This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

QoS Multicasting over Mobile Networks

Ai-Chun Pang*+ and Shun-Mao Wang+
Graduate Institute of Networking and Multimedia*,
Dept. of Comp. Sci & Info. Engr.+, National Taiwan University
Taipei, Taiwan, R.O.C.; Email: acpang@csie.ntu.edu.tw

Shiao-Li Tsao
Dept. of Comp. Sci & Info. Engr., National Chiao Tung University
Hsinchu, Taiwan, R.O.C.

***Abstract*— Based on multimedia broadcast/multicast service architecture(MBMS), this paper proposes an efficient QoS-based multicast approach for the Universal Mobile Telecommunications System (UMTS). Our approach adopts the existing scalable-coding technique to provide multiple levels of service quality to diverse mobile devices in the UMTS system. In this approach, two kinds of transmission modes are developed to utilize fully the network resources for wireline and wireless links. An analytic model is presented to investigate the performance of our approach and the 3GPP 23.246 approach. The numerical results indicate that in terms of transmission costs of core/radio networks, our approach outperforms the 3GPP 23.246 approach.**

I. Introduction

The existing point-to-multipoint (i.e., braodcasting and multicasting) services for wireline networks (e.g., Internet) allow transmission of data from a single source entity to multiple recipients. With the rapid growth of wireless/mobile subscribers, these services are expected to be used extensively over wireless/mobile networks. Furthermore, as multimedia applications (e.g., video streaming) are ubiquitous in the Internet world, multimedia broadcasting and multicasting is considered one of the most important services in 3G and future wireless/mobile communication systems.

3GPP 22.146 [2] has defined a *Multimedia Broadcast/Multicast Service* (MBMS) for *Universal Mobile Telecommunications System* (UMTS) networks. In 3GPP MBMS, two modes of operations are supported. The broadcast mode is a unidirectional point-to-multipoint transmission of multimedia data from a single source entity to all users in a broadcast service area. On the other hand, the multicast mode only transmits multimedia data to a specific group of mobile users (i.e., the multicast members). Both the broadcast and multicast modes are intended efficiently to use radio/network resources, which is achieved by the multicast tables of the networks nodes such as GGSN (Gateway GPRS Support Node), SGSN (Serving GPRS Support Node) and RNC (Radio Network Controller) described below. The multicast tables record the distribution of broadcast/multicast users among the service areas so that the multimedia data is only transmitted to the areas where the broadcast/multicast users reside [2], [6].

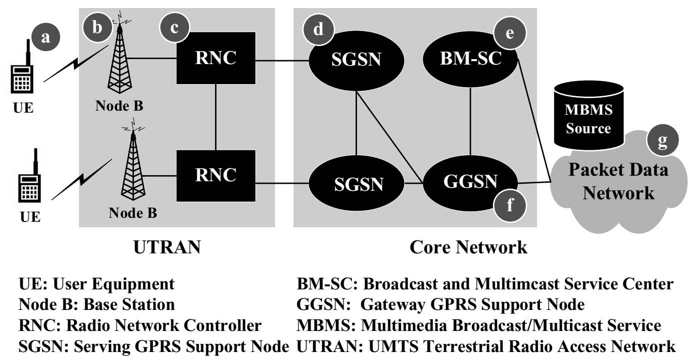

Fig. 1. The 3GPP MBMS Architecture

Figure 1 shows an example of MBMS architecture for UMTS networks [5]. UMTS networks consist of both a UTRAN(UMTS Terrestrial Radio Access Network) and a Core Network. This UMTS network connects to the *Packet Data Network* (PDN; see Figure 1 (g)) through the Core Network's SGSN (see Figure 1 (d))and the GGSN (see Figure 1 (f)). The SGSN connects to the radio access network. The GGSN provides interworking with the external PDN. GGSN is connected with SGSNs via an IP-based GTP (GPRS Tunneling Protocol) network.

To support MBMS, a new network node, *Broadcast and Multicast Service Node* (BM-SC; see Figure 1 (e)), is introduced to provide MBMS access control for mobile users. BM-SC connects to the GGSN via IP-based Gmb interface, and communicates with MBMS source located in the external PDN for receiving multimedia data. The *UMTS Terrestrial Radio Access Network* (UTRAN) consists of *Node B*s (the UMTS term for base stations; see Figure 1 (b)) and RNCs (see Figure 1 (c)). A *user equipment* (UE) or mobile device (see Figure 1 (a)) communicates with one or more Node Bs through the radio interface based on the Wideband CDMA technology [3].

As the number of mobile devices and the kinds of mobile applications explosively increases, the device types become diverse, and mobile networks are prone to be “Heterogeneous”. Broadcast/Multicast users with different kinds of mo-

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE GLOBECOM 2005 proceedings.

bile devices may request different quality levels of multimedia streams due to (1) users' preferences, (2) service charges, (3) network resources, and (4) device capabilities. Thus effectively delivering different quality levels of content to a group of users who request different QoS streams is quite challenging. This paper proposes an efficient QoS-based multimedia broadcasting/multicasting approach to transmit multimedia streams to broadcast/multicast users requesting different levels of service.

Our approach adopts the existing scalable-coding technique to provide multi-layered multimedia transmission for UMTS MBMS. Applying the scalable-coding technique to wireless transmission is intensively studied in the literature. In particular, Yang et. al propose a TCP-friendly streaming procotol, WMSTFP, which, over wireless Internet, reduces packet loss improving the system throughput. The issues for power consumption and resource allocation over wireless channels have also been investigated [9], [4], [10]. However, little work has been done in multimedia broadcasting/multicasting with scalable-coding support.

This paper is organized as follows. Section 2, describes the 3GPP 23.246 approach, and proposes a QoS-based multimedia multicasting approach for heterogeneous mobile networks. An analytic model is presented in Section 3 to investigate and compare these approaches. Section 4 uses numerical examples to evaluate the performance of the 3GPP 23.246 approach and our QoS-based multimedia multicasting approach. Section 5 concludes the paper.

II. QoS-Based Multimedia Multicasting for UMTS Networks

To support MBMS for mobile devices with diverse capabilities, the 3GPP 23.246 [1] proposes a multimedia multicasting[1] approach for UMTS networks. We assume that four QoS levels based on data rates(i.e., 32Kbps, 64Kbps, 96Kbps and 128Kbps) for multimedia streaming are provided to mobile multicast users.In this approach (Approach I; see Figure 2), multimedia (e.g., video and audio) streams are duplicated and encoded as different QoS levels at MBMS source. Then based on users' QoS profiles in the multicast tables maintained by GGSN, SGSNs and RNCs, the encoded video/audio streams of each QoS level are respectively transmitted to the multicast users requesting that quality. As shown in Figure 2, there are two SGSNs in the UMTS network: SGSN1 and SGSN2. SGSN1 covers routing areas RA1, RA2 and RA3[2]. SGSN2 covers routing areas RA4, RA5 and RA6.

To perform QoS-based multicasting in this approach, MBMS source duplicates multimedia streams, and encodes the duplicated streams at four levels. The four encoded streams are transmitted to the GGSN, and based on the multicast table, the GGSN forwards each stream to the SGSNs covering the

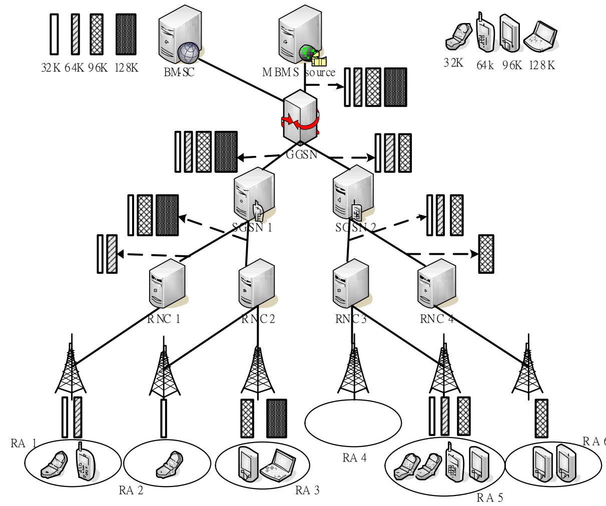

Fig. 2. The 3GPP 23.246 Approach I

multicast users with that request. In Figure 2, the streams of 32Kbps, 64Kbps and 96Kbps (through three GTP tunnels) are delivered to SGSN1, and SGSN2 receives four QoS level streams via four GTP tunnels. Similarly, the SGSNs relays the proper streams to the appropriate RNCs, and then to the RAs through radio channels. By using the 3GPP 23.246 approach, the transmitted streams fulfill the QoS level requested by each multicast user.

However, as the number of supported QoS levels increases (i.e., the types of mobile devices increase making the networks more "heterogeneous"), data duplication becomes more serious, which results in more resource consumption at both core and radio networks. Based on standard the 3GPP MBMS architecture [2], we propose more efficient multimedia multicasting approach (Approach II) to deliver scalable-coded multimedia to multicast users requesting specific levels of multimedia quality.

The goals of our multimedia multicasting approach are (1) to have a single multimedia stream source (i.e., no duplication at the MBMS source), (2) to transmit multimedia streams to all members in the multicast group with satisfactory quality, and (3) to utilize effectively the resources of core and radio networks. To achieve these goals, the existing scalable-coding technique is adopted to deliver multimedia streams. Figure 3 elaborates the basic concept for scalable coding. The scalable coding technique utilizes a layered coder to produce a cumulative set of layers where multimedia streams can be combined across layers to produce progressive refinement.

For example, if only the first layer (or base layer) is received, the decoder will produce the lowest quality version of the signal. If the decoder receives two layers, it will combine the second layer (or the enhancement layer) information with the first layer to produce improved quality. Overall, the quality progressively improves with the number of layers that are re-

[1]Broadcasting is a special case of multicasting.

[2]We assume that each RA is covered by one Node B.

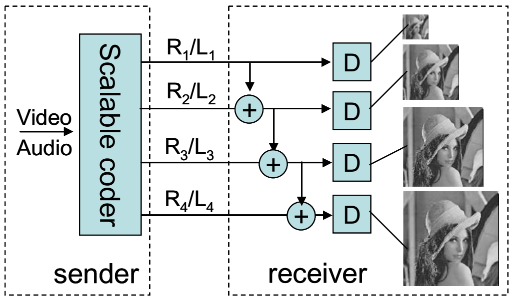

Fig. 3. The Scalable Coding Technique

ceived and decoded. With scalable coding, the requirement of single-source multimedia streams is fulfilled, and all multicast users can decode their preferred multimedia packets depending on the devices' capabilities.

However, effectively utilizing the resources of core and radio networks to transmit scalable-coded multimedia streams is still a challenging issue. We develop two transmission modes for our scalable-coding enabled multimedia multicasting: "Packed" mode Approach II_A and "Separate" mode Approach II_B. In the packed mode (see Figure 4), all layered multimedia data for one frame are packed into a packet at MBMS source. Then these packed packets are sequentially delivered in one shared tunnel (between GGSN and SGSN, and between SGSN and RNC) and then in one shared radio channel to all multicast users. Each packed packet consists of one frame of 4-layered multimedia data,as shown in Figure 4, packets are sent from GGSN to the SGSNs (i.e., SGSN1 and SGSN2), the RNCs (i.e., RNC1, RNC2, RNC3 and RNC4) and then the RAs (i.e., RA1, RA2, RA3, RA5 and RA6) where the multicast users reside. Upon receipt of 4-layered multimedia data, the multicast users select certain layers to decode based on their preferences.

For Approach II_A, our QoS-based multimedia multicasting can be easily implemented in UMTS networks without any modification of the existing GGSN, SGSNs and RNCs. Since the GGSN, SGSNs and RNCs are not aware of scalable coding and can not differentiate the layers of multimedia streams, 4-layered multimedia streames must be sent to all multicast users regardless of the users' request. This may result in extra resource (i.e., link bandwidth and channelization code) usage at the core and radio networks. Also, this leads to increased power consumption caused by mobile devices (e.g., the mobile phone in RA2) requesting low-quality multimedia streams.

Therefore "Separate" mode (Approach II_B) is developed to improve the transmission efficiency of scalable coded multimedia streams. Figure 5 shows the scenario of "Separate" mode for scalable-coded multimedia multicasting. In Approach II_B, each layered multimedia datum is encapsulated in one GTP packet, and all GTP packets are transmitted through a single tunnel. To deliver effectively the scalable coded multimedia streams, the GGSN, SGSNs and RNCs must

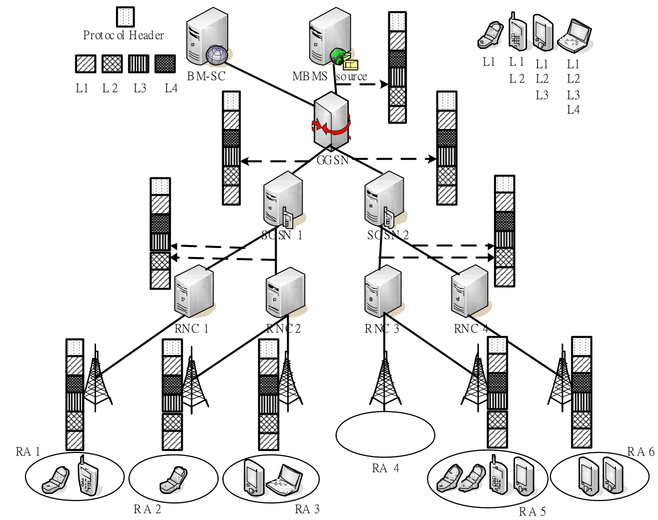

Fig. 4. Transmission Approach II_A for Our QoS-based Multimedia Multicasting

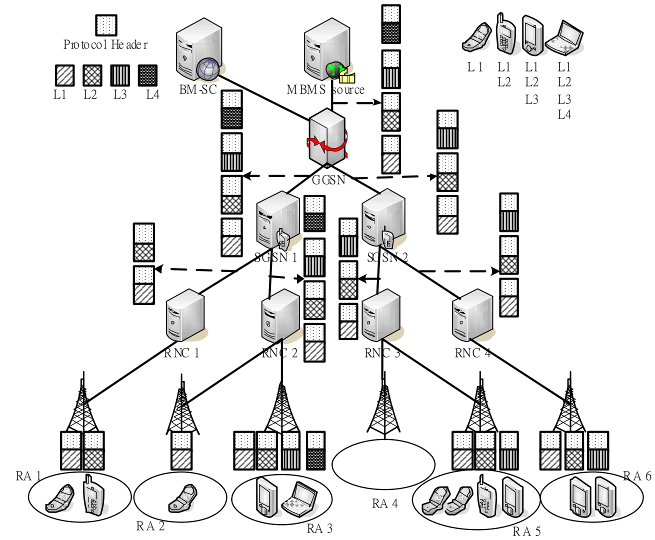

Fig. 5. Transmission Approach II_B for Our QoS-based Multimedia Multicasting

be modified to become aware of scalable coding. Importantly, these network nodes do not have to understand how scalable coding works. They only need to differentiate the layers of received multimedia streams, which can be accomplished through the tag of GTP packet headers.

Since the layer-differentiation can be done by the RNCs, each layer stream would be transmitted by one radio channel, and mobile devices can freely select and receive their preferred layers. This results in a significant reduction of the power consumption from mobile devices and the channel usage in radio networks. Based on the above discussion, Table I compares our proposed QoS-based multimedia multicasting approach (Approach II_A and Approach II_B) with 3GPP 23.246 Approach I. The following issues are addressed.

TABLE I
COMPARING OUR PROPOSED QOS-BASED MULTIMEDIA MULTICASTING WITH 3GPP 23.246 APPROACH

Issues \ Approaches	Approach I (3GPP 23.246)	Approach II_A	Approach II_B
Issue 1: QoS Maintenance for Multicast Users	Yes	No	Yes
Issue 2: Single Source for Heterogeneous Devices	No	Yes	Yes
Issue 3: Synchronization Problem (for UE)	No	No	Yes
Issue 4: Adaptation to Bandwidth Variation	No	No	Yes

1. QoS Request maintenance for Multicast Users :
   Both 3GPP 23.246 Approach I and Approach II_B select the multicasting path for a specific quality of multimedia streams based on the users' QoS profile. Thus the network nodes such as GGSN, SGSNs and RNCs in these two approaches have to maintain the QoS requests of mobile users. However, since all scalable-coded layers of multimedia streams are delivered to the multicast users, QoS maintenance for multicast users is not needed in Approach II_A.
2. Single Source for Heterogeneous Devices :
   For Approach I, the multimedia streams have to be duplicated and encoded as different qualities at MBMS source. On the other hand, since the scalable coding technique is used in Approach II, duplication can be avoided.
3. Synchronization Problem(for UE) :
   For Approach II_B, UEs may receive multiple layers of multimedia streams through several channels, which results in the synchronization problem between the received layered streams.
4. Adaptation to Bandwidth Variation :
   Approach II_B is capable of adapting to bandwidth variation especially for the bandwidth reduction of wireless links. When the bandwidth suddenly reduces, the transmission of multimedia streams for high quality can be temporarily suspended. At this point, the mobile devices with ongoing high-quality multimedia transmission can still receive low-quality streams without causing service interruption. This can not be achieved through Approach I and Approach II_A.

III. ANALYTIC MODELING

This section presents an analytic model to investigate the 3GPP 23.246 Approach I and our QoS-based multimedia multicasting Approach II including both Mode A and Mode B. Without loss of generality, we assume that there is one GGSN in the UMTS system. Let N_S be the number of SGSNs, K be the number of RNCs covered by each SGSN, and M be the number of Node Bs covered by each RNC. Thus the total numbers (N_R and N_B) of RNCs and Node Bs are respectively N_SK and N_SKM.

This analytic model is analyzed in terms of transmission cost of core and radio networks. The transmission costs (C_t) for these two approaches are measured by the following weighted function of bandwidth requirement of multimedia transmission for core and radio networks.

$$C_t = B_gC_g + B_sC_s + B_rC_r + B_bC_b, \quad (1)$$

where B_g, B_s, B_r and B_b respectively represent the total bandwidth requirements for multimedia multicasting between the GGSN and the SGSNs, between the SGSNs and RNCs, between the RNCs and Node Bs, and between the Node Bs and UEs. Similarly, C_g, C_s, C_r and C_b respectively denote the unit transmission costs between the GGSN and the SGSNs, between the SGSNs and RNCs, between the RNCs and Node Bs, and between the Node Bs and UEs. The weighting factors (i.e., C_g, C_s, C_r and C_b) are used since the link bandwidth of core networks is much larger than that of radio networks. Thus the unit transmission costs in core networks are generally lower than the equivalent costs in radio networks. Furthermore, the transmission cost over a wireless link is significantly higher than over a wireline link. From [7], the values of C_g, C_s, C_r and C_b are set to 0.2, 0.2, 0.5 and 5. To compute the total bandwidth requirements B_g, B_s, B_r and B_b, two classes of RAs are considered. Class 1 RAs cover complicated areas with dense population, and thus with diverse mobile devices. On the other hand, simple RAs (Class 2) have a uni-type of mobile devices. Let α be the portion of class 1 RAs, and assume that class 1 and class 2 RAs are uniformly distributed in the UMTS system. Note that our model can be easily extended to analyze other distributions of class 1 and class 2 RAs. Due to the page limitation, details of our analytic model are omitted, and the reader is referred to http://www.csie.ntu.edu.tw/~acpang/TR.htm.

IV. PERFORMANCE EVALUATION

Based on the analytic model in the previous section, we use some numerical examples to evaluate the performance of the 3GPP 23.246 Approach I, and QoS-based multimedia multicasting Approach II_A and Approach II_B. In our experiments, Foreman is used for test sequences, and the number of frames (with the size of 176x144 QCIF) is 400. MPEG-4 FGS and MPEG-4 respectively are used for scalable coding and non-scalable coding, and Codec adopts Microsoft MPEG-4 Reference Software [8]. Furthermore, uni-truncation (with equivalent bit-rate) is used for all enhancement layers of I-Frame and P-Frame. Six levels of service quality are provided in the experiments. For non-scalable coding, the six quality levels are defined by 120Kbps, 150Kbps, 180Kbps, 210Kbps, 240Kbps and 270Kbps bit rates. To have similar quality levels (i.e., similar PSNR values) for scalable coding, some experiments for quality measurement are made. For these experimental results, the reader is referred to

TABLE II
INPUT PARAMETERS

Variable	Description	Value
N_S	The number of SGSNs	10
K	The number of RNCs covered by each SGSN	10
M	The number of Node Bs covered by each RNC	50
n	The number of QoS levels	6
T	Playing time for test sequences	13.3sec
L_U	Header lengths of UDP	8 bytes
L_I	Header lengths of IP	20 bytes
L_G	Header lengths of GTP	12 bytes
L_P	Header lengths of PDCP	3 bytes

http://www.csie.ntu.edu.tw/~acpang/TR.htm. The results indicated that the bit rate of based layer ($L1$) for scalable coding would be 120Kbps, and the bit rates for accordingly enhancement layers (i.e., $L2$, $L3$, $L4$, $L5$ and $L6$) are 150Kbps, 120Kbps, 105Kbps, 90Kbps and 75Kbps. Table II shows the input parameters and their values used in our experiments. Figure 6 indicates the effect of α (i.e., portion of class 1 RAs covering diverse mobile devices) on the transmission costs C_T for Approach I, Approach II_A and Approach II_B. In this figure, the C_T value for Approach II_A remains the same as α increases (i.e., the number of complicated areas increase). On the other hand, the increase of α results in the increase of C_T for Approach I and Approach II_B. Specifically, the increasing rate for Approach I is much larger than that for Approach II. Furthermore, when $\alpha > 40\%$, Approach II (for both Mode A and Mode B) has a smaller C_T than Approach I. From this figure, we observe that when all RAs are class 1, Approach II_A has the lowest C_T. However, when α nearly equals 0, the performance of Approach I is better than that either Approach II_A or II_B. Also, this figure indicates that as t increases from 30ms to 90ms, the overhead for Approach II_B decreases marginally, and thus C_T slightly decreases.

V. CONCLUSIONS

As multimedia applications (e.g., video streaming) are ubiquitous around the Internet world, multimedia broadcasting/multicasting is considered one of the most important services in 3G and in future wireless/mobile communication systems. To support multimedia broadcast/multicast service for mobile devices with diverse capabilities, the 3GPP 23.246 proposed a multimedia multicast approach for *Universal Mobile Telecommunications System* (UMTS) networks. However, the multimedia streams for this approach have to be duplicated and encoded as different qualities at MBMS source, which results in significant resource consumption of core/radio networks. Thus based on the 3GPP multimedia broadcast/multicast service architecture, this paper proposes an efficient QoS-based multicast approach for UMTS. Our approach adopts the existing scalable-coding technique to provide multiple levels of service quality to diverse mobile devices in the UMTS system.

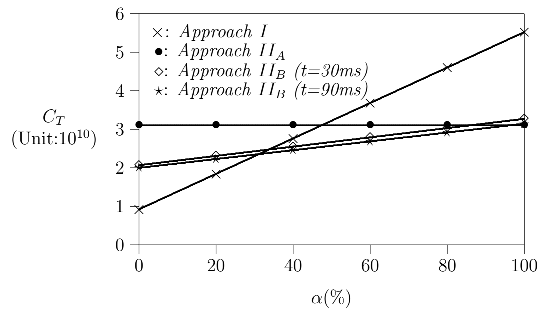

Fig. 6. Effect of α on C_T

In this approach, two transmission modes are developed to fully utilize the network resources for wireline and wireless links. An analytic model investigates the performance of our two approach modes and the 3GPP 23.246 approach. The numerical results indicate that in terms of transmission costs of core/radio networks, Approaches II_A and II_B outperform the 3GPP 23.246 Approach I.

VI. ACKNOWLEDGMENTS

We would like to thank Prof. Chia-Wen Lin and Mr. Jia-Ming Tsai for their valuable comments to improve the quality of this paper. Pang's work was sponsored in part by National Science Council under contracts NSC94-2213-E-002-048 and NSC94-2213-E-002-019, Intel and CCL/ITRI.

REFERENCES

[1] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description (Release 6). Technical report, 3GPP.

[2] 3GPP. 3rd Generation Partnership Project; Technical Specification Group Services and Systems Aspects; Multimedia Broadcast/Multicast Service (Release 6). Technical Report 3G TS 22.146 version 6.6.0 (2004-09), 3GPP, 2004.

[3] Holma, H., and Toskala, A. (edited). *WCDMA for UMTS*. John Wiley & Sons, 2000.

[4] Lee, T.-W., Chan, S.-H., Zhang, Q., Zhu, W. and Zhang, Y.-Q. Allocation of Layer Bandwidths and FECs for Video Multicast over Wired and Wireless Networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(12):1059–1070, 2002.

[5] Lin, Y.-B., Huang, Y.-R., Pang, A.-C. and Chlamtac, Imrich. All-IP Approach for Third Generation Mobile Networks. *IEEE Network*, 16(5):8–19, 2002.

[6] Pang, A.-C., Lin, Y.-B., Tsai, H.-M., Agrawal, P. Serving Radio Network Controller Relocation for UMTS All-IP Network. *IEEE Journal on Selected Aarea in Communications*, 22(4), 2004.

[7] Rummler, R., Chung, Y.-W. and Aghvami, A.-H. Modeling and Analysis of Efficient Multicast Mechanism for UMTS. *IEEE Trans. on Vehicular Technology*, 54(1), 2005.

[8] Wang, S.-H, Tung, Y.-S., Wang, C.-N., Chiang, T. and Sun, H. AHG Report on Editorial Convergence of MPEG-4 Reference Software. Technical Report JTC1/SC29/WG11 MPEG2003/M9632, ISO/IEC, 2003.

[9] Zhang, Q., Zhu, W. and Zhang, Y.-Q. Power-Minimized Bit Allocation for Video Communication over Wireless Channel. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(6):398–410, 2002.

[10] Zhang, Q., Zhu, W. and Zhang, Y.-Q. Channel-Adaptive Resource Allocation for Scalable Video Transmission over 3G Wireless Network. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(8):1049–1063, 2004.

