Detection of Content Adaptive LSB Matching (a Game Theory Approach)

Tomáš Denemark, Jessica Fridrich

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Content-adaptive steganography

- Every pixel is changed with probability

$$\beta_i = \frac{\exp(-\lambda\rho_i)}{1+\exp(-\lambda\rho_i)},$$

  where $\rho_i \geq 0$ are costs for each pixel and λ determined from the payload constrain $\frac{1}{n}\sum_{i=1}^{n} h(\beta_i) = \alpha$.
- Costs determined by image content $\implies$ approximately available to Warden who can adjust detector accordingly.
- How does this change Alice's embedding strategy?

Two fundamental approaches

1. **Omnipotent Warden** [Cachin, 1998]
   Warden knows payload and embedding probabilities for each pixel.
   Alice minimizes KL divergence between cover/stego distributions.
2. **Ignorant Warden** [Böhme, 2012]
   Warden knows only the payload.
   Alice can embed suboptimally (not minimize KL-div) to utilize mismatch of Warden's detector.

Our contribution

- We investigate modern steganography (LSBM).
- Warden uses LRT for detection.

Notation

Gaussian density with mean μ and variance σ^2:

$$f(x;\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

$\{-1,0,1\}$-mixture of Gaussian densities with a parameter $0 \le \beta \le 1/2$:

$$f_\beta(x;\sigma^2) = \frac{\beta}{2} f(x;-1,\sigma^2) + (1-\beta) f(x;0,\sigma^2) + \frac{\beta}{2} f(x;1,\sigma^2).$$

Cover Model

We assume cover is a sequence of n independent Gaussians X_i with unequal variances σ_i^2:

$$\mathbf{X} = (X_1, \dots, X_n), \quad X_i \sim \mathrm{N}(0, \sigma_i^2), \quad i = 1, \dots, n.$$

Embedding Method

- Alice uses LSBM with change rates $\beta_i^{(\mathrm{A})}$, $i = 1, \ldots, n$.
- Stego image $\mathbf{Y} = (Y_1, \ldots, Y_n)$,

$$\Pr(Y_i = x_i + s_i) = \begin{cases} \beta_i^{(\mathrm{A})}/2 & \text{for } s_i = -1, \\ 1 - \beta_i^{(\mathrm{A})} & \text{for } s_i = 0, \\ \beta_i^{(\mathrm{A})}/2 & \text{for } s_i = 1. \end{cases}$$

Therefore

$$Y_i \sim f_{\beta_i^{(\mathrm{A})}}(x, \sigma_i^2)$$

- Change rates must satisfy payload constraint

$$\sum_{i=1}^{n} h(\beta_i^{(\mathrm{A})}) = \alpha n$$

Warden's Detector

- Simple binary hypothesis test:

$$\begin{aligned} \mathrm{H}_0 &: \; X_i \sim f(x, 0, \sigma_i^2), \; \forall i, \\ \mathrm{H}_1 &: \; X_i \sim f_{\beta_i^{(\mathrm{W})}}(x, \sigma_i^2), \; \forall i, \end{aligned}$$

$\beta_i^{(\mathrm{W})}$ are change rates **assumed by Warden**

- Warden uses the Likelihood Ratio Test (LRT):

$$T(\mathbf{x}; \boldsymbol{\beta}^{(\mathrm{W})}, \boldsymbol{\sigma}^2) = \prod_{i=1}^{n} \frac{f_{\beta_i^{(\mathrm{W})}}(x_i, \sigma_i^2)}{f(x_i, 0, \sigma_i^2)}$$

$\boldsymbol{\beta}^{(\mathrm{W})} = (\beta_1^{(\mathrm{W})}, \ldots, \beta_n^{(\mathrm{W})})$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_n^2)$

Alice and Warden Play Game

- Players: Alice and Warden
- Strategies: $\boldsymbol{\beta}^{(\mathrm{A})} = (\beta_1^{(\mathrm{A})}, \ldots, \beta_n^{(\mathrm{A})})$ and $\boldsymbol{\beta}^{(\mathrm{W})} = (\beta_1^{(\mathrm{W})}, \ldots, \beta_n^{(\mathrm{W})})$
- Payoff function: total error probability

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \left(\frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}}) \right)$$

- The game solution is in Nash equilibrium.

Two Pixel Model

- Because of the computational and numerical complexity we limit ourselves to covers consisting of two pixels.
- Strategies: $(\beta_1^{(\mathrm{A})}, \beta_2^{(\mathrm{A})})$, $(\beta_1^{(\mathrm{W})}, \beta_2^{(\mathrm{W})})$ are in fact one-dimensional since the second beta is determined from payload.
- [Omnipotent Warden] KL divergence minimal at $(\beta_1^{(\mathrm{A},1)}, \beta_2^{(\mathrm{A},1)})$
- [Ignorant Warden] Nash equilibrium at $(\beta_1^{(\mathrm{A},2)}, \beta_2^{(\mathrm{A},2)})$

Solution

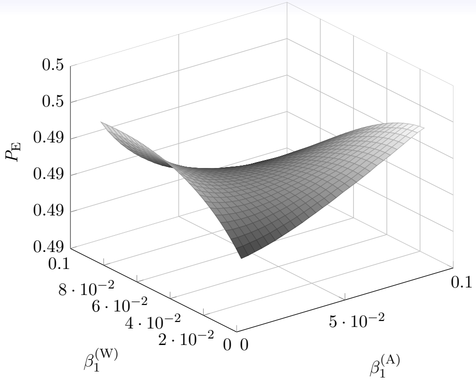

Figure: Payoff function $P_\mathrm{E}(\beta_1^{(\mathrm{A})}, \beta_1^{(\mathrm{W})})$ for $\alpha = 0.2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1.2$.

Smooth, with a unique saddle point ⇒ [Kuhn, 2003] solution exists in pure strategies, in said saddle point.

Results 1

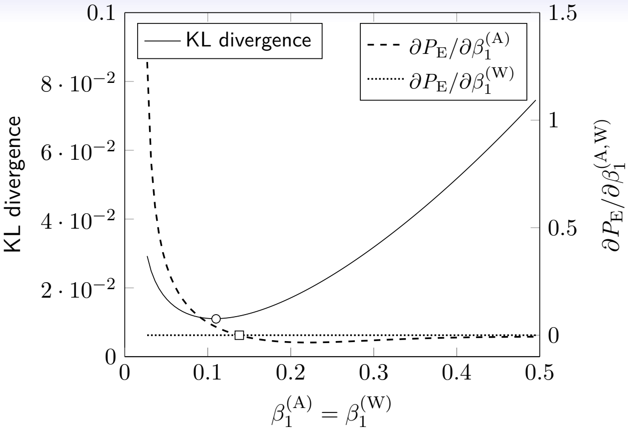

Figure: $\alpha = 0.2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1.2$.

Results 2

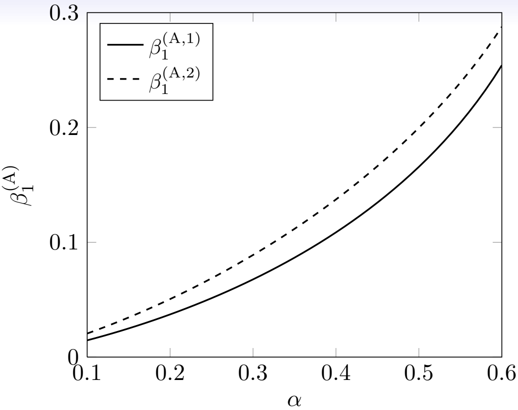

Figure: Alice's strategies under both scenarios $\beta_1^{(A,1)}$, $\beta_1^{(A,2)}$ as a function of α for $\sigma_1^2 = 1$ and $\sigma_2^2 = 1.2$.

Results 3

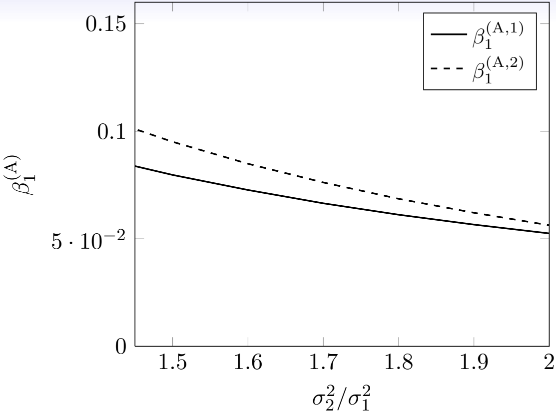

Figure: Alice's strategies under both scenarios $\beta_1^{(A,1)}$, $\beta_1^{(A,2)}$ as a function of content diversity measured by the ratio σ_2^2/σ_1^2 for $\alpha = 0.4$ and $\sigma_1^2 = 1$.

Results 4

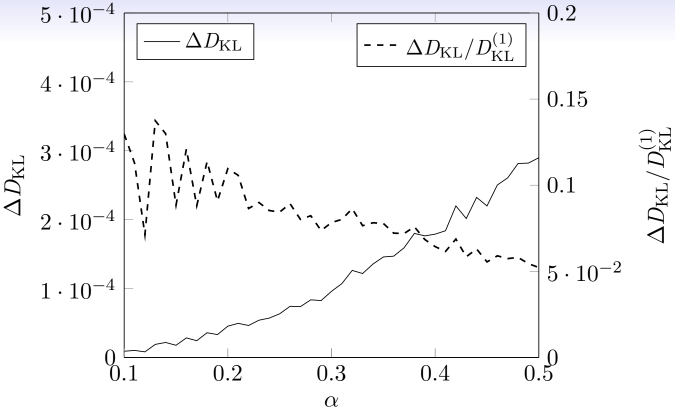

Figure: Warden's loss in her ability to detect Alice's embedding, $\triangle D_{\mathrm{KL}}(\ln T|\mathrm{H}_0||\ln T|\mathrm{H}_1)$ and $\triangle D_{\mathrm{KL}}/D^{(1)}_{\mathrm{KL}}$, as a function of α for $\sigma_1^2 = 1$ and $\sigma_2^2 = 1.2$.

Summary

- In practice Warden rarely has full access to the steganographic channel.
- Even the simplistic two pixel cover source reveals interesting phenomena:
  - Nash equilibrium $\neq$ point of minimal KL divergence.
  - It pays off for Alice to trade optimality for a mismatched detector.
- It is always advantageous for Alice to embed a slightly larger payload into the element with a smaller variance.
  - The difference between optimal strategies increases with increasing α.
  - The difference between optimal strategies decreases with increasing differences between σ_1^2 and σ_2^2.
- Computational complexity and numerical issues prevent scaling up this approach to realistic covers.