Selection-Channel-Aware Rich Model for Steganalysis of Digital Images

Tomáš Denemark, Vahid Sedighi, Rémi Cogranne, Vojtěch Holub, and Jessica Fridrich
Steganography and steganalysis

- Steganography is the art of secret communication

\[ \text{message } m \xrightarrow{\text{cover } X} \text{Emb}(X, m, k) \xrightarrow{\text{stego } Y} \text{Ext}(Y, k) \]

- Steganographer’s job
  Modify a cover image to stego image so that it contains a secret message (by flipping LSBs, changing DCT coefficients, ...).
  **Goal**: make the embedding changes statistically undetectable.

- Warden’s job: Distinguish between cover and stego images by building a detector. If cover source is known, the best detection is achieved using feature-based steganalysis and machine learning.
Steganography in practice

- **Sender**
  Specifies the cost of changing each pixel in the cover, $\rho_{ij} \geq 0$. Embeds the message by minimizing the distortion in the form of a sum of costs of all changed pixels, $\sum_{x_{ij} \neq y_{ij}} \rho_{ij}$. Problem is equivalent to source coding with a fidelity constraint.

  Can be implemented with syndrome-trellis codes that operate near the rate–distortion bound [Filler 2010].

- **Recipient**
  Extracts the secret message using the parity-check matrix of the shared syndrome-trellis code.
Content-adaptive steganography

- Embedding prefers changing pixels in textured / noisy areas

cover

stego changes
Content-adaptive steganography

- Embedding prefers changing pixels in textured / noisy areas

![Cover Image](image1.png)

![Stego Changes](image2.png)
Selection channel

- Formally, the selection channel are the probabilities of changing pixel \( ij \):
  \[
  p_{ij} = \frac{e^{-\lambda \rho_{ij}}}{1 + e^{-\lambda \rho_{ij}}},
  \]

- \( \lambda \geq 0 \) parameter controlling the payload
  - \( \rho_{ij} \) pixel “costs” computed from cover image \( x \)
  - costs dictated by content + noise

- Since stego changes are subtle: \( \rho_{ij} \) from cover \( \approx \rho_{ij} \) from stego image
Selection channel recoverability, WOW

[Holub, IEEE WIFS 2012] Designing Steganographic Distortion Using Directional Filters
Selection channel recoverability, S-UNIWARD

[Holub, EURASIP 2014] Universal Distortion Function for Steganography in an Arbitrary Domain
Selection channel recoverability, HILL

Using Selection Channel for Steganalysis

- [BOSS, IH 2011] no successful attack on HUGO based on approximate knowledge of the selection channel.
- [Schöttle et al., WIFS 2012] improved WS detector for naive content-adaptive LSB replacement.
- [Denemark, SPIE 2014] first successful attack on modern stego scheme that utilized an artifact in selection channel.
- [Denemark, WIFS 2014] maxSRMd2 (this presentation)
Spatial Rich Model (SRM)

cover X
Spatial Rich Model (SRM)

- $z_{ij} = x_{i,j} - \text{Pred}(\mathcal{N}(x_{ij}))$
- $\text{Pred}(\mathcal{N}(x_{ij}))$ ... pixel predictor on neighborhood $\mathcal{N}$
- linear and min/max filters
- $z_{ij}$ has narrower dynamic range
- better SNR (stego noise to image content)
Spatial Rich Model (SRM)

- $z_{ij} \rightarrow r_{ij} = Q_2(z_{ij})$
- $\mathbb{Q} = \{-Tq, -(T-1)q, \ldots, Tq\}$
- $T$ ... truncation threshold
- $q$ ... quantization step
  (SRM uses $q = 1, 1.5, 2$)

quantized residual $r$
Spatial Rich Model (SRM)

- collect quartets of values
- horizontal and vertical directions
Spatial Rich Model (SRM)

- $X$ → $Z$ → $r$
- $+1$
- $f_{N-2}$ $f_{N-1}$ $f_{N}$ $f_{N+1}$ $f_{N+2}$
- 4D co-occurrence matrix
- Symmetrization
- Co-occurrence vector
Co-occurrences in maxSRMd2

- collect quartets of values
- horizontal and vertical directions
- twice as many symmetries

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>X</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>L</td>
<td>C</td>
<td>2</td>
</tr>
<tr>
<td>-3</td>
<td>0</td>
<td>2</td>
<td>-1</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>-1</td>
<td>-3</td>
<td>2</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>-3</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
</tr>
</tbody>
</table>

The diagram illustrates the co-occurrence matrix with additional details on the quartets of values and symmetries.
Co-occurrences in maxSRMd2

\[ +\max(P(L),P(C),P(E),P(R)) \]

- 4D co-occurrence matrix
- utilize embedding probabilities
- symmetrization

<table>
<thead>
<tr>
<th>( f_{N-2} )</th>
<th>( f_{N-1} )</th>
<th>( f_N )</th>
<th>( f_{N+1} )</th>
<th>( f_{N+2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>([-1, -1, 0, 2])</td>
<td>([-1, -1, 0, 2])</td>
<td>([-1, -1, 0, 3])</td>
<td>(\ldots)</td>
<td>(\ldots)</td>
</tr>
</tbody>
</table>

co-occurrence vector
Detection gain w.r.t. SRM (WOW)
Detection gain w.r.t. SRM (S-UNIWARD)

![Graph showing detection gain with respect to SRM (S-UNIWARD)]
Detection gain w.r.t. SRM (HILL)

![Graph showing detection gain vs payload for SRM and maxSRMd2 with different values of \( \hat{\alpha} \).]
Co-occurrences in thresholded SRM (tSRM)

-1 -2 -1 3 -1 0 -1
0 3 L C E R -3
-2 -1 2 0 3 1 1
0 -3 -1 -2 0 -1 -2
2 2 2 -1 -2 0 -3
1 0 0 2 0 -3 1
-1 1 1 3 1 1 -1

- collect quartets of values
- horizontal and vertical directions
Co-occurrences in thresholded SRM (tSRM)

\[ \begin{array}{c}
X \\
\rightarrow \\
z \\
\rightarrow \\
r \\
\rightarrow \\
\rho \\
\end{array} \]

if \( \rho(L) < T \), +1

<table>
<thead>
<tr>
<th>( f_{N-2} )</th>
<th>( f_{N-1} )</th>
<th>( f_N )</th>
<th>( f_{N+1} )</th>
<th>( f_{N+2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( [-1, -1, 0, 2] )</td>
<td>( [-1, -1, 0, 2] )</td>
<td>( [-1, -1, 0, 3] )</td>
<td>( \cdots )</td>
<td></td>
</tr>
</tbody>
</table>

- 4D co-occurrence matrix
- utilize only some values
- symmetrization

co-occurrence vector
Comparison between maxSRMd2 and tSRM (WOW)

Gain in $\bar{p}_E$

Payload (bpp)
Comparison between maxSRM and tSRM (S-UNIWARD)

![Graph comparing Gain in $P_E$ with Payload (bpp)]
Summary

- maxSRM is a general-purpose feature set capable of utilizing the selection channel for detection of content-adaptive steganography
- Overly content-adaptive embedding hurts security (WOW)
- When designing steganography, selection-channel attacks need to be considered
  - often, improvement w.r.t. SRM leads to bigger loss w.r.t. maxSRM
- Matlab code available from http:\dde.binghamton.edu\download