IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Steganalysis using Partially Ordered Markov Models

Dr. Jennifer Davidson[a]

Jaikishan Jalan[b]

[a] Dept. of Mathematics, Dept. of Electrical & Computer Engineering
[b] Former student, Dept. of Computer Science

Iowa State University, Ames, IA

Overview of Talk

- Use of stochastic models for features in steganalysis
  - Feature selection in steganalysis: informal approach
  - Motivation to use stochastic models for steganalysis
- Define *partially ordered Markov models* and give a general problem solution for creating features for steganalysis using POMMs
- Experiments
  - Five JPEG embedding algorithms
  - Three additional steganalyzers
- Results
- Future research

Statistical steganalysis feature development

- The image *A* is modeled as a collection of random variables (r.v.s) with a probability distribution $P(A)$
- A vector $F(A) = (f_1(A), \ldots, f_n(A))$ of *feature values* is calculated from the image pixels, where *n* << the number of pixels in the image
- The functions $\{f_i(A)\}_{i=1}^{n}$ are chosen by the steganalyst using domain knowledge
- Features are selected to exploit known differences between stego and cover characteristics and used in targeted or blind pattern recognition systems

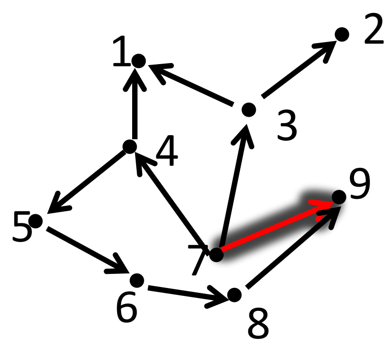
Statistical steganalysis feature development

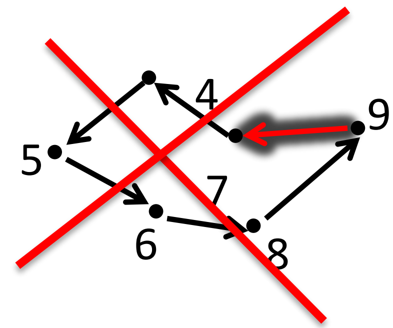
- Previously used probability distributions for features in steganalysis
  - Generalized Gaussian distribution for modeling mode histograms of DCT coefficients
  - Markov chains for pixels adjacent in the DCT domain and in the spatial domain (Shi et al. 2007, Pevný 2009)
  - We were motivated to investigate other stochastic models that could provide theoretical foundation for modeling steganographic changes to image

Acyclic directed graphs and partially ordered sets

- **Definition**. Let (*V,E*) be a finite acyclic directed graph:

Edges, (tail,head):
(4,1)
(4,5)
(7,9), etc.





No cycles of edges

- **Definition**. Let (*V*, ≤) be a *partially ordered set (poset)* where ≤ is a binary operation on *V*:

$$1.\ w \le w \text{ for all } w \in V \ \text{(reflexivity)}$$

$$2.\ w \le x,\ x \le y \Rightarrow w \le y \ \text{(transitivity)}$$

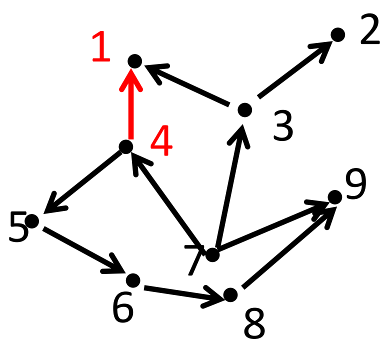
$$3.\ \text{If } w \le x \text{ and } x \le w \text{ then } w = x \ \text{(anti-symmetry)}$$

- Example: *V* = all subsets of a set, $\le = \subseteq$ (set inclusion)

Acyclic directed graphs and partially ordered sets

- **Def.** For V_i, V_j ε $(V,\leq)$, V_i *is covered by* V_j if $V_i < V_j$ and $V_i < V_k < V_j$ for no k.
- Given graph (V,E), construct poset $(V,\leq)$ by
  - (i,j) ε E implies V_i is covered by V_j in $(V,\leq)$.
  - This defines a partial order on V
  - In this case we write $V_i < V_j$

- Edge (4,1) defines the relation between V_4 and V_1, so V_4 is covered by V_1 and $V_4 < V_1$

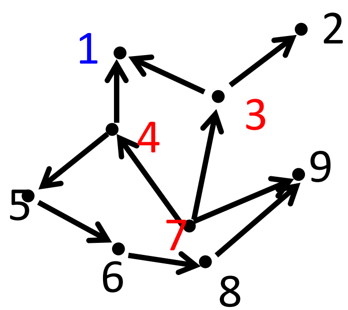


Definitions

$$cone(B) = \{C \in V : C \leq B, C \neq B\}$$

$$adj_{\leq}(B) = \{C : (C,B) \in E\} = \text{ all elements covered by } B$$

$$L^0 = \text{ set of minimal elements in } V \text{ (no edges incoming to vertices)}$$



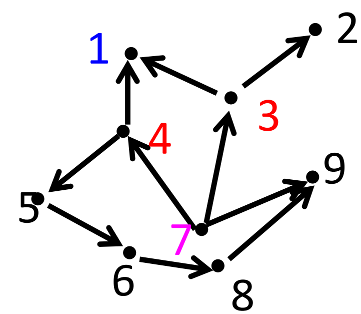
$$cone(V_1) = \{V_4, V_7, V_3\}$$

V_4 is covered by V_1

V_7 is covered by V_4

V_3 is covered by V_1

$$L^0 = \{V_7\}$$



$$adj_{\leq}(V_1) = \{V_4, V_3\}$$

V_4 is covered by V_1

V_3 is covered by V_1

Definition of partially ordered Markov model

- **Def.** Let *V* be a set of random variables and *B* ε *V*, where *V* is a finite acyclic digraph (*V*,*E*) with poset (*V*,≤). Let

$$Y_B = \{C : B \text{ and } C \text{ are not related under } \leq\}$$

Then (*V*,≤) is called a *partially ordered Markov model* (POMM) if for any $B \in V \setminus L^0$ and any subset $U_B \subseteq Y_B$ we have

$$P(B \mid cone(B, U_B)) = P(B \mid adj_{\leq}(B))$$

- The lower adjacent neighbors describe the "Markovian" property of the model

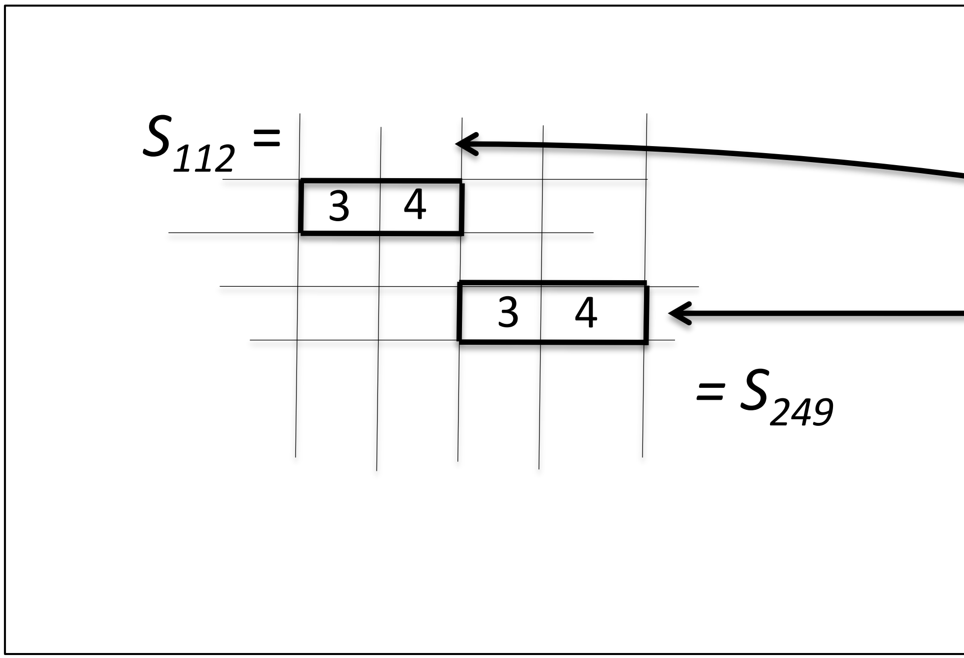
Our interest

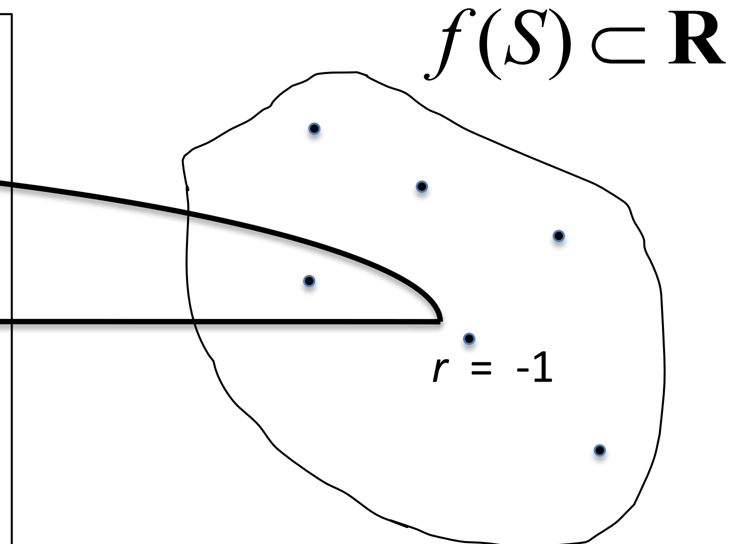
- $\boldsymbol{A} = \{A_{ij} : 1 \le i \le \mathrm{N}, 1 \le j \le \mathrm{M}\}$: set of r.v.s on array
- $\boldsymbol{S} = \{S_1, \ldots, S_t\}$ is a collection of subsets of r.v.s in $\boldsymbol{A}$ where each S_k is an ordered set
- Example: $\boldsymbol{S}^h$, $S_1^h = \{A_{11}, A_{12}\}$, $S_2^h = \{A_{12}, A_{13}\}$, etc.
- Introduce a function $f : \boldsymbol{S} \rightarrow \mathrm{R}$ the set of real numbers that gives quantifying information about the subsets
- Example: $$f(w_1, w_2) = w_1 - w_2$$
$$f(S_i^h) = f(A_{j,k}, A_{j,k+1}) = A_{j,k} - A_{j,k+1}$$

Our interest

- Create an acyclic digraph: $V = S \cup f(S), \quad E = \{E_i\}$
  where $E_i = (f(S_i), S_i)$ and has tail on $f(S_i)$ and head on S_i
- We call this the *function-subset acyclic digraph,* or *f-S*
- We use this acyclic digraph to construct a sequence of POMMs whose conditional probabilities are used as features
- If *f* is a useful function for the steganalyst, then the quantity $P(S_k | f(S_k))$, which is a measure of the frequency of occurrence of the pre-image of $f(S_k)$, can be used to distinguish between cover and stego images
- This is the motivation for using the *f-**S*** partial order/acyclic digraph as defined earlier

Diagram of *f*-*S* acyclic digraph





$$r = f(a_{ij}, a_{i,j+1}) = a_{ij} - a_{i,j+1}$$
$$= f(a_{kh}, a_{k,h+1}) = a_{kh} - a_{k,h+1}$$

- $P(S_k^h \mid f(S_k^h))$ measures frequency of occurrence of | 3 | 4 | given the difference value of -1
- $P(* \mid *)$ defines the POMM associated with thi horizontal $f - S$ model

Features

- Collect information in four directions: $\boldsymbol{S}^h$, $\boldsymbol{S}^v$, $\boldsymbol{S}^d$, $\boldsymbol{S}^m$
- Create a POMM for each of the four directions P^h, P^v, P^d, P^m
- Calculate conditional probabilities $P^*(S^*_k|f(S^*_k))$, * ε {h,v,d,m} using the quantized DCT array of values thresholded by value T
- Each direction gives a $(2T + 1)$ x $(2T + 1)$ feature matrix $F^*(w,z) = P^*(w,z|f(w,z)$
- Average over four directions to get $(2T + 1)^2$ intrablock feature values:

$$F^{\text{intra}}(w,z) = \frac{1}{4} \sum_{* \in \{h,v,d,m\}} P^*(w,z \mid w-z)$$

Features

- Also construct POMMs using interblock values from quantized DCT array in a similar manner
- There are 8*8 = 64 mode arrays
- Average over the 64 feature matrices to get another $(2T + 1)^2$ feature values

$$F^{\text{inter}}(w,z) = \frac{1}{64} \sum_{* \in \{h,v,d,m\}} P^{*}(w,z \mid w - z)$$

- Total number of features = $2*(2T + 1)^2$ and it depends on the value T

Experiments

- Used four databases: BOWS2 (10,000 images), a camera database (3164), Corel (8185), NRCS (2375)
- Created training and testing data from these DB
- Used three additional state of the art steganalyzers:
  - Shi's Markov model using intrablock values (Shi et al, 2007); "Markov324" (324 features)
  - Shi's Markov model using both intra and interblock values (Shi et al, 2008); "Markov486"
  - Pevný merged model with extended DCT features plus calibrated Markov values from Markov324 (Pevný et al., 2007) "Merged"

Experiments

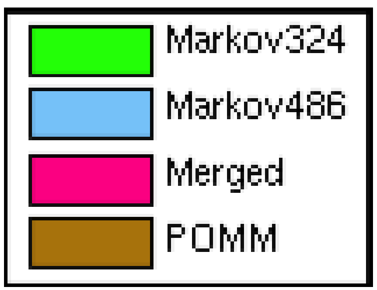
- Classifier: soft margin support vector machine with Gaussian kernel and grid-search method to determine training parameters (LIBSVM)
- Five embedding algorithms at four embedding rates each: Jsteg, OutGuess, F5, StegHide, and JPHide; bpnz = 0.05, 0.1, 0.2, 0.4 (except last one was omitted for OutGuess)
- Calculated detection accuracy using binary classifiers

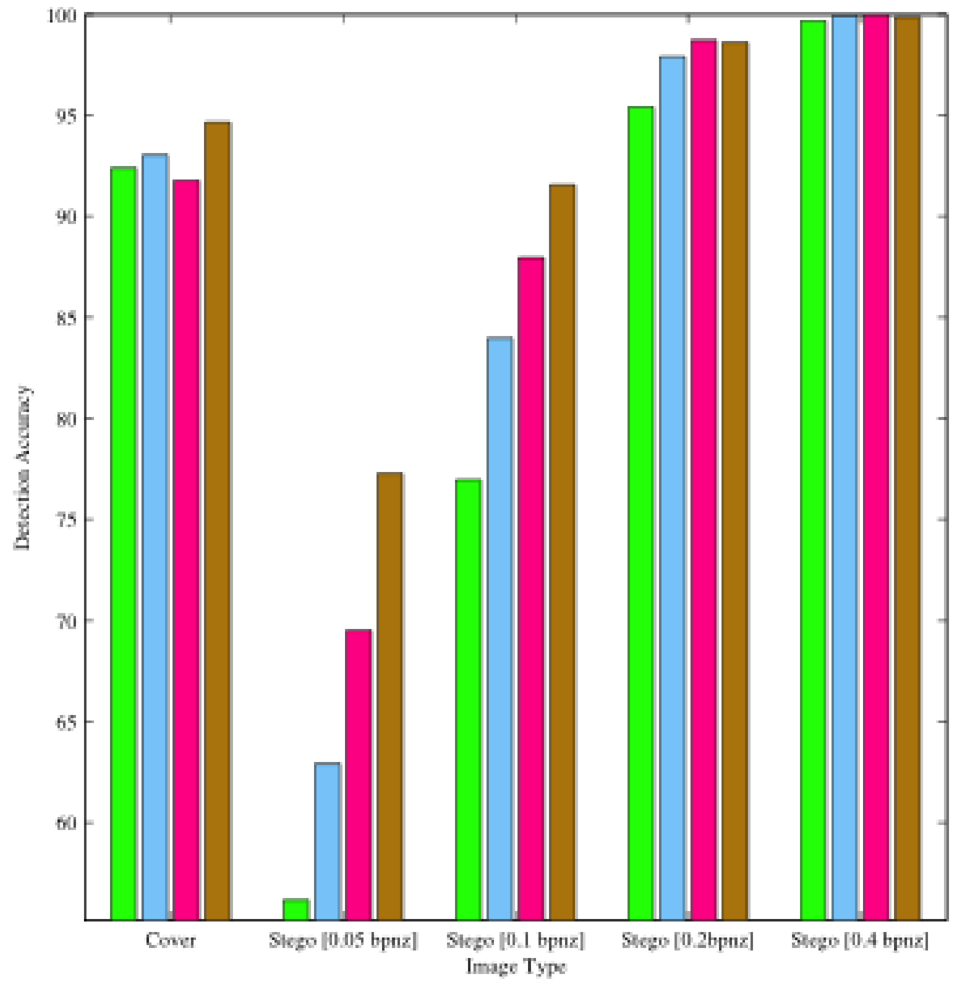
Method

- Tried five values for T: $T = 1, 2, 3, 4, 5$
- Overall best detection accuracy was achieved for $T = 3$; this gives a total of 98 features
- Developed binary classifiers for each case, total of 24 binary classifiers
- Half of data was used for training, other half for testing
- Tested each database separately

Results and discussion

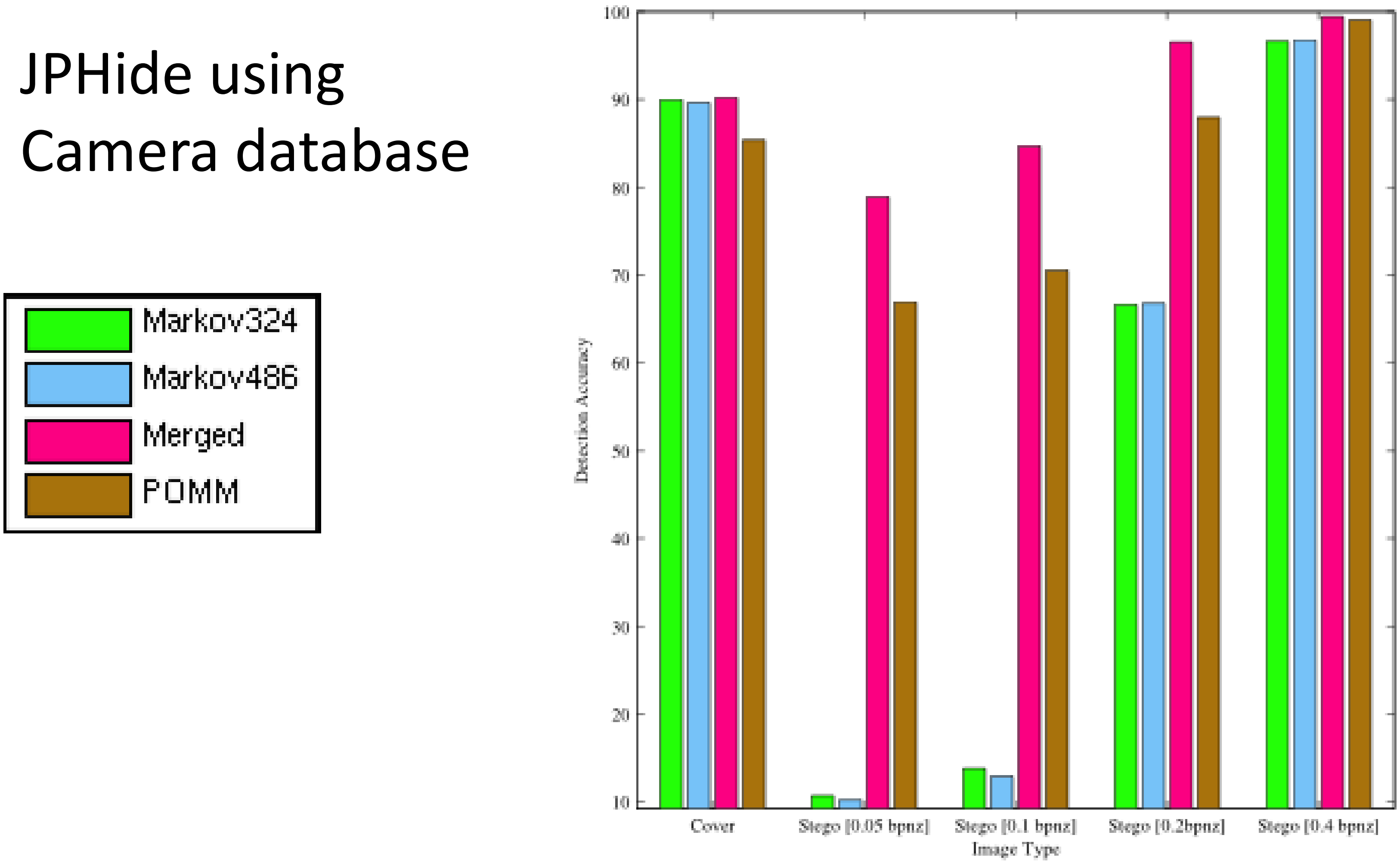
StegHide using BOWS2 database





Results and discussion

JPHide using Camera database



Discussion

- POMMs perform better almost without exception than either Markov model particularly at the lower embedding rates
- POMM performed better than Merged for Outguess and StegHide across all databases and all embedding levels
- Merged performed better than POMMs at lower embedding rates for F5 and JPHide across all databases
- At highest levels of embedding all algorithms performed similarly well

Discussion

- Another way to measure performance
- Criterion: Performed >greater than 1% better than any detector, or within 1% of top detector, on cover, 0.05 and 0.1 embedding rates (most difficult to detect)
- POMM: 17% of the time
- Merged: 18% of the time
- Other two steganalyzers were far beneath that

Conclusion

- Introduction of new modeling tool to measure embedding changes
- Allow steganalyst to create functions to detect changes
- Can use other measures of the probability distribution for features such as moments – mean, variance, etc.
- Possibility of using joint pdf in detection (MLE), as joint pdf is computationally efficient
- 98 features give equivalent detection to Merged steganalyzer
- Current and future research: double compression detector for use in police forensic GUI software
- Use of POMMs for spatial embedding detection
- Use of other functions f and subsets $\boldsymbol{S}$