Pooled steganalysis in JPEG: how to deal with the spreading strategy?

Ahmad ZAKARIA$^{1,2}$, Marc CHAUMONT$^{1,4}$, Gérard SUBSOL$^{1,3}$
LIRMM$^1$, Univ Montpellier$^2$, CNRS$^3$, Univ Nîmes$^4$, Montpellier, France

December 11, 2019

WIFS’2019, IEEE International Workshop on Information Forensics and Security,
December 9-12, 2019, Delft, The Netherlands.
Outline

Introduction

Pooled steganalysis architecture

Experimental protocol

Results

Conclusions and perspectives
Steganography / Steganalysis

Pooled steganalysis in JPEG: how to deal with the spreading strategy?

Introduction
Batch steganography / Pooled steganalysis

Alice:

- spreads a message \( m \in \{0, 1\}^{|m|} \),
- in multiple covers,
- using a strategy \( s \in S \).

A. D. Ker, “Batch steganography and pooled steganalysis,” in IH’06
Examples of possible spreading strategies

The 6 evaluated spreading strategies in this paper, \( S = \{ \text{IMS}, \text{DeLS}, \text{DiLS}, \text{Greedy}, \text{Linear}, \text{and Uses} - \beta \} \)
Pooled steganalysis: how to deal with the spreading strategy?

Many possibilities for Alice to spread the message;
What about Eve, the steganalyst?
Many possibilities for Alice to spread the message; What about Eve, the steganalyst?

Recent approaches opt for **pooling** individual scores (more general)
Many possibilities for Alice to spread the message; What about Eve, the steganalyst?

Recent approaches opt for pooling individual scores (more general)
Let us denote, \( f \), a **Single Image Detector (SID)**;
For example a payload predictor (quantitative steganalysis):

\[
f : \mathbb{R}^{r \times c} \rightarrow \mathbb{R}^+
\]
Pooled steganalysis: how to deal with the spreading strategy?

Many possibilities for Alice to spread the message; What about Eve, the steganalyst?

Recent approaches opt for **pooling** individual scores.
Recent studies

- [1] Hypothesis: Eve *does not know* the spreading strategy
  ⇒ best pooling strategy = *averaging* the individual scores

- [2] Hypothesis: Eve *does know* the spreading strategy
  ⇒ knowledge of the strategy = improves steganalysis results.

- [3] Hypothesis: Eve *does know* the spreading strategy
  ⇒ knowledge of the strategy = improves steganalysis results.


The addressed question

Hypothesis: *Eve does not know* the spreading strategy.

Can Eve “do better” than averaging the individual scores?
Outline

Introduction

Pooled steganalysis architecture

Experimental protocol

Results

Conclusions and perspectives
Given a vector of SID scores $z = \{f(x_1), ..., f(x_b)\}$:

$$h = \left[ \frac{1}{b} \sum_{f(x_i) \in z} k(f(x_i), c_1), ..., \frac{1}{b} \sum_{f(x_i) \in z} k(f(x_i), c_p) \right],$$

with $\{c_i\}_{i=1}^p$ a set of equally spaced real positive values, and $k(x, y) = \exp(-\gamma \|x - y\|^2)$.
T. Pevny and I. Nikolaev general architecture

- Histogram → can treat a bag of any dimension,
- Histogram → invariant to the sequential order in the bag.
The Single Image Detector (SID)

- Note: Alice embeds using J-UNIWARD (512×512 BossBase1.01 QF=75).
- Quantitative steganalysis in JPEG [1].
- GFR cleaned and normalized:
  - Gabor Features Residuals (GFR) of dimension 17 000 [2],
  - Clean cleaned from NaN values and from constant values → reduced to 16 750,
  - Normalize using random conditioning [3].

Learning: 5 000 covers + 5 000 stego per payload size (\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\} bpc).

- Note: M. Chen, M. Boroumand, and J. J. Fridrich, “Deep learning regressors for quantitative steganalysis,” in EI’2018 MWSF, is more efficient.
Outline

Introduction

Pooled steganalysis architecture

Experimental protocol

Results

Conclusions and perspectives
Alice: Batch spreading strategies

1. **Greedy strategy**: spreading into as few covers as possible.
2. **Linear strategy**: spreading evenly.
3. **Uses-β strategy**: spreading evenly across a fraction of covers.
4. **IMS strategy**: spreading in an unique artificial image.
5. **DeLS strategy**: spreading at the same deflection coefficient (MiPod model).
6. **DiLS strategy**: spreading at the same distortion.
Eve: Pooling strategies

- $g_{clair}$: Eve (clairvoyant) knows the spreading strategy. SVM learned on the known strategy $s \in S$.
- $g_{disc}$: Eve (discriminative) does not know the spreading strategy. SVM learned on all the strategies $S$.
- $g_{max}$: Maximum function AND $\tau_{max}$ by minimizing $P_e$ over $S$.
- $g_{mean}$: Average function AND $\tau_{min}$ by minimizing $P_e$ over $S$. 
Bags for the learning and for the test

$g_{clair}$ (clairvoyant) learning:
- Choose **one** bag size $b \in B = \{2, 4, 6, 10, 20, 50, 100, 200\}$,
- Choose **one** spreading strategies $s \in S$,
- Generate 5 000 cover bags and 5 000 stego bags (0.1 bptc).

$g_{clair}$ testing:
- Choose **the same** bag size $b$,
- Choose **the same** spreading strategies $s$,
- Generate 5 000 cover bags and 5 000 stego bags (0.1 bptc).
Bags for the learning and for the test

\( g_{\text{disc}} \) (discriminative), \( g_{\text{max}} \), and \( g_{\text{mean}} \) learning:

- Choose **one** bag size \( b \in B = \{2, 4, 6, 10, 20, 50, 100, 200\} \),
- Choose **all** the spreading strategies from \( S \),
- Generate 5 000 cover bags and 5 000 stego bags. **833 bags per strategy** (0.1 bptc).

\( g_{\text{disc}} \) (discriminative), \( g_{\text{max}} \), and \( g_{\text{mean}} \) testing:

- Choose **the same** bag size \( b \),
- Choose **one** spreading strategies \( s \in S \) (**unknown from Eve**),
- Generate 5 000 cover bags and 5 000 stego bags (0.1 bptc).
Outline

Introduction

Pooled steganalysis architecture

Experimental protocol

Results

Conclusions and perspectives
Alice: Spreading strategies comparison (Eve clairvoyant)

Figure: Spreading strategies comparison in the clairvoyant case (10 runs).
Eve: Pooling function comparisons

Figure: Pooled steganalysis comparison (10 runs).
Outline

Introduction

Pooled steganalysis architecture

Experimental protocol

Results

Conclusions and perspectives
Conclusions

Up-to-date algorithms:

▶ modern embedding (J-Uniward),
▶ 6 spreading strategies (3 moderns),
▶ modern (generic) pooling architecture.

→ Coherent results with past papers.

The take away messages:

▶ For Alice: DeLS is a really interesting spreading strategy.
▶ For Eve: $g_{disc}$ pooling can improve the detectability if Eve does not know the spreading strategy.
To be continued...

Future:

- DeLS with a DCT model,
- Robustness to the bag size variation (learn only once with various size),
- Robustness to the mismatch in the spreading strategy (uses a different strategy in the test; Examples in [1]),
- Minimize the Pe (for $g_{disc}$) differently for each strategy,
- Use something more powerful than an SVM,
- Extend to deep learning,
- Go toward a simulation of a game (GAN philosophy),