ENSURING FAIRNESS IN MULTIMEDIA MULTICAST STREAMING WITH OPTIMAL RATE ALLOCATION AND CLIENT BUFFER UTILIZATION

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

In this paper, we propose a rate control mechanism for multimedia multicast streaming in the Internet. A single and optimal transmission rate that is adaptive to the network congestion, buffer occupancies, and the playback requirements of the member clients in the multicast group is allocated. We optimize the multicasting delivery in the Internet by improving the buffer occupancy of all clients for on-time presentation while using the minimal transmission rate. Scalable playback is used for the clients with low bandwidth capacities in a heterogeneous environment. Simulation results show that the buffer capacity can be efficiently improved and fairness is ensured among the member clients.

1. INTRODUCTION

Real-time video transmission over the Internet is a challenging task. When streaming multimedia data over the network, multicast has the advantage of sending only a single copy of the data across the network while preserving the network bandwidth. Multicasting permits the server (sender) to deliver one video stream to a group of clients (receivers) in the network using a single transmission rate. Since the clients have different bandwidth and processing capacities, the allocation of the transmission rate should consider the playback requirements of all member clients, ensure the intra-session fairness among the members of the multicast group [4], and share the bandwidth with dominant TCP flows in the Internet with congestion control.

The server allocates a fixed transmission rate, which results in a mismatch between the fixed transmission rates and the heterogeneous link capacities. The heterogeneity makes rate control in multicast difficult. Layered coding and transmission has been proposed as an efficient solution to the heterogeneity problem in multicast. A video source is encoded into multiple layers of bit stream. The client can subscribe into different layers according to the capacity.

Rate allocation should also work with congestion control in multicast. Multicasting multimedia flows can share the bandwidth with TCP flows fairly if the streams adapt to network congestion in a TCP compatible way. In [2], a tree-based reliable multicast protocol that is similar to other approaches of adapting sliding window flow control to multicast was proposed. [6] introduced a sender-based approach for multicast congestion control for reliable bulk data transfer. Multirate multicasting was proposed in [3], where the receivers in a multicast group receive packets at different rates that were obtained to maximize the total receiver utility for multicast sessions.

The fairness of multicast video service means each client in a multicast group should receive video data at a rate commensurate with its processing capacity and its link capacity regardless of the capabilities of the other clients [1]. Hence, a lower capacity client will obtain a lower presentation quality, without sacrificing the presentation quality of the other clients with higher capacities. For this purpose, we propose a rate allocation mechanism for multicast in this paper, which tries to obtain a minimal sending rate that can maximize the expected buffer occupancy when averaged over all clients. It is designed for the current best-effort Internet transmission, since it is responsive to the network congestion.

The paper is organized as follows. In Section 2, our proposed multicast approach is presented. Simulations are provided in Section 3. Section 4 gives the conclusions.

2. PROPOSED MULTICAST APPROACH

Typically, the member clients in a multicast group are heterogeneous. If one client has a far lower capacity than the others, it is suitable for this client to have a reduced presentation quality and let the rest of the group achieve higher throughputs. Our approach addresses such a client bandwidth heterogeneity problem with adaptive playback. In our proposed approach, any client that cannot sustain the calculated optimal transmission rate needs to reduce the playback rate according to the packet arrival percent. For the lower capacity client, the playback rate changes with the arriving packets.

Because the packets will experience varying network delays and congestion before they arrive at different clients, each client will receive different numbers of packets at the same time interval. The multicast group can be viewed as a system with one input (the single transmission rate) and multiple outputs (different numbers of packets arriving at the clients). We can optimize the multicast delivery in terms of the whole multicast system, which tries to use a single transmission rate to optimize the presentation quality for all the clients.

Multimedia stream contains delay-sensitive data. Playback jitter may occur if the packets arrive with significant delay variations. Client buffer is used to absorb the jitter. The fully utilization of the buffer occupancy can efficiently decrease the playback jitter and the degraded video presentation. Hence, the determination of the transmission rate should be adaptive to the client buffer occupancy, the network delay and congestion. In this manner, the expected playback jitter can be reduced under the same network conditions. The mathematical formulization of our approach is given below.

For the j^{th} client in the multicast group at time interval k, define $q_{k,j}$ as the buffer occupancy, r_k as the packets transmitted from the server, $p_{k,j}$ as the packets arriving at the client buffer, l_k as the packets used for scheduled playback, and s_j as the allocated buffer size at the setup of the connection. Let $q_{k+1,j}$ denote the buffer occupancy at time interval $k+1$. Here, we assume that the values for l_k's are known. Since a single transmission rate is used and the playback schedule is the same for all the clients, r_k and l_k are the same for all the clients in the multicast group. Hence, we have

$$q_{k+1,j} = q_{k,j} + p_{k,j} - l_k \qquad (1)$$

However, because of the changeable network delays, $p_{k,j}$ is not equal to r_k at time interval k. Assume that the packets arriving at the client buffer at time interval k comprise of packets transmitted from the server at the time interval $k-d_j$, …, $k-d_j-i+1$, …and $k-d_j-n_j+1$, and let $b_{i,k,j}$ denote the percentage of the packets transmitted at r_{k-d_j-i+1} that arrives at the client buffer at time interval k to capture the scenario of changing network delays. Hence, $p_{k,i}$ can be represented as

$$p_{k,j} = b_{1,k,j} r_{k-d_j} + \ldots + b_{i,k,j} r_{k-d_j-i+1} + \ldots + b_{n_j,k,j} r_{k-d_j-n_j+1} \qquad (2)$$

Considering m clients in the multicast group as a system, we construct the following vectors to represent the whole multicast group:

$$Q_k = [q_{k,1}, \ldots, q_{k,j}, \ldots, q_{k,m}]^T$$
$$P_k = [p_{k,1}, \ldots, p_{k,j}, \ldots, p_{k,m}]^T$$
$$S = [s_1, \ldots, s_j, \ldots, s_m]^T$$
$$L_k = [l_k, \ldots, l_k, \ldots, l_k]^T$$
$$R_k = [r_k, \ldots, r_k, \ldots, r_k]^T$$

The difference between the buffer occupancy and buffer capacity is a scalar as defined in Equation (3).

$$e_k = W_p Q_{k+d_0} - W_q S \qquad (3)$$

Then the index function we try to minimize is

$$J_k = e_k^2 + (W_r R_k)^2 \qquad (4)$$

where d_0 is the maximal value of d_j, and the weighting coefficient vectors W_p and W_q for the buffer occupancy and the weighting coefficient vector W_r for the sending rates are defined as follows.

$$W_p = [w_{p,1}, \ldots, w_{p,j}, \ldots, w_{p,m}]$$
$$W_q = [w_{q,1}, \ldots, w_{q,j}, \ldots, w_{q,m}]$$
$$W_r = [w_{r,1}, \ldots, w_{r,j}, \ldots, w_{r,m}]$$

Here, $w_{p,j}$, $w_{q,j}$ and $w_{r,j}$ are the corresponding values for the j^{th} client, and in our proposed framework, the values for $w_{p,j}$ and $w_{q,j}$ are 1.

As shown in Equation (4), different optimization performances can be achieved with different combinations of the weighting coefficients. The detailed solution of the optimal sending rate can be seen in [7]. The rate obtained is actually an aggregate optimal rate after gathering all the feedback information from the group members. It takes into account all the buffer occupancies of the clients, and the different network delays and congestion situations those clients experienced.

Congestion control is implemented via adjusting the weighting coefficient vector w_r in Equation (4). If the packet loss rate experienced by the j^{th} client is larger than the threshold value, the $w_{r,j}$ value will be doubled, which will result in a decreased rate. If the packet loss ratio of this client is decreased to below the threshold value, the $w_{r,j}$ value will be increased by 1 in the next adjusting interval, which will result in a gradually increased rate. The $w_{r,j}$ value ranges between a lower bound and an upper bound. In this way, the congested link is reflected in the calculation and used to adjust the sending rate, which produces a multimedia stream adaptive to the network congestion and sharing the bandwidth fairly with other TCP traffic.

In our approach, fairness is referred to as when the clients with higher bandwidth capacities can maintain a higher buffer occupancy, which means more packets can be available for decoding and better perceptual video quality can be achieved. For a client with a relatively low bandwidth capacity, its buffer occupancy will be low if the normal playback schedule is used. When its bandwidth is much below the range of other members in the multicast group, a reduced playback schedule should be considered. The first item of the packet arrival percent ($b_{1,k,j}$) in Equation (2) is adopted as the scalable factor for the j^{th} client at time interval k. The larger the $b_{1,k,j}$, the less congested the current network is and more packets can be expected at the next time interval. Therefore, a better presentation quality can be provided without involving the buffer underflow in the next time interval.

3. SIMULATION RESULTS

We examine the performance of our approach using network simulator NS-2 [5]. In our simulations, we consider one server and four clients in the multicast group, each client with a different link delay from the server. All the clients stay in the multicast group throughout the period. All the queues use the FIFO drop-tail scheduling discipline. The packet size is 1000 bytes for the streaming traffic. The simulation parameters are shown in Table 1.

Table 1. Simulation Parameters

Packet Size	1000Bytes
Bottleneck Bandwidth	2Mbps
w_r lower_bound	4
w_r upper_bound	16
Adjusting Period	1 second
ACK Size	40Bytes
Size of Client Buffer	212500Bytes(=1.7×10^6bits)
Packet_loss_threshold	0.98

Simulations under two different scenarios for clients with heterogeneous link capacities are conducted. First, the link capacities of the clients in different ranges where packet losses occur in some low capacity links are considered. Second, we examine the cases where all the clients have low link capacities and experience different packet loss ratios. For those clients with lower capacities, adaptive playback which the playback rate changes dynamically with the packet arrival percent is adopted. The client generates a feedback report every 1 second. This feedback interval is suitable for multimedia transmission, and the feedback implosion can be efficiently reduced.

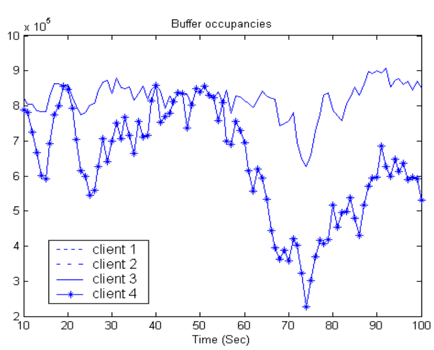

Figure 1. Buffer occupancies for the heterogeneous bandwidth links.

3.1. One Client with Packet Losses

In this simulation, the bandwidths for the four clients are 2.12 Mbps, 2.0Mbps, 2.1Mbps and 0.9Mbps, and the playback rate ranges in [8×10^5,1×10^6] bps. Hence, client 4 has the link capacity that is below the playback requirement and the adaptive playback is adopted. The buffer occupancies are shown in Figure 1. The buffer occupancies for clients 1, 2, and 3 are almost the same, and the buffer occupancy for client 4 is much lower than the others, which is fair to the higher bandwidth links.

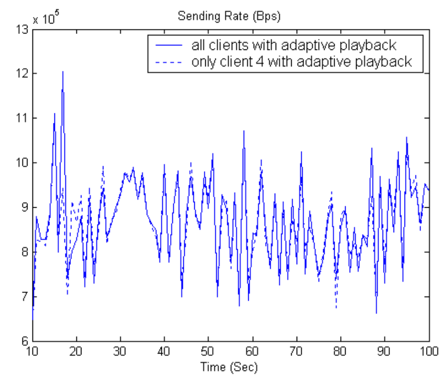

Figure 2. Sending rate comparison for the two cases.

3.2. All Clients with Different Packet Loss Ratios

In this simulation, the link bandwidths are 1.12Mbs for client 1, 1.1Mbs for client 2, 1.0Mbs for client 3 and 0.8Mbs for client 4, and the playback rate ranges in [7×10^5, 1.1×10^6] bps. All the four clients experience packet losses to some degrees. Hence, we further examine the performance under two cases: (1) adaptive playback used only for the client with the lowest capacity (i.e., client 4), and (2) adaptive playback used for all clients.

The sending rates in these two cases are compared in Figure 2, while the buffer occupancies of all clients are displayed in Figure 3 and Figure 4. As can be seen from the figures, the sending rates do not change much in these 2 cases, no overflow occurs in all clients, and the buffer occupancies of the clients are proportional to their link capacities. As can be seen from Figure 3, the buffer occupancies of client 3 are quite close to those of client 1 and client 2 since only client 4 uses adaptive playback. While in Figure 4, buffer occupancies of client 1 and client 2 are still quite close to each other, but client 3 and client 4 have much lower buffer occupancies since their link capacities are below the playback requirement. Hence, our proposed rate mechanism provides better fairness to the group members in the case when all the member clients adopt adaptive playback than in the case when only one client adopts adaptive playback.

In Figure 5, we further examine how the buffer occupancies of client 3 and client 4 change in both cases. Client 1 and client 2 are ignored since their buffer occupancies do not change much in both cases. In Figure 5, when all clients adopt adaptive playback, the buffer occupancies of client 3 are increased, while the buffer occupancies of client 4 are decreased. It is reasonable and fair since client 4 has a lower link capacity than client 3 and therefore has a more degraded quality of playback.

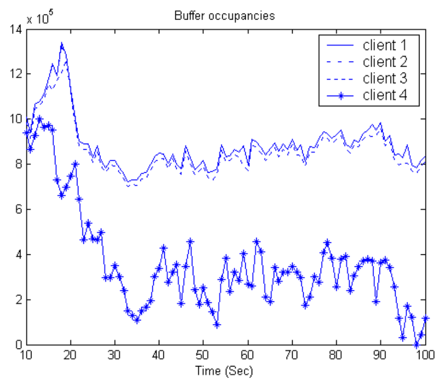

Figure 3. Buffer Occupancies when only client 4 uses adaptive playback.

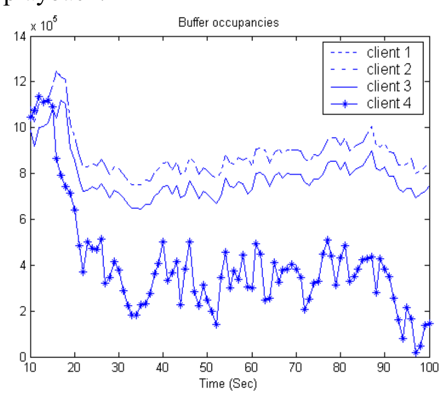

Figure 4. Buffer Occupancies when all clients use adaptive playback.

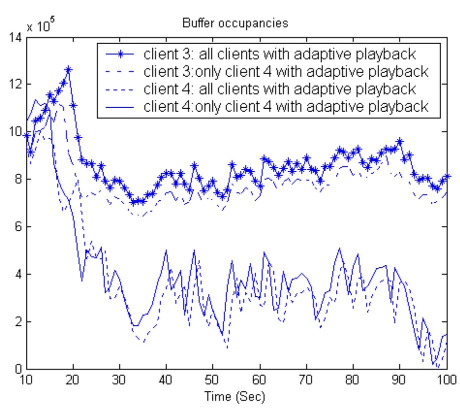

Figure 5. Buffer occupancies for client 3 and client 4 in both cases.

The simulation results demonstrate that our proposed multicast approach can provide better fairness and buffer utilization especially when more low capacities clients that experience different packet drop ratios adopt adaptive playback. In addition, overflows can be effectively avoided in our proposed approach. Also, computation overhead of our proposed approach is low, so the calculation of the optimal transmission rate is highly scalable and can be applied to multicast with large number of members.

4. CONCLUSIONS

In this paper, we propose an optimization criterion for rate allocation in multicast. Considering the member clients in a multicast group as a whole system, our proposed approach finds an optimal sending rate that can maximize the buffer occupancies and minimize the bandwidth allocation for all the clients. The produced multicasting stream is also adaptive to the network congestion and shares the bandwidth fairly with the TCP flows in the best-effort network. For the clients with low capacities, scalable playback is used, which is adaptive to the link capacity and network congestion. Hence, the fairness is achieved among all the clients in the multicast group.

5. ACKNOWLEDGEMENT

This work was supported in part by Telecommunications & Information Technology Institute (IT2)/FIU, IT2 BA01.

6. REFERENCES

[1] S. Cheung, M. H. Ammar, and X. Li, "On the Use of Destination Set Grouping to Improve Fairness in Multicast Video Distribution," *Proceedings of INFOCOM*, San Francisco, CA, pp. 553-560, March 1996.

[2] D. Chiu, M. Kadansky, J. Provino, J. Wesley, H. Bischof and H. Zhu, "A Congestion Control Algorithm for Tree-based Reliable Multicast Protocols," *Proceedings of INFOCOM*, New York, 2002.

[3] K. Kar, S. Sarkar and L. Tassiulas, "Optimization Based Rate Control for Multirate Multicast Sessions," *Proceedings of INFOCOM*, Anchorage, USA, April 2001.

[4] J. Liu, B. Li and Y-Q. Zhang, "A Hybrid Adaptation Protocol for TCP-Friendly Layered Multicast and Its Optimal Rate Allocation," *Proceedings of INFOCOM,* New York, 2002.

[5] S. McCanne and S. Floyd. Ns (network simulator). 1995. http://www-mash.cs.berkelry.edu/ns.

[6] S. Shi and M. Waldvogel, "A Rate-based End-to-End Multicast Congestion Control Protocol," *Proceedings of Fifth IEEE Symposium on Computers and Communications*, July 2000.

[7] M.-L. Shyu, S.-C. Chen, and H. Luo, "Optimal Bandwidth Allocation Scheme with Delay Awareness in Multimedia Transmission," *IEEE International Conference on Multimedia and Expo (ICME2002),* Lausanne, Switzerland, pp. 537-540, August 26-29, 2002.