The Challenges of Rich Features in Universal Steganalysis

Tomáš Pevný\textsuperscript{a} and Andrew D. Ker\textsuperscript{b}

\textsuperscript{a}Agent Technology Center, Czech Technical University in Prague, Czech Republic.
\textsuperscript{b}Department of Computer Science, Oxford University, England.

7th February 2013
Batch universal steganalysis
Batch universal steganalyzer

- Extract features.
- Calculate distances between actors (MMD).
- Identify the steganographer(s).
  local outlier factor (LOF)
Batch universal steganalyzer

- Extract features.
- Calculate distances between actors (MMD).
- Identify the steganographer(s).
- Local outlier factor (LOF)
Batch universal steganalyzer

- Extract features.
- Calculate distances between actors (MMD).
- Identify the steganographer(s).
- local outlier factor (LOF)

guilty
Batch universal steganalyzer

- Extract features.
- Calculate distances between actors (MMD).
- Identify the steganographer(s).

local outlier factor (LOF)

The method should work with any stego-sensitive features.
Accuracy with PF274 and $CF^*$ features

<table>
<thead>
<tr>
<th></th>
<th>PF274</th>
<th>$CF^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>dimension</td>
<td>274</td>
<td>8750</td>
</tr>
<tr>
<td>F5</td>
<td>14.6</td>
<td>9.5</td>
</tr>
<tr>
<td>nsF5</td>
<td>10.7</td>
<td>23.1</td>
</tr>
<tr>
<td>JP Hide&amp;Seek</td>
<td>7.8</td>
<td>16.2</td>
</tr>
<tr>
<td>OutGuess</td>
<td>1.9</td>
<td>5.7</td>
</tr>
<tr>
<td>Steghide</td>
<td>2.8</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Average rank of one guilty actor (out of 100) emitting payload 0.1 bpnc
Curse of dimensionality

- Anomaly detection estimates density: more difficult in high dimensions.
- In unsupervised learning cannot discard noise in features.
Curse of dimensionality

Our solution
Supervised dimensionality reduction.

Our aim
Steganographic features should be sensitive to embedding changes, yet insensitive to image content.
J. Fridrich, 2004
Dimensionality reduction

Prior art
- Principal component transformation
- Maximum covariance
- Ordinary least square regression

Proposed
- Calibrated least-squares
Principal component transformation (PCT)

\[
\arg\max_{w_k} \text{Var}(Xw_k)
\]

subject to

\[w_k \perp w_i, \; i \in \{1, \ldots, k-1\}.\]

\[X \in \mathbb{R}^{n,d} \quad \text{— matrix with features}\]

\[w_i \quad \text{— projections found}\]
Ordinary least square regression (OLS)

\[
\arg\max_{w_k} \text{Cov}(X^s w_k, Y^s) - \text{Var}(X^s w_k)
\]

subject to

\[w_k \perp w_i, \ i \in \{1, \ldots, k-1\}.
\]

\[X^s \in \mathbb{R}^{n,d} \quad \text{— matrix with stego features}
\]

\[Y^s \in \mathbb{R}^{n,1} \quad \text{— vector with payload}
\]

\[w_i \quad \text{— projections found}
\]
Maximum covariance (MCV)

\[
\arg \max_{w_k} \text{Cov}(X^s w_k, Y^s)
\]

subject to

\[w_k \perp w_i, \quad i \in \{1, \ldots, k - 1\}.\]

\(X^s \in \mathbb{R}^{n,d}\) — matrix with stego features

\(Y^s \in \mathbb{R}^{n,1}\) — vector with payload

\(w_i\) — projections found
Calibrated least squares (CLS)

$$\arg \max_{w_k} \text{Cov}(X^s w_k, Y^s) - \text{Var}(X^c w_k)$$

subject to

$$w_k \perp w_i, \ i \in \{1, \ldots, k - 1\}.$$
Experimental settings

- 3000 users of leading social network, 100 images from each
  - 1000 users for supervised feature reduction
  - 2000 users used for testing

- Guilty actor emits payload 0.1 bpnc
  - linear (in the paper) or greedy strategy
  - one of following algorithms:
    - F5, nsF5, JPHide&Seek (JP), OutGuess (OG), Steghide (SH)

- Steganalyst uses reduced \( CF^* \) features.

- Accuracy is measured by average rank of guilty actor.
  - \( 1.0 = \text{perfect}, \ 50.5 = \text{random guessing} \).
## Results

<table>
<thead>
<tr>
<th></th>
<th>PCT</th>
<th>MCV</th>
<th>OLS</th>
<th>CLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>F5</td>
<td>40.3</td>
<td>23.4</td>
<td>22.2</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td>(1)</td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>nsF5</td>
<td>38.0</td>
<td>26.6</td>
<td>5.8</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td>(1)</td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>JP</td>
<td>38.4</td>
<td>27.2</td>
<td>6.9</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(1)</td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>OG</td>
<td>26.5</td>
<td>31.6</td>
<td>2.4</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>(4)</td>
<td>(1)</td>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>SH</td>
<td>23.0</td>
<td>2.6</td>
<td>1.3</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>(1)</td>
<td>(1)</td>
<td></td>
</tr>
</tbody>
</table>
## Robustness

<table>
<thead>
<tr>
<th>PCT</th>
<th>CLS trained on</th>
<th>F5</th>
<th>nsF5</th>
<th>JP</th>
<th>OG</th>
<th>SH</th>
</tr>
</thead>
<tbody>
<tr>
<td>F5</td>
<td></td>
<td>40.3</td>
<td>1.6</td>
<td>8.8</td>
<td>6.6</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(4)</td>
<td>(3)</td>
</tr>
<tr>
<td>nsF5</td>
<td></td>
<td>38.0</td>
<td>1.8</td>
<td>10.1</td>
<td>10.9</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(1)</td>
<td>(1)</td>
<td>(4)</td>
<td>(3)</td>
</tr>
<tr>
<td>JP</td>
<td></td>
<td>38.4</td>
<td>8.9</td>
<td>7.2</td>
<td>15.5</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
<td>(2)</td>
</tr>
<tr>
<td>OG</td>
<td></td>
<td>26.5</td>
<td>3.7</td>
<td>11.8</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(6)</td>
<td>(2)</td>
<td>(1)</td>
<td>(1)</td>
</tr>
<tr>
<td>SH</td>
<td></td>
<td>23.0</td>
<td>5.2</td>
<td>9.1</td>
<td>1.2</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(6)</td>
<td>(2)</td>
<td>(1)</td>
<td>(1)</td>
</tr>
</tbody>
</table>
Optimal number of projections

The diagram illustrates the average rank for different methods as a function of the number of projections. The methods include F5, nsF5, JP, OG, and SH. The x-axis represents the number of projections, while the y-axis shows the average rank. As the number of projections increases, the average rank generally decreases, indicating improved performance or condensation of rich features.
Conclusion

- High dimensional features are not compatible with unsupervised steganalysis.
- Investigated dimensionality reduction to improve SNR of rich features.
- Validated the approach in universal batch steganalysis.
- The proposed method, CLS, exhibits robustness to embedding method.
F5 phenomenon

![Graph showing estimated change rate vs true payload for different datasets: F5, nsF5, JP, OG, and SH. The graph includes error bars for each dataset, indicating variability in the estimated change rate.]