A PREDICTIVE MULTI-CHANNEL MBAC TECHNIQUE FOR ON-LINE VIDEO STREAMING

Pietro Camarda, Cataldo Guaragnella, Domenico Striccoli
D.E.E. – Politecnico di Bari, via E. Orabona, 4, 70125 – Bari, Italy
camarda@poliba.it, guaragnella@poliba.it, d.striccoli@poliba.it

Keywords: Multimedia Streaming, Measurement Based Admission Control, Bandwidth Prediction.

Abstract: A measurement based admission control predictive technique is introduced for on-line streaming systems exploiting the GOP length demultiplexing of the aggregate bit stream in conjunction with a linear predictive algorithm. Due to the long latency of the statistical aggregate, the predictive technique is able to predict the bit rate over about two seconds of time. The prediction information is used in an admission control system to estimate the bit rate and the margin with respect to the channel capacity in the proposed streaming system. These measures have been used to estimate the overflow probability in a general aggregate situation. Tests conducted over real video sequences confirm the feasibility of the proposed technique.

1 INTRODUCTION

Multimedia streaming applications (e.g., audio-video streaming, Digital Video Broadcasting, UMTS services, etc.) are rapidly growing in the actual telecommunication world. They perform multimedia reproduction while taking into account some user feedback interactivity. In a multiplexed scenario, several multimedia flows with a high bit rate variability, data burstiness, self-similarity and Long Range Dependence (LRD) properties (Garrett and Willinger, 1994) are aggregated. Such a kind of Variable Bit Rate (VBR) traffic makes very difficult an optimal resource assignment and estimation, in terms of bandwidth occupation and client buffers, to properly manage video streaming.

An efficient aggregated VBR data transmission assumes that multimedia flows are delivered with an assigned Quality of Service (QoS), that is, end-to-end delay, data losses, and jitter specifications. To guarantee QoS requirements for each of the aggregated flows, Call Admission Control (CAC) schemes can be implemented, with the purpose to verify that each flow, sharing link bandwidth resources with other flows, can be delivered to destination while respecting QoS specifications. This task is performed whenever a new flow requests admission in the network link. If QoS requirements are respected, the new flow is admitted, otherwise is rejected.

Admission Control algorithms need the aggregate traffic to be accurately characterized, through a priori determined traffic parameters, or traffic measurements. In the first case, “parameter-based”, and in the second “measurement-based” admission control algorithms are implemented. As regards the first approach, the main flow characteristics are a priori set to estimate the amount of network resources needed by a flow aggregation, including the new flow requesting admission. The second approach is based on aggregate traffic measurements, to estimate the current link load.

Nevertheless, bursty compressed multimedia traffic is very difficult to characterize, due to the strong unpredictable nature of compressed aggregate traffic, and relative traffic descriptors could provide imprecise information for admission control purposes (Crovella and Bestavros, 1996 and Floyd, 1997). This is particularly true for the parameter-based approach, where traffic characteristics are a priori set. This makes an
accurate traffic characterization hard to perform, since admission control criterion does not properly take into account dynamics of the aggregate running into the network link. For bursty sources, this often translates in a network link underutilization, or a failed respect of QoS guarantees (Floyd, 1997). Measurement Based Admission Control (MBAC) algorithm class generally performs traffic evaluation and bandwidth estimation through a dynamic observation of traffic behaviour. This approach is useful to take into account aggregate bandwidth fluctuations by repeated “on-the-fly” measurements in past periodic time intervals, and subsequent actual bandwidth estimation, even if this generally comes at a cost of a computation complexity overload. This approach is particularly suitable in an on-line context, where only part of the aggregate data set is available for bandwidth estimation and actual load prediction.

An efficient MBAC scheme should thus provide an intelligent admission decision, based on network resource measurements, whose accuracy strongly affect MBAC efficiency and robustness. MBAC algorithms exploiting past traffic measurements and QoS parameters a priori specified decide, based on actual traffic estimation, the new flow admission or rejection. Several different strategies for traffic measurements have been developed, each of them strongly influencing the MBAC algorithm performance. In Jamin et al (1997a) three different MBAC techniques are compared. In Jamin et al (1997b) a MBAC algorithm is proposed for a packet delivery service with bounded delay requirements. In this algorithm, admission decision is based on a token bucket filter model, that is able to calculate worst case delays. The algorithm has been tested in a wide variety of scenarios, and including also LRD traffic. The MBAC proposed in Casetti et al (1996) presents a quite simple traffic estimation process in a single-link scenario, able to achieve a high bandwidth utilization, without violating, at the same time, QoS guarantees, specified as delay bounds. In Grossglauser and Tse (1999) the impact of some parameters, like estimation errors, flow dynamics (departures, arrivals, and bursty traffic time scales) and system memory, on MBAC robustness are analyzed. Their reciprocal interactions and impact on QoS specifications are considered for a unified framework. In Grossglauser and Tse (2003) a MBAC algorithm is proposed, based on a time scale decomposition of a flow aggregate, taking also into account flow arrivals and/or departures. Bandwidth fluctuations are decomposed into “fast time scale” and “slow time scale” contributes, separated by the “critical time scale” threshold. Aggregate fluctuations slower than the critical time scale are exploited for admission control policy while the faster are used to estimate the necessary amount of bandwidth needed to absorb high bandwidth fluctuations in a brief time period.

In this paper a novel MBAC algorithm is proposed. It is based on a set of aggregate traffic measurements performed in past time intervals, to derive statistically an estimation of the aggregate bandwidth and of the available bandwidth margin, necessary to absorb instantaneous peak rates, with the same approach used in Grossglauser and Tse (2003). Available bandwidth estimation is then exploited to admit new flows without violating QoS requirements of the whole aggregation.

The main novelty of the proposed algorithm lies in the criterion adopted for aggregate bandwidth estimation, that makes use of linear predictive techniques. In particular, aggregate traffic measurements are performed in time intervals preceding the new flow request of admission. This data set is exploited to statistically derive a prediction of the aggregate behaviour in a future time interval. This prediction is then combined with the QoS parameter, that is, overflow probability, to decide the new flow admission. In this context the overflow probability is the probability that the whole aggregate, included the new flow, exceeds the available link bandwidth.

The linear prediction technique dynamically performs an estimation of the statistical aggregate characteristics; the algorithm utilized is based on a “multi-channel” linear prediction, to determine the predictable characteristics of a set N different multiplexed video sources.

As it will be more clear in the next sections, this work aims at implementing a numeric filter that properly describes the predictable portion of aggregate traffic to quantify the overflow probability for admission decision.

The paper is structured as follows. In Section 2, the novel MBAC approach based on a bandwidth
prediction system, is presented and explained. In Section 2.1 the predictive system is analyzed in detail. Bandwidth prediction is then used in the MBAC algorithm, whose principles are explained in Section 2.2. In Section 2.3 the whole MBAC algorithm is implemented. Section 3 provides some significant numerical results. Finally, Section 4 gives some conclusions to the proposed work.

## 2 THE MBAC PREDICTIVE APPROACH

Statistical multiplexing of video sources allows an efficient use of the channel bandwidth for the video streaming application even if high frame variability is present (Salehi et al, 1998).

It can be observed that the bit rate of MPEG coded videos arranged in a composite bit stream (statistically multiplexed transmission system) presents a periodicity linked to GOP size: if the GOP length is N, the multiplexed video stream bit rate will still present a periodicity of N frames.

This particular case makes the bit rate time series “cyclostationary” over limited time intervals, and somehow “predictable”; the N frame-step stationarity can be exploited to predict the future behavior of the bit rate over a predictable time window.

The predictive nature of the video aggregate can be fruitfully exploited for designing MBAC system for on-line application in video streaming systems. It can be subdivided into two main parts:

- the predictive system, used to predict the bandwidth total allocation upon a future time, based on the bandwidth measures accumulated on a previous limited observation time;
- the admission control system.

Figure 1 pictorially represents the MBAC system.

### 2.1 The predictive multi-channel system

Predictive bandwidth estimation is carried out by standard predictive algorithm, with a multi-channel approach. The proposed system performs several predictions of the bit rates for several future time lags in order to estimate the whole bandwidth allocation required over a time-finite length observation window.

As it can be seen by figure 2, a quasi-periodicity is still present in the aggregate bit stream, and such characteristic comes from the intrinsic nature of the MPEG video encoded stream. The GOP quasi-periodicity allows to think the process as a cyclostationary process, so that statistical parameters can be computed separately on each stationary random process extracted by subsampling the aggregate bit rate time series with GOP length time step.

![Figure 2: Time evolution and spectrum of a 6 streams CIF aggregate.](image)

![Figure 3: the channel splitting system; 12 sub-channels depend on the 12 GOP length.](image)
long as the process we are dealing with is not properly stationary, several estimates of each channel correlation functions have been averaged to estimate the time correlation length. In the proposed approach, the time correlation length has been chosen as the time lag where the correlation function is greater than 0.5. Prediction can thus be carried out over time extending not further than 2 seconds. This means, at our full time resolution (25 frames/s), 25×2 = 50 frames, so that, for each channel, only 4 frames can be considered as the goal of our predictive technique.

The predictive scheme for time prediction of bandwidth allocation is reported in figure 5. After demultiplexing the data stream, filtering of each channel output is carried out. The filtering procedure is performed to extend process stationarity over longer time intervals, as subsampling tends to reduce the correlation between successive measures of bandwidth allocation (Shanmugan and Breipohl, 1988). This operation, in any case is straightforward and requires no computation: if a rectangular impulse response is used in the filter, the low-passed channel time series is at each time instant the sum of all the bandwidth measures along the widow length. This measure is simply obtainable by tracking the required transmission buffer filling, as the buffering at the transmission end cannot be avoided.

The data to be transmitted are simply rearranged in 12 sub-fluxes and each channel buffer filling is tracked to be used as the data input of the multi-channel prediction of the whole aggregate bit rate.

For each channel time series, a multistep predictive technique is implemented. Linear prediction assumes that, stated bₙ the n-th sample of a predictable time series in additive white gaussian noise (AWGN, wₙ), it can be estimated from its M previous samples xₙ₋₁, ..., xₙ₋₄ by the relation (Proakis and Manolakis, 1996):

\[
\sum_{k=1}^{M} c_k \cdot x_{n-k} + w_n
\]

The equation in (1) can be solved by means of a minimum mean squared error criterion. To obtain a compact description of the problem and a direct solution, matrix notation can be introduced:

\[
X \cdot \bar{c} = \bar{b}
\]

where X is the [N×M] data matrix of the signal time series, \( \bar{c} \) is the [M×1] linear prediction coefficient vector and \( \bar{b} \) is the [N×1] vector of the data to be predicted. Let us remember that N represents the length, in frames, of the aggregate periodicity. The number of equations of (2) is much higher than the number of coefficients of the linear prediction filter.

The solution in the mean squared error sense can be obtained by the use of the pseudo-inverse of X, giving:

\[
\bar{c} = (X^T \cdot X)^{-1} X^T \cdot \bar{b}
\]

This allows to compute the coefficients of the linear FIR predictor filter. This filter thus predicts xₙ (the elements in b vector) from its M preceding values [xₙ₋₁, ..., xₙ₋₄]. The same procedure can be established to predict the xₙ₊₁, xₙ₊₂ and so on in the future. Going a step further, we can perform a joint estimate of the next four samples [xₙ, ..., xₙ₊₃], basing on the observation of a certain number M of preceding channel bandwidth measures. At this aim,
equation (4) calculates a joint four-step prediction measure for each channel:

\[ C = \left( X^T \cdot X \right)^{-1} X^T \cdot B \]  

\[ (4) \]

In (4) C is a 4 columns vector of M coefficients. Each column in the vector contains the M coefficients of a linear prediction filters for the four successive channel bandwidth measures to be estimated.

The four bit rate measures predicted by (4) for the 12 channels are then summed together, thus obtaining the total number \( B_0 \) of bits predicted over a period T of two seconds. This gives us a predictive estimation of the total channel bandwidth, obtained as \( B_0 / T \).

This measure is then used to construct a predictive MBAC system for on line applications as described in the next section.

Figure 7 represents the cumulative bandwidth estimated in the predictable observation window both for the true aggregate and for the predicted. The difference is the prediction error.

2.2 The MBAC System

In this work, a look-ahead system, able to take care of the predictable information about the bit rate time evolution is proposed, to try to build up a Measurement Based Admission Control system.

The assumed transmission medium is characterized by a constant maximum channel capacity, C.

Both statistics of the prediction error and the predicted (low passed) composite video stream for all the cases of K-streams composite video have been computed.

The MBAC system, basing on the information about the acquired statistics of the composite video, should be able to accept or deny the request of a end user for the addition of a new stream in the video aggregate. The acceptance of the user request should not compromise the performances of the aggregate video stream of any user, so that a measurement based approach should be able to guarantee, within a given probability, that the inclusion of a new stream in the aggregate doesn’t lower too severely the whole system performances.

An efficient MBAC system should be able to use as much as possible the available channel capacity, so that the higher the number of videos in the aggregate bit streams, the higher the obtained performances of the streaming system.

Usually MBAC techniques use few statistic parameters such as the peak bit rate, the average bit rate and the standard deviation bit rate to define a procedure for the admission, so that a low bandwidth efficiency can be often experienced to be able to guarantee QoS.

In this work we deal with on line streaming. Not all the video stream is available, so that statistics for the generic N-video aggregate case are used to define the MBAC admission, in conjunction with the acquired statistics on the past observation of the process at hand.
The definition of the MBAC system needs a conservative margin with respect to \( C \), to manage all the unpredictable bit rate requirements in the K-streams composite video. Such a margin should be neither too strict, to avoid too frequent overflow of the available channel capacity \( C \), nor too mild, to avoid low bandwidth efficiency of the on-line streaming system.

The definition of a suitable margin with respect to the channel capacity \( C \) of the system is carried out basing on the prediction statistics for several \( N \)-aggregate streams, with different values of \( N \).

The channel capacity overflow probability event can depend on two possible causes:
- overflow coming from a peak in the prediction error;
- overflow coming from the peak of the required bandwidth to deliver the composite video stream.

The probability of the union of these events can be approximated by the sum of the two probabilities. As we are searching for a conservative MBAC algorithm, we can assume valid the following relation:

\[
P_q \approx P_s + P_i
\]

where \( P_q \) is the required overflow probability, \( P_s \) is the overflow contribute coming from the signal component and \( P_i \) is the overflow probability coming from the prediction error.

As a further requirement for the MBAC system, we assume the \( P_i \) probability be small with respect to \( P_q \):

\[
P_i = \alpha \times P_s
\]

with \( \alpha \) a coefficient to be defined, so that from (3.1):

\[
P_q = (1 + \alpha) \times P_s
\]

In our experiments, we assume that \( P_i \) be small if compared with the probability of the overflow event due to the signal specificity, so that we can choose \( P_i \) an order of magnitude smaller than \( P_s \); thus, we suppose \( \alpha = 0.1 \).

With such an assumption, we can draw the graphics for the \( P_q \) as a function of the bit rate for several orders \( K \) of the composite video stream.

By drawing the polynomial fitting curves of the bit rate as function of the \( P_q \), we can compute the bit rate of the composite video signal as a function of the overflow probability \( P_q \).

### 2.3 The MBAC implementation

The overflow probability can be used to define a MBAC system; the global bit rate of the K-streams composite video has to be considered; to take care of the two components of the bit rate, the predicted bit rate is computed as the sum of two components: the estimated bit rate in the prediction window and a margin with respect to the channel capacity, to avoid the frequent occurrence of the overflow event.

The predictable bit rate is estimated as previously described, while the margin depends on the number of aggregated streams and on the value of the predicted bit rate.

Two different measures have been carried out, one over the predicted signal and the other over the prediction error.

Let \( f_r(b) \) be the probability density function (pdf) of the prediction error with respect to the bit rate, and \( F_r(b) \) its corresponding cumulative distribution, the quantity \( F_{r,s}(b) = 1 - F_r(b) \) represents the probability that the event “the difference between estimated bit rate and the true one is greater than \( b \)”. This probability will be used in the proposed MBAC algorithm.

To accept a new streaming client, the available composite video bit rate must satisfy the relation:

\[
BR(N+1, P_s) + \Delta Br (\alpha \cdot P_s) \leq C
\]

where \( BR \) and \( \Delta Br \) are evaluated by (9). \( BR \) is a function of the overflow probability \( P_s \), and represents the composite video stream bit rate in the hypothesis the new stream is accepted and multiplexed together with the given K-streams composite video, to originate a (\( K+1 \))-streams composite video, and \( \Delta Br \) is the corresponding margin with respect to the channel capacity \( C \) to absorb the peaks of the \( BR \) due to the unpredictability of the phenomenon.

Whenever \( BR + \Delta Br < C \) the new client will be considered as admissible to the service: the probability of overflow will be estimated and sent to the client that, considering such information, might decide whether to accept the service or deny it.

The equality can be assumed in (8) as a limit condition to determine the searched \( P_s \) value.

For each \( K \), and defined \( \alpha \) in (6), the complementary curves of the probability as a function of the bit rate can be approximated by polynomial functions in the variable \( P_s \); we compute a single polynomial function for the unpredictability of the bit rate, \( \Delta Br \), and a family of curves for the \( N \)-stream cases of composite video streams.
Thus we have:

\[
\begin{align*}
\Delta Br(P_s) &= \Delta Br(\alpha \cdot P_s) = \sum_{n=0}^{N_s} c_n \cdot P_s^n \\
BR(N, P_s) &= \sum_{n=0}^{N_s} a_n (K) \cdot P_s^n
\end{align*}
\]  

(9)

Assuming to be in proximity of the channel capacity, with a K-streams composite video, the algorithm steps consist in:

1. Compute the probability \( P_{so} \) due to the \((K+1)\) case finding the zero of the polynomial equation \( BR(P_{so}) = C \):

\[
C = BR(K+1, P_{so}) = \sum_{n=0}^{N_s} a_n (K+1) \cdot P_{so}^n
\]

2. Compute the margin with respect to the channel capacity for the \((K+1)\)-streams composite video, \( Br(P_{so}) \):

\[
\Delta Br(K+1, \alpha \cdot P_{so}) = \sum_{n=0}^{N_s} c_n (K+1) \cdot P_{so}^n
\]

3. Compute the new maximum bit rate to be assumed for the smooth signal, \( BR_1 \):

\[
BR_1 = C - \Delta Br(\alpha \cdot P_{so})
\]

4. Compute the minimum value assumed at the new computed bit rate for the \((K+1)\) case, \( P_{min} \), as the solution of the polynomial equation \( BR_1 = BR(K+1, P_{min}) \):

\[
BR_1 = BR(K+1, P_{min}) = \sum_{n=0}^{N_s} a_n (K+1) \cdot P_{min}^n
\]

5. Compute the estimated overflow probability as:

\[
P_q = (1 + \alpha) \cdot P_s
\]

3 EXPERIMENTAL RESULTS

The experimental results we present all refer to real data of composite videos obtained by statistically multiplexing MPEG encoded YUV-CIF videos, with GOP of 12 frames and 30 fps. All the created composite video streams have been obtained randomly mixing sources with different statistics, and composite videos are limited to \(K=20\) sources composite video. Figure 8 refers to the obtained complementary cumulative distribution function for the cases of \(K\) multiplexed sources, with \(K\) in the set \([1,20]\).

Curves in figure 8 and 9, representing the complementary cumulative distribution functions for the aggregate predicted bit rate and errors respectively, have been polynomial fitted to give analytical expression in (9).

Two sets of \(K\)-streams composite videos time series bit rates have been produced: a training one, used to obtain curves in figures 8 and 9, and a test set, used to obtain experimental results.

The test of the proposed algorithm has been carried out on the test set of \(K\)-streams composite videos to estimate the expected overflow event.

A capacity \(C\) communication channel has been assumed, and a number of video, \(K\), in the composite video stream has been selected in order be close to the channel capacity and give a negligible overflow probability. A request of service by a \((K+1)\)-th client has been assumed, and the overflow probability for that client has been estimated as previously described.

Repeated experiments in the same conditions have been carried out. The experimental measure of the overflow probability for the test cases have been computed as the fraction between the overflow experienced time duration and the total duration of the \((K+1)\)-streams composite videos.

The obtained overflow probability is mostly well predicted by the proposed algorithm.

Figure 10 presents the histogram of the obtained probability values for 100 cases of \((K+1)\)-streams composite videos in the test set. The fraction of test cases showing an experienced overflow probability greater than the estimated one results lower than 15% in our simulations; also, the difference in probability value is never much higher than the estimated one.
Figure 9: \(1-F_n(BR)\): cumulative complementary probability functions of the prediction error (unpredictability) of the bit rates for several orders of composite videos.

4 CONCLUSIONS

In this paper, a new MBAC algorithm for VBR multimedia streaming has been proposed. It makes use of an innovative bandwidth prediction technique to more efficiently admit new flows in a network link with aggregate traffic already running, without violating the QoS requirements of the whole aggregation.

The proposed bandwidth prediction technique exploits the particular, quasi-stationary nature of aggregate traffic to predict bandwidth utilization in a relatively short time period. This approach is particularly useful to take into account aggregate dynamics quickly varying in time. Furthermore, this straightforward predictive technique can be fruitfully implemented in advanced video smoothing systems, where aggregate video transmission consists of Constant Bit Rate (CBR) pieces. The proposed predictive technique can thus be exploited to establish the maximum CBR size allowed for video smoothing in the next two seconds.

The predictive bandwidth estimation technique is the core of the MBAC algorithm. The predicted bandwidth information is utilized to perform a new flow admission, that respects the given aggregate overflow probability (QoS parameter).

Performed simulation show the effectiveness of the proposed MBAC for a wide number of simulation scenarios. Experimental results show that the observed overflow probability almost never exceeds the correspondent QoS parameter, testifying a good performance of the proposed algorithm and its ease of use in several practical scenarios.

5 REFERENCES


