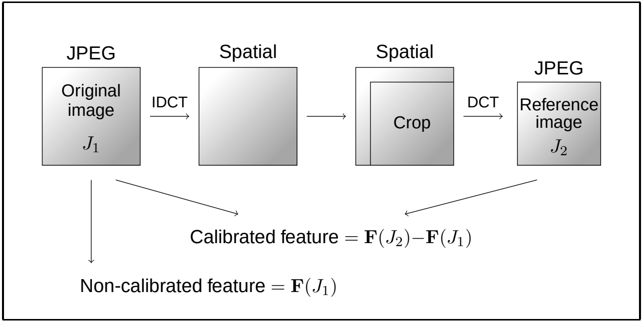
Calibration Revisited

Jan Kodovský, Jessica Fridrich

September 7, 2009 / ACM MM&Sec '09

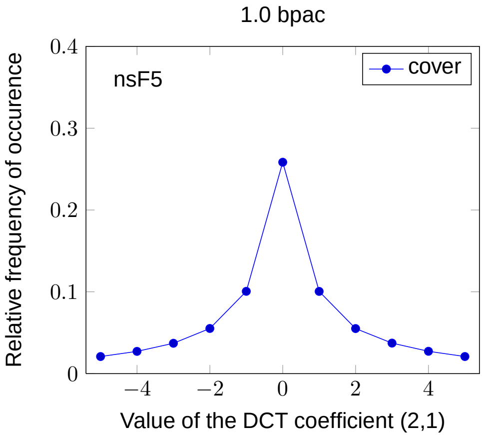
What is Calibration?

- 2002 - Calibration introduced (attack on F5)
- Part of feature extraction procedure for blind steganalysis
- Idea: estimate cover image statistics from the stego image



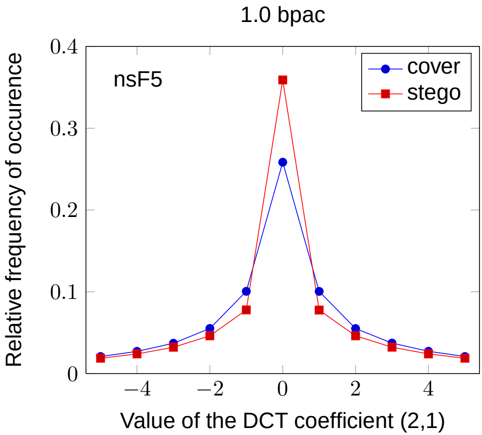
Motivation

- How well does calibration approximate cover?
- Experiment: local histograms (average over 6,500 images)



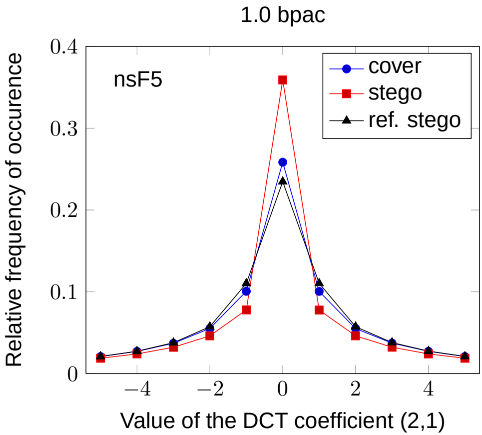
Motivation

- How well does calibration approximate cover?
- Experiment: local histograms (average over 6,500 images)



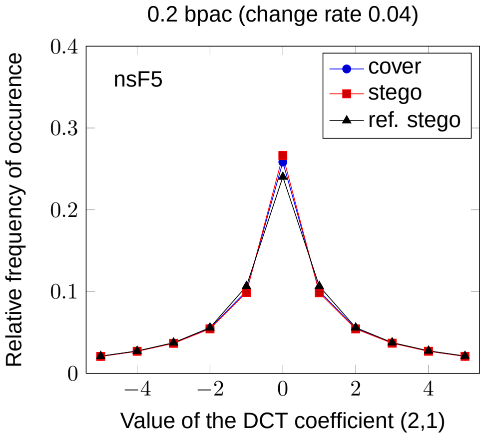
Motivation

- How well does calibration approximate cover?
- Experiment: local histograms (average over 6,500 images)



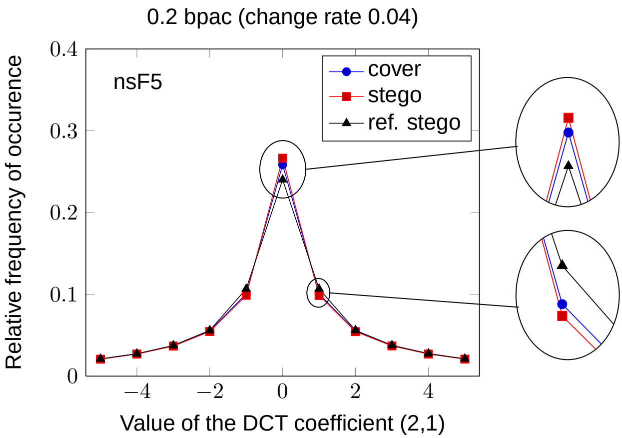
Motivation

- How well does calibration approximate cover?
- Experiment: local histograms (average over 6,500 images)



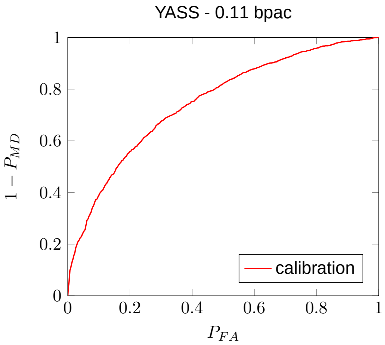
Motivation

- How well does calibration approximate cover?
- Experiment: local histograms (average over 6,500 images)



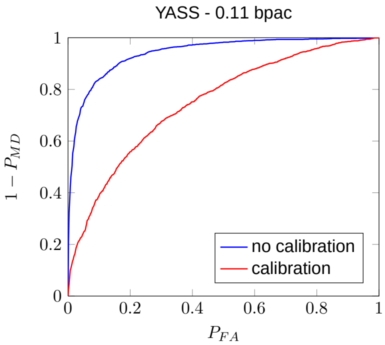
Motivation, cont'd

- Detectability of the steganographic algorithm YASS
- [Pevný 2007] - 274 merged features (Pevný Feature Set)
- SVM machine with Gaussian kernel, 6500 images



Motivation, cont'd

- Detectability of the steganographic algorithm YASS
- [Pevný 2007] - 274 merged features (Pevný Feature Set)
- SVM machine with Gaussian kernel, 6500 images



Challenges

Challenges

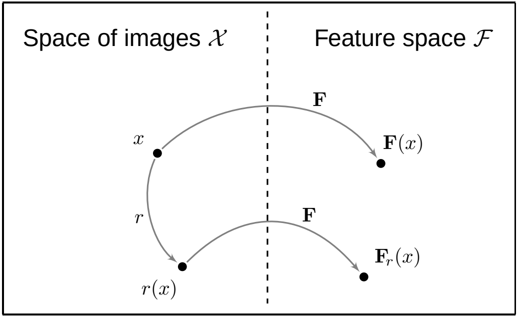
- How exactly does calibration affect detectability of steganographic algorithms?
- What is the real purpose of calibration?
- Does it make sense to calibrate all features?

Goals

- Create appropriate model for calibration
- Quantitative evaluation of the contribution of calibration to steganalysis performance

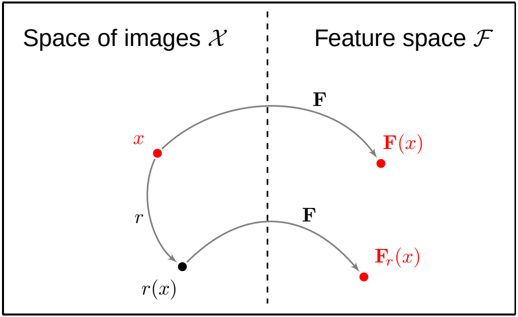
Notation

- Feature mapping . . . $\mathbf{F} : \mathcal{X} \to \mathcal{F}$
- Reference transform . . . $r : \mathcal{X} \to \mathcal{X}$
- Reference-feature mapping . . . $\mathbf{F}_r = \mathbf{F} \circ r : \mathcal{X} \to \mathcal{F}$

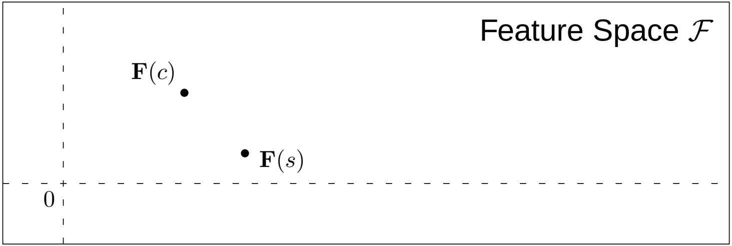


Notation

- Feature mapping … $\mathbf{F} : \mathcal{X} \to \mathcal{F}$
- Reference transform … $r : \mathcal{X} \to \mathcal{X}$
- Reference-feature mapping … $\mathbf{F}_r = \mathbf{F} \circ r : \mathcal{X} \to \mathcal{F}$



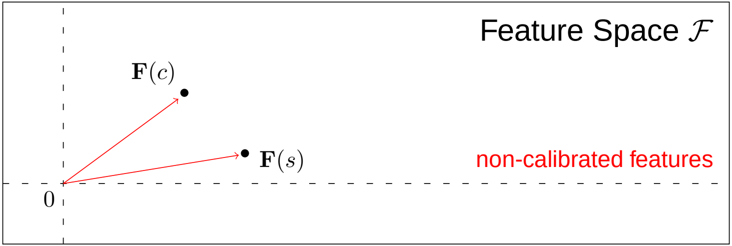
Basic Concept



$\mathbf{F}(c), \mathbf{F}(s) \ldots$ original features

$\mathbf{F}_r(c), \mathbf{F}_r(s) \ldots$ reference features

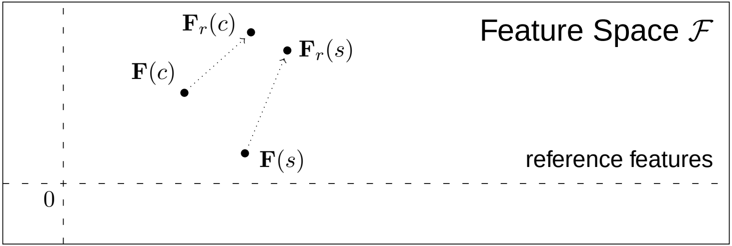
Basic Concept



$\mathbf{F}(c), \mathbf{F}(s) \ldots$ original features

$\mathbf{F}_r(c), \mathbf{F}_r(s) \ldots$ reference features

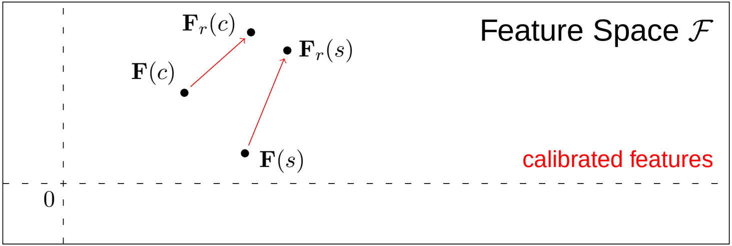
Basic Concept



$\mathbf{F}(c), \mathbf{F}(s) \ldots$ original features

$\mathbf{F}_r(c), \mathbf{F}_r(s) \ldots$ reference features

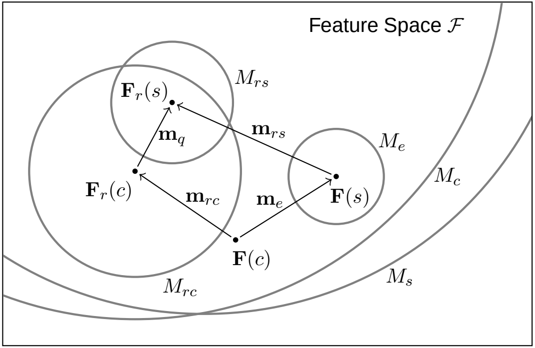
Basic Concept



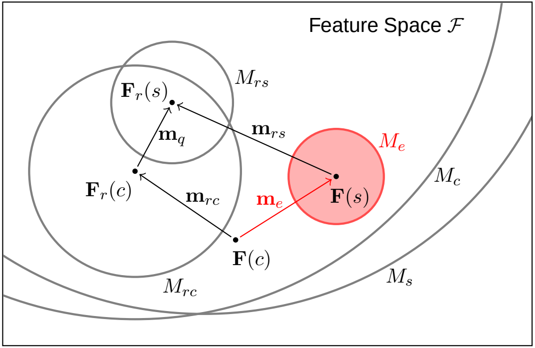
$\mathbf{F}(c), \mathbf{F}(s) \ldots$ original features

$\mathbf{F}_r(c), \mathbf{F}_r(s) \ldots$ reference features

Proposed Model

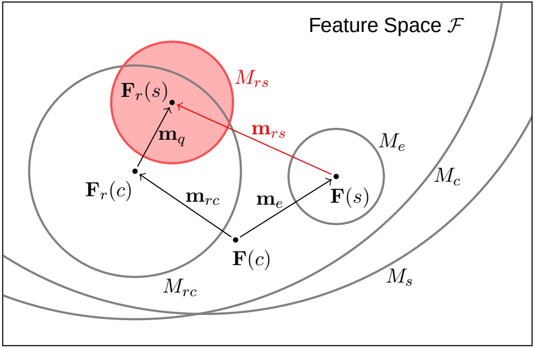


Proposed Model



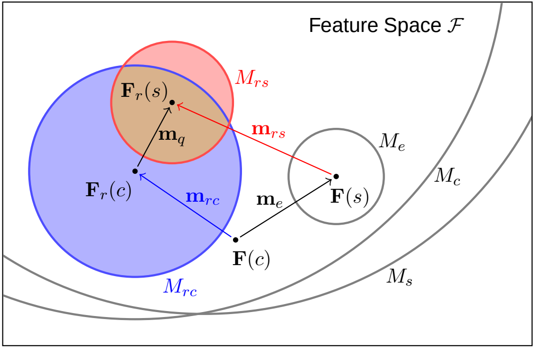
$$\begin{aligned} \mathbf{m}_e &= \text{median}\left[\mathbf{F}(s) - \mathbf{F}(c)\right], \\ M_e &= \text{median}\left[\|\mathbf{F}(s) - \mathbf{F}(c) - \mathbf{m}_e\|\right] \end{aligned}$$

Proposed Model



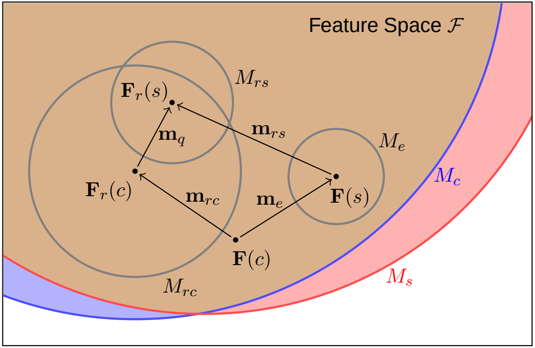
$$
\begin{aligned}
\mathbf{m}_{rs} &= \text{median}\left[\mathbf{F}(rs) - \mathbf{F}(s)\right], \\
M_{rs} &= \text{median}\left[||\mathbf{F}(rs) - \mathbf{F}(s) - \mathbf{m}_{rs}||\right]
\end{aligned}
$$

Proposed Model



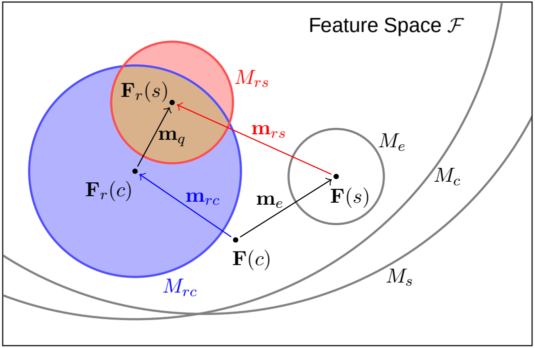
$$\mathbf{m}_{rc} = \text{median}\left[\mathbf{F}(rc) - \mathbf{F}(c)\right],$$
$$M_{rc} = \text{median}\left[\|\mathbf{F}(rc) - \mathbf{F}(c) - \mathbf{m}_{rc}\|\right]$$

Proposed Model



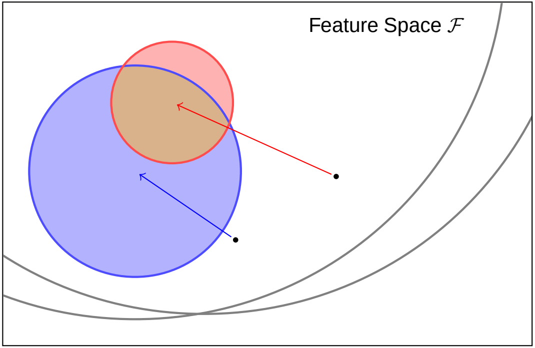
$$\mathbf{m}_{rc} = \text{median}\left[\mathbf{F}(rc) - \mathbf{F}(c)\right],$$
$$M_{rc} = \text{median}\left[\|\mathbf{F}(rc) - \mathbf{F}(c) - \mathbf{m}_{rc}\|\right]$$

Proposed Model



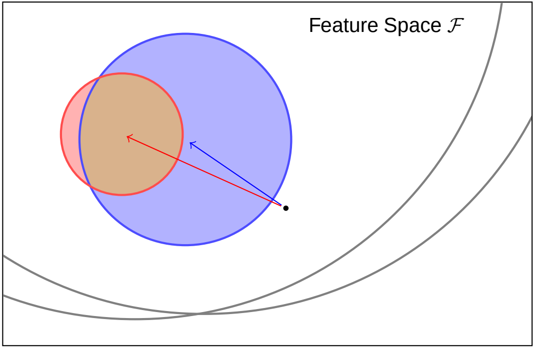
$$\begin{aligned}\mathbf{m}_{rc} &= \text{median}\left[\mathbf{F}(rc) - \mathbf{F}(c)\right], \\ M_{rc} &= \text{median}\left[\|\mathbf{F}(rc) - \mathbf{F}(c) - \mathbf{m}_{rc}\|\right]\end{aligned}$$

Proposed Model



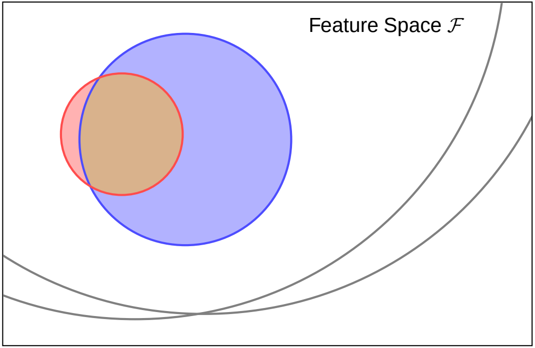
$$\begin{aligned}
\mathbf{m}_{rc} &= \text{median}\left[\mathbf{F}(rc) - \mathbf{F}(c)\right], \\
M_{rc} &= \text{median}\left[\|\mathbf{F}(rc) - \mathbf{F}(c) - \mathbf{m}_{rc}\|\right]
\end{aligned}$$

Proposed Model



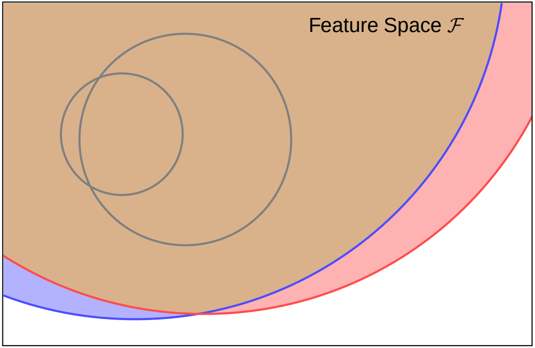
$$\begin{aligned}\mathbf{m}_{rc} &= \text{median}\left[\mathbf{F}(rc) - \mathbf{F}(c)\right],\\ M_{rc} &= \text{median}\left[\|\mathbf{F}(rc) - \mathbf{F}(c) - \mathbf{m}_{rc}\|\right]\end{aligned}$$

Proposed Model



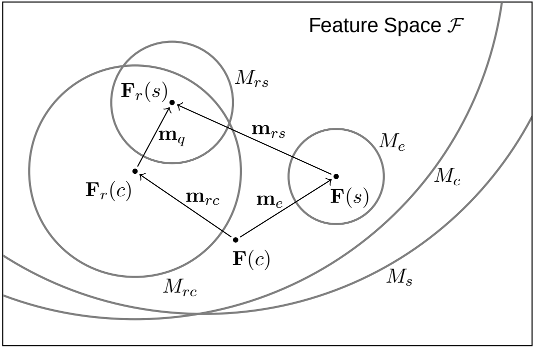
$$
\begin{aligned}
\mathbf{m}_{rc} &= \mathrm{median}\left[\mathbf{F}(rc) - \mathbf{F}(c)\right], \\
M_{rc} &= \mathrm{median}\left[\|\mathbf{F}(rc) - \mathbf{F}(c) - \mathbf{m}_{rc}\|\right]
\end{aligned}
$$

Proposed Model



$$
\begin{aligned}
\mathbf{m}_{rc} &= \operatorname{median}\left[\mathbf{F}(rc) - \mathbf{F}(c)\right], \\
M_{rc} &= \operatorname{median}\left[\|\mathbf{F}(rc) - \mathbf{F}(c) - \mathbf{m}_{rc}\|\right]
\end{aligned}
$$

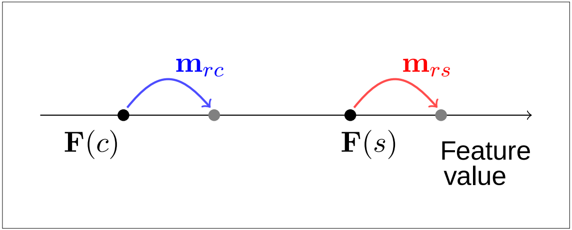
Proposed Model



$$
\begin{aligned}
\mathbf{m}_{rc} &= \operatorname{median}\left[\mathbf{F}(rc) - \mathbf{F}(c)\right], \\
M_{rc} &= \operatorname{median}\left[\|\mathbf{F}(rc) - \mathbf{F}(c) - \mathbf{m}_{rc}\|\right]
\end{aligned}
$$

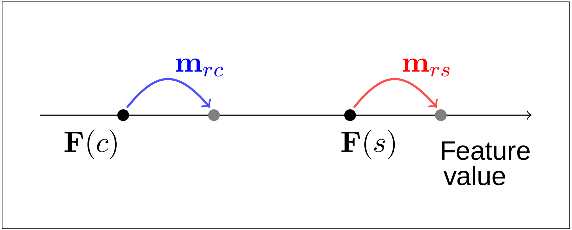
Parallel Reference

- $\mathbf{m}_{rc} \approx \mathbf{m}_{rc}$, $M_{rc} \approx M_{rs}$
- Calibration can be seen as a constant feature-space shift
- Calibration causes failure of steganalysis



Parallel Reference

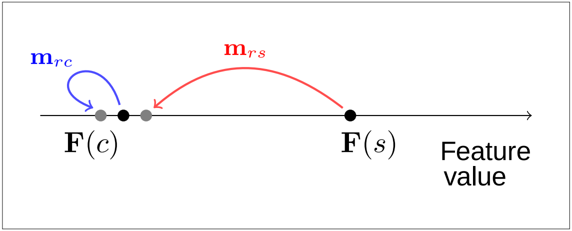
- $\mathbf{m}_{rc} \approx \mathbf{m}_{rc}$, $M_{rc} \approx M_{rs}$
- Calibration can be seen as a constant feature-space shift
- Calibration causes failure of steganalysis



- Experiments: observed often for YASS (robustness!)

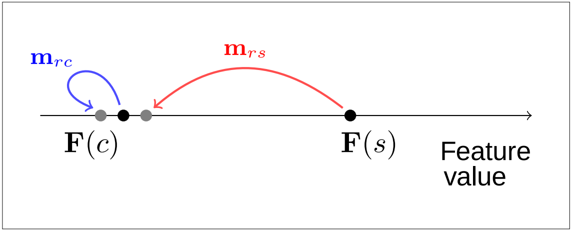
Cover Estimate

- Both $\mathbf{m}_{rc}$ and $\mathbf{m}_{rs}$ are close to cover feature $\mathbf{F}(c)$
- This stood behind the original idea of calibration
- Stego-image feature must differ from cover-image feature



Cover Estimate

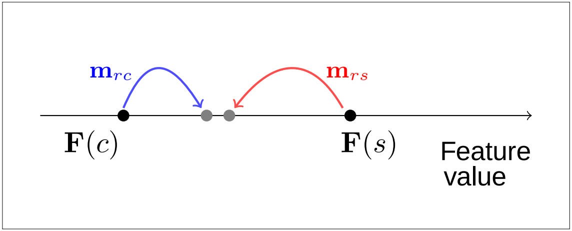
- Both $\mathbf{m}_{rc}$ and $\mathbf{m}_{rs}$ are close to cover feature $\mathbf{F}(c)$
- This stood behind the original idea of calibration
- Stego-image feature must differ from cover-image feature



- Experiments: easier to observe for larger payloads

Eraser

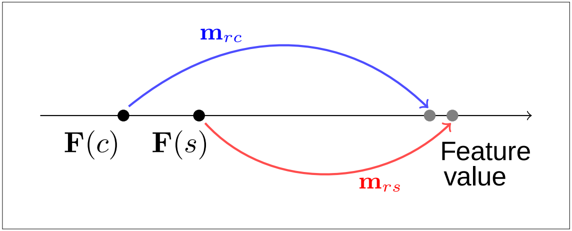
- Reference cover and stego features are close to each other
- Mapping r *erases embedding changes*
- $\mathbf{F}(c) \rightarrow \mathbf{F}(s)$ must be consistent in terms of direction



- Experiments: more frequent than cover estimate

Eraser

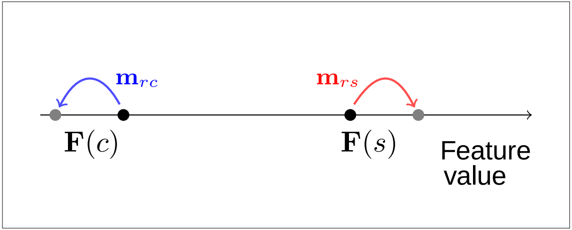
- Reference cover and stego features are close to each other
- Mapping r *erases embedding changes*
- $\mathbf{F}(c) \to \mathbf{F}(s)$ must be consistent in terms of direction



- Experiments: more frequent than cover estimate
- Different example: predictor in WS steganalysis

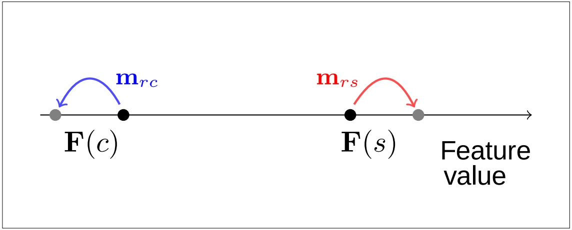
Divergent Reference

- $\mathbf{m}_{rc}$ must be different from $\mathbf{m}_{rs}$
- This situation essentially covers some of the previous ones
- Works even when $\mathbf{F}(c) = \mathbf{F}(s)$



Divergent Reference

- $\mathbf{m}_{rc}$ must be different from $\mathbf{m}_{rs}$
- This situation essentially covers some of the previous ones
- Works even when $\mathbf{F}(c) = \mathbf{F}(s)$



- Experiments: most frequent scenario
- Interesting example: histogram of zeros for JSteg

Lessons Learned

- Calibration does not have to approximate cover. Still, it might be benefitial to calibrate.

Lessons Learned

- Calibration does not have to approximate cover. Still, it might be benefitial to calibrate.
- Several different mechanisms may be responsible for a positive effect of calibration.

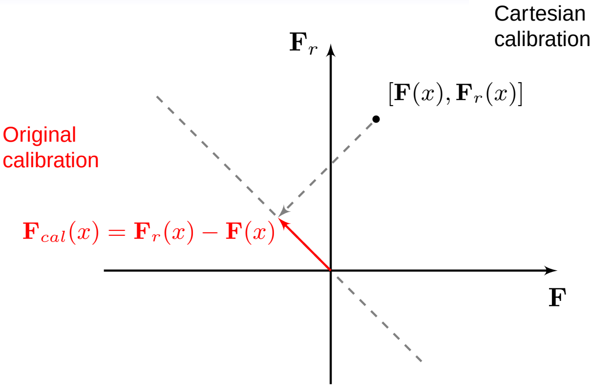
Lessons Learned

- Calibration does not have to approximate cover. Still, it might be benefitial to calibrate.
- Several different mechanisms may be responsible for a positive effect of calibration.
- Calibration may have a catastrophically negative effect on steganalysis as well (parallel reference).

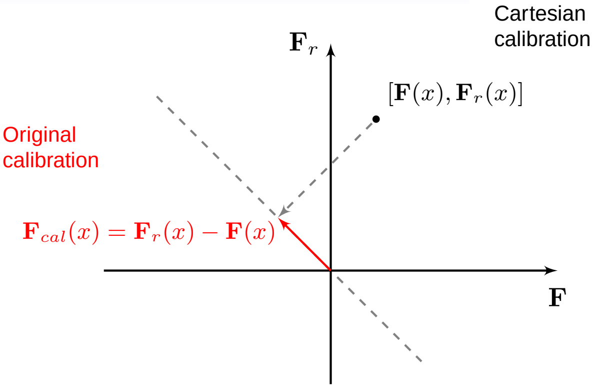
Lessons Learned

- Calibration does not have to approximate cover. Still, it might be benefitial to calibrate.
- Several different mechanisms may be responsible for a positive effect of calibration.
- Calibration may have a catastrophically negative effect on steganalysis as well (parallel reference).
- How to prevent steganalysis from such failures?

Different Point of View



Different Point of View



How well does Cartesian calibration perform in practice?

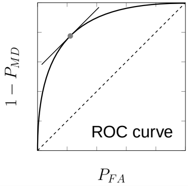
Cartesian Calibration Improves Steganalysis

Algorithm	bpac	P_E $\mathbf{F}$	$\mathbf{F}_r - \mathbf{F}$	$[\mathbf{F}_r, \mathbf{F}]$
nsF5	0.05	0.361	0.360	**0.331**
	0.10	0.202	0.218	**0.177**
	0.15	0.100	0.094	**0.077**
	0.20	0.048	0.040	**0.036**
Jsteg	0.02	0.097	0.132	**0.083**
	0.03	0.042	0.051	**0.032**
	0.04	0.022	0.021	**0.018**
	0.05	0.015	0.013	**0.010**
Steghide	0.02	0.114	0.127	**0.083**
	0.03	0.055	0.056	**0.043**
	0.04	0.031	0.031	**0.024**
	0.05	0.021	0.015	**0.011**
MME3	0.05	0.309	0.310	**0.277**
	0.10	0.187	0.207	**0.165**
	0.15	0.130	0.149	**0.107**
	0.20	0.023	0.017	**0.012**

Algorithm	bpac	P_E $\mathbf{F}$	$\mathbf{F}_r - \mathbf{F}$	$[\mathbf{F}_r, \mathbf{F}]$
JPHS	0.05	0.306	0.100	**0.094**
	0.10	0.160	0.066	**0.054**
	0.15	0.076	0.034	**0.022**
	0.20	0.039	0.014	**0.006**
YASS 1	0.110	0.133	0.317	**0.113**
YASS 2	0.051	0.179	0.347	**0.164**
YASS 3	0.187	0.102	0.121	**0.082**
YASS 4	0.118	0.120	0.303	**0.109**
YASS 5	0.159	0.075	0.241	**0.064**
YASS 6	0.032	0.269	0.342	**0.258**
YASS 7	0.078	0.244	0.298	**0.225**
YASS 8	0.138	0.211	0.251	**0.180**

Reported values of P_E are medians over 5 runs.

$$P_E = \min \tfrac{1}{2}(P_{FA} + P_{MD})$$



Calibration Revisited

- Shed more light on how, why, and when calibration works
- Introduced a new framework capable of both quantitatively and qualitatively capture behaviour of calibration in the feature space
- Supported our findings experimentally
- Proposed an improved way of calibration
  - Extractor of Cartesian-calibrated 274 merged features available

http://dde.binghamton.edu/ccmerged