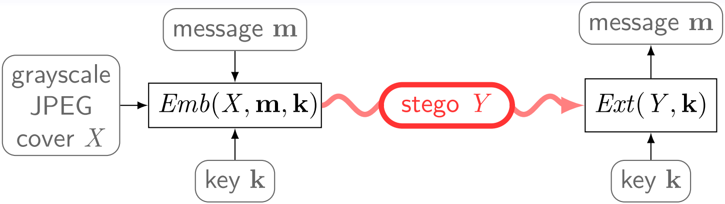
Phase-Aware Projection Model for Steganalysis of JPEG Images

Vojtěch Holub and Jessica Fridrich

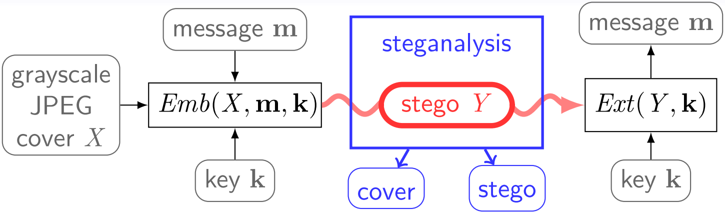
BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

JPEG steganography / steganalysis



- **JPEG Steganography**: Modify certain DCT coefficients of the image by ± 1 to communicate the message.

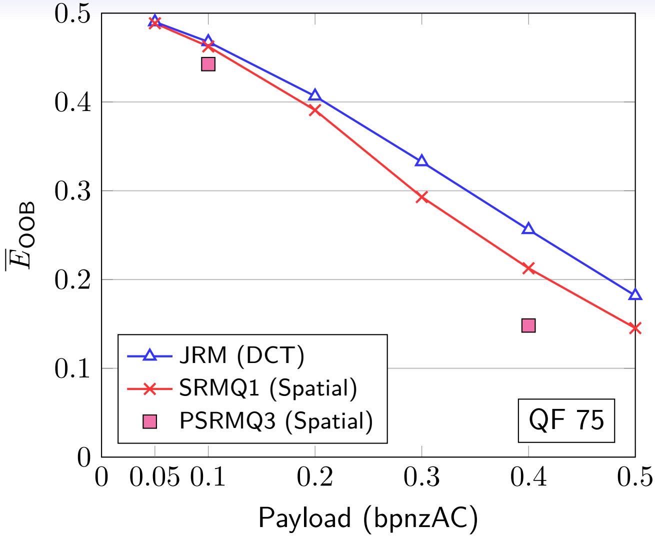
JPEG steganography / steganalysis



- **JPEG Steganography**: Modify certain DCT coefficients of the image by ± 1 to communicate the message.

- **Steganalysis**: Distinguish between cover and stego images by building a detector. If cover source is known and the steganographic scheme is not faulty, the best detection is achieved using feature-based steganalysis and machine learning.

Motivation – J-UNIWARD [Holub, 2014]



JPEG vs. spatial domain steganalysis

- Spatial domain steganalysis:
  - Analyzes dependencies among noise residuals.
  - Adjacent noise residuals put into a 4D co-occurrence $\implies$ treated as a stationary signal.
- JPEG domain steganalysis:
  - Analyzes dependencies among quantized DCT coefficients.
  - DCT coefficients extracted from 8×8 blocks, each mode uses a different DCT base and is quantized differently $\implies$ non-stationarity.

JPEG vs. spatial domain steganalysis

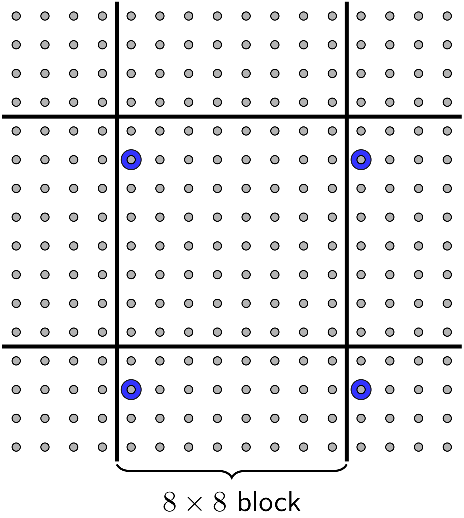
- Spatial domain steganalysis:
  - Analyzes dependencies among noise residuals.
  - Adjacent noise residuals put into a 4D co-occurrence $\Longrightarrow$ treated as a stationary signal.
- JPEG domain steganalysis:
  - Analyzes dependencies among quantized DCT coefficients.
  - DCT coefficients extracted from 8×8 blocks, each mode uses a different DCT base and is quantized differently $\Longrightarrow$ non-stationarity.

Fact

*After JPEG is **decompressed** into spatial domain, image pixels are **non-stationary**.*

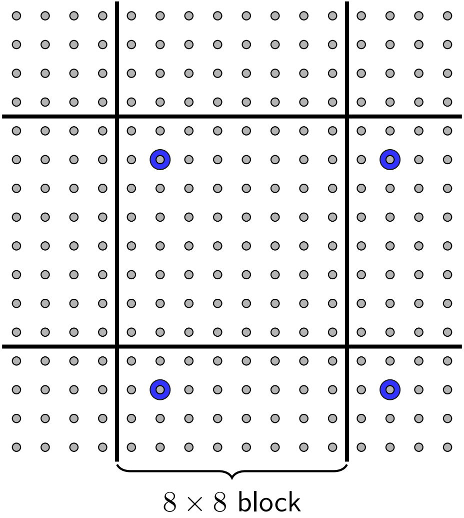
Which coefficients have the same statistics

JPEG (DCT) domain

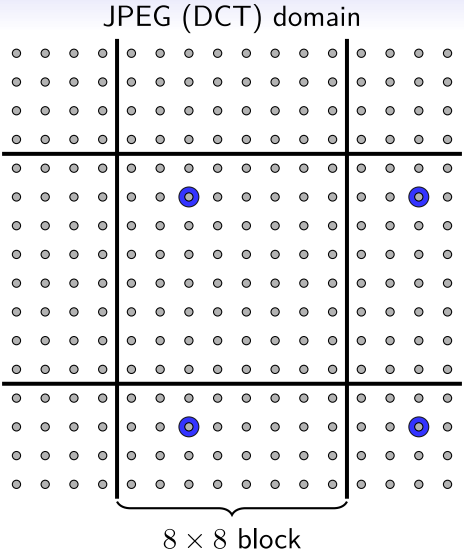


Which coefficients have the same statistics

JPEG (DCT) domain



Which coefficients have the same statistics



Projection Spatial Rich Model [Holub, 2013]

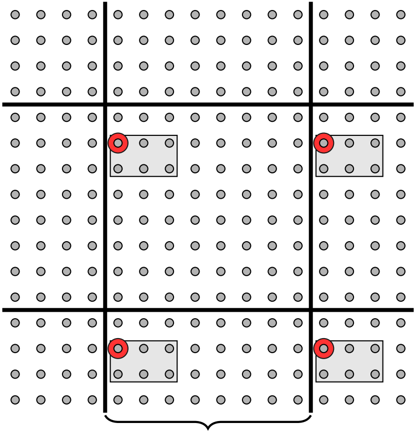
- Originally designed for spatial domain steganalysis.
- First extract multiple residuals using 39 different linear and non-linear (min-max) filters.
- Residuals are convolved with normalized random projection kernels $\Pi \in \mathbb{R}^{s_1 \times s_2}$, $s_1, s_2 \in \{1, \ldots, 8\}$.
- A histogram is built from the projection values for each residual and projection kernel.

The histogram is built from all projection values (all locations) $\Longrightarrow$ implicit assumption that residuals at all locations have identical statistics.

Can it be improved?

Phase-aware projections

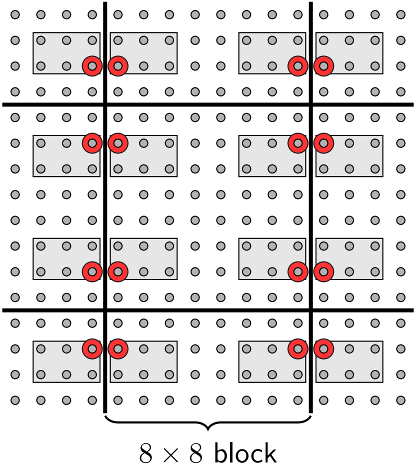
Residual domain of decompressed JPEG



8×8 block

Phase-aware projections

Residual domain of decompressed JPEG



Histogram built from absolute values of projections – symmetries.

Results on linear and minmax residuals

- Detection error $\overline{E}_{\text{OOB}}$
- J-UNIWARD at 0.4 bpnzac, BOSSbase 1.01, QF 75
- Two of PSRMQ3 submodels (linear and non-linear)

Feature type	'spam14h' & 'spam14v'		'minmax41'	
	$\nu = 110$ dim 660	$\nu = 1000$ dim 6000	$\nu = 110$ dim 660	$\nu = 1000$ dim 6000
Standard	0.2587	0.2034	0.3054	0.2257
Phase-aware	0.2576	0.1536	0.3323	0.2421
Phase-aware symmetrized	0.2292	0.1582	0.3292	0.2409

PHARM features

- Merger of 7 SPAM residuals (7 linear filters)

$$\begin{pmatrix} -1 & 1 \end{pmatrix} \quad \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 1 & -3 & 3 & -1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ -3 \\ 3 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \quad \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

- These 7 filters were obtained by a forward feature-selection algorithm using the $\overline{E}_{\mathrm{OOB}}$ estimate of the detection error from 25 prediction kernels.
- All PHARM parameters were optimized with respect to detection of J-UNIWARD
  - ν - number of random projections per residual
  - s - maximal size of the random projection matrix
  - T - number of histogram bins
  - q - quantization (width of histogram bins) – depends on JPEG quality factor

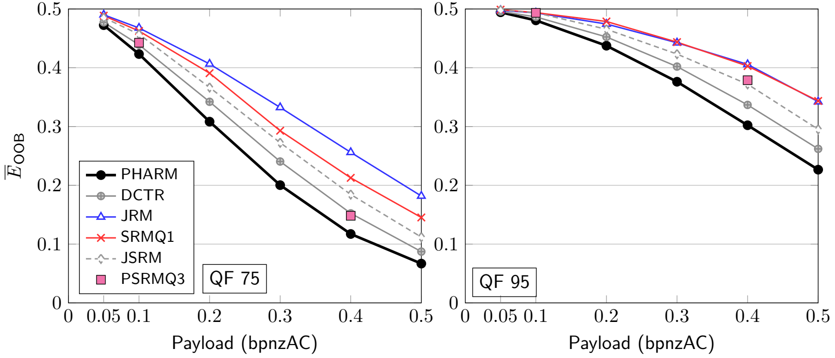
PHARM in numbers

- Total dimensionality: $7 \cdot T \cdot \nu = 7 \cdot 2 \cdot 900 = 12,600$ (dimensionality of PSRMQ3 is $12,870$)
- Quantization $q = \frac{65}{4} - \frac{3}{20} QF$ (QF 75: $q = 5$, QF 95: $q = 2$)
- Extraction time of 512×512 grayscale image, Intel i7 2 GHz laptop:

Feature set	PHARM	DCTR	JRM	SRMQ1	PSRMQ3
Dimensionality	12,600	8,000	22,510	12,753	12,870
Extraction time (s)	4.2	0.6	4.5	1.3	640

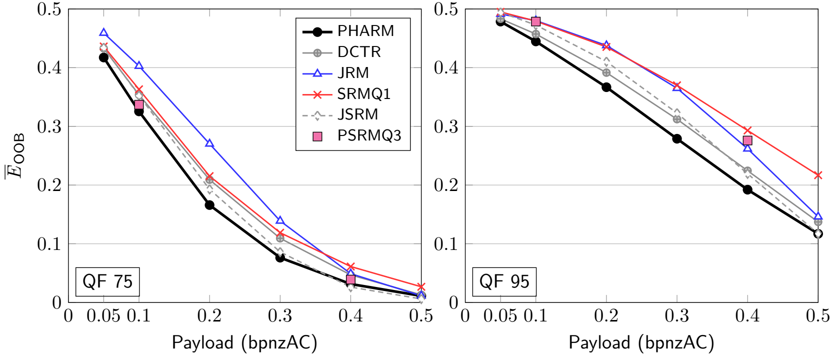
PHARM vs. JPEG steganography

J-UNIWARD [Holub, 2013]



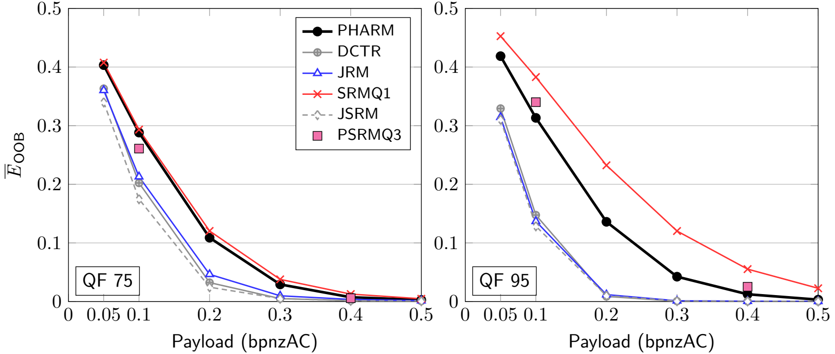
PHARM vs. JPEG steganography

UED [Guo,2014]



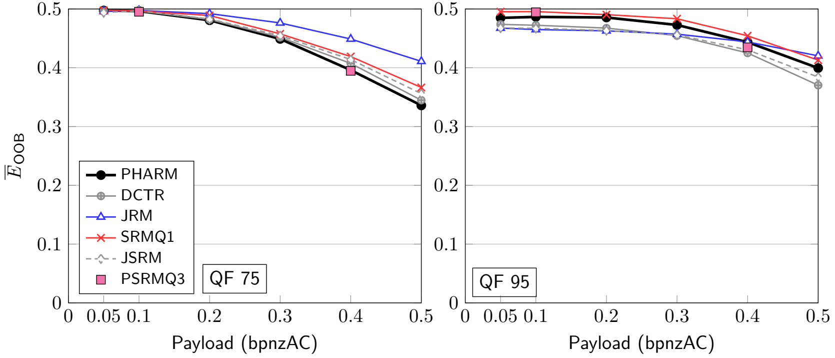
PHARM vs. JPEG steganography

nsF5 [Westfeld, 2001, Fridrich ,2007]



PHARM vs. JPEG steganography

SI-UNIWARD [Holub, 2013]



Conclusion

- Currently, the most reliable detection of modern JPEG stego schemes (J-UNIWARD, UED) is achieved by spatial domain steganalysis (PSRMQ3) – counterintuitive.
- Utilizing the knowledge of properties of decompressed JPEGs can further improve the detection.
- General approach using 'phase-aware' features is proposed.
- Its validity tested by building PHARM feature set based on the Projection Spatial Rich Model (PSRM).
- PHARM achieves superior detection of J-UNIWARD and UED with greatly reduced computational complexity over PSRM.
- Source code in Matlab and C++/MEX available at http://dde.binghamton.edu/download/feature_extractors/