Content-Adaptive Pentary Steganography Using the Multivariate Generalized Gaussian Cover Model

Vahid Sedighi, Jessica Fridrich, and Rémi Cogranne

Binghamton University
State University of New York

utt
université de technologie Troyes

Current steganography paradigm

- Define **distortion** $D(\mathbf{x}, \mathbf{y})$ between cover image $\mathbf{x} = (x_n)_{n=1}^{N}$ and stego image $\mathbf{y} = (y_n)_{n=1}^{N}$
- Most common is **additive** distortion defined using **costs** ρ_n of changing cover pixel x_n to y_n, $n = 1, \ldots, N$

$$D(\mathbf{x}, \mathbf{y}) = \sum_{\substack{n=1 \\ x_n \neq y_n}}^{N} \rho_n$$

- $D(\mathbf{x}, \mathbf{y})$ is the sum of costs of all changed pixels
- Costs should be designed to measure the “statistical impact” of embedding changes

Properties of the proposed features

Can be implemented using syndrome coding

1. Given $\mathbf{x}$, secret message $\mathbf{m} \in \{0,1\}^k$, and parity-check matrix $\mathbf{H} \in \mathbb{R}^{k \times N}$, the embedding algorithm communicates the message as a syndrome while minimazing distortion:

$$\mathbf{y} = \arg \min_{\mathbf{H}\mathbf{y}=\mathbf{m}} D(\mathbf{x}, \mathbf{y})$$

2. With $\mathbf{H}$ syndrome-trellis codes (STCs) [Filler et al. SPIE 2010, TIFS 2011], $D(\mathbf{x}, \mathbf{y})$ is very close to the minimum distortion determined by the corresponding rate–distortion bound

Distortion is not detectability

- Distortion is linked to statistical detectability only heuristically
- We should minimize statistical detectability rather than distortion
- Only possible if we adopt a model of images = hard because
  - Simple models may lead to suboptimal (deceiving) results
  - Complex models difficult to estimate, closed-form solutions unavailable
  - **Idea**: simple model but adapted to each pixel (multiparametric approach)

Generalized Gaussian image model

- Content (local pixel mean) can be estimated using predictors and subtracted

$$\mathbf{r} = (r_1, \ldots, r_N) = \mathbf{x} - F(\mathbf{x})$$

- $r_n \sim \mathcal{P}_{\sigma_n,\nu} = (p_{\sigma_n,\nu}(k))_{k\in\mathbb{Z}}$ independent with $\sigma_n^2 = b_n^2 \frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}$

$$p_{\sigma_n,\nu}(k) = \mathbb{P}(x_n = k) \propto \frac{\nu}{2b_n\Gamma(1/\nu)} \exp\left(-\frac{|k|^\nu}{b_n^\nu}\right)$$

- Notice the zero mean
- ν is the shape parameter (*fixed* over all pixels)
- Variance σ_n^2 contains **both** acquisition noise **and** modeling error (*estimated* for each pixel)

Stego image model

- Mutually independent pentary embedding
- Each pixel is changed by at most ± 2 with probabilities

$$\mathbb{P}(y_n = x_n + 1) = \beta_n \quad \mathbb{P}(y_n = x_n + 2) = \theta_n$$
$$\mathbb{P}(y_n = x_n - 1) = \beta_n \quad \mathbb{P}(y_n = x_n - 2) = \theta_n$$
$$\mathbb{P}(y_n = x_n) = 1 - 2\beta_n - 2\theta_n$$

- Stego residual follows pmf $\mathcal{Q}_{\sigma_n,\nu,\beta_n,\theta_n} = (q_{\sigma_n,\nu,\beta_n,\theta_n}(k))_{k\in\mathbb{Z}}$

$$\begin{aligned}&\mathbb{P}(y_n = k) = q_{\sigma_n,\nu,\beta_n,\theta_n}(k)\\&= (1 - 2\beta_n - 2\theta_n)p_{\sigma_n,\nu}(k) + \beta_n p_{\sigma_n,\nu}(k+1)\\&\quad + \beta_n p_{\sigma_n,\nu}(k-1) + \theta_n p_{\sigma_n,\nu}(k+2) + \theta_n p_{\sigma_n,\nu}(k-2)\end{aligned}$$

Embedding capacity

- Alice can embed a payload of R nats given by

$$R(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{n=1}^{N} H(\beta_n, \theta_n)$$

$H(x, y) = -2x \ln x - 2y \ln y - (1 - 2x - 2y) \ln(1 - 2x - 2y)$ is the pentary entropy function.

- We determine the change rates β_n, θ_n so that they minimize the power of the most powerful detector within the chosen Multivariate Generalized Gaussian (MVGG) model.

Deriving optimal detector

Assumptions (omniscient Warden)

1. Warden and Alice know variances σ_n^2
2. Warden knows change rates β_n and θ_n
3. Fine quantization limit $\sigma_n^2 \gg 1$
4. Large number of pixels $N \to \infty$

Hypothesis testing problem

- Due to our assumptions, we face a simple binary hypothesis test:

$$\begin{array}{ll} \mathcal{H}_0: & x_n \sim \mathcal{P}_{\sigma_n,\nu} \\ \mathcal{H}_1: & x_n \sim \mathcal{Q}_{\sigma_n,\nu,\beta_n,\theta_n} \end{array}$$

- We want a test $\delta : \mathbb{Z}^N \to \{\mathcal{H}_0, \mathcal{H}_1\}$, with the best possible performance.
- Best in the sense of Neyman–Pearson
    - Given the false-alarm probability $\alpha = \mathbb{P}(\delta(\mathbf{x}) = \mathcal{H}_1 | \mathcal{H}_0)$
    - Select δ that maximizes the detection power $\pi = \mathbb{P}(\delta(\mathbf{x}) = \mathcal{H}_1 | \mathcal{H}_1)$

Optimal steganalysis detector

- Log-likelihood ratio

$$\Lambda(\mathbf{x},\boldsymbol{\sigma},\nu)=\sum_{n=1}^{N}\Lambda_n=\sum_{n=1}^{N}\log\left(\frac{q_{\sigma_n,\nu,\beta_n,\theta_n}(x_n)}{p_{\sigma_n,\nu}(x_n)}\right)\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}}\tau$$

- Using our assumptions, the normalized log-LR

$$\Lambda^{\star}(\mathbf{x},\boldsymbol{\sigma},\nu)=\frac{\sum_{n=1}^{N}\Lambda_n-E_{\mathcal{H}_0}[\Lambda_n]}{\sqrt{\sum_{n=1}^{N}Var_{\mathcal{H}_0}[\Lambda_n]}}\overset{(D)}{\rightarrow}\begin{cases}\mathcal{N}(0,1) & \text{under } \mathcal{H}_0\\ \mathcal{N}(\varrho,1) & \text{under } \mathcal{H}_1\end{cases}$$

$$\varrho^2=\sum_{n=1}^{N}(\beta_n,\theta_n)\mathbb{I}_n\begin{pmatrix}\beta_n\\ \theta_n\end{pmatrix}$$

$\mathbb{I}_n$ is the 2×2 Fisher information matrix.

Obtaining the change rates

- β_n and θ_n determined by constrained optimization – minimizing the deflection coefficient ϱ with the payload constraint.
- Method of Lagrange multipliers states that β_n, θ_n, and λ must satisfy

$$\mathbb{I}_n \begin{pmatrix} \beta_n \\ \theta_n \end{pmatrix} = \frac{1}{\lambda} \begin{pmatrix} \ln(1 - 2\beta_n - 2\theta_n)/\beta_n) \\ \ln(1 - 2\beta_n - 2\theta_n)/\theta_n) \end{pmatrix} \; n = 1, \ldots, N$$

$$R = \sum_{n=1}^{N} H(\beta_n, \theta_n)$$

- We solve this using binary search over λ and Newton method parallelized over pixels

Embedding in practice

Alice embeds her payload using STCs while minimizing the distortion

$$D(\mathbf{x},\mathbf{y}) = 2\sum_{n=1}^{N}\left(\rho_n^{(1)}[x_n = y_n \pm 1] + \rho_n^{(2)}[x_n = y_n \pm 2]\right)$$

with costs of changing pixels by ± 1, $\rho_n^{(1)}$, and by ± 2, $\rho_n^{(2)}$, obtained by solving for each n

$$\beta_n = \frac{e^{-\lambda\rho_n^{(1)}}}{1 + 2e^{-\lambda\rho_n^{(1)}} + 2e^{-\lambda\rho_n^{(2)}}}$$

$$\theta_n = \frac{e^{-\lambda\rho_n^{(2)}}}{1 + 2e^{-\lambda\rho_n^{(1)}} + 2e^{-\lambda\rho_n^{(2)}}}$$

Experimental setup

- BOSSbase 1.01 (10,000 grayscale 512×512 images)
- FLD ensemble with
  - SRM (Spatial Rich Model) [Fridrich, TIFS 2011]
  - maxSRMd2 (selection-channel-aware SRM) [Denemark, WIFS 2014]
- Security evaluated using minimal total classification error probability under equal priors averaged over 10 random database splits

$$\overline{P}_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$$

- Separate classifier was trained for each embedding algorithm and payload to see the security across different payloads

Variance estimator

The most accurate estimator of the acquisition noise does not necessarily lead to the most secure steganography!

Stego Object

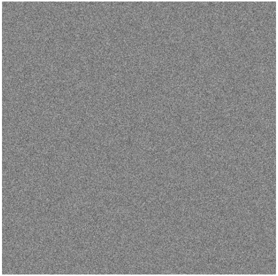
Requirements

- Modular (estimate modelling eror and acquisition noise)
- Fast (we need to embed a large number of images)

Variance estimator

The most accurate estimator of the acquisition noise does not necessarily lead to the most secure steganography!

Stego Object



Acquisition Noise

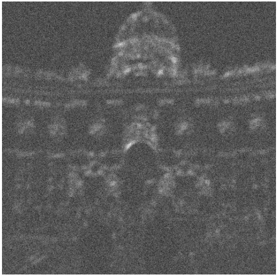
Requirements

- Modular (estimate modelling eror and acquisition noise)
- Fast (we need to embed a large number of images)

Variance estimator

The most accurate estimator of the acquisition noise does not necessarily lead to the most secure steganography!

Stego Object



Acquisition Noise
+
Modelling Error

Requirements

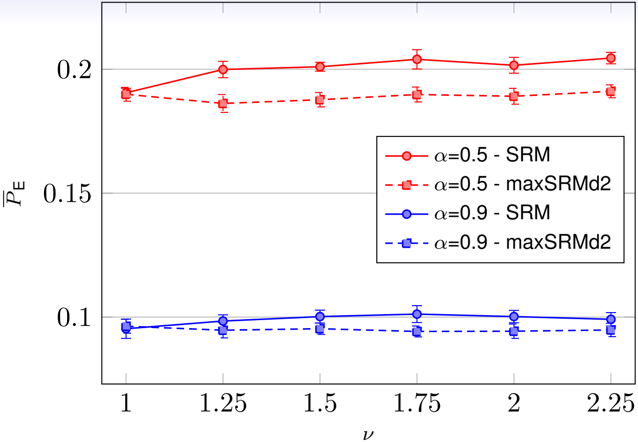
- Modular (estimate modelling eror and acquisition noise)
- Fast (we need to embed a large number of images)

Variance estimator (cont'd)

Design

- Extract noise $\mathbf{r}$ using Wiener filter W: $\mathbf{r} = \mathbf{x} - W(\mathbf{x})$
- Model residual content using pixel-wise linear model $\mathbf{r}_n = \mathbf{G}\mathbf{a}_n + \boldsymbol{\xi}_n$
  - $\mathbf{r}_n \in \mathbb{R}^{B^2}$ vector of residuals at pixel n
  - $\mathbf{G} \in \mathbb{R}^{B^2 \times q}$ modeling matrix (DCT modes)
  - $\mathbf{a}_n \in \mathbb{R}^q$ modeling parameters, $\boldsymbol{\xi}_n \in \mathbb{R}^{B^2}$ noise term
- Standard LSQ fit: $\widehat{\mathbf{a}}_n = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1} \mathbf{G}^{\mathrm{T}}\mathbf{r}_n$ and $\widehat{\mathbf{r}}_n = \mathbf{G}\widehat{\mathbf{a}}_n$
- $\widehat{\sigma}_n^2 = \max\left\{0.01, \frac{\|\mathbf{r}_n - \widehat{\mathbf{r}}_n\|^2}{p^2 - q}\right\}$ for numerical stability

GG shape parameter ν



Average detection error $\overline{P}_E$ of MVGG as a function the shape parameter ν using SRM and maxSRMd2 features for two different payloads

Prior art schemes

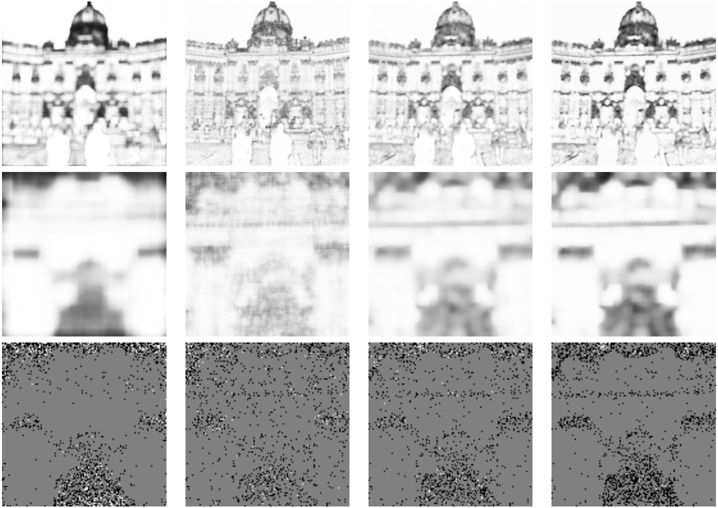
- S-UNIWARD [Holub et al., EURASIP 2013] implemented with stabilizing constant equal to 1
- HILL [Li et al., ICIP 2014] with 3×3 and 15×15 averaging filters
- Pentary versions of S-UNIWARD and HILL implemented with costs

$$\rho_n^{(\pm 2)} = D(\mathbf{x}, x_n \pm 2\, \mathbf{x}_{\sim n})$$

  where D is the distortion of the corresponding embedding algorithm and $x_n \pm 2\, \mathbf{x}_{\sim n}$ denotes the cover image in which only the nth pixel was modified by ± 2

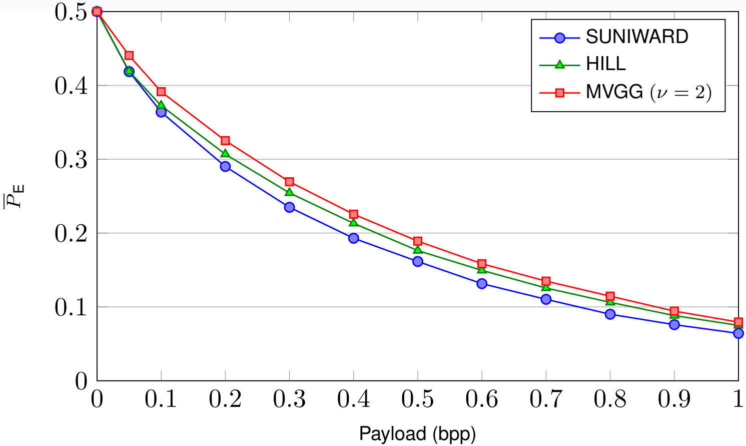
Embedding change probability $2\beta_n + 2\theta_n$

HILL | S-UNIWARD | MVGG ($\nu = 1$) | MVGG ($\nu = 2$)



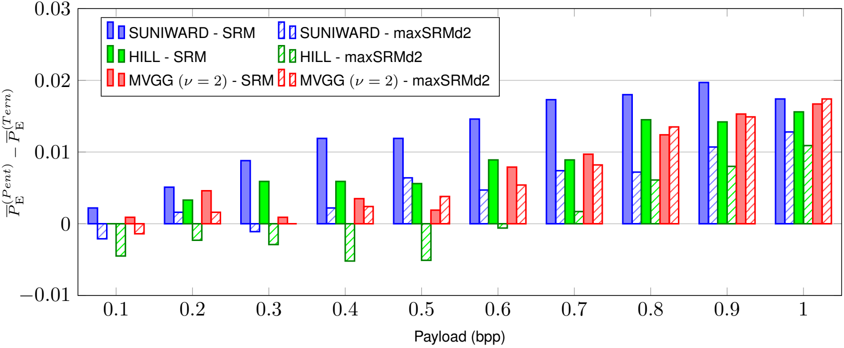
0.0860/0.0180 | 0.0834/0.0100 | 0.0817/0.0082 | 0.1012/0.0043

Comparison to prior art (maxSRMd2)



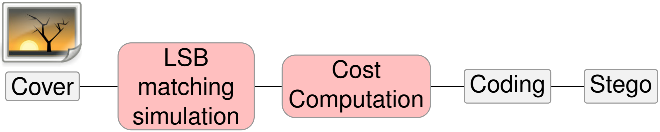
Average detection error $\overline{P}_\mathrm{E}$ for pentary versions of S-UNIWARD, HILL, and MVGG ($\nu = 2$) using maxSRMd2

Pentary vs. ternary

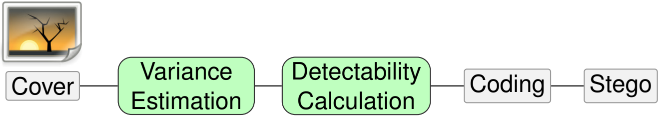


Average difference in detection error $\overline{P}_{\mathrm{E}}$ between pentary and ternary embedding as a function of payload for S-UNIWARD, HILL, and MVGG ($\nu = 2$) using SRM and maxSRMd2 features

Embedding is fundamentally different from prior art



Simplified flowchart of a typical prior-art content-adaptive steganography



Simplified flowchart of the proposed scheme

Summary

- Proposed model based steganography
  - Adapt the model for each pixel of the image
  - State-of-the-art steganalysis is insensitive to the shape parameter of the distribution (Further research in steganalysis)
- Used pentary embedding boosts ternary for large payloads
- Possible extension (and further security boost) to dependent adjacent pixels (jointly Gaussian). Potential problem with estimating the parameters (covariance).

Question