HTML Paper

OPTIMAL RESOURCE UTILIZATION IN MULTIMEDIA TRANSMISSION

Mei-Ling Shyu[1], *Shu-Ching Chen*[2], *Hongli Luo*[1]

[1]Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33124, USA
[2]Distributed Multimedia Information System Laboratory
School of Computer Science, Florida International University, Miami, FL 33199, USA

ABSTRACT

Efficient utilization of network resources is essential in the provision of quality multimedia services. Different approaches have been proposed to shape the multimedia streams as a transmission schedule with smoothed traffic burst. In this paper, we concentrate on the problem of efficient use of bandwidth and client buffer. For this purpose, an optimal transmission schedule is proposed to provide the minimal allocation of the bandwidth and maximal utilization of the client buffer. Simulation results show that the shaping results obtained can dynamically adjust the transmission rate according to the buffer packet sizes and playback rates at the client to avoid the loss of packets, and at the same time achieve the minimal bandwidth allocation and maximal utilization of the client buffer.

1. INTRODUCTION

The growing demand for distributed multimedia applications and the development of high-speed network has posed new challenges in networking. Efficient utilization of network resources is essential in the provision of cost-effective multimedia services where quality-of-service (QoS) [5][12] requirements for different network applications are met. Some dominating parameters for QoS are reliability, bandwidth, and jitter, to name a few. Multimedia information in general has a highly time-varying bandwidth requirement [11][13]. Media synchronization is also needed to guarantee QoS [1][9]. To achieve proper synchronization without prefetching a large part of the multimedia data, the reserved network resources must be sufficient to meet the presentation's peak throughput requirement. Static resource reservation schemes are based on fixed resource allocation at the connection stage. With large variations in bandwidth requirements, static allocation usually results in considerable wastage of network resources.

In this paper, we consider the end-to-end transmission of media streams. While a user request is presented, media data are first retrieved from the storage subsystem and then transmitted from the server to the

client continuously. On the client side, incoming data are temporarily stored in the client buffer and consumed. If the buffer is full when a packet arrives, this packet will get lost. We try to maximize the utilization of buffer and at the same time avoid sending too much data to overflow the client buffer because of the fixed client buffer size. 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, i.e. buffer overflow and underflow.

Normally the conventional network service will use a constant transmission rate that is determined by the peak data rate. Hence, it is a waste of bandwidth because of the varied playback throughput at the client. Much work has been done in the resource allocation and traffic control [6][7] and multimedia synchronization in transmission in multimedia network [9][10][14]. The provision of optimal networks from the point of view of reliability and efficiency is still a challenge.

In this paper, a transmission schedule that dynamically changes the transmission rate is designed, which can minimize the peak bandwidth, maximize the packet size stored in buffer at every time instant to keep the network utilization as large as possible. A network model is introduced to capture the relation of transmission rates, buffer packets, and playback rates.

The organization of this paper is as follows. Section 2 describes the network model we proposed and the optimal transmission rate schedule. Simulation results are given in Section 3. Conclusions and future work are presented in Section 4.

2. NETWORK FRAMEWORK

The network model we proposed is given in Figure 1. As can be seen from the figure, $R(t)$ is the transmission rate of the multimedia server, $R'(t)$ is the incoming packet rate at the client buffer. We assume that the packets transmitted from the server at a time instant are equal to the packets received at the client. That is, the propagation delay is ignored in this paper.
Therefore, we assume

$$R'(t) = R(t) \tag{1}$$

and $R(t)$ is used to represent $R'(t)$ in this paper.

Let Q_r be the allocated client buffer size at the setup of the connection, $Q(t)$ be the size of the packet in the client buffer, $L(t)$ be the playback rate, and $e(t)$ be the difference between the Q_r and $Q(t)$. In order to maximize buffer utilization, $Q(t)$ should be close to Q_r. So we try to minimize $e(t)$ (the difference between $Q(t)$ and Q_r) and at the same time, to minimize the bandwidth requirements (i.e., to minimize the transmission rate $R(t)$). The difference between the packet size in buffer and the allocated buffer size ($e(t)$) can be used as the feedback to control the transmission rate at the server side.

0-7695-1198-8/01/$10.00 (C) 2001 IEEE

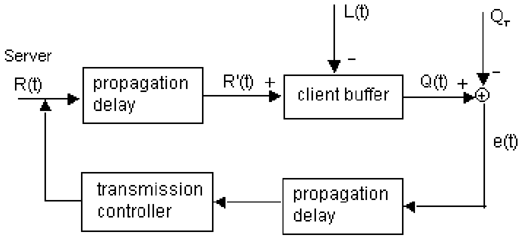

Figure 1. The Proposed Network Model

In our previous study, a *Multimedia Augmented Transition Network (MATN)* model [2][3][4] was proposed to obtain the playback schedule for multimedia presentations. 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. Therefore, *L(t)* is known since the playback schedule is modeled by the *MATN* model.

Instead of transmitting the packets in a fixed rate, transmission rate can be changed automatically according to the buffer packets and playback rates. 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.

2.1 LQ Tracker with Output Feedback

The LQ tracker [8] is applied to the system model given below, where *X(t)* is the status variable, *y(t)* is the output of the model, *u(t)* is the input of the model, *W(t)* is the known disturbance, and *A(t)*, *B(t)*, and *C(t)* are the time-variant matrices.

$$\dot{X}(t) = A(t)x(t) + B(t)u(t) + W(t) \quad (2)$$

$$y(t) = C(t)x(t) \quad (3)$$

We define the output difference to be $e(t) = y_r(t) - y(t)$, where $y_r(t)$ is the expected output.

Let the quadratic performance index function J of the system be

$$J = \frac{1}{2}e^T(t_f)Fe(t_f) + \frac{1}{2}\int_0^{t_f}[\mathrm{e}^T(t)G(t)e(t) + u^T(t)H(t)u(t)]dt$$

where G*(t)*>0, *F*>0, and H*(t)*>=0.

The optimal input is

$$u^*(t) = R^{-1}(t)B^T(t)[P(t) - K(t)x(t)] \quad (4)$$

where *K(t)* is the feedback gain that satisfies the Riccati equation [8] and the boundary condition given below.

$$\dot{P}(t) = -[A(t) - B(t)R^{-1}(t)B^T(t)K(t)]^T P(t) - C^T(t)H(t)y_r(t) + K(t)W(t) \quad (5)$$

$$P(t_f) = C^T(t_f)H(t_t)y_r(t_t) \quad (6)$$

Solving these equations, the optimal input *u*(t)* that satisfies the minimal performance index function *J* can be obtained.

2.2 Network Model

Consider the relationship among *Q(t)*, *R(t)*, and *L(t)* (defined earlier). Since the change rate of the buffer packets is the difference between the transmission rate and the playback rate, based on Equations (2) and (3), let *A(t)*=0, *B(t)*=1, *C(t)*=1, the transmission schedule model is defined as

$$\dot{Q}(t) = R(t) - L(t) \quad (7)$$

Let *Q(t)* be the variable we want to track, and *e(t)* be the difference between the tracked *Q(t)* and the allocated Q_r. We have

$$e(t) = Q_r - Q(t) \quad (8)$$

In order to make fully use of the client buffer to avoid the jitter, *e(t)* should be minimized. On the other hand, the transmission rate *R(t)* should be minimized to save the bandwidth. Let *F(t)*=0, *G(t)*=1, and *H(t)*=1, the performance index *J* that needs to be minimized is defined as

$$J = \frac{1}{2}\int_0^{t_f}[e^2(t) + R^2(t)]dt \quad (9)$$

Given the Riccati equation and the boundary condition below,

$$\dot{P}(t) = P(t) - Q_r + L(t) \quad (10)$$

$$P(t_f) = Q_r \quad (11)$$

we have

$$P(t) = Q_r - L(t) + \frac{e^t}{e^{t_f}}L(t) \quad (12)$$

According to the corresponding equation (Equation (4)), we can get the optimal transmission rate R*(t) as follows.

$$R^*(t) = Q_r + L(t) - \frac{e^t L(t)}{e^{t_f}} - Q(t) \quad (13)$$

Under this optimal transmission rate, the packets in client buffer will be

$$Q(t) = Q_r - \frac{1}{2}L(t)e^{(t-t_f)} + \frac{1}{2}e^{-t}(-2Qre^{t_f} + L(t))/e^{t_f} \quad (14)$$

0-7695-1198-8/01/$10.00 (C) 2001 IEEE

In our approach, in addition to minimize the performance index function J (given in Equation (9)), R_r, the maximal transmission rate due to the network bandwidth limitation, needs to be considered. When the optimal transmission rate obtained from Equation (13) is larger than R_r, R_r will be used as the transmission rate and the remaining packets in the buffer will be used for the playback to compensate the upper bound of the transmission rate. Under this design, we can optimize the network resource of the bandwidth and the client buffer.

3. SIMULATION RESULTS

In the simulations, different playback rates are examined to evaluate the effectiveness of the proposed model. The playback rate is generated randomly between 0.1MB per second (MBps) and 1MBps to simulate the actual playback scenario. Assume the allocated client buffer size is 1MB and it is run within the interval [1,200] seconds with the time increment of 1 second. The result is depicted in Figure 2 that shows the change of the transmission rates, packet sizes in the client buffer, and the playback rates. The maximal transmission rate (R_r) used in the simulation is 0.6MBps.

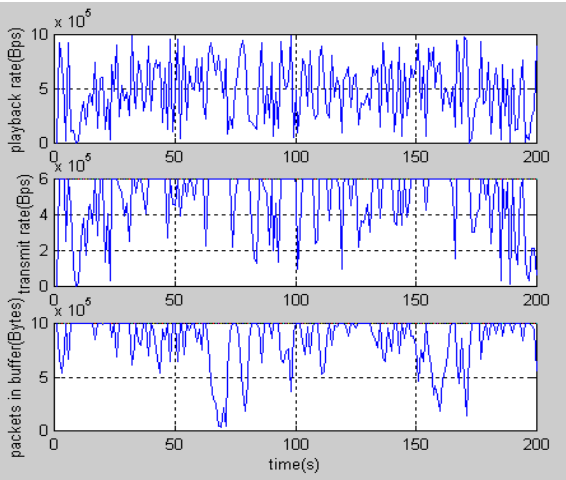

Figure 2. Transmission rate changes with the packet size in client buffer and playback rate (Q_r = 1MB, t = 1 to 200 second, and increment = 1 second).

As can be seen from the figure, in general, most of the times the size of the packets in the buffer is close to the allocated buffer size. This meets our goal to maximally utilize the client buffer as the reservoir for playback. When the playback rate is high and a high transmission rate is needed, instead of allocating the optimal $R^*(t)$ obtained from the optimal method in Equation (13), only the maximal transmission rate R_r can be used. If R_r cannot satisfy the playback rate $L(t)$, 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 ($Q(t)$) increases. By using the proposed network framework, later optimized transmission rates will be automatically obtained to alleviate this situation to allow fully utilization of the client buffer. Therefore, our proposed framework can avoid the loss of packets, and at the same time achieve the minimal bandwidth allocation and maximal utilization of the client buffer.

For simplicity, the simulation result in the interval [100,110] second with the time increment 0.2 second in Figure 3 is used to illustrate the effectiveness of our proposed framework. For example, at the time instant 101.8th second, the playback rate is very small, so the optimal transmission rate is used and the client buffer becomes full. At the next time instant (102th second), the playback rate increases tremendously, the obtained optimized transmission rate is larger than the maximal transmission rate (R_r) so that R_r is used, the packets in buffer is decreasing. Between 102th and 103th second, the playback rates remain at a high level, the obtained optimized transmission rates are larger than R_r. Therefore R_r is used to let the size of the packets in the buffer increase. The same explanation applies for other time duration.

Compared with the conventional methods, a precise and mathematically sound model is introduced and an optimal algorithm based on the model is proposed. The advantage of our method is that the transmission rate of the server is dynamically adjusted according to the application requirement of the client. The transmission rate is increased automatically when the playback rate is high. Consequently, the client buffer can be fully utilized under limited bandwidth requirement to provide jitter free playback. In our approach, the packet size in the buffer will never exceed the allocated buffer size, thus avoiding the overflow in client buffer and the loss of packets.

According to the performance index function we try to minimize, the transmission rate in the long range is the minimal, so more connections can be admitted under the same bandwidth capacity. Since our model considers the packet flow from the transmission to playback as a close-loop process, thus overflow and/or underflow situations can be solved immediately to provide effective and reliable transmission.

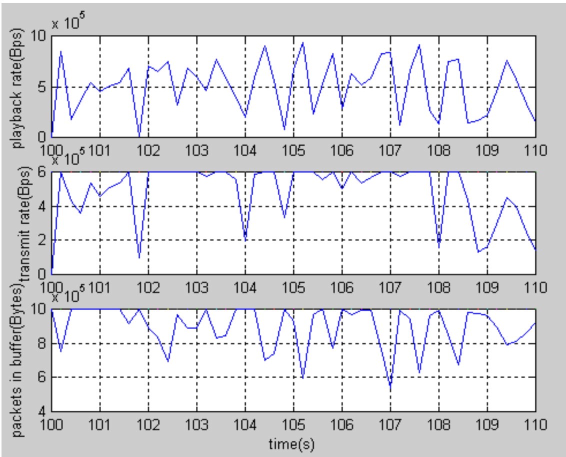

Figure 3. Transmission rate changes with the packet size in client buffer and playback rate (Q_r = 1MB, t = 100 to 110 second, and increment = 0.2 second).

0-7695-1198-8/01/$10.00 (C) 2001 IEEE

4. CONCLUSIONS

In this paper, a traffic shaping scheme is introduced to determine the suitable transmission rate of the server. Instead of giving a fixed transmission rate, our approach determines the rate dynamically to achieve the optimal utilization of network resources. LQ tracker with output feedback is used in our proposed framework to achieve the maximal utilization of client buffer and the minimal allocation of bandwidth. The simulation results show that our proposed framework can avoid the loss of packets due to overflow at the client buffer. Hence, our proposed framework is shown to be practical, efficient, and reliable to support continuous media transmission.

There are several issues to be considered for our future research. First, the network transmission delay should be considered. Second, when underflow occurs, the playback schedule should be delayed until the packets arrive in order to provide a jitter-free playback.

5. REFERENCES

[1] Blakowski G. and Steinmetz R., "A Media Synchronization Survey: Reference Model, Specification, and Case Studies," *IEEE Journal on Selected Areas in Communications*, Vol. 14, No. 1, January 1996, pp. 5-35.

[2] Chen S-C., Shyu M-L., and Kashyap R.L., "Augmented Transition Network as a Semantic Model for Video Data," accepted for publication, *International Journal of Networking and Information Systems*, Special Issue on Video Data.

[3] Chen S-C., Shyu, M-L., Zhang C., and Kashyap R.L., "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, November 13-15, 2000, pp. 428-435.

[4] Chen S-C., Sista S., Shyu M-L., and Kashyap R.L., "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, November 9-11, 1999, pp. 175-182.

[5] Grosky W., Jain R., and Mehrotra. R., editors. *The Handbook of Multimedia Information Management.* Prentice-Hall Professional Technical Reference, April 1997, pp. 335-363.

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

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

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

[9] Little T.D.C., and Ghafoor A., "Synchronization and Storage Model for Multimedia Objects," *IEEE Journal of Selected Areas in Communications*, Vol. 8, No. 3, April 1990, pp.413-427.

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

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

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

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

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