Effect of Saturated Pixels on Security of Steganographic Schemes for Digital Images

Vahid Sedighi and Jessica Fridrich
Binghamton University, USA

Quick background

- Steganography = data hiding or covert communication
- We hide in digital media by slightly changing pixel values:

$$\text{Cover} \rightarrow \text{Stego: } \mathbf{x} = (x_{ij}) \rightarrow \mathbf{y} = (y_{ij})$$

$$x_{ij}, y_{ij} \in \{0, \dots, 255\}$$

- Embedding is cast as source coding with fidelity constraint
- Given the cost of changing ijth pixel by $+1$, $\rho_{ij}^{(+)}$, and by -1, $\rho_{ij}^{(-)}$, the desired payload is embedded while minimizing the total distortion

$$D(\mathbf{x}, \mathbf{y}) = \sum_{x_{ij} \neq y_{ij}} \rho_{ij}^{(y_{ij} - x_{ij})}$$

- Pixel x_{ij} changes to $x_{ij} \pm 1$ with probability

$$\Pr\{y_{ij} = x_{ij} \pm 1\} = e^{-\lambda \rho_{ij}^{(\pm)}} / (1 + e^{-\lambda \rho_{ij}^{(\pm)}} + e^{-\lambda \rho_{ij}^{(\mp)}}) \triangleq \beta_{ij}^{(\pm)}$$

Dynamic Range Boundary Rules

- All state-of-the-art embedding schemes for the spatial domain that minimize additive distortion use symmetric costs, $\rho_{ij}^{(+)} = \rho_{ij}^{(-)}$
- When a cover pixel has a borderline value, $x_{ij} = 0$ or 255, embedding needs to be modified so that stego image stays within dynamic range
- It can be assured in at least three different ways:

- Rule 1: Embed and correct

$$\rho_{ij}^{(-)} = \rho_{ij}^{(+)} \implies \text{Embedding} \implies \begin{cases} y_{ij} = -1 & \implies y_{ij} = 2 \\ y_{ij} = 256 & \implies y_{ij} = 253 \end{cases}$$

- Rule 2: Forbid changes outside range

$$\begin{cases} x_{ij} = 0 & \implies \rho_{ij}^{(-)} = \infty \\ x_{ij} = 255 & \implies \rho_{ij}^{(+)} = \infty \end{cases} \implies \text{Embedding}$$

- Rule 3: Avoid saturated pixels altogether

$$(x_{ij} = 0 \text{ or } x_{ij} = 255) \implies \rho_{ij}^{(+)} = \rho_{ij}^{(-)} = \infty \implies \text{Embedding}$$

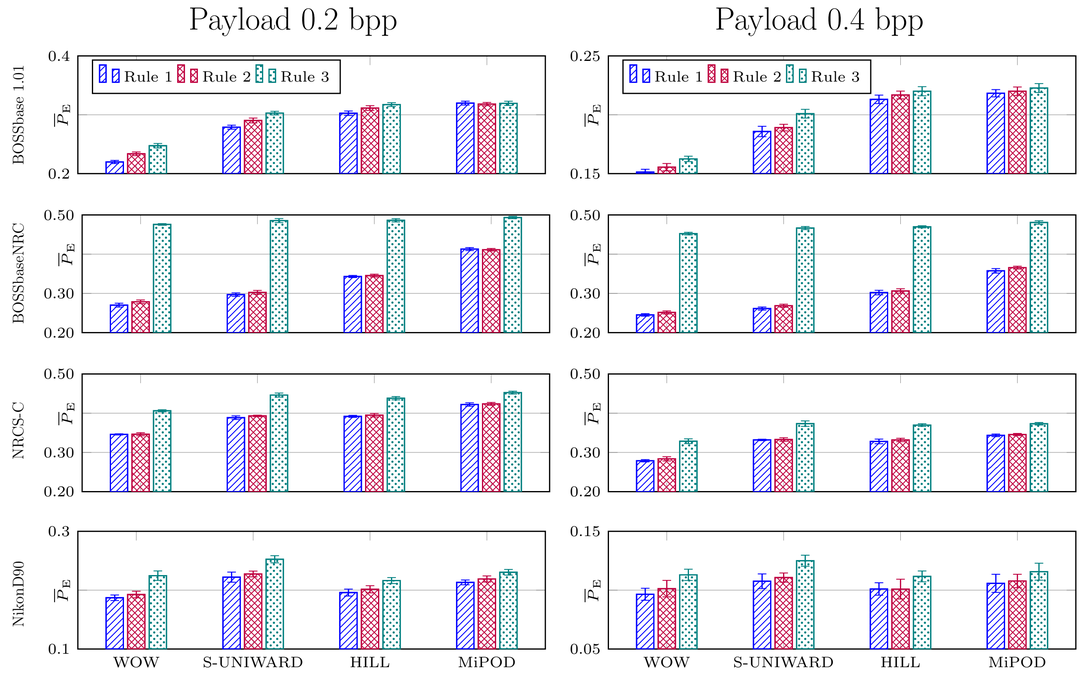
Setup of Experiments

- Image sources:
  - BOSSbase 1.01: 10,000 images, 7 different cameras, grayscale, downsampled and cropped to 512×512 pixels
  - BOSSbaseNRC (Non-interpolated Red Channel): 10,000 images, same RAW BOSSbase images, no color-interpolation and resizing, subsample only red channel by a factor of 2, gain and gamma adjustments
  - NRCS-C: 6,664 images, derived from NRCS database, RAW scans of nagatives from USDA Natural Resources Conservation Service, crop central 512×1024 part of each image and extract two 512×512 images
  - NikonD90: 2,276 images, subset of RAISE dataset taken with Nikon D90, converted to grayscale, cropped to 512×512
- Embedding schemes: WOW, S-UNIWARD, HILL, MiPOD
- Steganalysis features: selection-channel-aware Spatial Rich Model, maxSRMd2 , dim = 34,671 [1]
- Ensemble classifier [2]
- Detection reported using $\overline{P}_{\text{E}} = \min_{P_{\text{FA}}}(P_{\text{FA}} + P_{\text{MD}})/2$ averaged over ten random splits of the database into two halves

Image Source Facts

- BOSSbaseNRC contains the largest number of saturated pixels (2.2% on average)
- Saturated pixels form connected regions in all four sources
- The number of black pixels is comparatively much smaller and the pixels are more scattered across the image
- BOSSbaseNRC and NRCS-C are generally much noisier than BOSSbase 1.01 and NikonD90

Results



- On BOSSbaseNRC, using Rules 1 and 2, all embedding schemes become more detectable by about **20%** in comparison with Rule 3!
- High noise level of BOSSbaseNRC makes steganography undetectable everywhere except at the boundary of saturated regions where Rules 1 and 2 allow changes
- Rule 3 removes this flaw by avoiding saturated regions altogether

Location of Changes in Saturated Regions

Percentage of changed saturated pixels on the Boundary of Saturated Regions (BSR) and in the Middle of Saturated Regions (MSR) as the result of embedding with Rule 1 in NikonD90 at payload 0.4 bpp.

Saturation Type	WOW	S-UNIWARD	HILL	MiPOD
BSR	20.73	14.95	10.06	11.46
MSR	2.54	2.42	1.28	1.48

- Most changes in saturated areas are near the boundary
- All embedding schemes avoid these regions due to their high costs ρ_{ij}
- HILL and MiPOD avoid these boundaries more due to their cost design

Saturated Region Type

Detection error for Rule 1–3 and Rule 2 NE0 (No Embedding when $x_{ij} = 0$) but applies Rule 2 in saturated pixels, S-UNIWARD at 0.4 bpp.

	Rule 1	Rule 2	Rule 3	Rule 2 NE0
NikonD90	.1065±.0038	.1131±.0056	.1241±.0033	.1134±.0037
NRCS-C	.3195±.0028	.3337±.0030	.3740±.0035	.3387±.0042

- Increased detectability of Rules 1 and 2 is due to saturated pixels (Value 255) rather than black pixel (value 0)
- Saturation occurs even in very noisy images, while "underflow" is unlikely due to noise

Summary

- Addressed cost design in saturated regions
  - Three different treatments of these regions investigated
  - Statistical detectability can increase by 1%–20% with wrong rules
  - Conservative Rule 3 is the most secure and should be used in practice

References

[1] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich. Selection-channel-aware rich model for steganalysis of digital images. In *IEEE WIFS*, Atlanta, GA, Dec. 3–5, 2014.

[2] J. Kodovský, J. Fridrich, and V. Holub. Ensemble classifiers for steganalysis of digital media. *IEEE TIFS*, 7(2):432–444, 2012.