QUALITY OF SERVICE ADMISSION CONTROL FOR MULTIMEDIA END-SYSTEMS

A.G. MALAMOS
Dept. of Electronics and Computer Eng.,
Technical University of Crete,
University Campus,
GR-731 00 Chania,
GREECE

T.A. VARVARIGOU
Dept. of Electrical and Computer Eng.,
National Technical University of Athens,
Telecommunications Lab.,
Eroon Polytechniou 9,
GR-157 73 Zographou, Athens,
GREECE

E.N. MALAMAS
Dept. of Electronics and Computer Eng.,
Technical University of Crete,
University Campus,
GR-731 00 Chania,
GREECE

Abstract:- In recent years, several new architectures have been developed for supporting execution of multimedia applications in end-systems (servers, workstations and set top boxes). However Quality of Service admission control is an essential task that is still missing from these architectures. In this paper, we are first introducing a model that expresses, timing and format, scheduling parameters of distributed multimedia applications with respect to Quality of Service. We use this model to formulate an admission control scheme appropriate for handling effectively multiple concurrently received requests for multimedia services. Furthermore, we model the most common video and audio distributed multimedia applications and we are presenting the results of the simulations we performed to evaluate the effectiveness of the admission control scheme we propose.

IMACS/IEEE CSCC'99 Proceedings, Pages:1871-1879

Key-words:- Network services, multimedia, Quality of Service (QoS) , end-systems, admission control

1 Introduction

Distributed multimedia applications, and in particular continuous media presentation applications like video and audio services, have placed new requirements to the design of networks and related systems. These applications need not only guarantee correct delivery in acceptable time but also delivery in acceptable quality. This implies that when a user requests a continuous media service, he/she also has Quality of Service (QoS) expectations. So the user must be provided with a service which will constantly be within his expectations of quality. Therefore multimedia systems must provide not only Quality of Service but also Quality of Service guarantees. QoS that user experience is the quality that the entire infrastructure is able to provide. Thus QoS requirements for multimedia applications are typically end-to-end requirements. This imposes quality demands on both the involving parts of the service, i.e., the network as well as the end-systems (servers, workstations, set-top-boxes).

In most of the existing QoS management architectures, the level of quality that a user experiences is the one to whitch he/she and the provider agreed. Agreement is made, after negotiation [1][2], where the provider informs the user about the levels of QoS he can guarantee and user chooses the one that satisfies him/her. To extract the level of quality that can be guarantied, the provider has to consider network and endpoints' ability to reserve the necessary resources.

A lot of work has been done on the network management [1]-[3], where QoS reservation models and architectures have been introduced. On the other hand many scheduling frameworks have been implemented into operating systems (OS), to support continuous media applications in workstations and servers [4]-[11]. These scheduling frameworks are not directly related to the quality requirements of the scheduled media. They emphasize the protection of synchronization demands among parts of multimedia application (e.g. video, audio). Also, they place an emphasis on keeping information losses during a session within acceptable levels. This action is useful, mostly, when OS is not able to prevent end-system from over-loading. However, in network service environments that support QoS, end-systems have to be consistent with the QoS agreements and to provide guarantees. Thus, end-systems should support mapping of QoS characteristics to scheduling parameters, in order to participate in the QoS negotiations by introducing their own quality limitations in a way fully compatible to the rest communication infrastructure.

The existing scheduling frameworks comprise admission control schemes that require a-priori knowledge of certain characteristics of the continuous media tasks (threads), such as periodicity of the information

Table 1. Specifications and bandwidth requirements for uncompressed digital media.

Type	Best Quality Specifications	Best Quality Bandwidth requirements (Mbits/sec)
High Definition TV (HDTV) quality video	2000X1024 pixels/Frame 60 frames/sec 24 bits/pixel	2812.5
NTSC quality video	640X480 pixels/Frame 33 Frames/sec 8 bits/pixel	77.5
CD quality Audio	stereo (2 channels) 44100 samples/sec per channel 16 bit/sample	1.35
voice quality Audio	mono (1 channel) 8000 samples/sec 8 bit/sample	0.061

(e.g. frames per second for a video application) and computation time per period (e.g. CPU time per frame). However, this information is strongly related to provider-user QoS agreement. Thus, the current OS schemes are unable to contribute to provider-user negotiations, since the admission control schemes they employ can only check rather than decide about the level of quality that can be supported. Furthermore, most of the existing OS admission control algorithms are able to handle (by checking resource capacities) one request for service at a time. However, end-system (particularly server) admission control should be able to serve efficiently multiple requests received concurrently.

Servers and workstations are connected to the network infrastructure through network adapters (transducers), special purpose boards (ethernet boards, modems) and lines (coaxial, twist-pear, fiber optics lines). These devices have bounded bandwidth capability, which limits the ability of the end-system to transmit or receive information.

In this paper, we provide an applicable definition of QoS and according to this, we formulate a model to relate scheduling parameters, such as computation time, period, deadline, etc., to QoS of multimedia applications.

We use this model to introduce an admission control scheme, compatible to existing scheduling frameworks of operating systems, that provides the OS with the ability to decide on the QoS of incoming requests that can be admitted. This admission control scheme assigns, in an optimal manner, QoS to incoming multiple simultaneous arrived requests, considering end-system and application constraints.

We formulate the admission control as constraint optimization, with the objective to maximize the overall QoS provided by the end system. The quality that is computed through the optimization process, is the QoS that the end-system can guarantee. Hence, end-system becomes able to contribute to the provider-user QoS negotiations with its own constraints to the quality that the whole communication infrastructure can provide.

2 Background

2.1 Multimedia Application and CPU Schedulling

2.1.1 Classes of Multimedia information

When we refer to multimedia information, we mean types of data and processes (video, audio, text, control signals) that a multimedia application may use in order to provide a service. A scheduler that supports effectively multimedia applications has to manage efficiently all types of information, providing QoS guarantees with respect to the user, application and information demands and preferences. These types of information have in general different timing and processing requirements from the CPU.

In terms of timing requirements, multimedia information may be distinguished into three major classes: a) continuous media (CM) b) aperiodic information with real time constraints and c) aperiodic information without real time constraints.

CM class includes all types of data (video, audio, animation data etc.) that arrive, in a processing system with a continuous and periodic manner. Frames of CM are related to each other and there are precedence constraints in their processing. This category of multimedia information is strongly related to the quality of the service provided and consequently to user satisfaction.

The class of aperiodic information with real time constraints, includes flow-control or presentation data (photos, graphics, text, etc.) that have processing deadline. The way that their successful processing affects the satisfaction level of a user, depends on their type and the application they participate in.

Table 2. Implementation of QoS model to HDTV media

quality (q)	Vertical resolution (pixels)	Horizontal resolution (pixels)	Frame Rate (frames/sec)	Computation Rate	BW (Mbits/sec)
0	0	0	0	0	0
0.1	725	1416	12	1.23E-01	281.25
0.2	804	1571	19	2.46E-01	562.50
0.3	855	1670	26	3.69E-01	843.75
0.4	893	1743	32	4.92E-01	1125.00
0.5	923	1803	37	6.14E-01	1406.25
0.6	948	1852	42	7.37E-01	1687.50
0.7	971	1896	47	8.60E-01	1968.75
0.8	990	1934	51	9.83E-01	2250.00
0.9	1008	1969	56	1.11E+00*	2531.25
1	1024	2000	60	1.23E+00*	2812.50

The class of aperiodic information without real time constraints, includes all types of processes and applications (e-mail, off-line data transfer etc.), which do not demand any real time or synchronization guarantee.

2.2 QoS characteristics

QoS characteristics are performance characteristics that specify user, application and infrastructure (end-system, network) technical requirements so that the whole system can provide a level of quality. QoS characteristics are different at each individual layer of the distributed multimedia application. For example, the characteristics that specify the performance in network layer (delay, bandwidth, delay-jitter, etc.) are different than those of application layer (frame rate, media format, synchronization characteristics, etc.). Quality of service characteristics may be distinguished into three major classes [1][2][3].

The performance oriented characteristics specify performance metrics such as throughput, bandwidth, delay, jitter etc.. These characteristics are very important in the specification of QoS guaranties and resource allocation, mostly in the network layer of a multimedia application.

The format oriented characteristics are higher level parameters such as video resolution, compression format, frame rate etc. There are characteristics that are used to specify quality of service in the upper layers (user, application) of a multimedia system.

The synchronization oriented characteristics specify timing requirements, such as delay tolerance between video and audio or precedence constraints between parts of information in an application. These characteristics are mostly application and information oriented.

2.2.1 QoS characteristics and CPU scheduling

In the case of CPU scheduling, QoS characteristics and requirements must be translated into classical scheduling parameters such as computation time, period, deadlines and priorities. This mapping is a difficult issue due to the complicated relationship between the user, the application and the media requirements of quality. In the rest of this section we provide a general description of the relationship between QoS and scheduling parameters.

Computation time of a task that serves a piece of multimedia information, is related to format oriented characteristics such as video resolution, compression format and processing that is required to enhance (color adjustment, filtering, etc.) and deliver the information.

Timing constraints such as period and deadline are closely related to format QoS characteristics like frame rate or synchronization characteristics.

Importance, weights and priorities among tasks are related to priorities among applications but even more to priorities of individual types of information in the same application. Priorities and weights may be used in general to formulate user preferences and demands.

2.3 CPU Scheduling and Admission Control

In general, scheduler of an OS consists of the scheduling algorithm and the admission control scheme. Admission control is an essential part of any QoS oriented scheduling framework since it prevents processor from overloading and preserves the QoS provider-user agreement.

When a new service request arrives at the processor scheduler, the system, via the admission control mechanism, checks if there are enough resources to satisfy the new request without violating quality guarantees of services that are already agreed to and admitted.

Admission control is related to the scheduling algorithm that the system uses. The most common algo-

Table 3. Implementation of QoS model to NTSC media

quality (q)	Vertical resolution (pixels)	Horizontal resolution (pixels)	Frame Rate (frames/sec)	Computation Rate	BW (Mbits/sec)
0	0	0	0	0.00E+00	0
0.1	340	453	7	1.01E-02	7.734375
0.2	377	503	11	2.03E-02	15.46875
0.3	401	534	14	3.04E-02	23.203125
0.4	418	558	17	4.06E-02	30.9375
0.5	433	577	20	5.07E-02	38.671875
0.6	445	593	23	6.08E-02	46.40625
0.7	455	607	26	7.10E-02	54.140625
0.8	464	619	28	8.11E-02	61.875
0.9	472	630	31	9.12E-02	69.609375
1	480	640	33	1.01E-01	77.34375

rithms, among OS schedulers, are EDF (Earliest Deadline First) and RM (Rate Monotonic). The former is a dynamic priority algorithm where the highest priority for execution is assigned to task that is closer to its deadline. This algorithm has proved to be optimal, among all single processor scheduling algorithms, in the sense that if a set of tasks can be scheduled by any algorithm, it can be scheduled by the EDF algorithm [12]. On the other hand, Rate Monotonic is a fixed priority algorithm, where the highest priority is assigned to the task with the shortest period. RM has proved to be optimal, in the sense of processor utilization, among all fixed priority algorithms [12]. The admission control of an OS is based on the schedulability[1] condition of the algorithm that the end-system uses to schedule task execution. When EDF algorithm is used, then OS admission control is based on inequality (1), which is the schedulability condition of EDF introduced by Liu and Leyland in [12]. Where C_i (computation time) in (1) is the time that is necessary for the processor to execute the i-th task, T_i is the period of the task and n is the number of tasks.

$$U_{EDF} = \sum_{i=1}^{n} \frac{C_i}{T_i} = 1 \qquad (1)$$

When RM algorithm is used, then admission control is based on (2), which is the schedulability condition of RM algorithm [12]. We must notice here that U_{EDF} and U_{RM} express the processor utilization[2] in the case of EDF and RM algorithm respectively. When OS uses EDF algorithm, then full utilization (U_{EDF}=1) may be achieved. On the other hand, when RM algorithm is used, then the upper bound of processor utilization that preserves schedulability of the task set is less than 0.83 (0.83 is the upper bound in the case of a set of two tasks). Lehoczky et al. in [13] has improved schedulability condition (2) to achieve better processor utilization. However for simplicity many developers still use (2).

$$U_{RM} = \sum_{i=1}^{n} \frac{C_i}{T_i} \le n(2^{1/n} - 1) \qquad (2)$$

[1] Schedulability condition checks whether scheduling a set of tasks with an algorithm is feasible.

[2] Processor utilization is the fraction of processor time spent in the execution of the task set.

3 QoS Model for Admission Control Parameters

In this section we introduce a model that relates fundamental scheduling and technical parameters to QoS. The model relates multimedia application quality characteristics that are easy for a user to understand (format characteristics), to scheduling parameters of the processing system. This relation between parameters makes the system transparent to the user and allows him to express his requirements directly to the scheduling framework. Later we will use these parameters to formulate an admission control scheme suitable for multimedia oriented OS.

3.1 Definition of quality of a multimedia service.

Every network application has an upper limit of quality that can provide. This upper limit of quality is independent of the network or other temporal characteristics of the communication infrastructure. It rather depends on the service and the service provider. When a user asks for quality, he asks for a portion of the best quality that the system supports.

In this model, when we refer to quality (q) of a service (or of an application), we mean a real in [0,1] interval which expresses a fraction of the optimal quality that the application and the provider may provide.

When quality (q) is equal to one, then the user experiences the maximum quality that the service can pro-

Table 4. Implementation of QoS model to CD quality audio media

quality (q)	Sample Rate (stereo) (samples/sec)	Computation Rate	BW (Mbits/sec)
0	0	0	0.00
0.1	8820	8.82E-05	0.13
0.2	17640	1.76E-4	0.27
0.3	26460	2.65E-04	0.40
0.4	35280	3.53E-04	0.54
0.5	44100	4.41E-04	0.67
0.6	52920	5.29E-04	0.81
0.7	61740	6.17E-04	0.94
0.8	70560	7.06E-04	1.08
0.9	79380	7.94E-04	1.21
1	88200	8.82E-04	1.35

vide. When quality (q) is driven to zero, the user does not get any service at all. The relation between quality and application characteristics (e.g. resolution, frame rate, etc.) depends on the application and the provider.

3.2 Computation Rate and QoS

It is easy to see from (1), (2) that, what is really interesting about the task parameters in the case of scheduling, is the rate of computation time (C) and period (or deadline) (T). Both computation time and period (or deadline) of a multimedia information task, are related to QoS. In a video application for example, the higher the frame resolution and size , the higher the quality. However in this case, computation time is the processing time of a frame (e.g. number of pixels X processing time per pixel). Thus computation time is proportional to quality. Period of a video service is related to the frame rate of the video information. Period is inversely proportional to quality of the service, since frame rate may vary from 33 frames per second (best quality) with period 1/33 sec to less than 10 frames per second (poor quality) with period T=1/10 sec. In this model we express computation rate, instead of computation time and period individually, as a function of QoS. We assume linear relation between computation rate and QoS (3). In (3) q_i is QoS of i-th service and a_i is constant, specified by the type of the service.

$$R_i(q_i) = \frac{C_i(q_i)}{T_i(q_i)} = a_i \cdot q_i \qquad (3)$$

In section V there are some illustrative examples of how computation rate may be expressed in QoS characteristics (resolution and frame rate) of a multimedia service.

3.3 Bandwidth and QoS

The necessary bandwidth for a service is closely related to the size of information as well as time constraints such as period or deadline. According to the analysis of section III.B both parameters are related to QoS. Thus, bandwidth (4) is related to QoS in a manner similar to this of computation rate R

$$BW_i(q_i) = \frac{S_i(q_i)}{T_i(q_i)} = b_i \cdot q_i, \qquad (4)$$

where in (4) S_i is the size (in bits) of the information that arrives in the end-system and T_i is the time window (period, deadline) that the information could be transferred. Variable q_i is the quality of the i-th service, b_i is a constant related to the information.

3.4 Weights and QoS.

Weight is a constant that can be specified by both, provider and user and expresses the importance of the corresponding multimedia application. All the media involved in a multimedia application and moreover all applications are not of the same importance. Thus, considering user and application requirements, scheduling of multimedia applications has to embed mechanisms that support importance gradation, among services. Weights scale the influence of multimedia services on the overall quality. Furthermore, in a priced service environment, weights may be used to introduce the pricing policy of the application provider to the scheduling framework.

4 Admission Control Scheme

The scheme proposed in this paper admits not only requests for service but also the QoS that these requests will perform. The main concept of this model is that scheduling parameters (e.g. computation time, period, deadline) of multimedia information are closely related to the quality of this information. Thus, if these relations can be formulated in closed form expressions, then admission control may be formed as an optimization process. In general, optimization is a task of high computa-

Table 5. Implementation of QoS model to voice quality audio media.

quality (q)	Sample Rate (samples/sec)	Computation Rate	BW (Mbits/sec)
0	0	0	0.000
0.1	800	8.0E-06	0.006
0.2	1600	1.6E-05	0.012
0.3	2400	2.4E-05	0.018
0.4	3200	3.2E-05	0.024
0.5	4000	4.0E-05	0.031
0.6	4800	4.8E-05	0.037
0.7	5600	5.6E-05	0.043
0.8	6400	6.4E-05	0.049
0.9	7200	7.2E-05	0.055
1	8000	8.0E-05	0.061

tion complexity. However in the case of linear equations with real variables, due to Simplex algorithm [14], solutions can be achieved in acceptable time. Moreover, the idea of optimization process is made more realistic, if one considers that admission control takes place only once, in a multimedia service session, during provider-user negotiation when time constraints are loose.

The optimization model we use in this paper consists of the objective function and constraints imposed by the involved parts in the service.

As an objective function in our model we use the weighted overall quality (6) that the end-system provides. In (6) q_i is the quality of i-th multimedia application. Weight w_i is a constant that can be specified by the provider and the user and expresses the importance of the corresponding multimedia application. Weights scale the influence of multimedia services in the overall quality.

$$Q(\bar{q}) = \sum_{i=1}^{n} w_i \cdot q_i \tag{5}$$

We express schedulability conditions (1) and (2) as functions of QoS by using (3). Depending on the algorithm the OS employs, schedulability condition becomes inequality (6) for EDF and inequality (7) for RM.

$$U_{EDF}(\bar{q}) = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} a_i \cdot q_i \le 1 \tag{6}$$

$$U_{RM}(\bar{q}) = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} a_i \cdot q_i \le n(2^{1/n} - 1) \tag{7}$$

According to (4) the bandwidth is related to quality in a manner similar to that of processor utilization U.

$$B(\bar{q}) = \sum_{i=1}^{n} BW_i = \sum_{i=1}^{n} b_i \cdot q_i \le \underline{BW}, \tag{8}$$

where in (8), $\underline{BW}$ is the available bandwidth of the input or output devices. Inequality (8) could be expressed as the set of inequalities (9), in order to distinguish bandwidth constraints of individual input/output devices of the end-system.

$$B_j(\bar{q}) = \sum_{i} BW_i = \sum_{i} b_i \cdot q_i \le \underline{BW_j}, \quad j = 1..k \tag{9}$$

Although the above inequalities (8), (9) are for services with uncompressed information, they can be applied, as worst case estimation, into cases where compressed information is transferred.

User preferences of QoS may be expressed as bounds on the corresponding quality variables (q). In (10), $\underline{q_i}$ is the lower bound of quality that user demands for the i-th service.

$$\underline{q_i} \le q_i \le 1, \quad i = 1..n \tag{10}$$

According to the analysis of this section and expressions (5), (6), (7), (9) and (10) admission control scheme may be formulate as the linear optimization problem (11)-(15).

$$\textbf{maximize}\; Q(\bar{q}) = \sum_{i=1}^{n} w_i \cdot q_i \tag{11}$$

subject to

$$\text{(a)}\; U_{EDF}(\bar{q}) = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} a_i \cdot q_i \le 1 \tag{12}$$

or

$$U_{RM}(\bar{q}) = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} a_i \cdot q_i \le n(2^{1/n} - 1) \tag{13}$$

$$\text{(b)}\; B_j(\bar{q}) = \sum_{i} BW_i = \sum_{i} b_i \cdot q_i \le \underline{BW_j}, \quad j = 1..k \tag{14}$$

$$\text{(c)}\; \underline{q_i} \le q_i \le 1, \quad i = 1..n \tag{15}$$

Although analysis is made for Continues Media information, it is also applicable to aperiodic information with real time constraints.

4.1 The admission control strategy.

When either a single request or a set of n concurrently received requests arrive for service at the end-system, then admission controller formulates the corresponding linear optimization problem. If users have minimum preferences of quality, these can be introduced in the optimization as lower bounds in quality variables, see (15). If there are already admitted requests in the end-system, then the upper bounds of system utilization and bandwidth in (12), (13) and (14) are reduced to the available ones;

Solutions of the linear optimization problem are the quality levels of each service that the end-system can guarantee. Following, the system informs the network manager and the users, about the quality levels that can be admitted and the users have the ability to accept or refuse the offer. If scheduling of a set of requests is not feasible then the system may start a renegotiating procedure, where users should either reduce their minimum preferences of quality or withdraw (if they are not satisfied) their requests.

5 Simulation Results

We performed a number of simulations to evaluate the effectiveness of our model and admission control scheme. In this section we introduce some simple but illustrative examples on the performance of our solution. In these examples, we consider multimedia services, that involve a wide variety of media, from High Definition TV to voice quality audio. For the sake of simplicity we assume that CPU execution time per sample (for audio) or per pixel (for video) is constant and equals to 10 nano-second (10 nsec). This assumption focuses on workstations that include enhanced multimedia CPUs [15][16][17].

In Tables 2, 3,4 and 5 there are implementation examples of QoS model for admission control parameters for the media types of Table 1.

We must note here that, the exact way of mapping media characteristics onto levels of QoS is, in general, specified by the service provider.

5.1 Implementation of the admission control scheme.

We have implemented the admission control scheme, as it is formulated in (11)-(15), in MapleV environment. Using this implementation we simulate some illustrative scenarios of arrivals in the CPU scheduler.

Scenario a. In this scenario we assume that in the EDF scheduler of the end-system arrive fifty requests to be served. From these requests, one is for HDTV quality video service, nine are for NTSC quality video services, ten are for CD quality audio and the rest thirty requests are for voice quality services. We assume that all requests are of the same importance ($w_1=w_2=...w_{50}=1$) and users have no minimum preferences for the services. Furthermore we assume that all services arrive at the end-system through network with bandwidth capability of 3000 Mbit/sec. Thus linear optimization of (11)-(15) becomes:

$$\textbf{maximize}\ Q(\bar{q}) = \sum_{i=1}^{n} q_i \tag{16}$$

subject to

$$\text{(a)}\quad \begin{aligned} &U_{EDF}(\bar{q}) = \sum_{i=1}^{50} R_i = 1.23 \cdot q_1 + \sum_{i=2}^{10} 0.101 \cdot q_i + \\ &+ \sum_{i=11}^{20} 8.8 \cdot 10^{-4} \cdot q_i + \sum_{i=21}^{50} 8 \cdot 10^{-5} \cdot q_i \le 1 \end{aligned} \tag{17}$$

$$\text{(b)}\quad \begin{aligned} &B_1(\bar{q}) = \sum_{i=1}^{50} BW_i = 2812.5 \cdot q_1 + \sum_{i=2}^{10} 77.343 \cdot q_i + \\ &+ \sum_{i=11}^{20} 1.35 \cdot q_i + \sum_{i=21}^{50} 0.061 \cdot q_i \le 3000 \end{aligned} \tag{18}$$

$$\text{(c)}\quad 0 \le q_i \le 1,\ \ i = 1..50 \tag{19}$$

Simulation results of this scenario are placed in Table 6. From these results it is concluded that, the admission control scheme assigns maximum quality to all the user requests except the first one (HDTV quality service), which is the most resource consuming. Thus, when all requests are of the same importance, the proposed admission control scheme maximizes the number of requests that can be serviced, by rejecting these with high system utilization demands.

Scenario b. This is the same as *scenario a*, but, end-system scheduler uses Rate Monotonic (RM) algorithm instead of EDF. Thus constraint (17) becomes (20):

$$\begin{aligned} &U_{RM}(\bar{q}) = \sum_{i=1}^{50} R_i = 1.23 \cdot q_1 + \sum_{i=2}^{10} 0.101 \cdot q_i + \\ &+ \sum_{i=11}^{20} 8.8 \cdot 10^{-4} \cdot q_i + \sum_{i=21}^{50} 8 \cdot 10^{-5} \cdot q_i \le 50 \cdot \left(2^{1/50} - 1\right) \end{aligned} \tag{20}$$

Simulation results of this scenario are placed in Table 7. These results show that the proposed admission control scheme works properly with RM algorithm, since if all requests are of the same importance then admission control maximizes the number of requests that will be served by the end-system, in a way similar to that of scenario a.

Table 6. Simulation results of scenario a.

Type of request	**importance of request w**	**Bandwidth capability BW (Mbits/sec)**	**quality[1] q**	**processor utilization U**	**Necessary Bandwidth BW (Mbits/sec)**
HDTV quality video request 1	$w_1=1$		q1=0	0	0
NTSC quality video requests 2-10	$w_2=...=w_{10}=1$		$q_2=...=q_{10}=1$	0.909	696.0
CD quality audio requests 11-20	$w_{11}=...=w_{20}=1$		$q_{11}=...=q_{20}=1$	0.0088	13.5
voice quality audio requests 21-50	$w_{21}=...=w_{50}=1$		$q_{21}=...=q_{50}=1$	0.0024	1.83
Total	**3000**			**0.92**	**711.33**

Table 7. Simulation results of scenario b.

Type of request	**importance of request w**	**Bandwidth capability BW (Mbits/sec)**	**quality q**	**processor utilization U**	**Necessary Bandwidth BW (Mbits/sec)**
HDTV quality video request 1	$w_1=1$		q1=0	0	0
NTSC quality video requests 2-10	$w_2=...=w_{10}=1$		$q_2=0$, $q_3=0.8$, $q_4=0$, $q_5=...=q_{10}=1$	0.686	525.9
CD quality audio requests 11-20	$w_{11}=...=w_{20}=1$		$q_{11}=...=q_{20}=1$	0.0088	13.5
voice quality audio requests 21-50	$w_{21}=...=w_{50}=1$		$q_{21}=...=q_{50}=1$	0.0024	1.83
Total	**3000**			**0.698**	**541.26**

Table 8. Simulation results of scenario c.

Type of request	**importance of request w**	**Bandwidth capability BW (Mbits/sec)**	**quality q**	**processor utilization U**	**Necessary Bandwidth BW (Mbits/sec)**
HDTV quality video request 1	$w_1=2\ 10^4$		q1=0.5	0.615	1406.25
NTSC quality video requests 2-10	$w_i=10^3+i\ 100$ $w_2<w_3<..<w_{10}$		$q_2=...=q_6=0$ $q_7=0.7$ $q_8=...=q_{10}=1$	0.3737	286.17
CD quality audio requests 11-20	$w_i=10^2+i\ 10$ $w_{11}<w_{12}<..<w_{20}$		$q_{11}=...=q_{20}=1$	0.0088	13.5
voice quality audio requests 21-50	$w_i=i$ $w_{21}<w_{22}<..<w_{50}$		$q_{21}=...=q_{50}=1$	0.0024	1.83
Total		**3000**		**1**	**1707.7**

Scenario c. This scenario assumes that requests are of different importance compared to each other. Thus weights w, in the objective function, have values that express the importance of the corresponding request. The relations presented in the corresponding column of Table 8 give the values of weight w. Furthermore, CPU scheduler is assumed to use the EDF algorithm. The type and the number of the requests are the same as in scenario a. Simulation results of this scenario are presented in Table IV. When requests have different importance, then admission control assigns the best quality to those with the higher importance or with lower system utilization demands and rejects or degrades those with lower importance and high utilization demands.

From the Tables 2-5, it is obvious that we have chosen a discrete mapping between quality and media characteristics. Therefore, the quality computed by the optimization process, is rounded down to the nearest quality that is included in the mapping tables. Consequently, the implementation of the computed assignment of quality to applications tends to optimal as the discrete level of qualities, that the system supports, increases.

6 Conclusions

We have presented an applicable way to model scheduling parameters of distributed multimedia applications in respect to QoS. Furthermore we have presented an admission control scheme which in cooperation with this model, introduces QoS in the scheduling admission control of the end-systems (servers, workstations and set top boxes) of distributed multimedia services. We have formulated this scheme for the Earliest Deadline First (EDF) and the Rate Monotonic (RM) CPU scheduling algorithms, which are the most common, among all CPU scheduling algorithms, in operating systems. This admission control scheme, improves the capability of end-systems of supporting networking with QoS guarantees. By the use of this admission control framework, the end-system becomes able to introduce, in the user-provider negotiations for QoS, its own resource constraints. We have performed extended simulations, where we include the most common types of video and audio services. Simulation results show that, the framework proposed in this paper, can handle effectively the admission of multiple, concurrently arrived, multimedia service requests.

References

[1] A. Vogel, B. Kerhervé, G. v. Bochmann and J. Gecsei, "Distributed Multimedia Applications and Quality of Service- A Survey," *IEEE Multimedia*, vol. 2, no. 2, pp. 10-19, Summer 1995.

[2] G.v. Bochmann and A. Hafid, "Some principles for quality of service management", *Distributed Systems Engineering Journal* , vol. 4, no. 1, pp. 16-27, 1997.

[3] C. Aurrecoechea, A.T. Campbell, and L. Hauw, "A Survey of Quality of Service Architectures," *Multimedia Systems Journal* Special Issue on QoS Architecture, May 1998.

[4] L.C. Wolf, W. Burke, and C. Vogt, "Evaluation of a CPU Scheduling Mechanism for Multimedia Systems," *Sofware-Practice and Experience*, vol. 26, no. 4, pp. 375-398, April 1996.

[5] D.K.Y. Yau and S.S. Lam, "Adaptive Rate-Controlled Scheduling for Multimedia Applications," *IEEE/ACM Trans. on Networking*, vol. 5, no. 4, pp. 475-488, August 1997.

[6] G. Coulson and G. Blair, "Architectural Principles and Techniques for Distributed Multimedia Application Support in Operating Systems," *Operating Systems review*, vol. 29, no. 4, pp. 17-24, 1995.

[7] S.T.C. Chou and H. Tokuda, "System Support for Dynamic QoS Control of Continuous Media Communications", *Network and Operating Systems Support for Digital Audio and Video*, ED. by P. Venkat Ranfan LNGS 712, Springer Verlag Berlin, pp. 363-368, 1993.

[8] C.W. Mercer, S. Savage, and H. Tokuda, "Processor Capacity Reserves: Operating System Support for Multimedia Applications, " *Proc. of the Inter. Conf. on Multimedia Computing and Systems*, pp. 90-99, May 1994.

[9] P. Goyal, X. Guo, and H.M. Vin, "A Hierarchical CPU Scheduler for Multimedia Operating Systems," *Proc. of the 2nd USENIX symp. on OSDI*, pp. 107-122, Seattle,WA, Oct. 1996.

[10] M.B. Jones, D. Rosu and M-C. Rosu, "CPU, Reservations and Time Constraints: Efficient, Predictable Scheduling of Independent Activities," Proc. of the 16th ACM Symp. on Operating Systems Principles, pp. 198-211, Saint-Malo, France, Oct. 1997.

[11] J. Nieh and M.S. Lam, "The Design, Implementation and Evaluation of SMART: A Scheduler for multimedia Applications," Proc. of the 16th ACM Symp. on Operating Systems Principles, pp. 184-197, Saint-Malo, France, Oct. 1997.

[12] C.L. Liu and J.W. Layland, "Scheduling Algorithms for Multiprogramming in a Hard-Real-Time Environment," *J. of the Ass. for Computing Machinery*, vol. 20, no. 1, pp. 46-61, January 1973.

[13] J.P. Lehoczky, L. Sha, and Y. Ding, "The rate-monotonic scheduling algorithm: Exact characterization and average behavior," *IEEE Real-Time Systems Symp.*, pp. 166-171, 1989.

[14] C.H. Papadimitriou and K. Steiglitz, "Combinatorial Optimization-Algorithms and Complexity," ED. Prentice-Hall 1982.

[15] K. Diefendorff and P.K. Dubey, "How Multimedia Workloads Will Change Processor Design," *IEEE Computer Mag.*, pp. 43-45, September 1997.

[16] S. Undy, M. Bass, D. Hollenbeck, W. Kever, and L. Thayert, "A Low-Cost Graphics and Multimedia Workstation Chip Set," *IEEE Micro Mag*, pp. 10-22, April 1994.

[17] H. Veendrick, O. Popp, G. Postuma, and M. Lecoutere, " A 1.5 GIPS video signal processor (VSP)," *Proc. of the CICC*, pp. 95-98, San Diego, May 1994.