Steganalysis with Cover-Source Mismatch and a Small Learning Database

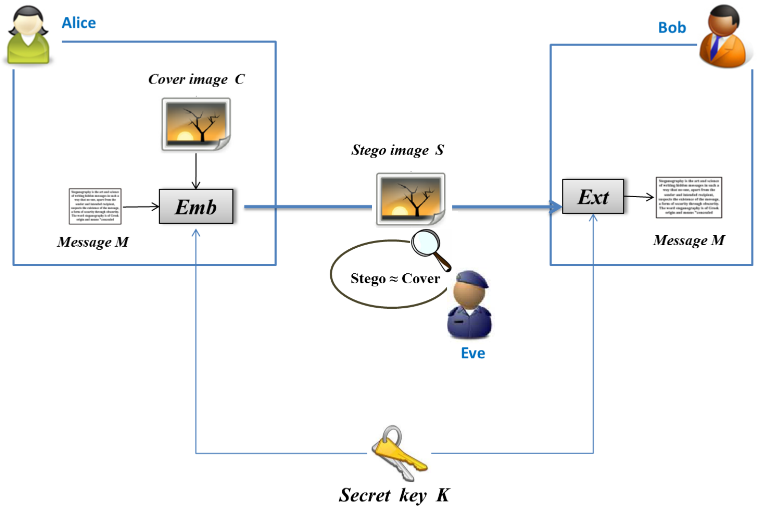
Jérome Pasquet, Sandra Bringay, Marc Chaumont
LIRMM, Montpellier, France

September 19, 2014

EUSIPCO 2014, 22nd European Signal Processing Conference 2014, Lisbon, Portugal, sep., 2014

Outline

Steganography vs Steganalysis



Eve (the steganalyst) job

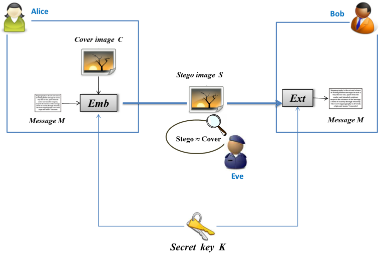
In the **clairvoyant scenario**, we decide that Eve knows:

- the algorithm(s) used by Alice,
- the payload (quantity of embedded bits) used by Alice,
- the size of images,
- quite well the distribution of Alice images.

Eve job is:

1. to learn to distinguish cover images from stego images = learning step,
2. to do the steganalysis = testing step.

Cover-Source Mismatch scenario (a closer step to reality)



In the **Cover-Source Mismatch scenario** ($\neq$ clairvoyant scenario), Eve, the steganalyst, has partial or erroneous knowledge of the cover distribution.

Definition: Cover-Source Mismatch phenomenon (= inconsistency)

Image model learned by Eve $\neq$ Image model used by Alice

History

- The cover-source mismatch phenomenon reported in 2008 [1],
- The only solution to manage cover source mismatch was proposed in 2012 by Lubenko and Ker [2, 3],
- Lubenko and Ker solution necessitate **million of images** for the learning step.

[1] G. Cancelli, G. J. Doërr, M. Barni, and I. J. Cox,
"A comparative study of +/-1 steganalyzers,"
in Workshop Multimedia Signal Processing, MMSP'2008.

[2] I. Lubenko and A. D. Ker,
"Going from small to large data in steganalysis,"
in Media Watermarking, Security, and Forensics III, Part of IS&T/SPIE Annual Symposium on Electronic Imaging, SPIE'2012.

[3] I. Lubenko and A. D. Ker,
"Steganalysis with mismatched covers: do simple classifiers help?,"
in ACM Multimedia and Security Workshop, MM&Sec'2012.

The proposition

Overcoming the cover-source mismatch problem

- We refute the hypothesis that millions of images are necessary to overcomes the problem of cover-source mismatch,
- We experimentally show that EC with post-features selection (EC-FS) [4] allows to obtain better results with 100 fewer images than [2, 3],
- We introduce an additional pre-processing that overcomes the problem of cover-source mismatch (the islet approach).

[4] M. Chaumont and S. Kouider,
"Steganalysis by ensemble classifiers with boosting by regression, and postselection of features,"
in IEEE International Conference on Image Processing, ICIP'2012.

Outline

Ensemble algorithms

The two competing algorithms:

- EAP : **Ensemble** Average Perceptron [3].
- EC-FS : **Ensemble** Classifier with Post-Selection of Features[4],

[3] I. Lubenko and A. D. Ker,
"Steganalysis with mismatched covers: do simple classifiers help?,"
in ACM Multimedia and Security Workshop, MM&Sec'2012.

[4] M. Chaumont and S. Kouider,
"Steganalysis by ensemble classifiers with boosting by regression, and postselection of features,"
in IEEE International Conference on Image Processing, ICIP'2012.

Ensemble Classifier: Definition of a weak classifier

An Ensemble Classifier (EAP or EC-FS) is made of L weak classifiers

- Let $\mathbf{x} \in \mathbb{R}^d$ a features vector,
- A weak classifier, h_l, returns -1 for cover, 1 for stego :

$$\begin{aligned} h_l : \mathbb{R}^d &\rightarrow \{-1, +1\} \\ \mathbf{x} &\rightarrow h_l(\mathbf{x}) \end{aligned}$$

Ensemble Classifier: Recall of how classification works.

Classification working using EAP [3] or EC-FS [4]:

1. Take an image to analyze (i.e. classify in cover or stego),
2. Extract the features vector $\mathbf{x} \in \mathbb{R}^d$,
3. Decide to classify cover or stego (majority vote):

$$C(\mathbf{x}) = \begin{cases} -1 \text{ if } \sum_{l=1}^{l=L} h_l(\mathbf{x}) \leq 0, \\ +1 \text{ otherwise.} \end{cases}$$

[3] I. Lubenko and A. D. Ker,
"Steganalysis with mismatched covers: do simple classifiers help?,"
in ACM Multimedia and Security Workshop, MM&Sec'2012.

[4] M. Chaumont and S. Kouider,
"Steganalysis by ensemble classifiers with boosting by regression, and postselection of features,"
in IEEE International Conference on Image Processing, ICIP'2012.

EC-FS

EC-FS (Ensemble Classifier with Post-Selection of Features):

- was presented at IEEE ICIP'2012 [4],
- is an extension of EC [5],
- increase the performance in the clairvoyant scenario,
- is scalable regarding the dimension of the features vector, has low computational complexity $O(d_{red}^2.L.N)$, has low memory complexity, is easily parallelizable.

[4] M. Chaumont and S. Kouider,
"Steganalysis by ensemble classifiers with boosting by regression, and postselection of features,"
in IEEE International Conference on Image Processing, ICIP'2012.

[5] J. Kodovský, J. Fridrich, and V. Holub,
"Ensemble classifiers for steganalysis of digital media,"
IEEE Transactions on Information Forensics and Security, TIFS'2012.

EC-FS: Selection of features...

Once a weak classifier learned:

Algorithm :

1. Compute a **score** for each feature; first database reading,
2. Define an order of selection of the features,
3. Find the best subset (lowest P_E)
   = suppress the features in order to reduce P_E;
   second database reading.

Order of complexity unchanged.

[4] M. Chaumont and S. Kouider,
"Steganalysis by ensemble classifiers with boosting by regression, and postselection of features,"
in IEEE International Conference on Image Processing, ICIP'2012.

EAP

EAP (Ensemble Average Perceptron):

- was presented at IS&T/SPIE'2012 and MM&Sec'2012 [2, 3],
- use the very old notion of perceptron (1957) = simplest network neuron,
- has very low computational complexity $O(d.L.N)$, has quasi null memory complexity (online algorithm), is easily parallelizable.
- **but** necessitates million of images in the cover-source mismatch scenario,

[2] I. Lubenko and A. D. Ker,
"Going from small to large data in steganalysis,"
in Media Watermarking, Security, and Forensics III, Part of IS&T/SPIE Annual Symposium on Electronic Imaging, SPIE'2012.

[3] I. Lubenko and A. D. Ker,
"Steganalysis with mismatched covers: do simple classifiers help?,"
in ACM Multimedia and Security Workshop, MM&Sec'2012.

EAP: Main concept

A weak classifier is an average perceptron:

$$\begin{array}{rcl} h_l : \mathbb{R}^d & \rightarrow & \{-1, +1\} \\ \mathbf{x} & \rightarrow & h_l(\mathbf{x}) = sign(\mathbf{w}^{avg}.\mathbf{x}) \end{array}$$

For an incoming features vector $\mathbf{x}_i$ with a class number $y_i \in \{-1, +1\}$, the weight vector $\mathbf{w}^{(i)}$ is updated such that:

$$\mathbf{w}^{(i)} = \begin{cases} \mathbf{w}^{(i-1)} & if \;\; y_i = sign(\mathbf{w}^{avg}.\mathbf{x}_i) \\ \mathbf{w}^{(i-1)} + y_i.\mathbf{x}_i & if \;\; y_i \neq sign(\mathbf{w}^{avg}.\mathbf{x}_i) \end{cases}$$

[2] I. Lubenko and A. D. Ker,
"Going from small to large data in steganalysis,"
in Media Watermarking, Security, and Forensics III, Part of IS&T/SPIE Annual Symposium on Electronic Imaging, SPIE'2012.

Outline

The idea

Reducing the heterogeneity before the learning process.

Before the learning step, there are two stages:

1. Partition the image database in a few clusters;
   $\rightarrow K$ vectors $\{\mu_k\}_{k=1}^{k=K}$,
2. Associate a classifier (EC-FS) to each cluster;
   $\rightarrow K$ classifiers.

During the learning step, each classifier learn and classify only vectors that belong to its cluster.

The classification process

During the testing step: Given a features vector $\mathbf{x}_i$ to be classified:

1. Select cluster **k** such that $k = \underset{k}{\arg} \min_{k \in \{1, \ldots k\}} dist(\mathbf{x}_i, \mu_k)$,
2. Use the $\mathbf{k}^{th}$ classifier (EC-FS) to classify $\mathbf{x}_i$ (into cover or stego).

Outline

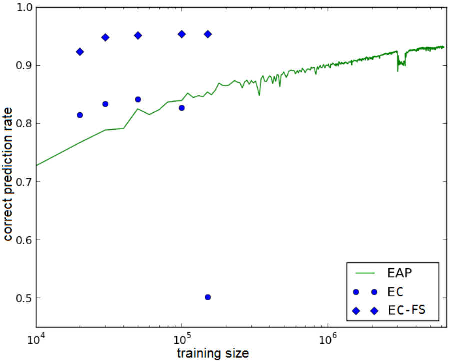
Experimental conditions

- 1 million of images from the TwitPic website,
- Images are decompressed, transformed, and cropped to 450×450,
- Spatial embedding with the HUGO [6] algorithm at 0.35 bpp,
- 3 steganalysis simulations,
- Features vector dimension $d = 34671$ features [7],
- Average P_E computed on 40 000 images never seen.

[6] T. Pevný, T. Filler, and P. Bas,
HUGO: "Using High-Dimensional Image Models to Perform Highly Undetectable Steganography"
in Information Hiding, IH'2010.

[7] J. Fridrich, J. Kodovský,
Rich models: "Rich models for steganalysis of digital images,"
in IEEE Transactions on Information Forensics and Security, TIFS'2012.

Steganalysis results



- Counter-performance of EC,
- EAP prediction rate converge around 93%,
- EC-FS prediction rate = 95% with only 50 000 learning.

Results for **Islet** approach

K islets	Training size per islet	Prediction rate
1	150 000	95.39
2	75 000	95.81% (+0.41%)
3	50 000	95.83% (+0.43%)
4	37 500	95.82% (+0.43%)
5	30 000	95.88% (+0.49%)
6	25 000	96.06% (+0.67%)
7	21 428	95.72% (+0.33%)

Table: Results of islets with EC-FS.

- Less samples per classifier but more homogeneity!
- When alone, EC-FS is converging to 95%;
  $\rightarrow$ The islets allow to overcome this bound,
- Non negligible improvement (we are close to 100%...).

Outline

Summary

- EC-FS is a very efficient tool for managing very heterogeneous data (overcomes the cover-source mismatch phenomenon),
- EC-FS gives better prediction rate than EAP (+2,3%),
- EC-FS requires a learning set 100 times smaller than EAP (experiments may require High Performance Computing Architectures),
- The islet approach is an additional efficient technique (+0.67%) (it acts on increasing homogeneity).