Video streaming modelling via Hidden Markov Processes

Ing. Lorenzo Rossi – Second Year Ph.D Student

INFOCOM Dep.t Università di Roma “La Sapienza”

Outline

- Video streaming services
- H.264 Switching Pictures in video streaming
- Model description
- Model assessment
- Simulation results
- Conclusion and Further Works

Video streaming services - 1

- Video traffic is historically generated in the framework of two services:
  - **broadcasting services – typically VBR**

    large available bandwidth and medium-to-high required quality, with loose coding delay requirements
  - **conversational services – typically CBR**

    characterized by significantly reduced bandwidth and quality, with severe coding delay requirements

- Video streaming services are characterized by
  - contemporary download and playout of the video content
  - short initial buffering delay.

- The typical architecture for video streaming consists of a server that transmits data extracted from
  - pre-encoded sequences
  - stored along with meta-data, such as:
    - packetization instructions (Hint Tracks)
    - session description informations

Video streaming services - 2

- The traffic must be adapted in order to comply network variations:
  - peak and average bandwidth
  - decoding buffer size
  - physical channel changes

- Server should react to adapt the rate of the transmitted video sequence to accommodate these variations

- Rate adaptation achieved without re-encoding but **switching dynamically between pre-encoded video streams** of different bandwidth and quality

  => *streams must have periodical access points (e.g. Intra frames)*

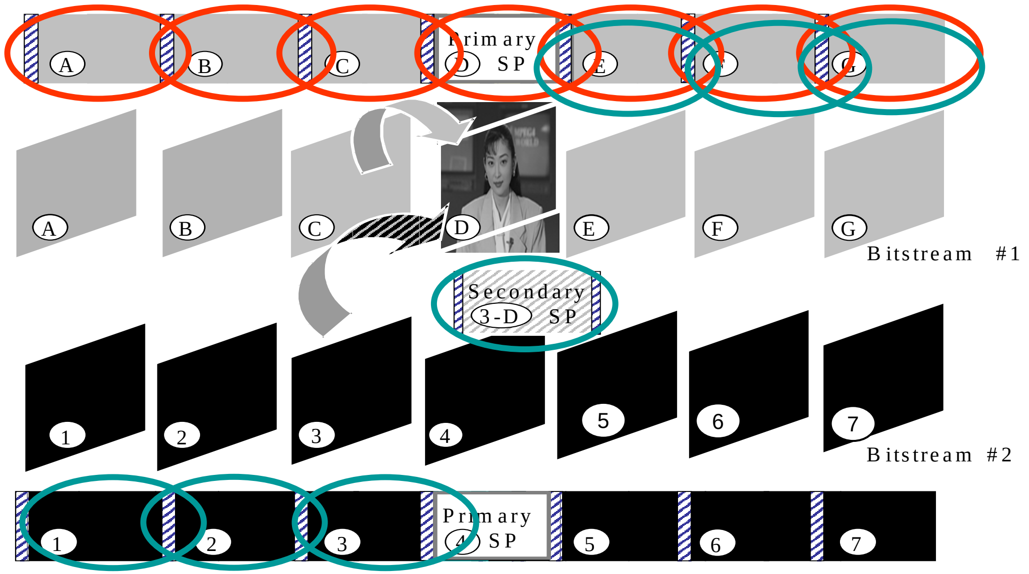
- H.264 enables perfect frame reconstruction even when different reference frames are used for prediction by means of Switching Pictures (SP)

- SP pictures provide an access point to different bitstreams, at a lower coding cost than classical Intra coded pictures

H.264 video streaming with SP frames

- SP pictures are placed at periodical interval in the video stream
  - **primary SP:** used when the stream is not going to change
  - **secondary SP:** allows to switch between two streams

- Example: a server is transmitting the same video content to two users.



Traffic transmitted to User 1 (bitstream #1)

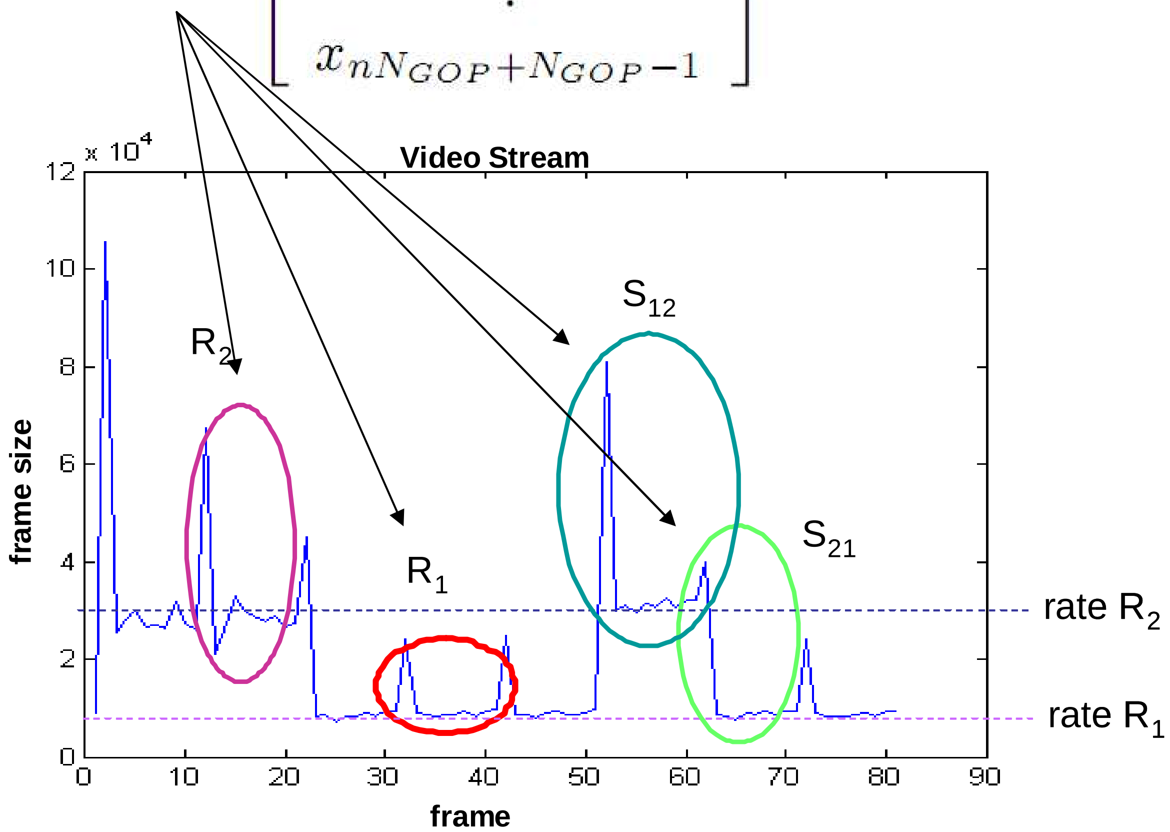
Traffic transmitted to User 2 (Switching from Bitstream #2 to #1)

Aim of the work

- Video is typically the most expensive media in bandwidth allocation, then a tight modeling of a video source is desired to predict the impact of the video traffic on the network
  - A theoretical model offers analytical framework for admission control strategies
  - A large amount of synthetic traffic can be generated according to the model and may be employed in the phase of network design; in fact measuring a huge amount of video data traces, in a variety of encoding conditions, is a heavy task, especially when bandwidth adaptation is required
- We address the modeling of a H.264 video streaming source (in terms of the frames size sequence) that employs SP frames
- The video is arranged as a group of consecutive frames (Group Of Pictures, GOP) beginning with a random access unit
- We design an optimal strategy for parameter set estimation from a real video sequence according to a Maximum Likelihood estimation criterion

Data vectors representing GOP frame sizes

$$\mathbf{x}[n] \stackrel{\text{def}}{=} \begin{bmatrix} x_{nN_{GOP}} \\ x_{nN_{GOP}+1} \\ \vdots \\ x_{nN_{GOP}+N_{GOP}-1} \end{bmatrix}$$



Hidden Markov Process - 1

- We model the video source by a Hidden Markov Process (HMP)
  - each state of the chain represents a different kind of GOP
  - the Markov chain emits data vectors representing GOP frame sizes
  - state-dependent conditional observations' probability density is found by modulating the mean and the covariance matrix of a multivariate white Gaussian process

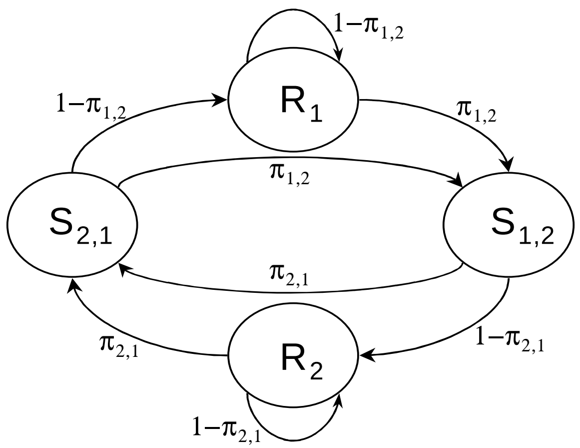
- First and second order moments of the HMP have been evaluated

- ML parameter set estimation is performed by means the Expectation-Maximization (EM) algorithm

- Experimental results show that, despite of a few simplifying modeling assumptions, the HMP model well captures the statistical characteristics of the source at a GOP Layer and allows to apply the Expectation-Maximization algorithm

Hidden Markov Process - 2

- $\lambda_n = \lambda$ is the state sequence of the Markov chain which models the GOP sequence
  - Each state of the chain represents a different kind of GOP
- e[n] is a realization of a white, Gaussian, N_{GOP} -dimensional random process independent from the state sequence λ_n



$$E\{\mathbf{e}[n]\} = \mathbf{0}$$

$$E\{\mathbf{e}[n]\mathbf{e}[n-m]^{\mathrm{T}}\} = I \cdot \delta[m]$$

$$\mathbf{x}[n] = \Sigma_\lambda \mathbf{e}[n] + \mathbf{c}_\lambda$$

- It follows that the state conditioned density probability is normal:

$$p(\mathbf{x}[n] | \lambda_n = \lambda) = \mathcal{N}\left(\mathbf{c}_\lambda, \Sigma_\lambda \Sigma_\lambda^{\mathrm{T}}\right)$$

HMP expected values

- If the chain is stationary, irreducible and aperiodic, the model is ergodic

$$\mu_{\mathbf{X}} = \sum_{\lambda=1}^{N_s} p_\lambda \mathbf{c}_\lambda$$

$$R_{\mathbf{X}}[m] = \sum_{\lambda_1=1}^{N_s} \sum_{\lambda_2=1}^{N_s} p_{\lambda_1} ||\Pi^m||_{\lambda_1\lambda_2} \mathbf{c}_{\lambda_2} \mathbf{c}_{\lambda_1}^T + \delta[m] \sum_{\lambda=1}^{N_s} p_\lambda \Sigma_\lambda \Sigma_\lambda^T$$

- The assumptions we made about the source permit to derive a ML estimation procedure

- Let us denote Θ the parameter set

$$\Theta \stackrel{\mathrm{def}}{=} \{\Pi, \Sigma_1, \ldots, \Sigma_{N_s}, \mathbf{c}_1, \ldots, \mathbf{c}_{N_s}\}$$

EM algorithm

- All the existing models estimate parameter set through heuristic procedures

- Here we develop a ML parameter set estimation through EM Algorithm

A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood From Incomplete Data via the EM Algorithm", J. Roy. Statist. Soc. B, 1977.

- $\mathbf{x}_0^{N-1} \stackrel{\text{def}}{=} \{\mathbf{x}[n], n = 0 \cdots N-1\}$ the observed video sequence

- $\lambda_0^{N-1} \stackrel{\text{def}}{=} \{\lambda_n, n = 0 \cdots N-1\}$ the (unknown) state sequence

- EM algorithm consists in calculating an auxiliary function and then maximizing with respect to its argument

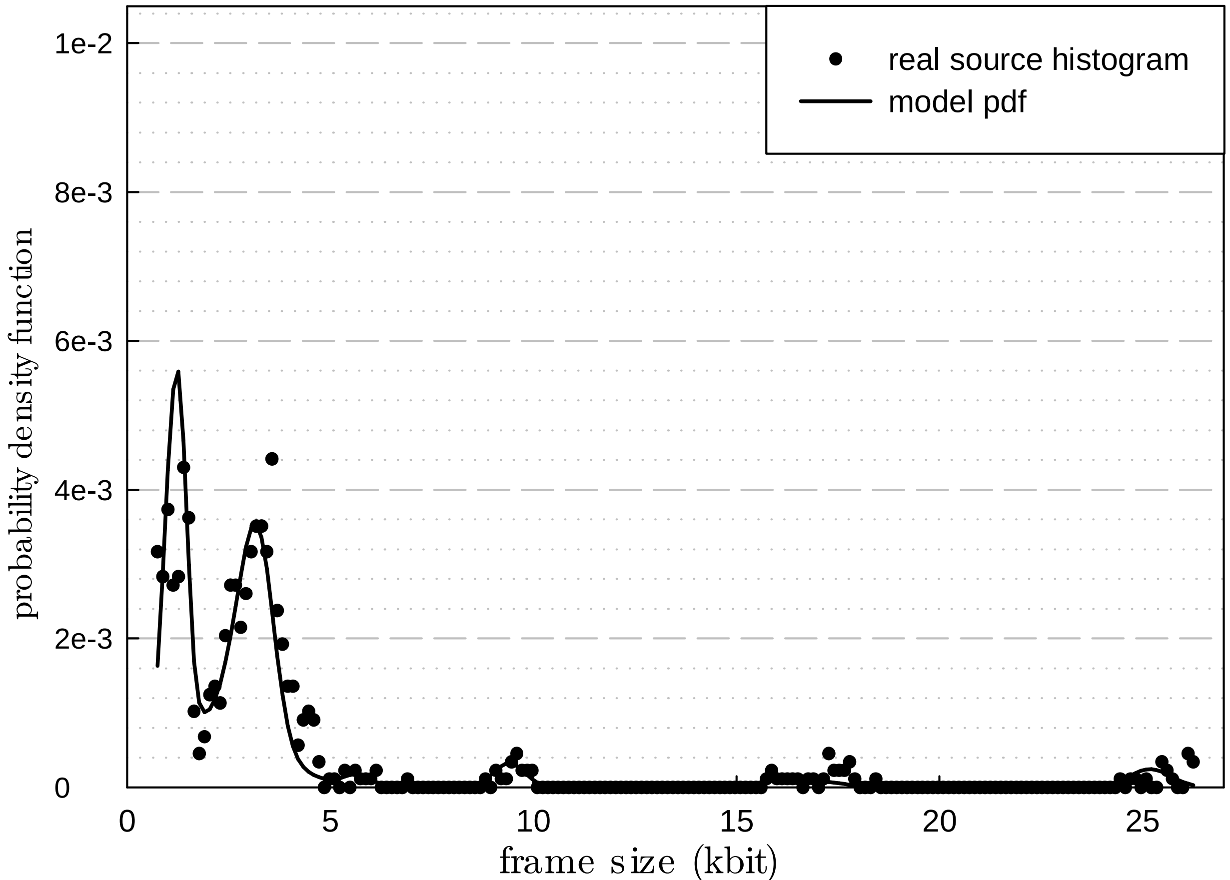
$$Q(\Theta, \Theta^{(k)}) \stackrel{\text{def}}{=} E\{\log(p(\lambda_0^{N-1}, \mathbf{x}_0^{N-1}|\Theta)) | \mathbf{x}_0^{N-1}, \Theta^{(k)}\}$$

$$\Theta^{(k+1)} = \arg\max_{\Theta} Q(\Theta, \Theta^{(k)})$$

- Simulations show that EM algorithm converges in few iterations

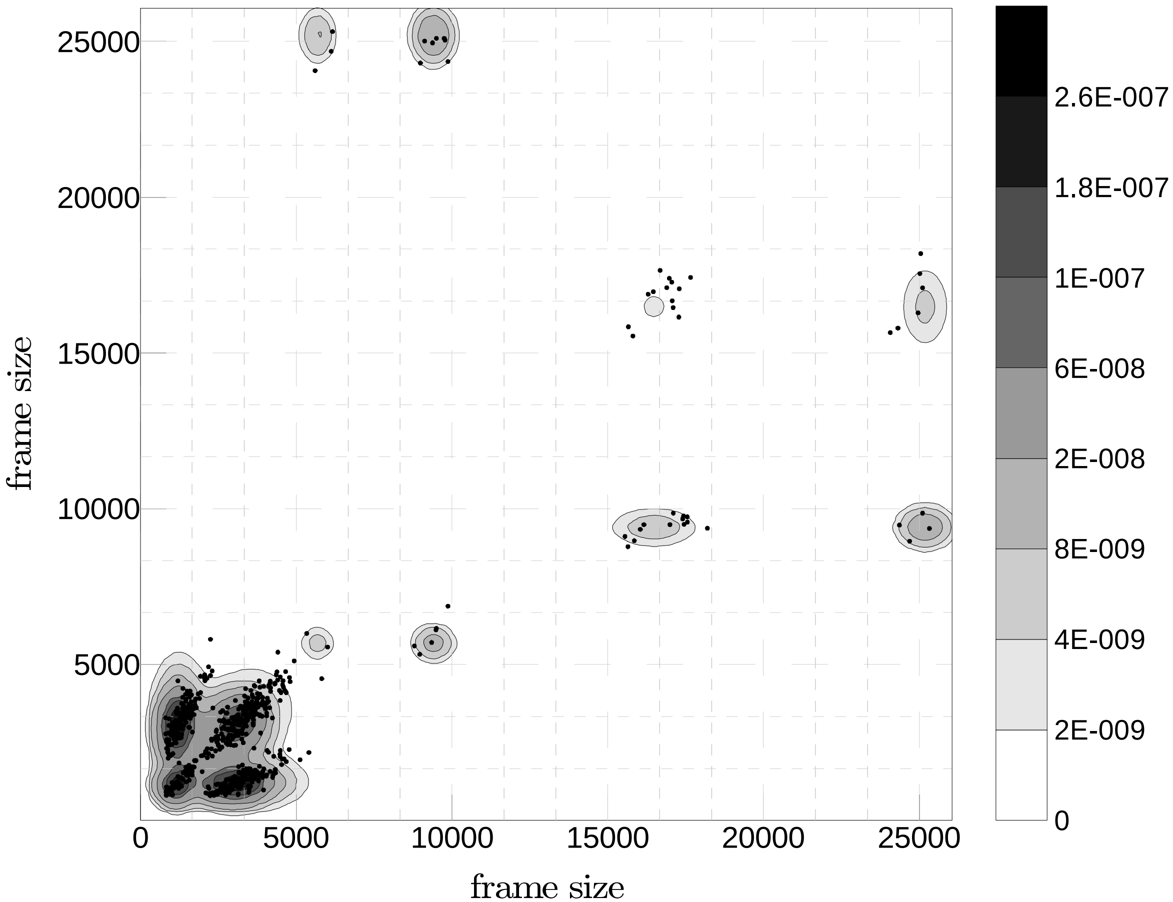
Model Assessement - 1

- Model parameter set is estimated by observing only a segment (the initial GOPs) of a real video sequence (Bridge far)
- Histograms of the real video sequence are compared to the model pdfs



Model Assessment - 2

- A bidimensional pdf $p(x^{(HMP)}[m], x^{(HMP)}[m+N_{GOP}])$
- Circles mark the observed pairs for the real video source



Autocorrelation function - 1

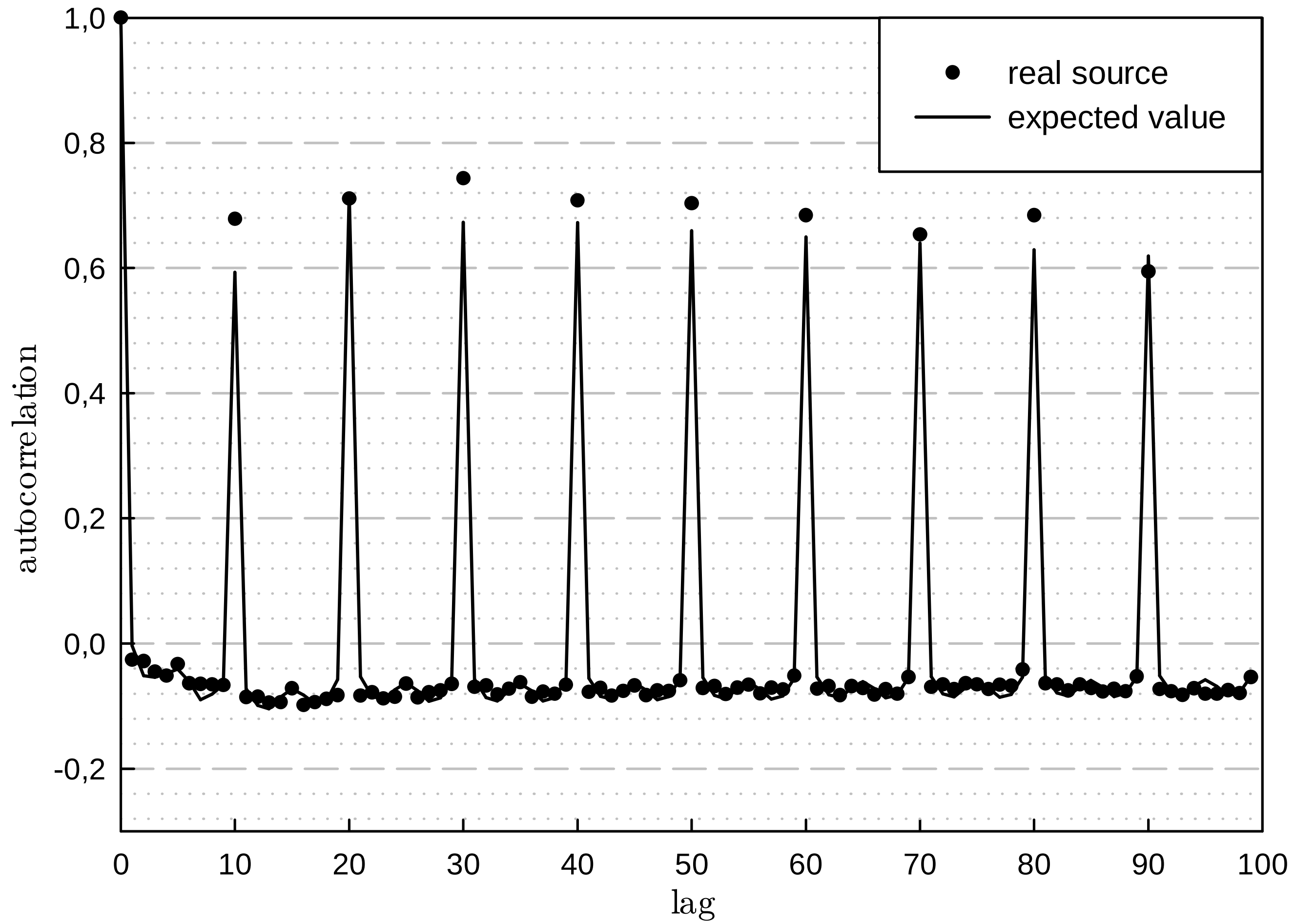
- We compare the autocorrelation function (ACF) of the observed video sequence with the expected value of the model
- $x_n^{(H.264)}$ is the frame size sequence for the real source

$$\rho_{x^{(\mathrm{H.264})}}[k] = \frac{\sum_{m=0}^{M-k-1} \left(x^{(\mathrm{H.264})}[m] - \overline{x}^{(\mathrm{H.264})}\right)\left(x^{(\mathrm{H.264})}[m+k] - \overline{x}^{(\mathrm{H.264})}\right)}{\sum_{m=0}^{M-1} \left(x^{(\mathrm{H.264})}[m] - \overline{x}^{(\mathrm{H.264})}\right)^2} \qquad \overline{x}^{(\mathrm{H.264})} = \frac{1}{N} \sum_{m=0}^{N-1} x^{(\mathrm{H.264})}[m]$$

$$\begin{aligned} E\{\rho_{x^{(\mathrm{HMP})}}[k]\} \approx & \left(\sum_{j=0}^{N_{\mathrm{GOP}}-1} \sum_{n=0}^{\lfloor \frac{M-k-1}{N_{\mathrm{GOP}}} \rfloor -1} \left(\|R_{\mathbf{x}}[\lfloor \frac{k+j}{N_{\mathrm{GOP}}} \rfloor]\|_{(k+j) \bmod N_{\mathrm{GOP}}, j} - \frac{1}{M} \sum_{p=0}^{N_{\mathrm{GOP}}-1} \sum_{l=0}^{G-1} \left(\|R_{\mathbf{x}}[n-l]\|_{j,p} \right.\right.\right. \\ & \left.\left. + \|R_{\mathbf{x}}[n + \lfloor \frac{k+j}{N_{\mathrm{GOP}}} \rfloor - l]\|_{(k+j) \bmod N_{\mathrm{GOP}}, p} \right)\right) + \sum_{j=0}^{(M-k-1) \bmod N_{\mathrm{GOP}}} \left(\|R_{\mathbf{x}}[\lfloor \frac{k+j}{N_{\mathrm{GOP}}} \rfloor]\|_{(k+j) \bmod N_{\mathrm{GOP}}, j} - \frac{1}{M} \sum_{p=0}^{N_{\mathrm{GOP}}-1} \sum_{l=0}^{G-1} \right. \\ & \left(\|R_{\mathbf{x}}[\lfloor \frac{M-k-1}{N_{\mathrm{GOP}}} \rfloor - l]\|_{j,p} + \|R_{\mathbf{x}}[\lfloor \frac{M-k-1}{N_{\mathrm{GOP}}} \rfloor + \lfloor \frac{k+j}{N_{\mathrm{GOP}}} \rfloor - l]\|_{(k+j) \bmod N_{\mathrm{GOP}}, p} \right)\Bigg) \\ & \left. + \frac{M-k}{M^2} \sum_{j=0}^{N_{\mathrm{GOP}}-1} \sum_{n=0}^{G-1} \sum_{p=0}^{N_{\mathrm{GOP}}-1} \sum_{l=0}^{G-1} \|R_{\mathbf{x}}[n-l]\|_{j,p} \right) \cdot \left(G \sum_{i=0}^{N_{\mathrm{GOP}}-1} \|R_{\mathbf{x}}[0]\|_{i,i} - \frac{1}{N_{\mathrm{GOP}} G} \sum_{n=0}^{G-1} \sum_{i=0}^{N_{\mathrm{GOP}}-1} \sum_{m=0}^{G-1} \sum_{j=0}^{N_{\mathrm{GOP}}-1} \|R_{\mathbf{x}}[n-m]\|_{i,j} \right)^{-1} \end{aligned}$$

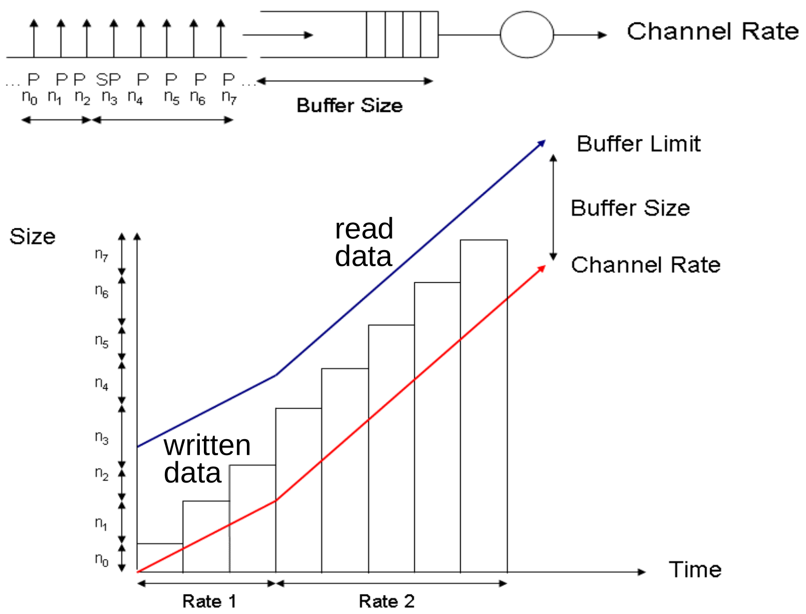
Autocorrelation function - 2

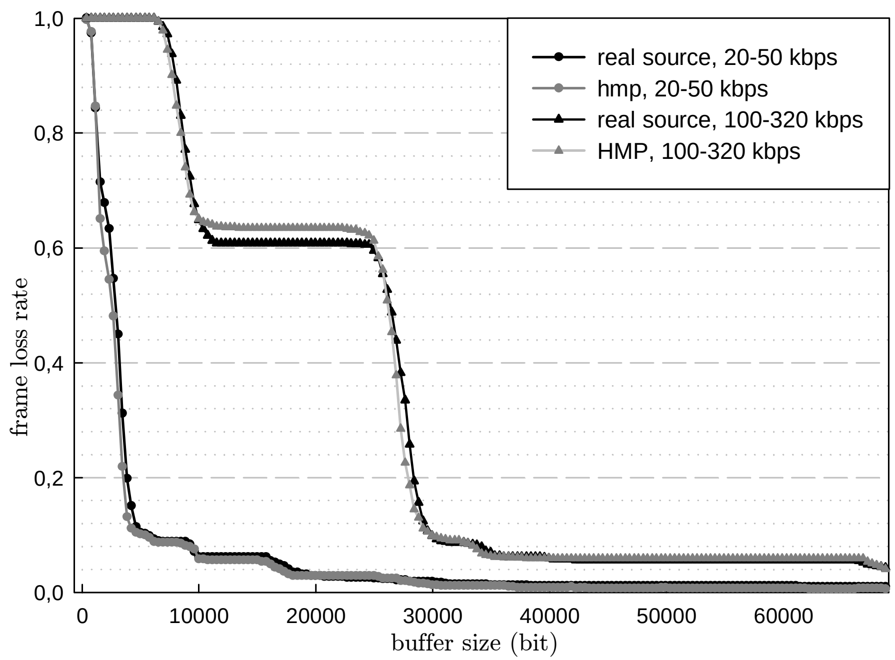
- Black, circle – real source
- Black, dashed – model expected value



Buffer load

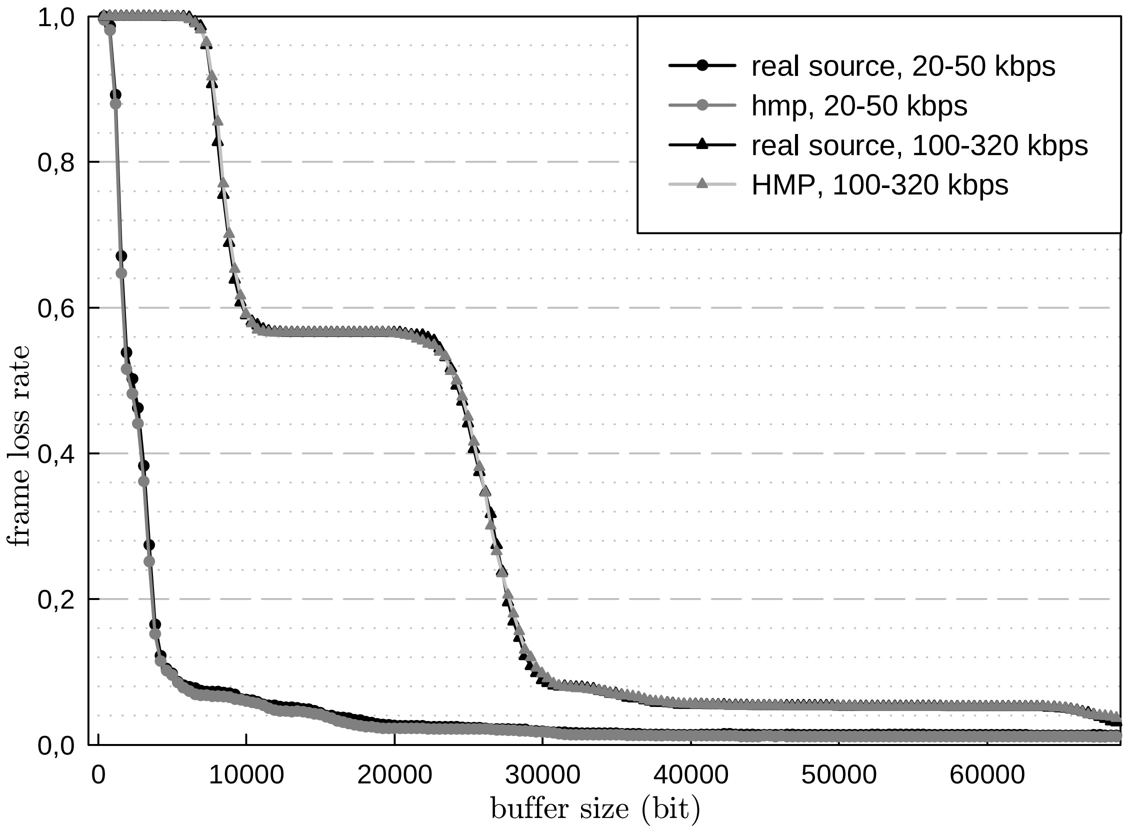
• We compared the buffer load (in terms of frame loss rate) for a real video source and a synthetic source whose parameter set is estimated by observing the first GOPs of the real video traffic





Compound Sequence

- The HMP achieves a good match in terms of frame loss rate also for a real source which exhibites a few scene changes
- Parameter set is estimated by observing the whole sequence



Conclusions and Further works

• We modelled a video source performing bitstream switching by a Hidden Markov Process whose states represent different kind of GOPs

• Consequent of the assumptions we made, we developed a ML parameter set estimation algorithm based on EM algorithm and evaluated the convergence

• We assessed the model by comparing the statistical characteristics of the model with respect to the real source

– **S. Colonnese, S. Rinauro, L. Rossi, G. Scarano, "H.264 Video Traffic Modeling Via Hidden Markov Process", EUSIPCO 2009.**

– **S. Colonnese, S. Rinauro, L. Rossi, G. Scarano, "Markov Model of H.264 Video Traffic", ISIVC 2008.**

– **S. Colonnese, S. Rinauro, L. Rossi, G. Scarano, "Hidden Markov Models of H.264 video sources using the Expectation-Maximization Algorithm",** ***submitted at IEEE Trans. on Image Processing***

•Future work on the argument :

– a recursive parameter estimation algorithm

– The study of different HMPs, such as Switching Markov Autoregressive Processes

–Application on Scalable Video Source