References:

[1] D. Reininger and D. Raychaudhuri, “Bit-Rate Characteristics of a VBR MPEG Encoder for ATM Networks”, in Proc. IEEE ICC'93, Geneva, Switzerland, May 1993.

[2] ATM: User-Network Interface Specification, Version 3.0. The ATM Forum, PTR Prentice Hall, 1993.

[3] D. Reininger, G. Ramamurthy and D. Raychaudhuri “VBR+, an enhanced service class for multimedia traffic,” Contribution 94-0353, ATM Forum, Munich, Germany, May 1994.

[4] D. Raychaudhuri, D. Reininger, M. Ott and G. Welling, “Multimedia Processing and Transport for the Wireless Personal Terminal Scenario,” in Proc. of SPIE Visual Communications and Image Processing Conference, VCIP'95, Taipei, Taiwan, May 1995.

[5] M. Ott, D. Reininger and W. Luo, “Adaptive and Scalable QoS for Multimedia using Hierarchical Contracts,” in Proc. of ACM Multimedia, Boston, MA, November 1996.

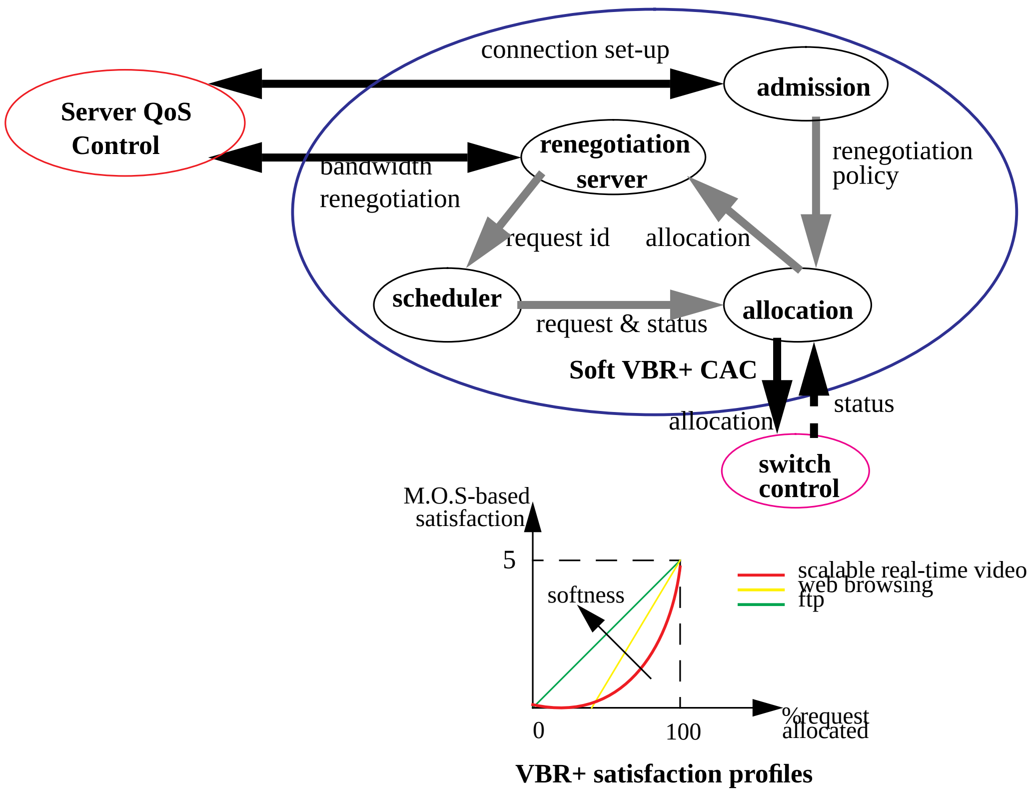

Figure 2: Soft bandwidth allocation based on satisfaction profiles

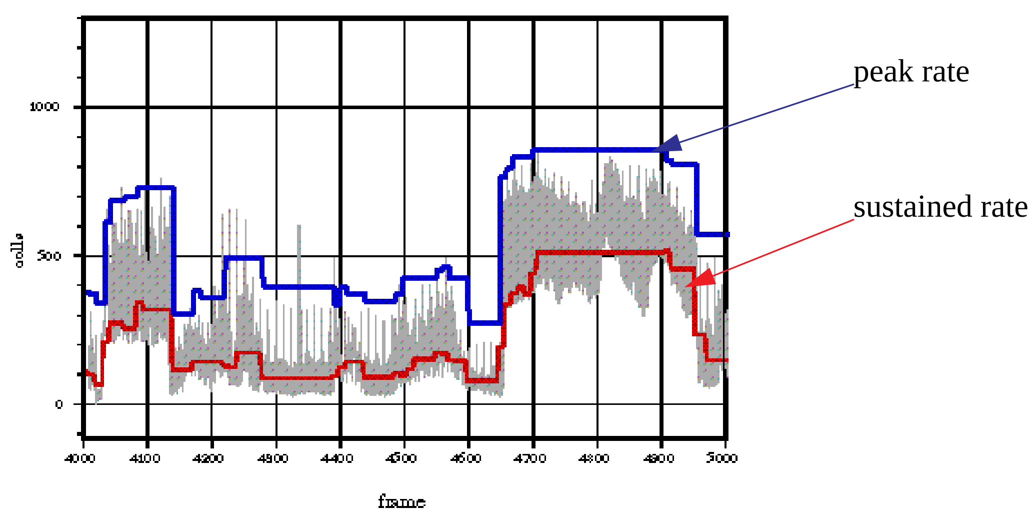

Figure 3: Dynamic bandwidth allocation process for VBR MPEG-2 video

that the new connection could have on the satisfaction of previously admitted connections. After evaluating the cost/benefits of the connection's request, the *admission module* makes an admission decision. Once a connection is admitted, bandwidth renegotiations are sent to the *renegotiation server.* The *renegotiation server* maintains a table with the connections' status and registers outstanding requests for bandwidth renegotiation with the scheduler passing a *request id*. The *scheduler* prioritizes requests to match processing resources at the node controller. Priority policies and their effect on system's performance are open issues for research. Once a connection has a renegotiation request scheduled, its *request and status* are passed to the *allocation module*. The allocation module executes a dynamic programming algorithm to allocate bandwidth by maximizing overall class satisfaction subject to a target minimum. When bandwidth is not available to maintain the target minimum, the allocation model can force some connections with large allocations and significantly high satisfaction, to release bandwidth. We call this procedure “network initiated renegotiation”.

In many ways this approach to bandwidth allocation constitutes a paradigm shift from current bandwidth allocation policies. In the proposed soft allocation scheme, established connections do not “own” bandwidth like they do in the traditional “telco” model; connections dynamically “rent” bandwidth according to their needs. The network accommodates the aggregate demand using a best-effort approach subject to minimum user satisfaction levels. Thus, at a given instant, it may not be possible to accommodate the bandwidth demand of a given connection. Then, connections must “self-accommodate” to their currently assigned bandwidth. As a result, individual user satisfaction will decrease during that period. To avoid severe degradation during these periods, a hybrid model where connections “own” some minimum amount of bandwidth and renegotiate any additional demand could be used. This minimum reservation will depend on the application's satisfaction profiles and softness.

3. Implementation

A “proof-of-concept” implementation of the described architecture has been completed. The client application consists of a multimedia browser and interactive scalable video retrieval with adaptive QoS and feedback of network usage and cost. The components of this application are active objects implemented in C++ which are rendered as three dimensional objects on the screen using the Xlib library of the X-Window system. The user interacts with the application by means of direct manipulation only, pushing buttons, assembling compound objects, and dragging widgets on the screen. The hardware test-bed consists of two Sun’s UltraSparc machines connected with Zeitnet ATM NICs to an NEC Model5 switch controlled by a Sun Sparc 10. The video server is implemented in C++ using a series of media streaming modules [5]. The server QoS control and the CAC & bandwidth allocation modules are implemented in Java. The bandwidth renegotiation signalling between the server and the switch currently takes place outside the ATM Forum’s UNI VBR signalling using a dedicated control channel.

Figure 3 shows a sample trace of the renegotiation process for VBR MPEG-2 video. The peak and sustained rate curves show how the system is able to track the dynamics of the instantaneous bitrate We are in the process of gathering statistics to characterize the system’s performance. We hope that this characterization will help finding quantiative answers to questions such as: what are appropriate softness profiles for multimedia, how do softness profiles impact bandwidth allocation, how does renegotiation frequency impact QoS and how much processing power is required.

The server's QoS manager determines and configures the control services required by generating appropriate sub-contracts. For example, in the case of VBR video, the QoS manager decides it needs a media and a traffic control. The media contract specifies a range for quantization, frame-rate and image resolution. The transport contract is specified in terms of tolerable delay, cell-loss, and expected bit-rate range.

The manager and control modules operates on different time-scales. Control modules operate when they receive alert signals from the data flow domain or contract renegotiations from managers. On the other hand, managers are isolated from the data flow and they react to alert signals from control modules and contract renegotiations from higher services. While the control modules operate within the terms specified in their contracts, no alert signals are generated. This allows service manager modules to be active only when QoS renegotiation from clients occur. For the network, the renegotiation functionality provided by VBR+ service class is relatively more difficult to achieve compared to CBR service class. Its successful implementation constitutes a challenging network engineering problem. The main complexity comes from the design of a robust bandwidth allocation algorithm and the associated protocol and signalling support. At the same time, the potentials for significant statistical multiplexing gain from a large number of users that would find the service to have the right cost-benefit performance are an incentive to network operators. Since the network does not know the sources' instantaneous requirements, at some times it may not be possible to meet the connections' required bandwidth. Thus, the VBR+ service class price-performance is suitable for users willing, and applications able, to tolerate occasional degradation in QoS in exchange for a favorably priced “on-demand” bandwidth assignment.

The bandwidth allocation algorithm integrates the applications' ability to tolerate degradation in terms of *satisfaction profiles*. The satisfaction profiles express the mean-opinion-score (MOS) satisfaction of applications' users as a function of the percentage of requested bandwidth allocated. As shown at the bottom in Figure 2, different applications, such as file-transfer, web browsing or scalable multimedia including video, have different tolerance to server-network bandwidth mismatches reflecting their inherent characteristics. These characteristics are incorporated in our model using the concept of application's *softness*. For example, in Figure 2, file-transfer is consider the softest application since, by being not interactive, it has no real-time constraints and is more delay tolerable. We consider its satisfaction profile linear with the allocated bandwidth. For browsing, the interactivity requires a critical bandwidth to be allocated before the service is usable; after that critical bandwidth is assigned, increments contribute linearly to the MOS. Browsing is therefore consider to have a smaller softness than non-interactive file-transfer. Finally, the least soft of the applications considered is real-time scalable multimedia applications; the satisfaction profile is non-linear since it requires significantly different bandwidth assignments to achieve a given MOS increment depending on the MOS region.

Figure 2 also shows the modules of the *soft* bandwidth allocation algorithm. In the model, the *server QoS control* establishes a VBR+ connection with the switch at the edge of the network. The *admission module* estimates the potential demand generated by the new connection using information provided during *connection set-up*. This information includes the connection's *satisfaction profile* and when available, a probable traffic envelope. Using the potential demand estimate, the *admission module* checks the available bandwidth at the node and the potential effects

2. Architecutre

In the proposed architecture, clients specify their QoS requirements in terms of dynamic contracts. At the server, the QoS control maps the client's contract from the application level description to source level parameters such as quantization, frame-rate, resolution and delay requirements. In addition, the QoS control adapts the media QoS to match the allocated bandwidth, and dynamically renegotiates with the network for the necessary bandwidth to maintain the QoS contract. Bandwidth renegotiation between the server and the network takes place in the context of a new network service class, called VBR+ [3]. A "soft" network bandwidth allocation scheme determines the bandwidth assignment to each renegotiation based on the total available bandwidth and the relative tolerance to QoS degradation (softness) of active connections. A binding architecture of distributed objects provides hierarchical services that renegotiate and control resources. This objects have a defined contract-oriented application program interfaces (APIs).

Figure 1 shows the architecture modules. The *application* [1] at the *multimedia terminal* specifies an *application QoS contract* in terms relevant to the application's components. In the case of video, the contract is an abstract description of desired quality (e.g., from excellent to poor). The abstract quality specfication is maped to a target picture detail and frame-rate. The picture detail is further maped to spatial resolution and quantization. The network and media objects can optionally modify the QoS contract to adapt the bit-stream to the terminal's capabilities. For example, the network control module at the terminal could temporarily lower the QoS contract to cope with buffer overflows, or it could request a change in transport protocol (e.g, from UDP to TCP) when a significant error rate is detected. In the case of CPU-based decoding, the media layer can match the frame-rate and picture detail to the processing power available [4]. The *client QoS manager* collects the contracts from the application, the media and transport/network objects and decides on the appropriate *connection's QoS contract*.

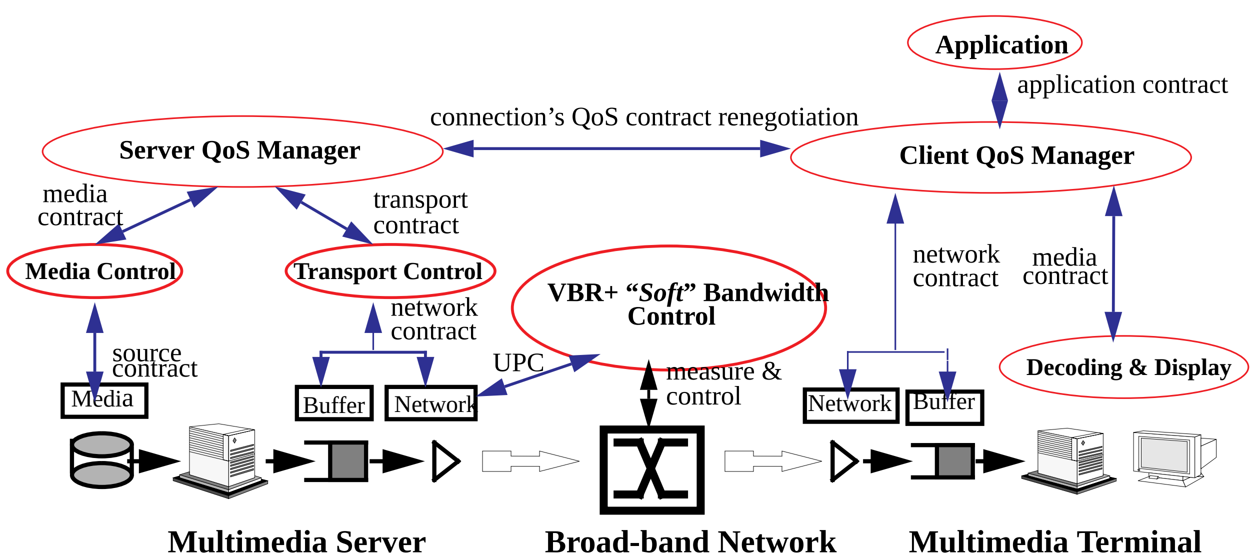

Figure 1: Architecture for Dynamic Bandwidth Control with Adaptive QoS

1. Words in italics refer to elements depicted in the figure.

Dynamic Bandwidth Allocation for Distributed Multimedia with Adaptive QoS

D. Reininger, M. Ott, G. Michelitsch and G. Welling

NEC USA. Inc.
4 Independence Way
Princeton, New Jersey 08540
e-mail: djr@ccrl.nj.nec.com

Extended Abstract

1. Introduction

Dynamic bandwidth allocation is necessary for distributed multimedia systems that support application-level control of media quality and network usage/cost. Traditional distributed multimedia systems lack integrated means of providing bandwidth-on-demand to match the non-stationary requirements of interactive multimedia. This paper describes the architecture of a multimedia system that provides these means by integrating adaptive client Quality-of-Service (QoS) renegotiation support, server source rate control and dynamic network bandwidth allocation. The description of and experiences with a “proof-of-concept” implementation of the system are given.

Multimedia (MM) applications dynamically demand different QoS grades from servers and networks during a session. Users of an interactive multimedia application may want to resize a video window and consequently request a larger image resolution; a ``trick mode" such as fast-forward, rewind, pause or jog can be requested from a video server; or the user may suddenly start to browse an image data-base. The variable bit-rate (VBR) broad-band network service is ideal to support application-level QoS since its traffic descriptors naturally tune to multimedia traffic burstiness [1].

In broadband networks that support QoS, like ATM, the connection model requires a set of traffic descriptors to be declared by the source at connection set-up. In ATM these parameters are referred to as Usage Parameter Control (UPC) set. The UPC consists of peak-rate, burst size and sustained rate [2]. Based on this UPC and the desired QoS a network’s connection admission control (CAC) process either accepts or rejects the connections’ request.

The traditional bandwidth allocation model must be modified to efficiently provide QoS to distributed interactive multimedia applications. QoS renegotiation requires global architectural support with flexible connection management and a suitable binding architecture which allows to tailor communication services to applications requirements. The use of a fixed UPC does not allow QoS support and high channel utilization since VBR multimedia traffic varies significantly over different time-scales. This becomes an even more critical issue in the mobile wireless multimedia scenario where it is difficult to commit fixed resources to VBR traffic for the duration of a connection while maintaining high network utilization.