Predictive rate control for realtime video streaming with network triggered handover

Susumu Ishihara Kenji Koyama
Shizuoka University
3-5-1 Johoku, Hamamatsu, 432-8561, Japan

Goh Miyamoto Masahiro Kuroda
National Institute of Information
and Communications Technology
3-4 Hikarinooka, Yokosuka, 239-0847, Japan

***Abstract*— We evaluated the effectiveness of a predictive QoS control scheme for real-time video streaming on a mobile network that supports network-triggered handover as the first step to develop a network controlled resource assignment architecture for mobile networks. This scheme is based on a network and application cooperation and enables video streaming applications on mobile devices and the correspondent hosts to control the data transmission rate before handover of mobile devices according to the estimated value of the available bandwidth after handover. This scheme essentially have ability to control the transmission rate of not only a device on handover but other devices to offer fairness to use limited shared bandwidth. The simulation results revealed that the scheme could avoid picture frame losses and starvation of the playback buffer on mobile devices after handover.**

I. Introduction

There have been many reports on integrating heterogeneous radio systems into all IP networks in future wireless communication systems [1][2]. The beyond-3G-system wireless network integrates various types of radio systems including 3G and wireless LANs. In such networks, mobile devices hand over various types of wireless access where characteristics such as delay, bandwidth, and packet-loss rate are different.

Handover in these networks has a large impact on the application QoS of real-time video streaming applications like that in video conferencing. Packet loss during handover and change in bandwidth not only cause temporal frame losses but discontinuity in video. Conventional receiver-report based approaches [3][4] to adapt the transmission rate of streaming data have not been able to avoid delays in adaptation and temporal or persistent degradation in application QoS like frame losses, degradation of space-time resolution etc.

Handover affects not only the handing over device but other devices. When a mobile device M sending/receiving at high transmission rate hands over to a congested cell, data transmission of mobile devices have been associated with the cell will be affected by the traffic of M because of the delay of M's transmission rate control. This may cause temporal excessive degradation of the transmission rate of the preexisting devices. However, if these mobile devices knew M's handover, they would be able to avoid some problems. This will lead the high degree of satisfaction of the users on networks with limited resources. Based on this idea, we aim to develop a network-controlled resource assignment architecture for mobile networks.

For the first step of network-controlled resource assignment for mobile network in order to solve these problems, we proposed a scheme to control the application QoS of video streaming in mobile environments with the help of a network-triggered handover mechanism. We evaluated the effectiveness of the scheme in this paper.

II. Impact of handover on video streaming

The available bandwidth for a mobile device drastically changes if it hands over from a quiet cell to a congested cell, or from a wireless LAN access point (AP) to a 3G AP. If the sender application for video streaming for the mobile device maintains a high data transmission rate for video frames after handover, these frames may be dropped because of congestion or low bit rate in the new AP. If the sender application does not slow the transmission rate, such undesirable conditions will continue. Furthermore, in this case, the mobile device on handover influences on the other devices which associated to the new AP.

Many adaptive rate control schemes have been developed [3][4][5] to adapt the data transmission rate of video streaming applications to the varying traffic load conditions over the Internet. Conventional adaptive-rate control schemes for streaming applications depend on messages from the receiver device measuring round trip time (RTT), packet loss rate, and the number of available frames in the playback buffer. Because of this, on a handover of a receiver mobile device, the transmission data rate is changed *after* receiving a message sent from the receiver device *after* it detects of the influence of the handover. As a result, conventional schemes cannot prevent the negative influence of handover, e.g., burst video frame loss and playback buffer starvation, and use excessive communication resources compared to other mobile devices that were associated with the new AP before handover.

III. Predictive QoS control scheme with network-triggered handover

In this section, we outline the scheme that controls the application QoS of video streaming on handover on a mobile network that supports network triggered handover.

A. Assumed conditions

Before describing our scheme, we will discuss the assumed network architecture and conditions. First, we assume handover of a mobile device is triggered by a administration device on the network. *Mobile Ethernet* [6], a beyond-3G wireless network candidate, has such function. Figure 1 shows the architecture for the Mobile Ethernet. It provides a mobility management mechanism for widely deployed Ethernet

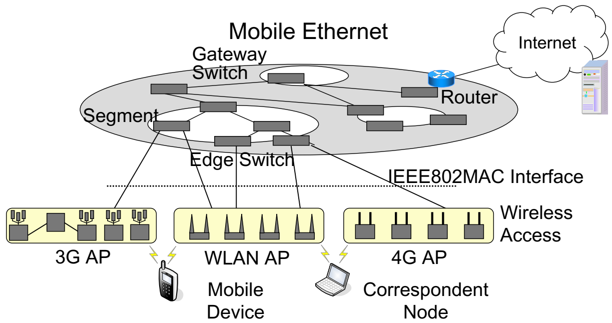

Fig. 1. Mobile Ethernet

networks, and accommodates heterogeneous radio access systems, including 3G. The Mobile Ethernet has features enabling fast handover to be attained across heterogeneous radio access systems using the mobility management mechanism at Layer 2. To achieve this, a signaling server on the Mobile Ethernet determines the AP where each mobile device will switch and the handover timing by gathering radio status information from each device. It then instructs mobile devices to hand over. In the following description, we assume that the mobile network has an equivalent functions of Mobile Ethernet. Furthermore, we assume a signaling server on Mobile Ethernet or an equivalent host can predict the available bandwidth of each AP from the collected information. Note that we allow the error of the prediction in the following discussion.

B. Predictive QoS Control Scheme with Network-Triggered Handover

This scheme is based on cooperation of application program and signals from mobile network, while conventional rate control schemes for streaming rely on traffic and buffer measurement on mobile devices. In this scheme, handover is predicted not only for layers 2 and 3 like conventional approaches that reduce handover latency, but also for layer 7 that adapts the data transmission rate at this layer. Application programs on mobile devices adapt the data transmission rate of streaming video according to handover notification and the estimated bandwidth of the new AP from the signaling server. Through this prediction, applications on mobile device can avoid the delay in adapting the data transmission rate to the conditions of the new APs, and reduce frame losses on handover caused by depleted buffered frames.

There is an overview of the scheme in Fig. 2. First, handover is notified by the signaling server to the MAC of a mobile device about to hand over as a layer 2 signal. This notification signal includes the estimated value of the available bandwidth at the vicinal AP where the mobile device is going to do the handover. After the mobile device receives notification, it is sent from layer 2 to layer 7 as an OS signal. Streaming applications use the information on this signal to adapt their transmission and receiving data rate and also request applications at their correspondent hosts to change their transmission data rate. After this procedure, the signaling server requests the mobile device to hand over, which it then initiates. The last request may not be used if the mobile device decides the timing of the handover by itself after the first notification.

Conventional adaptive-rate control schemes depend on messages from the receiving device after communication conditions have changed. Because of this, the sender cannot avoid the negative influence of changes in communication conditions. The proposed scheme, on the other hand, changes the transmission data rate just before handover. Therefore, the negative influence of handover, e.g., frame loss and playback buffer starvation decrease, and its influence on other device is mitigated.

If the handover of a mobile device and the data sending rate that the device has used are notified to mobile devices that have been associated to the new AP, application programs on these mobile devices can also adapt their data transmission rate before the real handover to avoid the impact of the handover.

C. Measuring available bandwidth

The proposed scheme depends on estimating the available bandwidth of each mobile device at the AP. Therefore, collecting information from all APs is important to make the proposed scheme work efficiently.

Recently, wireless APs have begun to perform some additional or higher-layer functions. Some of these are capable of collecting detailed statistics on the communications of each mobile device, such as the maximum bit rate, the number of successful packets for each device, and the number of associated devices. For example, if the signal strength of the mobile device, the maximum bit rate of the link and the number of the associated devices are known in an IEEE802.11 network, we can roughly estimate the available bandwidth including safe margin for the mobile device.

Standards for communication between these intelligent APs have also been discussed. For example, the Light Weight Access Point Protocol (LWAPP) proposed by IETF [7] provides access routers with the ability to obtain any statistical information collected by the APs. Using these functions, we can approximately estimate the available bandwidth for each mobile device after handover.

Collecting statistical values from many APs and estimating the available bandwidth accurately for all mobile devices impose burdens on the signaling server. To reduce these, it is important to allow some leeway in the estimated value for rate control.

D. Notifying available bandwidth

The proposed scheme depends on the messages from a signaling server to mobile devices. Especially, the timings of the arrivals of the messages from a signaling server and the accuracy of the information included by the messages are important.

a) Interval between arrival of rate control message and real handover: This will affect the adaptation of the frame data transmission rate. If the rate control message arrives earlier than handover that reduces available bandwidth, the data transmission rate will become adequately low and may not cause buffer starvation. If the rate control message arrives after handover, on the other hand, the rate will be changed according to the conventional rate control scheme running with the proposed scheme. It may cause buffer starvation and frame losses.

When the available bandwidth becomes larger after handover, even if the rate control message arrives earlier than the handover, the mobile device should not change the sending data rate larger, because the earlier acceleration of the sending rate may result in queuing at the old access point and frame losses.

b) Effect of accuracy of estimation: In a practical environment, the accuracy of the available bandwidth estimated at the new AP will be low. Because of this, the predictive rate control mechanism should work effectively even if this accuracy is low. Intuitively, if the estimated value of the available bandwidth is smaller than the real value, predictive rate control will not result in buffer starvation, but the frame data rate will be excessively small. On the other hand, if the estimated value is larger than the real value, this will cause congestion of the network and playback buffer starvation for a long time.

IV. Performance Evaluation

A. Simulation Model

To demonstrate the effectiveness of predictive QoS control, we evaluated the effect of the network and application cooperation scheme through simulations.

1) Network model: We used a MIRAI-SF simulator developed to assess the Mobile Ethernet to evaluate the proposed scheme. Figure 3 outlines the network topology for the simulation model. A mobile device is receiving video streaming data from a host outside the Mobile Ethernet. The mobile device is moving from a cell covered by AP1 to a cell covered by AP2 while the transmission is streaming. The bandwidth of the down-link from AP1 to the mobile device is 1 Mbps. The RTT between the mobile device and correspondent host was 60 ms.

We can see the bandwidth of the down-link from AP2 to the mobile device is 384 kbps. The bandwidth of the up-link was the same as the down-link. In the simulation, the behavior of the MAC layer protocol was neglected. Traffic caused by other mobile devices, packet loss and jitter caused by wireless links were also neglected. Packet loss only occurred because of delay in the MAC address table update occurring with handover and the saturated transmission queue on APs. The queue length was 10 packets.

Link update messages concerning Mobile Ethernet operation to update the MAC address table in switches were sent when a mobile device handed over during the simulation. These messages were not neglected.

2) Video data transmission model: The video streaming server on the correspondent host sent video streaming data to the mobile device as UDP packets. Up-link video data transmission was neglected. All video frames were assumed as I frames like in motion JPEG. In other words, the relationship between frames that MPEG video has was neglected. We assumed the frame size could be changed with respect to each frame. The frame rate was fixed to 24 frames/sec.

The proposed scheme was intended to be used for bidirec-

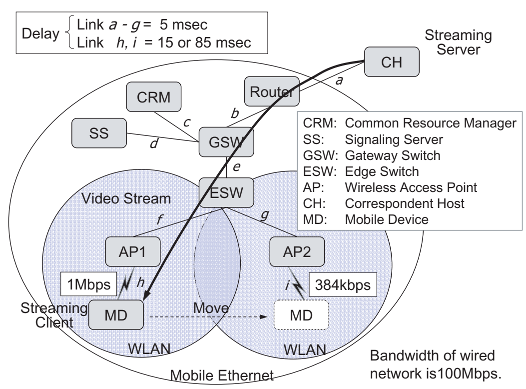

Fig. 3. Simulation Model

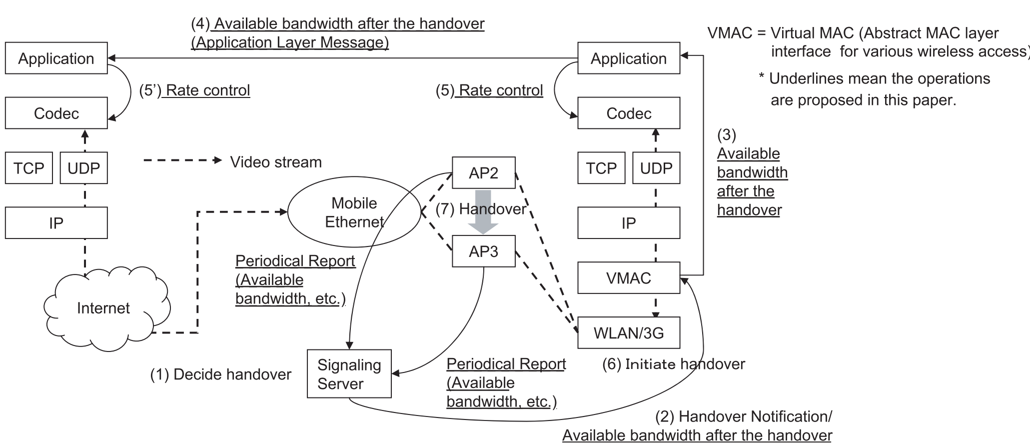

Fig. 2. Predictive QoS control scheme with network-triggered handover

 0-7803-8966-2/05/$20.00 © 2005 IEEE

tional real-time video streaming. However, the most critical situation with the scheme is a case where a mobile device sends a rate control message after receiving a bandwidth-change and handover notification from the signaling server and sends a rate control message to the correspondent host outside the Mobile Ethernet. Because of this, we only simulated this situation.

3) Model of conventional adaptive data rate control scheme: If the proposed prediction scheme is not used, the data rate from the sender host should be adapted to the conditions of the wireless link with the conventional scheme. Even if the proposed method is used, such a scheme should be used.

In our simulation, we used a scheme based on the ones proposed in [5] and [3] for this purpose. The frame data size in this scheme was scaled according to the following procedure.

Consider the situation at the playback buffer, which contains correctly received video frames. The video session starts with a preloading phase in which $Q^* + 1$ frames are prefetched into the buffer before playback commences. The preloading phase provides a cushion against variations in the frame arrival rate at the playback buffer. A value of Q^*, which is referred to as the playback buffer threshold, is selected depending on the target video quality, limited by real-time bidirectional communication channel.

The goal of this rate control scheme is to try to maintain the playback buffer occupancy around Q^*. For real time bidirectional multimedia applications like video chats and video conferencing, the preload time Q^* should be short to allow smooth interaction. For example, in ITU-T Recommendation G.114 [8], the acceptable delay is 150 ms for most real-time user applications in traditional telephony using G.711.

Once the preloading phase is completed, video playback can commence at a rate of f_p frames per second. Let $Q(i)$ be the number of frames in the playback buffer right after the playback of the ith video frame, $i = 1, 2, \ldots$. Note that $Q(1) = Q^*$. Frames not contained in the playback buffer account for the calculation of $Q(i)$ if newer frames than these have arrived. In other words, $Q(i)$ means the time that video can be played allowing some packet loss.

The occupancy of the playback buffer evolves according to:

$$Q(i+1) = \max\left\{0, Q(i) - 1 + \frac{f_r(i)}{f_p}\right\}, \qquad (1)$$

where $f_r(i)$ is the average rate at which frames are correctly received in the interval between the playback times of the ith and $(i+1)$th frames. Under ideal conditions, $f_r(i) = f_p$, and hence $Q(i+1) = Q(i) = Q^*$. However, when the channel is in a "bad" state (i.e., congested shared link or packet loss caused by handover), we are likely to have $f_r < f_p$, causing the playback buffer to underflow, increasing the backlog at the transmitter buffer. Such underflow is compensated for by means of rate control that allows the sender application to drain its backlogged queue and catch up with the frame encoding process. During this compensation period, we have $f_r(i) > f_p$. Due to channel uncertainties and the predictive nature of the rate control algorithm at the receiver, the rate controller may end up over-compensating, leading to $Q(i+1) > Q^*$.

The receiver sends the recommended data rate that the sender should use just after a video frame is played back. Define $S(i)$ as the recommended frame data rate reported to the sender before the ith frame is played back. Essentially, $S(i)$ is selected so that the buffer content is kept at around the threshold, Q^*.

Depending on $Q(i)$, the receiver selects the value of $S(i+1)$ for the next frame as follows.

- **Case I:** $Q(i) > Q^*$ (Stable Regime): In this case, $S(i+1)$ is set to $S(i)+\alpha$. Here, the value of α is small to enlarge the data rate of video frames when the link is capable of a larger data rate.
- **Case II:** $0 \leq Q(i) < Q^*$ (Underflow Regime): In this case, $S(i+1)$ is set to $\max\{\ S(i)/(Q^* - Q(i)),\ S_{\min}\ \}$.

In the simulation, the value of α and $S_{\min}$ were 15 kbps and 16 kbps respectively.

4) Model of the proposed scheme: In the discussion on these, we assume that a mobile device moves from a cell covered by an AP with a large available bandwidth for mobile devices to a cell with a small available bandwidth.

In the simulations, the mobile device updates $S(i+1)$ before handover according to the estimated available bandwidth for transmission data to the mobile device. Then the device sends this updated $S(i+1)$ to the correspondent host as a UDP message. In a real network, the estimated value is notified by the signaling server before handover.

To evaluate the effect of the interval between the transmission of a control message and real handover, we changed the interval in the simulation. Figure 4 shows the timing for the transmission of the rate control message and handover in the simulation. Handover occurred at 2,100 ms (= elapsed time from handover is 0 ms from the start of the data transmission). A rate control message to the correspondent node of the mobile device is sent at 1,900 ms (= -200 ms), 2,100 ms (= 0 ms) or 2,300 ms (= +200 ms).

To evaluate the effect the accuracy of the estimation of available bandwidth, we used several different estimated values of the available bandwidth at the new AP during the simulation. The estimated value was set to 80% (the estimated available bandwidth for a new AP was about 300 kbps), 100% (384 kbps), 120% (460 kbps), and 150% (576 kbps) of the real value.

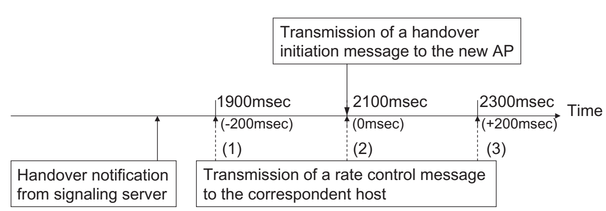

Fig. 4. Time chart for simulation model

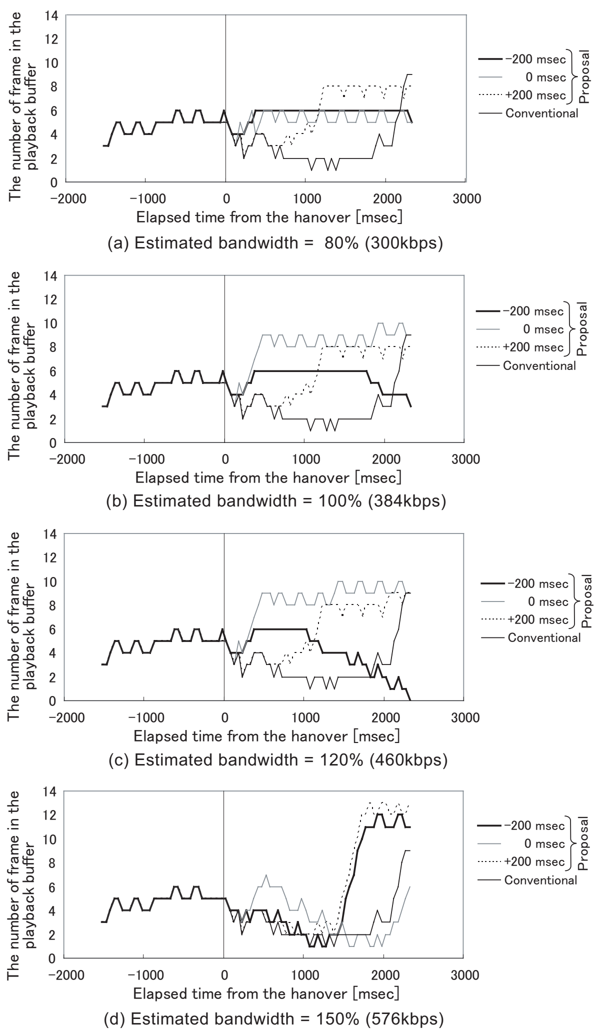

Fig. 5. Number of frames in playback buffer.

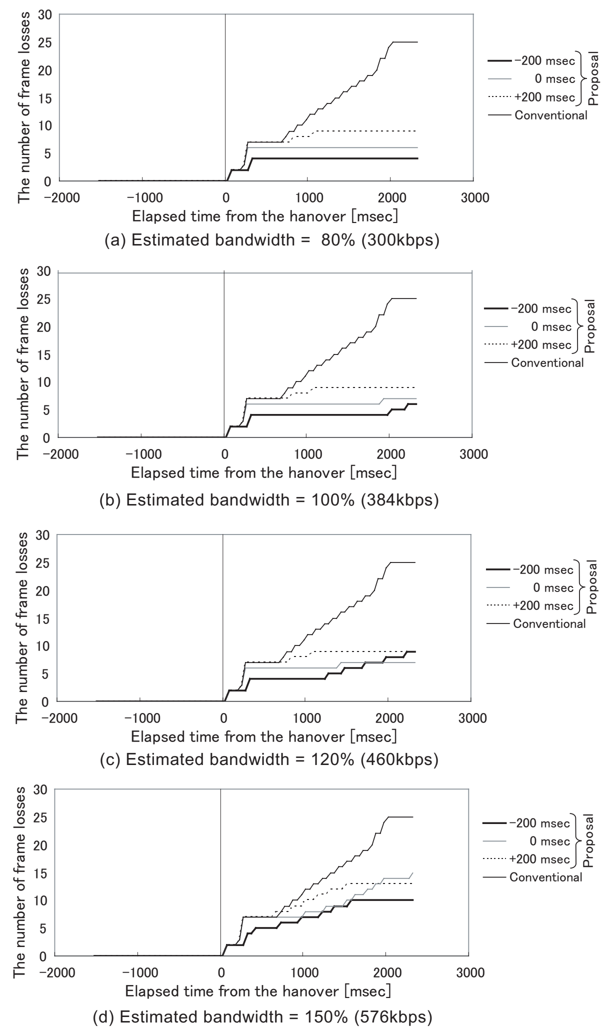

Fig. 6. Cumulative number of frame losses.

B. Simulation Results

Figure 5 plots the transition in the number of video frames in the playback buffer when the available bandwidth was changed from 1 Mbps to 384 kbps by handover. Q^* was set to 5. When the proposed scheme was not used, there were a smaller number of frames in the buffer just after handover.

When the rate control message was sent before handover, and the estimated bandwidth was smaller than the real value or much the same (Fig. 5(a), (b) and (c)), the number of frames returned to Q^* within a shorter time. The faster the rate control message was sent, the faster the buffer returned to the ideal state and became stable. However, when the estimated bandwidth was much larger than the real value (Fig. 5(d)), the number of frames did not return to Q^* for a long time. Note that even if the error in predicting the bandwidth is large, the number of frames in the playback buffer returns to Q^* faster than with the conventional scheme.

Figure 6 plots the transition in the cumulative number of picture frame losses. These losses occurred because the packets making up the frame did not arrive before the frame played back. The simulation conditions were the same as in Fig. 5. When the proposed scheme was not used, more than 25 frames were lost for 2000 ms after handover. When the estimated bandwidth was not too large (Fig. 6(a), (b), (c)), almost no frame was lost 400 ms after handover though some frames were lost just after handover. The later the rate control message was sent, the more frames were lost. When the estimated value was larger than the real value, differences between the conventional and proposed schemes became small.

Figure 7 plots the transition in the data transmission rate from the correspondent host. This value was calculated from the data size in transmitted frames. When the proposed method was not used, the data transmission rate suddenly decreased at 800 ms after handover. And it reached the minimum value defined in the simulation. On the other hand, when the estimated bandwidth was not too large (Fig. 5(a), (b), (c)), the data transmission rate decreased around the timing of handover

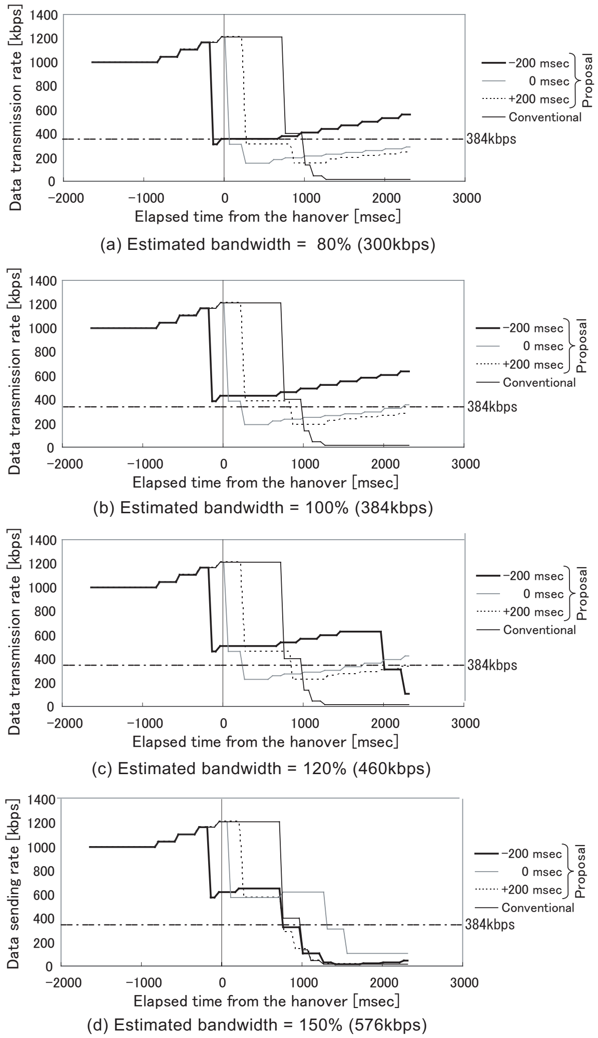

Fig. 7. Data transmission rate.

and the value quickly approached the real available bandwidth. The faster the rate control message was sent, the faster the data transmission rate improved. When the estimated value was too large (Fig. 5(d)), even if the predicted bandwidth was sent to the correspondent host, the data sending rate decreased to the minimum value, because the too large data sending rate caused packet losses and decreased the number of the frames in the playback buffer.

Once the data transmission rate become small, it took a long time to reach the real available bandwidth. Of course, this speed depends on parameter α to enlarge frame data size $S(i)$. If this value is larger than the value we used in simulation, $S(i)$ will become large sooner. However, as a larger α may create much larger transmission rate oscillations and more packet losses. α should be selected carefully. For example, in (Fig. 7(c)), when the rate control message was sent at 200 ms before handover, the data transmission rate decreased at 1800 ms after handover. This was caused by too aggressive augmentation and too large predicted value.

Although the results are not shown, we simulated cases when the RTT between a mobile device and the correspondent host at the new AP was 200 ms on assumption that a mobile device handed over from a cell covered by a wireless LAN AP to a cell covered by a 3G AP. In the results, there was the same tendency as in Figs. 5–7.

These results revealed that even if the estimated available bandwidth differed somewhat from the real value, our scheme, which could predict the available bandwidth quickly and accurately, could prevent picture frame losses and the playback buffer from being starved on mobile devices. However, too early or too pessimistic predictions will greatly decrease the application QoS of streaming video. Finding suitable timing and predictive values remains a serious problem that needs to be solved.

V. Conclusions

We evaluated the effectiveness of a predictive QoS control scheme for real-time video streaming on a mobile network with network triggered handover. In the scheme, application programs on a mobile device with the network and application cooperation scheme control the data transmission rate according to the estimated value of the available bandwidth after handover. This scheme is essentially have an ability to benefit not only mobile devices going to handover but also ones that has been associated with vicinal APs. This cannot be achieved with conventional-receiver report-based approaches. The simulation results revealed that fast and pessimistic rate control before handover works effectively. Additionally the proposed scheme could be said to be effective in real-time streaming video as compared with the conventional schemes.

Although we neglected in the simulation, to investigate the behavior of devices other than the mobile device on handover is important. If these devices could know the handover and the effect in advance, they would be able to control their application's data transmission rate appropriately. Comprehensive study including this function will be our future work.

References

[1] H. Lach, C. Janneteau and A. Petrescu, "Network Mobility in Beyond-3G Systems," *IEEE Communication Magazine*, pp. 52–57, July 2003.

[2] H. Yumiba, K. Imai, and M. Yabusaki, "IP-Based IMT Network Platform," *IEEE Personal Communication Magazine*, vol. 8, no. 5, pp. 18–23, October 2001.

[3] D. Wu, Y. T. Hou, W. Zhu, H-J. Lee, T. Chiang, Y-Q. Zhang, H. J. Chao,"On end-to-end architecture for transporting MEPG-4 video over the internet," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 6, 2000.

[4] T. Turletti and C. Huitema, "Videoconferencing on the Internet," *IEEE Trans. Networking*, vol. 4, pp. 340–351, 1996.

[5] M. M. Krunz and M. Hassan, "Adaptive rate control scheme for video streaming over wireless channels," in *Proc. of the Data Compression Conference (DCC'04)*, 2004.

[6] M. Kuroda, M. Inoue, A. Okubo, T. Sakakura, K. Shimizu, F. Adachi, "Scalable Mobile Ethernet and Fast Vertical Handover," in *Proc. of IEEE Wireless Communication and Networking Conference (WCNC2004)*, A27-3, Mar. 2004.

[7] P. Calhoun, B. O'Hara S. Kelly, R. Suri, M. Williams, M. Vakulenko, "Light Weight Access Point Protocol (LWAPP)," internet draft `draft-ohara-capwap-lwapp-00`, Mar. 2004.

[8] ITU-T Recommendation G.711 (11/88), "Pulse code modulation (PCM) of voice frequencies," 1988.