A Weighted Stego Image Detector for Sequential LSB Replacement

Andrew Ker
adk@comlab.ox.ac.uk

Royal Society University Research Fellow
Oxford University Computing Laboratory

Data Hiding for Information and Multimedia Security, Manchester
30 August 2007
A Weighted Stego Image Detector for Sequential LSB Replacement

Outline

• Spread and sequential LSB replacement
• The Weighted Stego Image (WS) method
• WS for sequential embedding
• Performance
**LSB Replacement**

- The **cover** is a stream of \( N \) words (e.g. pixel values in image, audio samples).
- The **payload** is arranged as stream of \( M \) bits.
- The cover object’s least significant bits are overwritten by the payload to form the **stego object**.

1. **Spread embedding**: overwrite a pseudorandom sequence of \( M \) LSBs. Each cover word has LSB flipped, independently, with probability \( M/2N \).

2. **Sequential embedding**: overwrite the first \( M \) LSBs. First \( M \) cover words have LSB flipped, independently, with probability \( 1/2 \).

In either case,
- the same number of LSBs are flipped by the embedding process,
- modifications are invisible to the eye,

*but sequential embedding “ought to be” easier to detect statistically.*
Steganalysis of LSB Replacement

There are many detectors* of LSB replacement:

- “RS” Fridrich, Goljan, & Du, 2001
- “Sample Pairs” Dumitrescu, Xu, & Wang, 2002
- “Pairs” Fridrich, Goljan, & Soukal, 2003
- “Least-Squares” Lu, Luo, Tang & Shen, 2004
- “Triples” Ker, 2005
- “ML Structure” Ker, 2007
- “Chi-Square” Westfeld & Pfitzman, 1999
- “Max. Likelihood” Dabeer et al, 2004
- “Empirical PMF” Draper et al, 2005
- “Weighted Stego” Fridrich & Goljan, 2004

*payload size estimators
Steganalysis of LSB Replacement

- “RS”
- “Sample Pairs”
- “Pairs”
- “Least-Squares”
- “Triples”
- “ML Structure”

Use “structural” analysis of LSB flipping.
- Highly sensitive to spread LSB replacement.
- Ineffective against sequential LSB replacement.
- Cannot be adapted to work in the sequential case.

- “Chi-Square”
- “Max. Likelihood”
- “Empirical PMF”

Based on signal-processing techniques.
- Poor sensitivity for sequential or spread embedding.
- Can sometimes be specialised to the sequential case, but remain weak.

- “Weighted Stego”

The subject of this talk.
- Quite good sensitivity for spread LSB payload.
- About equally good against sequential LSB payload.
The WS Method

Cover image: \( c_1, c_2, \ldots, c_N \)

Stego image: \( s_1, s_2, \ldots, s_N \)

"Weighted stego image": \( w_1^\alpha, w_2^\alpha, \ldots, w_N^\alpha \)
(real-valued)

Flip proportion \( M/2N \) of LSBs

Move \( \alpha \) towards flipping all LSBs
\[
    w_i^\alpha = \alpha \overline{s_i} + (1 - \alpha)s_i
\]

Theorem [Fridrich & Goljan, 2004]

The function \( E(\alpha) = \sum_{i=1}^{N} (w_i^\alpha - c_i)^2 \) is minimized at \( \alpha = M/2N \).

WS Steganalysis

1. Estimate cover by filtering stego image: \( \hat{c}_i = \text{average of surrounding four } s_i \).
2. Estimate size of payload
\[
    \hat{M} = 2N \arg\min_{\alpha} \sum_{i=1}^{N} (w_i^\alpha - \hat{c}_i)^2 = 2 \sum_{i=1}^{N} (s_i - \hat{c}_i)(s_i - \overline{s_i}).
\]
Sequential WS

**Cover image:** $c_1, c_2, \ldots, c_N$

**Stego image:** $s_1, s_2, \ldots, s_N$

**Weighted stego image:** $w_1^j, w_2^j, \ldots, w_N^j$

Theorem

The function $F(j) = \sum_{i=1}^{N} (w_i^j - c_i)^2$ is minimized at $j = M$.

Sequential WS Steganalysis

1. Estimate cover by filtering stego image: $\hat{c}_i = \text{average of surrounding four } s_i$.
2. Estimate size of payload

$$\hat{M} = \arg\min_{j} \left( \sum_{i=1}^{j} \left( \frac{1}{2} s_i + \frac{1}{2} \overline{s}_i - \hat{c}_i \right)^2 + \sum_{i=j+1}^{N} (s_i - \hat{c}_i)^2 \right).$$
Efficient Implementation

We need to determine

\[ \hat{M} = \text{argmin}_j \left( \sum_{i=1}^{j} \left( \frac{1}{2} s_i + \frac{1}{2} \bar{s}_i - \hat{c}_i \right)^2 + \sum_{i=j+1}^{N} (s_i - \hat{c}_i)^2 \right) \]  

(\ast)

The naïve implementation is \( O(N^2) \) ...

... but the recurrence

\[ g_0 = 0 \]
\[ g_j = g_{j-1} + \left( \frac{1}{2} (s_j + \bar{s}_j) - \hat{c}_j \right)^2 - (s_j - \hat{c}_j)^2 \]

satisfies \( F(j) = g_j + \text{constant} \) thus (\ast) can be found in linear time.
Performance: Spread Embedding

Experimental data from:
- 3000 grayscale bitmap cover images 0.3Mpixels,
- 20 different-sized payloads in each, creating 60000 stego images.

Proportionate payload $p = \frac{M}{N}$
**Performance: Sequential Embedding**

Experimental data from:
- 3000 grayscale bitmap cover images 0.3Mpixels,
- 20 different-sized payloads in each, creating 60000 stego images.
Conclusions

• Sequential LSB replacement is one of the worst possible choices to embed data secretly.

The embedding procedure has structure, and the payload is located predictably.

• There was no previous sensitive detector for it.

The most sensitive ("structural") detectors for spread LSB replacement do not adapt to sequential embedding.

• The WS detector can be adapted, and the new detector’s performance is superior.

  • 1000 1.5Mpixel grayscale RAW images from digital cameras;
  • Payloads of 500000 bits embedded sequentially;
  • Sequential WS payload estimates: over 90% were within 120 of 500000.
End

adk@comlab.ox.ac.uk