

A study of QoS Performance for Real Time Applications over a Differentiated Services Network

E. Tsolakou, E. Nikolouzou, I. S. Venieris

National Technical University of Athens

9 Heroon Polytechniou, 15773 Athens, Greece

tel: +30 10 7722424, fax: +30 10 7722534

email: {evi, enik}@telecom.ntua.gr, ivenieri@cc.ece.ntua.gr

Abstract: The Differentiated Service Architecture (DiffServ) is currently a popular research topic as a low-cost method to provide Quality of Service to the different applications in the Internet. This paper addresses the definition and deployment of specific network services in a DiffServ environment. The proposed network services are described, while our main focus is the real-time delay-sensitive applications such as voice and video. The corresponding services for those applications are further analyzed and simulated. The analysis includes the influence of different algorithms that could implement the corresponding services for voice and video as well as the QoS performance under these algorithms. Simulation outcomes prove that the fundamental principles of the network services are fulfilled and provide an evaluation of the proposed algorithms.

Keywords: *Network Services, QoS, Traffic Classes, Scheduling Algorithms.*

I. INTRODUCTION

Motivated by the rapid change of QoS requirements of the new introduced network applications, the Internet has been evolving towards providing a wide variety of services, in order to meet the qualities of information delivery demanded by the applications. Since the best effort nature of the current Internet is not sufficient to cope with the requirements of this traffic, in terms of throughput, delay, jitter and packet loss,

two major efforts focusing on providing different levels of guarantee in quality of service have been introduced. The Integrated Services (IntServ) architecture [1] was the first significant step for the introduction of QoS in the Internet. IntServ uses the Resource Reservation Protocol (RSVP) [2] for the explicit setup of reservation state on each network node along the path from the sender to the receiver. One major dilemma faced by this approach is that in the core of the Internet, where exist several millions of flows, it may not be feasible to maintain and control the forwarding states efficiently. These scalability and management problems are addressed recently by the DiffServ [3] approach.

The focal point of the DiffServ model lies in the differentiation of flows at an edge router of a DS-domain and the aggregation of those flows of the same service class at a core router of the DS-domain. At each ingress interface of an edge router, packets are classified and marked into different classes, using Differentiated Services CodePoint (DSCP) [3]. Complex traffic conditioning mechanisms such as classification, marking, shaping, and policing are pushed to network edge routers. Therefore, the functionalities of the core routers are relatively simple - they classify packets and then forward them using corresponding Per-Hop Behaviors (PHBs). In this sense, PHB is a means by which a node allocates resources to behavior aggregates, and it is on top of this basic hop-by-hop resource allocation mechanism that useful differentiated services may be constructed. PHBs are implemented in nodes by means of some buffer management and packet scheduling mechanisms and the parameters associated with those mechanisms are closely related to those of traffic conditioning.

Under the DiffServ architecture two PHBs are proposed: the Expedited Forwarding (EF) [4] and Assured Forwarding (AF) [5]. Without per-flow states in the backbone nodes and with optional signaling protocol for admission control, EF PHB promises to deliver a "virtual lease line (VLL)"-like end-to-end service with a low loss, low latency, low jitter, assured bandwidth through DS domains. AF PHB group provides assurance of quality according to the relative ordering between classes, rather than absolute service level for each class. Different PHBs have different constraints and requirements that must be fulfilled, and hence, they often require supports of certain traffic conditioning functions. Exploiting those mechanisms described in the DiffServ architecture, and since there was not any standard implementation, the Aquila architecture designed a new set of Traffic Classes, based on the QoS characteristics available in the routers. The main focus was to study the whole QoS picture by

offering to the users a limited set of Network Services, which can accommodate traffic with different QoS requirements. Therefore, the proposed Network Services follow the concepts of the IETF EF and AF PHBs, but they additionally propose and exploit a specific implementation, which alleviates their deployment in real networks. In order to provide the different applications with the required QoS demands and protect the core network from bottlenecks, Admission Control functions should also apply to the edges of the network. Those functions will protect already admitted flows from performance degradation and will also protect the core network from congestion and starvation of resources.

This paper provides an overview of the proposed services, while an elaborative description will be given for services supporting real-time delay sensitive applications. In the conducted simulations the QoS performance of the services appropriate for such applications, e.g. voice and video, is thorough examined and the different algorithms for implementing them in a real network are evaluated.

The paper is structured as follows. In Section II, the Network Services are presented, in Section III the concept of Traffic Classes is introduced and described in detail as well as the corresponding Admission Control Algorithms. Section IV presents the different algorithms that could implement the Services suitable for voice and video applications accompanied by the corresponding simulations results. Finally, we conclude in Section V.

II. NETWORK SERVICES

In order to provide QoS guarantees in a DiffServ network it is essential to assure QoS differentiation. Therefore, a set of five Network Services (NS) has been specified and implemented in our framework [6], which comprises the services sold by the provider to the potential customers, either end-users or other providers. They describe the QoS treatment a user's traffic experience within a network. The specified NSs are: Premium Constant Bit Rate (PCBR), Premium Variable Bit Rate (PVBR), Premium Multimedia (PMM), Premium Mission Critical (PMC) and Standard Best Effort (STD BE). Applications can be grouped into this relatively small number of services, with the

applications in each service having similar requirements on the network in order to perform effectively and flows in each service having similar characteristics.

The PCBR network service is intended to support applications that require VLL-like services. Therefore, it is appropriate for voice flows, voice trunks or interactive multimedia applications. That kind of flows is usually characterized by an almost constant bit rate (CBR) and low bandwidth requirements, while a great number of them are unresponsive (UDP). In addition, they should have small packets (<256Bytes), so as not to provoke long transmission delays. It requires and expects to receive low delay, very low jitter and very low packet loss. The targeted quantitative value for end-to-end delay is less than 150msec for 99.99% of the packets, while packet loss is expected to be no more than 10^{-6} .

The PVBR network service mainly copes with unresponsive variable bit rate (VBR) sources with medium to high bandwidth requirements. The intention is to separate those possibly high bandwidth VBR flows from the low bandwidth VBR and CBR flows in PCBR. This is caused by the fact that peak rate allocation is inefficient for the high bandwidth VBR flows, contrary to the flows belonging to PCBR. Typical candidate applications are real time video and teleconferencing. The requirements are similar to the PCBR network services but with a less strict need concerning the jitter and packet loss. They are characterized by large packet size, which oscillates from 256-1024 bytes. The targeted end-to-end delay is limited to less than 250msec for 99.99% of the packets, while packet loss should be less than 10^{-4} .

The PMM is expected to carry a mixture of TCP and non-TCP traffic. These flows require a minimum bandwidth, which must be delivered at a high probability. Independently of the transport protocol, flows are expected to implement some kind of congestion control mechanism and their aggressiveness should be similar to the one of TCP, assuming that they are roughly TCP-friendly [7]. This NS is supposed to serve adaptive applications (TCP), like low-quality video, streaming multimedia applications or file transfer (FTP). By nature, these flows are usually responsive, greedy and reflected to long-lived connections. They require throughput guarantees, which are translated into low packet loss for in-profile packets ($\leq 10^{-3}$), while there are no QoS guarantees for out-of-profile packets.

PMC is targeting to non-greedy adaptive applications that have great sensitivity concerning packet loss. It is thus suitable for transaction-oriented applications and interactive applications such as online games and chat-like applications. The main characteristics are the non-greediness of the flow, the responsive nature (TCP), the low use of bandwidth and the short life of the connection. As mentioned above, the most important requirement is very low packet loss for in-profile packets ($\leq 10^{-6}$), while no QoS guarantees are expected for out-of-profile packets. Nevertheless, low queuing delay is also desired, in order to retain the meaning of interactivity. Finally, packets of the STD BE receive no special treatment in the network.

III. TRAFFIC HANDLING

The description of the proposed Network Services was given in the previous section. In the rest of the paper the focal point is the exploration of the PCBR and PVBR services, which transport voice and video traffic over a DiffServ network. Therefore, our focus is the real-time delay-sensitive applications, which will be further examined and simulated in terms of QoS performance.

A. Traffic classes

The implementation of the Network Services is realized with the use of some network's mechanisms, which are the Traffic Classes (TCLs). A TCL is defined as a composition of a set of admission control rules, a set of traffic conditioning rules (Fig. 1) and a per-hop behavior (PHB). In the proposed architecture five TCLs are introduced: TCL1, TCL2, TCL3, TCL4 and TCL5 which correspond to PCBR, PVBR, PMM, PMC and STD BE. In general, each TCL maintains a separate queue at the router output ports and allocates one or more DSCPs in order to enable differentiation of packets in the core network. A PHB implemented in the output port of a router is realized in the network with the use of scheduling and buffer management algorithms.

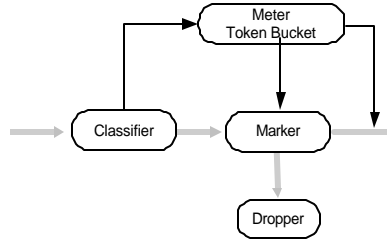


Fig. 1. Traffic Conditioning Mechanisms

In this paper only PCBR-TCL1 and PVBR-TCL2 will be closely examined since they serve real-time applications with strict QoS guarantees. The implementation of the traffic conditioner regarding TCL1 and TCL2 is described below and the applicable scheduling algorithms are analyzed in the next section.

The traffic profile for TCL1 is described in terms of a Single Token Bucket, which polices the peak rate of the flows. Admission control functions are also based on the peak rate of allocations for TCL1, since those flows are usually of low bandwidth. The single TB operates as both meter and dropper. Since TCL1 is characterized with strict QoS requirements, packet exceeding the declared profile should be dropped. The single TB is configured with token rate r equal to the Peak Rate (PR) of the flow, and bucket size b equal to a multiple x of the maximum allowed packet size (M), which is called Bucket size for PR, (BSP). The value of x lies in the range of $\{1,5\}$; a possible value could be $x=1$, while a larger value would allow a small amount of burstiness. The traffic conditioning mechanism is realized in the routers with the use of the Committed Access Rate (CAR) mechanism.

Peak rate allocation is not appropriate for TCL2, since it is characterized with high bandwidth flows. Therefore, admission control function is based on both the peak and sustainable rate of the flows and a dual TB as meter and dropper is proposed. The first TB is configured with r equal to the Sustained Rate (SR) of the flow, and b equal to the Bucket Size for SR in bytes (BSS). The second TB is configured with r equal to the PR of the flow and b equal to a multiple x of the maximum allowed packet size (M), (BSP). The value of x is in the range of $\{1,5\}$ and a value of $x=2$ is recommended. The depth of the first bucket defines the burstiness allowed for the sender's flow (BSS), and its value ranges from 10 to 20 times the M . A packet is marked as in-profile if there are enough tokens in the first and second TB to accommodate it, otherwise it is dropped. The intention is to limit the sender's

traffic in order to be conformant to the profile of the first TB (SR , BSS), while the second TB (PR) allows an amount of burstiness.

B. Admission Control Functions

Admission Control (AC) plays a significant role in ensuring the requested quality of service to user traffic. It is mainly responsible for limiting the access to the network, so that the already admitted flows do not anticipate any deterioration in their quality contract. Therefore, a bottleneck is prohibited to arise in the edge-link (i.e. the link between a core network and the ingress or egress router) as well as in any of the internal-links. In this way, the network is protected from congestion and an overall network stability is provided. Setting the AC limits is based on the target network utilization as well as the target performance of each Network Service. Although it is not the aim of this paper to analyze the AC procedures, the AC functions used for the PCBR and PVBR services is briefly discussed.

According to the network service, a specific formula is used in each case. The proposed AC algorithms are derived from the results developed in the context of ATM traffic control and described in detail in [6]. The assignment of AC Limits (ACL) to each edge router for each TCL represents a resource assignment to the relevant traffic aggregates. The request for network resources is accepted or rejected based on the admission decision made only at the network ingress and, in some cases, at the egress point. This makes the AC decision more critical, as link-by-link verification of resource availability is not possible. To perform the admission control at the ingress or egress, the single link model was considered with capacity C and buffer size B . Furthermore, for simplicity, the isolation between all traffic classes was assumed. Whenever below the parameters C or B are mentioned, they correspond to the capacity and buffer size dedicated to serve the given traffic class.

TCL1 is described in terms of a single TB, which polices the peak rate. Since for TCL1 negligible packet delay variation is assumed [8], the worst case traffic pattern for the superposition of a number of TCL1 flows takes the form of poissonian stream (with the mean rate equal to the sum of the PR parameters of the particular flows). If it is assumed that C_{TCL1} capacity is dedicated for TCL1, and N_l flows with $\{PR_1, PR_2, \dots, PR_{N_l}\}$ are currently admitted and active, then a new flow with PR_{new} as its peak rate is admitted if the following condition is satisfied:

$$PR_{new} + \sum_{i=1}^{N_1} PR_i \leq rC_{TCL1} \quad (1)$$

Parameter ρ ($\rho < 1$) specifies the admissible load of capacity allocated to the TCL1. The value of ρ is calculated from the analysis of M/D/1/B system taking into account the assumed target packet loss ratio and buffer size [9].

In case of TCL2 traffic class the Rate Envelope Multiplexing (REM) multiplexing scheme is assumed for guaranteeing low packet delay [10]. Therefore, the only QoS parameter that requires concern is the packet loss rate. In the REM multiplexing, the buffer (relatively small) has to be dimensioned for absorbing, the so-called packet scale congestion (simultaneous arrival of packets from different sources). For this purpose the N*D/D/1 queuing system analysis is useful. In the TCL2 class, each flow is characterized by the parameters of the dual token bucket and the proposed admission method for TCL2 is based on the notion of effective bandwidth. There are many methods for calculating effective bandwidth [10]. For simplicity reasons, the methods proposed in [11] are chosen. In this method the value of effective bandwidth, $Eff(.)$, is calculated on the basis of PR, SR and BSS parameters, taking into account the target packet loss rate. Let us assume that the capacity dedicated for TCL2 class is C_{TCL2} . In the case, when N_2 flows with $\{Eff_1, Eff_2, \dots, Eff_{N_2}\}$ are currently in progress, a new flow with Eff_{new} is admitted if the following condition is satisfied:

$$Eff_{new} + \sum_{i=1}^{N_2} Eff_i \leq C_{TCL2} \quad (2)$$

IV. IMPLEMENTATION OF TCL1 AND TCL2

A. Scheduling Issues

The selection of the appropriate scheduling algorithm regarding the implementation of TCLs compromises a major problem. Therefore, different scheduling algorithms have been implemented and studied:

- **Priority Queuing (PQ):** In PQ [12], a number of distinct queues is created and a level of priority is assigned to each one. Packets are scheduled from a particular priority queue in First Come-First Served order only when all queues of higher priority are empty.

- **Weighted Fair Queuing (WFQ):** In WFQ [13, 14, 15], the allocation of link bandwidth is considered fair, since the bandwidth is allocated in proportion to the weights associated with each queue and free bandwidth is fairly shared between the queues, based also on their weights.
- **PQ-WFQ:** It is a combination of the PQ and WFQ, where TCL1 is transmitted with high priority, while the other TCLs are transmitted through the WFQ scheduler.

Based on [12], it is concluded that serving each TCL with PQ, meaning giving to TCL1 the highest priority and to TCL5 the lowest one, does not favour the TCL5 traffic, which is possible to experience starvation, while even TCL3 and TCL4 could be blocked. From the other hand, adapting the WFQ algorithm would guarantee to each TCL a minimum amount of resources, starvation problems may be overridden and in advance isolation between the flows would be achieved. Nevertheless, WFQ would not be able to guarantee the strict delay requirement for TCL1 and TCL2. Finally, PQ-WFQ overcomes the limitations introduced by the PQ and WFQ, offering a more integrated approach. Under this approach, TCL1 is served with the highest priority and therefore strict delay guarantees are provided, while a minimum amount of resources may be reserved for TCL5, resulting in prohibiting starvation. Moreover, WFQ has been proposed within the IETF as the reference server model for guaranteeing service differentiation in the Internet [14].

Having in mind the QoS requirements of TCL1 and TCL2, appear to have a lot in common but from a closer look, they also have big differences. TCL1 requires very strict QoS performance while in addition requires a very small packet size. These differences are actually taken into consideration when trying to provide different approaches for the scheduling of these two TCLs. Additionally, it is studied whether or not separate queues should be configured for the transmission of these TCLs.

B. Simulation Model

In this section, the topology chosen for performing the different simulation scenarios is described. The OPNET Simulation Tool was used for realizing the network topology, which is depicted in Fig. 2.

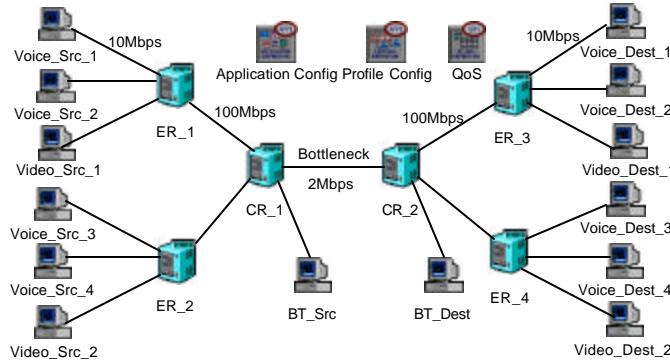


Fig. 2: Simulation Network Topology

This topology consists of seven traffic generators with their destinations respectively; four of them set up voice flows (Voice_Src), two of them set up video flows (Video_Src) and there is one for the background traffic (BT_Src). All traffic generators and their destinations are Ethernet Workstations. There are four edge routers (ER) and two core routers (CR). Regarding the router, it was used the CISCO 3620 model. The access links between the source/destination and the edge routers are 10Mbps Ethernet links. The links between the edge and core routers have a data rate of 100Mbps and a 2Mbps link between the core routers is considered. The Routing Information Protocol (RIP) is considered as the routing protocol for the network.

Assuming the 2Mbps access links, the recommended AC limits for TCL1 and TCL2 as well as the maximum permitted traffic are configured as following: for TCL1, ACL1=19% (380kbps), for TCL2, ACL2=25% (500kbps) and the remaining bandwidth is used by the Background Traffic (BT) (1.12Mbps). Concerning BT, it could be compromised of traffic belonging to TCL3, TCL4 and TCL5. Usually, networks are configured in such a way that premium services would not occupy more than 50 percent of the total link capacity, so as to ensure that QoS requirements are met and to avoid bottlenecks in the core network. Therefore a basic rule for determining the ACL1, is that TCL1 traffic, which is consisted of voice traffic, should not occupy more than 20 percent of the link capacity in order its strict QoS parameters to be assured. Ancillary, even video traffic should be limited to 25 percent of the link capacity. In this paper, the highest values of AC Limits are considered, in order to inject to the network a relative large amount of traffic of TCL1 and TCL2 and to produce a worst-case scenario.

The performance of TCL1 was validated assuming target packet loss ratio (P_{Loss}) to be equal to 10^{-2} . According to the specified admission control algorithm the maximum admissible load in this case (according to the M/D/1 system analysis [7]) is $\rho=0.68$, which is equivalent to 380 kbps. There were used 4 CBR flows of 65kbps data rate each with packets having a size of 80Bytes. Therefore, the single TB for TCL1 (CAR) was configured with PR=65kbps and BSP=160Bytes ($2*M$).

Assuming that the ACL2 is 500kbps and the target packet loss is equal to $7*10^{-1}$, the effective bandwidth for each admitted flow, according to [11], is 250kbps, when each flow is characterized by PR=500kbps, SR=400kbps and packet size 500Bytes. Based on the AC algorithm the number of admitted flows is 2, with $\sum_{i=1}^2 Eff_i = 500kbps$ for a 2Mbps access link. The dual TB (CAR) was consequently configured for each flow, with PR = 250kbps, BSP = 1000B ($2*M$), SR = 200kbps and BSS = 5000B ($10*M$).

The data rate of Background Traffic is constant with of 1.6Mbps and packets of size 1000Bytes. A CBR model was selected for the BT traffic, as the worst case traffic model, which continuously utilizes the bandwidth allocated for it. Additionally, it is assumed that BT tries to send more than the bandwidth allocated for it, and in fact 40 percent more, in order to provoke congestion in the core link. The UDP is actually considered as the transport protocol for both TCLs and BT.

C. Study of TCL1

Regarding the TCL1-PCBR, our purpose is to investigate the QoS parameters of the Premium CBR service that could guarantee both low packet delay and packet loss ratio, using different schedulers into the output interface of the routers as shown in Fig. 3 and Fig. 4. In Fig. 3, the WFQ scheduler is used for the transmission of TCL1, TCL2 and BT, while under the PQ-WFQ scheme, TCL1 is transmitted with the highest priority and furthermore two queues are configured in the WFQ for TCL2 and BT traffic. The FIFO drop tail algorithm was used as the buffer management algorithm for all queues, and this algorithm is being used for all next scenarios.

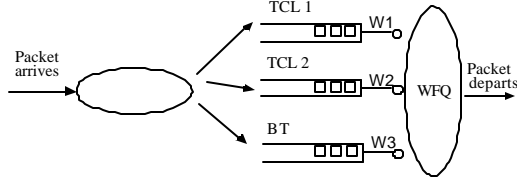


Fig. 3: WFQ Scheme

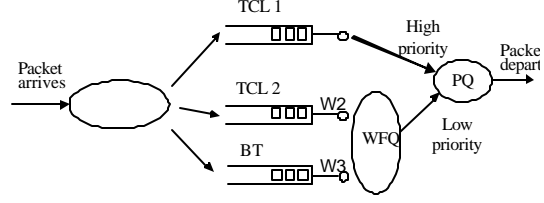


Fig. 4: PQ-WFQ Scheme

The buffer capacity in the routers was set to 69 packets, while for TCL1 was set to 5 packets in order to guarantee low packet delay requirements. The end-to-end delay, delay variation and packet loss was measured in the network for TCL1, using different schedulers (WFQ and PQ-WFQ), with their characteristics, as depicted in Table 1. The weights of the WFQ were configured based on the corresponding AC Limits, regarding TCL1 and TCL2.

Traffic Class		Weight		Maximum Queue Size (packets)
		WFQ	PQ-WFQ	
TCL1	w 1	0.19	-	5
TCL2	w 2	0.25	0.25	5
BT	w 3	0.56	0.75	59

Table 1: WFQ and PQ-WFQ Configuration

The minimum, maximum and average end-to-end delay of the voice flows for both schedulers are shown in Fig. 5, while the maximum and average delay variation are depicted in Table 2.

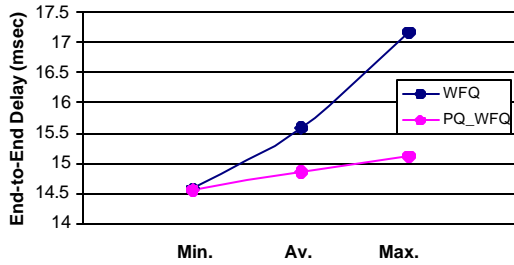


Fig. 5: End-to-End Delay of TCL1

Scheduler	Av. Delay Variation (μ sec)	Max. Delay Variation (μ sec)
WFQ	6	9
PQ-WFQ	4.6	4.7

Table 2: End-to-End Delay Variation of TCL1

The maximum end-to-end delay using PQ-WFQ is about 2 msec (13%) less than the WFQ one. Also, there is a main difference in the delay variation between the two schedulers; it is two times less using PQ-WFQ than using WFQ. Therefore, TCL1 should become somehow 'independent', having a higher priority from other TCLs, in order to achieve its strong QoS requirements. It is also worth mentioning that the packet loss for TCL1 was measured zero for both cases, which is much smaller than the target one (10^{-2}), deducing that the used AC functions are really very conservative.

Furthermore, the packet size of TCL1 flows should additionally be small, since it affects its QoS parameters and in particular the end-to-end delay. This impact is depicted in Fig. 6, where it is shown the average end-to-end delay of voice packets having different packet sizes (152B, 245B, 500B, 1000B), using the PQ-WFQ scheduler. Concluding, the value of the delay variation is getting bigger when the packet size is increased and therefore flows of TCL1 should be characterised by small packet sizes.

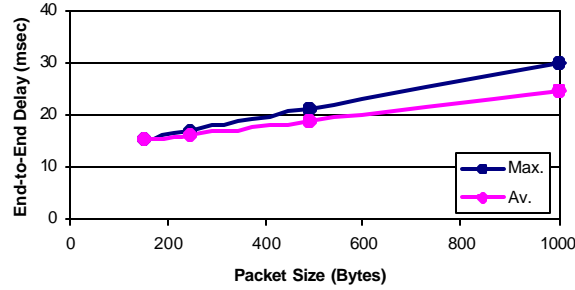


Fig. 6: End-to-End Delay of TCL1 for different Packet Sizes of TCL1

D. Study of TCL2

As far as TCL2-PVBR is concerned, the QoS parameters should guarantee both low packet delay and packet loss ratio, as well. Based on the above conclusion, we use the PQ-WFQ scheduler and examine the impact in the QoS parameters of TCL2 setting different values of w_2 (0.25, 0.5 and 0.99) using the same traffic model for TCL2. The average end-to-end delay is depicted in Fig. 7.

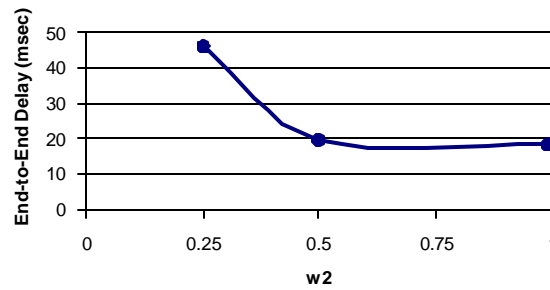


Fig. 7: Average End-to-End Delay of TCL2 for different values of w_2

The TCL2 sends traffic with a rate of 250kbps, which comprises a percentage of 25% of the Bottleneck link (2Mbps). Increasing the w_2 weight in the PQ-WFQ scheduler, the bandwidth allocated for TCL2 is also being increased. So, the value of the end-to-end delay is being 50% decreased when the weight is being doubled, while a further increase equal to 99% results in a slightly decreased value.

Concerning the value of packet loss ratio, it is being decreased 10,000 times when the weight is 99%, as depicted in Table 3. Under all the three different configurations of w_2 , a lower packet loss ratio is achieved than the target one (7×10^{-1}), determining that the used AC functions for TCL2 are also conservative.

w2	Packet Loss Ratio of TCL2
0.25	1e-01
0.5	1e-03
0.99	0.9e-04

Table 3: Packet Loss Ratio of TCL2 for different values of the weight w_2

Based on the above simulations, a conclusion concerning TCL2 is that increasing the weight dedicated for TCL2 the QoS parameters are upgraded. One reason for this is that allocating double or quadruple bandwidth than transmitting for TCL2, a fatter link is therefore used for video flows. That fat link could transmit individual packets much faster, hence introduce shorter delay, even though the link utilization is similar in all cases.

In addition, if the objective is the optimization of QoS characteristics of TCL2, a relative high weight should be chosen, especially in the case of a low bandwidth link, as the one depicted in Fig. 2. Furthermore, assigning to TCL2 a high weight of value equal to 0.99, results in a kind of prioritization of TCL2. According to this indirect way, TCL2 is regarded as a TCL with the second highest priority after TCL1.

Moreover, a very low amount of bandwidth is allocated to BT in order to avoid blocking that kind of traffic. It is also worth mentioning, that any unused bandwidth allocated to TCL2 is used by the BT (TCL3, TCL4 and TCL5), in case the last needs additional bandwidth than the one allocated to it.

E. Study of TCL1&TCL2 sharing the same queue

However, the above results create an impression that if TCL2 shares the same (high priority) queue with TCL1, then better values of the QoS parameters could be achieved. Therefore, in this scenario the QoS parameters of TCL1 and TCL2 are investigated when they share the same queue. The BT (all other traffic classes) is still in the low priority (WFQ) queue using the rest bandwidth. The maximum queue size for TCL1/ TCL2 is set to 10 packets and for BT to 59 packets.

Firstly, the Average End-to-End delay and the Average Delay Variation of TCL1 setting different packets sizes of TCL2 are measured, when they are in separate queues ('sep') and when they share the same queue ("same"), as depicted in Fig. 8 and Fig. 9.

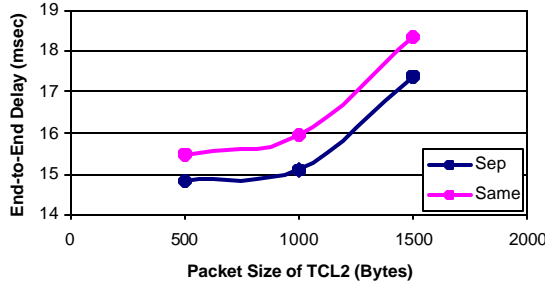


Fig. 8: Average End-to-End Delay of TCL1 under different packets of TCL2

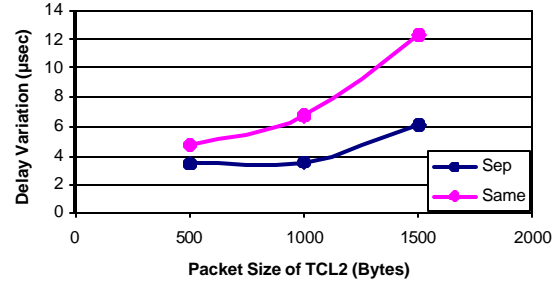


Fig. 9: Average Delay Variation of TCL1 under different packets of TCL2

It can be remarked that the end-to-end delay of TCL1 is increased when it shares the same queue with TCL2, and particularly when the packet size of TCL2 flows is getting larger. Regarding the delay variation of TCL1, it is almost the same when two queues are used, while it is increased dramatically when TCL1 shares the same queue with TCL2. The packet loss ratio of TCL1 is depicted in Table 4. Sharing the same queue with TCL2, actually concludes in a relatively high packet loss for TCL1, which is increased proportionally to the packet size of the TCL2 flows. Therefore, the performance of TCL1 is degraded and moreover influenced by the packet size of TCL2 flows when one queue is used. This can be regarded as a verification that in the "sep" case the two services are in a degree independent, not influencing each other.

Packet Size of TCL2 (Bytes)	Packet Loss Ratio of TCL1	
	Separate Queues	Same Queue
500	0	0
1000	0	0.4e-04
1500	0	1.7e-04

Table 4: Packet Loss Ratio of TCL1 for different packet sizes of TCL2

In addition, the performance of TCL2 under different packet sizes is studied for both cases ("sep" and "same"), as depicted in Fig. 10 and Fig. 11.

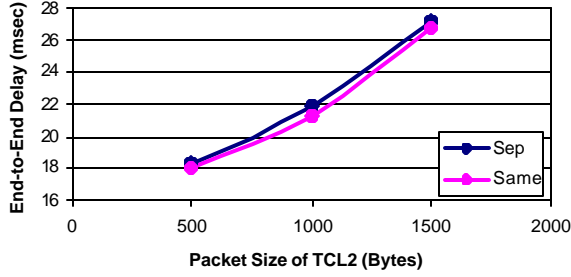


Fig. 10: Average End-to-End Delay of TCL2

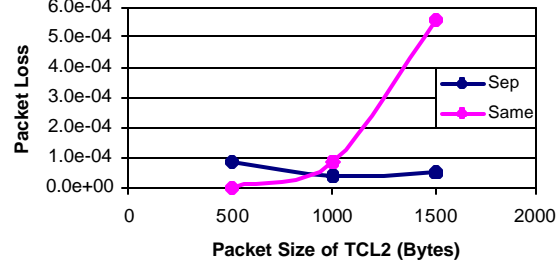


Fig. 11: Packet Loss Ratio of TCL2

The end-to-end delay of TCL2 is a little lower in the “same” case than in the “sep” case. Moreover, the value of packet loss ratio is much greater when same queue is used and is getting bigger when the packet size of TCL2 flows is increased. In the “sep queue” case, the packet loss for TCL2 remains almost the same, and it is not being influenced by the different packet sizes.

Putting TCL1 and TCL2 in the same queue results in achieving almost the same QoS parameters for both TCLs, as concluding from the above results. Therefore, service differentiation, as the targeted one is not accomplished, since both TCLs share almost the same characteristics. The main conclusion from the above simulations is that the impact of TCL2 in QoS Parameters of TCL1 is not meaningless. Furthermore, the value of the end-to-end delay is getting greater when the packet size of TCL2 is increased. Additionally, the end-to-end delay parameters of TCL2 are not upgrading effectively and the packet loss ratio has worse values than in the other case. So, TCL1, as well TCL2, should have their own queues.

V. CONCLUSIONS

In this paper we deal with the definition and deployment of a set of Network Services, within a DiffServ environment. The Network Services, which are implemented in the network through the traffic handling mechanisms offered by the respective Traffic Classes, target at different kinds of user traffic that exhibit similar QoS requirements and characteristics, and therefore they demand analogous treatment within the network. We described a specific implementation for voice and video applications, which corresponds to the PCBR and PVBR Network Services respectively, including admission control algorithms, traffic conditioning mechanisms and scheduling algorithms. In particular, a different set of mechanism is used for TCL1 and TCL2, based on flows characteristics and the corresponding QoS requirements. The corresponding services for those real time applications

are further analyzed and simulated. Subsequently, simulation results proved that the proposed traffic handling mechanisms as well as the introduced PQ-WFQ scheduling algorithm are adequate for the proposed PCBR and PVBR network services. Therefore, the correctness of our design was verified, since the target QoS performance was achieved for both TCL1 and TCL2. Future work would focus on the study of implementation of the other traffic classes (TCL3 and TCL4), accompanied by simulations, which would examine their QoS parameters.

REFERENCES

- [1] R. Braden, D. Clark and S. Shenker, "Integrated Services in the Internet Architecture: an Overview", RFC 1633
- [2] R. Braden, L. Zhang, S. Berson, S. Herzog and S. Jamin, "Resource ReSerVation Protocol (RSVP)", RFC 2205
- [3] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang and W. Weiss, "An Architecture for Differentiated Services", RFC 2475
- [4] V. Jacobson, K. Nichols and K. Poduri, "An Expedited Forwarding PHB", RFC 2598
- [5] J. Heinanen, F. Baker, W. Weiss and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597
- [6] Deliverable D1301, Specification of traffic handling for the first trial, AQUILA project consortium, <http://www.ist-aquila.org/>, September 2000
- [7] M. Allman, V. Paxson, and W. Stevens, "TCP Congestion Control", RFC 2581

- [8] Brichet et al., Stochastic ordering and the notion of negligible CDV, Proc. of 15th International Teletraffic Congress, Washington D.C., USA, 1996
- [9] Final Report COST 257, Impact of New Services on the Architecture and Performance of Broadband Networks (COST 257 Management Committee, Phuoc Tran-Gia, Norbert Vicari eds.), ISBN-Nr. 3-930111-10-1, compuTEAM, Wuerzburg, Germany 2000
- [10] Final report COST 242, Broadband network teletraffic: Performance evaluation and design of broadband multiservice networks (J. Roberts, U. Mocci, J. Virtamo eds.), Lectures Notes in Computer Science 1155, Springer 1996
- [11] K. Lindberger, Dimensioning and design methods for integrated ATM networks, Proc. of 14 the International Teletraffic Congress, Antibes, 1994
- [12] R.Guerin, V.Peris, "Quality-of-Service in packet networks: basic mechanisms and directions", Computers Networks, February 99, Vol 31, No.3, pg.169-189
- [13] A.Demers, S.Keshav, S.Shenker, "Analysis and simulation of a fair queuing algorithm", Journal of Interworking Research and Experience, p.3-26, October 1990
- [14] S.Shenker, R.Braden, D.Clark, "Integrated services in the internet architecture: an overview", RFC 1633, June 1994
- [15] J. Bennett, H. Zhang, "Hierarchical packet fair queuing algorithms," Proc of *SIGCOMM 1996*, August 1996, pg.143-156.