Toss that BOSSbase, Alice!

Vahid Sedighi, Jessica Fridrich, and Rémi Cogranne
Current steganography paradigm

- Content-adaptive steganography
  - Embed in textured/noisy areas that are harder to model and steganalyze
Current steganalysis paradigm

- Images represented with rich media models that use knowledge of the selection channel
- Classifiers trained on examples of cover and stego images
Benchmark

- **BOSSbase 1.01**
  - 10,000 images taken in the RAW format
  - Seven different cameras
  - Converted to grayscale, downsampled using the Lanczos resampling algorithm with antialiasing turned OFF
  - Cropped to the final size of $512 \times 512$ pixels

- The sole source on which the steganographers based their design

- BOSSbase images are far from what many would consider natural

Toss that BOSSbase, Alice!
**BOSSbase Oddities**

- Aggressive downsizing of the original full-resolution RAW images
  - complex content with weak dependency among pixels
  - suppress color interpolation artifacts
- It contains a lot of under exposed, out of focus, and dark images

- Makes the design overoptimized and suboptimal to other sources

Toss that BOSSbase, Alice!
Two New Versions of BOSSbase

- **BOSSbaseC (C as in Cropped)**
  - Same script as BOSSbase 1.01 but resizing skipped
  - Images are centrally cropped to 512 x 512 pixels after conversion from RAW format to grayscale
  - Less textured source but do contain acquisition noise

- **BOSSbaseJQF (J as in JPEG, QF is the JPEG quality factor)**
  - Formed from BOSSbase 1.01 images
  - JPEG compressing with quality factor $QF \in \{75, 85, 95\}$ and then decompressing to the spatial domain as an 8-bit grayscale
  - The low-pass character of JPEG compression makes them less textured and less noisy
Sample Images

1013.pgm  1014.pgm  1015.pgm  1016.pgm

BOSSbase 1.01

BOSSbaseC

BOSSbase-IQF85

Toss that BOSSbase, Alice!
Embedding Schemes

- **Cost based schemes**
  - Wavelet Obtained Weights (WOW) [Holub et al., WIFS 2012]
  - UNIversal WAvelet Relative Distortion (S-UNIWARD) [Holub et al., IH 2013]
  - High-Low-Low (HILL) [Li et al., ICIP 2014]

- **Model based scheme**
  - Minimizing the power of the most POwerful Detector (MiPOD) [Sedighi et al., SPIE 2015]
Experimental Setup

- FLD ensemble [Kodovsky et al., TIFS 2012] with SRM (Spatial Rich Model) [Fridrich et al., TIFS 2011] and maxSRM (selection-channel-aware SRM) [Denemark et al., WIFS 2014]

- Security evaluated using minimal total classification error under equal priors averaged over 10 random 5000/5000 database splits:

\[
\overline{P}_E = \min_{P_{FA}} \frac{1}{2} (P_{FA} + P_{MD})
\]
HILL and MiPOD are the most secure schemes

WOW is the least secure embedding method on both feature sets
All embedding schemes exhibit similar security using SRM features.

Using maxSRMd2 features, S-UNIWARD becomes the most secure scheme.
WOW is the most secure embedding method on both features sets.

MiPOD is the least secure embedding scheme on both feature sets.
Embedding in regions with an “edge” in the horizontal, vertical, and both diagonal directions

Three directional filters with $8 \times 8$ kernels denoted $K^{(k)}$, $k \in \{h, v, d\}$ are used to extract three directional residuals $R^{(k)} = K^{(k)} \star X$

Embedding suitabilities: $\xi^{(k)} = |R^{(k)}| \star |K^{(k)}|$

The embedding cost is obtained using the reciprocal Hölder norm

$$\rho_{i,j}^{(k)} = \left( \sum_{k=1}^{3} |\xi_{i,j}^{(k)}|^p \right)^{-p} \text{ with } p = -1$$
Optimizing S-UNIWARD

- Pixel embedding costs are obtained from a distortion function defined as the sum of relative absolute differences between wavelet coefficients of cover and stego images.

- Denote the \( u, v \)th wavelet coefficient of \( X \) in \( k \in \{ h, v, d \} \) subband with \( W_{uv}^{(k)}(X) \), \( W^{(k)} = K^{(k)} \ast X \), \( u, v \) of the same range as image pixels.

- S-UNIWARD uses the same kernels formed from 8-tap Daubechies wavelets as WOW.

- Non-additive distortion between the cover \( X \) and the stego image \( Y \) is used in UNIWARD.

\[
D(X, Y) = \sum_{k \in \{h,v,d\}} \sum_{u,v} \frac{|W_{uv}^{(k)}(X) - W_{uv}^{(k)}(Y)|}{\sigma + |W_{uv}^{(k)}(X)|}
\]

- \( \sigma = 1 \) is the stabilizing constant.
Optimizing HILL

- This algorithm originated from WOW
- Three directional kernels are replaced with one non-directional high-pass $3 \times 3$ KB kernel $H$.
- HILL thus uses a single residual $R = X \star H$
- The pixel costs are then computed using the following formula:

$$\rho = \frac{1}{|R| \star L_1} \star L_2$$

- $L_1$ is an averaging filter of support $3 \times 3$ and $L_2$ is another averaging filter of support $15 \times 15$
Optimizing MiPOD

- \( r_n \sim \mathcal{N}(0, \sigma_n^2) = (p_{\sigma_n}(k))_{k \in \mathbb{Z}} \) independent with

\[
p_{\sigma_n}(k) = \mathbb{P}(r_n = k) \propto (2\pi\sigma_n^2)^{-1/2} \exp\left(-\frac{k^2}{2\sigma_n^2}\right)
\]

- Warden faces a simple binary hypothesis test:

\[
\begin{align*}
\mathcal{H}_0 : & \quad x_n \sim \mathcal{P}_{\sigma_n} \\
\mathcal{H}_1 : & \quad x_n \sim \mathcal{Q}_{\sigma_n, \gamma_n}
\end{align*}
\]

\[
\quad \Rightarrow \quad \varrho = \frac{\sqrt{2} \sum_{n=1}^{N} \beta_n \gamma_n \sigma_n^{-4}}{\sqrt{\sum_{n=1}^{N} \gamma_n^2 \sigma_n^{-4}}}
\]

- \( \beta_n \) determined by minimizing the deflection coefficient \( \varrho \) with payload constraint using method of Lagrange multipliers.

- Estimate variance using local fitting with a two-dimensional DCT filter with degree 8 in a 9 \( \times \) 9 sliding window.

- The Fisher information is low-pass filtered with an averaging filter of size 7 \( \times \) 7.
Little gain when steganalyzing with SRM

S-UNIWARD remains the most secure embedding scheme with maximum gain from the search
The least and most secure embedding schemes keep their places swapped after the search.

The search reveals that smaller support for the residuals is better.
Synchronizing Embedding Changes

- Empirical security of embedding schemes built around an additive distortion function can be increased by synchronizing the polarity of embedding changes
- CMD [Li et al., TIFS 2015] and Synch [Denemark et al., IHMMSSec 2015]
- Higher change rate but ultimately better security
- The same four embedding algorithms are investigated on the new sources
CMD works slightly better than Synch

HILL and MiPOD benefit the most using maxSRMd2 feature set
Synchronization on BOSSbaseC

- Biggest synchronization impact with up to 3.6% for HILL using maxSRMd2 feature set
- The ranking does not change as the result of synchronization
WOW is the most secure embedding method on both features sets

MiPOD is the least secure embedding scheme on both feature sets
Summary

- We study the effect of using different sources on the empirical security of the state-of-the-art steganographic schemes.

- Statistical properties of pixels can change dramatically after filtering, compression, and resizing of images.

- Even after optimization of embedding schemes to a new cover source, different embedding schemes rank differently.
  - The least secure embedding scheme on BOSSbase 1.01, WOW, becomes the most secure on BOSSbaseJQF.
  - The most secure scheme on BOSSbase 1.01, MiPOD, becomes the least secure on BOSSbaseJQF.

- The effectiveness of certain boosting measures, such as synchronizing the polarity of the embedding changes vastly change across sources.
Questions

Matlab code available from http:\\dde.binghamton.edu\download

Toss that BOSSbase, Alice!