1000-9825/2002/13(05)0900-07 ©2002 Journal of Software 软 件 学 报 Vol.13, No.5

A Dynamic Bandwidth Allocation Algorithm for MPEG Video Streams*

JU Yi-ming, ZHANG Yi, PENG Yu-xing, CHEN Fu-jie

(*School of Computer, National University of Defence Technology, Changsha* 410073, *China*)

Received January 15, 2001; accepted December 5, 2001

Abstract: Asynchronous Transfer Mode (ATM) is an emerging standard for broadband networks, packetized video is likely to be one of the most significant high-bandwidth users of such networks. The transmission of variable bit-rate (VBR) video offers the potential promise of constant video quality but is generally accompanied by packet loss, delay and delay jitter which significantly diminishes this potential, so it is necessary to provide requested QoS. In this paper, two new dynamic bandwidth allocation algorithms for MPEG videos over ATM are proposed, one is dynamic measure algorithm, another is dynamic virtual buffer algorithm, in which the authors dynamic adjust the bandwidth of MPEG video sources with a specified allowed cell loss ratios to provide QoS guarantees in an ATM network.

Key words: variable bit-rate video; dynamic bandwidth allocation algorithm; cell loss ratios; QoS

MPEG videos traffic in networks requires control of such quality of service (QoS) parameters as average delay, delay variation, and loss rate. To ensure that QoS requirements will be met, the normal approach is to require the application to describe its traffic, using such statistics as the period, deadline, average and worst-case processing time or packet size, etc. These statistics are often conservative estimates, and are statically declared. Based upon these statistics, a calculation is made of the resources (buffer space, network and disk bandwidth, etc.) needed to provide the requested QoS.

QoS guarantees for video traffic are difficult for following reasons: first, when the traffic is processed in real-time, it is impossible to predict precisely what the actual behavior will be. Second, the traffic itself may be highly variable, which makes a peak-rate model excessively conservative. Third, compressed video can exhibit long-range dependency[1,2], which implies that the standard statistical traffic models are probably inadequate. For these reasons, efficient resource allocation, along with good control of QoS, is a challenging problem.

To date, a variety of bandwidth allocation methods for MPEG videos have been proposed, which can be categorized as either off-line methods or on-line methods. Off-line methods will make allocation decisions before any traffic is transmitted, an example of a static off-line method is the peak-rate allocation which is used for most real-time applications. This approach has the advantages of simplicity and predictability, but suffers from the

* Supported by the National Natural Science Foundation of China under Grant No.69933030(国家自然科学基金)

JU Yi-ming was born in 1956. He was a senior engineer at School of Computer, National University of Defence Technology. His current research areas include distributed multimedia and multimedia network. **ZHANG Yi** was born in 1973. She is a teacher at School of Computer, National University of Defence Technology. She received her Ph.D. degree at Computer Science from National University of Defence Technology in 2000. Her current research areas include distributed multimedia system, and network security. **PENG Yu-xing** was born in 1963. He is a professor at School of Computer, National University of Defense Technology. His current research areas include distributed multimedia and multimedia network. **CHEN Fu-jie** was born in 1935. He is a professor and doctoral supervisior at School of Computer, National University of Defence Technology. His current research areas include distributed multimedia and multimedia network, computer construction, real-time system and network security.

problems noted above, as well as low resource utilization if the peak-to-mean ratio is high.

Real-time videos can also be supported by using on-line measurement based approaches to dynamically allocate the required bandwidth to support the QoS. Some methods[5,10] use the previous frame sizes to predict the size of the future frames and future bandwidth requirements. While these methods which allocate bandwidth on a frame-by-frame basis are efficient in theory, they may not be feasible in practice since it may be difficult for the network to dynamically allocate bandwidth as frequently as each frame period (i.e. 24~30 times per-second). Other methods[6] use a large time window to reduce the frequency of bandwidth allocation, a problem with almost all these approaches is that they are not easily scalable because they focus only on a single video source.

In Refs.[9,11], they have studied dynamic bandwidth allocation based on measured cell loss ratios (CLR). The measurement time interval of these methods is inversely proportional to the CLR. When the target CLR is small, a very large time window will be required to collect the statistics for CLR estimation. For example, measuring a CLR of 10^{-5} with meaningful statistics requires at least 10^7 cell arrivals to be monitored. However, it is not feasible to use very large time windows to dynamically allocate bandwidth since the traffic conditions may have changed significantly before accurate measurement can be obtained. Furthermore, these mechanisms do not control short-term QoS violation, which may lead to poor quality in video transmission for a short period of time.

In this chapter, we propose two dynamically bandwidth allocation algorithms(dynamic measure algorithm and dynamic virtual buffer algorithm)to the aggregate traffic of videos. Our approach has several advantages. First, considering the aggregate traffic reduces the complexity associated with keeping track of the individual sources and also improves the accuracy of the bandwidth prediction. Second, bandwidth can be reallocated within a reasonable time window (2s to 10s). Third, resources are allocated based on measured QoS so that it does not require the source to specify precise traffic models.

1 Dynamic Bandwidth Allocation Algorithm for MPEG Videos

1.1 System model

Network must allocate a variety of resource in order to provide the desired QoS for MPEG videos, in this paper, the QoS parameter we are most interested in is cell lost ratio (CLR). For simplicity, We assume that the QoS (CLR) requirement for each video source is the same. Our objective is to allocate the minimum bandwidth which will meet the specified cell loss ratio parameter Q for each video source. Q is a constant value which is pre-designed.

The algorithm dynamically renegotiates with the access node, and a proper bandwidth would be allocated in order to meet a desired QoS. Cell arrivals and losses from the queue are monitored throughout the duration of the application. Bandwidth changes are renegotiated at discrete instances of time. We denote the n^{th} re-negotiation instant as t_n, and the interval between re-negotiation points t_n and t_{n+1} as the n^{th} update interval $\zeta(t_n,t_{n+1})$. The bandwidth during $\zeta(t_n,t_{n+1})$ is constant and is denoted $C(n)$. Bandwidth is reallocated at time t_i (i=1,2,3,…) to accommodate the highly bursty nature of VBR video. Let $A(n)$ and $L(n)$ denote the number of cells that arrive at the buffer and the number of cells lost due to buffer overflow during the n^{th} measurement period. The CLR during the n^{th} measurement period is $P_n=L(n)/A(n)$. The cumulative cell loss ratio for all intervals up to and including the n^{th} interval is denoted:

$$\wp = \sum_{i=1}^{n} L(i) / \sum_{i=1}^{n} A(i) . \tag{1}$$

1.2 Dynamic measure algorithm (DMA)

The goal of DMA algorithm is to adjust the bandwidth to a minimum value C^*, while the CLR approaches Q, another goals is simply to be implemented.

Figure 1 shows the pseudo code of DMA algorithm at one renegotiating instant t_i. DMA algorithm updates the network bandwidth $C(i)$, based on the observed CLR during the i^{th} interval. This measurement along with the desired **CLR** value, Q are used in the error function $ln(P_i/Q)$. The non-linear error function provides either a positive or negative feedback value based on the observation taken during the most recent interval. To keep the bandwidth at a more stable value, the feedback becomes smaller as measured CLR approaches the desired value. The error function is also appropriate due to the very small loss rates that are normally desired. In Fig.1, γ represents the bandwidth re-adjustment factor. It can be set according to experience.

As shown in the Fig.1, the re-negotiation interval is lengthened in two cases: if the cumulative CLR & and the CLR during the most recent interval P_i are both better than required, or when P_i and P_{i-1} are on different sides of (one greater than, the other less than) the desired CLR. Doubling the interval length also reduces the number of re-negotiations required over time.

if (($\wp \le Q$) AND ($P_i \le Q$)) then
$\zeta(t_{i+1}, t_{i+2}) \leftarrow 2 \times \zeta(t_i, t_{i+1})$;
else
{
/* adjust the re-negotiation interval */
if ($\ln(P_{i-1}/Q) \times \ln(P_i/Q) \le 0$) then
$\zeta(t_{i+1}, t_{i+2}) \leftarrow 2 \times \zeta(t_i, t_{i+1})$;
/* update the bandwidth */
if ($P_i > Q$) then
$C(i+1) \leftarrow C(i) + \gamma \times \ln(P_i/Q)$
else
$C(i+1) \leftarrow C(i) + \gamma/2 \times \ln(P_i/Q)$
}

Fig.1 DMA algorithm

1.3 Virtual buffer algorithm

In DMA algorithm, it requires that Q be large enough to be measured accurately within a measurement period, if the Q is small or the measurement is short, the variance of the measured CLR in each measurement period can be very large. To reduce the variance, we propose another method to estimate small CLR. In Refs.[9,11] presented measurement-based schemes to calculate CLR, however, those methods require that the Q be large enough to be measured accurately within a measurement period. If the Q is small or the measurement period is short, the variance of the measured CLR in each measurement period can be very large. To reduce the variance, we apply the virtual buffer method to estimate small CLR.

Here, we introduce the Dynamic Virtual Buffer (DVB) method for performance monitoring of small Q value. We consider an ATM multiplexer loaded with N MPEG video source, the multiplexer has a finite buffer of size K cells and serves cells on a FIFO-served basis at rate C cells per-second. Here, the DVB uses a parallel set of counters to simulate virtual cells being multiplexed into several virtual buffers with different service capacities, It is conceptually similar to the parallel virtual buffers described in Ref.[9]. The major difference is that the Q in Ref.[9] must be large enough to be measured directly ($\ge 10^{-3}$). Our algorithm does not have this limitation because the small Q (e.g., 10^{-5}) is obtained by prediction instead of direct measurement.

In DVB algorithm, There are M parallel virtual buffers with services rates C_m, m=1,...,M. Without loss of generality, we assume that $0 \le C_1 \le C_2 \le ... \le C_M \le C_{res}$, where C_{res} is the reserved capacity according to the CAC mechanism when a video is accepted into system. This reserved bandwidth is dedicated to the VBR video sources and cannot be used by other bandwidth guarantee services. At the beginning of each measurement period, the buffer occupancy of each virtual buffer is set equal to the actual buffer occupancy. The number of cells that arrive $A(n)$ and cells that are lost $L_m(n)$ are measured for the m^{th} virtual buffer in the n^{th} measurement period. For the m^{th} virtual buffer with capacity C_m, we calculate the CLR $P_m(n) = L_m(n)/A(n)$. Those virtual buffers with smaller service capacities have higher virtual CLRs.

To relate the statistics of cell losses of these virtual buffers to those of physical buffer, we need to find a relationship between the service capacity and CLR. The simplest relationship in the literatures is the Gaussian approximation[13], which basically provides a quadratic relationship between log (CLR) and service capacity C. Thus, we assume the following quadratic function to relate log (CLR) and C:

$$\log(\text{CLR}) = a_0C^2 + a_1C + a_2, \tag{2}$$

where a_i's are constants to be determined by minimizing the least squares error as follow:

$$\min\sum_{i=1}^{M'}\left[\log(P_i) - \left(a_0C_i^{\,2} + a_1C_i + a_2\right)\right]^2, \tag{3}$$

where M' is an positive integer $\le M$.

Once we have determined the coefficients a_i's, we can use equation (2) to estimate the required bandwidth to support the target Q value. The proposed algorithm is flexible since it makes no assumptions about the arrival process of the video sources. The regression method only requires to know the CLRs of the multiplexer when the service capacities are small. These are obtained from measurement using the set of parallel virtual buffers, i.e. (C_m,P_m), $m=1,2,\dots,M$.

The proposed bandwidth allocation mechanism operates as follows. For every measurement period, we apply the DVB method to obtain M pairs of (C_m,P_m). These data are used to determine a_i's in equation (2). Then Eq.(2) is used to estimate the bandwidth $C_e(n)$ that guarantees the Q during the n^{th} measurement period. The allocated bandwidth for the $(n+1)^{th}$ measurement period is given by:

$$C_a(n+1) = \alpha C_a(n) + (1-\alpha)C_e(n), \tag{4}$$

where $\alpha \in [0,1]$ is the weight for the estimated bandwidth, and $C_a(n)$ is the allocated capacity in the n^{th} measurement period. We set $C_a(1) = C_{res}$. Formulation 4 is a simple exponential forecasting commonly used in time series prediction. Accordingly, the residual bandwidth that can be used by other services is simply equal to $C_{res} - C_a(n)$ in the n^{th} measurement period.

The advantages of this approach are: (1) it is efficient and flexible since it attempts to allocate the minimum required bandwidth to satisfy the target QoS without assuming any particular traffic model; (2) it allows a much lower CLR to be specified while using a reasonable measurement interval.

1.4 CLR control in short period

Real time video is highly bursty and difficult to predict. When there is a sudden jump in the arrival rate due to scene changes, we may under-allocate the required bandwidth in one particular measurement period. Excessive cells may be lost in this particular period. Although error correction technique, such as error concealment, can handle to certain degrees this kind of short-term cell loss, it may cause a sudden drop in picture quality for a short period of time. This may be totally unacceptable from the users' point of view. Furthermore, the cumulative CLR may also violate the guaranteed QoS due to this large short-term cell loss. Since we do not control such short-term cell losses in our proposed algorithm in Section 1.3, we call this algorithm to be Uncontrolled Dynamic Virtual Buffer

Algorithm (U-DVBA). In the following, we suggest two methods based on U-DVBA to control such short-term cell loss due to under-allocation of bandwidth.

(1) ρ-factor allocation method

The ρ-factor Allocation method simply specifies ρ, a target bandwidth utilization, and allocates $C_a(n)/\rho$ instead of $C_a(n)$ in the n^{th} measurement period. We find that setting ρ between 0.85~0.95 is sufficient to handle the effect of sudden surges in arrival rates for most cases. While this method is simple, it wastes bandwidth because the U-DVBA algorithm only occasionally under-allocates bandwidth.

(2) Re-Allocation method

The Re-Allocation (RA) method avoids excessive cell loss by changing the bandwidth within a measurement period. For each measurement period, a target value of short-term cell loss L_{short} is specified. If insufficient bandwidth is allocated during the measurement period, more than L_{short} cells will be lost, which triggers the immediate increase of bandwidth to the reserved bandwidth C_{res} for the remaining interval.

When the bandwidth allocation is accurate (i.e., the bandwidth allocation is approximately equal to the input rate), the expected number of cell losses allowed in a measurement period is $Q\times\Delta\times C_a(n)$, where Q is the target long-term CLR, $C_a(n)$ is the allocated capacity in n^{th} measurement period and Δ is the length of the measurement period. Suppose the U-DVBA algorithm fails β percent of the time. To maintain the cumulative CLR to satisfy the target value Q, the maximum number of cells that are allowed to lose in a measurement period can be approximated by $Q\times\Delta\times C_a(n)/\beta$.

Generally speaking, larger β gives smaller L_{short} and hence bandwidth reallocation is activated more frequently. This in turn gives smaller cumulative CLR but requires more bandwidth.

As mentioned previously, a fixed amount of bandwidth C_{res} has already been reserved for the VBR connections in the call admission phase. In the n^{th} measurement period, we allocate $C_a(n)$ to the VBR traffic and the residual bandwidth $C_{res}-C_a(n)$ is released to support other services. If more than L_{short} cells are lost in a measurement period, reallocation of bandwidth will be activated.

2 Simulation and Results

Real video traces are used to study the performance of our bandwidth allocation mechanisms. Those video traffic characteristics are summarized in Table 1.

2.1 Numerical result of DMA algorithm

In this section, the performance of DMA algorithm is investigated using MPEG-compressed videos, which are listed in Table 1. The desired QoS was a CLR of $1*10^{-3}$. Table 1 shows the performance of DMA and an off-line peak-rate allocation algorithm.

For DMA algorithm, the initial re-negotiation interval was 4 seconds, γ was 100Kbps. As indicated in Table 1, DMA algorithm requires fewer bits than peak-rate allocation. Savings of DMA algorithm as compared to peak allocation ranged from 15.3% to 134.1%. In each experiment DMA algorithm reduced the bandwidth allocated while maintaining the desired QoS.

2.2 Simulation results of DVBA algorithm

In this subsection, we evaluate the performance of VBA algorithms using MPEG video sources (videos' characterization are list in Table 1). We consider an ATM multiplexer with a finite buffer of size K. There are N video sources multiplexed to form the incoming traffic. Simple FIFO policy is used and cells which arrive at a full buffer are simply dropped. We assume that within each source, the cell arrivals are equally spaced within each frame interval.

Table 1 Video traffic and DMA algorithm performance

Video	Peak/Mean	DMA algorithm (100M bit/s)	Peak-Rate algorithm (100M bit/s)
Tennis	8.69	5.61	7.64
Talk	7.20	2.46	4.62
News	9.01	4.69	7.78
Term	7.29	1.88	3.18
Soccer	6.9	6.50	7.49
Bowl	5.99	4.01	5.63
MTV1	9.31	6.39	9.18
MTV2	6.02	5.89	8.98
StarWars	14.3	6.23	10.08
Simpsons	12.9	4.83	9.62
Dino	10.1	5.20	9.79
Mr.Bean	13.0	3.92	9.18
Race	9.15	2.91	4.78
Lambs	18.4	3.12	5.37
Movie	12.1	4.85	7.0
Bond	6.31	6.32	9.34
Asterix	6.59	3.71	5.92
Talk2	7.01	5.34	8.30

We consider 20 MPEG video sources being multiplexed into a buffer of size 500 cells with target cell loss ratio be 10^{-5}. We thus apply the algorithm proposed in section 2.3 and 2.4 to allocation bandwidth dynamically. In the simulation, the weight α is set to 0.9 and measurement period Δ=4s. We also set the parameters ρ=0.90, and $\beta_{\min}$=10% for the two methods respectively.

Table 2 summarizes the performance of the ρ-factor Allocation method, Re-Allocation method, and U-DVBA algorithm.

Table 2 Comparison along U-DVBA, ρ-factor allocation and re-allocation methods

Bandwidth allocation methods	Allocated mean bandwidth(cells/s)	Cumulative CLR	Worst CLR in a measurement period
U-DVBA	3.05×10^5	1.3×10^{-4}	3.98×10^{-2}
ρ-factor allocation method	3.23×10^5	6.3×10^{-6}	4.43×10^{-3}
Re-Allocation method	3.02×10^5	1.01×10^{-5}	1.13×10^{-3}
Static allocation	3.54×10^5	5×10^{-5}	1.10×10^{-2}

3 Conclusions

In this paper, we introduce two dynamic bandwidth allocation algorithms (DMA and DVBA), which efficiently allocate resource to provide a required QoS. They adjust the bandwidth of MPEG video sources with a specified allowed cell loss ratios.

We also discuss the potential problem of short-term cell losses with our U-DVBA scheme and then introduce two solutions, namely the ρ-factor Allocation and Re-Allocation schemes, to control cell losses. Through simulation, we show that the Re-Allocation scheme can be more efficient than the ρ-factor Allocation scheme if dynamic adjustment of bandwidth within a measurement window is allowed.

Using numerical examples, we show that our proposed dynamic bandwidth schemes is more efficient than the static bandwidth allocation scheme, which requires accurate traffic information. Finally, we examine the effects of different parameters on the performance of our proposed algorithms.

In summary, the major advantages of our proposed measurement-based schemes are: they do not require the user to specify accurate traffic parameters; they can be applied to a wide variety of traffic sources; they adapt quickly to the traffic variation within small measurement intervals; the complexity of the algorithms is small.

References:

[1] Erramilli, A., Narayan, O., Willinger. W. Experimental queuing analysis with long-range dependent packet traffic. *I*EEE/ACM Transaction on Networking, 1996,4(2):209~223..

[2] Garret, M.W., Willinger, W. Analysis, modeling and generation of self-similar VBR video traffic. In: SXCOMMT4. London, 1994. 269~280.

[3] Rose, O. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. Technical Report 101, University of WUrzburg Institute of Computer Science, Feb. 1995.

[4] The ATM Forum. Traffic Management Specification. Version 4.0, March 1996.

[5] Adas, A. Supporting real time VBR video using dynamic reservation based on linear prediction. In: Proceedings of the IEEE INFOCOM'96. 1996. 1476~1483.

[6] Chong, S., Li, S.-Q. Predictive dynamic bandwidth allocation for efficient transport of real-time VBR video over ATM. IEEE JSAC, 1995,13(1):12~23.

[7] Cruz, R.L. A calculus for network delay, part I: network elements in isolation. IEEE Transactions on Information Theory, 1991, 37(1):114~141.

[8] Garrett, M.W., Willinger, W. Analysis, modeling and generation of self-similar VBR video traffic. In: Proceedings of the SIGCOMM'94. 1994. 269~280.

[9] Hsu, I., Walrand, J. Dynamic bandwidth allocation for ATM switches. Journal of Applied Probability, 1996,33(3):758~771.

[10] Pancha, P., Zarki, M.El. Bandwidth-Allocation schemes for variable-bit-rate MPEG source in ATM networks. IEEE Transactions on Circuits and Systems for Video Technology, 1993,3(3):190~198.

[11] Rampal, S., Reeves, D.S., Viniotis, I. Dynamic resource allocation based on measured QoS. Technical Report TR 96-2, North Carolina State University, 1996.

[12] Rose, Q. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. University of Wuerzburg, Institute of Computer Science Research Report Series. Report No.101. Feb.1995.

[13] Guerin, R., Ahmadi, H., Naghshineh, M. Effective capacity and its application to bandwidth allocation in high-speed networks. IEEE Journal on Selected Areas Communications, 1991,9(7):968~981.

基于 **MPEG** 视频流的动态带宽分配算法

鞠怡明, 张 怡, 彭宇行, 陈福接

(国防科学技术大学 计算机学院,湖南 长沙 410073)

摘要: ATM 是兴起的宽带网标准,基于报文的视频正是宽带网上的一个典型应用.变位率视频的传输方法为恒定的视频质量提供了可能,但是在传输过程中,报文的丢失、延迟和延迟抖动等因素都会影响视频质量,因此有必要提供 QoS 保证.提出了两种新的在 ATM 网上进行 MPEG 视频传输的动态带宽分配算法,一种是动态计算算法,另一种是动态虚拟缓存算法.在这些算法中,根据允许的丢包率,动态地调整 MPEG 视频带宽以保证 QoS.

关键词: 变速率视频;动态带宽分配算法;丢包率;QoS

中图法分类号: TP393 文献标识码: A