EFFICIENT FAIR QoS CONTROL USING STOCHASTIC MULTIOBJECTIVE RESOURCE ALLOCATION STRATEGIES

Menotti Machado

Telecommunications National Institut/RST
9 Rue Charles Fourier 91011 - EVRY Cedex - France
phone: +33-1-60-76-6664 fax: +33-1-60-76-4291
{Menotti.Machado@int-evry.fr }

ABSTRACT

In this paper, we present a model to evaluate and to control the Quality of Service end-to-end (QoS) in packet networks by using optimal resource allocation techniques. This is done controlling the QoS for each Class of Service (CoS), subject to QoS requirements. We consider fixed trade-offs among the QoS of the classes. Our QoS model is based on three traffic performance metrics. We evaluate and compare six strategies to solve this problem by resource allocation. The technics based on stochastic multiobjective optimization allocated the network resources of form to supply good Network mean QoS values and fair CoS mean QoS values. The simulation was realized on an IP network considering a multiplexer placed on the network edge as the control dispositive.

1. INTRODUCTION

The growth and the popularization of the Internet turned real the possibility of offering new services, especially the ones of the multimedia type. In a first observation, it was verified that the approach of sending packets based on the *better effort* showed inapt to the statistic characteristics of that type of traffic. With the purpose of enabling the Internet to offer these services, several concepts are being incorporated to its strategies of traffic control. In the scope of this work we define end-to-end Quality of Service (QoS) in Internet [1] in two aspects: the objective aspects related to traffic characteristics and network state and the subjective aspects observed by end users and network operators. For defining our functional QoS we use three performance metrics [2] calculated end-to-end: the mean delay , jitter, the mean loss of packets . Each performance metric is calculated considering the stochastic traffic behavior and the availability of network resources. We assumed that all network nodes support the RSVP protocol [3] for transmitting the traffic contracts and the periodic information about the state of all network components. Our formulation allows the differentiation of the class of service (DiffServ) [4], by the information obtained in the traffic contracts and QoS satisfaction levels desired by network operator. The experiments were realized on a component placed in the edge of the network, namely multiplexer.

In this work, we compare six strategies for the QoS control by resources allocation. We analyzed the results observing the adaptive capacity of the strategies to the traffic stochastic behavior and the fairness obtained among QoS levels of the CoS. By the results obtained, we concluded that the techniques based on stochastic multiobjective optimization represent a good alternative strategy to solve this kind of problem.

2. THE MODELS

2.1. End-to-End Quality of Service (QoS) Model

The control of Quality of Service in the Internet is focused as a fundamental problem to allow the support for the multimedia traffic. We defined the network mean QoS as a combination of the mean values obtained to the QoS of all N CoS,

$$QoS(\cdot) = f(E\{QoS_1(\cdot)\}, \ldots, E\{QoS_N(\cdot)\}). \quad (1)$$

We assumed that the network operator is capable to state the relative importance of the QoS of the CoS i to the network QoS by the (ω_i) weighting terms. These factors represent the CoS differentiation indexes and the fairness wanted in the resource allocation process from the point of view of QoS. We defined the QoS for each CoS as a combination of the mean values of the their M performance metrics,

$$QoS_i(\cdot) = f(E\{\Psi_{i1}(\cdot)\}, \ldots, E\{\Psi_{iM}(\cdot)\}), \quad (2)$$

calculated end-to-end. We assumed that the network operator is also capable to define the relative importance (γ_{ij}) of the performance metrics j for the QoS function of each CoS i. The performance metrics Ψ_{ij} were defined by the traffic statistical characteristics and the state of the network resources. In order to define the QoS to each CoS, we used a combination of the mean values of three metrics of performance, the transmitting delay $D(\cdot)$, loss of packet $P(\cdot)$ and jitter $J(\cdot)$,

$$\Psi_{i1} = \{P_i(\cdot)\},\ \Psi_{i2} = \{D_i(\cdot)\},\ \Psi_{i3} = \{J_i(\cdot)\}, \quad (3)$$

where $E\{\cdot\}$ represents the mathematical expectation.

In order to calculate the end-to-end values of each parameter of QoS for each CoS, we considered the additive form to the mean delay and to the jitter and the multiplicative form to the probability of loss of packets [5]. We tested the QoS control strategies considering the model of classes of *Differentiated Services (DiffServ)*, proposed to IETF in [6] and [7]. This model defines their CoS by of priorities to delivering the packets of the aggregate traffics. In our model, the differentiation of the CoS is obtained by the factors ω_i, that form the priority levels that provide QoS to the several classes.

2.2. Network and Traffic Model

We consider an IP multihop network model where the control of end-to-end QoS is made by the components placed on the edges (access points). We suppose this component is a multiplexer with finite bandwidth capacity R and finite buffer capacity B. Its buffer is organized in individual FIFO (First-In First-Out) queues, one for each CoS (see Figure 1). This device multiplexes N independent and identically distributed (i.i.d.) classes of traffic of packets onto their output link. The packet arrival intervals follow a *pareto* probability distribution with parameter α_i. The size of the packets in the classes are i.i.d. and follow a binomial probability distribution with parameters M_i and s_i. Each class i presents an equivalent bandwidth r_i expressed in bits per second. The service probability law depends on the scheduling scheme adopted.

The multiplexer presents a control mechanism for taking periodical decisions about the optimal values of the bandwidths and of the spaces in the buffer for each CoS. According to these values, each queue is served during an interval of time T_i, in which it receives the whole bandwidth of the output link. Limited on this time, the packets in the queue i enter immediately in service after the end of the service of the precedent packet. For the minimization of the loss packets, we consider that the buffer capacity B_i of each queue is also calculated for each interval of time T_i. We assumed that a packet can not fragmented. The control mechanism proposed presents an interface developed to receive the statistical measures of the multiplexer, the QoS contract information and the external performance measures. We consider that information is transported by the PATH and RESV messages of the IP signalling protocol RSVP.

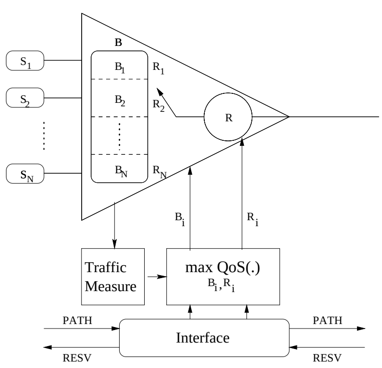

Figure 1: The Multiplexer.

We consider that the components in the core of the network are responsible just for routing and for reclassifying the packets, when necessary.

3. PROBLEM FORMULATIONS

3.1. Stochastic Multiobjective Formulation

The main objective of this formulation is to represent the QoS network as a function of the traffic stochastic characteristics and the network stochastic state of a representative form to the network operator. For your resolution we used the optimal resource allocation approach subject to constraints for maintaining the QoS network in the their best possible values and the QoS classes in the fairest values. The optimal values of the resources are obtained respecting simultaneously all the established traffic contracts, named here *QoS contracts*. To achieve that, we maximize the normalized mean values of QoS of each class i, $\max E\{(qos_i)\}$. This network problem was modelled as a multiobjective optimization problem where multiple stochastic criteria are evaluated simultaneously under a group of constraints [8], [9]. This problem is presented in the following way:

$$\begin{aligned} &\max_{x \in \mathcal{X}} QoS = E\{< \omega_i, QoS_i(x) >\} \\ &s.t.\ E\{QoS_i(x) - QoS_{i_{min}}(x)\} \geq 0 \\ &\forall i = 1, \ldots, N, \end{aligned} \quad (4)$$

where,

$$\begin{aligned} &QoS_i = E\{< \gamma_{ij}, \Psi_{ij}(x) >\} \\ &E\{\Psi_{ij}(x) - \Psi_{ij_{max}}(x)\} \leq 0, \forall i = 1, \ldots, N. \end{aligned}$$

We considered that the functions, QoS, QoS_i and Ψ_{ij}, $\forall i, j$, are convex on $\mathcal{X}$ convex. And $QoS_{i_{min}}$ is the minimum QoS value admitted for each class i. The space of the network resources x is represented by $\mathcal{X}$. In this strategy, the allocated resources were represented by the relative bandwidth R_i (with relation to times of services T_i) and by the fractions of the buffer capacity, B_i, attributed to each queue. This formulation allows that several procedures of fairness and of sharing can be associated to.

3.2. Resource Allocation Strategies

The strategies of QoS control consider only the active connections (connections already admitted).

Strict Priority (S): In this strategy, we attributed a fixed service priority to each CoS. Classes of high priority degree are served before classes of lesser priority degree. In case of loss of the packet, the less priority classes lost packets before classes high priority. In ours experiments, we adopted the values of CoS differentiation factors ω_i to define the priority degrees. In the case of the differentiation scheme with equal factors, we adopted the random selection.

Queues with Guaranteed Resources (G): In this scheme, we defined previously fixed fractions, R_i and B_i of global resources, R and B, to each CoS using the fixed weighting factors α_i and β_i. These factors were calculated by the values of CoS differentiation weight ω_i,

$$\sum_{i=1}^{N} R_i \leq R, \quad R_i \leq \alpha_i R, \ \ \forall i = 1, \ldots, N,$$
$$\sum_{i=1}^{N} B_i \leq B, \quad B_i \leq \beta_i B, \ \ \forall i = 1, \ldots, N,$$
$$\alpha_i = \beta_i = \frac{\omega_i}{\sum \omega_j}, \ \ \forall i, j = 1, \ldots, N$$
$$\sum_{i=1}^{N} \beta_i = 1, \quad \sum_{i=1}^{N} \alpha_i = 1$$

Weighted Fair Queue (W): In this method, we calculated dynamically the weighting factors, α_i and β_i, based on the of of CoS differentiation factors ω_i by the measured CoS mean QoS, $Q\hat{o}S_i$. The resources were allocated in function of this factor,

$$\alpha_i = \beta_i = \frac{\omega_i \cdot Q\hat{o}S_i}{\sum \omega_j \cdot Q\hat{o}S_j}, \ \ \forall i, j = 1, \ldots, N$$

Multiobjective Weighted Fair Queue (MO): This technique represents a variant of the previous case, where we incorporated the aspects of multi-objective optimization theory (described in Section 3.1) to the resource allocation.

Complete Sharing (C): Excepting for multiplexing strategy with guaranteed resources, in the others strategies we used the complete sharing of the resources. In this case, every CoS used all the resources, R and B, to their exhaustion,

$$\sum_{i=1}^{N} R_i \leq R \ \ 0 \leq R_i \leq R, \ \ \forall i = 1, \ldots, N$$
$$\sum_{i=1}^{N} B_i \leq B \ \ 0 \leq B_i \leq B, \ \ \forall i = 1, \ldots, N$$

4. SIMULATION

4.1. Discrete Event Model

For the simulation we implemented an event discrete simulator [10] considering the *Arrival First* (AF) model [11] in order to account the simultaneity of the arrival and departure instants of the packets in the multiplexer. To help the resource allocation allocation decision making, the controller considers the following sequence of events: the packet arrivals, the multiplexer statistical measures, the reception of the external information, the calculation the resources allocation and the packet service. This allocation resource decision is hold until a new alteration is needed. In the simulations, we used a periodic interval T for the alteration of the allocation of resources.

Parameters of Multiplexer		
R (Mbps)	B (Mbits)	T (s)
30	9	1

Table 2: Parameter of Multiplexer.

	CoS			
Parameters	1	2	3	4
M_i(Kbits)	12	12	12	12
s_i	0.67	0.67	0.67	0.67
α_i	2.50	2.50	2.50	2.50
r_i(Mbps)	1-27	9	6	3
$D_{i max}$(s)	0.30	0.30	0.30	0.30
$J_{i max}$(s)	0.10	0.10	0.10	0.10
$P_{i max}$(%)	0.10	0.10	0.10	0.10
$D_{i tot}$(s)	0.01	0.01	0.01	0.01
$J_{i tot}$(s)	0.01	0.01	0.01	0.01
$P_{i tot}$(%)	0.01	0.01	0.01	0.01

Table 3: Description of the Traffic.

4.2. Scenario Description

In the simulation, we considered three *differentiation schemes* named $W1$, $W2$ and $W3$. For each scheme we assumed that values of ω_i are known *a priori* and represented by fixed values (static differentiation) [9] and [8], presented in Table 3. We considered that the performance metrics are equally important for all classes ($\gamma_{ij} = 1/3, \forall i, j$) and that their values accumulated from other components present a maximal value represented by Ψ_{tot}. These suppositions simplified the simulation and allowed a better analysis about the advantages of the used method with relation to the characteristics of traffic of the CoS. We used four traffic source representing the CoS that generated packets with the characteristics described in Section 2.2 and presented in Table 3. Table 13 describe the used symbols. The physical parameters of the multiplexer are described in Table 2. We changed the network load by the increasing the equivalent bandwidth of the CoS 1. Analysis was performed to compare the obtained values to network mean QoS, the dispersion of the mean QoS to all classes, considering six resource allocation strategies. We used a time of simulation of 200 seconds, with 10 repetitions for each experiment.

4.3. Results

We compared the strategies of resource allocation considering different combinations to ways of sharing R and B based on the strategies described in Section 3. The results are presented in Figures 6 12, considering the strategy symbols of Table 5.

4.4. Analysis of the Results

To compare the resource allocation strategies we consider the mean values obtained to the QoS, the output link utilization rate and the buffer occupation rate at the network and class levels considering an input load between 0.80 and 0.98. We used the

	CoS Differentiation Indexes			
Differentiation Schemes	1	2	3	4
$W1$	0.25	0.25	0.25	0.25
$W2$	0.10	0.20	0.30	0.40
$W3$	0.40	0.30	0.20	0.10

Table 4: Differentiation Schemes.

Acronyms	**Strategies of Allocation**
GRB	Guaranteed R_i and B_i;
WRB	Weighted R_i and B_i;
SRB	Strict Priority R_i and B_i;
$MORGB$	Multiobjective R_i and Guaranteed B_i;
$MORWB$	Multiobjective R_i and Weighted B_i;
$MORB$	Multiobjective R_i and B_i;

Table 5: Symbols of the Strategies.

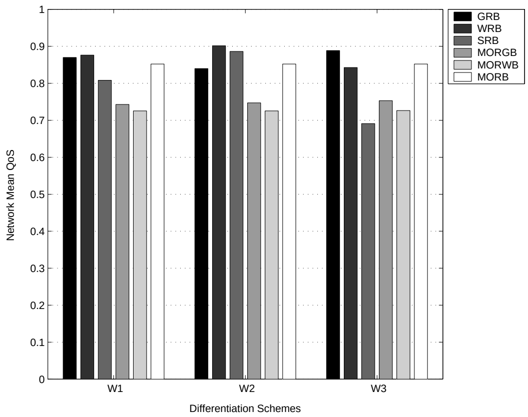

Figure 6: Network Mean QoS: Differentiation Schemes $W1$,$W2$ and $W3$.

dispersion of the mean values obtained in class level as a fairness criteria. According to Figure 6, we observed that GRB, WRB and $MORB$ strategies present the best values of mean QoS at network level considering the three differentiation schemes. The WRB strategy presented the best QoS values when the CoS differentiation indexes are equal or larger to CoS with less intense traffic. The GRB strategy presented the best value to network QoS when the CoS that required most QoS received the largest fraction of resources like in the differentiation $W3$. The $MORB$ strategy supplied the intermediate values of network QoS to all differentiation schemes. The SRB strategy presented good network QoS values when the service sequence was random (without differentiation) or prioritized the CoS with less intense traffic, because more classes has probability of service. The $MORGB$ and $MORWB$ strategies presented the worst values to network QoS showing that this association did not supply optima resource allocation.

The Figures 7, 8 and 9 present the CoS mean QoS obtained to $W1$, $W2$ and $W3$ differentiation schemes. The Figure 10 presents the dispersion of the CoS mean QoS to differentiation schemes ones and allow analyze the strategies fairness criteria. We observed that the $MORB$ is the fairest strategy considering the three differentiations. In the $W1$ differentiation scheme SRB strategy was the fairest because selected the first class to server in random form. In the $W3$ differentiation schema WRB strategy was the fairest because it allocated the resources attributing with the largest weight to the CoS that required more QoS. In the $W2$ differentiation scheme $MORB$ strategy was the fairest because it was capable of identifying the CoS that require more QoS and to allocate the optima resources, independently

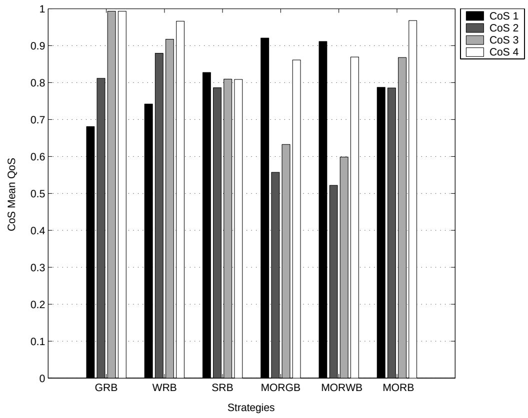

Figure 7: CoS Mean QoS: Differentiation Scheme $W1$.

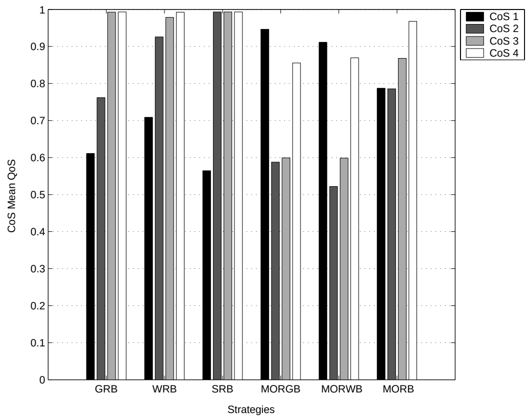

Figure 8: CoS Mean QoS: Differentiation Scheme $W2$.

of the importance attributed for it. The others strategies consider allocate the resources considering mainly the differentiation indexes. These factors did not express the traffic real characteristics what allowed unfair resource allocation.

The Figure 11 presents the output link utilization rate to the three differentiation schemes. We observed that MORB, MORWB and MORGB strategies did not supply the largest mean utilization rates, because they considering the QoS constraints. It means for these strategies that is better to store the packets respecting the delay requirements than send them with violation of jitter requirements. The Figure 12 reinforces this conclusion showing the buffer occupation rate to the three differentiation schemes. The The SRB and MORB strategies utilize completely the buffer to the three differentiation schemes. The MORB use continuity the buffer to respect the requirements of the QoS parameters of the CoS. The SRB strategy uses the buffer when the network packet arrival rate overflows the multiplexer service rate without to consider the QoS constraints. The others strategies use partially the buffer only respecting the constraints that are imposed directly over your share, in guaranteed or weighted forms.

5. CONCLUSIONS AND PERSPECTIVES

In this work, we studied the application of the stochastic multiobjective optimization technique to end-to-end QoS control by allocation of resources in pa-

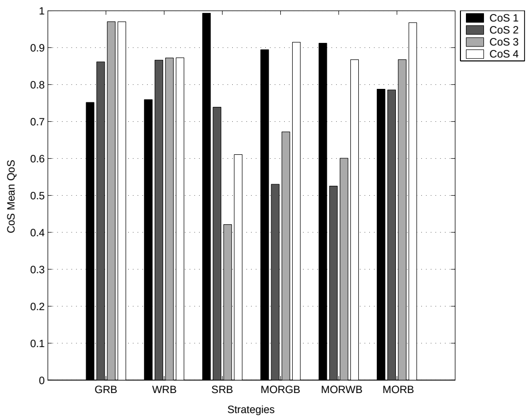

Figure 9: CoS Mean QoS: Differentiation Scheme $W3$.

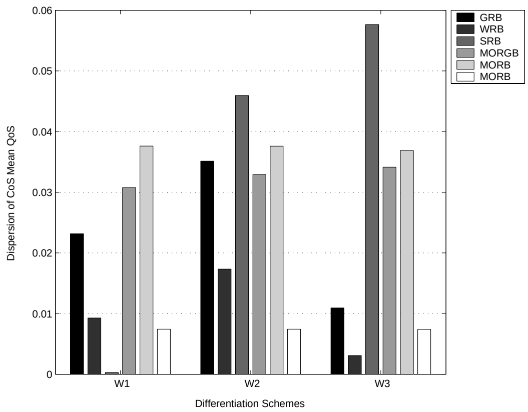

Figure 10: Dispersion of CoS Mean QoS: Differentiation Schemes $W1$,$W2$ and $W3$.

cket networks. We differentiated four CoS using weight factors that represent their relative importance in the network operator vision. We imposed constraints for limiting the CoS mean QoS values by the traffic contract specifications. We compared six strategies of allocation of resources where the obtained results showed that the proposed model is valid to describe this type of problem. We observed that the strategy based on multiobjective optimization presented a good trade-off between the obtained Network mean QoS value and the fairness among the CoS mean QoS values.

We intend to adapt this method to the Control of Admission to Connections (CAC) problem. Thus, it will be possible to profit from the fairness in the allocation of resources at the moment of deciding which connections to admit and which to reject.

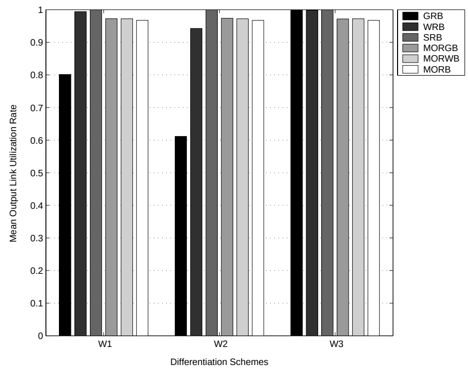

Figure 11: Mean Output Link Utilization Rate: Differentiation Schemes $W1$,$W2$ and $W3$.

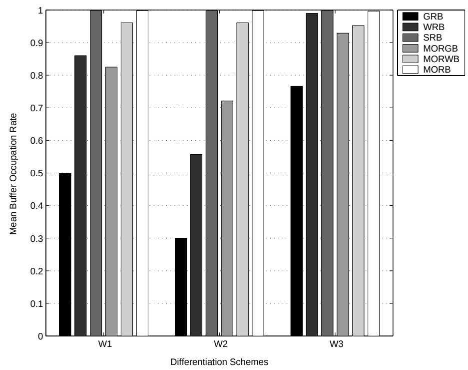

Figure 12: Mean Buffer Occupation Rate: Differentiation Scheme $W1$,$W2$ and $W3$.

References

[1] S. Shenker, C. Partridge, and R. Guerin. Specification of Guaranteed Quality of Service. *RFC 2212*, Sep 1997. IETF.

[2] V. Paxson, G. Almes, J. Mahdavi, and M. Mathis. Framework for IP Performance Metrics. *RFC 2330*, May 1998. IETF.

[3] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. Resource ReServation Protocol(RSVP) - Version 1 : An Overview. *RFC 2205*, Sep 1997. IETF.

[4] S. Black, D. Carlson, E. Davis, Z. Wang, and W. Weiss. An Architecture for Differentiated Service. *RFC 2475*, Dec 1998. IETF.

[5] Y.A. Korilis, A.A. Lazar, and A. Orda. *Achieving Network Optimal Using Stackelberg Routing Strategies. IEEE/ACM Transaction on Networking*, 5(1):161–172, Feb. 1997.

[6] T. Finland J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski. An Assured Forwarding PHB Group. *RFC 2597*, Jun 1999. IETF.

[7] V. Jacobson, K.Nichols, and K. Poduri. An Expedited Forwarding PHB. *RFC 2598*, Jun 1999. IETF.

[8] C. Lee, J. Lehoczky, R. Rajkumar, and D. Seiwieorek. On Quality of Service Optimization with Discrete QoS Options. *Fifth IEEE Real-Time Thechnology and Applications Symposium*, Jun 1999. Vancouver, Canada.

[9] R. Rajkumar, C. Lee, J. Lehoczky, and D. Siewiorek. A Resource Allocation Model for QoS Management. *Proceedings of the IEEE Real-Time Systems Symposium*, Dec. 1997.

[10] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, Singapore, second edition, 1991.

[11] G. Hebuterne and A. Depuis. Microscopic Models of ATM Multiplexing. *Sixth IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*, Jul 1998. Ilkley, West Yorkshire, U.K.

Appendix A

Symbols	Traffic parameters to the class i
M_i	Size of the largest packet;
s_i	Frequency of the largest packet;
α_i	Pareto probab. distribution parameter;
r_i	Equivalent packet arrival rate;
ω_i	Differentiation index;
W	Differentiation scheme.
	Maxima admitted values of the class i
D_{imax}	End-to-end delay;
J_{imax}	End-to-end jitter;
P_{imax}	Loss probability.
	Maxima accumulated values of the class i
D_{itot}	Delay (other components);
J_{itot}	Jitter (other components);
P_{itot}	Loss probability (other components).
	Allocated resources for the class i
R_i	Rate of service;
B_i	Buffer capacity.
	Multiplexer parameters
R	Bandwidth capacity of the multiplexer;
B	Buffer capacity of the multiplexer;
T	Period to the decision algorithm.

Table 13: Adopted Symbols.