Supporting All Service Classes in ATM: A Novel Traffic Control Framework

[1]Tamás Marosits, [1]Sándor Molnár, [2]Gábor Fodor
[1]Department of Telecommunications and Telematics
Technical University of Budapest
Pázmány Péter sétány 1/D, Budapest, Hungary H-1117
{molnar, marosits}@ttt-atm.ttt.bme.hu
[2]Mobile Networks and Systems Research Department
Ericsson Radio Systems
Kistagangen 20-26, Kista, Stockholm, Sweden SE-16480
Gabor.Fodor@era-t.ericsson.se

Keywords: ATM Quality of Service, cell loss, cell delay, cell delay variation, service classes dynamic cell scheduling algorithm

Edited by: Marcin Paprzycki

Received: January 31, 1999 **Revised:** July 31, 1999 **Accepted:** August 31, 1999

This paper presents a general traffic control framework for Asynchronous Transfer Mode (ATM) networks with its performance evaluation. The proposed traffic control scheme can incorporate all the recently considered ATM service classes including Constant Bit Rate (CBR), real time Variable Bit Rate (rtVBR), non-real time Variable Bit Rate (nrVBR), Available Bit Rate (ABR) and Unspecified Bit Rate (UBR) services. The control is based on a complete buffer partitioning architecture and on the associated buffer scheduling rule with adaptive weighting functions. We present the formulation of the traffic control as an optimization problem in a 3-dimensional Quality of Service (QoS) state space. A solution approach based on dynamic programming is also suggested. A comprehensive performance evaluation of the method has been performed based on simulations and results are presented with several examples. The QoS dependence on CBR load, VBR load, VBR burstiness, UBR load are investigated and results are demonstrated with explanations.

1 Introduction

Since ATM networks are to support CBR, rtVBR, nrVBR, ABR and UBR service classes, the simplest 2-level priority based control policies become inadequate [11, 19]. Furthermore, within these service classes different VCs may require different cell loss, cell delay and cell delay variation parameters. It is therefore essential that the traffic control strategies be capable for the provision of the negotiated QoS parameters and for high network utilization by statistical multiplexing traffic classes with strict (CBR and VBR) or limited (ABR) QoS guarantees with a pure best effort type service class (UBR). Indeed, one of the key issues in the success of ATM is the traffic integration, and specifically, the design and analysis of control strategies which make the integration possible [9, 15, 18, 20, 21] . The need for a fine granularity of traffic control in ATM has been recognized e.g. in [19] and [11] where a 4-level priority based control mechanism is proposed. This model classifies traffic classes as "sensitive" or "less sensitive" to loss and delay. Gelenbe et al. concentrates on minimizing the impact of cell loss where cells belonging to different classes are assigned a cost function representing the importance of cells belonging to different "sessions" i.e. VCC's [12]. However, the priority based control algorithms are simple and easy to evaluate, their behaviour are static, they cannot be adapted to variable traffic. In addition, there are too many loser services because of static rules. The mixed approaches, such as Partial Buffer Sharing [14] are better, but they are not suitable for both loss and delay sensitive traffic, for example rtVBR. In a static priority system the cells with higher priority level can completely push out the lower priority traffic. This is an important problem in the case the network can provide ABR traffic. The ABR cannot have the lowest priority because it also has QoS guarantees. However, if it has a medium priority level, it can push out all the lower priority cell streams that are neglected in the control. ABR increases its rate to utilize the full bandwidth left by higher priority classes.

It is envisaged that second generation ATM switches will employ the Generalized Processor Sharing (GPS) and its packetized version PGPS, as the basic principle for buffer management [16]. However GPS is a static rule which means that it is reconfigured only when a new connection is established. There is no adaption of the fluctuation of traffic between two reconfigurations and the instantaneous QoS parameters have not been taken into account. The ABR traffic has some problems also with this scheduling discipline, because it can change its rate during the connection setup. In the more theoretical vein, recently there has been a growing interest in devising stochastic control methods, which can serve as a theoretical basis in the engineering of control enabling traffic integration in ATM, see e.g. [13, 17].

The purpose of this paper is to present a traffic control framework which allows for arbitrary degree of granularity in terms of guaranteed QoS parameters in an ATM network where service classes with and without QoS guarantee, with and without congestion control, and with and without real time guarantees are present. This means that the control cannot be based on a single bit, like the Cell Loss Priority (CLP) bit found in the ATM header. This is not only because a single bit cannot contain enough information on a complex QoS measure, but also because the continuous ("real time") QoS monitoring of VCC's is both required and feasible by current and next generation ATM switches. We propose a scheme where the control is based on the (1) negotiated QoS parameters, (2) instantaneous (current) QoS of the VCC under control, and (3) network resources allocated for the VCC under control. This is achieved by (1) defining a 3-dimensional QoS state space where the QoS parameters of each VCC can exactly (or with arbitrary precision) be described and (2) by a complete buffer partitioning with complete link capacity sharing [14] architecture of ATM multiplexers, which allows for "an individual handling" of VCC's requesting sharply different QoS measures from the network. It follows that a buffer partition arbitration algorithm is needed, which decides (possibly at each time slot) which partition's cell(s) gets served next. The basic requirements to this algorithm are that (1) it should guarantee the negotiated QoS parameters to each VCC and (2) it should optimize the "overall" network performance in the sense that provided that (1) is kept, each VCC gets the highest possible quality of service while network utilization is also kept high. Since the algorithm is to be executed real time, it should be simple and feasible by current technologies. The authors published the basic idea of this traffic control algorithm in [10] with. This paper presents the performance evaluation of this scheme.

The paper is organized as follows. Section 2 presents the reference architecture of the ATM buffers and the reference model of a VC connection, which serve as the basic model for the traffic control scheme under study. The QoS state space, is described in Section 3. In this space the contracted traffic parameters define an acceptance region, within which the points representing the QoS of the current VCC's must fall. The buffer partition arbitration algorithm is formulated as a dynamic programming optimization problem. Next, in Section 4, simulation results are presented, where the impact of the so called weighting function settings on network behavior is studied. The role of the weighting functions assigned to the partitions is to define the partition to be scheduled next. The section discusses numerical results and relates some results to related work. Section 5 draws conclusions and outlines future extensions of the model.

2 The Scheduling Technique

2.1 Buffer architecture

The traffic control method based on the complete buffer partitioning architecture (see Figure 2.1) where the total switch memory is divided to five FIFO buffers according to the presently considered service classes: CBR, rtVBR, nrVBR, UBR and ABR [4]. For the ABR class the end-to-end rate based EFCI mechanism [8] is used to reduce the cell loss. The motivation behind choos-

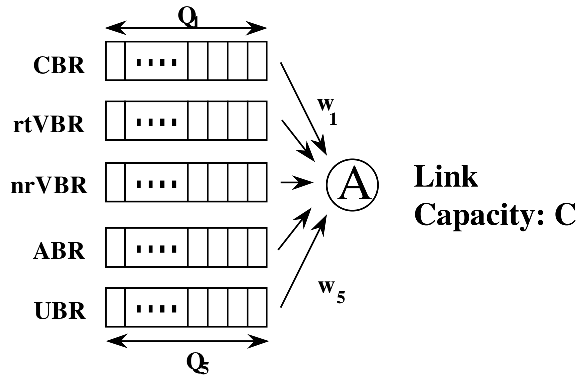

Figure 2.1 The buffer architecture

ing the complete buffer partitioning is that previous works clearly demonstrated that the presence of all these inherently different services having also rather diverse QoS requirements the complete buffer partitioning based schemes are superior over shared buffering techniques [11, 14, 19]. The Connection Traffic Descriptors should be selected in corporation with the proper dimensioning of Usage Parameter Control (UPC) [1, 4, 7], see below. An efficient Call Admission Control (CAC) can be performed with buffer partitioning and bandwidth allocation based on the listed parameters. In our model a weighting function (W_i) is dedicated to each buffer. The cell scheduling rule is the following: all weighting functions are evaluated at each time slot and the Head of Line (HOL) cell of the buffer with the greatest weighting function value is forwarded to the output link. The appropriate choice of the weighting function is a crucial point of the control. The weighting function uses the Connection Traffic Descriptors, QoS Requirements, Network Resource settings and also the current (instantaneous) QoS information of the VCC under control. This idea of weighting functions allows us to set flexible and adaptive control method. An application example of setting the weighting functions can be found below. We use the results of the CAC as starting point so we get the buffer sizes (Q_i) as input parameters to our traffic control method.

2.2 Traffic Control Parameters

The main goal of traffic control is to protect the network and the user in order to achieve network performance objectives with optimum allocation of network resources [2]. To fulfill these objectives QoS requirements, traffic descriptors and network information needed for the generic traffic control functions. We have chosen the following parameters for our traffic control framework which is in agreement with the standardization work of ATM Forum [3, 4] and ITU-T [1, 2]:

- **Connection Traffic Descriptors**: Peak Cell Rate (PCR), Cell Delay Variation Tolerance (CDVT), Sustainable Cell Rate (SCR), Maximum Burst Size (MBS), Minimum Cell Rate (MCR) and the conformance definition: the Generic Cell Rate Algorithm (GCRA) [2]

- **Quality of Service Parameters**: Cell Loss Ratio (CLR), average Cell Transfer Delay (CTD), peak-to-peak Cell Delay Variation (CDV)

- **Network Resources**: link capacities (C), memory size for buffering (Q)

3 The QoS Control

3.1 QoS specification

Recently there are five service classes with different traffic descriptors and QoS requirements defined in ATM. Correlation can be discovered between descriptor parameters and QoS requirements, which are specified in Table 3.1. Our assignment is mainly based on the ATM Forum specification [4].

The weighting function related to a service class should reflect the parameters specified in the appropriate column of Table 3.1. The end-to-end performance objectives of traffic contract should be allocated among the connection portions. We use the allocation principles specified in the standardization works, i.e. the CLR and CTD objectives are allocated by additive rules and CDV objectives are determined by the square root rule [1, 5]. In this way, our control method can be performed locally in the switches, because each switch has the performance objectives after the above decomposition for itself. Note that using the local performance objectives, local resource settings and the instantaneous local QoS information with the given Connection Traffic Descriptors the general end-to-end traffic control problem can be handled as a local traffic control problem in each switch. We avoid the overload caused by the transmission of lot of information necessary for a global control too. Also note, that this control can coexist with the end-to-end ABR control mechanism.

Attribute	CBR	rtVBR	nrVBR	UBR	ABR
Traffic Parameters					
PCR, CDVT	X	X	X	X	X
SCR, MBS, CDVT		X	X		
MCR					X
QoS Parameters					
CDV	X	X			
CTD	X	X			
CLR	X	X	X		X

Table 1: Table 3.1 Parameters of traffic contract and QoS requirements

3.2 The 3-dimensional abstract QoS space

To connect the quality of service requirements negotiated by the traffic contract to weighting function parameters we define a 3-dimensional state space with co-ordinates of measures of cell loss, delay and delay variation characteristics [8]. We choose for these measures the instantaneous CLR, CTD, and CDV parameters of a connection. That means each connection represented as a point of this space in each time slot (see Figure 3.1).

In this state space the QoS evolution of VCs can be observed where acceptance region can also be identified based on the negotiated QoS requirements. We define a cost function as an abstract distance of the actual QoS from the origin in the state space. The task of the traffic control method is thereafter formulated as to find the appropriate weighting functions such that:

- the actual QoS values for each VC should be within the negotiated region
- the total cost of all VC connections should be minimal

To fulfill these objectives we face with an optimization problem. One of the possible solutions

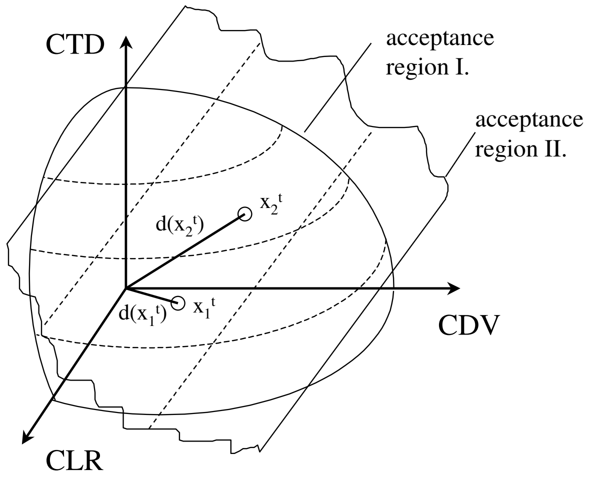

Figure 3.1 The abstract QoS state space

is to define weighting functions, which parameters are evaluated with a dynamic programming algorithm [10].

3.2.1 System equation

Consider the buffer partition in Figure 3.2. The lower index i refers to the service class ($i = 1..5$, i=1: CBR, i=2: rtVBR, i=3: nrVBR, i=4: ABR, i=5: UBR), the upper index (k) refers to the kth time slot. The state of the system at each time slot is specified by the following variables assigned to the ith queue at time (k): $n_i^{(k)}$: the number of processed cells; $l_i^{(k)}$: the number of discarded cells; $q_i^{(k)}$: the length of the queue (i.e. the number of cells in the queue); $\tau_i^{(k)}$: time stamp of the current HOL cell. This stamp is assigned to this HOL cell when entering the buffer. The aggregation of these quantities gives the system state column vector $(X_i^{(k)})^T = (n_i^{(k)}, l_i^{(k)}, q_i^{(k)}, \tau_i^{(k)})$. Q_i is the buffersize of the ith partition, $a_i^{(k)}$ is the number of arriving cells and $u_i^{(k)}$ is the number of served cells during the kth time slot. In this paper, for simplicity, we do not allow batch arrivals or batch departures, i.e. the latter two quantities are either 0 or 1. Note the a will correspond to the random disturbance while u to the control in the DP algorithm. Also note that all these variables are quantities which can be stored locally at the switch, which ease the formulation of a local control law.

With the above notation and assuming $l_i^{(0)} = 0$

lower index i	index of the service class
upper index (k)	number of the current time slot
$n_i^{(k)}$	the number of processed cells
$l_i^{(k)}$	the number of discarded cells
$q_i^{(k)}$	the number of cells in the queue
$\tau_i^{(k)}$	the time stamp of the current HOL cell
Q_i	the buffersize of the ith partition
$a_i^{(k)}$	the number of arriving cells
$u_i^{(k)}$	number of served cells

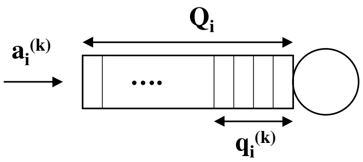

Figure 3.2 Notations of cell buffering

the system dynamics is described by the following discrete time system equations:

$$
\begin{aligned}
q_i^{(k+1)} &= min(q_i^{(k)} - u_i^{(k)} + a_i^{(k)}, Q_i) \\
l_i^{(k+1)} &= max(q_i^{(k)} - u_i^{(k)} + a_i^{(k)} - Q_i, 0) + l_i^{(k)} \\
n_i^{(k+1)} &= n_i^{(k)} + a_i^{(k)}
\end{aligned}
$$

Note that an alternative to (1) could be:

$$n_i^{(k+1)} = n_i^{(k)} + u_i^{(k)},$$

but we will assume that $n_i^{(k+1)}$ is large enough to make this difference irrelevant. Obviously, we have the following control constraint: $\sum_i u_i^{(k)} \leq 1$ for all k, indicating that at most one queue can get served at any one time.

3.2.2 Instantaneous QoS Characteristics

We proceed with relating cell loss probability $c_{i1}^{(k)}$, average (instantaneous) cell delay $c_{i2}^{(k)}$ and average (instantaneous) cell delay variation $c_{i3}^{(k)}$ to the system variables. These definitions are important, because they map the system description to the abstract model of the QoS state space $(i = 1..5)$.

$$c_{i1}^{(k+1)} = \frac{l_i^{(k)}}{n_i^{(k)}}$$

$$c_{i2}^{(k+1)} = \begin{cases} \frac{n_i^{(k)} * c_{i2}^{(k)} + (k - \tau_i^{(k)})}{n_i^{(k)}} & \text{if } u_i^{(k)} = 0 \\ c_{i2}^{(k)} & \text{if } u_i^{(k)} = 1 \\ & \text{OR } q_i^{(k)} = 0 \end{cases}$$

$$c_{i3}^{(k+1)} = \begin{cases} \frac{n_i^{(k)} * c_{i3}^{(k)} + (k - \tau_i^{(k)} - c_{i2}^{(k)})}{n_i^{(k)}} & \text{if } u_i^{(k)} = 1 \\ c_{i3}^{(k)} & \text{if } u_i^{(k)} = 0 \end{cases}$$

3.2.3 Cost Functional

In this subsection we define the cost functional applicable for optimization with dynamic programming, since it facilitates a straightforward formulation of the Bellman equation [6]. Throughout we restrict our attention to the case when a single traffic source generates traffic to each queue, i.e. each traffic class is represented by a single source (one VCC per queue). This assumption is not especially restricting, since with proper aggregated traffic models the multiple source per service class case can readily be represented. To obtain an overall scaler valued cost functional suitable for optimization, we first need the cost function of a single session, i.e. the cost assigned to a VCC:

$$
\begin{aligned}
J_i^{(k)}(c_i^{(k)}) &= J_i^{(k)}(c_{i1}^{(k)}, c_{i2}^{(k)}, c_{i3}^{(k)}) = ||J_i^{(k)}|| = \\
&= \sqrt{f_{i1}^{(k)^2}(c_{i1}^{(k)}) + f_{i2}^{(k)^2}(c_{i2}^{(k)}) + f_{i3}^{(k)^2}(c_{i3}^{(k)})}
\end{aligned}
$$

where the $f_{i1}^{(k)}(.)$, $f_{i2}^{(k)}(.)$, $f_{i3}^{(k)}(.)$ functions are the loss, delay and delay-variation weighting functions, respectively, of VCC_i, and $c_i^{(k)}$ is a column vector, corresponding to the QoS of VCC_i at time k, $(c_i^{(k)}) = (c_{i1}^{(k)}, c_{i2}^{(k)}, c_{i3}^{(k)})$. Assuming stationarity and equal VCC (session) weighting functions for all i, we can neglect the dependency of the weight functions from the i_1, i_2 and i_3 parameters, as well as from the time index (k). In other words, we simply have:

$$J_i(c_i^{(k)}) = \sqrt{{f_1}^2(c_{i1}^{(k)}) + {f_2}^2(c_{i2}^{(k)}) + {f_{i3}}^2(c_{i3}^{(k)})}$$

Let Γ denote the overall cost functional vector, and C denote the cost matrix:

$$\Gamma(C^{(k)}) = (J_i(c_i^{(k)})); C^{(k)} = (c_i^{(k)}); c_i^{(k)} = (c_1^{(k)}, c_2^{(k)}, c_3^{(k)})$$

With this notation the overall scaler cost functional to be minimized takes the form:

$$\gamma(C^{(k)}) = \sum_{i=1}^{5} J_i(c_i^{(k)}) \text{ subject to } C^{(k)} \leq (QoS)_{i,j}$$

where $i = 1..5, j = 1..3$, and the QoS matrix contains the negotiated QoS parameters (CLR, CTD, CDV), see Subsection 2.2. With the above formulation of the optimization problem we are facing the challenge of a DP task, where we wish to find the optimal (and feasible) control law μ which only depends on the state of the physically observable state vector X, such that the cost functional is optimized over the 3 dimensional QoS state space represented by the C matrix. The other way is to solve the optimization problem is using direct cost functions in the algorithm. This means that instead of weighting functions cost functions are evaluated before departure and the HOL cell of the most "expensive" queue must be sent. The cost function should be discounted in order to slowly forget the past. We are recently working on this topic.

3.3 An example for the set of weighting functions

These objectives are mathematically formulated as follow [10]:

$$
\begin{aligned}
W_1 &= a_1 * \frac{LC_1}{SUM_1 * CLR_1} + b_1 * \frac{T_1}{CTD_1} + \\
&\quad + c_1 * max(T_1 - CTD_1 - \frac{2}{3} * CDV_1, 0) \\
W_2 &= a_2 * \frac{LC_2}{SUM_2 * CLR_2} + b_2 * \frac{T_2}{CTD_2} + \\
&\quad + c_2 * max(T_2 - CTD_2 - \frac{2}{3} * CDV_2, 0) \\
W_3 &= a_3 * \frac{LC_3}{SUM_3 * CLR_3} \\
W_4 &= a_4 * \frac{LC_4}{SUM_4 * CLR_4} \\
W_5 &= \begin{cases} w_5 & \text{if } K_1,\ K_2,\ K_3,\ K_4 \text{ are all} > 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

where:

$$K_1 = \frac{d_1 * (a_1 + b_1 + c_1)}{W_{CBR}}; K_2 = \frac{d_2 * (a_2 + b_2 + c_2)}{W_{rtVBR}};$$

and

$$K_3 = \frac{d_3 * a_3}{W_{nrVBR}}; K_4 = \frac{d_4 * a_4}{W_{ABR}}$$

The value of a weighting function is equal to minus infinity if the queue is empty. Let be LC_i is the number of lost cells of class i, SUM_i is the total number of cells of class i, and T_i is the waiting time of HOL cell in the queue of class i. Note that $[x]^+$ is equal to x if $x > 0$ else 0. These weighting functions obtained by heuristics based on Table 3.1, that means they reflect the different service classes sensitivity to cell loss, delay and delay variations and also take into account the required QoS parameters. Specifically, the weighting parameters a_i, b_i, c_i and d_i are to determine the relative "importance" of a given QoS parameter in the weight of a given service class, while the constants CLR_i, CTD_i, and CDV_i are the negotiated (contracted) cell loss ratio, cell transfer delay and cell delay variation of the respective VCC's. These latter three parameters are referred to as QoS in this paper.

4 Performance Evaluation

4.1 The input traffic

In the next following simulation scenarios we consider a link of capacity 45 Mbps, and a multiplexer with 5 input ports corresponding to the 5 service classes. The basic state of the traffic sources is the following: The CBR source is of 1.5 Mbps representing DS-1 circuit emulation. The rtVBR, nrVBR and UBR sources are all bursty and modeled as Interrupted Bernoulli Processes (IBPs) and are characterized by their peak and sustainable cell rates. The ABR source is assumed to be of rate based and is also modeled by an IBP. It is characterized by its peak and minimum cell rate (see Table 4.1). We have given the burstiness parameters of all services measured by the squared coefficient of variation of the interarrival time (i.e. the c^2 parameter).

	PCR	SCR	MCR	c^2
CBR	1.5	-	-	0
rtVBR	15.0	3.0	-	9.44
nrVBR	22.5	1.0	-	20.75
UBR	45.0	5.0	-	26.06
ABR	22.5	-	4.5	-

Table 4.1 Basic input traffic characteristics (the rates are given in Mbps)

Note that with the above link capacity a time slot in our discrete time model correspond to 9.422 μs, which will be used as the time unit in the CTD and CDV values below. Tables 4.2-4.4 display the QoS requirements of different services, the buffer sizes available for different service classes and an appropriate parameter set for weighting functions, respectively. These weighting function parameters obtained by heuristics. Note that no delay or delay variation parameters are negotiated for the nrVBR or the ABR service classes and no QoS requirements are given for the UBR service.

	CBR	rtVBR	nrVBR	UBR	ABR
CLR_i	10^{-5}	10^{-6}	10^{-7}	-	10^{-7}
CTD_i	3.0	5.0	-	-	-
CDV_i	1.0	2.0	-	-	-

Table 4.2 The QoS requirements
(CTD and CDV requirement are given in time unit)

Service class	CBR	rtVBR	nrVBR	UBR	ABR
Buffersize	5	8	12	250	80

Table 4.3 Buffer sizes in cells

	a_i	b_i	c_i	d_i
CBR	0.1	0.6	0.9	0.5
rtVBR	0.2	0.3	0.4	0.5
nrVBR	0.6	-	-	0.7
UBR	-	-	-	-
ABR	0.4	-	-	0.6

and $w_5 = 6.0$

Table 4.4 The parameter set of weighting functions

4.2 The QoS dependence on CBR load

Figures 4.1-4.3 display simulation result on CLR, CTD and CDV respectively, when we increase the CBR load from 1.5 Mbps up to 7.5 Mbps and the other sources are in basic state. In this example we consider a single multiplexer with the weighting function parameter set described above. Due to the lower utilization of the connections (between 0.73 and 0.87) there is a considerable decrease in the QoS parameters of the traffic classes, which have a strict traffic contract with the network. We can see that all the negotiated QoS parameters met their requirements. The CLP and CDV of the CBR service class is slightly increasing according to the increasing load, but this increase effects the increase of the value of the weighting function of CBR class, i.e. the CBR service class gets more bandwidth and the QoS parameters finally rest within the negotiated region.

UBR service class has no any QoS requirements, so the load change causes changes only in the QoS parameters of this service, as it can be seen in the Figures 4.1-4.3.

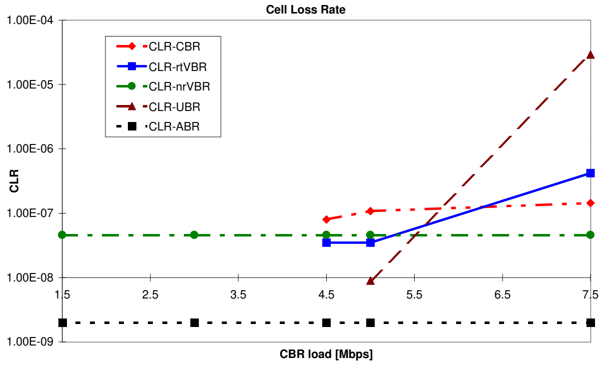

Figure 4.1 Cell Loss Ratio vs. CBR load

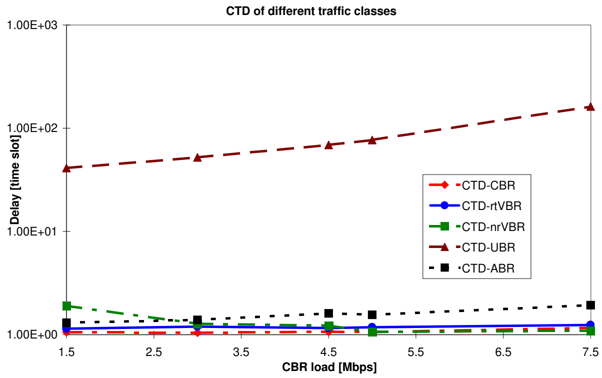

Figure 4.2 Cell Transfer Delay vs. CBR load

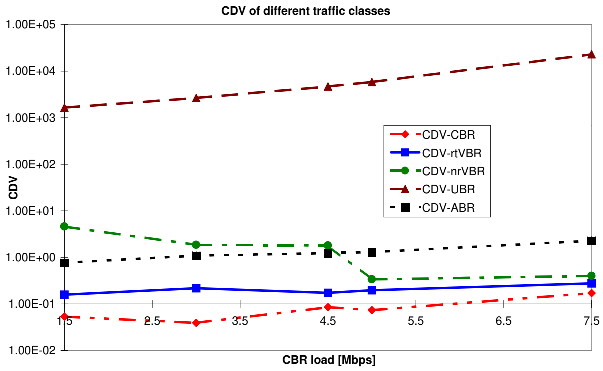

Figure 4.3 Cell Delay Variation vs. CBR load

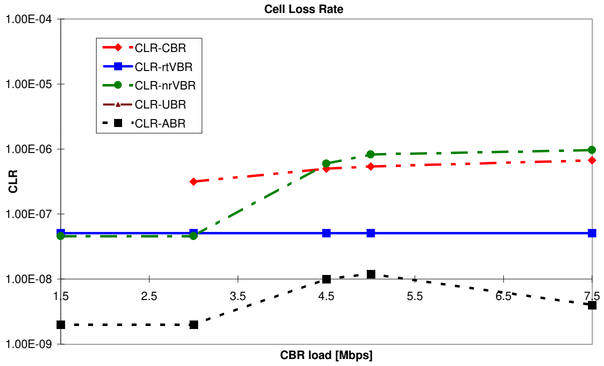

Figure 4.4 Cell Loss Rate vs. CBR load under heavy UBR traffic

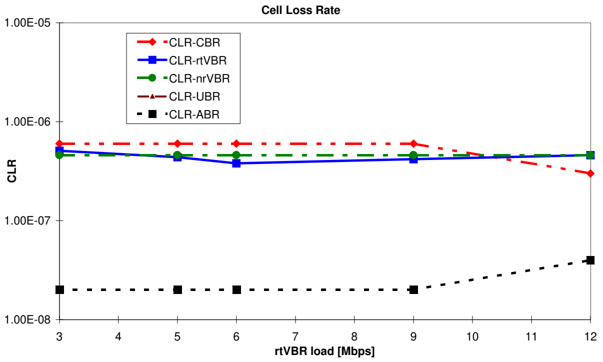

Figure 4.7 Cell Loss Rate vs. rtVBR load under heavy UBR traffic

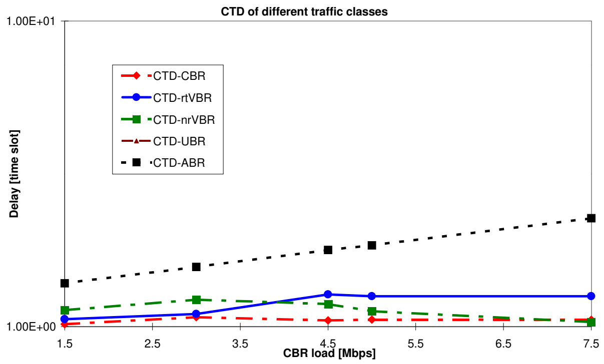

Figure 4.5 Cell Transfer Delay vs. CBR load under heavy UBR traffic

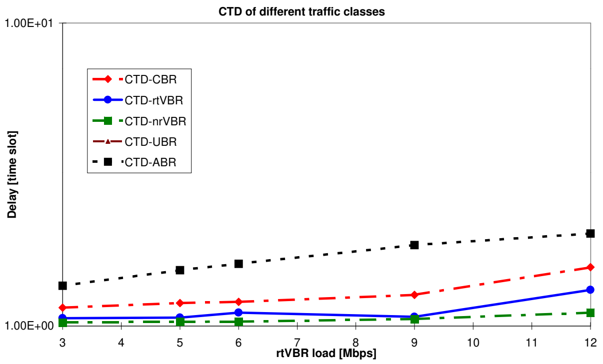

Figure 4.8 Cell Transfer Delay vs. rtVBR load under heavy UBR traffic

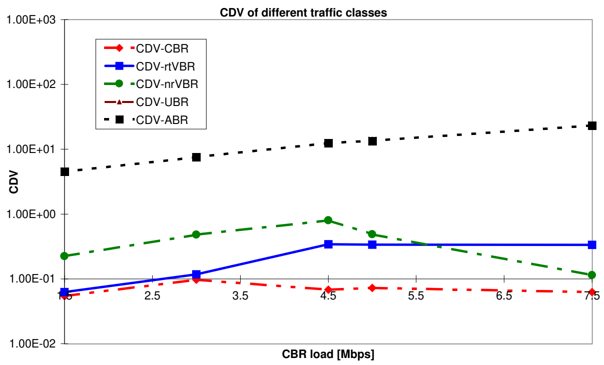

Figure 4.6 Cell Delay Variation vs. CBR load under heavy UBR traffic

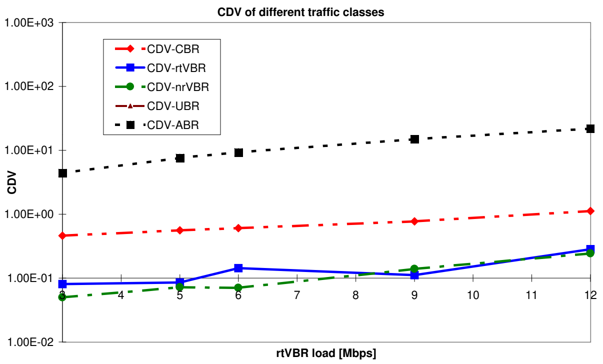

Figure 4.9 Cell Delay Variation vs. rtVBR load under heavy UBR traffic

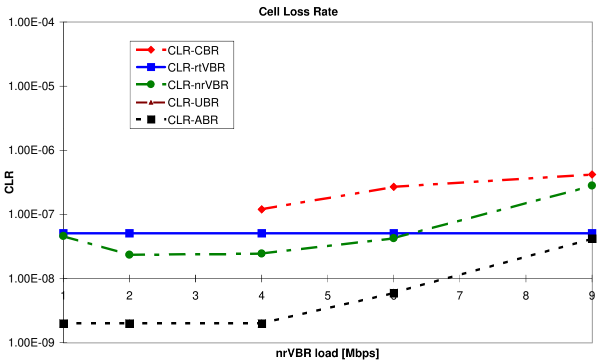

Figure 4.10 Cell Loss Rate vs. nrVBR load under heavy UBR traffic

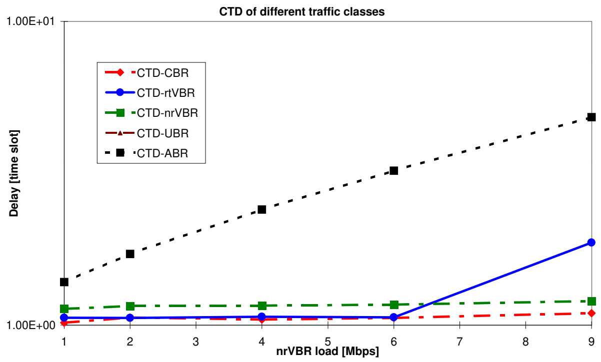

Figure 4.11 Cell Transfer Delay vs. nrVBR load under heavy UBR traffic

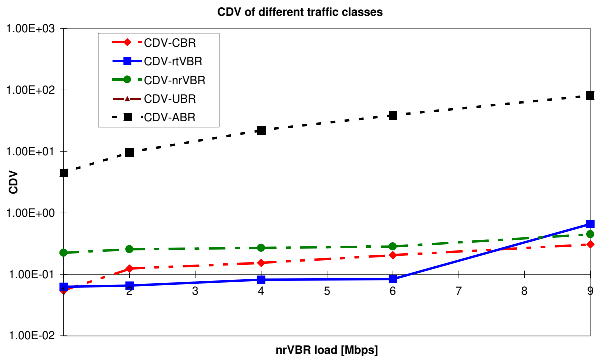

Figure 4.12 Cell Delay Variation vs. nrVBR load under heavy UBR traffic

In the following scenario, we increase the sustainable cell rate of UBR traffic to 12 Mbps end we set the parameter d_3 to 0.9. The remaining three sources are in basic state and the other parameters are the same as in the previous scenario.

Figures 4.4-4.6 display the QoS parameters of a highly utilized link. The utilization goes from 88% up to 95%. The CLP parameters are similar to the previous case. The guaranteed services have constant cell loss except CBR, which has an increase by a decade. This resulted in the slow decreasing of the CDV parameter. The CDV of the other regarded class (rtVBR) is normal. The nrVBR traffic class has no CDV assurance; the non-monotony of the curve comes from the abrupt step of its CLP at the same point.

Observe that the load increase affects the CLR of UBR only, as desired, since all other classes have strictly prescribed CLR values. The same behaviour can be observed for the CTD and CDV parameters of CBR and rtVBR classes. The ABR class is congestion controlled and sensitive to CLR only so its CTD and CDV behaviour is determined by the other classes.

4.3 The QoS dependence on VBR load

In the next following simulation studies we examine the dependence of QoS parameters on the increasing load of VBR traffic. In Figures 4.7-4.9 the load of rtVBR goes from 3 Mbps up to 12 Mbps. The sustainable cell rate of UBR source is set to 15 Mbps and the d_3 is set to 0.9; the other sources and parameters are in basic state.

The utilization is about 0.97 in the Figures 4.7-4.9. In this cases the CLP requirements of nrVBR and ABR classes are increased to 10^{-6} and 10^{-7}, respectively. Because rtVBR is a bursty traffic, there are more significant changes in the QoS parameters of the guaranteed classes. The CDV of CBR class gets in the near of QoS requirement (1.0). This, in consideration of the increasing average delay of CBR, effects the decreasing of CLP at the last measuring point. The other curves meet their QoS requirements. In Figures 4.10-4.12 the load of nrVBR goes from 1 Mbps up to 9 Mbps. The sustainable cell rate of UBR source is set to 18 Mbps and the d_3 is set to 0.7; the other sources and parameters are in basic state.

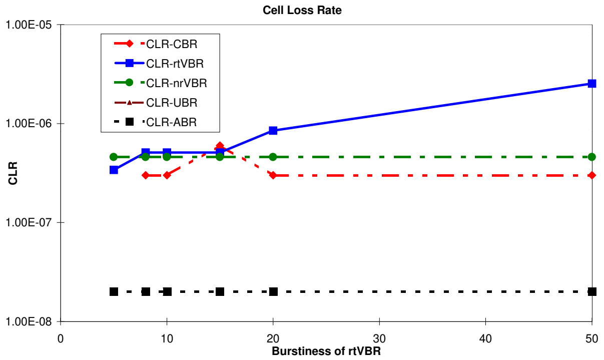

Figure 4.13 Cell Loss Rate vs. rtVBR burstiness under heavy UBR traffic

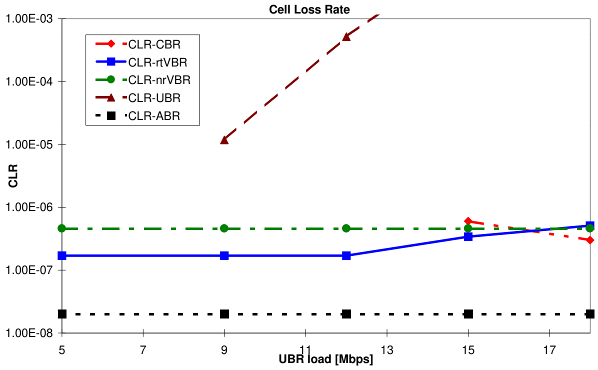

Figure 4.16 Cell Loss Rate vs. UBR load

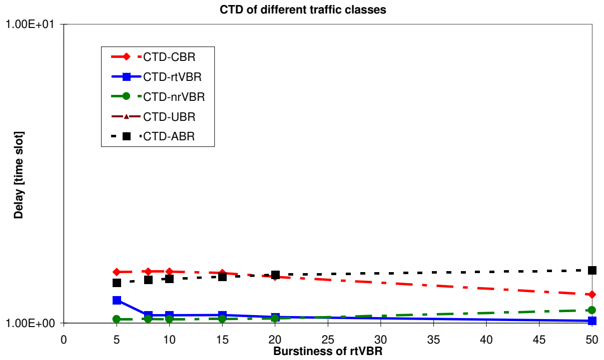

Figure 4.14 Cell Transfer Delay vs. rtVBR burstiness under heavy UBR traffic

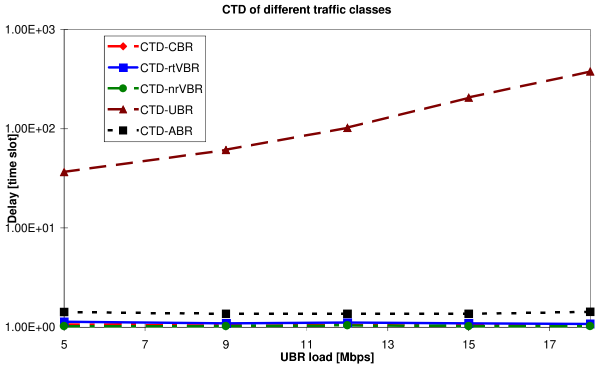

Figure 4.17 Cell Transfer Delay vs. UBR load

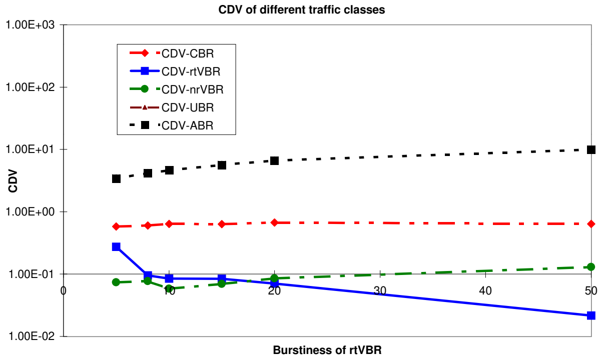

Figure 4.15 Cell Delay Variation vs. rtVBR burstiness under heavy UBR traffic

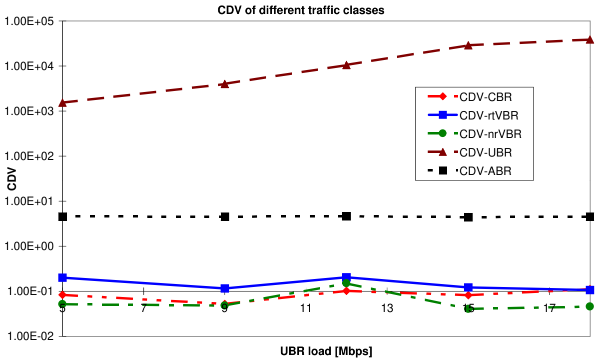

Figure 4.18 Cell Delay Variation vs. UBR load

4.4 The QoS dependence on the burstiness on VBR

In the next following simulation studies we examine the dependence of QoS parameters on the increasing burstiness of VBR traffics. In Figures 4.13-4.15 the burstiness of rtVBR (measured by the squared coefficient of variation of the rtVBR interarrival time) goes from 5 up to 50. The sustainable cell rate of UBR source is set to 18 Mbps and the load of rtVBR source is 3 Mbps; the other sources and parameters are in basic state.

In the Figures 4.13-4.15 can be seen excellently, that the weighting functions handle the different services independent from each other. Real-time VBR traffic with increasing burstiness is arriving to the short buffer described in Table 4.3. The CLP of the rtVBR has linear increase with the burstiness. This causes a decreasing in the CTD and CDV of the rtVBR, but for other classes it seems to be neutral.

4.5 The QoS dependence on UBR load

At the last we show how is the dependence of the QoS parameters of service classes on the increasing load of UBR traffic. In Figures 4.16-4.18 the load of UBR goes from 5 Mbps up to 18 Mbps. Note that the burstiness of UBR traffic is constantly 26.06 in all cases. The other sources and parameters are in basic state.

The increase of the UBR load does not have any impacts on the QoS parameters of the other classes. It can be seen in Figure 4.17. The average cell transfer delay of UBR traffic significantly increases, while other classes have the same CTD. Note that in our model the UBR service is not totally transparent for the other services. However, there are cells of other classes in the buffer, it maybe delivered an UBR cell, because of the adaptability of our model. We give a chance to the UBR if all other classes meet their QoS with a given reserve. Although, the UBR has poor prestige in the network, if the other services needs the bandwidth.

4.6 Comparison to static scheduling schemes

We made simulations to compare the performance of our algorithm with static scheduling rules - FCFS and Round Robin. Our results show that to achieve the same CLP the static traffic control schemes need 15-20% more buffer space for the guaranteed services, moreover, the utilization of the network decreases. With FCFS the delay requirements can not be fulfilled. In the Round Robin-case the delay is limited by the number of served queues. We have an ongoing research on the detailed comparison of traffic control methods.

5 Conclusions

We have considered the issue of optimal cell scheduling in an integrated services ATM network and proposed a general traffic control framework which is based on a complete buffer partitioning architecture and on an adaptive weighting function based buffering schedule. The method can incorporate all the presently considered service classes with their diverse QoS requirements and it is capable of providing an optimal scheduling considering also the temporary traffic load at the switches with a simple information processing which requires only summation and multiplication.

We suggest a dynamic programming solution for the optimization problem. Moreover, in the paper a performance evaluation study of the control framework is demonstrated with several examples investigating the QoS dependence on CBR load, VBR load, VBR burstiness and UBR load. ¿From the results we can conclude that the QoS characteristics of each service are within the negotiated QoS region and that the remaining resources are efficiently used by best effort type service classes. We can see that the utilization is achieved by keeping the actual QoS characteristics close to the negotiated parameters rather than overfulfilling them. The examples also show the advantage of statistical multiplexing of the five different service classes sharing only 45 Mbps capacity instead of 106.5 Mbps, which would be the case of peak rate allocation.

We can conclude that the proposed traffic control scheme is capable to keep traffic contracts for the classes with strict QoS requirements by an optimal resource sharing and also distributes resources to the best effort type service classes as such resources become available. In our future research we concentrate to find optimal weighting functions for different scenarios, and apply and analyze the method for different traffic environments with using real traffic sources.

References

[1] ITU-T Recommendations I.356 *B-ISDN ATM Layer Cell Transfer Performance*, 1996.

[2] ITU-T Recommendations I.371 *Traffic Control and Congestion Control in B-ISDN*, 1995.

[3] ATM Forum, *ATM User Network Interface Specification Version 3.1*, September 1994.

[4] ATM Forum, *Traffic Management Specification Version 4.0*, April 1996.

[5] ATM Forum, *B-ISDN Inter Carrier Interface (B-ICI) Specification Version 1.0*, August 1993.

[6] Dimitri P. Bertsekas, "Dynamic Programming, Deterministic and Stochastic Models", *Prentice-Hall Inc., Englewood Cliffs, N.J. 07632*, 1987.

[7] S. Blaabjerg, S. Molnár, "Methods for UPC Dimensioning of a CDV Perturbated Cell Stream", RACE BRAVE Workshop, Milano, Italy, June 14-15, 1995.

[8] F. Bonomi, K. W. Fendick, "The Rate-Based Flow Control Framework for the Available Bit Rate ATM Service", *IEEE Network*, p 25-39, March/April 1995.

[9] I. Cidon, L. Georgiadis, R. Guérin, A. Khamisy, "Optimal Buffer Sharing", *IEEE J-SAC*, vol. 13, No. 7, September 1995.

[10] G. Fodor, T. Marosits, S. Molnár, "A General Traffic Control Framework in ATM Networks", in *Proceedings of the IEEE ICCS/ISPACS '96*, p 160-164, Singapore, November 25-29, 1996.

[11] P. A. Ganos, M. N. Koukias, G. K. Kokkinakis, S. A. Kotsopulos, "A Novel Dynamic Priority Scheduling Method for Multiple Classes of ATM Traffic in an ATM Statistical Multiplexer", in *Performance Modelling and Evaluation of ATM Networks*, Vol. 1, Ed. Demetres Kouvatsos, IFIP, Chapman and Hall, 1995.

[12] Erol Gelenbe, Vilay Srinivasan and Sridhar Seshadri, "Single Node and End-to-End Buffer Control in Real Time", in *Performance Modelling and Evaluation of ATM Networks*, vol. 1, Ed. Demetres Kouvatsos, IFIP, Chapman and Hall, 1995.

[13] Harold J. Kusnher, "Analysis of Controlled Multiplexing Systems via Numerical Stochastic Control Methods", *IEEE J-SAC*, vol. 13, No 7, September 1995.

[14] A. Y.-M. Lin, J. A. Sylvester, "Priority Queueing Strategies and Buffer Allocation Protocols for Traffic Control at an ATM Integrated Broadband Switching System", *IEEE J-SAC*, vol. 9, No. 9, p 1524-1536, December 1991.

[15] J. W. Mark, Jing-Fei Ren, "A Traffic Management Framework for ATM Networks", *IFIP Performance Modelling and Evaluation of ATM Networks*, Chapman and Hall, 1996

[16] A. K. Parekh, R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single Node Case", *IEEE/ACM Transactions on Networking*, 1, p 344-357, 1993.

[17] S. Rajagopal, V. G. Kulkarni and S. Stidham, Jr., "Optimal Flow Control of a Stochastic Fluid-Flow System", *IEEE J-SAC*, vol. 13, No 7, September 1995.

[18] J.-F. Ren, J. W. Mark, J. W. Wong, "A Dynamic Priority Queueing Approach to Traffic Regulation and Scheduling in B-ISDN",*Proceeding of the IEEE Globecom*, p 612-618, 1994.

[19] Chen ShanZi and Yan Liemin, "A New Priority Control of ATM Output Buffer", *Telecommunication System Journal*, 4, p 61-69, 1995.

[20] S. Valaee, M. A. Kaplan, "Congestion Control in a Constraint-Worst-case Framework",*IFIP Performance Modelling and Evaluation of ATM Networks*, Chapman and Hall, 1996

[21] Hui Zhang, "Service Disciplines for Guaranted Performance Service in Packet-Switching Networks", *Proceedings of the IEEE*, 83(10), October 1995.