WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2005; **5**:95–111
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/wcm.279

Scalable multiple description coding and distributed video streaming in 3G mobile communications

Ruobin Zheng, Weihua Zhuang*,† and Hai Jiang

Department of Electrical and Computer Engineering, Center for Wireless Communications, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Summary

This paper proposes a distributed multimedia delivery mobile network for video streaming in 3rd generation (3G) mobile communications. The joint design of layered coding (LC) and multiple description coding (MDC) is employed to address the bandwidth fluctuations and packet loss problems in the wireless network and to further enhance the error resilience tools in MPEG-4. A new Internet protocol (IP) differentiated services (DiffServ) video marking algorithm is presented to support an unequal error protection of the LC components. Both intra-RAN (radio access network) handoff and inter-RAN handoff procedures are discussed, which provide path diversity to combat streaming video outage due to handoff in the universal mobile telecommunications system (UMTS). Computer simulation results demonstrate that: (1) the newly proposed IP DiffServ video marking algorithm is more suitable for video streaming in an IP mobile network as compared with the previously proposed algorithm, and (2) the proposed handoff procedures have better performance in terms of handoff latency, end-to-end delay and handoff scalability than that in UMTS. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: video streaming; multiple description coding (MDC); layered coding (LC); handoff; universal mobile telecommunications system (UMTS); Internet protocol (IP); differentiated services (DiffServ)

1. Introduction

With the emergence of broadband wireless networks and increasing demand for multimedia information on the Internet, wireless video communications have received great interests from both industry and academia [1–4], and wireless multimedia services are foreseen to become widely deployed in this decade. Real-time transport of live video or stored video is the predominant part of real-time multimedia. Video streaming is the main approach for delivery of stored video over wireline networks such as the Internet [5–8], where the streaming video is partitioned into packets and played out simultaneously during video delivery. In comparison with video download, video streaming has the advantages of a low (initial) delay and requiring a small storage space. To provide quality of service (QoS) over future Internet, the

*Correspondence to: Weihua Zhuang, Department of Electrical and Computer Engineering, Center for Wireless Communications, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.
†E-mail: wzhuang@bbcr.uwaterloo.ca

Contract/grant sponsor: Natural Science and Engineering Research Council (NSERC) of Canada (Strategic Research Project); contract/grant number: STPGP 257682-02.

differentiated services (DiffServ) approach [9] has emerged as an efficient and scalable solution based on handling of limited traffic classes [10].

In this paper, we investigate video streaming over a hybrid cellular wireless (i.e. universal mobile telecommunications system, UMTS) and IP-based DiffServ wireline network. It is technically very challenging due to the hostile wireless propagation environment, user mobility, and dynamic nature of video traffic. Because of its real-time nature, video streaming typically has QoS requirements in bandwidth, delay and transmission error rate. However, unreliability, bandwidth fluctuations and high bit error rate of a wireless channel can cause severe video quality degradation. In a cellular network, the importance of seamless handoffs is well known; but it is largely unexplored in the applications of streaming media. In particular, issues associated with media streaming during seamless handoff include handoff latency (or media stream interruption), end-to-end delay (or service delivery time), media synchronization and handoff scalability. However, the handoff procedures in UMTS [11,12,40] may not satisfy the requirements of seamless handoff for media streaming services. Furthermore, it is challenging to provide QoS attribute translation and mapping between the wireline IP domain and the wireless UMTS domain for the end-to-end QoS provisioning.

To meet the challenges, we propose a distributed multimedia delivery mobile network (D-MDMN) model for video streaming over 3G wireless networks, where media streaming services are pushed to the edge of core network so that the streaming media is sent over a shorter network path. It reduces the media service delivery time, the probability of packet loss and the total network resource consumption with relatively consistent QoS. A UMTS-to-DiffServ QoS mapping scheme and its marking algorithm for MPEG-4 video are used to support the unequal error protection for layered video. The system employs a novel scalable multiple description coding (SMDC) framework, where video layered coding (LC) and multiple description coding (MDC) [16,17] are jointly designed to overcome the bandwidth fluctuation and packet loss. The LC components of the proposed SMDC scheme can support the classification and priority assignment in the DiffServ network. The intra-RAN handoff and inter-RAN handoff procedures in D-MDMN are studied. Simulation results show that the proposed video marking algorithm and handoff procedures achieve performance improvements as compared with the previously proposed video marking algorithm and the original UMTS handoff solutions.

This paper is organized as follows. Section 2 reviews related works in video steaming over UMTS with IP DiffServ backbone. Section 3 describes the proposed D-MDMN system model for video streaming over a hybrid UMTS and IP DiffServ environment. Section 4 presents the details of the proposed SMDC and IP DiffServ MPEG-4 video marking algorithm. Section 5 discusses the handoff procedures in the D-MDMN. Computer simulation results are presented in Section 6 to demonstrate the performance of the proposed techniques, followed by concluding remarks in Section 7.

2. Video Streaming Over UMTS With DiffServ Backbone

So far, layered coding (also called scalable coding) with transport prioritization has emerged as the most popular and effective scheme for video transmission over wireline or wireless networks. In LC, a raw video sequence is coded into multiple layers: the base layer contains the most important features of the video and has the ability to provide coarse visual quality independently, while the enhancement layers can refine reconstructed visual quality when decoded together with the base layer. Depending on the way the video information is partitioned, there are four scalable mechanisms to implement LC: temporal scalability, spatial scalability, signal-to-noise-ratio (SNR) scalability and data partitioning [16]. Further, by exploiting scalable coding, the fine granularity scalability (FGS) [13,14] and its variation progressive fine granularity scalability (PFGS) [15] can provide more flexibility to adapt to different access link bandwidth. The enhancement bit stream can be truncated anywhere to achieve the target bit rate. To serve as an error resilience tool, LC must be paired with unequal error protection (UEP) in the transport system, so that the base layer is protected more strongly, for example by assigning a more reliable sub-channel, using stronger forward error correction (FEC) codes, allowing more retransmissions, or allocating more resources. LC has the ability to adapt to unpredictable bandwidth fluctuations, combat both network heterogeneity and receiver heterogeneity and provide error resilience. It has no feedback channel requirement and therefore does not induce extra delay.

In LC, the base layer is critically important. If error-free transmission cannot be guaranteed, severe distortion can be induced because of packet losses in the base layer. An alternative is to use MDC [16,17], where several substreams, termed descriptions, are coded from the source streams and transmitted through independent paths from the source to the destination. Any single description should provide a basic level of quality, and more descriptions together will provide improved quality. MDC does not require special mechanisms in the network to provide a reliable sub-channel as in the case of LC. On the other hand, the coding efficiency is low in MDC, due to the necessary information overlap in different descriptions [16]. In content delivery networks, MDC combined with path diversity can provide improved error resilience for streaming media transmission [18]. However, when wireless links are included in the content delivery networks, the time-varying capacity of wireless links will degrade the performance of MDC.

For multimedia services over the wireless domain, UMTS specifications [19] define four QoS classes: conversational, streaming, interactive and background. The main distinguishing factor among these classes is delay sensitivity. The conversational class is the most sensitive, while background is the least sensitive. On the other hand, an all-IP DiffServ platform [9] is the most promising architecture to interconnect the different wireless access networks with Internet backbone to provision broadband and seamless access to end users [39]. In DiffServ edge routers, packets are classified into a limited number of service classes, according to the service level agreement (SLA) negotiated with the Internet service provider (ISP). In a core router, packets from different classes are aggregately differentiated (based on packet classification) by different per-hop behaviors (PHBs): (1) expedited forwarding (EF) [21], which provides a low delay, low loss and a guaranteed bandwidth; (2) assured forwarding (AF) [20], which provides services with minimum rate guarantee and low loss rate and (3) best effort (BE) with no QoS guarantee. Thus complex functionality (e.g. traffic marking, traffic conditioning etc.) is pushed to the edge routers, which makes DiffServ scalable. Using class-based approaches, DiffServ is well suitable to implement UEP for layered coded video flows. When a UMTS network is interconnected with a DiffServ backbone network, to provision end-to-end QoS, one important issue is to provide QoS attribute translation and mapping between the IP-DiffServ world and the UMTS world. Specifically, AF PHB can be used to provision end-to-end QoS for streaming class [37].

For UMTS networks, seamless handoff is required to dynamically support terminal migration. The handoff procedure in UMTS is discussed in References [11,12,40]. However, it may not satisfy the requirements of streaming video applications. The media providers are separated from the UMTS networks by the core network. The media streams should first get through the core network and then feed into the UMTS network. Thus, the transfer delay requirement of streaming video may not be satisfied under the current model. In addition, during handoff, the buffered data in the old radio network subsystem (RNS) need to be forwarded to the new RNS, resulting in relatively large handoff latency (defined as time between the last packet transmitted from the old base station and the first packet transmitted from the new base station) or media stream interruption, and a large amount of additional traffic.

To address the problems of provisioning QoS to video streaming over UMTS with DiffServ backbone, a distributed multimedia delivery mobile network model (i.e. D-MDMN), a scalable multiple description coding and its marking scheme over DiffServ networks and handoff procedures for the D-MDMN network model are proposed, as follows, in this paper.

3. Distributed Multimedia Delivery Mobile Network (D-MDMN)

Consider distributed multimedia delivery for video streaming services in a hybrid wireless UMTS and wireline IP-DiffServ environment. Figure 1 illustrates the system model for D-MDMN. Each radio access network (RAN) consists of several possibly interconnected RNSs. An RNS contains one radio network controller (RNC) and at least one node B. The RNC is in charge of the overall control of logical resources provided by the node Bs. The node Bs are responsible for radio transmission from/to user equipments (UEs) in the cells. Two types of general packet radio service (GPRS) support nodes are also included: the serving GPRS support node (SGSN) and the gateway GPRS support node (GGSN). The rest of the IP-based core network consists of regular routers and switches that forward packets on the basis of the user-level IP addresses. The proposed D-MDMN model is an extension of the content delivery network [18] originally proposed to overcome network congestion and server overload in the star-type network topology. It includes a set of complementary distributed media description servers (MDSs) to interact and collaborate with each

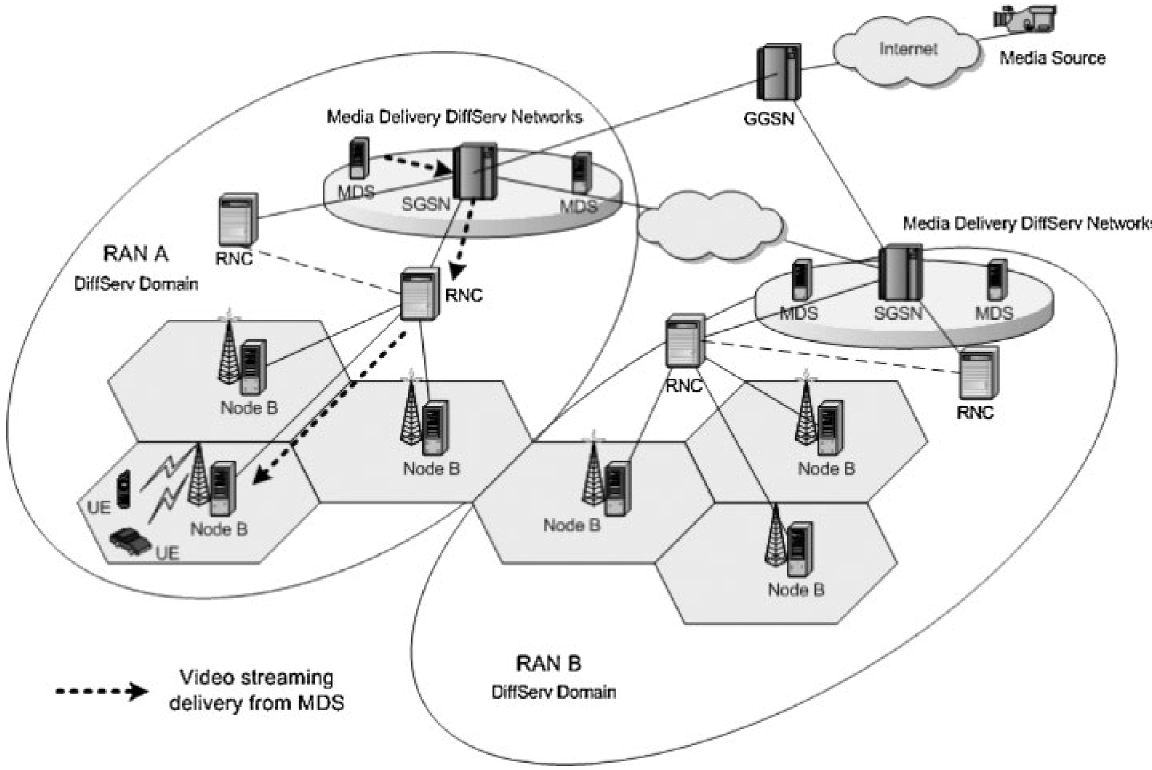

Fig. 1. The D-MDMN system model for video streaming over UMTS/IP DiffServ.

SGSN for media delivery to UEs in the RAN. Each MDS keeps one complementary description of media streams that was originally downloaded from the service provider during the streaming service publication. In the system model, the streaming media is delivered from the closest MDS and not from the origin server. Such a shorter network path for streaming media reduces the media service delivery time (end-to-end delay), the probability of packet loss and the total network resource consumption.

Figure 2 shows the proposed protocol stack for video services, where the Uu, Iub and Iu interfaces are defined in Reference [22]. This protocol stack is similar to the one shown in Figure 2b of [23], where the user datagram protocol (UDP) is used instead of the transmission control protocol (TCP). Since TCP retransmission introduces delays that are not acceptable for streaming applications with stringent delay requirements, UDP is typically employed as the transport layer protocol for video streams. In addition, since UDP does not guarantee packet delivery, the receiver needs to rely on the upper layer, i.e. real-time transport protocol (RTP) [24], to detect packet loss. RTP is a data transfer protocol designed to provide end-to-end transport functions for supporting real-time applications. It does not guarantee QoS or reliable delivery, but rather, provides the following functions in support of media streaming: time-stamp-

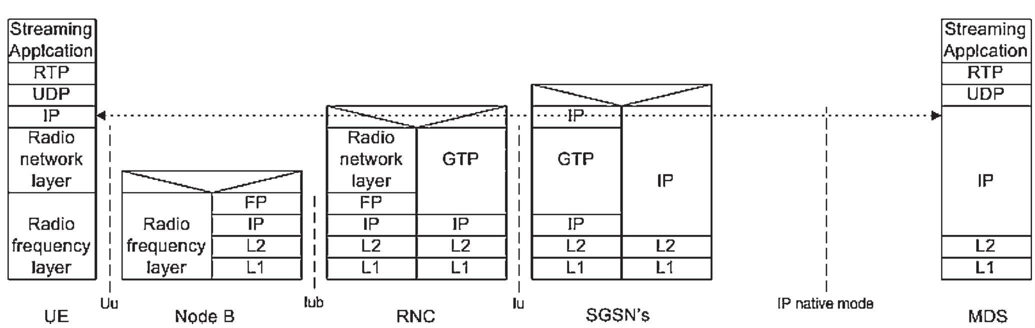

Fig. 2. Protocol stacks for video services in D-MDMN (data plane).

ing, sequence numbering, payload type identification and source identification. It works in conjunction with the RTP control protocol (RTCP) [24].

In the protocol stack, over Uu interface, radio network layer (RNL) frames (termed transport blocks) are transmitted, while over the Iub interface, an RNL framing protocol (FP) is used to encapsulate a set of transport blocks. The video IP packets generated from the video provider are supposed to pass two different tunnels: a GTP (GPRS tunneling protocol) tunnel existing between SGSN and RNC, and an FP tunnel starting from RNC and terminating at node B in the RAN. This is referred to as the transport mode of IP deployment. In other parts of the core networks, native node of IP deployment is used, where no other intermediate layers are involved [23].

4. Scalable Multiple Description Coding (SMDC)

For the purpose of solving the handoff problems in media streaming, a combination of SMDC with distributed video storage is proposed to support streaming video handoff in the D-MDMN. It provides path diversity to combat outage due to handoff. The coded video stream consists of MDC components and LC components. In the proposed D-MDMN, MDC components enhance the robustness to losses and bit errors of LC components through path diversity and error recovery. MDC components also reduce the storage, reliability and load balancing requirement among distributed media edge servers (i.e. MDSs). At the same time, LC components not only deal with the unbalanced MD operation at the server end, but also combat the fluctuation of wireless resource availability due to the time-varying wireless channels and user mobility.

4.1. The Proposed Architecture

The architecture of the proposed SMDC framework is depicted in Figure 3. It is an object-based coding, which jointly employs LC based on PFGS [15] and MDC based on the multiple state recovery (MSR) [16,17], provided that it is compatible with the MPEG-4 codecs.

Similar to the human visual system mechanism, the smallest entity in SMDC is each object in a picture with its associated shape, texture in the interior of the shape and motion. The original video input to the encoder is segmented into a set of individual video objects (VOs). Each VO is then individually compressed through shape encoding and PFGS texture encoding. For the support of two descriptions, the encoder stores the last two previously coded frames (instead of just the last one) and chooses which previously coded frame to be used as the reference

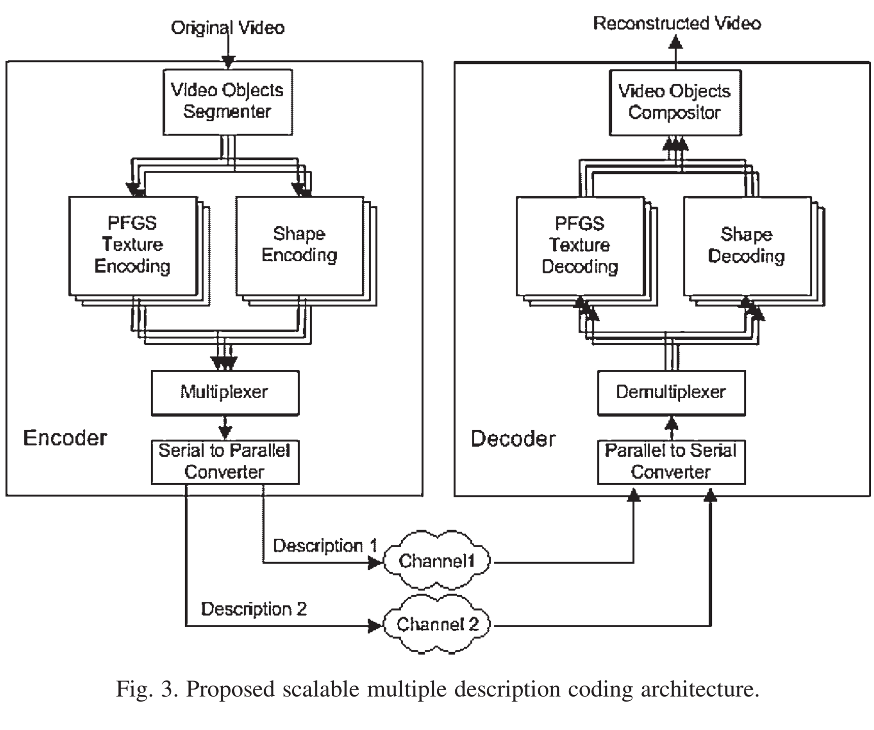

Fig. 3. Proposed scalable multiple description coding architecture.

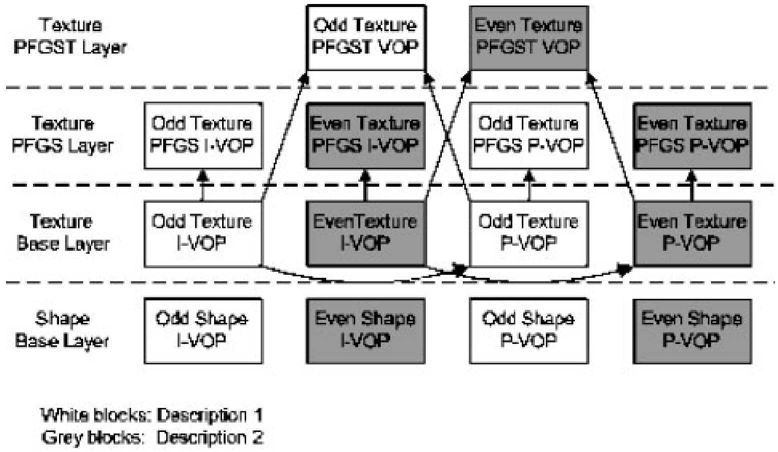

Fig. 4. Scalability structure of the scalable multiple description coding.

for the current prediction [27]. After multiplexing, one video stream with four different layers can be generated. This video stream is further partitioned into two subsequences of frames: odd numbered video frames (Description 1) and even numbered video frames (Description 2), as shown in Figure 4.

The different descriptions are transmitted over different wireless channels experiencing independent error effects, in order to minimize the chances that both video streams are corrupted at the same time. In fact, the video stream can be partitioned into N (> 2) complementary frame subsequences if there are N independent channels between the encoder and the decoder. However, it also adds complexity of the multiple description (MD) assembling at the decoder. For presentation simplicity, two descriptions and two corresponding channels are used in the following discussion.

At the decoder, the processing procedures reverse in accordance. Similarly, the decoder alternates the previous erroneously decoded frame used as the reference for the next prediction. If both descriptions are received erroneously after parallel to serial converter (MD assembler) and demultiplexing, the shape and texture information of VOs are restored from shape and PFGS texture decoder for final composition into reconstructed video. If there is an error in a stream, the error propagation will happen in that stream due to motion compensation and differential encoding.

The SMDC framework can employ any shape coding [25,26], for example binary shape coding or grayscale coding. The texture coding techniques are discrete cosine transform based coding for arbitrary shaped objects. The concept of object based representation makes it possible to exploit the content redundancy in addition to the data redundancy and to improve the coding efficiency for the very low bit-rate transmission.

4.2. Scalability Structure of SMDC

As shown in Figure 4, the proposed SMDC scalability structure is as follows:

- the shape base layer, consisting of shape information of VOs in the intra-coded video object plane (I-VOP) or shape and motion information of VOs in the predictively coded VOP (P-VOP);
- the texture base layer, consisting of basic texture information of VOs contoured by the shape base layer;
- the texture PFGS layer, consisting of texture information of SNR scalable enhancement for the texture base layer and
- the texture PFGS temporal (PFGST) [41] layer, consisting of motion-compensated residual frames predicted from the texture base layer for temporal scalable enhancement. In comparison, the motion-compensated PFGST frames in SMDC take the place of B-frames in the multilayer FGS-temporal scalability structure presented in Reference [13].

For instance, suppose the first three layers are implemented and the texture PFGST layer is left as an option. Thus, playing only one description with only the shape base layer gives a black and white (or grayscale) video at the half frame rate. Playing only one description with the shape base layer and the texture base layer gives a color video in a basic quality at the half frame rate. Playing only one description with all the three layers yields a color video in a better

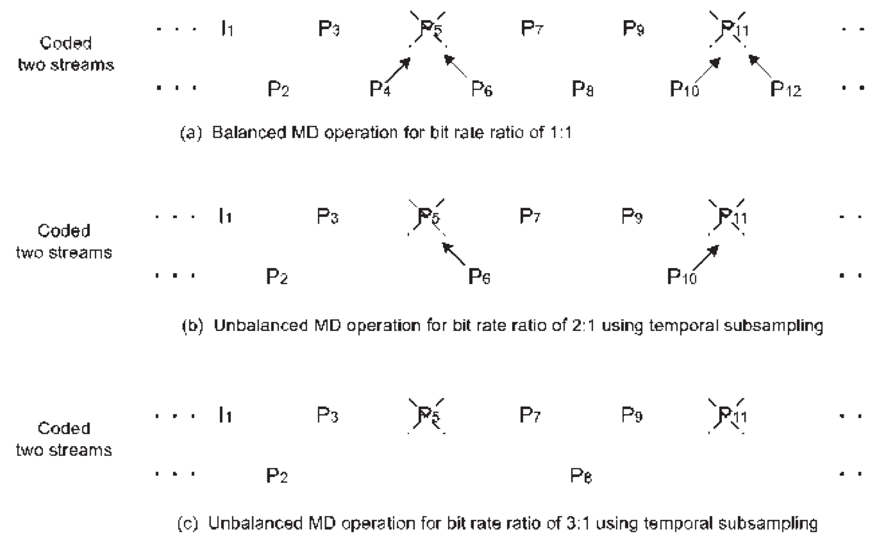

Fig. 5. Balanced and unbalanced MD operations. (a) Balanced MD operation for bit rate ratio of 1:1. (b) Unbalanced MD operation for bit rate ratio of 2:1 using temporal subsampling. (c) Unbalanced MD operation for bit rate ratio of 3:1 using temporal subsampling.

quality at the half frame rate. In the same way, if both two descriptions with all the three layers can be decoded correctly, it yields a color video in the best quality at the full frame rate. Note that the layering in SMDC is more flexible than that given in the example. If no sufficient resources, the texture PFGS layer needs not be discarded as a whole. The enhancement bit stream can be truncated anywhere to achieve the target bit-rate. This benefit of achieving continuous rate control comes from the bitplane coding in the PFGS encoder [15] for the enhancement stream.

4.3. Advantages of the Proposed LC Components

In MDC, the coded descriptions are transmitted via different and independent paths. To fully utilize the available transmission capacity along its path, the bit rate of each description should be adapted accordingly. Therefore, unbalanced MD operation [27] is required in case of different bandwidth availability levels among different paths. Adaptive quantization, spatial resolution or frame rate are possible solutions to achieve unbalanced operation. In order to avoid perceiving of a quality variation (flicker) at half of the original frame rate, each stream should preserve approximately equal quality level. Adaptive quantization approach is effective when a small rate change is required, for example 10% rate reduction can be achieved with a cost of 0.5 dB. However, it may lose the effectiveness in case of large rate changes. Adaptive spatial resolution approach is not good due to the potential flicker [27]. Adaptive frame rate (i.e. temporal subsampling) approach [27] has the ability to adapt to different available transmission capacity and at the same time preserve the quality per frame, as illustrated in Figure 5. However, if the frame rate of one stream is decreased too much, the quality of that stream cannot be closely preserved. Also, the unbalanced MD operation will fail if the bit rate ratio of these two streams is larger than 2:1, as illustrated in Figure 5.

Consider that the bit rate of the upper stream is bigger than that of the lower stream in Figure 5, where P_x denotes the P-frame X,[‡] and ⤫ means that a frame is discarded or damaged. The balanced MD operation is shown in Figure 5(a), where the damaged P-frame 5 can be recovered or concealed from P-frames 4 and 6, and damaged P-frame 11 is recovered or concealed from P-frames 10 and 12. In Figure 5(b), the frame rate of the lower stream has to be decreased by 50% for a bit rate ratio of 2:1. That is, P-frames 4 and 8 have to be discarded. The damaged P-frame 5 can be

[‡]In MPEG coding, there are three types of compressed frames: intra-coded (I) frame, coded independently of all other frames; predictively coded (P) frame, coded with reference to a previous I-frame or P-frame and bi-directionally predicted (B) frame, coded with reference to both previous and future I- or P-frames.

recovered but only from P-frame 6, and P-frame 11 can be recovered but only from P-frame 10. It is clear that the error recovery capability of 2:1 unbalanced MD operation illustrated in Figure 5(b) is lower than that of the balanced MD operation illustrated in Figure 5(a). In Figure 5(c), the frame rate of the lower stream has to be decreased by 67% for a bit rate ratio of 3:1. In this case, the damaged P-frames 5 and 11 in the high bit rate stream cannot be recovered from the low bit rate stream because their adjacent previous P-frames 4 and 10 and their adjacent future P-frames 6 and 12 have to be discarded.

In addition, to support unbalanced MD operation, the approaches with adaptive quantization, spatial resolution or frame rate generally require close-loop feedback channels to indicate the available capacity in the transmission paths. This task is challenging in the wireless domain due to the time-varying capacity of the wireless channel. The induced delay in the feedback channel also affects the effectiveness of these approaches.

On the other hand, the proposed SMDC has the ability to address the unbalanced MD operation effectively with no need of a close-loop feedback channel. The capability of error recovery of MDC and SMDC are compared under the bit rate ratio of 3:1 in Figure 6. As discussed above, for adaptive frame rate approach, the errors occurred in the high bit rate stream cannot be recovered or concealed from the low bit rate stream in the case that the bit rate ratio is larger than 2:1. Suppose that the bit rate of the upper stream in Figure 6 is three units and that of the lower stream is only one unit. Instead of temporal subsampling illustrated in Figure 6(a), the LC is introduced in Figure 6(b) for the unbalanced MD operation. As to the path of low bandwidth, part of the enhancement layers can be dropped so that the original frame rate can be preserved. In Figure 6(b), P_x' denotes the remaining part of P-frame X after layer-dropping in order to adapt to the bandwidth limitation. Thus, the damaged P-frames 5 and 11 in the upper stream can still be recovered from the frames of the lower stream, shown in Figure 6(b), in the same manner as the balanced MD operation shown in Figure 5(a). In other words, the unbalanced MD operation using LC does not affect the error recovery capability of SMDC.

4.4. IP DiffServ Marking Algorithm for SMDC

In video coding, the coded information may not have the same importance level. In MPEG encoders, different frames in a video sequence do not have the same importance as some frames are dependent on others. Intra-coded frames (I-frames) are more important than predictive frames (P-frames). In each frame, different types of information (e.g. shape, motion and texture) also have different importance levels, for example for a P-frame, the shape and motion information is of more importance than texture information [28]. Generally, different importance levels can be explored in the implementation of UEP, taking advantage of the error resilience [29] and concealment [16] tools provided by MPEG-4. Therefore, to fully utilize the error resilience and concealment capacity in MPEG-4, it is desired that the network can distinguish frame types, coded layer types and information types. Usually, IP video traffic is classified into an AF class in DiffServ networks [20,30], and random early detection (RED)-based queue management architectures [31,32] have been proposed to combine random dropping of packets with IP precedence. The RED-based approaches

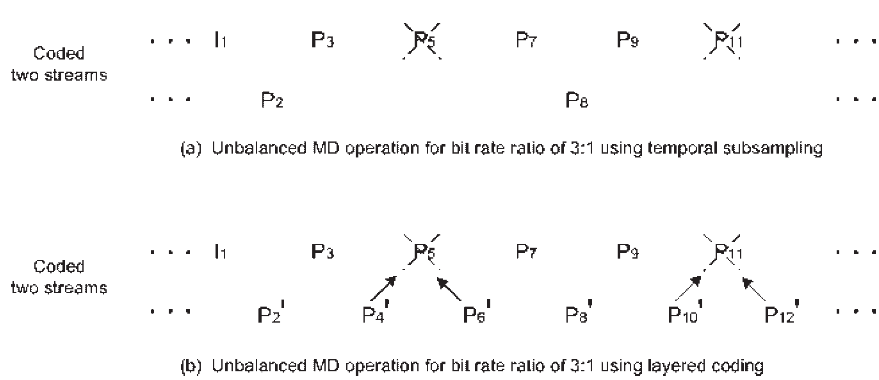

Fig. 6. Comparison of unbalanced MDC with unbalanced SMDC: (a) Unbalanced MD operation for bit rate ratio of 3:1 using temporal subsampling. (b) Unbalanced MD operation for bit rate ratio of 3:1 using layered coding.

Table I. Proposed IP DiffServ MPEG-4 video marking algorithm.

Stream	Control information	Shape base layer	Texture base layer	Texture PFGS layer	Texture PFGST layer
OD and BIFS	EF or AF1	—	—	—	—
I-VOP	—	AF1 (Class I stream)	AF1 (Class I stream)	AF2 (Class II stream)	—
P-VOP	—	AF1 (Class I stream)	AF2 (Class II stream)	AF3 (Class III stream)	—
PFGST VOP	—	—	—	—	AF3 (Class III stream)

take advantage of the TCP retransmission mechanism. However, UDP is more suitable for video streaming, where all the randomly dropped packets are considered as packet loss and are not retransmitted. Because of error propagation in video streaming, the effect of packet loss gets worse. Since MPEG-4 video is predictive inter-frame coded and layered coded, artifacts due to random packet dropping can persist for many frames or layers [33]. If an error occurs while transmitting the base layer, its enhancement layers have to be discarded. It means that stochastically isolated single packet loss or bit error is converted to a burst of lost packets or bit errors. Therefore, early random packet dropping before congestion is not suitable for video streaming.

A novel marking algorithm to support UEP implementation of LC components in SMDC is proposed in Table I. In the proposed SMDC scheme, different types of information in different coded layers and different VOPs are re-organized and packetized into Classes I, II and III streams, and marked into AF1, AF2 and AF3 classes in DiffServ domain respectively. In comparison with the IP DiffServ video marking algorithm (DVMA) proposed in Reference [30], where there is only one queue with three different levels of precedence for video stream, each AF class in the proposed algorithm has one separated queue. This algorithm can be implemented by class-based weighted fair queuing (WFQ) [34], where RED should be disabled in each class. Moreover, MPEG-4 introduces extra data control streams, such as the object descriptor (OD) and scene description (binary format for scene, BIFS). These signaling streams are very loss- and jitter-sensitive and need to be protected and marked as EF, or AF1 if EF PHB is not available.

5. Handoff Procedures

This research focuses on the hard handoff procedures. Soft handoff may provide better performance for media streaming. However, hard handoff is required when there are no connections between source RNC and target RNC within the mobile network, especially under the consideration of network heterogeneity and receiver heterogeneity, such as interworking between UMTS and global system for mobile communications (GSM)/enhanced data rates for GSM evolution (EDGE) radio access networks, or UMTS and IEEE 802.11 (wireless local area networks).

5.1. Intra-RAN Handoff Procedure

Figure 7 illustrates the intra-RAN handoff scenario in the D-MDMN. The intra-RAN handoff procedure consists of three phases, as shown in Figure 8.

Phase I—preparation for RNS handoff and resource allocation: the control plane of the proposed handoff phase I is basically the same as that in UMTS [11,12,40], except that the current position (offset) of the received media stream should go along with measurement report (signal no. 1) given by the UE. It is the only state information required for session migration, which is small enough to be hidden inside the handoff signaling and to be relayed to the SGSN. Based on the measurement report from the UE and its own measurement and on current traffic conditions, the source RNS (sRNS) makes the handoff decision and sends an HO-required message (signal no. 2) to the SGSN, indicating the selected target RNS (tRNS) to which the handoff should be performed. The SGSN then sends to the selected tRNS an HO-request message (signal no. 3). If there are sufficient resources to accommodate the UE, the tRNS allocates a physical channel for the coming UE, and returns an HO-request-ack message (signal no. 4) to the SGSN. Upon receipt of the message, the SGSN sets up a link (i.e. GTP tunnel) to the tRNS, and sends to the UE (via the sRNS) an HO-command message (signal no. 5 and 6), which also contains the radio interface message in the tRNS. Upon receiving the signal no. 5 at the end of the preparation phase, the sRNS stops transmitting downlink data to the UE. Unlike the case in UMTS [11,12,40], no data forwarding is required in D-MDMN, thus the sRNS does not store all downlink

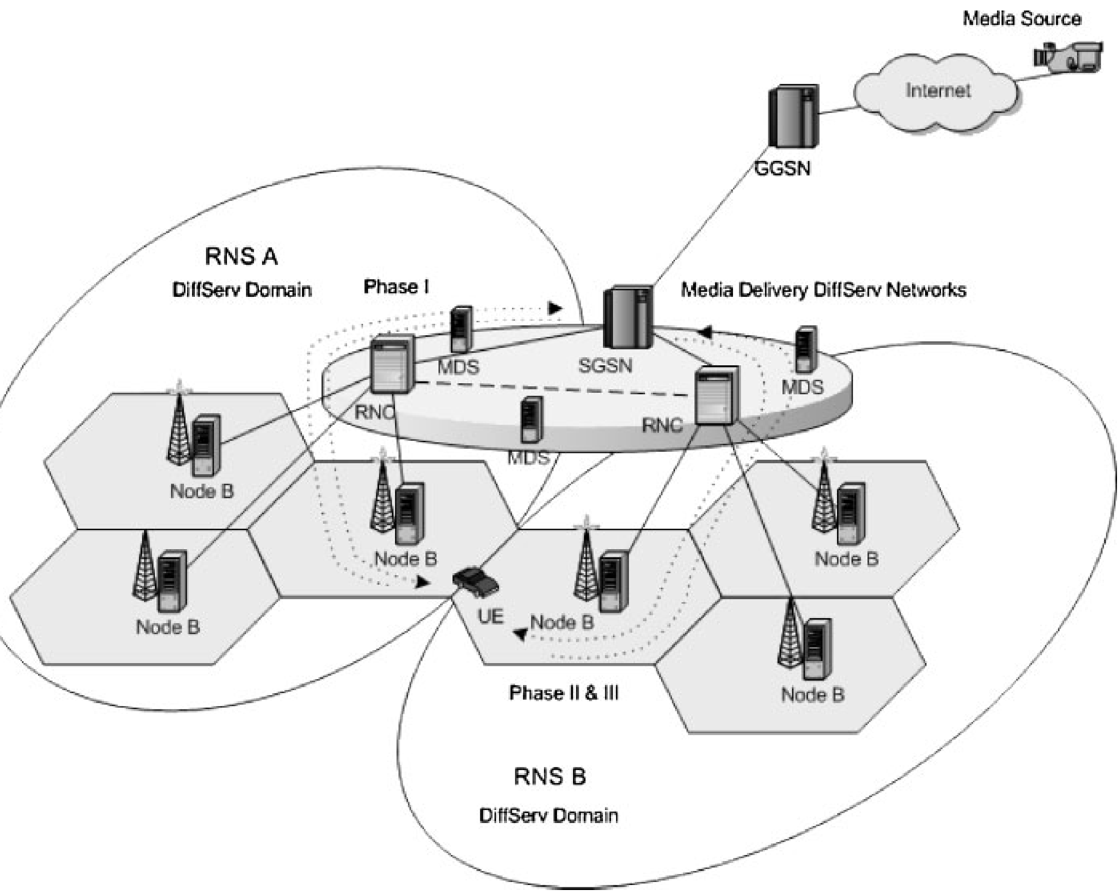

Fig. 7. Proposed network model of the intra-RAN handoff (data plane) in D-MDMN.

data, which continue to arrive at the source RNC from the SGSN.

Phase II—moving the serving RNS role to target RNS: upon receiving the HO-command (signal no. 5), the sRNS issues to the UE the radio interface message HO-command (signal no. 6), which includes a handover reference number previously allocated by the tRNS. The UE then breaks its old radio link and accesses the new radio resource by the HO-access message (signal no. 8), which contains the handover reference number. After the verification of the number, the tRNS shall send an HO-detect message (signal no. 9) to the SGSN. These are similar to those in the UMTS handoff procedure [11,12,40]. The most important difference between the proposed handoff phase II and that in UMTS is that the Stream Re-establishing takes the place of the Media Stream Forwarding upon receiving the signal no. 5 in Phase I. There are no buffered data required to be forwarded. As soon as the GTP tunnel is created between the tRNS and the SGSN, the SGSN initiates the MD-request message (signal no. 7) and the user flow is switched from the old path to the new path. Upon receiving the MD request, the set of MDSs surrounding the SGSN starts the downlink media delivery from the offset point of the same stream at the handoff decision according to subscriber service bindings (i.e. how many descriptions and layers the UE subscribes). In other words, the media stream is re-established. The MD-request message contains the offset information at the handoff decision point.

Phase III—releasing resource reservation in the old path: after successful connection with the tRNS, the UE sends an HO-complete message (signal no. 10) to the tRNS, which then sends an HO-complete message (signal no. 11) to the SGSN. Upon receiving the message, SGSN releases the resources previously allocated to the UE in the sRNS by using the clear-command message (signal no. 12) and clear-complete message (signal no. 13). The most important difference between the proposed handoff phase III and that

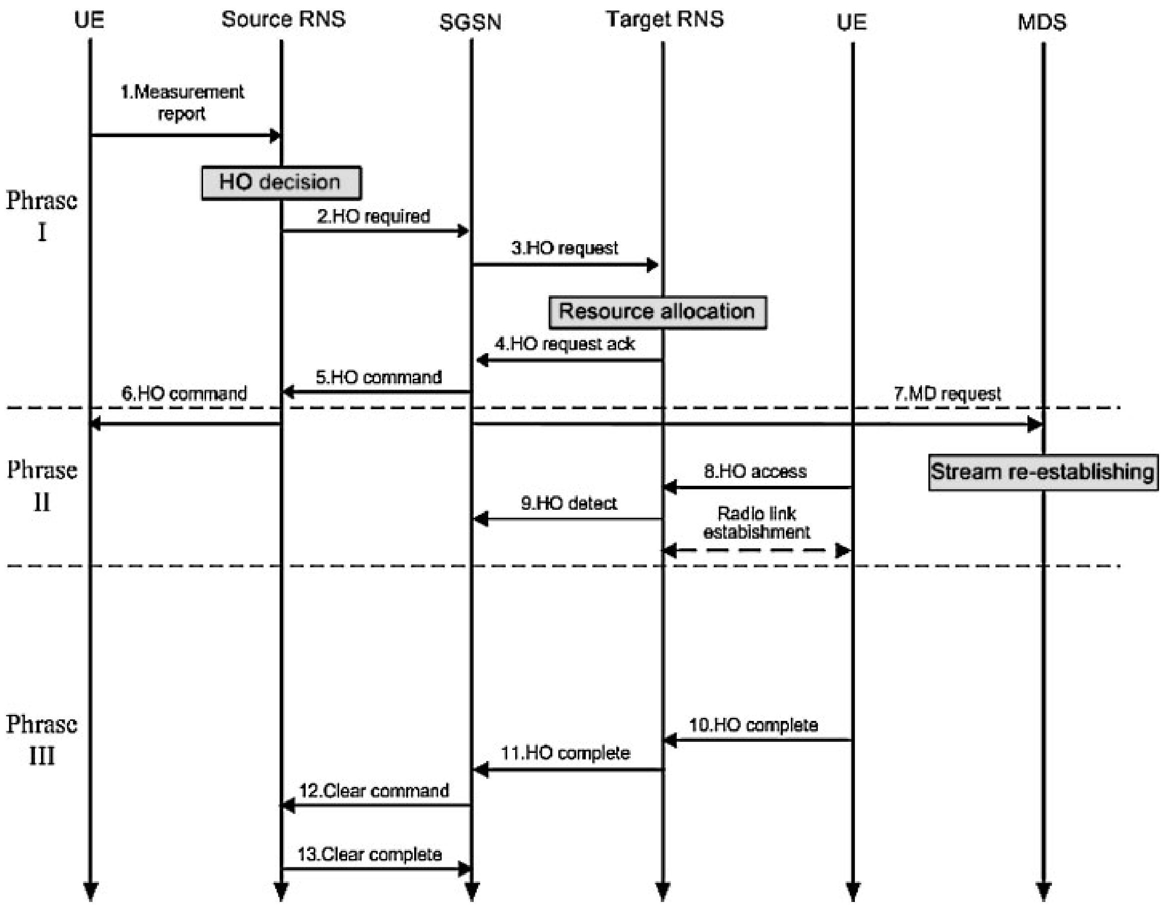

Fig. 8. Proposed intra-RAN handoff procedure (control plane) in D-MDMN.

in UMTS [11,12,40] is that there is no buffer requirement in the RNSs for data forwarding and resequencing. Only a small buffer is needed in the RNSs for absorbing the delay jitter of a video stream and for reordering due to changes in the routing paths. The functionality of multiple description assembly is implemented in the UEs.

5.2. Inter-RAN Handoff Procedure

The inter-RAN handoff scenario is illustrated in Figure 9. The proposed inter-RAN handoff procedure for media streaming also consists of three phases, as shown in Figure 10. Phase I is for preparation for RNS handoff and resource allocation. There is no tunneling required between the SGSNs as compared to UMTS [11,12,40], since no data forwarding is required. In Phase II, the serving RNS role is moved to the target RNS. Phase III is to release resource reservation in the old path.

5.3. Handoff Enhancement for Streaming Services

The advantages of the proposed handoff approach for media streaming are summarized as follows:

- Due to the replacement of data forwarding by stream re-establishing, the handoff latency can be reduced. In addition, there is no buffer requirement in the RNSs for data forwarding and resequencing in case of stream re-establishing. Only a small buffer is needed in the RNSs for absorbing the delay jitter of a video stream.
- The proposed handoff procedure takes advantage of the existence of the MDSs. The streaming media can be delivered over a shorter network path, which reduces the transfer delay and delay jitter of media service delivery, the probability of packet loss and the total network resource occupancy.
- It has relatively consistent QoS in all scenarios (i.e. with handoff scalability enhancement). For

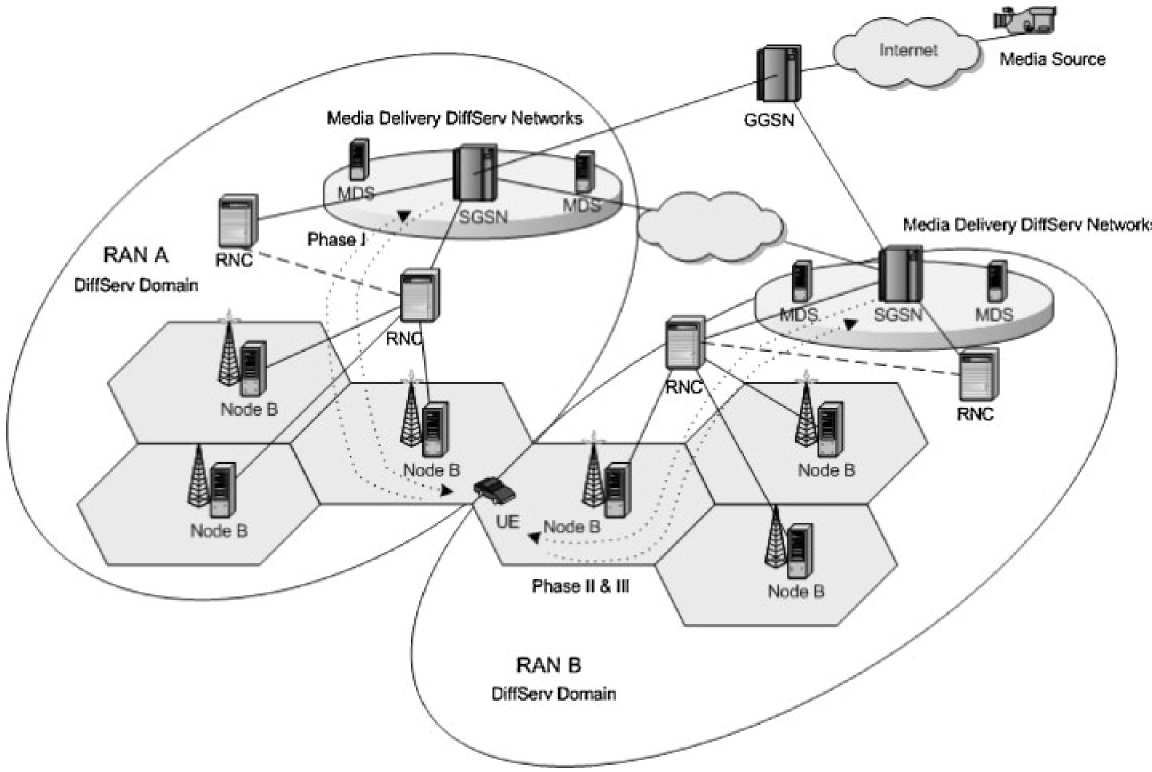

Fig. 9. Proposed network model of inter-RAN handoff (data plane) in D-MDMN.

handoff procedure in UMTS [11,12,40], the values of handoff latency vary with the length of the data-forwarding path in different handoff scenarios. Also, the end-to-end delay varies with different delivery paths and different locations of the media providers, which is outside the core network and far from the mobile hosts. However, due to the introduction of the MDSs, the values of handoff latency and end-to-end delay in different handoff scenarios depend mainly on the length of media delivery path from MDSs to SGSN, and then to UE. Usually, the SGSN and the MDSs are neighbor nodes. The length of media delivery path from the MDSs to the UE is relatively consistent in different handoff scenarios.

6. Performance Analysis

In the following, we present numerical results to demonstrate the performance of the proposed IP DiffServ MPEG-4 video marking algorithm and the streaming video handoff procedures, based on computer simulation using OPNET [36]. The video traffic is made up of Class I, Class II and Class III layer-coded video streams [35], which are classified through the proposed IP DiffServ MPEG-4 video marking algorithm. The video traffic, which is generated by OPNET is subjected to the requirements of UMTS bearer service attributes of streaming class [19]. Table II lists the traffic parameters, where voice and web traffic flows are classified into EF and BE classes respectively.

6.1. Video Marking Algorithm

To implement the proposed IP DiffServ video marking algorithm in each node, we use WFQ discipline to classify and schedule the incoming packets into and out of each EF, AF1, AF2, AF3 or BE queue. In order to compare the proposed IP DiffServ video marking algorithm with the DVMA scheme [30], three scenarios are tested: (1) the best effort model using a drop-tail BE queue in each node, (2) the DiffServ model using DVMA which is based on class based queuing (CBQ) with multi-level random early detection (MRED) AF queue [35] and (3) the DiffServ model using the proposed IP DiffServ video marking

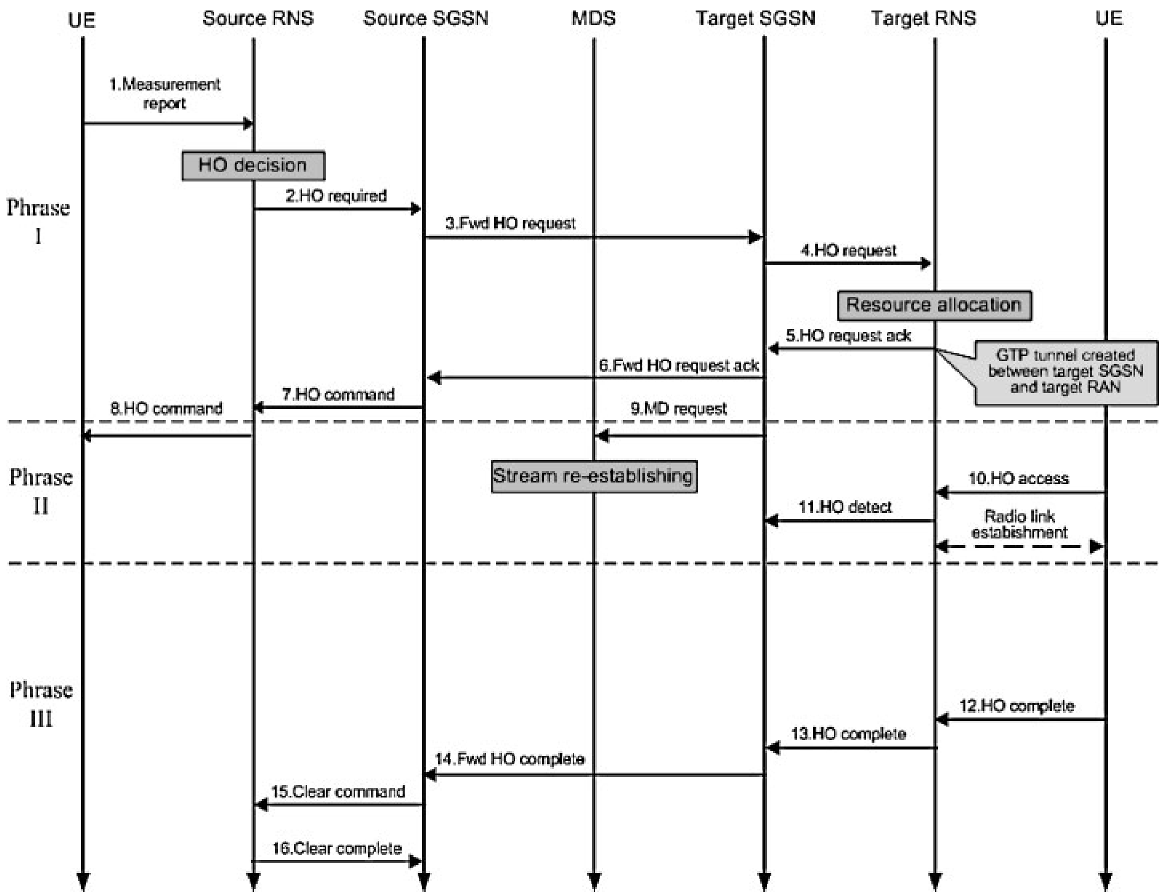

Fig. 10. Proposed inter-RAN handoff procedure (control plane) in D-MDMN.

Table II. Traffic profile of each cell for evaluation of video marking/handoff algorithms.

Traffic type	User number	Sending rate (mean)	Packet length (mean)	Standard
Video stream[‡]	2	240 Kbps per client	1 Kbytes	MPEG-4
Voice stream	25/5	16 Kbps per user	200 Bytes	G.728
Web traffic	—	400 Kbps	1 Kbytes	HTTP

[‡]Frame rate (mean) = 20 frame/s, frame length (mean) = 1 Kbytes.

Table III. Queue scheduling configurations.

Solution	Queue scheduling	Queue classification	Queue size	Normalized bandwidth (%)
Best effort	FIFO	BE queue	45	100
DVMA	MRED + CBQ	BE queue	15	4.5
		AF queue	15	45.5
		EF queue	15	50
Proposed	WFQ	BE queue	15	4.5
		AF3 queue	5	9.1
		AF2 queue	5	13.7
		AF1 queue	5	22.7
		EF queue	15	50

algorithm with the drop-tail AF1, AF2 and AF3 queues. Table III lists the configurations of queue scheduling in these three scenarios. For all the test scenarios, a Rayleigh fading channel is assumed with $\mathrm{BER} = 10^{-4}$.

To evaluate the quality of reconstructed video, peak signal to noise ratio (PSNR) is an objective metric. For different transmitted video sequences, different PSNR values may be obtained even with the same transmission bit error rate (BER) or frame error rate (FER)[§] due to the different motion characteristics and different video complexity degrees [38]. As a small FER is very likely to lead to a high PSNR, FER is used in this

[§]In this paper, FER is referred to as the service data unit (SDU) error ratio [35].

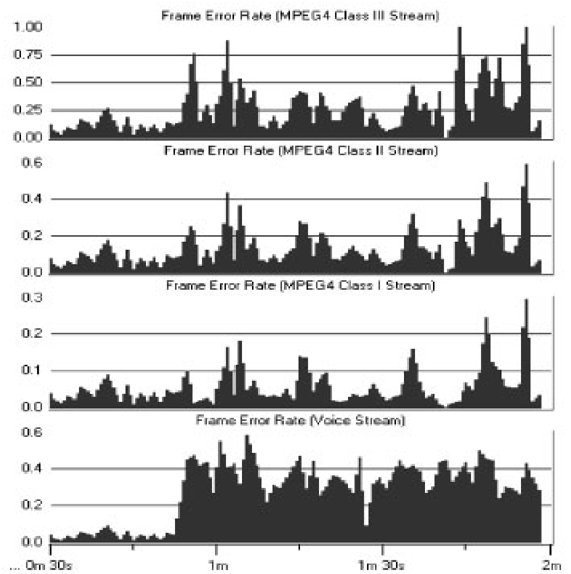

Fig. 11. FER (test case a: best effort, FIFO).

paper to evaluate the quality of received and reconstructed video. The upper bound of FER allowed for streaming class in UMTS is 10% [19]. Effects of the different video marking algorithms on FER are shown in Figures 11–13. The frame errors are caused by both the packet loss at the RNSs due to congestion and the wireless channel bit errors. Due to the employment of MRED for proactive packet-dropping in DVMA, the packet loss in DVMA begins earlier than that in the proposed solution. However, the packet loss in the tail-dropping BE solution occurs even earlier than that

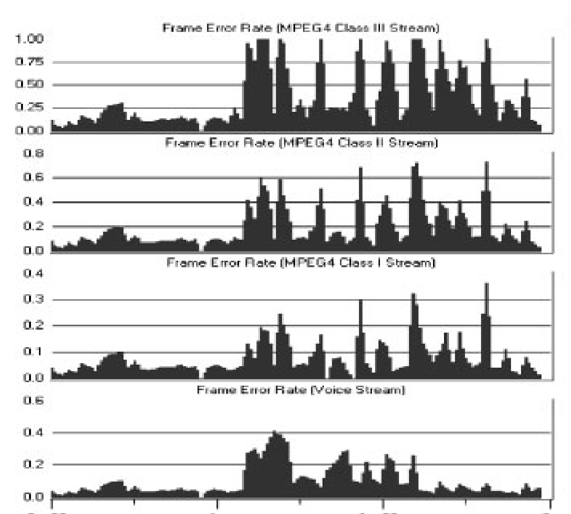

Fig. 12. FER (test case b: DiffServ-DVMA, CBQ + MRED).

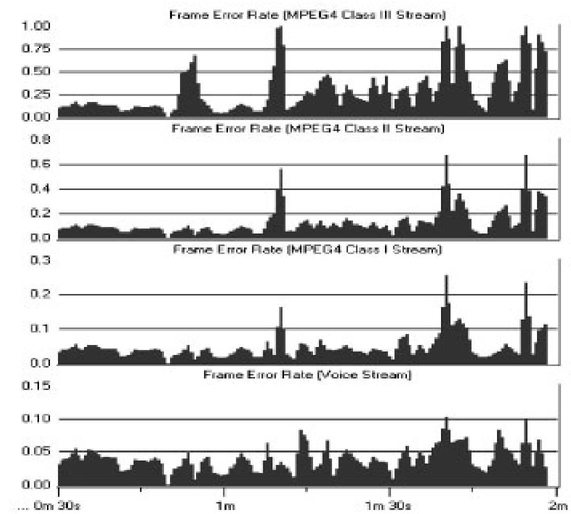

Fig. 13: FER (test case c: DiffServ—the new algorithm, WFQ).

in DVMA. That is because the background traffic and the video traffic enter into the same queue, resulting in the congestion to happen earlier. Note that the background traffic and the video traffic are separated in different queues in the DiffServ-based solutions. Since we re-organize the shape, motion and texture video information into different layers, unequal error protection is introduced and results in differentiated services. If a bit error or a packet loss occurs in the MPEG-4 Class I stream, the corresponding bits or packets in the MPEG-4 Class II and Class III streams have to be considered erroneous or lost. Similarly, if a bit error or a packet loss occurs in the MPEG-4 Class II stream, the corresponding bits or packets in the MPEG-4 Class III stream also have to be considered erroneous or lost. Some packets may arrive late and are also considered lost. If the higher priority traffic is protected, less packet loss (i.e. FER) will occur. Among the three solutions, the protection of both voice stream and MPEG-4 Class I stream in the proposed solution is the best. This also results in the least FER among all the scenarios. In addition, the protection of EF traffic in DVMA is better than that in BE, at the cost of high FER of AF traffic in DVMA.

In the simulations, we choose the maximum allowed transfer delay for streaming class in UMTS, 300 ms [19], as the maximum delay requirement for D-MDMN under the test conditions. The maximum jitter is set to be of the duration of one video frame, i.e. 50 ms (with a video frame rate 20 frames/s). The effects of different video marking algorithms on the

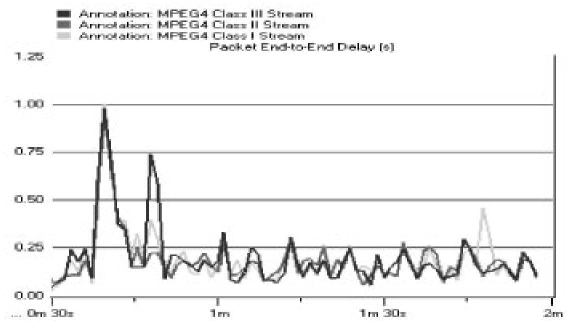

Fig. 14. End-to-end delay (test case a: best effort, FIFO).

end-to-end delay are presented in Figures 14–16. In the BE solution and the DVMA solution, the end-to-end delay of video traffic in different classes cannot be differentiated, because different classes video streams go through the same queue (e.g. BE queue or AF queue). As expected, in the proposed solution, the performance of MPEG-4 Class I stream is better than Class II, and Class II better than Class III. At BER $= 10^{-4}$, with a sudden increase of the background traffic at time 36 s, the end-to-end delay of video streams jump sharply up to a higher level. The proposed

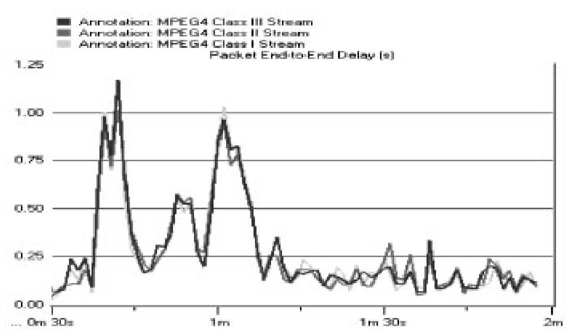

Fig. 15. End-to-end delay (test case b: DiffServ-DVMA, CBQ + MRED).

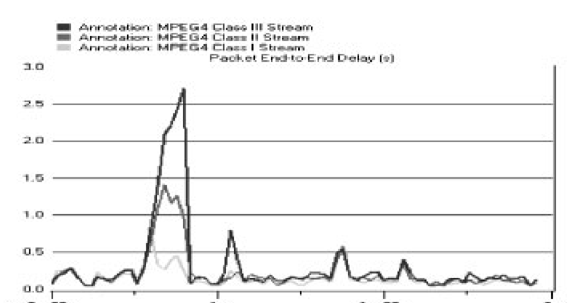

Fig. 16. End-to-end delay (test case c: DiffServ—the new algorithm, WFQ).

solution delays by 10 s the start point of performance degradation, compared with BE and DVMA. After the RNSs begin to drop packets, the performance of end-to-end delay turns better. In DVMA, the video delay jitters of all three classes are not acceptable [35], though Class III and Class II streams in DMVA are better than those in the proposed solution. In comparison, the MPEG-4 Class III and Class II streams in the proposed solution are sacrificed in order to guarantee the QoS of the Class I stream.

6.2. Streaming Video Handoff

The performance of the proposed streaming video handoff in D-MDMN is examined and compared with that in UMTS model under the scenario of the proposed IP DiffServ video marking algorithm. Eight test cases of streaming video handoff are considered, as shown in Table IV, where BER is 10^{-5} in Cases I–IV and 10^{-4} in Cases i–iv. As indicated previously, the maximum end-to-end delay and delay jitter of video streaming allowed in both UMTS and D-MDMN are 300 ms and 50 ms respectively. The handoff latency also should be below 50 ms as a delay jitter. In all the UMTS test cases, the handoff latency cannot satisfy the 50 ms bound. Note that, because the distance between the video provider and GGSN is unknown, we choose the best case in the UMTS simulation model. That is, the distance between them is only one hop. If the video provider is farther away from the UMTS core network, the performance of maximum end-to-end delay in UMTS is unlikely to meet the QoS requirement anymore. On the other hand, with the proposed stream re-establishing handoff solution, all the test cases in D-MDMN meet the performance requirement of the maximum end-to-end delay and handoff latency. Because the video provider is distributed as the media databases inside the core network, the handoff performance remains stable in D-MDMN and there is no distance problem as mentioned above in UMTS. Table IV summarizes the handoff performance in UMTS and D-MDMN. It shows that the improvement of handoff performance ascends with the increase of the scale of the mobile core network. For example, the handoff latency improvement in the intra-RAN scale is 26 or 38 ms under different conditions of wireless transmission BERs, but in the inter-RAN scale it is 45 or 57 ms. Furthermore, from Table IV, it can be seen that the proposed stream re-establishing handoff performance in D-MDMN is relatively consistent in all the scenarios. This validates the enhancement of handoff scalability in D-MDMN.

Table IV. Comparison of handoff performance in UMTS and D-MDMN.

Test case	End-to-end delay (mean) (ms)	Delay improvement of D-MDMN (ms)	Handoff latency (mean) (ms)	Handoff latency improvement of D-MDMN (ms)
I: UMTS, intra-RAN	74	32	68	26
II: D-MDMN, intra-RAN	42		42	
III: UMTS, inter-RAN	75	30	91	45
IV: D-MDMN, inter-RAN	45		46	
i: UMTS, intra-RAN	79	34	82	38
ii: D-MDMN, intra-RAN	45		44	
iii: UMTS, inter-RAN	131	75	104	57
iv: D-MDMN, inter-RAN	56		47	

7. Conclusion

This paper proposes a distributed multimedia delivery mobile network model for the UMTS core network. The model combines the concepts of content delivery network and scalable multiple description coding into the UMTS network in order to solve the video handoff problem and meet the stringent QoS requirements of video streaming in 3G wireless communications. In the network model, the media streaming services are pushed to the edge of core network, thus reducing the media service delivery time, the probability of packet loss and the total network resource consumption, and achieving relatively consistent QoS. With the joint design of LC and MDC, the MDC components enhance the robustness to losses and bit errors of the LC components through path diversity and error recovery, while the LC components not only deal with the unbalanced MD operation at the server end, but also combat the bandwidth fluctuations of the time-varying wireless channel. A new MPEG-4 video marking algorithm is presented for the hybrid wireless UMTS and wireline IP DiffServ environment, which provides service differentiation to different classes of video packets. The proposed handoff procedures employ the principle of video stream re-establishing to replace the principle of data forwarding in UMTS. Computer simulation results demonstrate the effectiveness of the proposed video marking algorithm and handoff procedures.

References

1. Wu D, Hou YT, Zhang Y-Q. Scalable video transport over wireless IP networks. In *Proceeding of IEEE PIMRC* 2000; vol. 2; pp. 1185–1191.
2. Karrer R, Gross T. Dynamic handoff of multimedia streams. In *Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, 2001.
3. Dogan S, Eminsoy S, Sadka AH, Kondoz AM. Personalised multimedia services for real-time video over 3G mobile networks. In *Proceeding of Third International Conference on 3G Mobile Communication Technologies* 2002; pp. 366–370.
4. Ha R, Zhuang W. Performance of MPEG-4 video streaming over 3G networks via WAP. In *Proceedings of 3G wireless* 2002; pp. 213–218.
5. Wu D, Hou YT, Zhu W, Zhang Y-Q, Peha JM. Streaming video over the internet: approaches and directions. *IEEE Transactions on Circuits and Systems for Video Technology* 2001; **11**(3): 282–300.
6. Ahmed T, Mehaoua A, Buridant G. Implementing MPEG-4 video on demand over IP differentiated services. In *Proceedings of Globecom* 2001; vol. 4; pp. 2489–2493.
7. Shin J, Kim JW, Kuo CCJ. Quality-of-service mapping mechanism for packet video in differentiated services network. *IEEE Transactions Multimedia* 2001; **3**(2): 219–231.
8. Lee JYB. Optimizing channel allocation in a unified video-on-demand system. *IEEE Transactions on Circuits and Systems for Video Technology* 2002; **12**(10): 921–933.
9. Blake S, Black D, Carlson M, Davies E, Wang Z, Weiss W. An architecture for differentiated services. *IETF RFC 2475*, December 1998.
10. Xiao X, Ni LM. Internet QoS: a big picture. *IEEE Network* 1999; **13**(2): 8–18.
11. *Handover Procedures*. ETSI TS 100 527 V7.0.0, 1999.
12. *Handover Procedures*. 3GPP TS 23.009 V5.8.0, 2004.
13. Radha HM, van der Schaar M, Chen Y. The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP. *IEEE Transactions on Multimedia* 2001; **3**(1): 53–68.
14. Li W. Overview of fine granularity scalability in MPEG-4 video standard. *IEEE Transactions on Circuits and Systems for Video Technology* 2001; **11**(3): 301–317.
15. Wu F, Li S, Zhang Y-Q. A framework for efficient progressive fine granularity scalable video coding. *IEEE Transactions on Circuits and Systems for Video Technology* 2001; **11**(3): 332–344.
16. Wang Y, Zhu Q-F. Error control and concealment for video communication: a review. *Proceedings of the IEEE 1998*; **86**(5): 974–997.
17. Goyal VK. Multiple description coding: compression meets the network. *IEEE Signal Processing Magazine* 2001; **18**(5): 74–93.
18. Apostolopoulos J, Wong T, Tan W, Wee S. On multiple description streaming with content delivery networks. In *Proceedings of IEEE Infocom* 2002; vol. 3; pp. 1736–1745.
19. *Quality of Service (QoS) Concept and Architecture*. 3GPP TS 23.107 V6.1.0, 2004.

20. Heinanen J, Baker F, Weiss W, Wroclawski J. Assured Forwarding PHB Group. IETF RFC 2597, 1999.
21. Davie B, Charny A, Bennett JCR, Benson K, Le Boudec JY, Courtney W. *An Expedited Forwarding PHB (per-hop behavior)*. IETF RFC 3246, 2002.
22. UTRAN Overall Description. 3GPP TS 25.401 V6.4.0, 2004.
23. Chiussi FM, Khotimsky DA, Krishnan S. Mobility management in third-generation all-IP networks. *IEEE Communications Magazine* 2002; **40**(9): 124–135.
24. Schulzrinne H, Casner S, Frederick R, Jacobson V. RTP: a transport protocol for real-time applications. *IETF RFC 1889*, 1996.
25. Katsaggelos AK, Kondi LP, Meier FW, Ostermann J, Schuster GM. MPEG-4 and rate-distortion-based shape-coding techniques. *Proceedings of the IEEE* 1998; **86**(6): pp. 1126–1154.
26. Brady N. MPEG-4 standardized methods for the compression of arbitrarily shaped video objects. *IEEE Transactions on Circuits and Systems for Video Technology* 1999; **9**(8): 1170–1189.
27. Apostolopoulos JG, Wee SJ. Unbalanced multiple description video communication using path diversity. In *Proceeding of International Conference on Image Processing* 2001; vol. 1; 966–969.
28. Liu J, Shao H-R, Li B, Zhu W, Zhang Y-Q. A scalable bitstream packetization and adaptive rate control framework for object-based video multicast. In *Proceedings of IEEE Globecom* 2001; vol. 3; pp. 2020–2025.
29. Talluri R. Error-resilient video coding in the ISO MPEG-4 standard. *IEEE Communications Magazine* 1998; **36**(6): 112–119.
30. Ahmed T, Buridant G, Mehaoua A. Encapsulation and marking of MPEG-4 video over IP differentiated services. In *Proceeding of IEEE Symposium on Computers and Communications* 2001; pp. 346–352.
31. Floyd S, Jacobson V. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking* 1993; **1**(4): 397–413.
32. Ziviani A, de Rezende JF, Duarte OCMB, Fdida S. Improving the delivery quality of MPEG video streams by using differentiated services. In *Proceedings of 2nd European Conference on Universal Multiservice Networks* 2002; pp. 107–115.
33. Jill MB, Gaglianello RD. Packet loss effects on MPEG video sent over the public internet. *ACM Multimedia 98-Electronic Proceedings* 1998.
34. Bennett JCR, Zhang H. WF^{2}Q: worst-case fair weighted fair queuing. In *Proceedings of IEEE INFOCOM* 1996; pp. 120–128.
35. Zheng R. Scalable multiple description coding and distributed video streaming over 3G mobile networks. MASc Thesis, University of Waterloo, 2003.
36. OPNET Modeler 9.0 Documentation. OPNET Technologies, Inc., 2002.
37. Maniatis SI, Nikolouzou EG, Venieris IS. QoS issues in the converged 3G wireless and wired networks. *IEEE Communications Magazine* 2002; **40**(8): 44–53.
38. Kim IM, Kim HM. A new resource allocation scheme based on a PSNR criterion for wireless video transmission to stationary receivers over Gaussian channels. *IEEE Transactions on Wireless Communications* 2002; **1**(3): 393–401.
39. Moon B, Aghvami H. Diffserv extensions for QoS provisioning in IP mobility environments. *IEEE Wireless Communications* 2003; **10**(5): 38–44.
40. *Handover for Real-Time Services From PS Domain*. 3GPP TR 25.936 V4.0.1, 2001.
41. Sun X, Wu F, Li S, Gao W, Zhang Y-Q. Macroblock-based progressive fine granularity scalable (PFGS) video coding with flexible temporal-SNR scalabilities. In *Proceedings of Int. Conference on Image Processing* 2001; vol. 2; pp. 1025–1028.

Authors' Biographies

Ruobin Zheng received his B.E. degree from Xiamen University, China in 1996 and the M.A.Sc. degree from the University of Waterloo, Canada in 2003, both in Electrical Engineering. He is with Huawei Technologies, Co., Ltd, a telecom equipment manufacturer in China since 1999, first as an ASIC engineer involved in ATM processor design, then as a project manager in NGN media gateway product, and presently a senior system engineer in 802.16 broadband wireless access products.

Weihua Zhuang received her B.Sc. and M.Sc. degrees from Dalian Maritime University, Liaoning, China, and the Ph.D. from the University of New Brunswick, Fredericton, NB, Canada, all in Electrical Engineering. Since October 1993, she has been with the Department of Electrical and Computer Engineering, University of Waterloo, Ont., Canada, where she is a full professor. She is a co-author of the textbook *Wireless Communications and Networking* (Prentice Hall, 2003). Her current research interests include multimedia wireless communications, wireless networks and radio positioning. Dr. Zhuang received the Premier's Research Excellence Award (PREA) in 2001 from the Ontario Government for demonstrated excellence of scientific and academic contributions. She is an associate editor of *IEEE Transactions on Vehicular Technology* and *EURASIP Journal on Wireless Communications and Networking*.

Hai Jiang received his B.S. degree in 1995 and an M.S. degree in 1998, both in Electrical Engineering, from Peking University, Beijing, China. He is currently working towards Ph.D. at the University of Waterloo, Canada. His current research interests include QoS provisioning and resource management for multimedia communications in all-IP wireless networks.