Some Significant Steganalysis Algorithms

LSB Embedding and the Histogram Attack

Giving a relative message length $q = m/n$:

$$E\{\mathbf{T}_s[2i]\} = (1-\frac{q}{2})\mathbf{T}_c[2i] + \frac{q}{2}\mathbf{T}_c[2i+1]$$
$$E\{\mathbf{T}_s[2i+1]\} = \frac{q}{2}\mathbf{T}_c[2i] + (1-\frac{q}{2})\mathbf{T}_c[2i+1].$$

- Ineffective for random work embedding.
- Improvements:
  - Sliding window.
  - ...

Sample Pairs Analysis

A very clever method!

- Use spatial correlation within images.
- More reliable and accurate.

Basic Idea

Giving a sequence of values $s_1, s_2, \cdots, s_n$.

- All adjacent pairs
  $\mathcal{P} = \{(u,v) = (s_i, s_{i+1}), 1 \le i \le n\}$.

  $$(s_1, s_2), (s_2, s_3), \cdots, (s_{n-1}, s_n).$$

- Partition of $\mathcal{P}$:

	$v\%2 = 0$	$v\%2 = 1$
$u = v$	$\mathcal{Z}$	$\mathcal{Z}$
$u < v$	$\mathcal{X}$	$\mathcal{Y}$
$u > v$	$\mathcal{Y}$	$\mathcal{X}$

Partition of $\mathcal{P}$

Continue portioning $\mathcal{Y}$ into $\mathcal{W}, \mathcal{V}$.

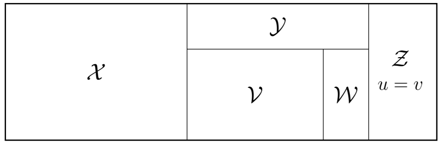
- $\mathcal{W}$: A small subset of $\mathcal{Y}$.

  $$\{(u = 2k, v = 2k{+}1) \vee (u = 2k{+}1, v = 2k), k \in \mathbb{Z}\}.$$

- $\mathcal{V} = \mathcal{P} - \mathcal{W}$.

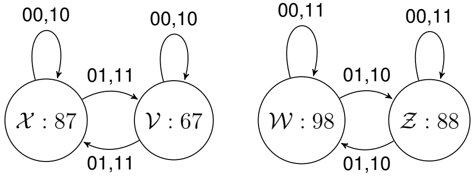
The bin of LSB: $\mathcal{W} + \mathcal{Z}$.

Partition of $\mathcal{P}$



A Finite State Machine

Notice that the **modification** pattern $\pi \in \{00, 01, 10, 11\}$ is not message binary sequence.

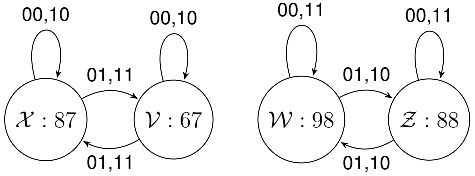


Transition Probability

Giving relative message length q, expectation of modification (i.e. 1) is $q/2$:

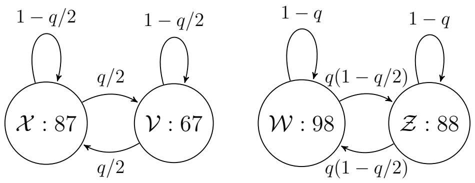
$$\rho(00, \mathcal{P}) = \left(1 - \frac{q}{2}\right)^2$$
$$\rho(01, \mathcal{P}) = \rho(10, \mathcal{P}) = \frac{q}{2}\left(1 - \frac{q}{2}\right)$$
$$\rho(11, \mathcal{P}) = \left(\frac{q}{2}\right)^2.$$

Put Them Together



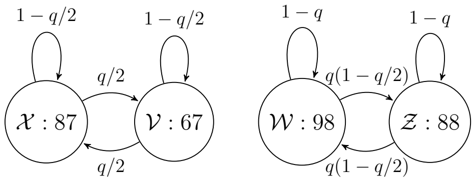
$$\begin{array}{l} \rho(00,\mathcal{P}) = \left(1-\frac{q}{2}\right)^2 \\ \rho(01,\mathcal{P}) = \frac{q}{2}\left(1-\frac{q}{2}\right) \\ \rho(10,\mathcal{P}) = \frac{q}{2}\left(1-\frac{q}{2}\right) \\ \rho(11,\mathcal{P}) = \left(\frac{q}{2}\right)^2 \end{array} \Rightarrow \begin{cases} 00,10: & \rho_{00}+\rho_{10} = 1-q/2 \\ 01,11: & \rho_{01}+\rho_{11} = q/2 \\ 00,11: & \rho_{00}+\rho_{11} = 1-q \\ 01,10: & \rho_{01}+\rho_{10} = q(1-q/2) \end{cases}$$

Put Them Together



$$\begin{aligned}\rho(00,\mathcal{P}) &= \left(1-\frac{q}{2}\right)^2\\ \rho(01,\mathcal{P}) &= \frac{q}{2}\left(1-\frac{q}{2}\right)\\ \rho(10,\mathcal{P}) &= \frac{q}{2}\left(1-\frac{q}{2}\right)\\ \rho(11,\mathcal{P}) &= \left(\frac{q}{2}\right)^2\end{aligned} \Rightarrow \begin{cases}00,10: & \rho_{00}+\rho_{10} = 1-q/2\\ 01,11: & \rho_{01}+\rho_{11} = q/2\\ 00,11: & \rho_{00}+\rho_{11} = 1-q\\ 01,10: & \rho_{01}+\rho_{10} = q(1-q/2)\end{cases}$$

Put Them Together



Count in and out:

$$\begin{aligned}|\mathcal{X}'| &= |\mathcal{X}|(1-q/2)+|\mathcal{V}|q/2\\ |\mathcal{V}'| &= |\mathcal{V}|(1-q/2)+|\mathcal{X}|q/2\\ |\mathcal{W}'| &= |\mathcal{W}|(1-q+q^2/2)+|\mathcal{Z}|q(1-q/2).\end{aligned}$$

Some Math

To solve q, we have equalities

$$\begin{aligned}|\mathcal{X}'| &= |\mathcal{X}|(1-q/2)+|\mathcal{V}|q/2\\ |\mathcal{V}'| &= |\mathcal{V}|(1-q/2)+|\mathcal{X}|q/2\\ &\Rightarrow\\ |\mathcal{X}'|-|\mathcal{V}'| &= (|\mathcal{X}|-|\mathcal{V}|)(1-p)\\ &= |\mathcal{W}|(1-p)\\ |\mathcal{W}'| &= |\mathcal{W}|(1-q+q^2/2)+|\mathcal{Z}|q(1-q/2)\\ &= |\mathcal{W}|(1-q)^2+(|\mathcal{W}|+|\mathcal{Z})|q(1-q/2)\\ &= |\mathcal{W}|(1-q)^2+\gamma q(1-q/2)\\ &= (|\mathcal{X}'|-|\mathcal{V}'|)(1-q)+\gamma q(1-q/2)\end{aligned}$$

Continue

Because $|\mathcal{W}'| = |\mathcal{P}|-|\mathcal{X}'|-|\mathcal{V}'|-|\mathcal{Z}'|$:

$$\begin{aligned}&|\mathcal{P}|-|\mathcal{X}'|-|\mathcal{V}'|-|\mathcal{Z}'| =\\ &\quad (|\mathcal{X}'|-|\mathcal{V}'|)(1-q)+\gamma q(1-q/2)\\ &\frac{\gamma}{2}q^2+(|\mathcal{P}|-|\mathcal{X}'|-|\mathcal{V}'|-|\mathcal{Z}'|) =\\ &\quad (|\mathcal{X}'|-|\mathcal{V}'|)-(|\mathcal{X}'|-|\mathcal{V}'|)q+\gamma q\\ &\frac{\gamma}{2}q^2+(|\mathcal{P}|-2|\mathcal{X}'|-|\mathcal{Z}'|) =\\ &\quad -(|\mathcal{X}'|-|\mathcal{V}'|-\gamma)q\\ &\frac{\gamma}{2}q^2+(|\mathcal{X}'|-|\mathcal{V}'|-\gamma)q+(|\mathcal{P}|-2|\mathcal{X}'|-|\mathcal{Z}'|) = 0.\end{aligned}$$

More Compacted Form

$$\begin{aligned}0 &= \frac{\gamma}{2}q^2+(|\mathcal{X}'|-|\mathcal{V}'|-\gamma)q+(|\mathcal{P}|-2|\mathcal{X}'|-|\mathcal{Z}'|)\\ &= \frac{\gamma}{2}q^2+(|\mathcal{X}'|-|\mathcal{V}'|-|\mathcal{W}'|-|\mathcal{Z}'|)q\\ &\qquad +(|\mathcal{X}'|+|\mathcal{Y}'|+|\mathcal{Z}'|-2|\mathcal{X}'|-|\mathcal{Z}'|)\\ &= \frac{\gamma}{2}q^2+(2|\mathcal{X}'|-|\mathcal{P}|)q+(|\mathcal{Y}'|-|\mathcal{X}'|).\end{aligned}$$

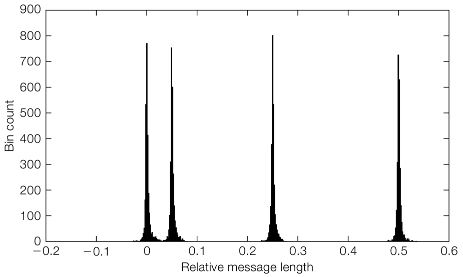
The Solution

- If $\gamma = 0$, $|\mathcal{X}| = |\mathcal{X}'| = |\mathcal{Y}| = |\mathcal{Y}'| = |\mathcal{P}|/2$.

  $$0q^2+0q+0 = 0.$$

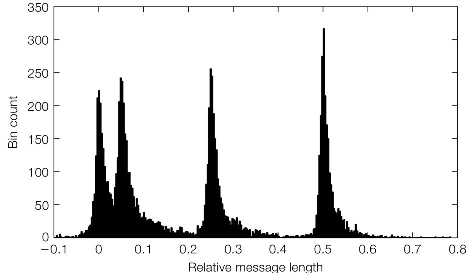
- If two complex conjugate roots:
  - Taking the real parts.
- If has a negative root:
  - $p = 0$.

JPEG



$q = 0.05, 0.2, 0.5.$

Raw Scan



$q = 0.05, 0.2, 0.5.$

Analysis

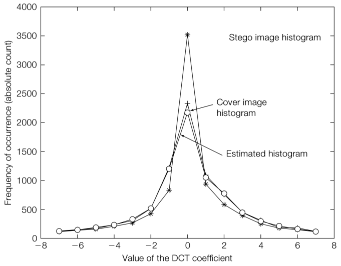
- Noisy has negative influence.
- Estimation for short message is not robust.
- Sample
  - Local is better
  - Thus neighboring pairs.

Extension

- One point: histogram
- Sample pairs.
- Sample more: 2×2 neighboring pixels.

Blind Steganalysis Using Calibration

- Shift 4 pixels and re-compress.



In General

$$f_i = \|F_i(J_1) - F_i(J_2)\|.$$

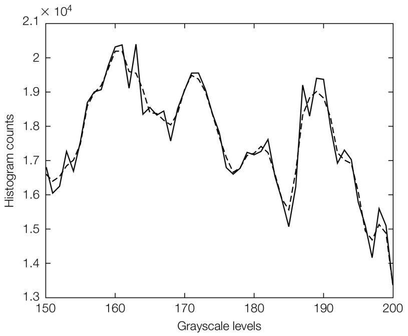
- J_1: stego JPEG image.
- J_2: shift and re-compress stego JPEG image.
- Find efficient F_i or training.

In Spatial Domain

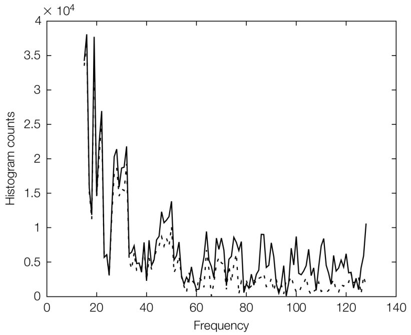
Just using different feature.

- Steganographic method: adding noise.
- Smooth the work a little bit and check the difference.

Illustration



Illustration



A Basic Method

Compute the noise residual from a smoother F:

$$\mathbf{r} = \mathbf{s} - F(\mathbf{s}).$$

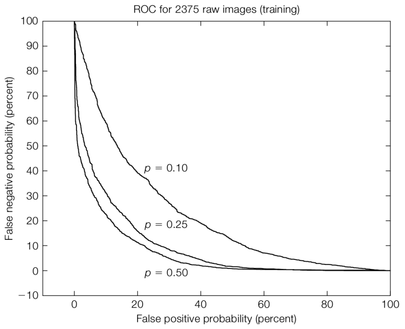
Then use $k = 1, 2, \cdots$ moments as the feature:

$$\mu_k = \sum (\mathbf{r} - \bar{\mathbf{r}})^k.$$

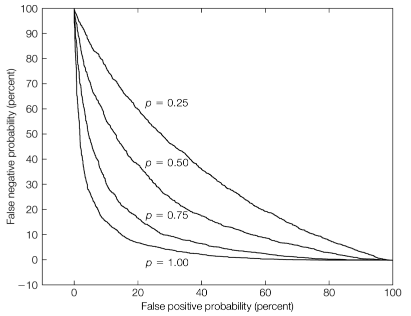
Classification via Fisher linear discriminant.

More details in the book.

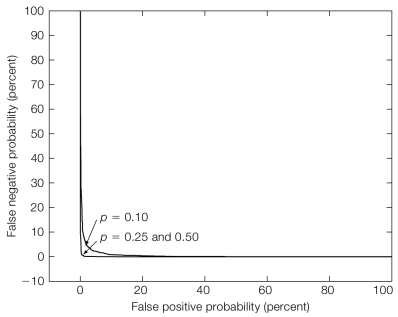
Raw Digital Camera



Raw Scans



JPEG



Analysis

- Noise!
  - It is better to pick noise image as the cover for steganography.