A New Paradigm for Steganalysis via Clustering

Andrew D. Ker[1] and Tomáš Pevný[2]

[1]University of Oxford, UK
[2]Czech Technical University in Prague

26th January 2011

Outline

Classical steganalysis

scenario

- Independently tests objects for stego content.
- Consider only one actor.
- Rarely happens in practice.

usual approach

- Train a classifier on examples of cover and stego images.
- The classifier can detect only some steganographic algorithms.
- Might have problems with model mismatch.

Batch steganalysis

scenario: monitoring a network

- *Multiple* actors transmit *multiple* objects.
- Guilty actors include some *stego* objects.
- Most actors are innocent.

new paradigm

1. Extract steganalytic features from all objects.
2. Calculate distance between actors.
3. Cluster actors.

Comparison

advantages of the new paradigm

- It does not need to be trained.
- It is potentially universal.
- It removes the problem of model mismatch.

limitations of classical steganalysis

- Train a classifier on examples of cover and stego images.
- The classifier can detect only some steganographic algorithms.
- Might have problems with model mismatch.

Comparison

advantages of the new paradigm

- It does not need to be trained.
- It is potentially universal.
- It removes the problem of model mismatch.

limitations of classical steganalysis

- Train a classifier on examples of cover and stego images.
- The classifier can detect only some steganographic algorithms.
- Might have problems with model mismatch.

Comparison

advantages of the new paradigm

- It does not need to be trained.
- It is potentially universal.
- It removes the problem of model mismatch.

limitations of classical steganalysis

- Train a classifier on examples of cover and stego images.
- The classifier can detect only some steganographic algorithms.
- Might have problems with model mismatch.

Outline

Agglomerative clustering

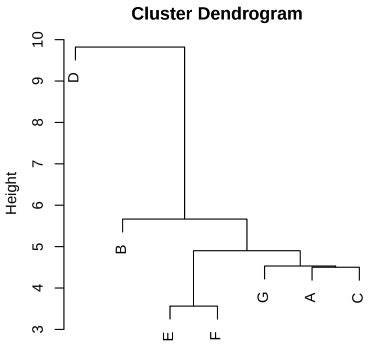
algorithm

Initialization

Each actor is in one cluster.

Iteration

Join two closest clusters.



Distances between clusters

single linkage	$D_{\mathrm{SL}}(X,Y) = \min_{\substack{x\in X \\ y\in Y}} d(x,y)$
complete linkage	$D_{\mathrm{CL}}(X,Y) = \max_{\substack{x\in X \\ y\in Y}} d(x,y)$
centroid	$D_{\mathrm{CEN}}(X,Y) = \frac{1}{\lvert X\rvert\cdot\lvert Y\rvert} \sum_{x\in X,} \sum_{y\in Y} d(x,y)$
average linkage	$D_{\mathrm{AVG}}(X,Y) = \frac{1}{(\lvert X\cup Y\rvert)(\lvert X\cup Y\rvert-1)} \sum_{x\in X,} \sum_{y\in Y} d(u,v)$

Outline

Maximum mean discrepancy (1)

definition

$$\mathrm{MMD}(\mathscr{P},\mathscr{Q}) = \max_{f\in\mathscr{F}} \left|\mathbf{E}_{X\sim\mathscr{P}}[f(X)] - \mathbf{E}_{X\sim\mathscr{Q}}[f(X)]\right|$$

$\mathscr{P}$ and $\mathscr{Q}$ are probability distributions
$\mathscr{F}$ is a unit ball in Reproducing Kernel Hilbert Space

unbiased estimator

$$\mathrm{MMD}^2(\mathbf{X},\mathbf{Y}) = \frac{1}{n(n-1)}\sum_{i\neq j} k(x_i,x_j) + k(y_i,y_j) - k(x_i,y_j) - k(x_j,y_i)$$

$k(\cdot,\cdot) : D\times D \mapsto \mathbb{R}$ is universal bounded kernel,
and $\{x_i\}_{i=1}^n \sim \mathscr{P}$, $\{y_i\}_{i=1}^n \sim \mathscr{Q}$

Maximum mean discrepancy (2)

examples of kernels

Gaussian kernel	$k(x,y) = e^{-\gamma\|x-y\|_2^2}$
Linear kernel	$k(x,y) = x^t y$

some practical remarks

- Features are normalized to have zero mean and unit variance.
- Kernel and normalization parameters have to be fixed.

Proposed method

algorithm

1. Extract features from all objects.
2. Normalize features & calculate MMD between actors.
3. Cluster actors by agglomerative clustering.

assumptions

- *Multiple* actors transmit *multiple* objects.
- Guilty actors include some *stego* objects.
- Most actors are innocent.

Experimental setup

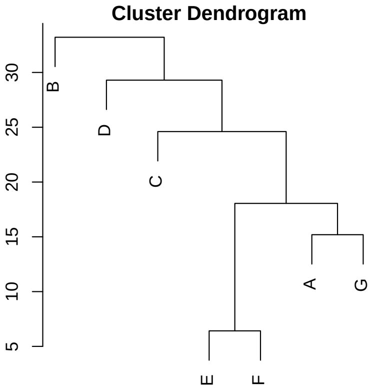
- Actors are simulated by different digital cameras.
- All actors use single-compressed JPEGs with quality factor 80.
- Guilty actor uses F5 with shrinkage removed by wet paper codes and matrix embedding turned off.
- The steganalyst uses PF274 features.
- Result presented here used centroid clustering (more in the paper).

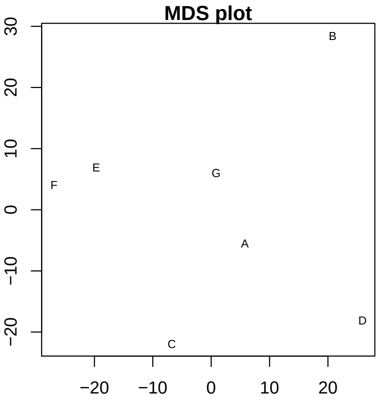
Outline

Actors and their cameras

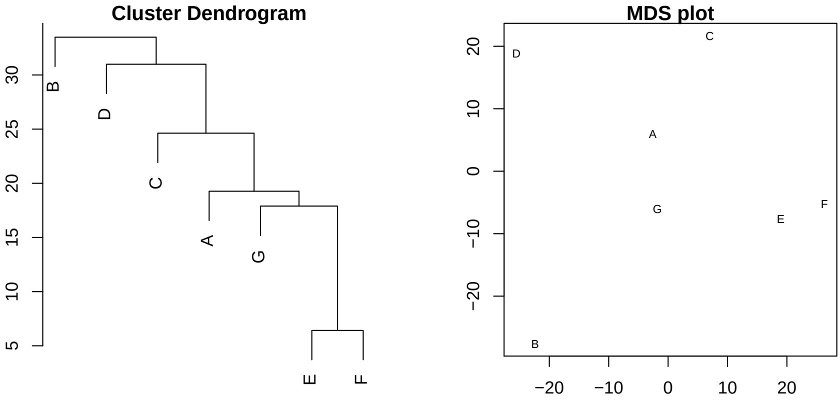
actor	camera	resolution
A	Olympus c765	2288×1712
B	Nikon D100	3008×2000
C	Sigma SD9	2268×1512
D	Minolta DiMage A1	2000×1500
E	Canon Powershot G2	2272×1704
F	Canon Powershot S40	2272×1704
G	Kodak DC290	1792×1200

All actors emit cover objects

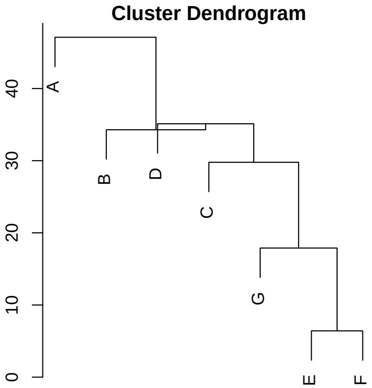


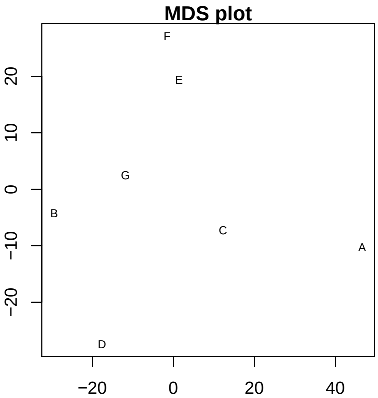


Actor A emits payload 0.1bpnz in 10% images

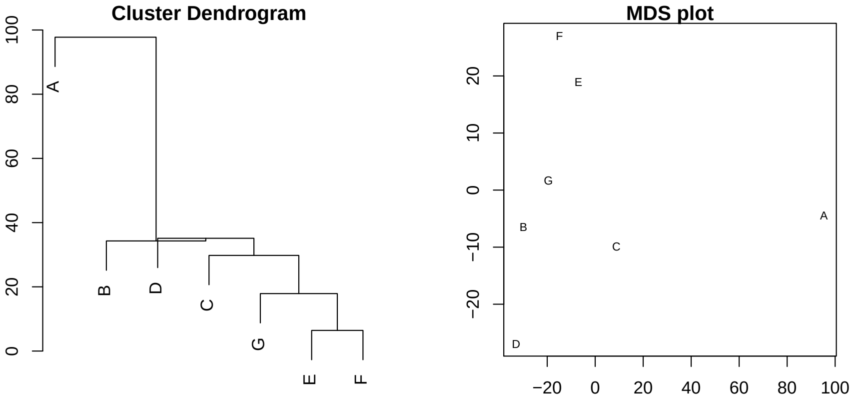


Actor A emits payload 0.2bpnz in 20% images

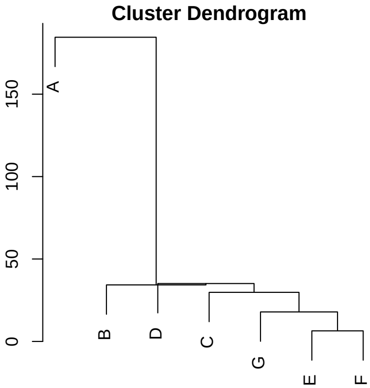


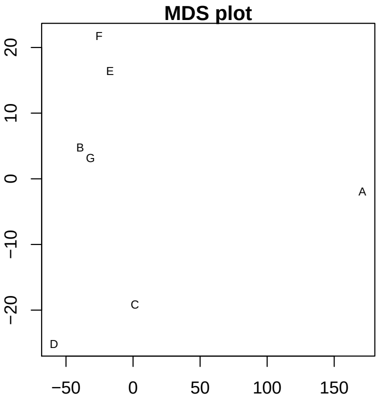


Actor A emits payload 0.25bpnz in 25% images

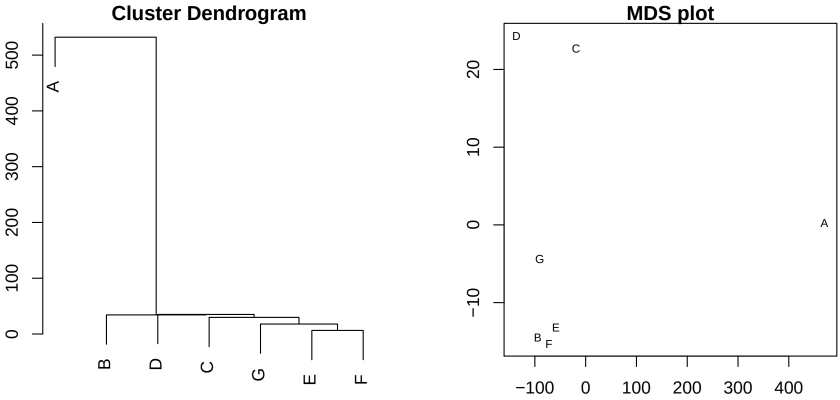


Actor A emits payload 0.3bpnz in 30% images

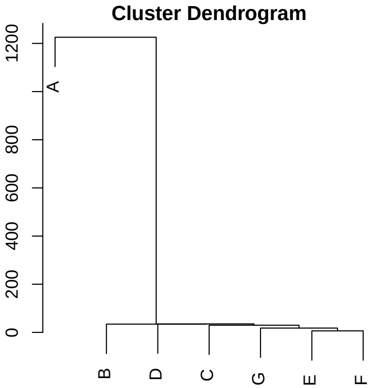


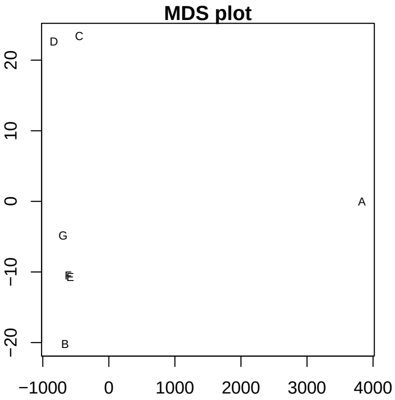


Actor A emits payload 0.4bpnz in 40% images

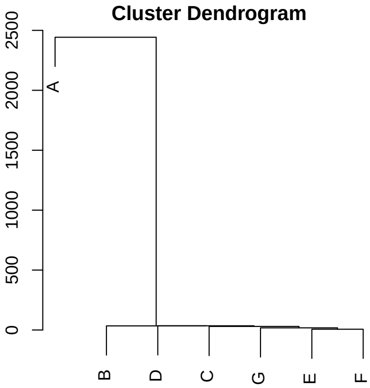


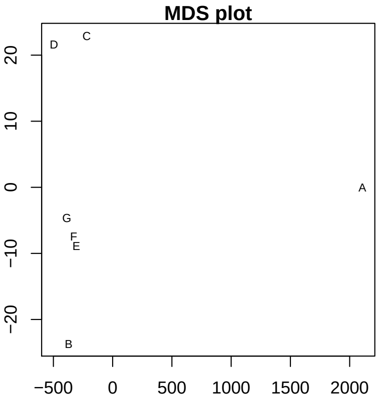
Actor A emits payload 0.5bpnz in 50% images



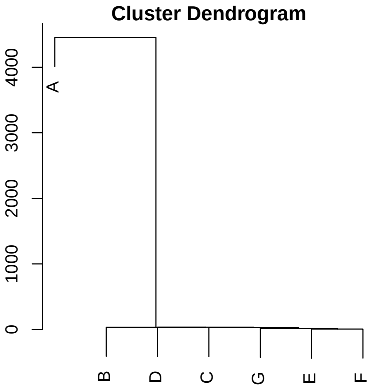


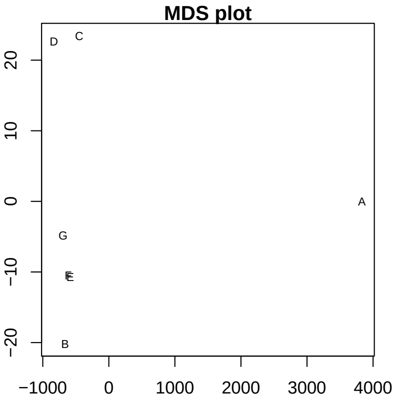
Actor A emits payload 0.6bpnz in 60% images





Actor A emits payload 0.7bpnz in 70% images





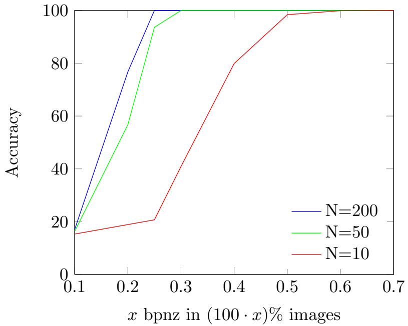
7 actors, one guilty — experimental results

proposed detector

Almost perfect detection at 0.0625bpnz.
N=200, 0.25bpnz in 25% images

targeted detector

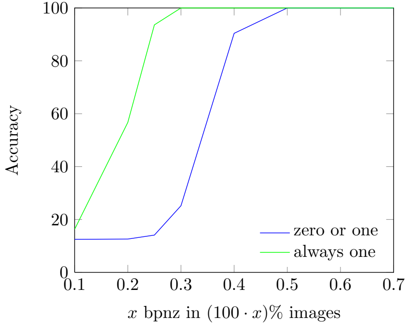
Almost perfect detection at 0.05bpnz.



Outline

7 actors, zero or one guilty

- Zero false positive rate.
- Zero incorrect positive rate (accusation of innocent person).
- Decrease in accuracy due to false negatives.



7 actors, zero or one guilty

Guilty actor	Identified actor: none	A	B	C	D	E	F	G
none	100	0	0	0	0	0	0	0
A	96	4	0	0	0	0	0	0
B	51	0	49	0	0	0	0	0
C	68	0	0	32	0	0	0	0
D	99	0	0	0	1	0	0	0
E	93	0	0	0	0	7	0	0
F	91	0	0	0	0	0	9	0
G	100	0	0	0	0	0	0	0

Tab: 0.3bpnz in 30% images
overall accuracy 25.3%
N=50

Guilty actor	Identified actor: none	A	B	C	D	E	F	G
none	100	0	0	0	0	0	0	0
A	11	89	0	0	0	0	0	0
B	2	0	98	0	0	0	0	0
C	0	0	0	100	0	0	0	0
D	22	0	0	0	78	0	0	0
E	11	0	0	0	0	89	0	0
F	7	0	0	0	0	0	93	0
G	24	0	0	0	0	0	0	76

Tab: 0.4bpnz in 40% images
overall accuracy 90.4%
N=50

Outline

Cameras of new actors

actor	camera	resolution
H	Canon EOS 400D	3906×2602
I	Pentax K20D	4688×3124
J	Canon EOS 7D	5202×3465
K	Canon Digital Rebel XSi	4290×2856
L	Leica M9	5216×3472
M	Nikon D70	3039×2014

13 actors using images with native resolution

- native resolution
- 0.7bpnz in 70% images
- overall accuracy 43%
- 100 images per actor
- One guilty actor
- MMD with linear kernel
- Centroid clustering

Guilty actor	Identified actor A	B	C	D	E	F	G	H	I	J	K	L	M
A	64	0	0	0	0	0	0	0	0	0	36	0	0
B	0	43	0	0	0	0	0	0	0	0	57	0	0
C	0	0	64	0	0	0	0	0	0	0	36	0	0
D	0	0	0	71	0	0	0	0	0	0	29	0	0
E	0	0	0	0	18	0	0	0	0	0	82	0	0
F	0	0	0	0	0	18	0	0	0	0	82	0	0
G	0	0	0	0	0	0	21	0	0	0	79	0	0
H	0	0	0	0	0	0	0	55	0	0	45	0	0
I	0	0	0	0	0	0	0	0	2	0	98	0	0
J	0	0	0	0	0	0	0	0	0	100	0	0	0
K	0	0	0	0	0	0	0	0	0	0	100	0	0
L	0	0	0	0	0	0	0	0	0	0	100	0	0
M	0	0	0	0	0	0	0	0	0	0	97	0	3

Tab: N=100

13 actors using cropped images

- Images cropped to 1792×1200
- 0.3bpnz in 30% images
- overall accuracy 85.1%
- 100 images per actor
- One guilty actor
- MMD with linear kernel
- Centroid clustering

Guilty actor	Identified actor A	B	C	D	E	F	G	H	I	J	K	L	M
A	97	0	0	0	0	0	0	0	0	0	0	0	3
B	0	97	0	0	0	0	0	0	0	0	0	0	3
C	0	0	78	0	0	0	0	0	0	0	0	0	22
D	0	0	0	93	0	0	0	0	0	0	0	0	7
E	0	0	0	0	94	0	0	0	0	0	0	0	6
F	0	0	0	0	0	90	0	0	0	0	0	0	10
G	0	0	0	0	0	0	79	0	0	0	0	0	21
H	0	0	0	9	0	0	0	51	1	0	0	0	39
I	0	0	0	0	0	0	0	0	93	0	0	0	7
J	0	0	0	0	0	0	0	0	0	92	0	0	8
K	0	0	0	1	0	0	0	0	0	0	62	0	37
L	0	0	0	0	0	0	0	0	0	0	0	80	20
M	0	0	0	0	0	0	0	0	0	0	0	0	100

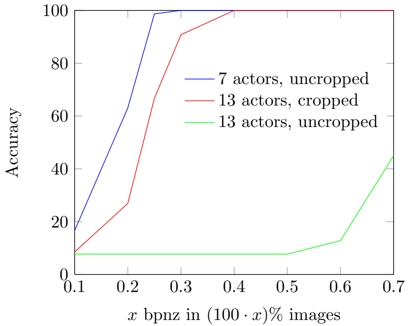
Tab: N=100

Scenarios compared

- MMD with linear kernel
- Centroid clustering

Problem

PF274 features do not scale well with respect to image size.



Conclusion

- New paradigm for pooled steganalysis was introduced.
  - Uses clustering rather then classification.
  - Identifies actors rather then objects.
- Advantages
  - Does not need training.
  - Is universal.
  - Mitigates the model mismatch.
- Good accuracy for payload $0.04 - 0.16$bpnz.

Future directions

- Investigate other clustering techniques.
- Examine distance metric based on MMD.
- Examine robustness of the PF274 features.
- Develop more robust features.