Fast Convergent Layered Multicast in Fixed and Mobile Networks

Kenichi Kashibuchi†*, Tarik Taleb*‡*, Yoshiaki Nemoto***, and Nei Kato*†

Graduate School of Information Sciences, Tohoku University, Japan

E-mails: †{buchiken, kato}@it.ecei.tohoku.ac.jp, ‡talebtarik@ieee.org, *nemoto@nemoto.ecei.tohoku.ac.jp

Abstract — **In the near future, multimedia streaming services will be popular in both fixed and mobile environments. It is said that layered multicast is an effective method for streaming multimedia data to a large number of receivers with highly different network conditions. However, the conventional layered multicast approach converges slowly to the optimal rate. In this paper, we propose a scheme that enables a new session to promptly converge to the appropriate rate relevant to the network condition, which provides a better quality of service (QoS) to multimedia streaming traffic. The proposed scheme is based on the fact that the layered multicast uses the priority packet dropping. Performance of the proposed scheme is evaluated and compared with the conventional method. The results show that the proposed scheme achieves appropriate bandwidth utilization from the beginning of the session. It is also shown that the proposed scheme is effective in managing handoff in mobile networks, thus achieving better QoS in heterogeneous network environments.**

Keywords: multimedia streaming system, receiver-driven layered multicast, priority dropping, handoff management.

1. Introduction

According with the widespread use of wireless broadband Internet access technologies, such as wireless LAN (WLAN), Worldwide Interoperability for Microwave Access (WiMAX), and 3^{rd}–4^{th} generation cellular systems (3 – 4G), users will desire to receive realtime multimedia streaming services not only in fixed networks but also in mobile networks. For providing these services to many users, multicast will be generally used from the standpoint of network utilization efficiency. In general IP multicast, since the server streams at only a single rate, multimedia data are transmitted with same contents and same quality (bit rate) for all users. However, in heterogeneous network environments, each user may have very different available bandwidth. For instance, a user A can access the Internet at several tens of Mbps whereas another user B may access at only a few Mbps. In order to deliver data to as many users as possible, streaming rate should take into account the available bandwidth of users. In this case, streaming rate should be adjusted to user B, and user A is to receive data at low quality.

In order to deal with such network diversity, the cumulative layered multicast approach has been proposed [1]. In layered multicast, streaming servers encode the multimedia data in a base layer and several other enhancement layers. They transmit each layer on a different multicast group. In case of receiver-driven congestion control, the receiver assesses whether to add or drop layers (*i.e.* to join or leave multicast group) according to the network condition. In layered encoding, the more layers users receive and decode with, the higher quality they can get. However, since the enhancement layers depend on lower layers, the user cannot decode data without lower layers. If packets of some lower layers are dropped in the network, packets from higher layer become useless. Therefore, since lower layers provide more important data and should be given higher priority, routers should be equipped with a priority-based packet dropping policy [2], [3]. Among receiver-driven layered multicast approaches, Receiver-driven Layered Multicast (RLM) is a notable example [4]. However, in the beginning of the session, RLM user just joins the base layer and then joins the higher enhancement layers one-by-one. RLM requires a long time until an appropriate layer is found.

Moreover, when providing multimedia streaming services for mobile users, users' mobility should be considered. Indeed, when a mobile node performs handoff between two base stations, if the two cells have different available bandwidth, the user cannot receive the stream at a rate suitable for the new cell immediately after the handoff. This bandwidth disparity can be due to the difference in traffic load in both wireless cells, or use of different wireless access techniques with different link speeds. As a result, it brings network congestion or a waste of resources in the new network.

To solve these issues, we utilize the fact that layered multicast generally uses priority dropping policy, and propose a scheme that quickly converges to the appropriate rate to the network condition at the beginning of the session. The mechanism of the proposed scheme is extended further to mobile networks by taking into account the mobility of users.

The remainder of this paper is structured as follows. Section 2 surveys related works on multimedia streaming techniques in heterogeneous environment. The mechanism of the proposed scheme is described in section 3. Section 4 reports the results of the performance evaluation. Following this, the paper concludes in section 5.

2. Related work

To tackle issues related to network diversity, we can envision the two types of multicast streaming approaches: replicated streaming and layered streaming. In the former [5], which is also called as simulcast approach, a server provides multi-

  September 17–20, 2006

ple streams carrying the same content with different qualities (different bit rates). Each stream is multicast on a different multicast group. Each user joins the multicast group that best satisfies its needs. In the latter, the server divides its data into layers: a base layer and several other enhancement layers. It transmits each layer to a different multicast group. The base layer can be independently decoded and provides the basic-level quality. The enhancement layers can be decoded together with the base layer. Users join the base layer and join as many enhancement layers as the network condition permits. In general, the layered multicast approach is more advantageous than the simulcast approach in bandwidth utilization, especially in highly heterogeneous networks.

Among layered multicast approaches, RLM and packet Pair receiver-driven cumulative Layered Multicast (PLM) [6] are notable examples. In RLM, a server transmits data packets of each layer using different multicast group addresses and users join as many layers as their available bandwidths allow. If a user has sufficient bandwidth, he/she joins the next enhancement layer. In contrast, if he/she detects packet losses due to network congestions, he/she quits receiving the highest enhancement layer. This procedure is called "join-experiments". At the beginning of a session, RLM assesses its available bandwidth in a slow start-like fashion. That is, users converge to the optimal rate by subscribing to next enhancement layers incrementally until congestion occurs. Therefore, it needs a long convergence time. Furthermore, joining higher layers may yield packet losses and may degrade the streaming quality.

To rectify this issue, PLM has been proposed. It estimates the available bandwidth based on a packet pair mechanism without join-experiments. All packets on all layers are transmitted in pairs. At the beginning of a session, users simply join the base layer. Users then estimate their available bandwidth C seconds ($C = 1$ in [6]) after receiving the first packet pair and check whether they should add or drop layers. PLM can converge to the optimal rate faster than RLM, without inducing losses. However, PLM requires that all routers in the network should implement a fair scheduler. Moreover, as the bandwidth estimation is based on an accurate estimation of propagation delays, PLM may fail in the bandwidth estimation in case the network load becomes heavy.

On the other hand, in mobile networks, users freely perform handoff upon changing their point of attachment to the network. To guarantee smooth handoff, several approaches have been proposed. In [7] multiple paths are established between the server and a mobile node during handoff. Admittedly this scheme provides smooth handoff for streaming media. Nevertheless it uses multiple paths during the time a mobile node exists in the cell overlapping area. As in the case of a handoff between 3G and WLAN, if the distance of the cell overlapping area is long, it is unacceptable to use multiple paths for a long time as it causes redundant transmissions of important data.

3. Proposed scheme

3.1. Preconditions

The proposed scheme is based on cumulative layered streaming approach. All routers are assumed to be multicast-capable and to support some priority disciplines.

Input multimedia data are encoded into a set of n cumulative layers $(L_1, ..., L_n)$. All subsets $\{L_1, ..., L_i\}_{i \le n}$ provide the same content. The provided quality increases as index i increases. Since the lowest layer contains the most important data, packets from the lowest layer are assigned higher priority. Indeed, the base layer L_1 has the highest priority (the lowest drop probability) among all layers. Higher layers have lower priority as index i increases. We assign these priorities at the streaming servers using the class field in the IPv6 packet header.

In the proposed scheme, two message packets are defined: "low priority join" and "normal join". The former is transmitted by users to local multicast routers in order to join a session or conduct join-experiments. Upon receiving a low priority join message, multicast routers subsequently decrease the priority of the forwarded streaming packets by one and forward them. In other words, the priority of layer L_i (P_i) is decremented to $(P_i - 1)$, not P_{i-1}. These streams with lower priority do not affect the network because their packets are discarded first when a downstream router is congested. A normal join message is transmitted by users to local multicast routers in order to receive packets with normal priority P_i. After receiving a normal join message, multicast routers forward packets without changing their priority.

All multicast routers maintain an internal table with information on source address, multicast group address, outgoing interface identifier, and priority level. In the table, the priority level field can be set to either "N" or "L". L indicates that the priority of streaming packets is decreased by one by this multicast router, while N indicates that this router forwards packets without changing their priority.

Additionally, when we apply our scheme to mobile users, it is assumed that wireless cells overlap with each other. To access two networks in parallel, a mobile node needs to be simply equipped with two wireless interfaces. Along with a further integration of wireless technologies, it will become normal for a mobile node to have an interface that can simultaneously access different types of wireless networks.

3.2. Mechanism of the proposed scheme

As previously mentioned, a server S encodes multimedia data into a set of n cumulative layers $(L_1, ..., L_n)$. Layer L_i is multicast to group G_i with priority P_i. Lower layers are transmitted with higher priorities (*i.e.* $P_i > P_{i+1}$).

Figure 1 shows the overview of the proposed scheme. Here, it is supposed that receiver R_1 has been joining the session from the server S. When the new receiver R_2 joins the same session, R_2 transmits "low priority join" messages

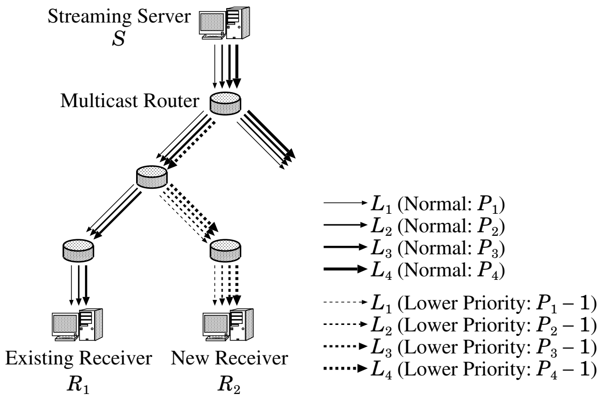


Figure 1: Approach of the proposed scheme.

$\{(S, G_i), i = 1, ..., n\}$ to the local multicast router. Upon receiving a low priority join message (S, G_i), multicast routers first check whether the stream corresponding to L_i is flowing or not (*i.e.* whether the entry $(S, G_i, *, N)$ exists in the internal table). Asterisk$(*)$ denotes unspecified interface. If a multicast router does not have the entry $(S, G_i, *, N)$, it adds the entry (S, G_i, IF, N) to its table and forwards the low priority join message (S, G_i) to its upper multicast router. IF indicates the interface identifier that received the low priority join message. On the contrary, if the multicast router has the entry $(S, G_i, *, N)$, it adds the entry (S, G_i, IF, L) to the table and subsequently decreases the priority of the forwarded packets from P_i to $(P_i - 1)$. It then transmits packets to the requesting users or lower multicast routers. In this manner, multicast routers forward the low priority join message (S, G_i) to upper routers until the expected stream is found.

After receiving the first packet of group G_i, which is transmitted with the lower priority $(P_i - 1)$, R_2 calculates the packet loss rate ρ_i experienced by packets of layer L_i for an interval of time T_D (detection timer). If ρ_k, the packet loss rate of L_k, is smaller than a predefined threshold θ, users transmit a normal join message (S, G_k) to the local multicast router. The setting of θ indicates the system tolerance level in terms of packet drops and depends on Forward Error Correction (FEC) redundancy. If ρ_k is above θ, R_2 stops receiving packets of layer L_k and higher layers by leaving multicast groups $\{G_i, i = k, ..., n\}$.

Upon receiving a normal join message (S, G_k), a multicast router verifies whether it has the entry (S, G_k, IF, L) or not. In case it does not have the entry (S, G_k, IF, L), it forwards the normal join message (S, G_k) to the upper multicast router. On the contrary, when it has the entry (S, G_k, IF, L), it modifies the entry to (S, G_k, IF, N) to indicate that it is forwarding packets with normal priority P_k.

After a given period of time T_J (join timer), R_2 sends low priority join message (S, G_k) to perform the join-experiment operation with lower priority. After receiving the first packet of layer L_k, R_2 calculates the packet loss rate ρ_k on the layer L_k for an interval of time T_D, as in the beginning of the session. When ρ_k is smaller than the threshold θ, R_2 transmits a normal join message (S, G_k) to the local multicast router and joins L_k with normal priority. In case ρ_k exceeds θ, R_2 interprets this join-experiment as failure and leaves layer L_k. R_2 then multiplicatively increases the join timer T_J as follows:

$$T_J \leftarrow \alpha \cdot T_J \tag{1}$$

where α is a constant. After that, every time T_J elapses, R_2 performs a join-experiment in the same way. Besides, during the session, if the packet loss rate on L_k which is not the layer added by join-experiments exceeds θ, users drop L_k and the higher layer, because that indicates a congestion occurrence.

3.3. Extension of the proposed mechanism to handoff

In the remainder of this section, we consider applying the proposed scheme to mobile networks. This application is similar in spirit to the idea presented in [8]. Figure 2 illustrates the idea. A mobile node (MN) instantly measures radio strength or link quality. Prior to handoff, MN receives data via base station BS_1 using wireless interface IF_1. When MN enters into the new cell of base station BS_2, a new network address is given to the MN's wireless interface IF_2. When radio strength or link quality through IF_1 goes down below a predefined threshold, MN transmits low priority join messages $\{(S, G_i),\ i = 1, ..., n\}$ to the local multicast router through IF_2. MN then receives all layers L_i with lower priority $(P_i - 1)$ through IF_2. After receiving the first packet of L_i, MN calculates ρ_i, the packet loss rate of L_i, for a time T_D. If ρ_k exceeds a predefined threshold θ, MN stops receiving the layer L_k and higher layers. Besides, it transmits normal join messages $\{(S, G_i), i = 1, ..., k-1\}$ to the local multicast router, in order to join the lower layer with normal priority. After receiving the stream with normal priority, MN leaves the multicast groups through IF_1. Henceforth, it conducts join-experiment with lower priority through IF_2 every T_J period of time.

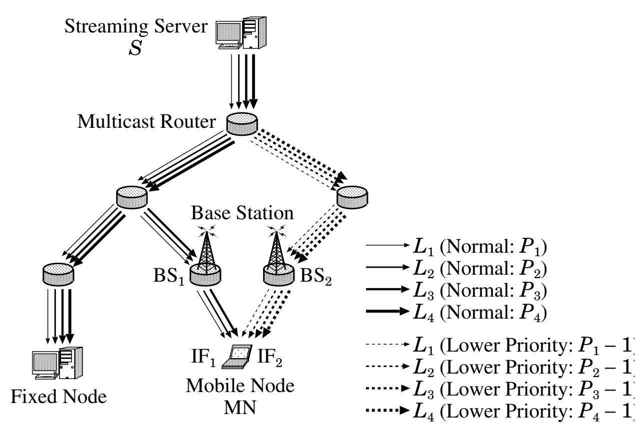


Figure 2: Application of the proposed approach for handoff.

4. Performance evaluation

We carried out several simulations with the Network Simulator (ns-2.28) [9] and compared the performance of our scheme against RLM. The cumulative bit rates are 128 kbps, 256 kbps, 512 kbps, 1 Mbps, 2 Mbps, 4 Mbps, 8 Mbps, 16 Mbps, 32 Mbps, and 64Mbps. They are exponentially spaced to simulate highly heterogeneous environments.

In this paper, Weighted Random Early Detection (WRED) [10] is used for queuing. WRED selectively discards lower priority traffic when a router begins to get congested. It provides also differentiated performance characteristics for different classes of service. Here, it should be emphasized that all routers in the network do not have to set the same parameter. On the other hand, RLM uses drop tail queue [4]. The maximum size of queues is set to 20 packets.

In the proposed scheme, we set both T_D and the initial value of T_J to one second. The multiplicative coefficient α, used in the computation of T_J (equation (1)), is set to two. RLM parameters are set to the same values as in [4]. The used multicast routing protocol is Distance Vector Multicast Routing Protocol (DVMRP). Both wired and wireless links are error-free throughout this paper.

Since the users cannot decode the higher layer without receiving all the lower layers in layered multicast, we use goodput to evaluate the performance of both schemes. Here, goodput is defined as throughput of the layers below the highest layer that users can sequentially achieve with less than $\theta = 10$ [%] of packet loss rate. This metric indicates the number of bytes received and actually decoded by users.

4.1. Fixed network

We first conducted the simulation with a simple network topology, as shown in figure 3. We consider a scenario where receivers R_1, R_2, R_3, R_4, and R_5 join a session provided by server S at time $t = 100$ [sec], $t = 200$ [sec], $t = 300$ [sec], $t = 400$ [sec], and $t = 500$ [sec], respectively. Figure 4 shows the goodput transition of each user. In this topology, R_1, R_2, R_3, R_4, and R_5 can use bandwidth up to 100 Mbps, 50 Mbps, 20 Mbps, 10 Mbps, and 5 Mbps, respectively. Figure 4 indicates that each user succeeded in receiving data at a rate appropriate to its available bandwidth in both schemes.

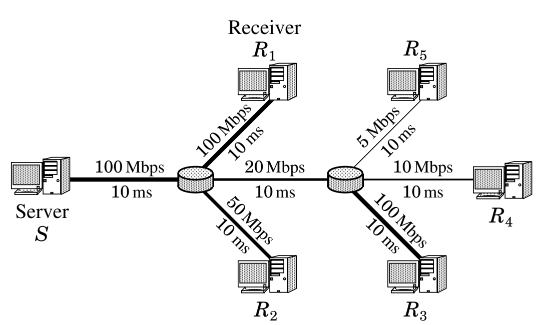
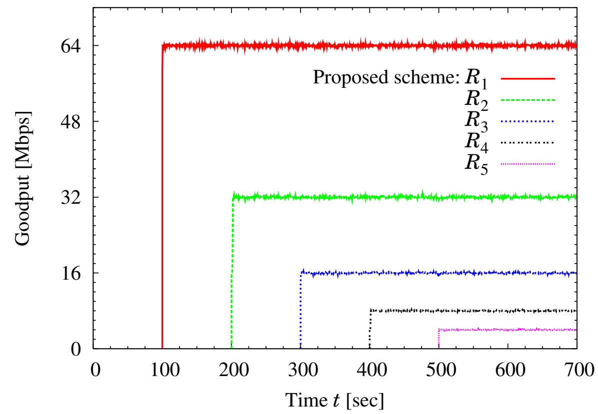
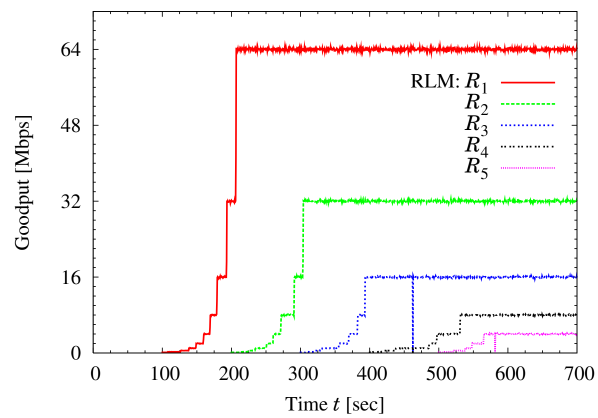


Figure 3: A fixed network topology.



(a) Proposed scheme



(b) RLM

Figure 4: R_1, R_2, R_3, R_4, and R_5 join a session at $t = 100$, $t = 200$, $t = 300$, $t = 400$, and $t = 500$, respectively.

Compared to the proposed scheme, however, RLM requires a significant time till it achieves the most appropriate rate. The reason behind this performance consists in the fact that unlike RLM, our scheme allows users to estimate the available bandwidth by joining all layers with lower priority. In RLM, users need to conduct several join-experiments till they find out the available bandwidth. During this query for the available bandwidth, significant packet drops occur. On the other hand, in the proposed scheme, users conduct join-experiments at only one time with lower priority. Accordingly, users can receive data at appropriate bit rates from the beginning of the session.

4.2. Mobile network

We secondly investigate the performance of our scheme in mobile environments. At the beginning of the simulation, a mobile node (MN) resides in the cell of BS_1. MN joins a session from server S at time $t = 100$ [sec]. It then moves into the cell overlapping area of BS_1 and BS_2, and performs a handoff at time $t = 300$ [sec] as shown in figure 5. In a similar way, MN performs handoffs from BS_2 to BS_3 and from BS_3 to BS_4 at time $t = 500$ [sec] and $t = 700$ [sec], respectively.

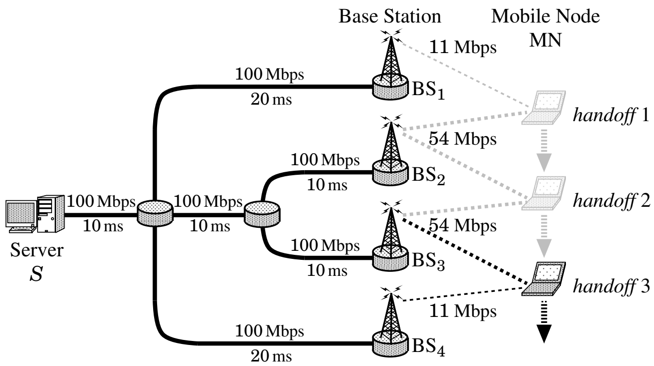


Figure 5: A mobile network topology.

Figure 6 graphs the goodput transition when MN performs handoff from 11 Mbps to 54 Mbps, between 54 Mbps, and from 54 Mbps to 11 Mbps. It demonstrates that the proposed scheme enables MN to receive the stream at rates suitable to the available bandwidth immediately after each handoff. On the other hand, RLM needs longer convergence time as it adds the upper layers one-by-one.

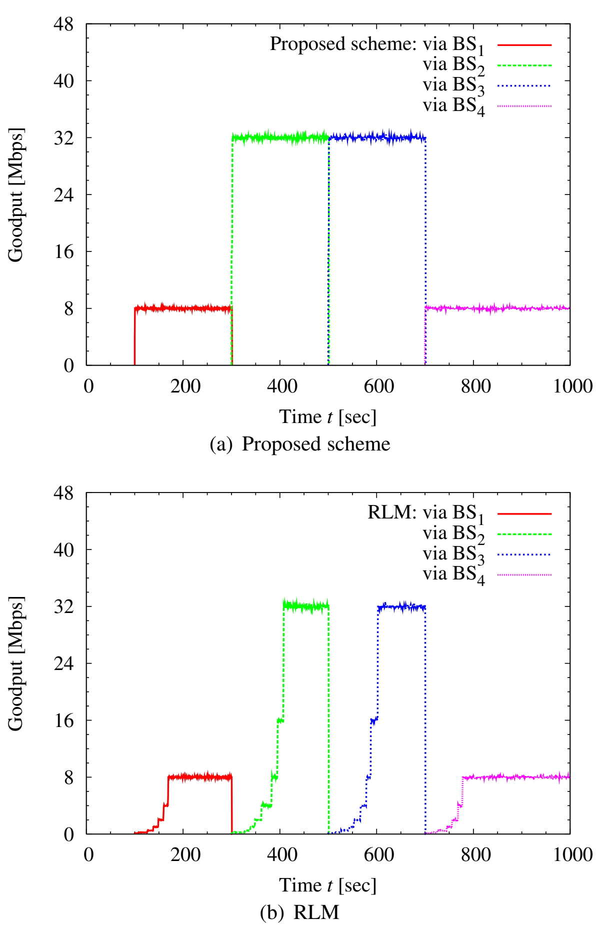


Figure 6: MN performs handoffs from 54 Mbps *to* 11 Mbps*, between* 54 Mbps*, and from* 54 Mbps *to* 11 Mbps.

5. Conclusion

In this paper, we proposed a layered multicast streaming scheme. The proposed scheme exploits the fact that layered multicast uses priority-based packet dropping policies. In our method, a user joins all layers with lower priority at the beginning of a session and then calculates the packet loss rate on each layer. If the packet loss rate on a certain layer exceeds a predefined threshold, the user stops receiving that layer and higher layers. The user then receives packets of lower layers with normal priority.

The performance of the proposed scheme was investigated through several simulations. The obtained results revealed that our scheme enables users to converge fast to the optimal bit rate (most suitable rate to the network conditions) from the beginning of a session. Furthermore, the proposed scheme can be used for mobile users. Our scheme enables mobile nodes to perform a handoff while receiving the stream at the rate available bandwidth allows.

References

[1] N. Shacham, "Multipoint Communication by Hierarchically Encoded Data," in *Proc. of IEEE INFOCOM '92*, Florence, Italy, May 1992, pp. 2107–2114.

[2] S. Bajaj, L. Breslau, and S. Shenker, "Uniform versus Priority Dropping for Layered Video," in *Proc. of ACM SIGCOMM '98*, Vancouver, BC, Canada, Sept. 1998, pp. 131–143.

[3] C.-K. Tham, Y.-S. Gan, and Y. Jiang, "Congestion Adaptation and Layer Prioritization in a Multicast Scalable Video Delivery System," in *Proc. of IEEE/EURASIP Packet Video 2003*, Nantes, France, Apr. 2003.

[4] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven Layered Multicast," in *Proc. of ACM SIGCOMM '96*, Stanford, CA, USA, Aug. 1996, pp. 117–130.

[5] S. Y. Cheung, M. H. Ammar, and X. Li, "On the use of destination set grouping to improve fairness in multicast video distribution," in *Proc. of IEEE INFOCOM '96*, San Francisco, CA, USA, Mar. 1996, pp. 553–560.

[6] A. Legout and E. W. Biersack, "PLM: Fast Convergence for Cumulative Layered Multicast Transmission Schemes," in *Proc. of ACM SIGMETRICS '2000*, Santa Clara, CA, USA, June 2000, pp. 13–22.

[7] Y. Pan, M. Lee, J. B. Kim, and T. Suda, "An End-to-End Multipath Smooth Handoff Scheme for Stream Media," *IEEE J. Select. Areas Commun.*, vol. 22, no. 4, pp. 653–663, May 2004.

[8] K. Kashibuchi, T. Taleb, A. Jamalipour, Y. Nemoto, and N. Kato, "A New Smooth Handoff Scheme for Mobile Multimedia Streaming using RTP Dummy Packets and RTCP Explicit Handoff Notification," in *Proc. of IEEE WCNC*, Las Vegas, NV, USA, Apr. 2006.

[9] The Network Simulator - ns-2. [Online]. Available: http://www.isi.edu/ nsnam/ns/

[10] Cisco IOS Quality of Service Solutions Configuration Guide, Release 12.2. [Online]. Available: http://www.cisco.com/ univercd/cc/td/doc/product/software/ios122/122cgcr/fqos_c/ qcfbook.pdf