Effect of Imprecise Knowledge of Selection Channel on Steganalysis

Vahid Sedighi and Jessica Fridrich

BINGHAMTON
UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Current steganography paradigm

- Content-adaptive steganography
  - Embed in textured/noisy areas that are harder to model and steganalyze

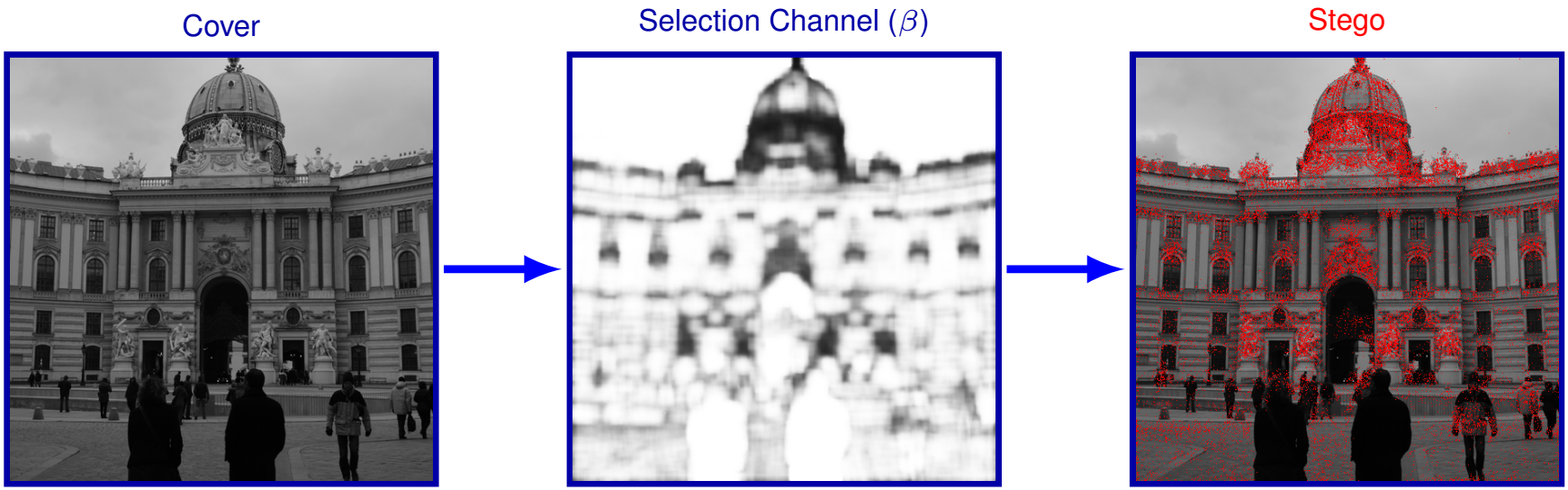
Cover

Current steganography paradigm

- Content-adaptive steganography
  - Embed in textured/noisy areas that are harder to model and steganalyze

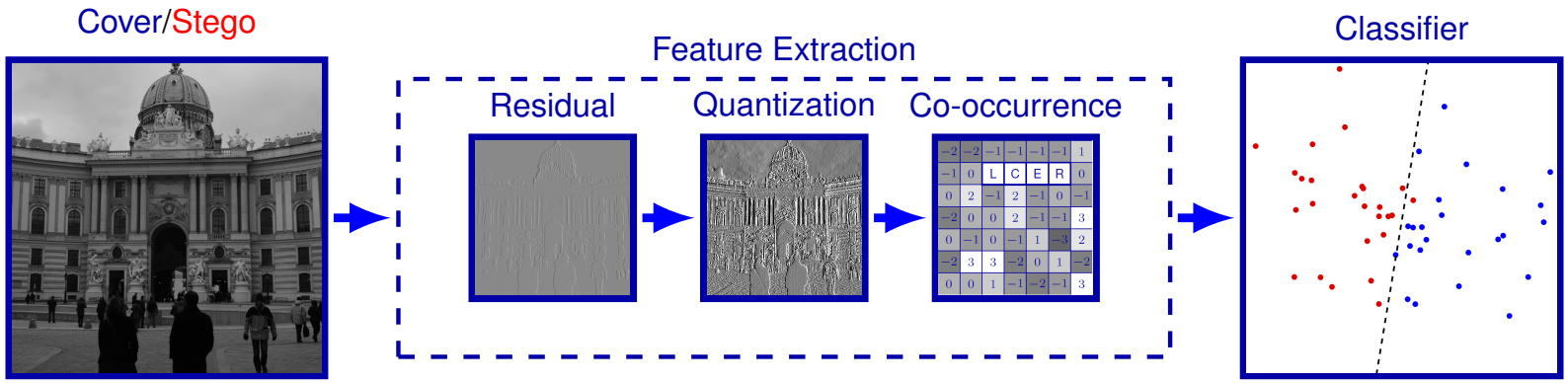
Current steganography paradigm

- Content-adaptive steganography
  - Embed in textured/noisy areas that are harder to model and steganalyze



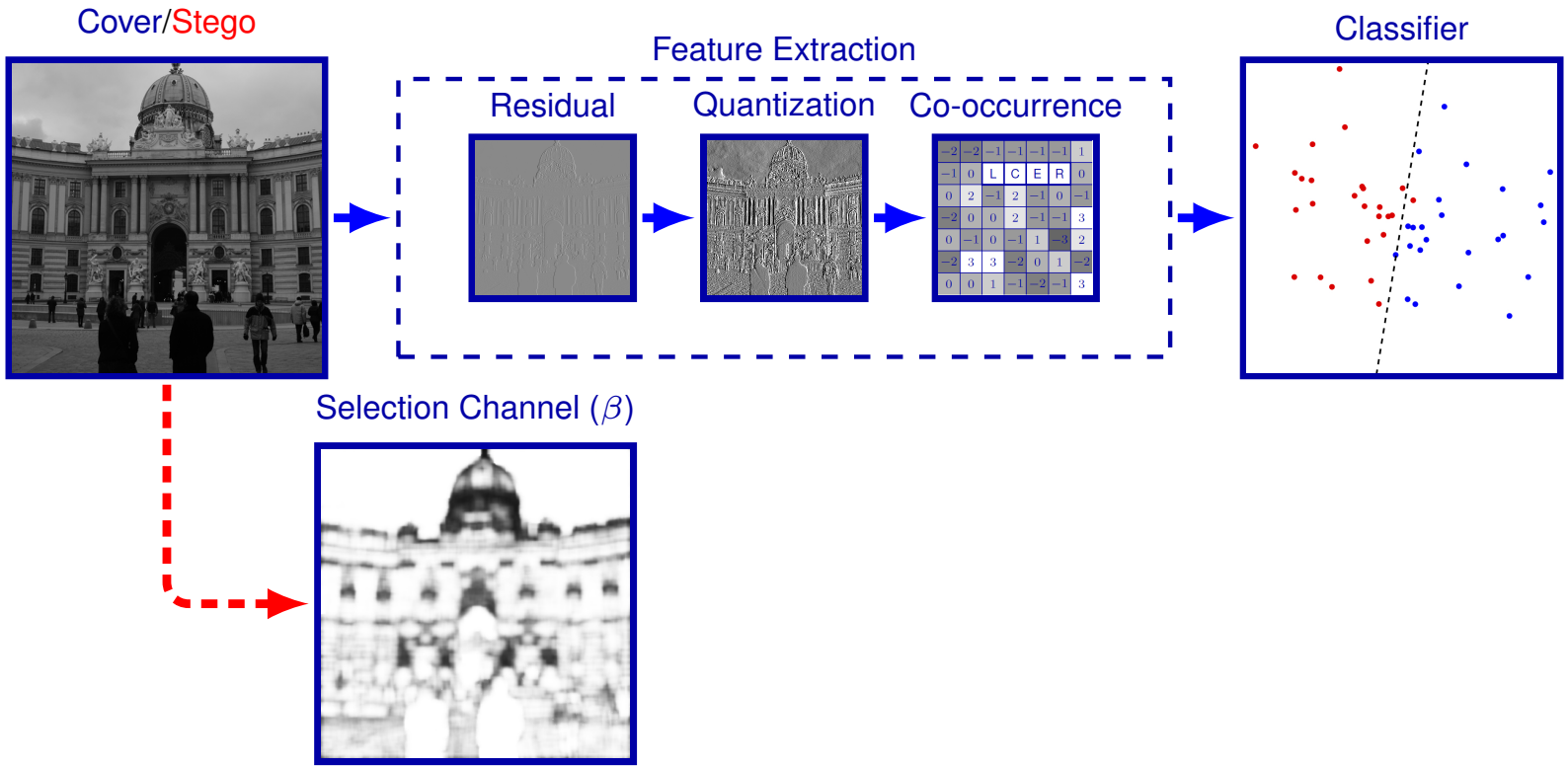
Current steganalysis paradigm

- Images represented with rich media models that use knowledge of the selection channel
- Classifiers trained on examples of cover and stego images



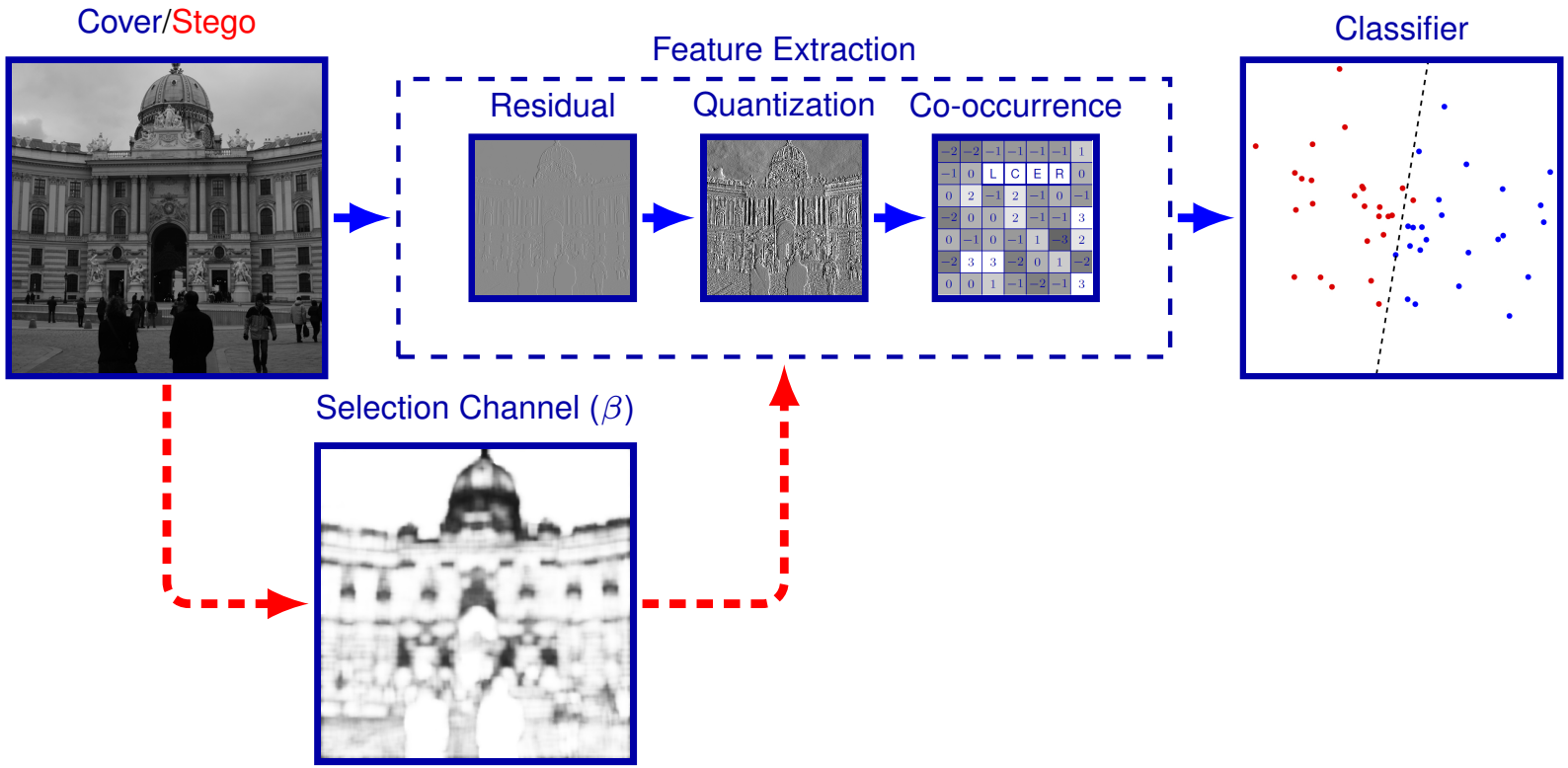
Current steganalysis paradigm

- Images represented with rich media models that use knowledge of the selection channel
- Classifiers trained on examples of cover and stego images



Current steganalysis paradigm

- Images represented with rich media models that use knowledge of the selection channel
- Classifiers trained on examples of cover and stego images



The goal of our study

Features depend on the selection channel (embedding probabilities β_n), which in turn depend on

1. Payload α:

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

2. Embedding itself:

$$\beta_n \text{ from stego } \neq \beta_n \text{ from cover}$$

$\Longrightarrow$ β_n will be known only approximately to the Warden $\Longrightarrow$ potentially negative impact on steganalysis

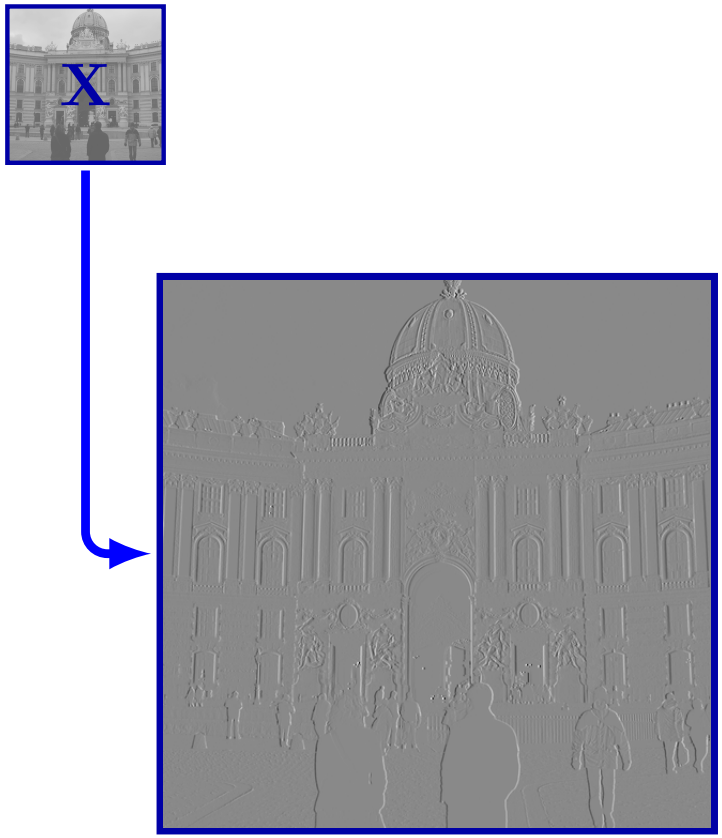
Outline

Outline

Spatial Rich Models (SRM-maxSRM)

Image ($\mathbf{X}$)

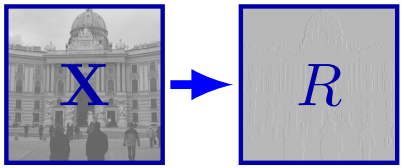
Spatial Rich Models (SRM-maxSRM)

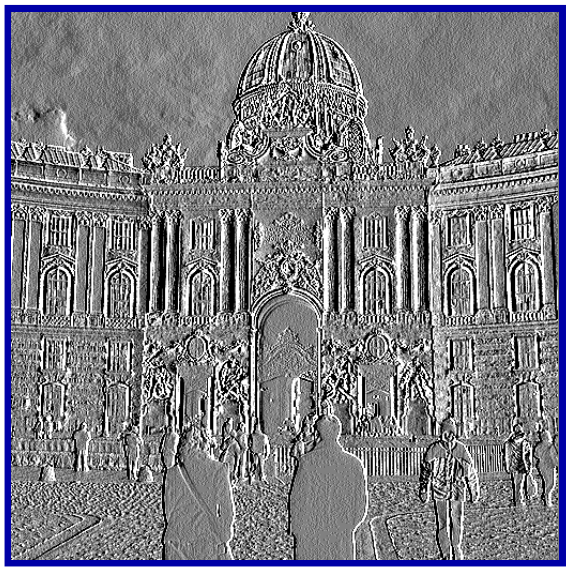


Noise Residual (R)

- $z_{ij} = x_{i,j} - \mathrm{Pred}(\mathcal{N}(x_{ij}))$

Spatial Rich Models (SRM-maxSRM)

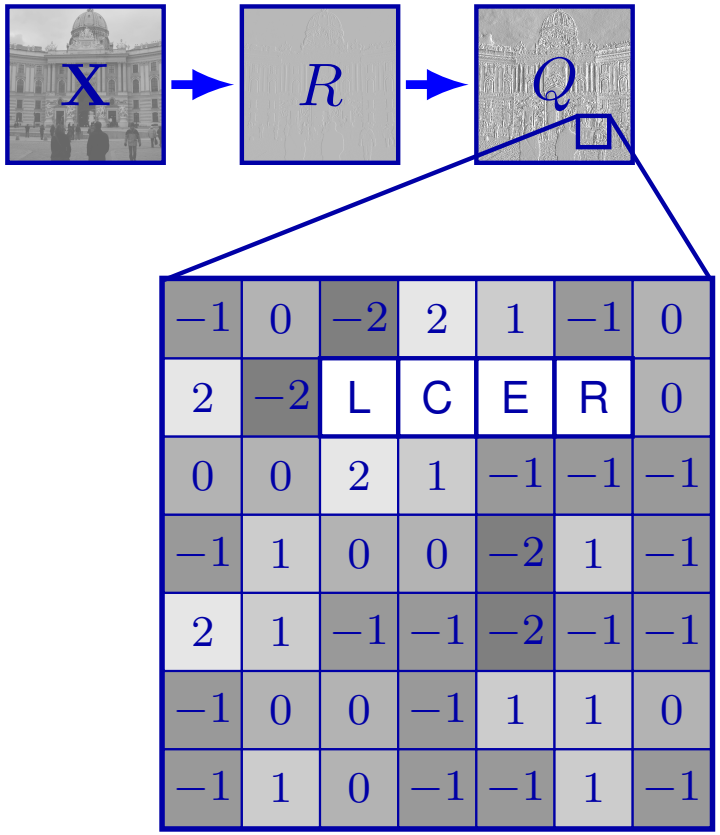




Quantized Residual (Q)

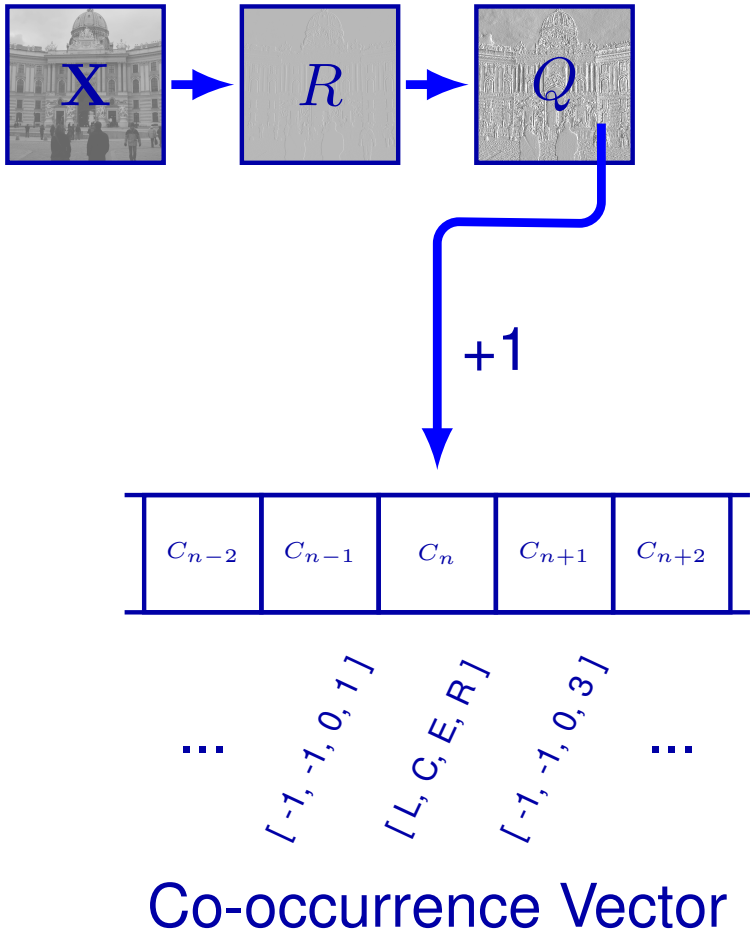
- $z_{ij} \rightarrow r_{ij} = Q_{\mathcal{Q}}(z_{ij})$
- $\mathcal{Q} = \{-Tq, -(T-1)q, \ldots, Tq\}$

Spatial Rich Models (SRM-maxSRM)



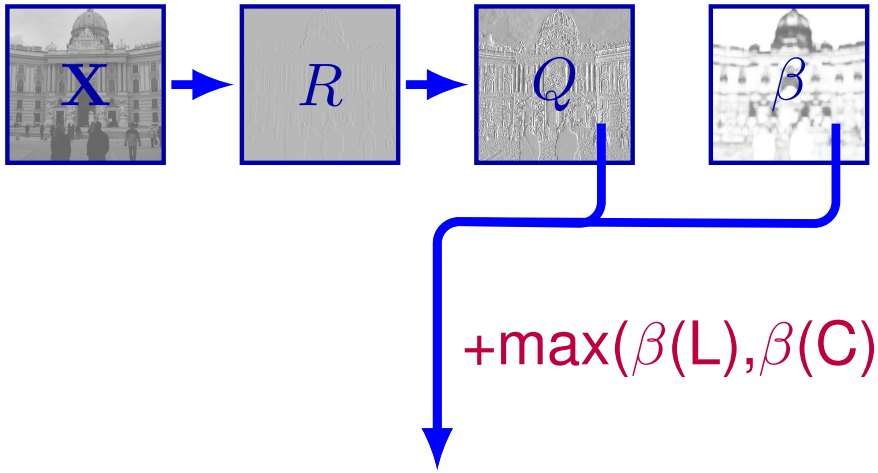
- collect quartets of values
- horizontal and vertical directions

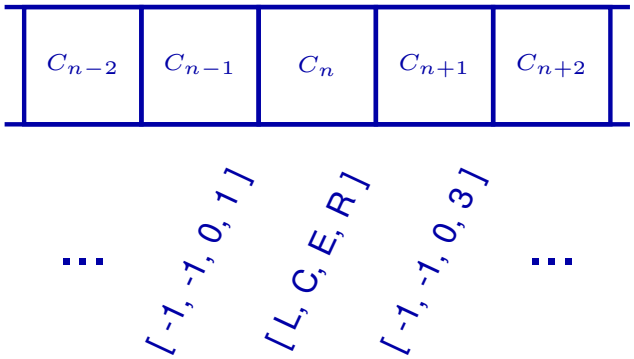
Spatial Rich Models (SRM-maxSRM)



- $4D$ co-occurrence matrix
- symmetrization

Spatial Rich Models (SRM-maxSRM)





Co-occurrence Vector

- $4D$ co-occurrence matrix
- symmetrization
- utilize embedding probabilities

Outline

Multivariate Gaussian image model

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

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

- $r_n \sim \mathcal{N}\left(0, \sigma_n^2\right) = \left(p_{\sigma_n}(k)\right)_{k \in \mathbb{Z}}$ independent with

$$p_{\sigma_n}(k) = \mathbb{P}(r_n = k) \propto \left(2\pi\sigma_n^2\right)^{-1/2} \exp\left(-k^2/(2\sigma_n^2)\right)$$

- Variance σ_n^2 contains **both** acquisition noise **and** modeling error (*estimated* for each pixel)

Stego image model

- Mutually independent ternary embedding (LSB matching)
- Each pixel is changed by at most ± 1 with probabilities

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

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

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

Assumptions for deriving optimal detector

1. Warden and Alice know variances σ_n^2
2. Warden uses changed rates $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)$ that might or might not coincide with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)$
3. Fine quantization limit $\sigma_n^2 \gg 1$
4. Large number of pixels $N \to \infty$

LRT

- Warden faces a simple binary hypothesis test:

$$\begin{aligned} \mathcal{H}_0 : &\quad x_n \sim \mathcal{P}_{\sigma_n} \\ \mathcal{H}_1 : &\quad x_n \sim \mathcal{Q}_{\sigma_n,\gamma_n} \end{aligned}$$

- Asymptotic form of normalized LRT:

$$\Lambda^\star(\mathbf{x},\boldsymbol{\sigma}) = \sum_{n=1}^{N} \log\left(\frac{q_{\sigma_n,\beta_n}(x_n)}{p_{\sigma_n}(x_n)}\right) \overset{(D)}{\to} \begin{cases} \mathcal{N}(0,1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\varrho,1) & \text{under } \mathcal{H}_1 \end{cases}$$

$$\varrho = \frac{\sqrt{2}\sum_{n=1}^{N}\beta_n\gamma_n\sigma_n^{-4}}{\sqrt{\sum_{n=1}^{N}\gamma_n^2\sigma_n^{-4}}} \text{ (deflection coefficient)}$$

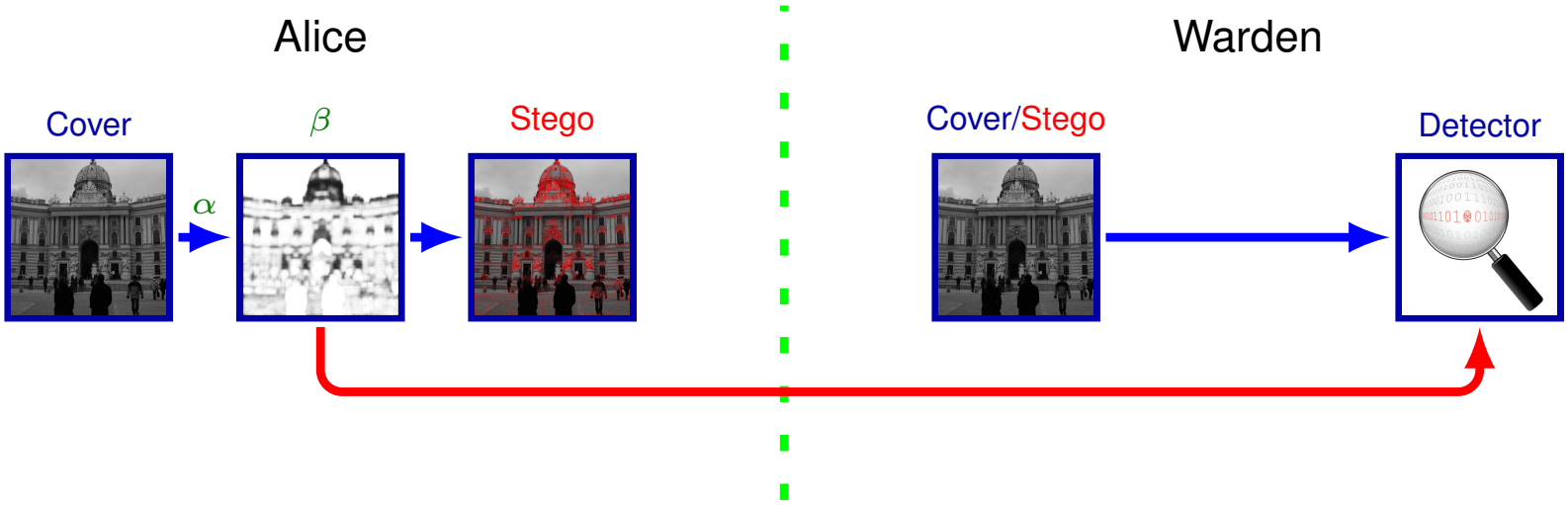
- For each image detection is completely described by its deflection coefficient:

$$\text{test power: } \pi(\alpha) = Q(Q^{-1}(\alpha) - \varrho)$$

Outline

Omniscient Warden (Omni)

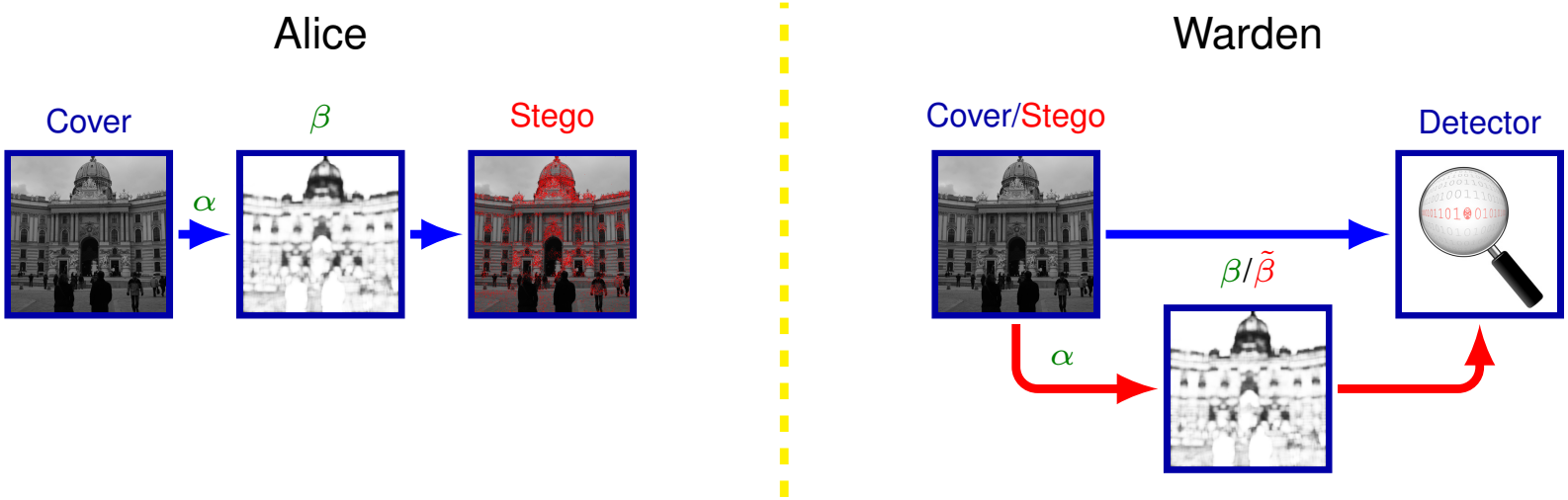
- Knows the exact actions of the sender executed during embedding.



- Empirical detector computes β_n from cover image assuming true payload size α
- LRT detector computes β_n from cover image ($\gamma_n = \beta_n$ for both cover and stego images)

Payload-Informed Warden (PI)

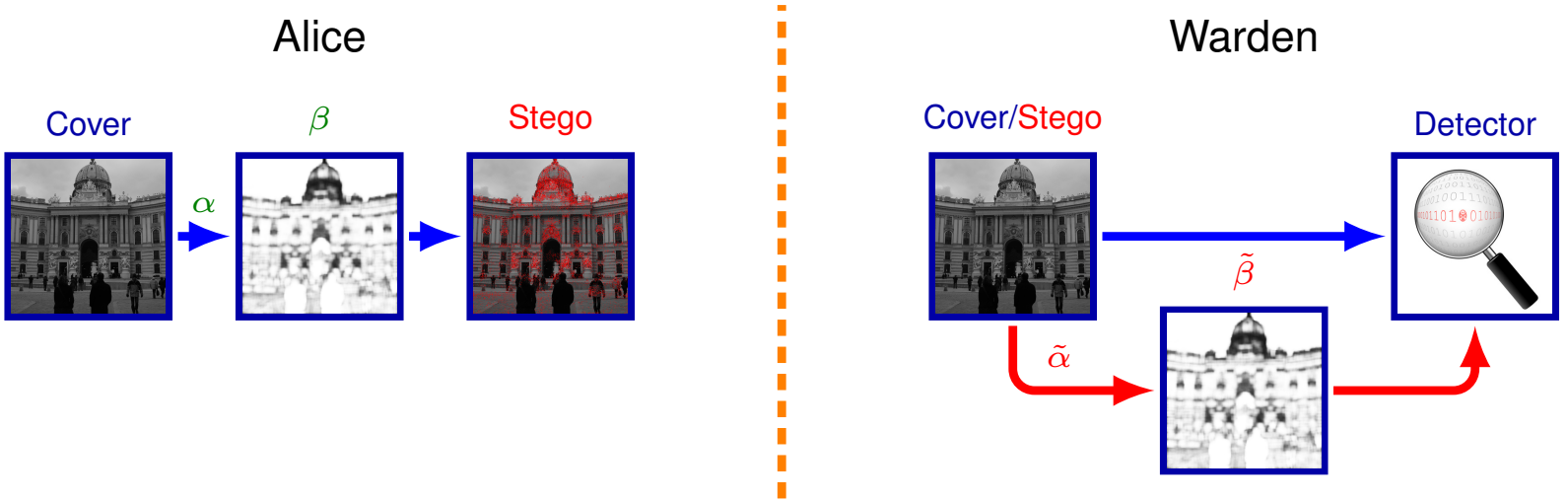
- Knows the size of the embedded payload α but has no access to cover image.



- Empirical detector computes β_n from the available image assuming true payload size α.
- LRT detector computes β_n from the available image ($\gamma_n = \beta_n$ only for cover images)

Fixed-Payload Warden (FP)

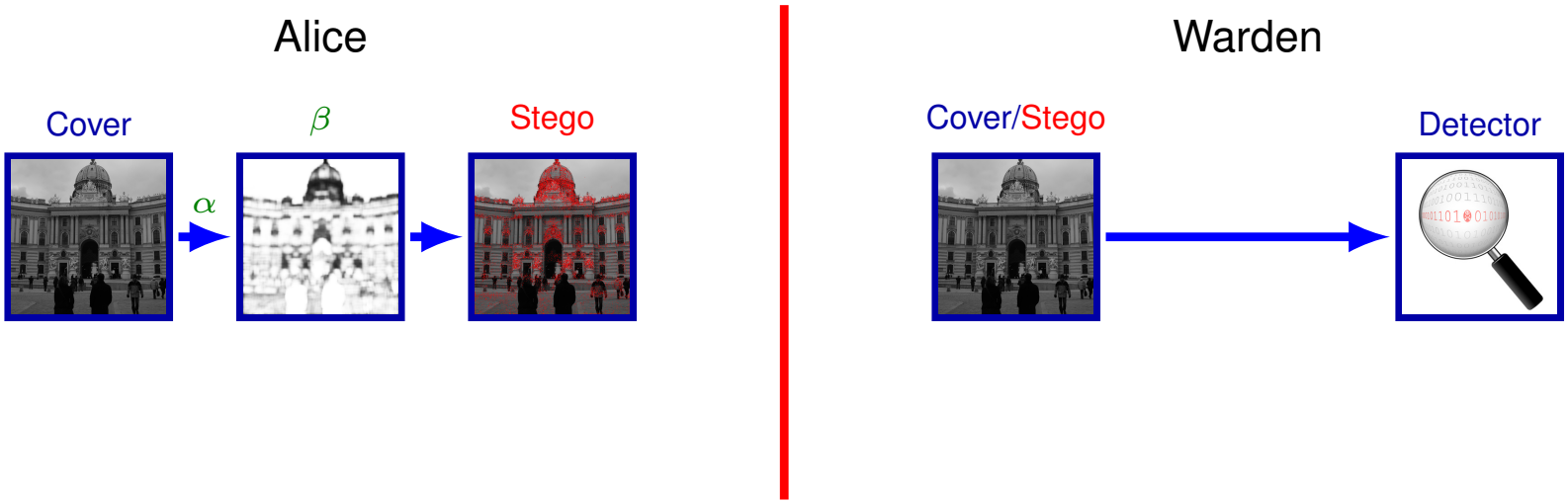
- Does not know payload size α, no access to cover image.



- Empirical detector computes β_n from the available image assuming a fixed payload size $\tilde{\alpha} \neq \alpha$.
- LRT detector computes β_n from the available image ($\gamma_n \neq \beta_n$ for both cover and stego images)

Indifferent Warden (Indif)

- Assumes no adaptive embedding.

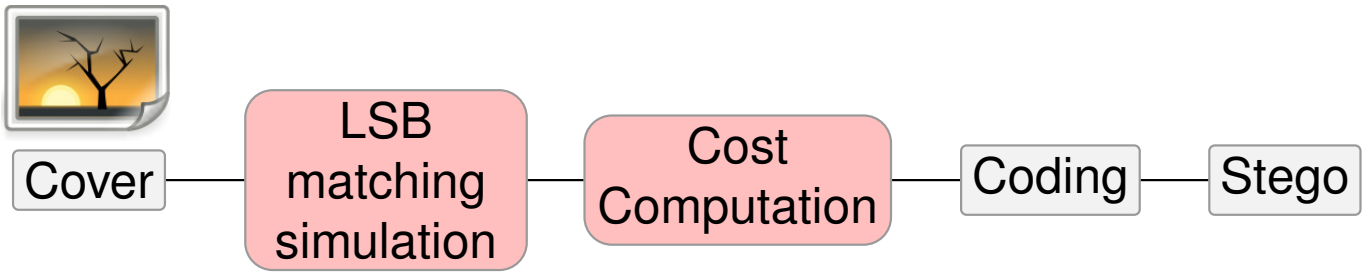


- Empirical detector uses SRM features
- LRT detector uses $\gamma_n = \gamma$

Outline

Cost-based Schemes

- WOW [Holub et al., WIFS 2012]
- SUNIWARD [Holub et al., IH 2013]
- HILL [Li et al., ICIP 2014]



- Alice embeds payload using STCs while minimizing embedding distortion

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

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

Model-based Schemes

- MVG [Sedighi et al., SPIE 2015]

- β_n determined by minimizing the deflection coefficient ϱ with payload constraint.
  - Method of Lagrange multipliers $\Longrightarrow$ β_n and λ must satisfy

$$\beta_n = \frac{1}{\lambda I_n} \ln \frac{1-2\beta_n}{\beta_n}, \; n = 1, \ldots, N$$

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

Outline

General setup

- BOSSbase 1.01 ($10,000$ grayscale 512×512 images)
- Two sender types
  - Payload Limited Sender (PLS): always embeds a message with a fixed relative length. The used payloads are small (0.05 bpp), medium (0.2 bpp), and large (0.5 bpp).
  - Random Payload Sender (RPS): payload size is chosen uniformly randomly from $[0.05, 0.5]$ bpp
- All embedding schems are simulated at their corresponding rate-distortion bounds

Setup of empirical detectors

- FLD ensemble [Kodovsky et al., TIFS 2012] with
  - SRM (Spatial Rich Model) [Fridrich et al., TIFS 2011]
  - maxSRM (selection-channel-aware SRM) [Denemark et al., WIFS 2014]
- Security evaluated using minimal total classification error under equal priors averaged over 10 random $5000/5000$ database splits:

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

Setup of the LRT

- Detects stenography in each individual image
- For each false alarm α, power is averaged over 10 random $5000/5000$ database splits:

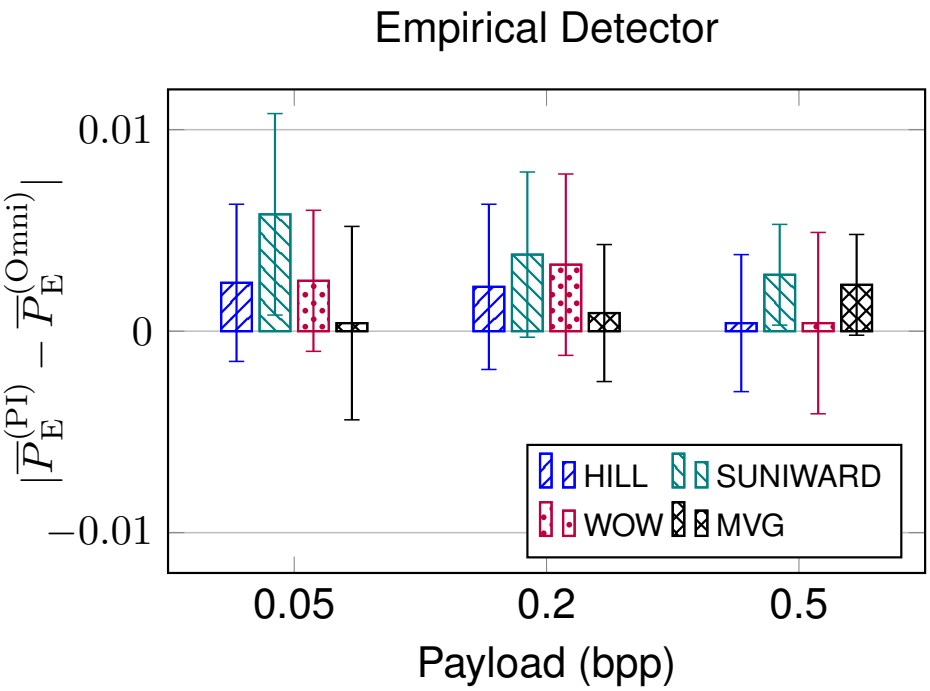
$$\overline{\pi}(\alpha) = \frac{1}{5000} \sum_{n=1}^{5000} \pi^{(n)}(\alpha)$$

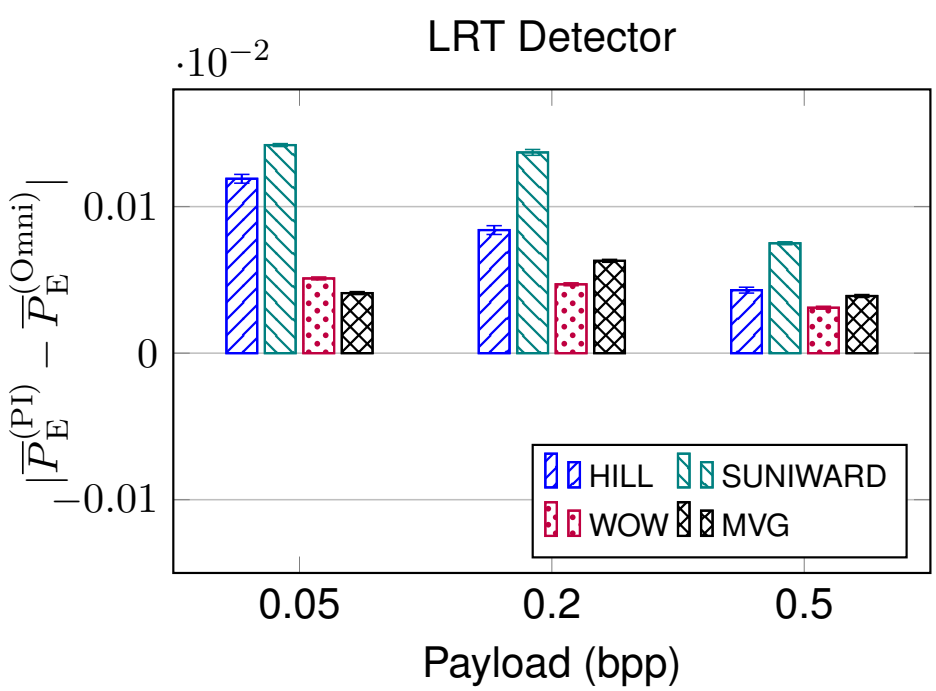
- Security measured again using P_{E}:

$$P_{\mathrm{E}} = \min_{0 \leq \alpha \leq 1} \tfrac{1}{2}(1 - \overline{\pi}(\alpha) + \alpha)$$

Outline

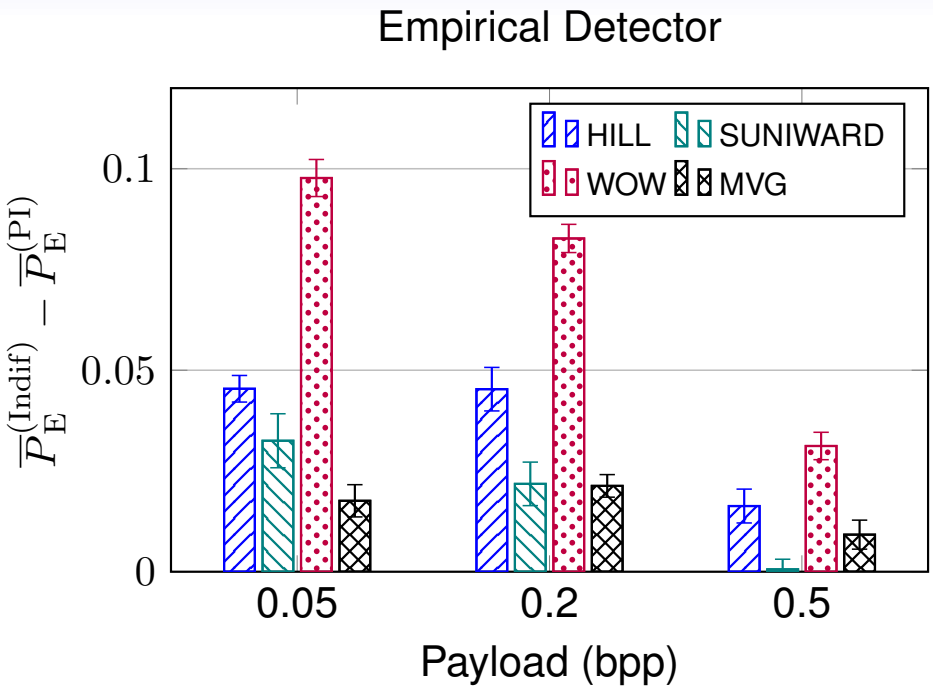
PLS - Omniscient vs Payload-Informed

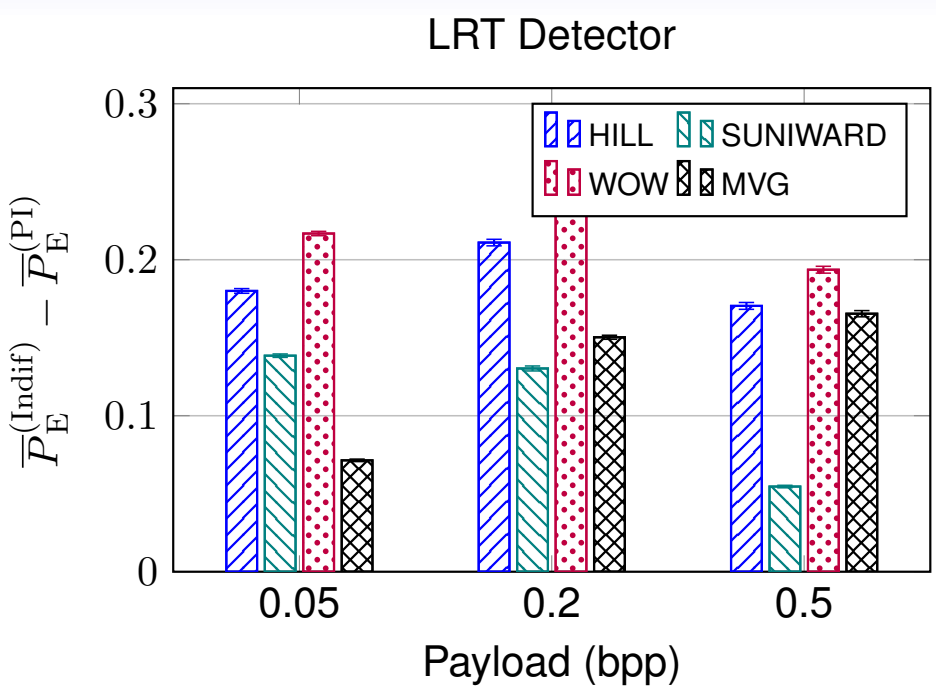




- Impact of "β_n from stego $\neq$ β_n from cover" is negligible
- For empirical detector, detection loss is within statistical spread

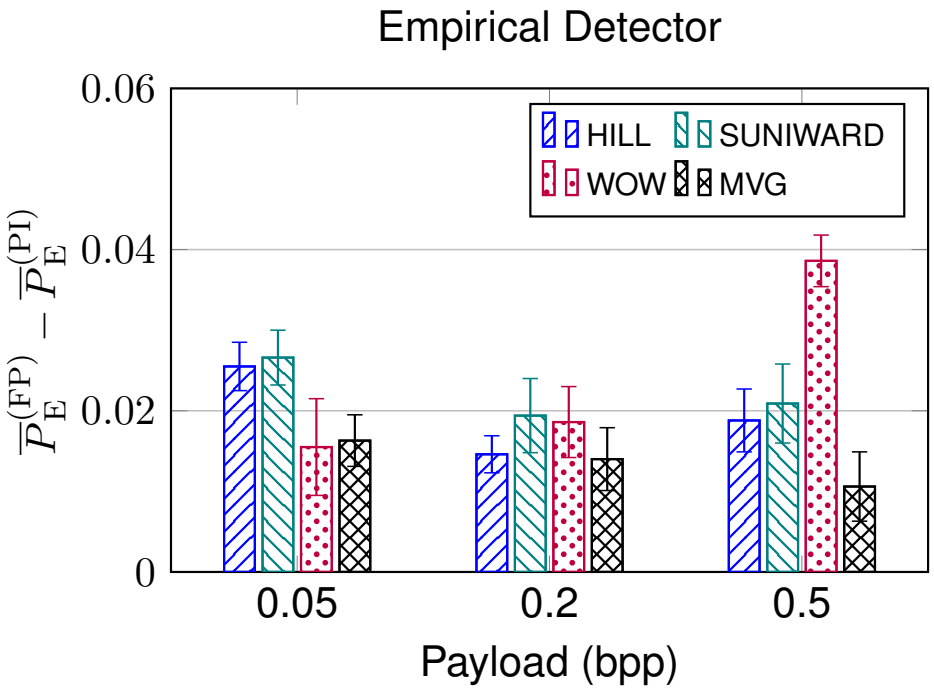
PLS - Payload-Informed vs Indifferent

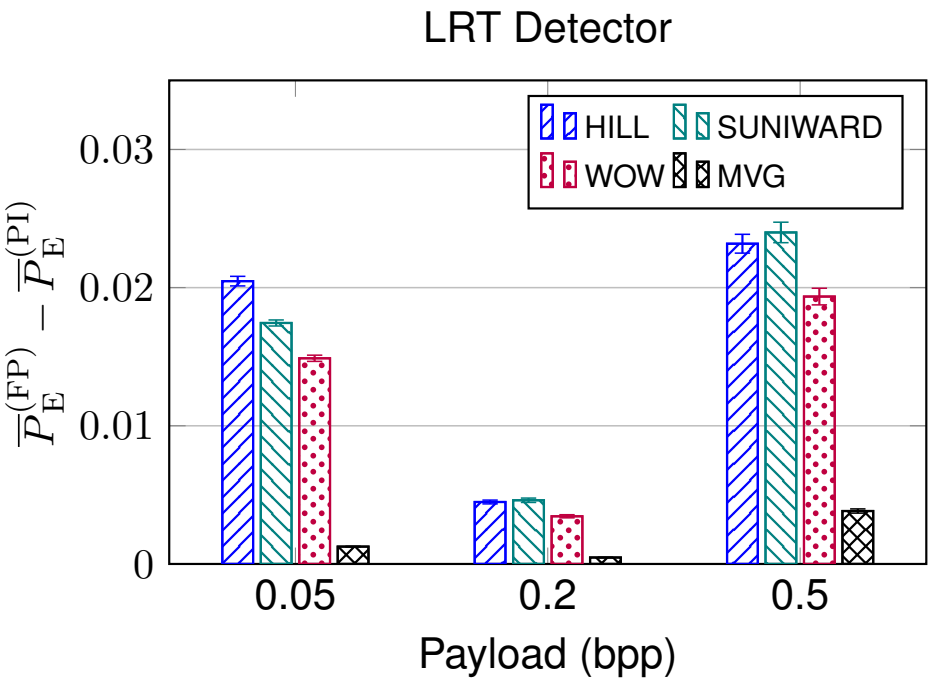




- Using the selection channel knowledge $\Longrightarrow$ substantial detection gain
- Relative comparison of embedding schemes is approximately preserved between detectors

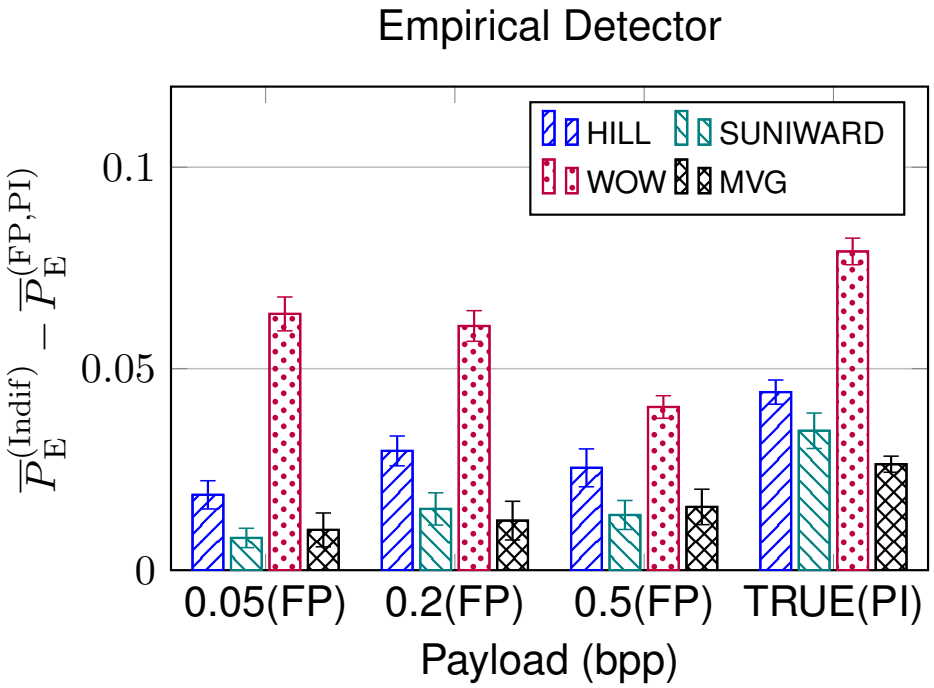
RPS - Payload-Informed vs Fixed-Payload

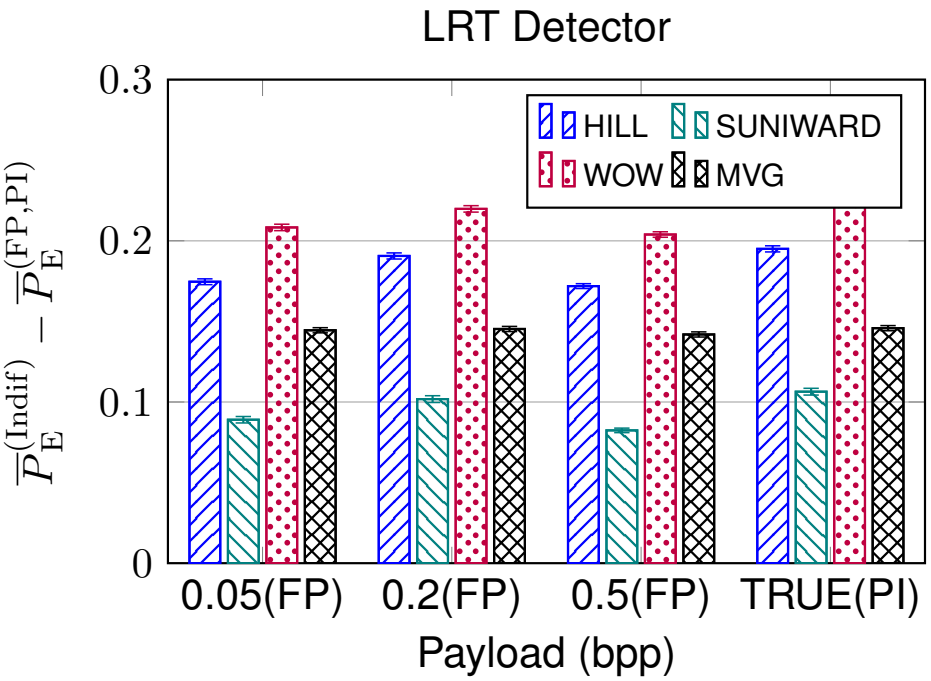




- Loss of detection accuracy when payload size not known
- Both detectors indicate that using medium fixed payload leads to smallest overall loss

RPS - Fixed-Payload & Payload-Informed vs Indifferent





- Gain in detection power by using partial/full knowledge of the selection channel vs. not using it at all
- Imprecise knowledge of selection channel better than not using it at all

Outline

Summary

- We study the effect of using inaccurate knowledge of embedding change probabilities on steganalysis
- Two completely different detectors and four Wardens with different levels of knowledge about selection channel
  - Both detectors exhibit qualitatively the same behavior
  - Loss of detection due to imprecise knowledge of selection channel is small
  - Impact of "β_n from stego $\neq \beta_n$ from cover" is negligible
- It is better to use imprecise embedding change probabilities than none!

Questions

Matlab code available from `http:\\dde.binghamton.edu\download`