STEGANOGRAPHY WITH TWO JPEGS OF THE SAME SCENE

BINGHAMTON UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Tomáš Denemark and Jessica Fridrich, Binghamton University

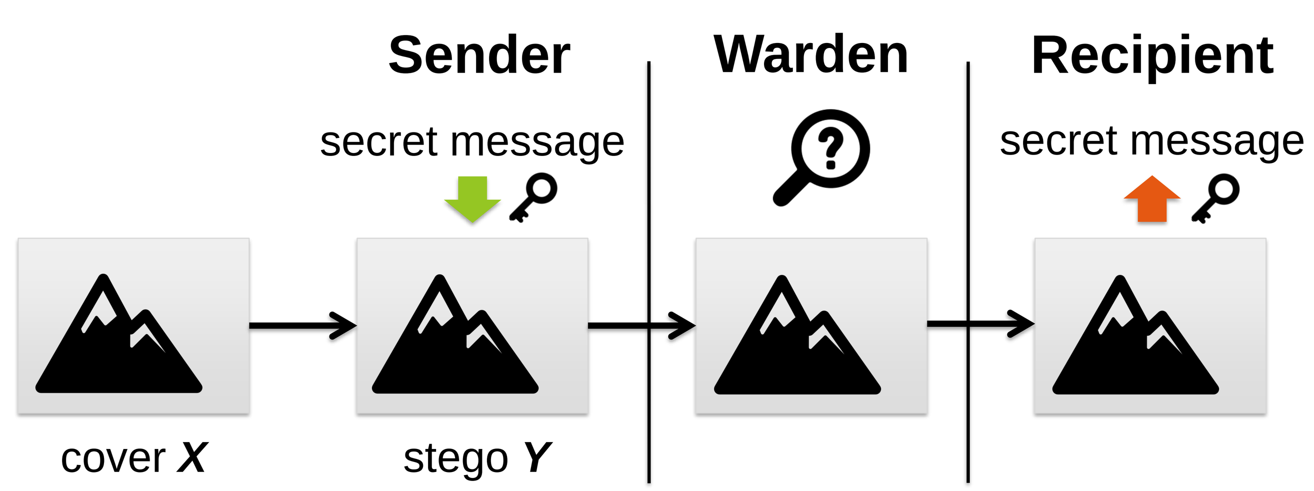
All source code available at **dde.binghamton.edu**

Abstract

Steganography is a private communication method in which secrets are hidden in innocuous objects, such as digital images. We developed a method in which the sender takes *two* JPEG pictures of the same scene, hides the message in one of them while using the second exposure as *side-information*. The differences between the two JPEG files inform the sender about which DCT coefficients are most sensitive to acquisition noise. The proposed steganography favors such changes to obtain a *substantial gain* in security w.r.t. steganography with a single JPEG.

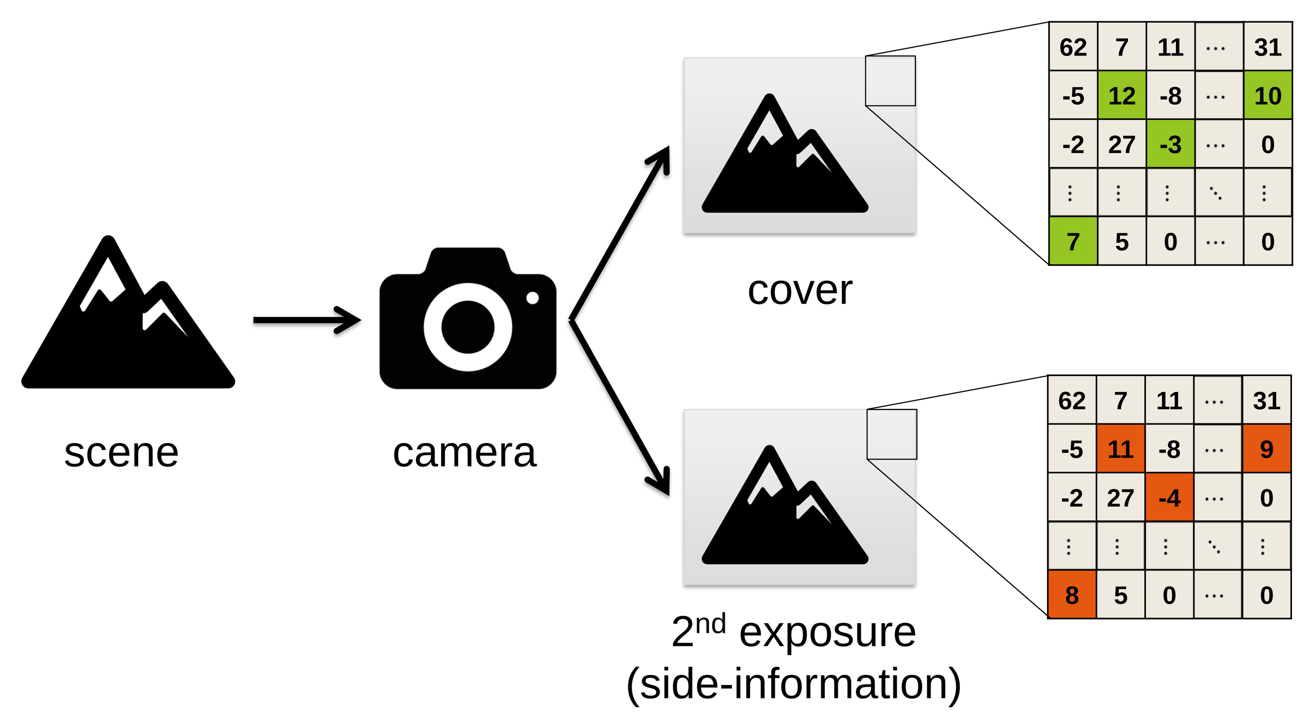
Steganography

Private, covert communication with a shared secret key.



The main idea

Exploit differences in JPEG DCT coefficients due to acquisition noise:



Embedding scheme (J2-UNIWARD)

Sender hides the secret message by modifying cover elements $x_{ij} \to x_{ij} \pm 1 = y_{ij}$ while minimizing the total embedding distortion

$$D(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{i,j} \rho_{ij}(y_{ij} - x_{ij}),$$

where $\rho_{ij}(0) = 0$ and $\rho_{ij}(-1), \rho_{ij}(+1) \geq 0$ are "costs" of changes determined by content complexity. This task is source coding with fidelity constraint.

Practical embedding can be implemented with syndrome-trellis codes [2], which operate near the payload-distortion bound. Recipient extracts secret message using a shared parity-check matrix H

$$\mathbf{H} \times \mathrm{LSB}\,(\boldsymbol{Y}) = \text{secret message}.$$

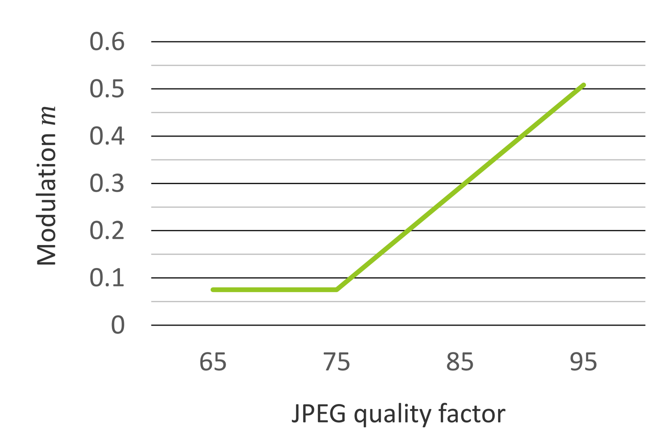
We start with costs from an existing stego method called J-UNIWARD [1]

$$\rho_{ij}^{(\mathrm{J})}(+1) = \rho_{ij}^{(\mathrm{J})}(-1) = \sum_{\mathcal{F} \in \mathcal{B}} \sum_{u,v} \frac{\left|\mathcal{F}(\boldsymbol{X})_{uv} - \mathcal{F}\left(\boldsymbol{X} \pm \delta_{ij}\right)_{uv}\right|}{1 + |\mathcal{F}(\boldsymbol{X})_{uv}|},$$

where $\boldsymbol{X}$ is the cover decompressed to spatial domain, $\mathcal{B}$ is a wavelet filter bank and δ_{ij} Kronecker delta.

The second exposure informs the sender about which elements in the cover (1st exposure) are most sensitive to acquisition noise. Their costs are *decreased* by a modulation parameter $0 \leq m \leq 1$ determined experimentally:

$$(■ = ■):\ \rho_{ij}(\pm 1) = \rho_{ij}^{(\mathrm{J})},$$
$$(■ < ■):\ \rho_{ij}(+1) = m\rho_{ij}^{(\mathrm{J})},\quad \rho_{ij}(-1) = \rho_{ij}^{(\mathrm{J})},$$
$$(■ > ■):\ \rho_{ij}(+1) = \rho_{ij}^{(\mathrm{J})},\quad \rho_{ij}(-1) = m\rho_{ij}^{(\mathrm{J})},$$



Empirical security evaluation

Warden's goal is to detect the *presence* of a secret. Currently, the best detectors are built as binary classifiers [6] trained on examples of cover and stego images represented using rich media models [3–5].

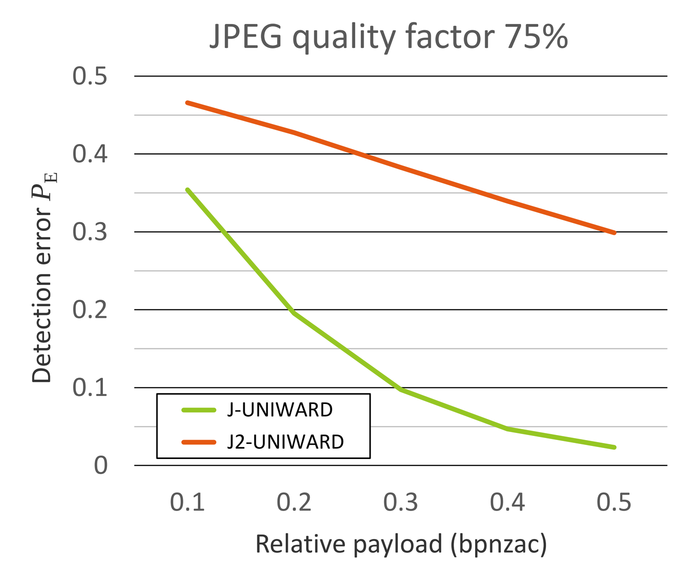
Security quantified as Warden's minimal total detection error under equal priors:

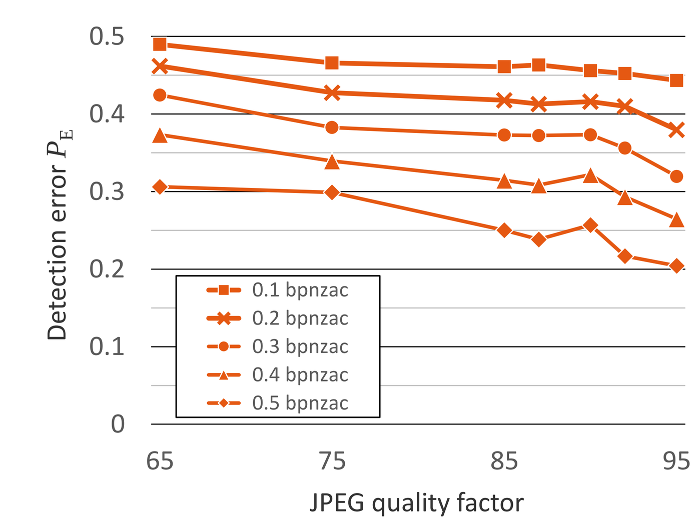
$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$$

averaged over 10 runs on different splits of the database into equal sized training and testing sets.

Experiments on BURSTbase

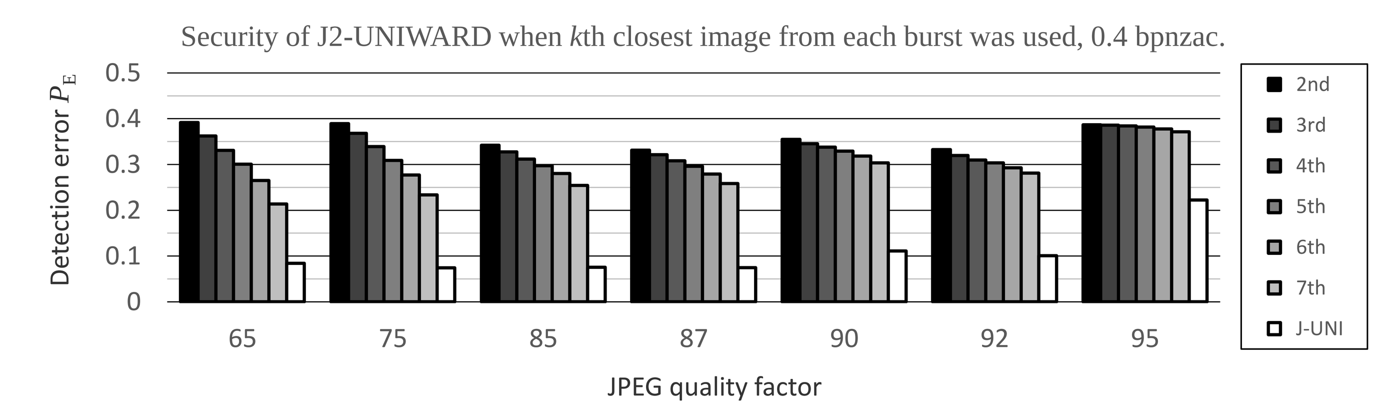
- 133 bursts of 7 pictures shot from a tripod, cut into 9310×7 tiles with 512×512 pixels
- From each burst, we selected two closest (MSE) images, one as the cover and the other as the 2nd exposure or side-information
- Classifier = linear LSMR [6], features = SRM + cc-JRM + GFR [3–5]
- bpnzac = bits per non-zero AC DCT coefficient





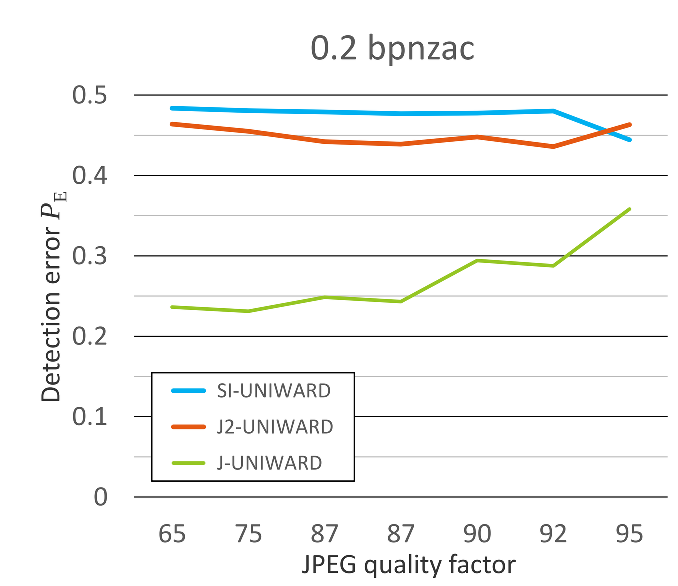
Robustness to camera shake

- Instead of the closest image use the *k*th closest as second exposure



Comparison to other side-info

- SI-UNIWARD [1] = sender has a single *uncompressed* image



Summary

- Steganography with two JPEGs much more secure than with a single JPEG
- Improvement across a range of JPEG quality factors
- Camera shake not a dealbreaker
- Method is practical, can be incorporated in mobile devices

Future directions

- Handheld bursts
- Consecutive frames from M-JPEG

[1] **Universal distortion function for steganography in an arbitrary domain**, V. Holub et al., *EURASIP Journal on Information Security 2014(1)*.

[2] **Minimizing additive distortion in steganography using syndrome-trellis codes,** T. Filler et al., *IEEE Transactions on Information Forensics and Security 6(3), 2011*.

[3] **Steganalysis of adaptive JPEG steganography using 2D Gabor filters,** X. Song et al., *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security.* ACM, 2015.

[4] **Rich models for steganalysis of digital images,** J. Fridrich et al., *IEEE Transactions on Information Forensics and Security 7.3*, 2012.

[5] **Steganalysis of JPEG Images using rich models,** J. Kodovský et al., *Proc. SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIV,* San Francisco, CA, 2012.

[6] **Modeling and extending the ensemble classifier for steganalysis of digital images using hypothesis testing theory,** R. Cogranne et al., *IEEE TIFS 10.2*, 2015.