(1) A New Approach of Reversible Acoustic Steganography for Tampering Detection
(2) Capacity Adaptive Synchronized Acoustic Steganography Scheme

Xuping Huang, huang-xp@nii.ac.jp
School of Multidisciplinary Science
The Graduate University of Advanced Studies [SOKENDAI],
National Institute of Informatics (NII), Japan
What is steganography?
- A technology to hide and transmit data through apparently innocuous carriers in an effort to conceal the existence of secret data.
Introduction of Steganography:
Usage of Steganography(1)

(1) **Authenticate use** of steganography: verify integrity and reliability of data (Tampering Detection) by hiding hash value
Introduction of Steganography: Usage of Steganography(2)

**Problem:**
- Personal security and privacy risks are increasing: Including defamation, network tracing and stalking, eavesdropping, malicious tampering, Gumblar, copyright abuse and blackmail, etc.

**Solution:**
- Embed secret speech data into another innocent acoustic data (cover data) in an imperceptible way to hide and transmit the secret data between trusted users;
Importance of Information Security
- Vast amounts of privacy-sensitive digital data are being collected today by organizations for a variety of reasons and highly available.
  - Personal information, identification, and even healthcare information, etc.
  - Mobile phone recordings, information recording (personal video records)
- However, criminals, spies and predators are exploiting information illegally
  - malicious modification, eavesdropping, etc

Security and privacy risks have increased
- 6,690 cyber crimes arrested in 2009, which has a creasing about 5.8%
  (369 cases) than crimes in 2008

[Research Scope and Background image]

According to document from National Police Agency on March 4th, 2010
A New Approach of Reversible Acoustic Steganography for Tampering Detection

Xuping Huang, huang-xp@nii.ac.jp

School of Multidisciplinary Science

The Graduate University of Advanced Studies [SOKENDAI], National Institute of Informatics, Japan
Outline of Contents

1. Research Scope and Background
2. Conventional works and Problems
3. Acoustic Steganography to Authenticate Use
4. Implementation
5. Experiment and Evaluation
6. Summary and Future Work
Social needs to protect data from tampering using Steganography

- **Objective**: maintaining integrity of content with probative importance

  - An increasing usage of digital recordings which are of probative importance (stored as proves):
    - **Audio**: interview, investigator tape, wills and testament tape (probate court), telephone recording, phone banking, emergency calls, air traffic communication, etc.
    - **Video**: documentary, operation video
    - **Image**: military image, medical image (X-ray, etc.)

- Focus on Audio data
  - Acoustic media is a necessary media and widely used for people’s communication and for data’s recording
  - Protection of audio data from tampering is highly required though not yet applied by conventional research
Function properties

- Feature value is embedded into cover data after spectrum transform, expansion and optimization.
- Function Properties and requirement:

Verifiability
- Possible to verify the integrity and reliability
- Embed data is related to cover data
  Feature value or data which can not be replaced easily

Reversibility
- Reversibility requires both the embedding and extracting algorithms to be lossless to make sure cover data blindly reversible

Imperceptibility
- When the feature data are camouflaged in the cover data, distortion needs to be kept below a perceptible level;
  Human Auditory System adaptive gain optimization
- Stego-audible without additional conversion
Outline of Contents

1. Research Scope and Background
2. Conventional works and Problems
3. Acoustic Steganography to Authenticate Use
4. Implementation
5. Experiment and Evaluation
6. Summary and Future Work
Conventional works

- Algorithm classification of conventional works
  - (1) Data compression based method[14][15]:
    - The original portions of cover data will be replaced by payload using compression and data alteration
    - Usage of “Unequal Bit Sensitivity”: a few bits can be substituted without introducing perceptual distortion.
    - Encode according to G.729 (applied to speech and VoIP system)
    - These works have proved redundancy positions were available, however, not applied to verification use of acoustic data
  - (2) Histogram bin shifting method[17]:
    - The embedding target is replaced by the histogram of a block.
    - A modified side match vector quantization (SMVQ) technique is proposed to confirm the reversibility of the proposed method
    - **Neglect the distortion** of stego-image since cover data is available due to the reversibility of the proposed method
(3) Method based on modifying difference expansion[16]:
- These schemes represent features of original data with small values
- Then the value is expanded to embed payload in LSB.
- Use a DE technique, which discovers extra storage space by exploring the redundancy in the image content
- Algorithm
  - Step1: calculating the difference values
  - Step2: partitioning difference
  - Step3: values into four sets
  - Step4: creating a location map
  - Step5: collecting original LSB values
  - Step6: data embedding by replacement
  - Step 7: an inverse integer transform
- It takes advantages of Human Visual System and the algorithm cannot be applied to acoustic system directly.
Conventional works

- **Philosophical Properties of Audio Steganography**
  - Verifiability: verify the malicious tampering on stego data
  - Reversibility: cover data can be available without any loss after computing and without cover data
  - Imperceptibility: distortion of stego must be controlled under perceptible level

- **Algorithm classification of conventional methods and their properties for integrity verification**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Verifiability</th>
<th>Reversibility</th>
<th>Imperceptibility</th>
<th>Acoustic Applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data compression based method[14][15]</td>
<td>YES</td>
<td>NO</td>
<td>No</td>
<td>YES</td>
</tr>
<tr>
<td>Histogram bin shifting method</td>
<td>YES</td>
<td>YES</td>
<td>Limited</td>
<td>YES</td>
</tr>
<tr>
<td>Method based on modifying difference expansion</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>
Target application: Scenarios and Social needs

- **recorder**
- **Original tape (deny charges)**
- **Cover (original)**
- **Embedding**
- **Stego**
- **Stego’**
- **Verification**
- **Stego**
- **Stego’**
- **Extraction**
- **Cover**

Illegal modification (confess to charges)

- With tampering
- Without tampering

Chief judge (verify videos)

ReQUIRES original data

A part of content has been replaced!! Not reliable!

According to the lossless algorithm, cover data is reversible if without attack

For example, investigator tapes, wills and testament tapes (probate court), telephone recordings

JFLI Workshop 2010
Outline of Contents

1. Research Scope and Background
2. Conventional works and Problems
3. Acoustic Steganography to Authenticate Use
4. Implementation
5. Experiment and Evaluation
6. Summary and Future Work
**Step 1:** Divide cover data into size-fixed frames:

- Extract feature value of each frame using SHA-1 algorithm (128 bits);

**Step 2:** Extract feature value of each frame using SHA-1 algorithm (128 bits);
**Embedding**

**Step3:** Transform signal of each frame from time domain to spectrum domain.

**Step4:** Divide low-frequency region and high-frequency region (reserved for embedding).

### Audible: 20Hz~20kHz
### Difficult: 18kHz~20kHz

### Sample:
44.1kHz-16bit-mono with 1323008 samples

<table>
<thead>
<tr>
<th>DCT size</th>
<th>Samples/frame</th>
<th>High-frequency</th>
<th>Low-frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>5168</td>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td>512</td>
<td>2584</td>
<td>256</td>
<td>256</td>
</tr>
<tr>
<td>1024</td>
<td>1292</td>
<td>512</td>
<td>512</td>
</tr>
<tr>
<td>2048</td>
<td>646</td>
<td>1024</td>
<td>1024</td>
</tr>
</tbody>
</table>
Step5: Amplitude extension for hiding and initialize the positional flag $\phi$

Amplitude expansion $20 \log_{10}(2) = 6$ dB

$N = \text{DCT size}, M = \text{length}(\phi)$

Length of $\phi = 16$;

Length of hash = 128;

Length for optimization = $N/2 - 128 - 16$;
Embedding

Countermeasure 1:
i) Optimize signal in the segment (-3dB)

ii) set φ component from 0 to 1

iii) Write optimized data into region for optimization

iv) If overflow happens again, repeat i) and ii) until embedding positions are not enough for hash

Countermeasure 2:
(without optimization)
i) Shift 1 bit right of expansion start index i; index’ = index + 1;

ii) Embed hash from index’ in spectrum domain

Length of segment = N/M

N = DCT size, M = length(φ)
Length of φ = 16
Length of hash = 128
Length for optimization = N/2 - 128 - 16;

Amp. (dB)
Amplitude expansion
20log_{10}(2) = 6dB

JFLI Workshop 2010
### Embedding: payload:

Payload embed region (in case high-frequency rate=1/2):

<table>
<thead>
<tr>
<th>DCT Size</th>
<th>Hash Region(No.)</th>
<th>Positional data’s region (length)</th>
</tr>
</thead>
<tbody>
<tr>
<td>512</td>
<td>256~495</td>
<td>496~512(16)</td>
</tr>
<tr>
<td>1024</td>
<td>513~1008</td>
<td>1009~1024(16)</td>
</tr>
<tr>
<td>2048</td>
<td>1025~2032</td>
<td>2033~2048 (16)</td>
</tr>
</tbody>
</table>

Payload calculation:

\[ \text{Payload} = (\text{embedLength} - \text{length(φ)}) \times \text{numberOfFlames} \]

Payload examples using 1323008 samples: (if no overflow)

<table>
<thead>
<tr>
<th>DCT Size</th>
<th>Embed Length</th>
<th>Length of φ</th>
<th>Number of flames</th>
<th>Payload</th>
</tr>
</thead>
<tbody>
<tr>
<td>512</td>
<td>256</td>
<td>16</td>
<td>2584</td>
<td>620160</td>
</tr>
<tr>
<td>1024</td>
<td>512</td>
<td>16</td>
<td>1292</td>
<td>640832</td>
</tr>
<tr>
<td>2048</td>
<td>1024</td>
<td>16</td>
<td>646</td>
<td>651168</td>
</tr>
</tbody>
</table>
**Step 1:** Divide cover data into size-fixed frames.

**Step 2:** Transform signal of each frame from time domain to spectrum domain.

**Step 3:** Read and analyze $\phi$ and extract and reconstruct hash information when $\phi=1$ to get hash'’;
2 extraction and verification

**Step 4:** Reconstruct expensed original data by
\[ \text{signal} = \text{signal} \times \sqrt{2} \] from expansion index;

**Step 5:** Reconstruct original data in low-frequency domain;

**Step 6:** Verification

\[ \text{hash( reconstructed cover data)} = \text{hash’} \]

compare hash’ with reconstructed hash”

if (hash’ != hash”)

verified=true;

Reconstruct original data in low-frequency domain.
Outline of Contents

1. Research Scope and Background
2. Conventional works and Problems
3. Acoustic Steganography to Authenticate Use
4. Implementation
5. Experiment and Evaluation
6. Summary and Future Work
Signal representations

Let $H(i)$ be the corresponding transform matrix:

$$H(i) = \sum_{t=0}^{N-1} h(t) \cos \frac{\pi (2t + 1)(2i + 1)}{4N}, \quad i = 0, 1, \ldots, N - 1$$

Let $C_{N}^{IV}$ be the corresponding transform matrix:

$$C_{N}^{IV} = \left( \cos \frac{\pi (2t + 1)(2i + 1)}{4N} \right), \quad i, t = 0, 1, \ldots, N - 1$$

**Integer DCT type-IV and inverse Integer DCT type-IV are used to transform data from time domain to frequency (spectrum) domain**

**Algorithm complexity:**

The total number of rounding operations is $2.5N$, where $N$ is DCT size (the length of a block).

$t$ is discrete time; $0 < t < N$.

$i$ is discrete frequency; $0 < i < N$. 

2010/10/15
Fig 1. Structure of the embedding phase

Fig 2. Extraction and verification phases

2010/10/15
Experiment Data Set:
- the RWC Music Database, 100 tracks.
- with an L-channel waveform with 44.1 kHz sampling and 16-bit quantization.
- The samples were cut to the initial 30 seconds of the playback time.

Evaluation:
- We evaluated the distortion in acoustic quality on the basis of the ITU-R BS.1387 (PEAQ) standard.
- objective difference grade (ODG), ranged from 0 to -4
  0 imperceptible
  -1 perceptible but not annoying
  -2 slightly annoying
  -3 annoying
  -4 very annoying.

Acoustic data were re-sampled at 48 kHz before the evaluation.
ODG depends on auditory categories and DCT size N.

Samples: RWC Music Database[20] with 100 tracks, using L-channel waveform with 44.1 kHz sampling and 16-bit quantization, were cut to the 30 seconds of playback time.

Objective Difference Grades of 100 Tracks, $N=2048$ with an average of $-0.959$

<table>
<thead>
<tr>
<th>DCT size</th>
<th>Payload (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1024</td>
<td>640832</td>
</tr>
<tr>
<td>2048</td>
<td>651168</td>
</tr>
<tr>
<td>4096</td>
<td>656336</td>
</tr>
<tr>
<td>4096 (stereo)</td>
<td>1314704</td>
</tr>
</tbody>
</table>

Payload = $(N/2-16) \times (\text{Length(signal)}/N)$

PAYLOAD WITH DIFFERENT DCT SIZES
Sample: RWC-MDB-G2001 No.10 with a 30 second playback time
Outline of Contents

1. Research Scope and Background
2. Conventional works and Problems
3. Acoustic Steganography to Authenticate Use
4. Implementation
5. Experiment and Evaluation
6. Summary and Future Work
Summary

- **Contribution:**
  - We proposed a tampering detectable approach using Steganography.
  - This scheme is a reversible way of guaranteeing the integrity and reliability of cover data.
  - Distortion of stego data is controlled to be **imperceptible** and stego data is audible without the need to apply additional.

- **Method**
  - A **hash function** was used to extract the feature value of the original cover data, and this feature value was used as the payload for verifying whether tampering had occurred.
  - **Amplitude expansions** were concentrated in the high-frequency spectrum domain to make the hidden data imperceptible.
  - A scheme to **optimize amplitude** prevented the amplitude of the stego signal from overflow.
Future Work

- **Improve the current method**
  - 1) Improve robustness for steganalysis:
    - Problem: embedding was concentrated in the high-frequency component after it was transformed to the spectrum domain, it was conspicuous of the hiding due to the boundary line.
    - Solutions:
      - Gradually increase amplification spectrum
      - PN Key for shifting (distortion must be considered)
  - 2) Fix the distortion problem (Improve ODG value)
    - Solution: to use Modified DCT instead of intDCT type IV

- **A new approach using lossless acoustic compression with HAS threshold estimation**
  - To use Human Auditory System to estimate the threshold for hiding (imperceptibility)
  - Use a lossless compression algorithm to create payload capacity
    - LPAC, La, MokeyAudio (60%, little distortion, fast), TTA, TAK, FLAC, WavPack
Capacity Adaptive Synchronized Acoustic Steganography Scheme

Xuping Huang, huang-xp@nii.ac.jp

School of Multidisciplinary Science

The Graduate University of Advanced Studies [SOKENDAI], National Institute of Informatics, Japan
Outline of contents

1. Overview
2. Details of Proposal: Modeling and Algorithm
3. Experiment
4. Results and Analysis
5. Future work
Overview - Background

- Requirement towards malicious attack
  - Protection of privacy
- Steganography, which supplies a philosophical concealed communication, has been focused on
  - Bits scrambling in cover data for imperceptible change
  - Companied with a secret key
  - Conceal the existence of secret
- Principal design challenge
  - Secret data must be recoverable towards considerable level of distortion and tamper attacks (secret sharing)
  - Takes advantage of the perceptual properties of the human auditory system (HAS)
Acoustic media, is required for usage in concealed communication subliminal channel
- Live broadcasting, Mobile communication, TV conferences

Synchronized process is required, whereas security issues are enhanced via recording the secret data in a real-time process

- **Steganography**
  - Hide and transmit acoustic data through apparently innocuous acoustic carriers to conceal the existence of the secret audio data

- **Real-time**
  - Secret audio data is recorded when embedding scheme starts to run and is synchronized with cover sequences dynamically

- **Multi-bit**
  - Embedding capacity and positions can be specified arbitrarily (key-sharing between trusted users)
An illustrative Multimedia Acoustic Synchronized Steganography Scheme

1. Run socket server for a connection
2. Choose a innocuous audio carrier
3. Record speech secret data
4. Embedding works with a key shared with Alice in advance
5. Send stego data in subliminal channel (socket server, etc)

1. Run socket client
2. Receive stego data
3. Extract secret data with a key, while one of three data: secret, stego, and cover can be played
4. Files can be generated if necessary
Synchronized Acoustic Steganography

Capacity Adaptive Synchronized Acoustic Steganography Scheme

- **Data fidelity and integrity**
- **Human Auditory System**
- **Positions and capacity specified arbitrarily**
- **Security Enhancement due to a real-time process**

**Proposed method**
- Divide data into size-fixed frames
- Record secret data while cover is playing
- PCM setting (Frequency-size-channel)
- Embed secret sampling bit stream

**Details**
- **Cover:** 32kHz-16bit-2channel
  - **Embed:** 16kHz-8bit-1channel
- two secret bits are embedded into specified comparatively insignificant [1st, 8th] in cover freely
- Extraction with a key shared between sender and trusted receiver
  - Positions, Bit numbers
Illustration of Algorithm with bit position (2,4) is specified

$$c_{mask} = (2^{(loc_1-1)} \text{ XOR } 0xFF) \text{ AND } (2^{(loc_2-1)} \text{ XOR } 0xFF)$$
JFLI Workshop 2010

Page 40

Insignificant 8 bits in a sampling point of cover data

\[
\begin{array}{cccccccc}
1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\
\end{array}
\]

Embed data \( i=1 \)

\[
\begin{array}{cccccccc}
1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
\end{array}
\]

\( e_1 \text{ shifts 5 bits} \)

\[
\begin{array}{cccccccc}
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
\end{array}
\]

(cleared to be 0)

\[
\begin{array}{cccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]

(shifted to position \( loc1 \))

\[
\begin{array}{cccccccc}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\end{array}
\]

c1

\[
\begin{array}{cccccccc}
1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
\end{array}
\]

e1

\[
\begin{array}{cccccccc}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\end{array}
\]

e2

\[
\begin{array}{cccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

\( c_{mask} \text{ AND} \)

\[
\begin{array}{cccccccc}
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
\end{array}
\]

Insignificant 8 bits in a sampling point of stego data

\[
\begin{array}{cccccccc}
1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{array}
\]
Data information for experiment

First “1” in each sampling point
Experiments

- **Group1: Extreme situation**
  - Cover: quiet weighted white noise
    - Recorded in the lab at midnight
  - Secret: Japanese male’s speech
    - Without any sampling points where only 0 are filled

- **Group2: Validation due to embedding positions**
  - (1)1,2  (2)1,4  (3) 1,8  (4) 4,5  (5) 5,8  (6) 7,8
  - Using music data (pop & jazz)

- **Group3: Validation due to acoustic source**
  - Embedding positions are (7th, 8th)

※ Using WAV format, with sampling of 30 seconds
Evaluation-SNR solution

- Objective evaluation of embedding performance
- calculate the differences between stego data and cover data of each sampling point
- c: cover data ; s: stego data, d: correlation difference
- n: numbers of sampling points in a certain time terminal

\[
SNR = 10 \log_{10} \frac{\sum_n c^2}{\sum_n d^2(c, s)} [dB]
\]
Experimental Result (Group 1)

Embed

Loud

Cover

natural noise with small amplitude

Frequency: JPN-Speech-8-8-1ch

Stego

Imperceivable

Frequency: Natural_Noise-32-16-2ch

Frequency: Speech_2_Noise-32-16-2ch

It is recorded and synchronized with cover sequence.
Experimental Result (Group 2)

- SNR ranks [52.15, 88.28]
- Imperceptible even when embedding positions locate in 7th, and 8th bit
Embedding positions addressed to (7th, 8th) pair
SNR ranks [51.68, 55.18]
Acoustic Categories are not effective aspects

Data consistency
Embedding quality performs better if cover and secret structurally correlated with each other
Contribution:
- A real-time Steganography scheme has been proposed to protect the secret data’s security
- Socket communication is used to transmit stego data
- Distortion has been controlled according to embedding positions
- Secret data recording, embedding, transmitting and extraction are in a real-time process

Future work
- Robustness:
  ■ Sophisticated algorithm is required to enhance the robustness against detection
- Data capacity:
  ■ enlarge data amount below the audible threshold in each sampling point
- Evaluation solution:
  ■ correlation analysis of significant bits between secret and cover sample
References

References


References