

Playback-Rate Based Streaming Services for Maximum Network Capacity in IP Multimedia Subsystem

Chin-Feng Lai, *Member, IEEE*, and Min Chen, *Senior Member, IEEE*

Abstract—In the shared environment of multimedia files of an IP multimedia subsystem (IMS) system, the issue of quickly obtaining the desired files from various heterogeneous networks presents a challenge. There are many studies on information-sharing mechanisms and bandwidth estimation methods in the IMS environment; however, most focused on how to effectively take advantage of network bandwidth to more quickly transmit data to the receiver by the Greedy approach. Nevertheless, with increasing numbers of users in the IMS environment, more and more connections and data transmission will easily result in too heavy network traffic load. Hence, this paper proposed cross-layer playback-rate based streaming services, which can maintain network transmission quality and receive data before playback reliably in IMS networks with many users. The experimental results show that the services could reduce the overall network load without the occurrence of dropped packets.

Index Terms—IP multimedia subsystem, maximum network capacity, playback-rate.

I. INTRODUCTION

ADVANCES in communications technology make the next generation network (NGN) core network architecture evolve into IP technology-based all-IP networks, with IP multimedia subsystem (IMS) as the core. The IMS technology can integrate mobile networks and fixed networks, which are original systems of different network transmission media. IMS uses session initiation protocol (SIP) technology as the communications standard protocol of various heterogeneous networks, which provide mobile communications users and service providers a heterogeneous multimedia service environment. However, in a heterogeneous network environment,

the multimedia service implementation performance may be lowered due to different network status, further affecting the quality of various network services. To address this problem, a cross-layer QoS design is required to ensure service quality. Most studies proposed mixing various heterogeneous networks to provide always-best-connection methods and architectures in order to transmit data to a receiver as quickly as possible, which may result in the overuse of network bandwidth. Cases of excessive network traffic load may result in data transmission errors and the inability of servers with rare multimedia files to satisfy the requests of a number of clients. Such a situation would be even more serious in the cases of integrated heterogeneous networks. Meanwhile, multimedia streaming technology requests that playback start upon downloading a few data segments. By using the concept of a client buffer, the data can be directly read, played back, and abandoned through a buffer without the actual storage to save local disk storage and a lot of download wait time. Therefore, this paper proposed the playback-rate based streaming services for maximum network capacity in IMS. By its cross-layer design, this method could dynamically provide appropriate multimedia services, based on different requirements, to transmit multimedia data and satisfy users QoS prior to playback, while maintaining network bandwidth quality to satisfy more needs of multimedia services.

This paper had the following three contributions.

- 1) This paper proposed a cross-layer design in the service/application layer, the IMS layer, and the transport layer in order to achieve a more efficient control method in the IMS architecture.
- 2) This paper proposed a method to evenly transmit multimedia data on the network bandwidth that could reduce transmission errors, and ensure real-time playback and completion of multimedia data transmission before playback.
- 3) As it was a playback rate-based calculation method, the proposed method could satisfy relatively more multimedia services.

Details were presented in later sections. The second section discussed IMS and QoS-related technologies, the third section elaborated on the overall architecture of playback-rate based streaming services in IMS and proposed the simulation experiment in an NS2 environment, the fourth section presented the

Manuscript received November 15, 2010; revised June 15, 2011; accepted July 5, 2011. Date of publication September 23, 2011; date of current version November 23, 2011. The work of C.-F. Lai was supported by the National Science Council of the Republic of China, under Grant NSC 99-2218-E-197-003. The work of M. Chen was supported by NAP of the Korea Research Council of Fundamental Science and Technology, and the IT Research and Development Program of KCA (10913-05004: Study on Architecture of Future Internet to Support Mobile Environments and Network Diversity).

C.-F. Lai is with the Institute of Computer Science and Information Engineering, National Ilan University, Ilan 260, Taiwan (e-mail: cinfon@ieee.org). M. Chen is with School of Computer Science and Engineering, Seoul National University, Seoul 151-742, Korea (e-mail: minchen@ieee.org).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSYST.2011.2165190

experimental data and results, while the fifth section offered the summary.

II. RELATED WORK

A. IMS

IMS is the technical standard proposed by the international cooperation organization, 3GPP (as shown in Fig. 1). Its main functionality is to integrate the circuit switched domain and the packet switched domain. Based on SIP, the IMS platform transmits multimedia data by peer to peer approach after establishing the session. Through its open and standard architecture, service providers can simultaneously provide voice, data, video, and other diversified application services on the IMS platform. Meanwhile, the IMS platform can serve as a common platform for circuit switched domain or packet switched domain networks, such as the fixed network, WLAN, WiMAX, GSM, and GPRS. The purpose of applying IMS is to provide mobile Internet services in the 3G network architecture, and ensure QoS during the conversion of different networks to establish multimedia sessions which allow the use of multimedia free from restrictions and with greater flexibility to provide more services to users [7]–[9], [13]–[19].

IMS divides the network architecture into layers, including the service/application layer, IMS layer, and transport layer. The application program server of the service/application layer provides IMS users with various services. The operators may develop programs or services through the standard IMS architecture and establish the application service layer to connect with the IMS core layer to provide users with services. The IMS layer is the core component of the call session control function, such as serving-call session control function (CSCF), interrogating-CSCF, and proxy-CSCF. User SIP request signals are all processed, analyzed, and transmitted through the component of this layer to the correct receiver, for example, transmitting the signal to the IMS of other service providers, sending and establishing connection signals, and sending application server request services. The IMS core network has a group of SIP proxy known as CSCF; after the integration with home subscriber system, CSCF can transfer any kind of SIP services to any core network to achieve the capability of crossing different core networks. In the transport layer, there are different types of access technologies. Users may achieve a core network connection through an IPv4 or IPv6 network, WiFi, GPRS/3GP, LTE, WiMax, and other technologies, and then connect with the IP multimedia subsystem to use the various services provided by the IMS. It can be known from the above that no matter the type of access technology, the user can connect with IMS and use the various resources provided by the system after connecting with the core network. Thus, it can be seen that the IP multimedia subsystem has a wide range of technical support for the terminal and access technology of various users [20]–[26].

B. Cross-Layer QoS Issue

Providing QoS in a mobile network is a very challenging problem. Changes in bandwidth and hand-off between base

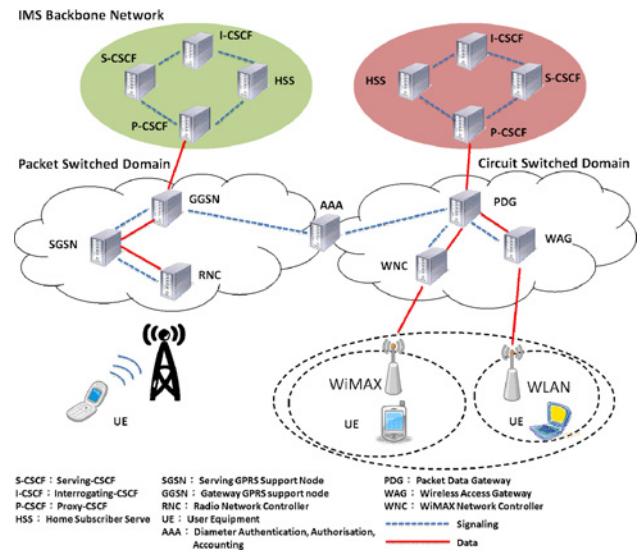


Fig. 1. Illustration of IP multimedia subsystem architecture.

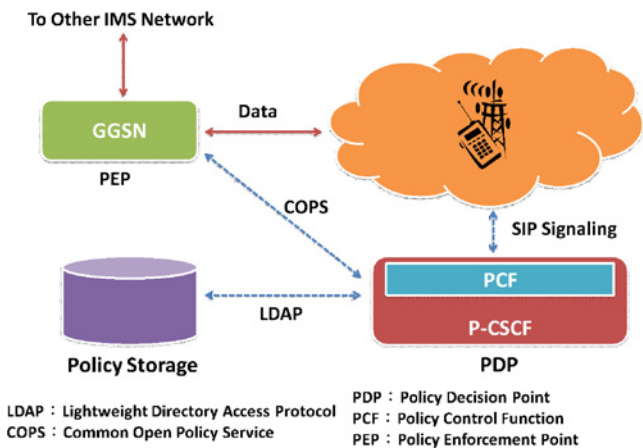


Fig. 2. Illustration of IMS policy decision function.

stations seriously affect the transmission of packets, making the mobile network real time applications quite vulnerable. In general wired networks, the transmission of packets is in the state of best efficiency. This state means that network will maintain the applications required bandwidth without any assurance, according to bandwidth availability and network congestion. The design cannot guarantee the QoS of the real-time application services of the mobile network. The IMS QoS mechanism is designed to replace the state of best efficiency to ensure transmission quality. To ensure packet transmission quality, the IMS QoS mechanism can acquire resources in advance through measuring network-related parameters, including transmission rate, gateway delay, and error rate. The user may designate the required QoS according to service forms and user environments. The IMS intelligent QoS request is termed as a policy decision function (PDF) model (the implementation architecture is shown in Fig. 2), where the PDF module interacts with the basic packet switching network and controls the basic packet switching network resource distribution (via go interface to GGSN).

Relevant studies on network QoS cross-layer enhancement, in [1] and [27]–[29], Camarillo *et al.* proposed an approach

to IMS policy control based on session policies that achieve transparent end-to-end session establishments between IMS terminals. In this paper, the proposal offers modularity and scalability properties that enable operators to establish policies and modify existing ones without major changes in the IMS core. In [2], the authors focused on a possible enhancement of ITU-T NGN architecture, tackling in a comprehensive way a few problems that have been identified within principal standardization bodies such as dynamic provisioning of QoS enabled connectivity across heterogeneous network technologies to support application services, resource virtualization, and service signaling made independently of the source node. This paper has proposed a novel NGN RACF entity, named TRCG-FE, able to control dynamically and to coordinate the QoS-guaranteed connectivity provided by CP-enabled transport networks to support NGN service provisioning. In [3], an optimized, network driven QoS management and provisioning model are proposed. It is also considering a localization mechanism, doubled by a QoS control of the access networks. This paper focuses on a single operator scenario, which deploys multiple access technologies using an IMS signaled 3GPP system architecture evolution. There are more works studying different problems in QoS on IMS [4]–[6], [30].

III. PLAYBACK-RATE BASED STREAMING SERVICES FOR MAXIMUM NETWORK CAPACITY IN IP MULTIMEDIA SUBSYSTEM

In an IMS environment, different communication transmission protocols of various domains make it difficult to estimate the cross-domain end-to-end (E2E) bandwidth. From the application perspective, the major factors of E2E QoS are as follows.

- 1) With relevant mechanisms and interfaces that allow applications to designate QoS requirements, for example, applications relating to installation or management tools allow the presetting of QoS parameters to the application.
- 2) Transmit the QoS requirements by E2E to various domains. A certain domain may use such information to set special functions and resources to enable the domain to provide the required QoS. When the certain domain connects with that particular domain, the QoS-related information will transmit the information to the new domain.
- 3) Each domain requires a special mechanism to support the QoS demands, including pre-flow admission control, resource reservation, scheduling, and a prioritization mechanism.

Traditionally, when establishing a communications conference, the parties should exchange information in the session layer and application layer. Such information exchange can enable the communicating parties to negotiate capabilities for providing services and special application-related attributes, such as the encoding format and version number. This way, the E2E communications can perform the appropriate network loading service. At this stage, the QoS information should be

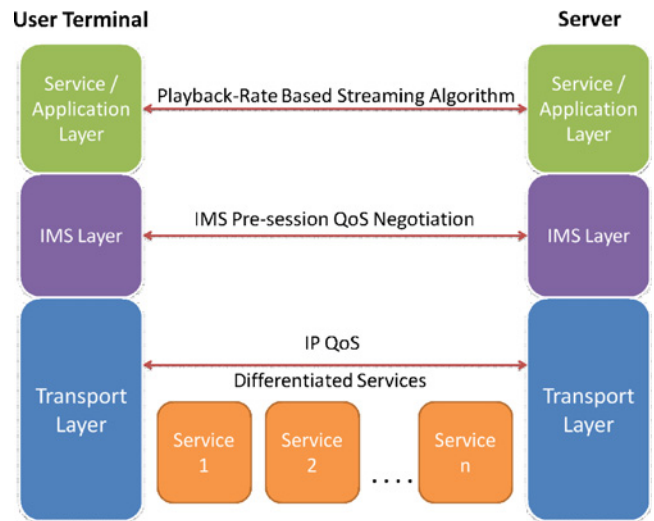


Fig. 3. Cross layer playback-rate based streaming services in IMS.

distributed in all domains on the E2E path. The E2E QoS information distribution mechanism requirements include the following points:

- 1) generality of various different QoS technologies;
- 2) controlled information delivery should be clear and distinctive;
- 3) independent from applications;
- 4) flow identification independence (user flow and flow identifier should be independent from each other);
- 5) simultaneously support one-way and two-way resource reservations;
- 6) able to reconstruct efficient services after hand-off. In fact, with protocols of mobility and hand-off, distributive mechanisms should cooperate seamlessly;
- 7) on the E2E path, admission control and resource reservation should enable the domain to communicate with involved domains after judgment.

Once the QoS information is established on the E2E path, each domain can set its own QoS mechanism according to the users data stream. Hence, this paper established the cross layer playback-rate based streaming services within the architecture shown in Fig. 3. First, in the IMS layer, the setting of QoS parameters was completed before establishing the two user environment sessions. The data transmission time in the transport layer was then estimated to predict the data transmission delay time from the sender user equipment to the receiver user equipment according to the features of the multimedia contents when transmitting multimedia streams. Finally, by the scheduling algorithm of the service/application layer, the multimedia content was transmitted to the receiver user equipment (UE).

A. Transport Layer

In the transport layer, QoS is mainly achieved with differentiated services due to IMS that connects mobile and fixed network, the two different systems of network transmission media. According to service quality, differentiated services will classify every data package transmission appearing in network. In data packages queuing in a router, the ones

with high priority will be processed earlier than others. Differentiated services decide priority of data packages by differentiated services code point defined in IP header, and deliver them to next router by per-hop behavior. Differentiated services is easier to implement because it applies the control of package classification, is involved in queue dispatching and buffer management, and never considers which of next hops a package should go to. In the network, differentiated services simply control and deliver data packages with the order of their priorities. In IMS, each application with the requirement of different QoS generates the packages of its priority; each kind of package delivered by each router has a priority of its own. Hence estimating QoS of E2E transmission depends on the waiting time of packages required. In this paper, we defined some symbols for easy illustration, let μ denote the average speed of process packages in the router. Let s denote the number of routers for the package's next hop and P_n denote the probability of n packages that precisely are waiting to be processed in the router. In this paper, W is deemed as the time of each package in the router (contains the process time) $= \frac{L_q}{\lambda}$. On the contrary, W_q is deemed as the time of each package in the router (does not contain the process time) $= W_q + \frac{1}{\mu}$. The expected number of packages in the router is defined as

$$L = \sum_{n=0}^{\infty} n P_n = \lambda(W_q + \frac{1}{\mu} = L_q + \frac{\lambda}{\mu}) \quad (1)$$

and L_q denotes the expected number of packages waiting to be processed (does not contain the packages that are processing). Then let B denote the allocation of the router's processing time and C denote the number of the transmitting packages. When package n is transmitted to the next router, the factor of package n , C_n , can be calculated by

$$C_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!}, & \text{for } n = 1, 2, 3, \dots, s \\ \frac{(\lambda/\mu)^s}{s!} \left(\frac{\lambda}{s\mu}\right)^{n-s}, & \text{for } n = s+1, s+2, \dots \end{cases} \quad (2)$$

If $\lambda < s\mu$, we can get the utilization factor of packages $\rho = \frac{\lambda}{s\mu} < 1$. Therefore

$$P_0 = \frac{1}{\left[1 + \sum_{n=1}^{s-1} \frac{(\frac{\lambda}{\mu})^n}{n!} + \frac{(\frac{\lambda}{\mu})^s}{s!} \sum_{n=s}^{\infty} \left(\frac{\lambda}{s\mu}\right)^{n-s} \right]} \quad (3)$$

$$= \frac{1}{\left[\sum_{n=1}^{s-1} \frac{(\frac{\lambda}{\mu})^n}{n!} + \frac{(\frac{\lambda}{\mu})^s}{s!} \frac{1}{1 - \lambda/(s\mu)} \right]}$$

According to C_n and P_0 , P_n can be calculated by

$$P_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!} P_0, & \text{if } 0 \leq n \leq s \\ \frac{(\lambda/\mu)^s}{s! s^{n-s}} P_0, & \text{if } n \geq s. \end{cases} \quad (4)$$

From P_n and L , we can get

$$L_q = \sum_{n=s}^{\infty} (n-s) P_0 = \sum_{j=0}^{\infty} j P_{s+j} = \sum_{j=0}^{\infty} j \frac{(\frac{\lambda}{\mu})^s}{s!} \rho^j P_0$$

$$= P_0 \frac{(\frac{\lambda}{\mu})^s}{s!} \rho \frac{d}{dp} \left(\sum_{j=0}^{\infty} \rho^j \right) = P_0 \frac{(\frac{\lambda}{\mu})^s}{s!} \rho \frac{d}{dp} \left(\frac{1}{1-\rho} \right) \quad (5)$$

$$= \frac{P_0 (\frac{\lambda}{\mu})^s \rho}{s! (1-\rho)^2}$$

However, when the packages with high priorities arrive and wait to be processed, the router will switch the packages with the low priorities to the waiting status in order to let the router could process the high-priority packages immediately. If we consider the priorities of packages for DiffServ, we let W_k denote the waiting time of the k th priority packages in the router, μ_k indicate the average speed of process the k th priority packages, λ_i denote the average arrival time of the i th priority packages, and L_k mean the expected number of the k th priority packages in the router (includes the packages that have been processing), where $B_0 = 1$, $B_k = 1 - \frac{\sum_{i=1}^k \lambda_i}{s\mu}$, $\lambda = \sum_{i=1}^N \lambda_i$ and $\gamma = \frac{\lambda}{\mu}$. The total expected waiting time can be estimated by

$$W_k = \frac{1}{B_{k-1} B_k \mu}, \quad \text{for } k = 1, 2, \dots, N. \quad (6)$$

And L_k still is equal to $\lambda_k W_k$.

B. IMS Layer

The IMS standards should explicitly regulate QoS parameters prior to the establishment of sessions between UE and server [10], the purpose of which was to check whether there was sufficient resources available when connecting between UE and server. After the confirmation of QoS parameters, IMS network requested the core network and the access network to reserve resources for the session. The implementation process of the SIP INVITE and UPDATE messages is shown in Fig. 4 [11].

When UE sent the INVITE message to server, the message would be entrained with the QoS proposal (request), which checked the subscription level in the serving-call session control functions (S-CSCF) of UE and server and determined QoS parameters. Afterward, server would return the QoS proposal, which would also check the subscription level of the S-CSCF in UE and server, and determine the QoS parameters accordingly. Finally, UE received the server QoS proposal, and started to establish the session or renegotiate according to the

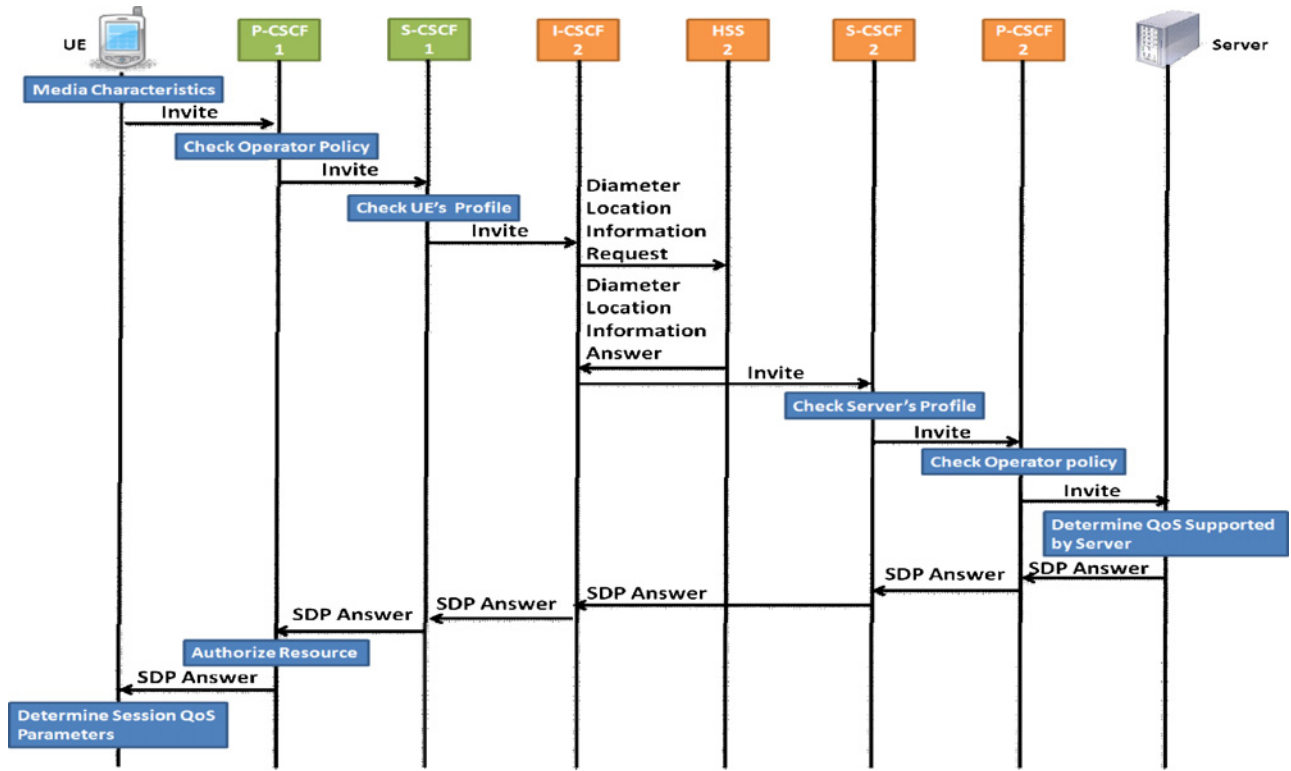


Fig. 4. Implementation process of the IMS QoS parameters update.

TABLE I
NOTATION

Symbol	Definition
T	The current time
T_s	The segment start to be played
T_t	The threshold time
J	A set of segments
I	A set of senders
b_i	The remaining bandwidth of sender i
b_c	The remaining bandwidth of receiver c
b_s	The bandwidth of segment are played per second
$m_{i,j}$	A binary variable $x_{i,j} = 1$ to denote that the j th segment is kept in sender i
J_b	The byte size of segment J
r_{xij}	The receiver is receiving the j th segment form sender i
x_{ij}	A binary variable $x_{i,j} = 1$ to denote that the j th segment is sent from sender i

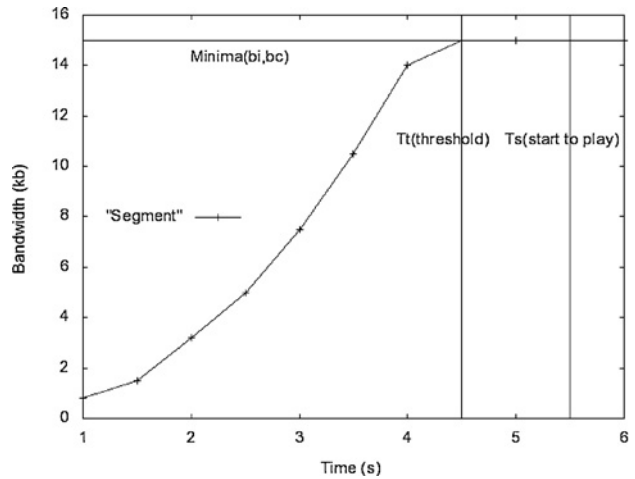


Fig. 5. Data were received at the average receiving speed.

SIP UPDATE message. By transmitting the QoS parameters via the SIP message, the session data can be transmitted by the session description protocol [12].

C. Service/Application Layer

In the IMS transmission mode, the receiver used the E2E network bandwidth to complete the segment data reception and reduce the network flow as possible. Regarding the satisfaction of the above conditions, this paper intended to discuss in the perspective of the segment and the E2E network architecture. Table I illustrates the description and meaning of the required variables.

To reduce the network traffic load and prevent congestion, this paper did not use the Greedy method to obtain the

segments, and instead, the data were received at the average receiving speed, as shown in Fig. 5. T_s denotes the starting time of the segment playback, while T_t denotes that the segment should be immediately transmitted to the thread hold; the time calculation is $T_t = J_b / \text{Min}(b_i, b_c)$. Prior to T_t , the transmission speed of the segment from the start to the end of playback is the calculated red curve $T_b = J_b / (T_s - T)$.

After storing the segment received by each node in the buffer, upon the request of other nodes, the sender would feed back to the client information regarding connection that allowed the client to calculate the time of data acquisition. After the completion of data reception, the sender would disconnect and reclaim the bandwidth.

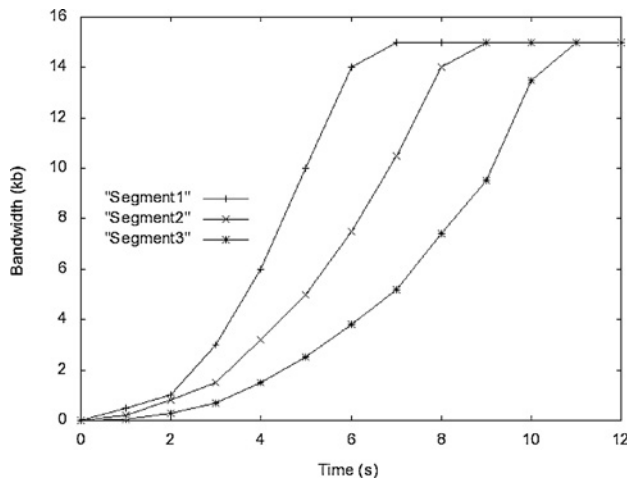


Fig. 6. Required transmission time for each segment.

TABLE II
PSEUDO-CODE FOR PLAYBACK-RATE BASED STREAMING
SERVICES (PRBSS)

```

01 procedure PRBSS( $I, J, T$ )
02 begin
03 Step1:
04  $T_b = J_b / (T_s - T)$ ;
05 if  $T < T_i$ 
06 Priority =  $(1 / b_s) * (\text{num}(I) / \text{num}(m_{i,j})) * (T_b)$ ;
07 else if  $T = T_i$ 
08 Priority is maximum;
09 else
10 Priority is Zero (Fail);
11 endif
12 Step2:
13 Update  $r_{xij}$ ;
14  $b_c = b_c \bar{T}_b$ ;
15 Step3:
16 Detect the biggest Priority of senders is which has cached the  $j$ th
17 segment;
18 Step4:
19 Select Sender  $i$  has maximum( $b_i$ );
20  $x_{ij} = 1, r_{xij} = 1$ ;
21 Step5:
22 While  $\Sigma T_b > b_s$ 
23 if the  $j$ th segment of  $i$ th server could be transmitted before  $T_i$ ;
24  $T_b = J_b / (T_s - T_i)$ ;
25  $m_{i,j} = 1, r_{xij} = 1$ ;
26  $b_c = b_c \bar{T}_b$ ;
27 else
28 Update  $b_c$ ;
29 endif
30 end

```

1) *E2E network architecture*. As the connection between the sender and the receiver has been established in the IMS layer, the receiver can immediately receive the media content evenly prior to playback, and according to the individual state of each segment. By obtaining the transmission curve of each segment through an algorithm, this paper can calculate the required transmission speed of all segments in the cache of the sender. Fig. 6 illustrates the required transmission time for each segment at the moment and Table II shows the algorithm proposed in this paper.

In Step 1, the receiver first calculated the priority of the segments not yet received. When the time arrived at the thread

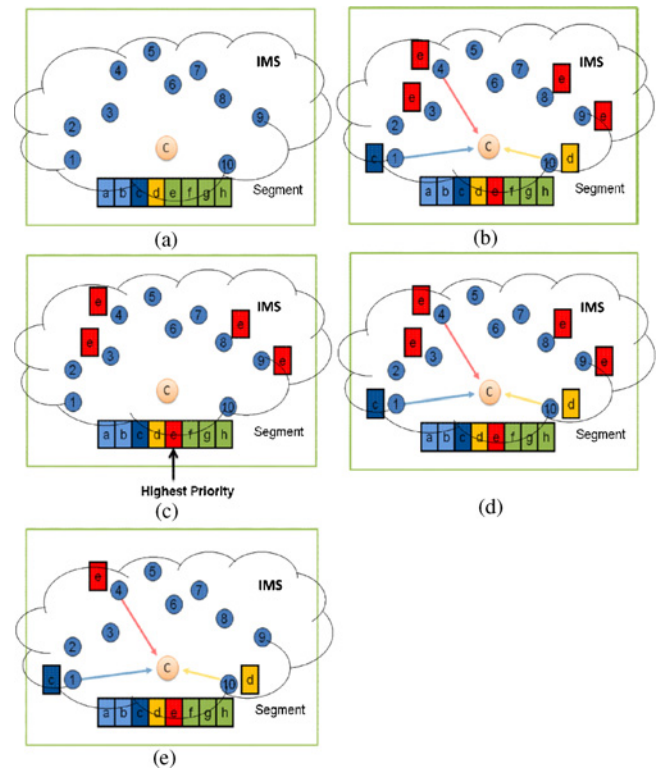


Fig. 7. Illustrations of playback-rate based streaming services. (a) Calculating the priorities of segments not yet received. (b) Checking which senders the client is receiving data from. (c) Choosing the highest priority segment to receive. (d) Receiving the highest priority segment from the maximum bandwidth node. (e) Restarting to calculate the priorities of segments not yet received.

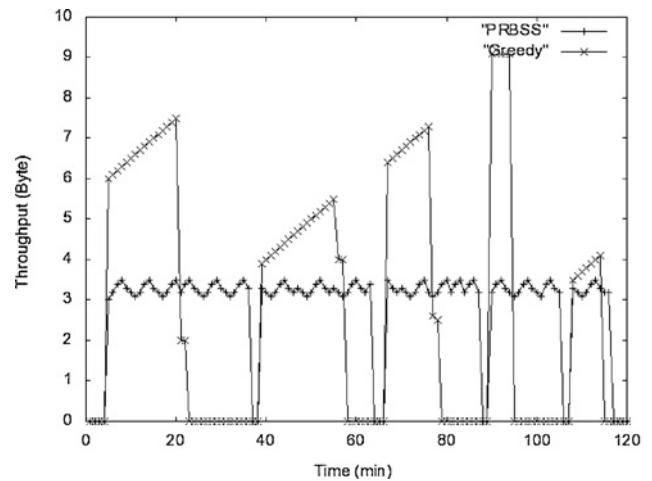


Fig. 8. Peak bandwidth of PRBSS and AVG.

hold (T_i), it meant the segment should be transmitted at the maximum connection speed due to the bandwidth limitation of the sender or the receiver. When the time was beyond T_i , it indicated that the segment could not arrive in time and would not be received due to b_i or b_c limitations. When the time has not arrived at T_i , the segment priority would be calculated, as based on current conditions, such as the available bandwidth of the sender and the distance to the thread hold. As shown in Fig. 7(a), the client has received segments a, b, c, and d, and started to calculate the priority of e, f, g, and h.

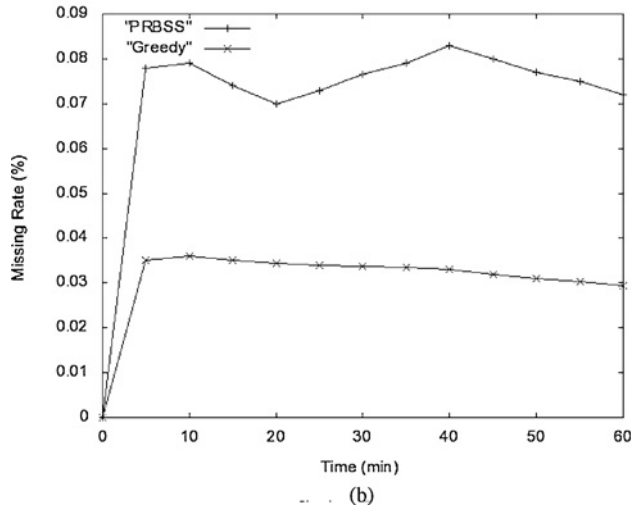
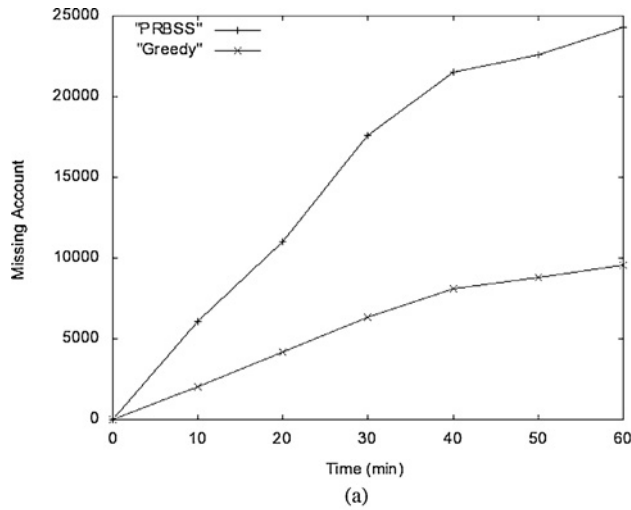


Fig. 9. Lost packets of the PRBSS and Greedy approach. (a) Missing amount. (b) Missing rate.

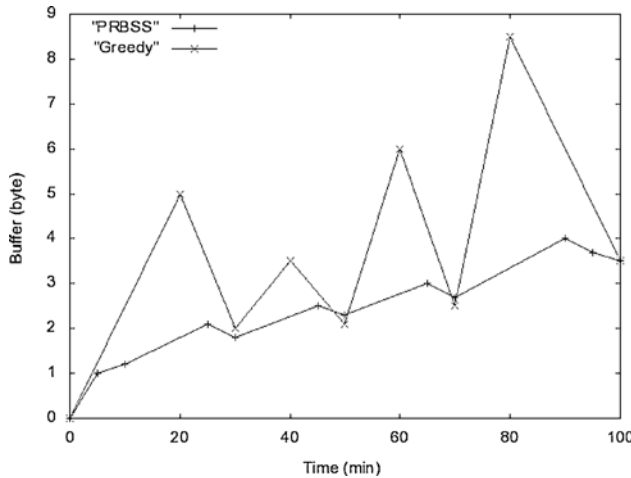


Fig. 10. Buffer size of PRBSS and the Greedy approach.

In Step 2, the data source of the receiver was checked and the available bandwidth of the receiver was calculated. As shown in Fig. 7(b), the receiver was receiving segments c and d from the sender. The downloading bandwidth of the receiver could be obtained by deducting the two connections. In Step 3, the segment j of the highest priority from the segments

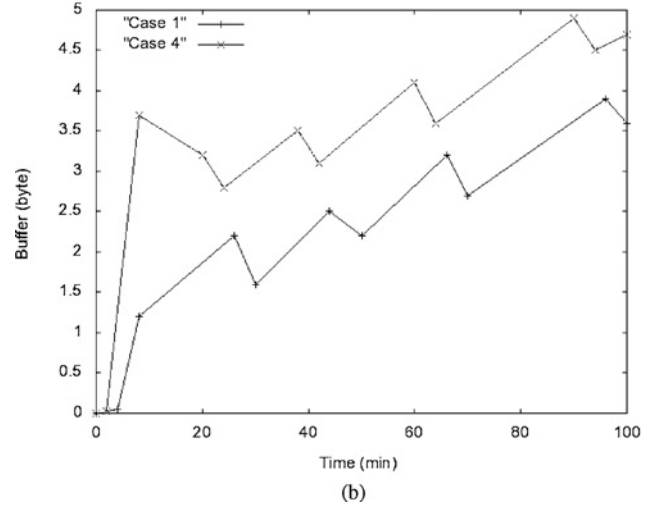
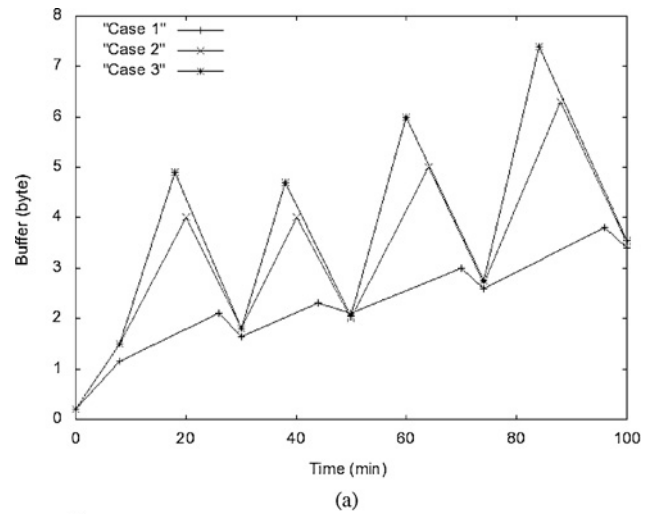


Fig. 11. Peer buffer size of multi-case. (a) Impact of α on buffer size. (b) Impact of β on buffer size.

that have not been received was selected. As shown in Fig. 7(c), after the calculation of segments e, f, g, and h, the priority of the segment e was the highest for the follow-up reception. In Step 4, the transmission speed T_b of the senders that have segment j was recalculated, and segment j was added into the current receiving list. As shown in Fig. 7(d), nodes 3, 4, 8, and 9 contained segment e. After calculations, node 4 could provide the maximum T_b to the client. In Step 5, whether the bandwidth of the current receiver was greater than the average transmission bandwidth was judged. If it was broader than the transmission bandwidth, it would wait until some transmissions have been ended in order to render the bandwidth smaller than the average transmission bandwidth. If it arrived at the T_i of a certain segment, transmission must be complete. If it was narrower than the transmission bandwidth, it would return to Step 1 to receive other segments. As shown in Fig. 7(e), the process restarted to calculate the priority of segments f, g, and h.

IV. PERFORMANCE EVALUATION

Fig. 8 illustrates the bandwidth use of PRBSS and the Greedy approach. Each network environment has its own

maximum transmission and reception capabilities. The figure illustrates the bandwidth use of the Greedy and PRBSS approach. The Greedy algorithm employed the maximum bandwidth of the node to transmit data; hence, there were many peaks, as shown in the figure. However, the PRBSS method transmitted data according to the playback speed; therefore, the transmission was more stable without the occurrence of peaks, as shown in the figure.

Fig. 9(a) and (b) demonstrates the lost packets of the PRBSS and Greedy approach. Due to the transmission characteristics of the Greedy method, it transmitted data when the current network environment allowed; however, the PRBSS transmitted data according to the playback speed. Hence, the packet loss by the Greedy approach would be less serious than that of the PRRSA method. Regarding the overall network packet loss, it could be seen that the PRBSS had relatively slighter impact on the online playback despite the higher packet loss rate.

Fig. 10 illustrates the buffer use of the PRBSS and the Greedy approach. As the Greedy approach utilized the maximum transmission capability, the receiving end must reserve a large buffer for the incoming data. In contrast, the PRBSS transmitted the data according to the transmission rate, and thus, the buffer would be smaller and node storage space could be saved.

Fig. 11(a) and (b) illustrates the comparison of the different α and β values of cases 1, 2, 3, and 4. In the case of various bandwidths, the maximum transmission and receiving capabilities of the network environments were compared. Fig. 7(a) illustrates the buffer use of cases 1, 2, and 3; in the cases of different α values, a greater α value represented a greater buffer was needed, with more buffer consumption. Fig. 7(b) demonstrates the buffer use of cases 1 and 4; in the cases of different β values, smaller β value represented an earlier time of buffer transmission to obtain packets earlier, with more buffer consumption.

V. CONCLUSION

While serving a real-time streaming service in IMS, a good segment scheduling algorithm should be considered the bandwidth to reduce the bandwidth peaks in IMS. Although previous real-time segment scheduling algorithms can maximize the received number of segments and minimize delay in segment transmission, they induce high peak bandwidth very frequently. Frequent occurrences of high peak bandwidth make the network unstable and adversely affect the performance of the network. They also suffer from large peer buffer size for storing not played segments. In this paper, we proposed cross layer playback-rate based streaming services architecture, for real-time streaming services in IP multimedia subsystem. Taking the playback-rate of the real-time streaming service into consideration, PRBSS schedules the segments of the service evenly transmitted into the network. As a consequence of the consideration, the limited bandwidth in the network can be used more efficiently. Through the simulation results, we showed that PRBSS outperforms the Greedy-based algorithm by down to 45%, 60%, and 33% in terms of peak bandwidth,

buffer size, and segment missing rate. While in streaming services testing, through experimental proof with the algorithm designed in this research it is found the relation among buffer size, missing rate, and bandwidth. In this paper, we did not consider the characteristic of video codec. As for future works, we will research streaming services with the video codecs for preventing the occurrence of peak in IMS precisely.

REFERENCES

- [1] G. Camarillo, T. Kauppinen, M. Kuperinen, and I. M. Ivars, "Toward an innovation oriented IP multimedia subsystem," *IEEE Commun. Mag.*, vol. 45, no. 3, pp. 130–136, Mar. 2007.
- [2] F. Baroncelli, B. Martini, V. Martini, and P. Castoldi, "Supporting control plane-enabled transport networks within ITU-T next generation network (NGN) architecture," in *Proc. NOMS*, 2008, pp. 271–278.
- [3] M. I. Corici, F. C. de Gouveia, and T. Magedanz, "A network controlled QoS model over the 3GPP system architecture evolution," in *Proc. AusWireless*, 2007, p. 39.
- [4] R. Good and N. Ventura, "End to end session based bearer control for IP multimedia subsystems," in *Proc. IM*, 2009, pp. 497–504.
- [5] M. Ageal, R. Good, A. Elmagosh, M. Ashibani, N. Ventura, and F. Ben-Shatwan, "Centralized policy provisioning for inter-domain IMS QoS," in *Proc. EUROCON*, 2009, pp. 1793–1797.
- [6] S. Tompros, C. Kavadias, D. Vergados, and N. Mouratidis, "A strategy for harmonized QoS manipulation in heterogeneous IMS networks," *Wirel. Personal Commun.*, vol. 49, no. 2, pp. 197–212, Aug. 2008.
- [7] L. Lifeng and L. Gang, "Cross-layer mobility management based on mobile IP and SIP in IMS," in *Proc. WiCom*, 2007, pp. 803–806.
- [8] M. A. Melnyk, A. Jukan, and C. D. Polychronopoulos, "A cross-layer analysis of session setup delay in IP multimedia subsystem (IMS) with EV-DD wireless transmission," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 869–881, Jun. 2007.
- [9] K. Kapitanova, Y. Wei, W. Kang, and S. Son, "Applying formal methods to modeling and analysis of real-time data streams," *J. Comput. Sci. Eng.*, vol. 5, no. 1, pp. 85–110, Mar. 2011.
- [10] G. Camarillo and M. A. Garcia-Martin, *The 3G IP Multimedia Subsystem, Merging the Internet and Cellular Worlds*. New York: Wiley, 2004.
- [11] J. Rosenberg, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler, *SIP: Session Initiation Protocol*, RFC 3261, IETF, Jun. 2002.
- [12] M. Handley and V. Jacobson, *SDP: Session Description Protocol*, RFC 2327, IETF, Apr. 1998.
- [13] K. B. Johansson and D. C. Cox, "QoS scheduling of mixed priority non-real-time traffic," in *Proc. VTC*, vol. 4, 2001, pp. 2645–2649.
- [14] J. Chen, T. Lv, and H. Zeng, "Joint cross-layer design for wireless QoS content delivery," in *Proc. ICC*, vol. 7, 2004, pp. 4243–4247.
- [15] N. Miguel, D. Yezid, and R. Viviana, "An IMS architecture with QoS parameters for flexible convergent services," in *Proc. ISCC*, 2010, pp. 640–645.
- [16] U. Iqbal, Y. Javed, S. Rehman, and A. Khanum, "SIP-based QoS management framework for IMS multimedia services," *Int. J. Comput. Sci. Netw. Security*, vol. 10, no. 5, pp. 181–188, May 2010.
- [17] S. R. Yang and W. T. Chen, "SIP multicast-based mobile quality-of-service, support over heterogeneous IP multimedia subsystems," *IEEE Trans. Mobile Comput.*, vol. 7, no. 11, pp. 1297–1310, Nov. 2008.
- [18] M. S. Siddiqui, S. O. Amin, and C. S. Hong, "A set-top box for end-to-end QoS management and home network gateway in IMS," in *Proc. ICCE*, 2009, pp. 1–2.
- [19] B. Raouyane, M. Bellafkih, and D. Ranc, "Qos management in IMS: DiffServ model," in *Proc. NGMAST*, 2009, pp. 39–43.
- [20] F. Xu, L. Zhang, and Z. Zhou, "Interworking of WIMAX and 3GPP networks based on IMS," *IEEE Commun. Mag.*, vol. 45, no. 3, pp. 144–150, Mar. 2007.
- [21] K. Knightson, N. Morita, and T. Towle, "NGN architecture: Generic principles, functional architecture, and implementation," *IEEE Commun. Mag.*, vol. 43, no. 10, p. 496, Oct. 2005.

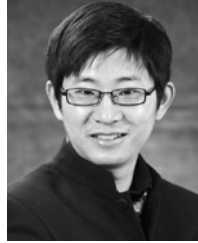
- [22] L. Zhou, N. Xiong, L. Shu, A. Vasilakos, and S.-S. Yeo, "Context-aware multimedia service in heterogeneous networks," *IEEE Intell. Syst.*, vol. 25, no. 2, pp. 40–47, Mar.–Apr. 2010.
- [23] T. K. Lee, T. Y. Chai, L. H. Ngoh, X. S. Teo, and J. C. M. L. Zhou, "An IMS-based testbed for real-time services integration and orchestration," in *Proc. APSCC*, 2009, pp. 260–266.
- [24] W. A. Aziz, S. H. El-Ramly, and M. M. Ibrahim, "IP multimedia subsystem (IMS) performance evaluation and benchmarking," in *Proc. SIBIRCON*, 2010, pp. 209–214.
- [25] L. Zhou, X. Wang, W. Tu, G. Mutean, and B. Geller, "Distributed scheduling scheme for video streaming over multi-channel multi-radio multi-hop wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 409–419, Apr. 2010.
- [26] L. Zhou and Y. Zhang, "Distributed media-service scheme for P2P-based vehicular networks," *IEEE Trans. Vehicular Technol.*, vol. 60, no. 2, pp. 692–703, Feb. 2011.
- [27] E. A. Jammeh, M. Fleury, and M. Ghanbari, "Rate-adaptive video streaming through packet dispersion feedback," *IET Commun.*, vol. 3, no. 1, pp. 25–37, 2009.
- [28] M. N. Islam and R. Adve, "Transceiver design using linear precoding in a multiuser multiple-input multiple-output system with limited feedback," *IET Commun.*, vol. 5, no. 1, pp. 27–38, 2011.
- [29] N. Mokari and K. Navaie, "Cross-layer resource allocation for push to talk service over orthogonal frequency division multiple access-based networks with heterogeneous traffic," *IET Commun.*, vol. 4, no. 15, pp. 1838–1846, 2010.
- [30] M. H. Tsai, J. T. Sung, and Y. M. Huang, "Cross-layer resource allocation for push to talk service over orthogonal frequency division multiple access-based networks with heterogeneous traffic," *IET Commun.*, vol. 4, no. 9, pp. 1108–1115, 2010.



Chin-Feng Lai (M'09) received the Ph.D. degree from the Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan, in 2008.

He has been an Assistant Professor with the Institute of Computer Science and Information Engineering, National Ilan University, Ilan, Taiwan, since 2011. After receiving the Ph.D. degree, he has authored and co-authored over 60 refereed papers in journals, conference, and workshop proceedings about his research areas within two years. His

current research interests include multimedia communications, sensor-based healthcare, and embedded systems.



Min Chen (M'08–SM'09) has been an Assistant Professor with the School of Computer Science and Engineering, Seoul National University (SNU), Seoul, Korea, since September 2009. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of British Columbia (UBC), Vancouver, BC, Canada, for three years from March 2006. Before joining UBC, he was a Post-Doctoral Fellow with SNU for one and a half years.

Mr. Chen is a Symposia Co-Chair for IEEE ICC

2012.