

FIRST STEGANOGRAPHIC METHODS

- Ancient Chinese wrote messages on fine silk, which then crunched into a tiny ball and covered in wax. The messenger would then swallowed the ball of wax.
- In the sixteenth century, the Italian scientist Giovanni Porta described how to conceal a message within a hard-boiled egg by making an ink from a mixture of one ounce of alum and a pint of vinegar, and then using it to write on the shell. The solution penetrates the porous shell, and leaves a message on the surface of the hardened egg albumen, which can be read only when the shell is removed.
- Special “inks” were important steganographic tools even during Second World War.
- During Second World War a technique was developed to shrink photographically a page of text into a dot less than one millimeter in diameter, and then hide this microdot in an apparently innocuous letter. (The first microdot has been spotted by FBI in 1941.)

FIRST STEGANOGRAPHY BOOKS

A variety of methods was used already in Roman times and then in 15-16 century (ranging from coding messages in music, and string knots to invisible inks).

In 1499 Johannes Trithemius, opat from Würzburg, wrote 3 out of planned 8 books “Steganographia”.
{Description of letters-words and systems+methods reminding telepathy.}

In 1518 Trithemius printed 6 books, 540 pages, on cryptography and steganography. Books’ title: **Polygraphiae**.

This is Trithemius’ most notorious work. It includes a sophisticated system of steganography, as well as angel magic. It also contains a synthesis of the science of knowledge, the art of memory, magic, an accelerated language learning system, and a method of sending messages without symbols or messenger.

In 1665 Gaspari Schotti published the book “Steganographica”, 400 pages. (New presentation of Trithemius.)

GENERAL STEGANOGRAPHIC MODEL

A general model of a cryptographic system has already emerged.

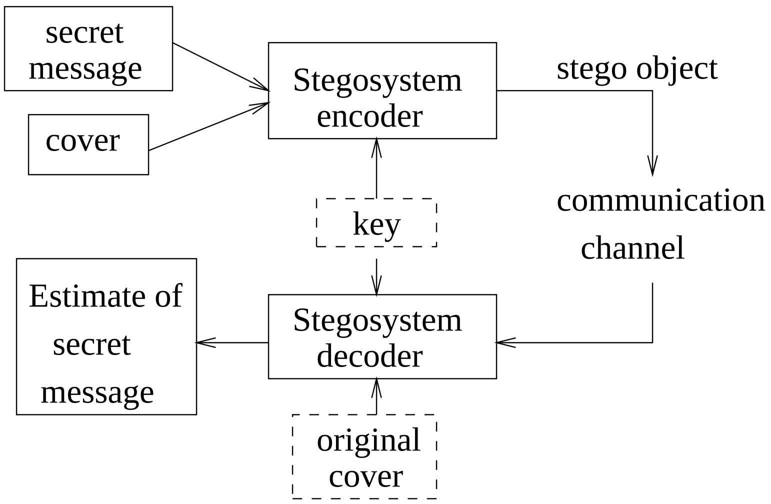

Figure 1: Model of steganographic systems

Steganographic algorithms are in general based on replacing *noise component* of a digital object with a to-be-hidden message.

Kirchoffov principle holds also for steganography. Security of the system should not be based on hiding embedding algorithm, but on hiding the key.

BASIC CONCEPTS of STEGOSYSTEMS

- **Covertext (cover-data - cover-object)** is an original unaltered message.
- **Embedding process** in which the sender, Alice, tries to hide a message by embedding it into a (randomly chosen) cover-text, usually using a key, to obtain a stego-text (stego-data or stego-object).

  The embedding process can be described by the mapping $E : C \times K \times M \rightarrow C$, where C is the set of possible cover- and stego-texts, K is the set of keys and M is the set of messages.
- **Stegotext (stego-data - stego-object)**
- **Recovering process** (or extraction process) in which the receiver, Bob, tries to get, using the key only, not the covertext, the hidden message in the stegotext.

  The recovery process can be seen as mapping $D : C \times K \rightarrow C$.
- **Security requirement** is that a third person watching such a communication should not be able to find out whether the sender has been active, and when, in the sense that he really embedded a message in the cover -text. In other words, stegotexts should be indistinguishable from covertexts.

BASIC TYPES of STEGOSYSTEMS

There are three basic types of stegosystems

- Pure stegosystems - no key is used.
- Secret-key stegosystems - secret key is used.
- Public-key stegosystems - public key is used.

Definition **Pure stegosystem** $\mathcal{S} = \langle C, M, E, D \rangle$, where C is the set of possible covertexts, M is the set of secret messages, $|C| \geq |M|$, $E : C \times M \rightarrow C$ is the embedding function and $D : C \rightarrow M$, is the extraction function, with the property that $D(E(c, m)) = m$, for all $m \in M$ and $c \in C$.

Security of the pure stegosystems depends completely on its secrecy. On the other hand, security of other two stegosystems depends on the secrecy of the key used.

Definition **Secret-key stegosystem** $\mathcal{S} = \langle C, M, K, E_K, D_K \rangle$, where C is the set of possible covertexts, M is the set of secret messages with $|C| \geq |M|$, K is the set of secret keys, $E_K : C \times M \times K \rightarrow C$, $D_K : C \times K \rightarrow M$ with the property that $D_K(E_K(c, m, k), k) = m$ for all $m \in M$, $c \in C$ and $k \in K$.

LINGUISTIC STEGANOGRAPHY

A variety steganography techniques allowed to hide messages in formatted texts.

- Acrostic. A message is concealed into certain letters of the text, for example into the first letters of some words.

  Tables have been produced, the first one by Trithentius, called Ave Maria, how to replace plaintext letters by words.

- An improvement of the previous method is to distribute plaintext letters randomly in the cover-text and then use a mask to read it.
- The presence of errors or stylistic features at predetermined points in the cover data is another way to select the location of the embedded information.
- Line shifting encoding.
- Word shifting encoding.
- Data hiding through justifications.
- Feature coding (for example in vertical lines of letters $b,\ d,\ h\ k$).

PERFECT SECRECY of STEGOSYSTEMS

In order to define secrecy of a stegosystems we need to consider

- probability distribution P_C on the set C of covertexts;
- probability distribution P_M on the set M of secret messages;
- probability distribution P_K on the set K of keys;
- probability distribution P_S on the set $\{E_K(c, m, k) \mid c \in C, m \in M, k \in K\}$ of stegotexts.

The basic related concept is that of the relative entropy $D(P_1, || P_2)$ of two probability distributions P_1 and P_2 defined on a set Q by

$$D(P_1||P_2) = \sum_{q \in Q} P_1(q) \lg \frac{P_1(q)}{P_2(q)},$$

which measures the inefficiency of assuming that the distribution on Q is P_2 if it is really P_1.

Definition Let $\mathcal{S}$ be a stegosystem, P_C the probability distribution on covertexts C and P_S the probability distribution of the stegocovers and $\varepsilon > 0$. $\mathcal{S}$ is called ε-secure against passive attackers, if

$$D(P_C||P_S) \leq \varepsilon$$

and **perfectly secure** if $\varepsilon = 0$.

PERFECTLY SECURE STEGOSYSTEM

A perfectly secure stegosystem can be constructed out of ONE TIME-PAD CRYPTOSYSTEM

Theorem There exist perfectly secure stegosystems.

Proof. Let n be an integer, $C_n = \{0,1\}^n$ and P_{C_n} be the uniform distribution on C_n, and $m \in C_n$ a secret message.

The sender selects randomly $c \in C_n$, computes $c \oplus m = s$. The resulting stegotexts are uniformly distributed on C_n and therefore $P_C = P_S$ from what it follows that

$$D(P_{C_n}||P_S) = 0.$$

In the extraction process the message m can be extracted from s by computation

$$m = s \oplus c.$$

DETECTING SECRET MESSAGES

The main goal of a passive attacker is to decide whether data sent to Bob by Alice contain secret message or not.

The above task can be formalized as a statistical hypothesis-testing problem with the test function $F : C \to \{0, 1\}$:

$$f(c) = \begin{cases} 1, & \text{if } c \text{ contains a secret message;} \\ 0, & \text{otherwise} \end{cases}$$

There are two types of errors possible:

- Type-I error - a secret message is detected in data with no secret message;
- Type-II error - a hidden secret message is not detected

Practical steganography tries to minimize probability that passive attackers make type-II error.
In the case of ε-secure stegosystems there is well know relation between the probability β of the type II error and probability α of the type I error.
Theorem let $\mathcal{S}$ be a stegosystem which is ε-secure against passive attackers, β the probability that the attacker does not detect a hidden message and α the probability that the attacker falsely detects a hidden message. Then

$$d(\alpha, \beta) \leq \varepsilon,$$

where $d(\alpha, \beta)$ is the binary relative entropy defined by

$$d(\alpha, \beta) = \alpha \lg \frac{\alpha}{1 - \beta} + (1 - \alpha) \lg \frac{1 - \alpha}{\beta}.$$

INFORMATION HIDING in NOISY DATA

One of the basic methods of steganography is to utilize the existence of redundant information in a communication process.

Images and digital sounds naturally contain such redundancies in the form of noise components.

For images and digital sounds it is naturally to assume that a cover-data are represented by a sequence of numbers and their least significant bits (LSB) represents noise.

If cover-data are represented by numbers

$$c_1, c_2, c_3, \dots,$$

then one of the most basic steganographic method is to replace in some of c_i's, chosen using an algorithm and a key, the least significant bits by the bits of the message to-be-hidden.

Unfortunately, this method does not provide high level of security and it can change significantly statistical properties of the cover-data.

ROBUSTNESS of STEGOSYSTEMS

Steganographic systems are extremely sensitive to cover modifications, such as

- image processing techniques (smoothing, filtering, image transformations, ...);
- filtering of digital sounds;
- compression techniques.

Informally, a stegosystem is **robust** if the embedded information cannot be altered without making substantial changes to the stego-objects.

Definition Let $\mathcal{S}$ be a stegosystem and $\mathcal{P}$ be a class of mappings $C \to C$. $\mathcal{S}$ is $\mathcal{P}$-robust, if for all $p \in \mathcal{P}$

$$D_K(p(E_K(c,m,k)),k) = D_K(E_K(c,m,k),k) = m$$

in the case of a secret-key stegosystem and

$$D(p(E(c,m))) = D(E(c,m)) = m$$

in the case of pure stegosystem, for any m, c, k.

- There is a clear tradeoff between *security* and *robustness*.
- Some stegosystems are designed to be robust against a specific class of mappings (for example JPEG compression/decompression).
- There are two basic approaches to make stegosystems robust:
  - By foreseeing possible cover modifications, the embedding process can be robust so that possible modifications do not entirely destroy embedded information.
  - Reversing operations that has been made by an active attacker.

ACTIVE and MALICIOUS ATTACKS

At the design of stegosystems special attention has to be paid to the presence of active and malicious attackers.

- Active attackers can change cover during the communication process.
- An attacker is malicious if he forges messages or initiates a steganography protocol under the name of one communicating party.

In the presence of a malicious attacker, it is not enough that stegosystem is robust.

If the embedding method does not depend on a key shared by the sender and receiver, then an attacker can forge messages, since the recipient is not able to verify sender's identity.

Definition A steganographic algorithm is called secure if

- Messages are hidden using a public algorithm and a secret key. The secret key must identify the sender uniquely.
- Only the holder of the secret key can detect, extract and prove the existence of the hidden message. (Nobody else should be able to find any statistical evidence of a message's existence.)
- Even if an enemy gets the contents of one hidden message, he should have no chance of detecting others.
- It is computationally infeasible to detect hidden messages.

TYPES of STEGO–ATTACKS

Stego-only attack Only the stego-object is available for stegoanalysis.

Known cover attack The original cover-object and stego-object are both available.

Known message attack Sometimes the hidden message may become known to the stegoanalyser. Analyzing the stego-object for patterns that correspond to the hidden message may be beneficial for future attacks against that system. (Even with the message, this may be very difficult and may even be considered equivalent to the stego-analysis.)

Chosen stego attack (The stegoanalysis generates a stego-object from some steganography tool or algorithm from a chosen message. The goal in this attack is to determine corresponding patterns in the stego-object that may point to the use of specific steganography tools or algorithms.

Known stego attack The steganography algorithm is known and both the original and stego-objects are available.

BASIC STEGANOGRAPHIC TECHNIQUES

Substitution techniques substitute redundant part of the cover-object with a secret message.

Transform domain techniques embed secret message in a transform space of the signal (e.g. in the frequency domain).

Spread spectrum techniques embed secret messages adopting ideas from spread spectrum communications.

Statistical techniques embed messages by changing some statistical properties of the cover-objects and use hypothesis-testing methods in the extraction process.

Distortion techniques store secret messages by signal distortion and measure the deviation from the original cover in the extraction step.

Cover generation techniques embed messages in the way a cover-object for secret communication is created.

COVER DATA

A **cover-object** or, shortly, a **cover** c is a sequence of numbers c_i, $i = 1, 2, \ldots, |c|$.

Such a sequence can represent digital sounds in different time moments, or a linear (vectorized) version of an image.

$c_i \in \{0, 1\}$ in case of binary images and, usually, $0 \leq c_i \leq 256$ in case of quantized images or sounds.

An **image** C can be seen as a discrete function assigning a color vector $c(x, y)$ to each pixel $p(x, y)$.

A color value is normally a three-component vector in a **color space**. Often used are the following color spaces:

RGB-space - every color is specified as a weighted sum of a *red*, *green* and a *blue* component. A vector specifies intensities of these three components.

YCbCr-space It distinguishes a *luminance* Y and two *chrominance* components (Cb, Cr).

Note A color vector can be converted to YCbCr components as follows:

$$Y = 0.299R + 0.587G + 0.114B$$

$$Cb = 0.5 + (B - Y)/2$$

$$Cr = 0.5 + (R - Y)/1.6$$

BASIC SUBSTITUTION TECHNIQUES

- LSB substitution - the LSB of an i-th binary block c_{k_i} is replaced by the bit m_i of the secret message.

  The methods differ by techniques how to determine k_i for a given i.

  For example, $k_{i+1} = k_i + r_i$, where r_i is a sequence of numbers generated by a pseudo-random generators.
- Substitution into parity bits of blocks. If parity bit of the block c_{k_i} is m_i, then the block c_{k_i} is not changed; otherwise one of its bits is changed.
- Substitution in binary images. If image c_i has more (less) black pixels than white pixels and $m_i = 1$ ($m_i = 0$), then c_i is not changed; otherwise the portion of black and white pixels is changed (by making changes at those pixels that are neighbors of pixels of the opposite color).
- Substitution in unused or reserved space in computer systems.

HISTORY of WATERMARKING

Paper watermarks appeared in the art of handmade papermarking 700 hundred years ago.

Watermarks were mainly used to identify the mill producing the paper and paper format, quality and strength.

It was a perfect technique to eliminate confusion from which mill paper is and what are its parameters.

Legal power of watermarks has been demonstrated in 1887 in France when watermarks of two letters, presented as a piece of evidence in a trial, proved that the letters had been predated, what resulted in the downfall of a cabinet and, finally, the resignation of the president Grévy.

Paper watermarks in bank notes or stamps inspired the first use of the term *water mark* in the context of digital data.

The first publications that really focused on watermarking of digital images were from 1990 and then in 1993.

EMBEDDING and RECOVERY SYSTEMS

in WATERMARKING SYSTEMS

Figure 2 shows the basic scheme of the **watermarks embedding systems**.

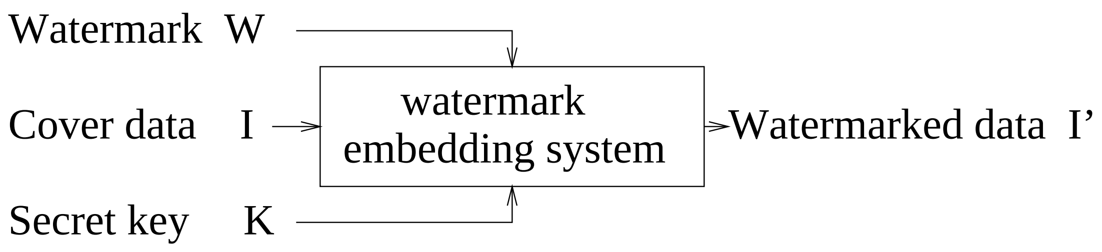

Figure 2: Watermark embedding scheme

Inputs to the scheme is the **watermark**, the **cover data** and an optional **public or secret key**. The output are **watermarked data**. The key is used to enforce security.

Figure 3 shows the basic scheme for **watermark recovery schemes**.

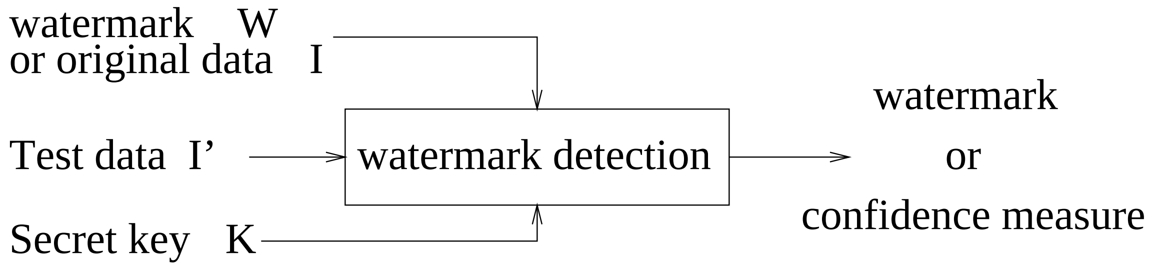

Figure 3: Watermark recovery scheme

Inputs to the scheme are the **watermarked data**, the **secret or public key** and, depending on the method, the **original data and/or the original watermark**. The output is the recovered watermarked W or some kind of **confidence measure indicating how likely it is for the given watermark at the input to be present in the data under inspection**.

TYPES of WATERMARKING SCHEMES

Private (non-blind) watermarking systems require for extraction/detection the original cover-data.

- Type I systems use the original cover-data to extract the watermark from stego-data and use original cover-data to determine where the watermark is.
- Type II systems require a copy of the embedded watermark for extraction and just yield a yes/no answer to the question weather stego-data contains a watermark.

Semi-private (semi-blind) watermarking does not use the original cover-data for detection, but tries to answer the same question. (Potential application of blind and semi-blind watermarking is for evidence in court ownership,....)

Public (blind) watermarking – neither cover-data nor embedded watermarks are required for extraction – this is the most challenging problem.

INVISIBLE COMMUNICATIONS

We describe some important cases of information hiding.

Subliminal channels. We have seen how to use a digital signature scheme to establish a subliminal channel for communication.

Covert channels in operating systems. Covert channels can arise when one part of the system, operating at a specific security level, is able to supply a service to another system part with a possibly different security level.

Video communicating systems. Steganography can be used to embed secret messages into a video stream recorded by video conferencing systems.

Data hiding in executable files. Executable files contain a lot of redundancies in the way independent instructions are scheduled or an instruction subset is chosen to solve a specific problem. This can be utilized to hide messages.

SECRET SHARING by SECRET HIDING

A simple technique has been developed, by Naor and Shamir, that allows for a given n and $t < n$ to hide any secret (image) message m in images on transparencies in such a way that each of n parties receives one transparency and

- no $t-1$ parties are able to obtain the message m from the transparencies they have.
- any t of the parties can easily get (read or see) the message m just by stacking their transparencies together and aligning them carefully.

TO REMEMBER!!!!

There is no use in trying, she said: *one cannot believe impossible things.*

I dare to say that you have not had much practice, said the queen,

When I was your age, I always did it for half-an-hour a day and sometimes I have believed as many as six impossible things before breakfast

Lewis Carroll: *Through the Looking-glass*, 1872