Slice-Based VBR Video Traffic Classification

Natalia Markovich

Institute of Control Sciences
Russian Academy of Sciences, Moscow, Russia

joint work with Astrid Undheim, NTNU, Trondheim, Norway

COST, September, 22-23, 2008, Samos, Greece
Description of the slice-based encoded video data

The mode type selection for the frame-based and the slice-based video encoding scheme.

- For the frame-based video encoding the first frame in GoP is intracoded;
- For the slice-based video encoding a slice of each frame in GoP is intracoded.
- The burstiness due to the large intracoded frames is hence reduced.
We consider
- a bufferless fluid model of a communication system;
- non-aggregated flow.

We deal with frame sizes of a single flow

Size of frame = number of packets in the frame × packet size

- One frame is transmitted every 1/30 second.
- The packet size is constant apart from the last packet in the frame.
- The frame size is a random variable since the number of packets inside each frame is random.
Our aims are

- to classify the frame sizes to the domains of homogeneity and stationarity;
- to estimate the mean loss due to the exceedances over the capacity of the channel and,
- to estimate high quantiles (e.g., 99%, 99.9%) of the exceedances that determine the quantiles of losses, the amount of losses may exceed these quantiles with a very small probability.
The frame-sizes of the slice-based encoded video stream, together with the mean frame size $u = 8.781$ Kbytes (solid horizontal line) and the 95% (dotted) and 99% (dot-dashed) empirical quantiles.
Classification of the size-frame data

Classification may be done by

- average frame size;

Four classes are determined: for $i \in [0, 4000] \cup [6461, 7198]$, $i \in [4001, 6039]$, $i \in [6040, 6210] \cup [6335, 6460]$ and $i \in [6211, 6334]$, $i$ is a number of frame.

- scene changes frequency (Undheim, Lin & Emstad 2007);

- the change of marginal distribution;

- the tail index (it shows the heaviness of tails of the class);

- the extremal index (it shows the dependence of the class).
The 80%, 95% and 99% empirical quantiles for four classes separated by average frame size.
Scene change detection

Motivation to divide the video traffic traces into scenes

- to have approximately constant bit rate (Undheim, Lin, Emstad, 2007);
- to have independent blocks (scenes) required for estimation of the extremal index and the mean loss per cluster.
Scene change detection

Let $X_j$ be the $j$th frame size and the minimum scene length is 12 frames, e.g. one GoP.

Methods of scene detection

Heyman, Lakshman 1996: the scene change is detected if the inequality is fulfilled for a fixed parameter $\lambda$

$$\frac{(X_{n+1} - X_n) - (X_n - X_{n-1})}{\left(\frac{1}{6}\right) \sum_{j=n-5}^{n} X_j} < -\lambda$$

Undheim, Lin, Emstad 2007: $\lambda = 0.4$ based on visual inspection

Markovich, Undheim, Emstad 2008: quantile method of scene detection
The first frame which size exceeds a fixed quantile of frame sizes (we take 80% to have enough data) is detected as a scene change frame.
Comparison of methods of scene change detection

Scatter plot of the logarithm of the scene lengths in Bytes versus logarithm of the size of the scene change frames in Bytes for the whole trace. Left: original method. Right: quantile method

The dependence scene lengths - change frame sizes is not homogeneous.

The proposed quantile method of the scene detection is more sensitive to the selection of the classes with different average frame sizes than the original method used before.
Test of dependence of scene lengths and scene change frame size

The ACF of the scenes of the entire sample:
left: the original method; right: 80 % quantile scene change detection. Since all ACFs are located inside the Bartlett’s 95% Gaussian confidence interval with the bounds $\pm 1.96/\sqrt{n}$, one may assume that both the scene lengths and the size of the scene change frames are independent. Hence, the selected scenes may be independent.
The cluster exceedance structure showing that the clusters correspond to the congestion periods.

A cluster or a congestion period is defined as a period in which one or more frame sizes are larger than the threshold. The mean loss is determined by means of exceedances.
Let $X_i, i = 1, 2, \ldots$ be a stationary process with a marginal distribution function $F(x)$. We have $n$ measurements from this process.

According to the theory of extremal values, for large $n$ and $u_n$:

$$P\{\max (X_1, \ldots, X_n) \leq u_n\} \approx F^{n\theta}(u_n),$$

where $\theta \in [0, 1]$ is the extremal index that reflects the dependence in the sequence.

For independent, identically distributed sequences $\theta = 1$. 
Estimates of the extremal index $\theta$

The blocks estimator: a cluster is defined as a block of the data with at least one exceedance over the threshold $u$.

$$\bar{\theta}^B(u) = \frac{k^{-1} \sum_{j=1}^{k} \mathbb{1}(M(j-1)r, jr > u)}{rn^{-1} \sum_{i=1}^{n} \mathbb{1}(X_i > u)},$$

$M_{i,j} = \max(X_{i+1}, \ldots, X_j)$, $k$ is the number of blocks, $r = \lfloor n/k \rfloor$ is the number of observations in each block.

The runs estimator: a cluster is defined as a block of data with some number of exceedances over the threshold and the following $r$ observations are all below the threshold $u$.

$$\bar{\theta}^R(u) = \frac{(n-r)^{-1} \sum_{i=1}^{n-r} \mathbb{1}(X_i > u, M_{i+1,i+r} \leq u)}{n^{-1} \sum_{i=1}^{n} \mathbb{1}(X_i > u)}.$$
Interpretation of the extremal index estimators

The $1/\bar{\theta}^B(u)$ ($1/\bar{\theta}^R(u)$) shows the mean number of exceedances in the cluster.

$$\text{The inversion of the blocks (runs) estimator}$$

$$= \frac{\text{the number of observations that exceed the thresholds } u}{\text{the number of clusters}}$$

The overall bit loss over all clusters for a fixed threshold $u$

$$= \text{cumulative exceedance over } u \text{ of the entire trace}$$

$$= \text{the mean excess } \times \text{the mean number of exceedances in the cluster}$$

$$\times \text{the number of clusters}$$

Runs estimate has a better asymptotic bias than the blocks estimate. Smith and Weissman 1994.
Main problem. Two parameters determine the clusters:

- the threshold (capacity) \( u \) and,
- the number of blocks \( k \) (or the number of observations \( r \) below \( u \)).

Runs estimate against variable thresholds \( u \) and the parameter \( r \).

The parameters \( u \) and \( r \) can be selected corresponding to the stability interval. The plot is stable for \( u \in [20, 30] \) and for any value \( r \) from \([70, 150]\).
Scene blocks estimator of the extremal index $\theta$

New idea

We shall use scenes as blocks.

In this case, we deal with blocks of unequal size.

New scene blocks estimator

$$\theta_S^B(u) = \frac{\sum_{j=1}^{k} 1(M_{\sum_{m=0}^{j-1} r_m, \sum_{m=1}^{j} r_m > u})}{\sum_{i=1}^{n} 1(X_i > u)}$$

where $r_j$ is the number of frames in $j$th scene, $\sum_{j=1}^{k} r_j = n$, $r_0 = 0$ and $k$ is number of scenes.
The extremal index $\theta$ plays a double role:

- **It implies** dividing the video traffic according to the **dependence** of frames in the scenes of each class.

- **It detects a non-stationarity** in the data, since the distribution of the inter-exceedance times is exponential with intensity equal to the extremal index.
Classification of dependence of frames by the extremal index

The inverse scene blocks estimate $1/\theta_B^S$ is calculated sequentially for moving window containing $m$ scenes for Classes 1-3 (from left to right).

Threshold $u$ equal to the 80% quantiles of frame sizes of the corresponding class. Moving window contains $m$ scenes: $m = 10$ for Classes 1 and 2, and $m = 2$ for Class 3.

Classes 2 and 3 are more stationary than Class 1 with respect to the more homogeneous inter-exceedance times distribution.
Classification by the extremal index

The channel capacity required for each class to satisfy a given bit loss, equal to 3%, is shown.

Table: The 3 % overall bit loss, mean loss per cluster, inter-cluster length, and the corresponding threshold for each class, in Kbytes.

<table>
<thead>
<tr>
<th>Class</th>
<th>Threshold</th>
<th>Overall bit loss</th>
<th>Mean loss per cluster</th>
<th>Inter-cluster length in number of frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>1033.8</td>
<td>24.2341</td>
<td>110.14</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>109.8</td>
<td>3.3936</td>
<td>63.00</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>258.1</td>
<td>21.4438</td>
<td>24.51</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>74.68</td>
<td>25.7223</td>
<td>42.41</td>
</tr>
</tbody>
</table>

The occurrence of a few “heavy” clusters or more frequent “light” clusters may influence the QoS and required capacity of channel.
The mean excess function of frame sizes

\[ \hat{e}(u) = \frac{n}{\sum_{i=1}^{n} \mathbb{1}\{X_i > u\}} \sum_{i=1}^{n} (X_i - u) \mathbb{1}\{X_i > u\} \]

determines the mean bit loss over a time \( t \).

The empirical mean excess function \( \hat{e}(u) \) for the whole trace against the threshold \( u \).

Class structure of the frame size data regarding the threshold \( u \): the values \( u \in [0, 8], u \in [9, 20], u \in [21, 37], u \in [38, 50] \) Kbytes.

The increasing (or decreasing) plot indicates that the data are distributed with a heavy- (or light-) tail.
The tail index $\alpha$ indicates the shape of the distribution tail, the smaller the value of $\alpha$, the heavier is the tail:

$$1 - F(x) = x^{-\alpha} \cdot \ell(x),$$

$\ell(x)$ is a slowly varying function. Examples: positive constants, $\ln x$, $\ln \ln x$.

Hill’s estimator

$$\hat{\alpha}^H(n, k_0) = \left( \frac{1}{k_0} \sum_{i=1}^{k_0} \ln X(n-i+1) - \ln X(n-k_0) \right)^{-1},$$

$X_1 \leq X_2 \leq \ldots \leq X_n$ are the order statistics of the sample $X^n = \{X_1, X_2, \ldots, X_n\}$ and $k_0$ is a smoothing parameter.
## Classification of frame sizes

### Table: Classification by the average frame size and the tail index

<table>
<thead>
<tr>
<th>Class number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hills estimate</td>
<td>4.1248</td>
<td>11.1003</td>
<td>18.4467</td>
<td>5.8846</td>
</tr>
</tbody>
</table>

### Table: Classification by the mean excess value for a given $u$

<table>
<thead>
<tr>
<th>Contributor classes</th>
<th>Threshold interval</th>
<th>Distribution type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 4</td>
<td>$u \in [0, 8]$</td>
<td>Light-tailed</td>
</tr>
<tr>
<td>1 – 4</td>
<td>$u \in [9, 20]$</td>
<td>Pareto-like</td>
</tr>
<tr>
<td>1, 3, 4</td>
<td>$u \in [21, 37]$</td>
<td>Light-tailed</td>
</tr>
<tr>
<td>1, 3</td>
<td>$u \in [38, 50]$</td>
<td>Exponential</td>
</tr>
</tbody>
</table>
High quantile estimation of losses

Let us consider positive exceedances

\(\{ Y_i = X_i - u, i = 1, 2, \ldots \}, \{ X_i, i = 1, \ldots, n \}\) are measured frame sizes for a fixed threshold \(u\).

\(\{ Y_i \}\) form the losses in the bufferless model.

We are looking for the estimation of their quantiles and, particularly, high quantiles close to 100%.

Weissman’s estimator of \((1 - p)\)th \((p\) is close to 0) high quantile

\[ x_p^w = Y_{(n-k_0)} \left( \frac{k_0 + 1}{(n + 1)p} \right)^{1/\alpha}, \quad k_0 = 1, \ldots, n - 1 \]

Main assumptions are

the independence and stationarity of \(Y_1, \ldots, Y_n\);

positivity of the tail index \(\alpha\).
High quantile estimation of losses

Table: Estimate of high quantiles for the positive exceedances data.

<table>
<thead>
<tr>
<th>Data Class</th>
<th>Tail index</th>
<th>High quantile estimate</th>
<th>Empirical quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class estimate</td>
<td>99%</td>
<td>99.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2.1234</td>
<td>[1.8975, 2.4104]</td>
<td>21.725</td>
</tr>
</tbody>
</table>
Conclusions

- We classify a video stream by average frame sizes.
- The distributions of the selected classes can only be mixtures of classical heavy- and light-tailed distributions.
- The dependence structure within the classes is variable because of the video encoding.
- A new quantile method for scene change detection is employed. The scenes selected by this method reflects the classification by the average frame sizes.
- The mean bit loss per cluster and the overall bit losses in the bufferless model are estimated.
- High quantiles of bit losses which determine the upper limits of losses for a fixed capacity are evaluated.
- We found the capacity of channel required to give the maximum allowed loss rate for each class.
- This methodology is applied for a test flow of slice-based video data. It can be extended to an aggregated flow.
Reference