Self-Adjusted Network Transmission for Multimedia Data

Mei-Ling Shyu
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33124
USA
shyu@miami.edu

Shu-Ching Chen
School of Computer Science
Florida International University
Miami, FL 33199, USA
chens@cs.fiu.edu

Hongli Luo
Department of Electrical and Computer Engineering
University of Miami
Coral Gables, FL 33124
USA
h.luo@miami.edu

Abstract

High bandwidth requirements in multimedia transmission make the efficient use of limited network resource a challenging task, especially when multiple clients make their requests to the server simultaneously. In this paper, we propose a self-adjusted network transmission mechanism for multiple clients. Instead of assigning a fixed bandwidth for each client, the server determines the transmission rate for each client connection according to the buffer packets and playback rate at each client. Transmission rates are adjusted when the total requested bandwidth is larger than the network bandwidth. In addition, the proposed mechanism can minimize the bandwidth allocation and maximize the client buffer utilization. A simulation is performed and the simulation results show that the proposed mechanism can dynamically change the transmission rate for each client to avoid overflow of the client buffer, and achieve the optimal utilization of the limited network resource in multiple client network environments.

1. Introduction

The development and use of distributed multimedia applications are growing rapidly in these years. Typical examples are video-conferencing, video-on-demand, digital library and distant learning. High network bandwidth is usually needed in order to provide high quality delivery of information. Efficient utilization of network resources is essential in the provision of cost-effective multimedia services where quality-of-service (QoS) [14] requirements are met. Parameters for QoS are reliability, bandwidth, jitter, and etc.

Multimedia information in general has a highly time-varying bandwidth requirement since media data are variable bit rate (VBR) in nature due to the coding and compression technologies applied [4][7]. For example, in an MPEG-1 movie, the average frame size is usually less than 25% of its maximal frame size [9]. There are several approaches to address the bandwidth requirements of multimedia transmission. One approach is the static resource reservation schemes that allocate a constant bit rate (CBR) channel to transmit the VBR stream by the peak data rate. With large variations in bandwidth requirements of multimedia data, static allocation usually results in considerable wastage of network resources. Another approach is rate adaptation [15] that adjusts the bandwidth used by a transmission connection according to the existing network conditions. The adaptive approach can better utilize the available network resource that changes with time in comparison with the static allocation approach.

Media synchronization is also needed to guarantee jitter-free playback requirement [11][13]. Specifically, the delivery of videos imposes two challenges for multimedia transmission, namely a high-bandwidth requirement and a real time delivery constraint. In order to provide jitter-free video playback, each video frame must arrive at the client buffer before the time it is scheduled to be displayed. The data arrives at the client and is stored temporarily in the client buffer. If the buffer is full when a packet arrives, this packet will get lost and overflow occurs. If the packet cannot arrive at the client before the schedule time, underflow occurs and jitters in playback happen. The client buffer acts as a reservoir to regulate the difference between the transmission rate and the playback rate. It is an important resource for users to prevent playback jitters. Since most of the clients have limited buffer space, the network must control the transmission rate to avoid overflow or underflow to guarantee the end-to-end QoS.

Much work has been done in the resource allocation, traffic control, and multimedia synchronization [5][6][10] [16] in transmission in multimedia network. The provision of an efficient and reliable transmission rate scheduler that can optimize the utilization of buffer and bandwidth at the same time is still a challenge. In our previous work [12], we proposed a closed-loop framework for multimedia transmission and investigated the single client situation. Linear Quadratic (LQ) tracker was used to obtain the optimal transmission rate for one client connection. It was proved that the proposed closed-loop framework efficiently utilizes the network resources such as the bandwidth and the client buffer. In this paper, a self-adjusted transmission mechanism that dynamically changes the transmission rates for multiple clients is designed. A network model is introduced to capture the relationships among the transmission rates, buffer packets, and playback rates. Simulation results show that the proposed mechanism can minimize the bandwidth allocation and maximize the packet size stored in client buffer at every time instant.

The organization of this paper is as follows. The proposed network framework and the self-adjusted network transmission mechanism for multiple clients are presented in Section 2. Simulation results are given in Section 3. Section 4 concludes this paper.

2. Self-adjust network transmission

In this section, the self-adjusted network transmission mechanism for multiple clients and the network model are presented.

2.1. Network model for a single client

The network model for a single client (e.g., the j^{th} client) is given in Figure 1. Define the variables as below:

- k: time instant;
- $Q(k)$: packet size in buffer at time instant k;
- $R(k)$: packets transmitted from the server at time instant k;
- $R'(k)$: packets arriving at the client buffer at time instant k;
- $L(k)$: packets used for playback at time instant k; and
- Q_r: allocated client buffer size at the setup of the connection.

In this framework, we consider the situations that the propagation delay is quite small and can be negligible. So it can be assumed that the packets transmitted from the server at a time instant are equal to the packets received at the client. Therefore, we have

$$R'(k) = R(k) \tag{1}$$

and $R(k)$ is used to represent $R'(k)$ in this paper. Consider the relationship among $Q(k)$, $R(k)$, and $L(k)$. The buffer packet at the time instant k+1 is

$$Q(k+1) = Q(k) + R(k) - L(k) \tag{2}$$

Let $Q(k)$ be the variable we want to track, and $e(k)$ be the difference between the tracked $Q(k)$ and the allocated Q_r.

$$e(k) = Q(k) - Q_r \tag{3}$$

In order to make fully use of the client buffer, $e(k)$ should be minimized. On the other hand, the transmission rate $R(k)$ should also be minimized to save the bandwidth.

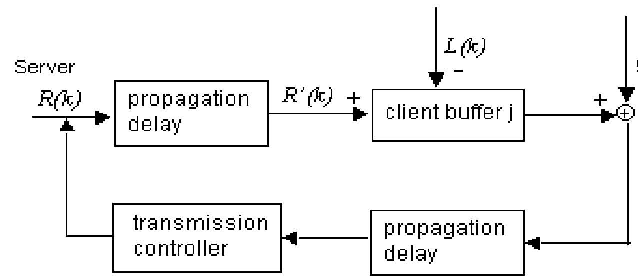

Figure 1. Proposed network model for a single client (e.g., the j^{th} client).

The *Multimedia Augmented Transition Network (MATN)* model that was proposed in our previous research [1][2][3] can provide the playback schedule for the multimedia video before the transmission starts. At the time of creation of a multimedia presentation, the *MATN* model can be used to specify the temporal constraints among various multimedia data that must be observed at the time of playback. The multimedia presentation is modeled by *MATN* and stored at a server. That is, the value for $L(k)$ is known. When the client requests the information from the server, the playback schedule is obtained by the server and can be used to determine an optimal transmission rate to satisfy the playback requirement.

Instead of transmitting the packets in a fixed rate, transmission rate can be changed automatically according to the buffer packets and playback rates in the clients. In order to maximize the buffer utilization and minimize the bandwidth allocation, Linear Quadratic (LQ) tracker [8] is used to design the transmission scheduler to get the optimal transmission rate for each client connection.

2.2. Discrete linear quadratic tracker

First, an optimal control law [8] that forces an object to track a desired reference trajectory over a specified time interval is presented. Given an object that can be represented in a linear status space form, and a known disturbance d_k, we have

$$x_{k+1} = Ax_k + Bu_k + d_k \quad (4)$$

where A and B are constant matrices describing the object, u_k is the control input, and x_k is the status of the object to be tracked. Let the cost function J of the system be

$$J = \frac{1}{2} x_N^T S_N x_N + \frac{1}{2}\sum_{k=1}^{N-1}\left(x_k^T G x_k + u_k^T H u_k\right) \quad (5)$$

with $S_N \geq 0$, $G > 0$, $H > 0$, where G and H are the weighting matrices, S_N is for the boundary conditions, and [1, N] is the period we want to track the object.

The optimal control u_k is given by

$$K_k = \left(B^T S_{k+1} B + H\right)^{-1} B^T S_{k+1} A, \quad S_N \text{ given,} \quad (6)$$

$$S_k = A^T S_{k+1}\left(A - BK_k\right) + G, \quad (7)$$

$$v_k = \left(A - BK_k\right)^T v_{k+1} - \left(A - BK_k\right)^T S_{k+1} d_k, \quad v_N = 0, \quad (8)$$

$$F_k = \left(B^T S_{k+1} B + H\right)^{-1} B^T, \quad (9)$$

$$u_k = -K_k x_k + F_k v_{k+1} \quad (10)$$

where K_k is the feedback gain, F_k is the feedforward gain, and S_k and v_k are auxiliary sequences when calculating the optimal u_k.

Applying the control input sequence u_k to the object, we can get a sequence of object status that minimizes the quadratic cost function J.

2.3. Optimal rate transmission mechanism for multiple clients

Assume that there are m clients that request data from the server simultaneously. For the j^{th} client, let J_j be the cost function, $Q_{r,j}$ be the allocated buffer size of the j^{th} client at the connection, $e_j(k)$ be the difference between the allocated buffer size and the buffer packet of the j^{th} client at the time instant k, $Q_j(k)$ be the buffer packet at the time instant k, and $R_j(k)$ be the transmission rate of the j^{th} client. Define the difference between the allocated buffer size and the buffer packets as

$$e_j(k) = Q_j(k) - Q_{r,j} \quad (11)$$

To maximize the client buffer $Q_j(k)$ is to minimize the $e_j(k)$.

The optimization function of the server side should be

$$J = \sum_{j=1}^{m} J_j = \sum_{j=1}^{m}\left(\frac{1}{2}\sum_{k=1}^{N-1}\left(e_j^2(k) + R_j^2(k)\right)\right) \quad (12)$$

Since the m clients are independent, if the J_j of the j^{th} client is minimal, then the sum of the J_j functions is also the minimal.

$$J_j = \frac{1}{2}\sum_{k=1}^{N-1}\left(e_j^2(k) + R_j^2(k)\right) \quad (13)$$

For each client, the LQ tracker is used to achieve the optimal transmission rate.

After the server calculates the optimal bandwidth for each client, an effective bandwidth W_j for the connection for the j^{th} client should satisfy the requirement that

$$\sum_j W_j \leq BW \quad (14)$$

where BW is the network bandwidth.

Since the transmission rate is optimized individually, it is possible that the sum of all the requested bandwidth is greater than the network bandwidth. Therefore, we need to reallocate the bandwidth to each client connection. To provide fairness among all clients, the bandwidth reallocated should be proportional to the real requirement of each client. An algorithm for the multimedia server is given below in the form of the pseudo code. Let *rate(j)* be the transmission rate for the j^{th} client and *Total_rate* be the sum of the transmission rates for all the m clients.

Pseudo code for rate adjustment at time instant k on the sever is given as follows.

```
Total_rate = Σ_{j=1}^{m} rate( j);
if  Total_rate > BW
    for  j = 1 to m do
        rate(j)=rate(j)*(BW/Total_rate);
    endfor
endif
```

Sometimes when the playback rate is low and the transmission rate is large, overflow may occur. Since the playback schedule is known, we can predict the situation and avoid it. The buffer size can be increased automatically at the possible overflow instant to accommodate those packets and be decreased to the normal value when overflow period ends.

For each client, we have the following buffer difference equation (15) of the client buffer that can be used to obtain the optimal transmission rate. Since for each client, we follow the same computation procedure, in order to simplify the notation, *e(k)*, *R(k)*, and *L(k)* are used to represent the corresponding values of the j^{th} client at time instant *k*.

$$e(k+1) = e(k) + R(k) - L(k) \qquad (15)$$

According to the corresponding equations (4) and (5), we set the values $A = 1$, $B = 1$, $S_N = 0$, $G = 1$ and $H = 1$. Then, the optimal transmission rate *R(k)* can be obtained by solving the following equations.

$$K_k = (S_{k+1} + 1)^{-1} S_{k+1}, \quad S_N = 0, \qquad (16)$$

$$S_k = S_{k+1}(1 - K_k), \qquad (17)$$

$$v_k = (1 - K_k)^T v_{k+1} + (1 - K_k)^T S_{k+1} L_k, \quad v_N = 0, \qquad (18)$$

$$F_k = (S_{k+1} + 1)^{-1} \qquad (19)$$

Therefore the optimal transmission rate at time instant *k* is

$$R(k) = -K_k e(k) + F_k v_{k+1} \qquad (20)$$

and the packets in the buffer at time instant *k* +1 is

$$Q(k+1) = e(k+1) + Q_r \qquad (21)$$

In the proposed algorithm, the transmission rate for the whole transmission period can be calculated before the transmission starts. Lots of the parameter calculation can be done off-line, which simplifies the implementation of the transmission rate controller. That is, the sequences S_k, K_k and F_k can be computed off-line before the transmission rate applies. The gains K_k and F_k can be stored for use when the actual transmission proceeds. The only work left to do is to compute the optimal control *R(k)* with equation (20).

3. Simulation results

In this paper, a simulation is performed to illustrate how the proposed self-adjusted network transmission for multiple clients can be achieved. For simplicity, assume that there are 3 clients simultaneously connecting to the host and requesting the services. The simulation is run within the interval [1, 50] seconds with the time increment of 1 second and uses the randomly generated playback rates to simulate the real playback scenario. Assume that the playback rate is between 0.1MB per second (MBps) and 1MBps, the allocated client buffer size is 1MB, and the network bandwidth is 2 MBps. The transmission rates of the clients and the total transmission rate of the server are depicted in Figure 2. In Figure 2, the transmission rates 1, 2, and 3 indicate the optimal transmission rates for client connections 1, 2, and 3 after the rate adjustment.

As can be seen from the figure, at some time instant, the optimal transmission rate should be adjusted when the

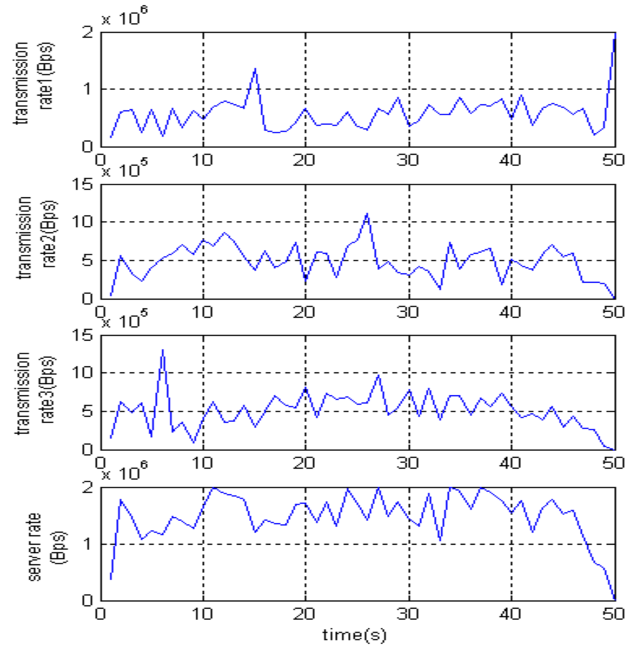

Figure 2. Transmission rate changes for each client (clients 1, 2, and 3) and the server (network bandwidth = 2 MBps, t = 1 to 50 second, and increment = 1 second).

total transmission rate of all the clients is larger than the network bandwidth (for example, at time instants 34^{th} second and 37^{th} second). After the rate adjustment, the total transmission rate will be equal to the network bandwidth, which means the limited bandwidth is utilized maximally to satisfy the playback requirements of all the clients. If the transmission rate is not large enough to satisfy the playback rate *L(k)* after the adjustment, the packets in the client buffer will be consumed so that the difference between the allocated buffer size (Q_r) and the packet size in the buffer (*Q(k)*) increases. The later optimized transmission rates will be automatically obtained to alleviate this situation and satisfy the playback requirement.

Each client has its own playback schedule and has its peak rate at different time instants. If we allocate the bandwidth according to the peak rate, it is a waste of bandwidth since most of the time the playback requirement is much below the peak rate. In addition, there will be fewer clients that can be serviced by the server. In our approach, the proposed self-adjusted network transmission mechanism dynamically allocates the bandwidth to each client according to its playback requirement and buffer packets. At a certain time instant, more bandwidth can be assigned to those clients whose playback requirements are high and less bandwidth can be assigned to those clients with low playback requirement, under the constraint that the sum of bandwidth does not exceed the network bandwidth. Under this mechanism, the playback requirement of each client can be satisfied and the underflow situation can be avoided. At the same time, service for more clients can be provided with the same limited network bandwidth. Compared with the traditional fixed bandwidth allocation approaches, our approach is more efficient in utilizing the network bandwidth.

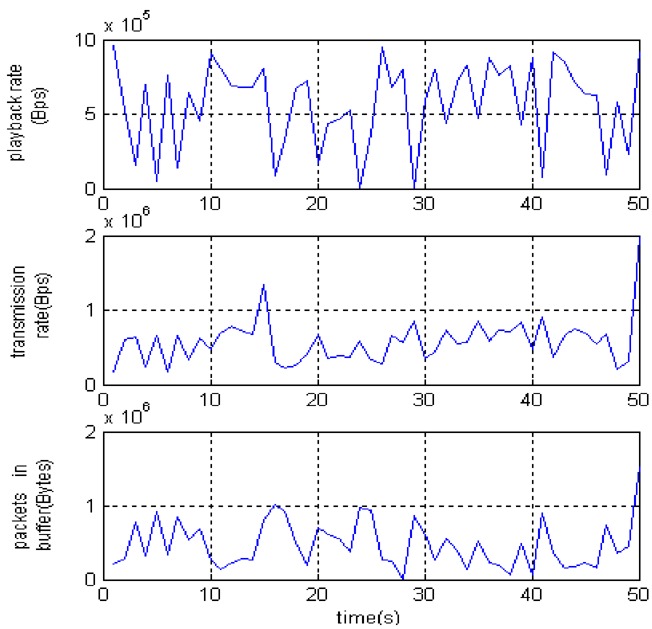

Figure 3. Transmission rate changes with the packet sizes in the client buffer and the playback rates for client 1 (Q_r = 1MB, t = 1 to 50 second, and increment = 1 second).

To illustrate how the rate changes according to the playback rate and buffer packet for every client connection, the changes of the transmission rates, packet sizes in the client buffer, and the playback rates for one single client (i.e., client 1) are shown in Figure 3.

Usually, the transmission rate is determined according to the playback rate requirement and the existing buffer packet. The rate is determined so as to satisfy the playback requirement while at the same time to keep the rate as minimum as possible. For example in Figure 3, at the time instant 16th second, since the client buffer is full and the existing buffer packets can provide for the playback requirement, the transmission rate can be adjusted at a low level to save the bandwidth. Also, the playback rate begins to increase and keeps at a high level starting from the time instant 16th second, and the transmission rate increases when the buffer packets cannot provide enough packets for the playback requirement. It needs to be noted that at the current stage, the proposed mechanism does not put any constraint on the final time instant in our optimization function. Therefore, it is possible that overflow occurs at the final time instant.

From the simulation results, it is shown that our proposed self-adjusted network transmission mechanism can avoid the loss of packets, and at the same time achieve the minimal bandwidth allocation and maximal utilization of the client buffer. The advantage of the proposed mechanism is that the transmission rate is dynamically adjusted according to the playback rates of all the clients. The transmission rate is increased automatically when the playback rate is high. Consequently, the client buffer can be fully utilized under the limited bandwidth to provide jitter free playback.

According to the performance index function we try to minimize, when the minimal index function for each client is achieved, the minimal index function at the server side is obtained, so that the bandwidth utilization is optimized. Therefore, more connections can be admitted under the same network bandwidth.

4. Conclusions

When there are multiple clients requesting data from a multimedia server simultaneously, how to efficiently allocate the bandwidth to each client and satisfy the QoS requirements of the client applications is a challenging task. In this paper, we propose a self-adjusted network transmission mechanism that can obtain the maximum utilization of the network resource. Instead of giving a fixed transmission rate for each client, the transmission rate for each client is determined dynamically to achieve the maximal utilization of client buffer and the minimal allocation of bandwidth. The optimized transmission rate can be adjusted to satisfy the constraint of the network bandwidth. With the limited bandwidth resource of the network, QoS requirements for multiple clients can be met using the proposed mechanism. In addition, since lots of the parameter computation can be done off-line, the proposed mechanism is also practical and efficient.

5. References

[1] S.-C. Chen, M.-L. Shyu, and R.L. Kashyap, "Augmented Transition Network as a Semantic Model for Video Data," *International Journal of Networking and Information Systems*, Special Issue on Video Data, Vol. 3, No. 1, pp. 9-25, 2000.

[2] S.-C. Chen, M.-L. Shyu, C. Zhang, and R.L. Kashyap, "Object Tracking and Augmented Transition Network for Video Indexing and Modeling," the *12th IEEE International Conference on Tools with Artificial*

Intelligence (ICTAI 2000), Vancouver, British Columbia, Canada, pp. 428-435, November 13-15, 2000,

[3] S.-C. Chen, S. Sista, M.-L. Shyu, and R.L. Kashyap, "Augmented Transition Networks as Video Browsing Models for Multimedia Databases and Multimedia Information Systems," *the 11th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'99)*, Chicago, IL, USA, pp. 175-182, November 9-11, 1999.

[4] W. Effelsberg, O. Spaniol, A. Danthine, and D. Ferrari, editors. *High-speed Networking for Multimedia Applications*. Kluwer Academic Publishers, 1996.

[5] J. Y. Hui, "Resource Allocation for Broadband Networks," *IEEE Journal on Selected Areas in Communication*, pp. 1598-1608, Vol. 6, No. 9, 1988.

[6] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and a Survey," *Computer Networks & ISDN Sys.*, Vol. 28, No. 13, pp. 1723-1738, 1996.

[7] H. Kanakia, P. Mishra, and A. Reibman, "Adaptive Congestion Control Scheme for Real-time Packet Video Transport," *SIGCOMM Symposium on Communications Architecture and Protocols*, San Francisco, California, pp. 20-31, ACM/IEEE, Sept. 1993.

[8] F. Lewis, and L. Syrmos, *Optimal Control*. John Wiley & Sons, INC., 1995.

[9] M. Garrett, and W. Willinger, "Analysis, Modeling and Generation of Self-similar VBR Video Traffic," *ACM SIGCOMM*, pp. 269-280, August 1994.

[10] S. Ramanathan, and P. V. Rangan, "Adaptive Feedback Techniques for Synchronized Multimedia Retrieval over Integrated Networks," *IEEE/ACM Transactions on Networking*, Vol. 1, No. 2, pp. 246-260, April 1993.

[11] J. Roberts, U. Mocci, and J. Virtamo, editors. *Broadband Network Teletraffic, Final Report of Action Cost 242*, Springer, 1996.

[12] M.-L. Shyu, S.-C. Chen, and H. Luo, "Optimal Resource Utilization in Multimedia Transmission," *IEEE International Conference on Multimedia and Expo (ICME)*, Waseda University, Tokyo, Japan, pp. 880-883, August 22-25, 2001.

[13] R. Steinmetz, "Human Perception of Jitter and Media Synchronization," *IEEE Journal on Selected Areas in Communications*, Vol. 14, pp. 61-72, January 1996.

[14] M. Tatipamula, and B. Khasnabish, editors. *Multimedia Communications Networks Technologies and Services*. Artech House, pp. 133-173, 1998.

[15] X. Wang, and H. Schulzrinne, "Comparison of Adaptive Internet Multimedia Applications," *IEICE Trans. Commun.*, Vol. E82-B, No. 6, pp. 806-818, June 1999.

[16] M. Woo, N. U. Qazi, and A. Ghafoor, "A Synchronization Framework for Communication of Pre-Orchestrated Multimedia Information," *IEEE Network*, Vol. 8, No. 1, pp. 52-61, January/February, 1994.