Steganography Models
CSM25 Secure Information Hiding

Dr Hans Georg Schaathun
University of Surrey

Spring 2007

The Steganography System

What steganography is

The data hiding system

Watermarking System

Embedding → Extractor

- What would Kerckhoffs think of this stego-system?
- Security depends on the confidentiality of the algorithm.

Watermarking vs. Steganography

- Watermarking: the cover-image is essential
  - Two receivers:
    - One observes the cover-image
    - One extracts the hidden message
  - Minimum distortion is important
- Steganography: What is the use of cover-image at receiver?
  - Bob wants the message
  - The image is a red herring

Two key differences

Watermarking vs. Steganography

- Cover-image
  - Important in watermarking
  - Meaningless in steganography
- Attacker
  - Steganography: determine whether secret information exists or not
  - Watermarking: various other goals
    - Change cover-text
    - Remove watermark
    - Change watermark
### Real Steganography

#### Mathematical definition

A **Secret-Key stego-system** is \( S = (C, M, K, E, D) \) where

- \( C \): set of cover texts
- \( M \): set of messages
- \( K \): key space (set of possible keys)
- \( E \) is an encoding function, \( E : K \times M \rightarrow C \)
- \( D \) is a decoding function, \( D : K \times C \rightarrow M \)

such that

- \( \Pr(D(k, E(k, m)) = m) \approx 1 \).

---

### Compression

- \( \mathbb{F}^* \) is set of binary strings of arbitrary length

**Definition**

A compression system is a function \( c : \mathbb{F}^* \rightarrow \mathbb{F}^* \), such that \( E(\text{length}\, \tilde{m}) > E(\text{length}(c(\tilde{m}))) \) when \( \tilde{m} \) is drawn from \( \mathbb{F}^* \).

- The compressed string is expected to be shorter than the original.

**Definition**

A compression \( c \) is **perfect** if all target strings are used, i.e. if for any \( \tilde{m} \in \mathbb{F}^* \), \( c^{-1}(\tilde{m}) \) is a sensible file (cover-text).

- Decompress a random string, and it makes sense!

---

### Steganography by Perfect Compression

**Anderson and Petitcolas 1998**

- A perfect compression scheme.
- A secure cipher.

\[ \begin{array}{c}
\text{Encryption} \quad \text{Key} \quad \text{Decrypt} \\
\downarrow \quad \downarrow \quad \downarrow \\
C \quad \quad \quad \quad \quad \quad C \\
\downarrow \quad \downarrow \\
\text{Decompress} \quad S \quad \text{Compress} \\
\end{array} \]

- Steganography without data hiding.
Pure Steganography
Mathematical definition

A Pure stego-system is $S = (C, M, E, D)$ where
- $C$ : set of possible cover files, i.e. insuspicous files
- $M$ : set of possible (secret) messages
- $E$ is an encoding function, $E : C \times M \rightarrow C$
- $D$ is a decoding function, $D : C \rightarrow M$

such that
- $P_{C,M}(D(E(c, m)) = m) \approx 1.$

Secret-Key Steganography
Mathematical definition

A Secret-Key stego-system is $S = (C, M, K, E, D)$ where
- $C$ : set of cover texts
- $M$ : set of messages
- $K$ : key space (set of possible keys)
- $E$ is an encoding function, $E : K \times C \times M \rightarrow C$
- $D$ is a decoding function, $D : K \times C \rightarrow M$

such that
- $P_{C,M}(D(k, E(k, c, m)) = m) \approx 1.$

Distortion in Watermarking

- The distortion is the difference between an $N \times M$ cover image $X$ and marked image $Y$
- Often measured by PSNR

$$\text{PSNR} = 10 \log \left( \frac{(\max_{i,j} X)^2 \cdot N \cdot M}{||X - Y||^2} \right).$$

- High distortion $\iff$ low PSNR
- The higher the distortion, the less usable is $Y$ as replacement for $X.$
- In steganography, $X$ has no value, and $Y$ need not replace it.

To remember

- The cover-text is a red herring in steganography.
- The standard definitions of pure steganography and secret-key steganography apply to a very limited class of steganography based on data hiding.
- Cover-text irrelevant $\Rightarrow$ distortion irrelevant.
  - PSNR used to measure distortion in Watermarking.
Steganalysis: first steps

Classification by question asked

Traditional steganalysis:
- Does this file contain a hidden message?

Extended steganalysis:
- How long is the embedded message?
- What is the contents of the hidden message?
- Which system has been used for the message?
- What is the secret key used for the embedding?

Other attacks:
- Disable the message.

Classification by information available

- Stegogramme (always)
- System and algorithm (always by Kerckhoffs’ principle)
- Known message
- Known covertext
- Chosen message
- Chosen covertext
- Chosen stegogramme

Cryptoanalysis

- Assumptions
  - Known algorithm
- Aims to recover, either
  - Key, or
  - Message

- Scenario classes
  - Ciphertext only
  - Known plaintext
  - Chosen plaintext
  - Chosen ciphertext

Steganalysis classes

- Targeted steganalysis
  - Specialised to detect stegotext from a single system
- Uniclass steganalysis (fully blind)
  - Not related any particular system
- Multiclass steganalysis (quasi-blind)
  - Can identify a range of different systems
Exercise

Classify the $\chi^2$ steganalysis test, with which you have worked, in terms of the different classifications we have discussed.

- Question asked
- Information required
- Steganalysis classes
- Can you think of other classifications?

Blind and targeted

- Targeted steganalysis
  - Taylor-made for a specific stego-system
  - Extremely accurate
  - Completely inflexible
  - Important step in the evaluation of stego-systems
- Blind steganalysis
  - Intended to work against any stego-system
  - Can (potentially) identify the stego-system used
  - Rarely as accurate as targeted techniques

Blind steganalysis

- Blind steganalysis = classification problem
- Common approach is machine learning
  - Define heuristics
  - Train on objects from each class
  - Empirical data $\rightarrow$ choose threshold
  - Threshold used for decision in new cases
- Uniclass steganalysis
  - Cover-image or stegogramme
  - Train on cover-images
- Multiclass steganalysis
  - One class per known stego-system + cover-images
  - Train on cover-images and stegograms

Multiclass steganalysis

- Targeted attacks may be added for extra scrutiny
- Less blind than uniclass categorisation
- More complicated training process
Limitations

- **Sensitive to image type**
  - Cover-images are widely different
  - Photos vs. drawing
  - Scanning vs. digital camera
  - Landscape vs. portrait
  - Computer graphics and cartoons
  - Unknown type of cover-images could hit as stegogramme(!)

- **Targeted blind steganalysis**
  - Uses side information
  - e.g. we know the camera which took the cover-images