

ISA 666

Internet Security Protocols

Lecture #12

Information Hiding, Steganography, Watermarking

ISA 666

By Dr. Xinyuan (Frank) Wang

1

Outline

- Introduction to Information Hiding
- Steganography
 - Definition and History
 - Applications
 - Principles
 - Techniques
- Watermarking
 - Definition and History
 - Applications
 - Principles
 - Attacks on Watermarking
- Information Theoretic Point of View
 - Modeling information hiding as a communication problem
 - Capacity of discrete channels
 - Choice, uncertainty and entropy

ISE at George Mason University

ISA 666 By Dr. Xinyuan (Frank) Wang

2

Information Hiding

- What is Information Hiding?
 - Not hiding the implementation details in OO ☺
 - It is about secretly embedding information in some cover (carrier) signal
- Multidisciplinary
 - Communication theory
 - Information theory and coding theory
 - Signal processing
 - Cryptography
 - Probability & statistics
 - Human perception

ISE at George Mason University

ISA 666 By Dr. Xinyuan (Frank) Wang

3

Information Hiding Branches

- Covert channels
- **Steganography** – about concealing the very **existence** of some message (information)
 - Linguistic steganography
 - Technical steganography
- Anonymity
- **Watermarking** – embed some (visible or invisible) information about the carrier
 - Robust watermarking
 - Visible
 - Invisible
 - Fragile watermarking

ISE at George Mason University

ISA 666 By Dr. Xinyuan (Frank) Wang

4

Why Want Information Hiding?

- Unobstrusive communication – make it hard to be detected or jammed by enemy
 - Military
 - Intelligence agencies
 - Criminals
 - USA today reported that Osama Bin Laden used steganography to communicate with operative
- Copyright protection
 - USA Today, Jan. 2000: Estimated lost revenue from digital audio piracy US \$8,500,000,000.00
- Ownership proof
 - Traitor tracing
- Anonymous communication
 - Private voting
 - Privacy

ISE at George Mason University

ISA 666 By Dr. Xinyuan (Frank) Wang

5

Information Hiding Requirements

- Imperceptibility
 - The cover signal should be perceptually identical before & after information embedding
- Robustness
 - The embedded information should survive normal (natural) signal distortion
- Security
 - Ability to resist intentional tempering
 - Hard for the adversary to detect the existence of the hidden information
 - Hard to recover embedded information without proper key
 - Hard to forge hidden information without proper key
 - Hard to remove hidden information without proper key
 - Hard to corrupt the hidden information without proper key
- Capacity
 - Maximize the rate of embedded information

ISE at George Mason University

ISA 666 By Dr. Xinyuan (Frank) Wang

6

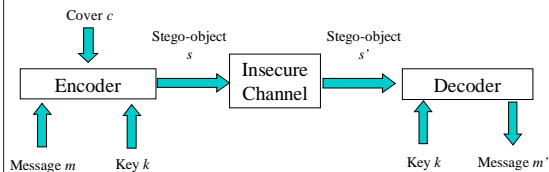
Steganography Definition and History

- **Steganography**
 - Derived from Greek words **steganos** which means “covered” and **graphia** which means “writing”
 - **Covered writing**
- **Different from Cryptography, which means “secret writing”**
 - Cryptography is about concealing the content of messages
 - **Steganography** is about concealing the very existence of the message
- **Steganography has a long history**
 - Writing on shaved head
 - Writing underneath the wax of a writing tablet
 - Invisible ink
 - Microscopic images
 - ...

Simmons' Prisoners' Problem

- Simmons formulated steganography as the “Prisoners' Problem” in Crypto 1983
- Alice and Bob are in jail and wish to hatch an escape plan
- All the communication between Alice and Bob pass through the warden Willie
 - Willie will throw Alice & Bob into solitary confinement if he detects any encrypted message!
 - Willie can be passive – only passively monitoring the passing message and either pass or stop it
 - Willie can be active – may change the content of the passing message and forward it
- Alice and Bob want to hide their message in an innocuous looking cocontext

Steganography Model



- How much hidden information can be transferred?
- What's the error probability given certain distortion?

Steganography Techniques: Substitution

- Substitute redundant parts of a cover with message m
- **Least Significant Bit (LSB) substitution**
 - Choose a subset of cover elements and substitute least significant bits of each element with message bits
 - Message could be encrypted or compressed before hiding
 - May use pseudorandom number generator to randomly spread the message over the cover
 - Is it robust against random perturbation over LSBs?
 - Assuming the adversary can change the LSBs of all elements in cover
 - Adversary does NOT know which elements have been used by the information hider

Steganography Techniques: Domain Transformation

- Embed the message m in a transformed domain (i.e. frequency domain) of the cover
- **Embedding in the Discrete Cosine Transform (DCT) domain**
 - Split the cover image into 8×8 blocks. Each block is used to encode one bit
 - Blocks are chosen in a pseudorandom manner
 - The relative size of two pre-defined DCT coefficients is modulated using the message bit
 - The two coefficients are chosen from middle frequencies
 - Why?

Steganography Techniques: Spread Spectrum

- Use ideas from spread spectrum communication where a signal is transmitted in a bandwidth much larger than the minimum necessary to send the information
- The message embedding is spread over a wide frequency spectrum
- The SNR in every frequency band is small – difficult to detect
- Even if parts of the message are removed from several bands, there is still enough information in other bands to recover the message
- It is difficult to remove the message completely without destroying the cover – robustness ☺
 - Other attacks?

Steganography Techniques: Statistical

- Embed information by signal distortion
- The encoder applies a sequence of modifications to the cover. The sequence corresponds to the message.
- The decoder measures the differences between the original cover and the distorted cover to detect the sequence of modifications and recover the message accordingly
- Limitations
 - Decoder need the original (undistorted) cover to decode

Steganography Techniques: Distortion

- Encode information by changing some statistical properties of the cover
- The cover is split into blocks, and each block is used to embed 1 bit
- If the message bit is "1", then the cover block is modified, otherwise the cover block is not changed
- Challenges
 - How to tell whether a cover block is modified or not?

Steganography Techniques: Cover Generation

- Encode information by ways a cover is generated
- Automated generation of English text
 - Use a large dictionary of words categorized by different types, and a style source which describes how words of different types can be used to form a meaningful sentence
 - Transform message bits into sentences by selecting words out of the dictionary which conforms to a sentence structure in the style source.
 - It is based on the redundancy of natural language
 - What is the redundancy in English in term of percentile?

Steganalysis

- What is steganalysis?
 - The art of detecting and decoding messages that have been hidden steganographically
 - Methods of countering steganography.
- Two goals of steganalysis:
 - Discover the presence of the secret message embedded in cover work.
 - Corrupt the hidden message

Watermarking Definition

- What is watermarking?
 - The practice of imperceptibly altering a cover to embed a message about that cover
- Closely related to but different from steganography
 - In watermarking, the message is about the cover
 - In steganography, the message is arbitrary between the sender and receiver – a point-to-point covert channel
 - Steganography emphasizes more on stealthiness than robustness
 - Watermarking is often used where people know the existence of the hidden information and have a interest in removing it
 - Copyright information
 - Watermarking emphasizes more on robustness and resilience than stealthiness
- Watermarks should be inseparable from the cover in which it is embedded – unlike cryptography, watermark protects content even after it is decoded

Watermarking History

- More than 700 years ago
 - Watermarks were used in Italy to indicate the paper brand and the mill that produced it
- By the 18th century
 - Watermarks began to be used as anti-counterfeiting measures on money and other documents
- The term watermark was introduced near the end of 18th century
 - Probably because the marks resemble the effects of water on paper
- The first example of a technology similar to digital watermarking is a patent filed in 1954 by Emil Hembrooke for identifying music works
- Digital watermarking becomes an active research area since 1995

Principle of Digital Watermarking

- Embedded hidden information, which travels with the watermarked data, even after copying and redistribution.
- How can information be hidden in digital data? By exploiting “perceptual headroom”
 - Human perception is imperfect
 - Make modification to the original data without changing its perceptual quality, exploit masking principle (JND).
 - Modifications can be detected via signal processing.
- What happens if there exist a perfect compression algorithm?

Digital Watermarking Categories

- Robust watermark
 - Used for copyright protection.
 - Requires the watermark be permanently intact to the watermarked signal, removing the watermark result in destroying the perceptual quality of the signal.
- Fragile watermark
 - Used for tamper detection or somewhat digital signature.
 - Requires the watermark to be very easy to break under any modification of the watermarked signal.
- Semi Fragile watermark
 - Used for data authentication.
 - Requires the watermark to be robust to some benign modifications, but easy to break by other attacks.
 - Provide information about the location and nature of attack

Types of Digital Watermarking

- **Non-blind**
 - Need the original cover to detect watermarks
- **Semi-blind**
 - Does not use the original cover but use some side information and/or the original watermark.
- **Blind**
 - Does not use the original cover or any side information (most challenging).

Digital Watermarking Applications

- Copyright protection
 - Most prominent application
 - Embed information about the owner to prevent others from claiming copyright
 - Requires very high level of robustness
- Copy protection
 - Embed watermark to disallow unauthorized copying of the cover
 - For example, a compliant DVD player will not palyback or copy data that carry a “copy never” watermark
- Content authentication
 - Embed a watermark to detect modifications to the cover
 - The watermark needs to be “fragile”

Digital Watermarking Applications (cont'd)

- Transaction tracking
 - Embed a watermark to convey information about the legal recipient of the cover
 - Useful to illegally produced copies of the cover
 - Often referred to as “fingerprint”
- Broadcast monitoring
 - Embed watermark in the cover and use automatic monitoring to verify whether cover was broadcasted as agreed

Limitations of digital watermarking

- Digital watermarking **does not** prevent copying or distribution.
- Digital watermarking **alone is not** a complete solution for access/copy control or copyright protection.
- Digital watermarks **cannot** survive every possible attack.

Watermarking Techniques

- **Spatial domain watermarking**- Watermark embedded by directly modifying the pixel values. Usually use spread spectrum approach.
- **Transform domain watermarking**- Watermark embedded in the transform domain e.g., DCT, DFT, wavelet by modifying the coefficients of global or block transform.

Spread Spectrum Watermarking

- **Spatial domain technique**
- Use spread spectrum to spread the watermark all over the host image.

$$S(x, y) = \phi(x, y) \alpha b_i + I(x, y)$$

Where

b_i is the watermarking bit,

α is a scaling factor,

$\phi(x, y)$ is two-dimensional pseudo-random sequence of "1" or "-1",

$I(x, y)$ is the original image

Spatial watermarking example



Original image

Watermarked image

NEC Watermarking Scheme

- We take an example of Cox's Spread Spectrum watermarking in this attack.
- For example, in DCT transformed domain, selecting 1,000 largest AC coefficients v_i ($i=1,2,\dots,1000$), embedding the watermark $S=\{s_1, s_2, \dots, s_{1000}\}$ as follows

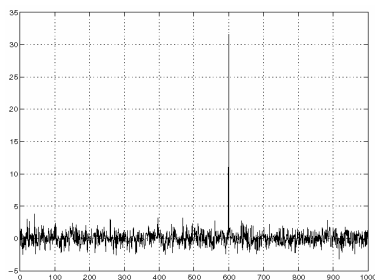
$$v'_i = v_i (1 + \alpha s_i)$$

$$v_i = v'_i / (1 + \alpha s_i)$$

- Detection function

$$\text{Sim}(S, S^*) = \frac{S^* \cdot S}{\sqrt{(S^* \cdot S^*)}}$$

Watermark Detection



DCT Phase Modulation

- **Embedding algorithm**
 - Randomly select a group of low frequency DCT coefficients using a key.
 - Generate a binary message as a watermark.
 - Set the phase of the selected coefficients in accordance with the embedded watermark.
- **Decoding algorithm**
 - Use the same key to select the coefficient.
 - Extract the sign of the selected coefficients and decode according to the embedding rule.

Watermark Attacks

- **Robustness attacks:** Intended to remove the watermark
 - Lossy compression (JPEG, MPEG), filtering (high/low pass, linear/non-linear), noise removal, cropping, histogram equalization, additive noise, A/D & D/A conversion (print-scan)
- **Presentation attacks:** To cause watermark detection failure.
 - Geometric transformation, rotation, scaling, translation, change aspect ratio, line/frame dropping, affine transformation, mosaic, synchronization etc.
- **Counterfeiting attacks:** Render the original image useless
 - generate fake original, dead lock problem.

Synchronization Attack

- By disturbing the synchronization, the adversary can mask the watermark signal.
- Simple examples include delay and time scaling for audio and video, and rotation, scaling, and translation for images and video.
- More complex distortions include sample removal in audio, and shearing, horizontal reflection, and column or line removal in images.
- Even more complicated distortions include nonlinear warping of images,....

Geometric Attack

- Is a kind of synchronization attack that tries to fool the detector
 - stirMark distort (stretch) image
- **Resynchronization methods**
 - **Searching techniques**
 - Computationally expensive, probability of false alarm
 - **Use of templates (embed a pattern of peaks)**
 - Can be located by attacker, can be removed
 - Finding pattern is like finding a watermark
 - **Invariant representation (DFT, Fourier-Mellin)**
 - Not robust to other attacks

An Information Theoretic View

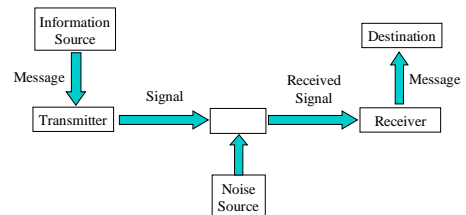
- Model information hiding as a communication with sided information
- Model countermeasure as a kind of noise
- Fundamental questions
 - Does there exist undetectable information hiding?
 - Does there exist an universal detection scheme that can detect all hidden information?
 - How much information can be embedded without being detected?
 - How much information can survive the removal attack?
 - What's the tradeoffs between embedding strength, attack strength, achievable embedding rate

Introduction to Information Theory

- C E. Shannon published the ground-breaking paper in 1948
 - *A mathematical theory of communication*
- Later preprinted as book in 1949
 - *The mathematical theory of communication*
- Introduced a number of key concepts and theorems that have profound impact
 - Channel capacity
 - Entropy, equivocation
 - Coding theorems in discrete/continuous noiseless/noise channel
- Arguably the greatest engineering paper ever written

The Fundamental Problem of Communication

- Is that of reproducing at one point either exactly or approximately a message selected at one point. – Shannon



Shannon's General Communication Schematic Diagram

Capacity of Discrete Channel

- The capacity of a discrete channel is the maximum possible error free transmission rate (in term of symbol/sec or bit/sec)
 - It is determined by the characteristics of the channel itself

$$C = \lim_{T \rightarrow \infty} \frac{\log N(T)}{T}$$

where $N(T)$ is the number of allowed signal of duration T

Choice, Uncertainty

- The discrete information source is modeled as a Markoff process
 - Can we define a quantity that can measure how much information is "produced" by such a process? Or better, at what rate information is produced?
- How to measure information?
- Suppose we have a set of $n > 0$ possible events whose probabilities of occurrence are p_1, p_2, \dots, p_n .
 - How much choices is involved in the selection of the event?
 - How uncertain we are of the outcome?

Measure of Choice, Uncertainty

- Suppose there is such a measure $H(p_1, p_2, \dots, p_n)$, it is reasonable to require of it the following properties
 - H should be continuous in the p_i
 - If all $p_i = 1/n$, then H should be a monotonic increasing function of n . The more events, the more choices
 - If a choice be broken down into two successive choices, the original H should be weighted sum of the individual values of H .

$$H(1/2, 1/3, 1/6) = H(1/2, 1/2) + 1/2H(2/3, 1/3)$$

- The only H satisfying the three above requirements is of the form

$$H = -K \sum_{i=1}^n p_i \log p_i \text{ where } K > 0$$

Entropy

- Shannon defines the measure of uncertainty of a set of probabilities p_1, p_2, \dots, p_n as **entropy**

$$H = -\sum_{i=1}^n p_i \log_2 p_i$$

- Shannon "invented" "bit" as the unit of measurement of information
- Entropy is the logarithm of the "average" number of choices according to the probability of each choice
 - No choice means entropy is 0
- Entropy is the **minimum descriptive complexity** of a random variable

Conditional Entropy

- What happens to the entropy of some random variable after we know something?
 - The uncertainty or choices could be less given the knowledge of something else
- Shannon defines the **conditional entropy** as the average of the entropy of y for each value of x , weighted according to the probability of getting that particular x

$$H_x(y) = -\sum_{i,j} p(i,j) \log_2 p_i(j) \text{ where } p_i(j) = \frac{p(i,j)}{\sum_j p(i,j)}$$

- Conditional entropy measures how uncertain we are of y on the average when we know x

Properties of Entropy

- $H=0$ iff all the p_i but one are zero, the non-zero one has probability 1
- For a given n , H is a maximum and equal to $\log n$ when all the $p_i = 1/n$.
 - This is intuitively the most uncertain situation
- For two events x and y , $H(x, y) \leq H(x) + H(y)$
 - The uncertainty of a joint event is less than or equal to the sum of the individual uncertainties
- Any change toward equalization of the probabilities p_1, p_2, \dots, p_n increases H
- $H(x, y) = H(x) + H_x(y)$
 - The uncertainty of a joint event is the uncertainty of x plus the uncertainty of y when x is known
- $H(y) \geq H_x(y)$
 - The uncertainty of y is never increased by knowledge of x

The Fundamental Theorem for a Discrete Noiseless Channel

- Let a source have entropy H (bits per symbol) and a channel have a capacity C (bits per second).
 - Then it is possible to encode the output of the source in such a way as to transmit at the average rate $CH - \epsilon$ symbols per second over the channel where ϵ is arbitrarily small.
 - It is not possible to transmit at an average rate greater than C/H
- H determines the channel capacity required with most efficient coding
 - Justify using H as the rate of generating information
 - H is the minimum number of bits to determine the uncertainty or choices

Example of Source Entropy

- Suppose a source produces a sequences of letters chosen from among A, B, C, D with probabilities 1/2, 1/4, 1/8, 1/8, what is the entropy of each symbol that encodes the 4 letters
 - $H = -[(1/2)\log(1/2) + (1/4)\log(1/4) + (2/8)\log(1/8)] = 7/4$ bits
- Naive coding would code
 - A as 00, B as 01, C as 10, D as 11, which would require 2 bits
- Clever coding would code
 - A as 0, B as 10, C as 110, D as 111
 - which would require $1/2 + 1/4 \times 2 + 2/8 \times 3 = 7/4$ bits per symbol

Impact of Noisy Channel

- Now we consider the case that the channel has noise that may change the signal probabilistically
 - No longer possible to transmit a signal completely error-free
 - Want to find way of transmitting that is optimal in combating the noise
- Suppose we are transmitting 1000 bits per second through a noisy channel, and the noise would cause, on the average, 1 out of 100 bits flipped (0 as 1 or 1 as 0). What is the rate of transmission of information?
 - Intuitively might think the rate is 990 bits
 - But how do you know which bits are flipped by noise?
 - If the noise is so big that the received signal is completely independent from the original signal sent
 - Statistically we still might have 500 bits sent correctly
 - But there is no useful information transmitted at all

Equivocation in Noisy Channel

- What matters here is the uncertainty about what has actually been sent when we have received a signal
 - This is the information that is missing from the received signal
 - We need this information to correct the error of received signal
- It is the conditional entropy $H_y(x)$, where x is the signal sent and y is the received signal
 - Shannon named it as **equivocation**
- The rate of transmission of information on a noisy channel is

$$R = H(x) - H_y(x)$$
- In previous example
 - $H_y(x) = -[0.99\log 0.99 + 0.01\log 0.01] = 0.081$ bits/symbol
 - Effective (error-free) transmission rate is $1000 - 81 = 919$ bits/second

Capacity in Noisy Channel

- The capacity of a noisy channel should be the maximum possible rate of transmission

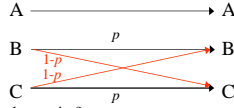
$$C = \text{Max} [H(x) - H_y(x)]$$
- Apparently we want the equivocation $H_y(x)$ as small as possible
 - Intuitively we know that increasing redundancy of transmitted signal will decrease the probability of error (or the equivocation)
 - But increasing redundancy would decrease the effective information transmission rate
 - One would think that in order to make the error probability approach to 0, we have to increase the redundancy infinitely
- Whether we could achieve asymptotically error free transmission rate C ?

The Fundamental Theorem for A Discrete Noisy Channel

- Shannon surprised people by proving that it is indeed possible to achieve asymptotic error free information transmission rate at the channel capacity C of a noisy channel
- Suppose a discrete channel has a capacity C and a discrete source has entropy H .
 - If $H \leq C$, there exists a coding system such that the output of the source can be transmitted over the channel with an arbitrarily small frequency of errors.
 - If $H > C$,
 - It is possible to encode the source so that the equivocation is less than $H - C + \epsilon$ (where ϵ is arbitrarily small).
 - It is not impossible to achieve an equivocation less than $H - C$
- This theorem gives a precise and startlingly simple description of the utmost dependability one can ever achieve in a noisy channel

Example of A Discrete Noisy Channel

- Consider a discrete channel sending 3 symbols: A, B, C, and symbol A is never affected by noise, and B & C have probability p of being transmitted correctly, and have probability $1-p$ of being changed to C and B respectively



- What is the channel capacity?
 - $H(x) = -p \log p - 2(1-p) \log(1-p)$
 - $H_y(x) = 2(1-p)[-p \log p + (1-p) \log(1-p)]$
 - $C = \log \frac{\beta + 2}{\beta}$ where $\beta = 2^{-[p \log p + (1-p) \log(1-p)]}$
 - If $p = 1$, $\beta = 1$ and $C = \log 3$. Noiseless channel
 - If $p = 1/2$, $\beta = 2$ and $C = \log 2$. B & C are indistinguishable and considered as one symbol

Information Theoretic View of Information Hiding

- Fundamental questions
 - What is the ultimate information hiding rate given the statistics of distortion?
 - What is the practical error-correction coding scheme that can achieve the hiding capacity?
 - Is it possible to have non-zero information hiding capacity if the distortion of embedding is less than the distortion by adversary?