Steganalysis with Cover-Source Mismatch and a Small Learning Database

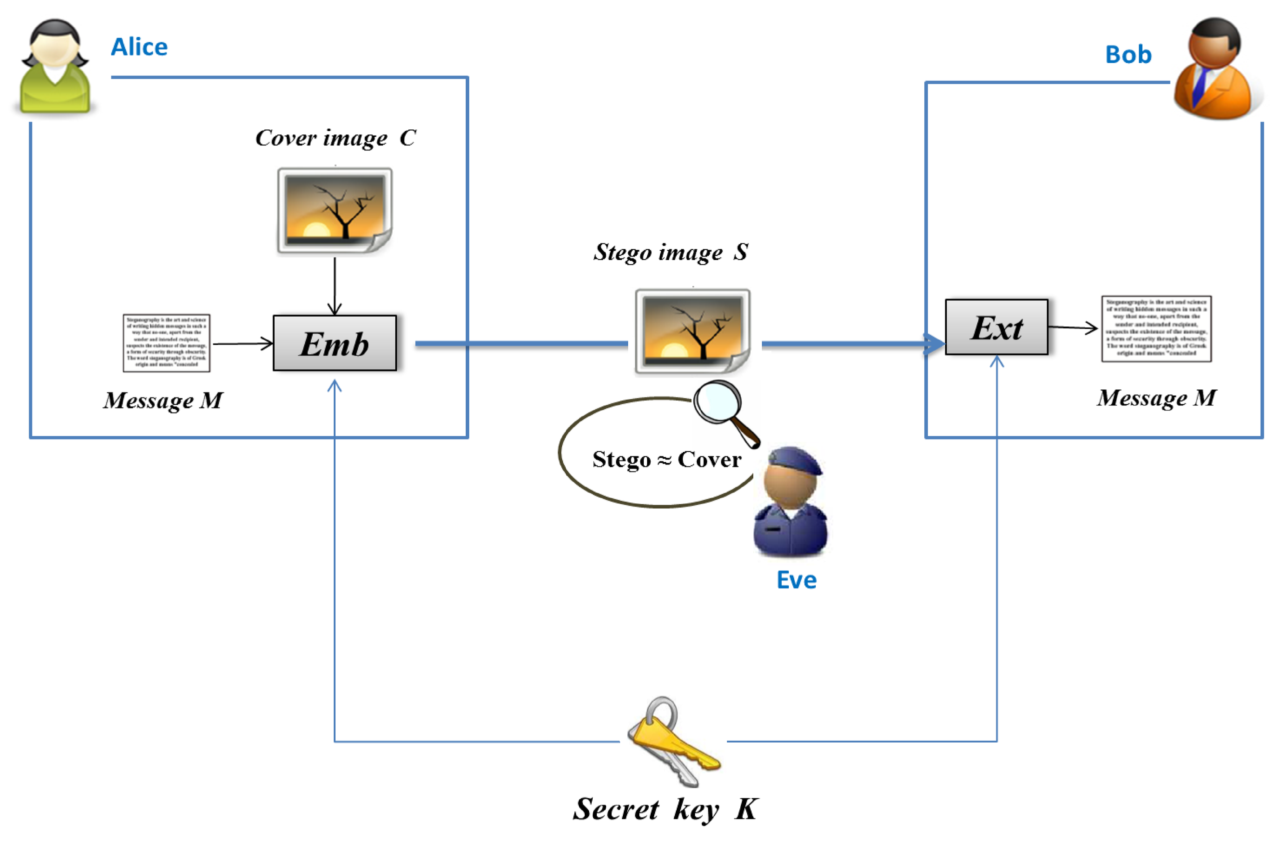
Pasquet Jérôme[2,3], Sandra Bringay[2,3,4] et Marc Chaumont[1,2,3]

[1] UNIVERSITE DE NIMES, F-30021 Nîmes Cedex 1, France
[2] UNIVERSITE MONTPELLIER 2, UMR5506-LIRMM, F-34095 Montpellier Cedex 5, France
[3] CNRS, UMR5506-LIRMM, F-34392 Montpellier Cedex 5, France
[4] AMIS, UNIVERSITE MONTPELLIER 3, Route de Mende 34199 Montpellier Cedex 5, France

{jerome.pasquet, sandra.bringay, marc.chaumont}@lirmm.fr

Steganalysis

Steganalysis is the study of detecting messages hidden in a support.



Eve (the steganalyst) job

Eve's Job is :

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

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 sizes of images,
- quite well the distribution of Alice images.

far from the reality.

↓ A closer scenario to reality

Using the Cover-Source Mismatch scenario [1]

Definition: Cover-Source Mismatch phenomenon (= inconsistency)

Image model learned by Eve and image model used by Alice are differents

The proposition to overcome the cover-source mismatch problem

- We **refute** the hypothesis that millions of images are necessary to overcome the problem of cover-source mismatch.
- Experiment 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 preprocessing technique that overcomes the problem of cover-source mismatch (the islet approach).

Islet approach

Main Idea : Reducing the heterogeneity before the learning process.

Before the learning step, there are two stages:

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

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

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

1. A cluster k is selected such that $k = \arg\min_{k \in \{1,\dots,K\}} \mathrm{dist}(\mathbf{x}_i, \mu_k)$,
2. The k^{th} classifier (EC-FS) is used to classify x i (into cover or stego).

Ensemble algorithms

An Ensemble Classifier is made of L weak classifiers

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

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

The two competing algorithms:

EAP [3] Ensemble Average Perceptron of Features	EC-FS [4] Ensemble Classifier with Post-Selection
• 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 $O(d_{red}.L.N)$ and quasi null memory complexity (online algorithm), • but necessitates million of images in the cover-source mismatch scenario.	• was presented at IEEE ICIP'2012, • 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 $O(d^2_{red}.L.N)$ and low memory complexity.

EAP:

The weak classifier is an average perceptron :

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

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

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

EC-FS:

Once a weak classifier is learned :

Algorithm :

1. Compute a score for each feature
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

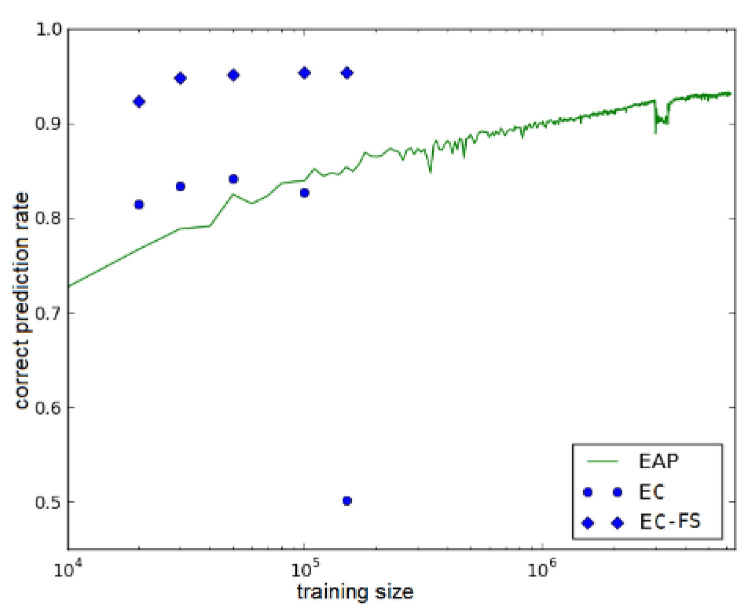
Order of complexity unchanged.

Results

Experimental conditions:

- 1 million 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 is d = 34671 features [7],
- Average P_E computed on 40 000 images never seen.

Steganalysis results:



- Counter-performance of EC
- EAP prediction rate converges 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!
- EC-FS alone converges to 95%
  → The islets allow to overcome this bound
- Non negligible improvement (we are close to 100%)

Summary

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

[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.

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

[6] T. Pevny, 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. Kodovsk y, Rich models: "Rich models for steganalysis of digital images," in IEEE Transactions on Information Forensics and Security, TIFS'2012.