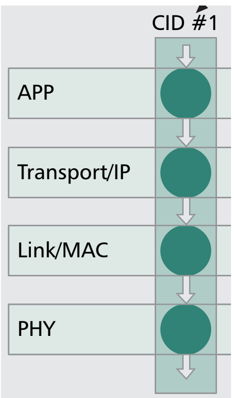
DOWNLINK SCHEDULING FOR MULTIMEDIA MULTICAST/BROADCAST OVER MOBILE WIMAX: CONNECTION-ORIENTED MULTISTATE ADAPTATION

HONGFEI DU, JIANGCHUAN LIU, AND JIE LIANG, SIMON FRASER UNIVERSITY



The authors examine the design issues and the state-of-the-art of the multimedia downlink scheduling in the multicast/broadcast-based WiMAX system. They also propose a viable end-to-end framework.

ABSTRACT

With its comprehensive QoS support and ubiquitous coverage, the Mobile WiMAX network offers promising opportunities for unwiring the last mile connectivity to Internet. However, stringent QoS demands of multimedia applications entail studies on service-oriented radio resource management. This article systematically examines the design issues and the state of the art of multimedia downlink scheduling in the multicast/broadcast-based WiMAX system. We propose a viable end-to-end framework, connection-oriented multistate adaptation, by considering cross-layer adaptations in source coding, queue prioritization, flow queuing, and scheduling. Its performance is confirmed by simulations on important metrics, showing that the framework can effectively accommodate heterogeneity in link variations, queue fluctuations, and reception diversities.

INTRODUCTION

For years, the IEEE has devoted continuous efforts to develop the wireless metropolitan area network (MAN) 802.16 standard, streamlined as the Worldwide Interoperability for Microwave Access (WiMAX) by the WiMAX Forum. This standard has since attracted a great deal of attention in both the research and industry communities, and is touted as the next killer technology that promises to offer *multiplay* services in the future wireless multimedia marketplace. The main advantages of WiMAX lie in its cost-competitive deployment and comprehensive quality of service (QoS) support for large numbers of heterogeneous mobile devices with high-data-rate wireless access. Since 2004, WiMAX has established its relevance as a wireless extension (or alternative) to conventional wired access technologies, such as T1/E1 lines, cable modems, and digital subscriber line (xDSL), extending the reach to remote areas. Mobile WiMAX, based on the IEEE 802.16-2004 [1] and IEEE 802.16e amendment [2], fills the gap between the wireless local area network (WLAN) and third-generation (3G) cellular systems with respect to their data rate and coverage trade-offs, and acts as a strong competitor to the current 3G Partnership Project (3GPP) long-term evolution (LTE) on the road to 4G wireless broadband markets.

On the other hand, recent advances in digital multimedia broadcasting (DMB) have offered the network operator a platform to deliver multimedia services to a mass market in a spectrum-efficient cost-effective way. A variety of initiatives, such as multimedia broadcast/multicast services (MBMS) [3], digital video broadcasting-handheld/satellite services for handhelds (DVB-H/SH), and terrestrial/satellite-DMB (T/S-DMB), and media forward link only (MediaFLO), have been envisioned to provide one-to-many content distribution to mobile users. The multimedia multicast broadcast service (MBS) [4] specified in 802.16e leverages on the successful features of the above technologies, while offering comprehensive and flexible support for mobile TV, IP audio/video streaming, live sport/entertainment events broadcast, and so on.

Notably, WiMAX introduces a novel concept of *connection-oriented* service flow, supporting fine-granularity QoS for both uplink and downlink on a per service flow basis. Unlike the *access-oriented* design on network QoS in conventional packet-oriented transmission, such a concept motivates us to better address the service-oriented design on per-service-flow QoS, as the service requirements are thus easy to comply with when they are associated with a connection rather than individual packets spanning the network. To this end, we exploit this unique feature in the downlink scheduling for multimedia MBS service, which is based on the 3GPP MBMS framework [3]. Furthermore, the advances of scalable video coding (SVC) technologies, such as H.264/MPEG advanced video coding (AVC), allow the delivery of video clips with a considerably reduced payload burden while accommodating a wide range of underlying network/link vibrations as well as receiver diversities.

 1536-1284/09/$25.00

Most of the existing works on WiMAX concentrate on uplink scheduling with diverse objectives [5–9]. We, however, pay attention to multisession MBS scheduling in multicast/broadcast (MC/BC)-based WiMAX, where fundamental design issues in downlink scheduling during session transmission must be addressed in a more adaptive and scalable manner, in order to incorporate the performance dynamics induced from channel variations, queue fluctuations, and terminal heterogeneities. In this article we propose a viable end-to-end cross-layer framework, connection-oriented multistate adaptation (CMA), which adopts the service-oriented design on per-service-flow connections carrying multisession MBS. The scheme features the following salient novelties. First, we explore essential optimization criteria for performing downlink scheduling in MC/BC-based WiMAX, while simultaneously taking into account multiple performance dynamic metrics. Second, we incorporate the video scalability feature in the application (APP) layer in conjunction with the unique connection-oriented system nature, where differentiated adaptations are performed for each connection carrying a single video substream. We identify several key design issues in the optimization objective function, and apply appropriate adaptations on source coding, prioritization, and reception evaluations. To our knowledge, such joint cross-layer adaptations on downlink scheduling in MC/BC WiMAX have not been well addressed in the literature [4–14].

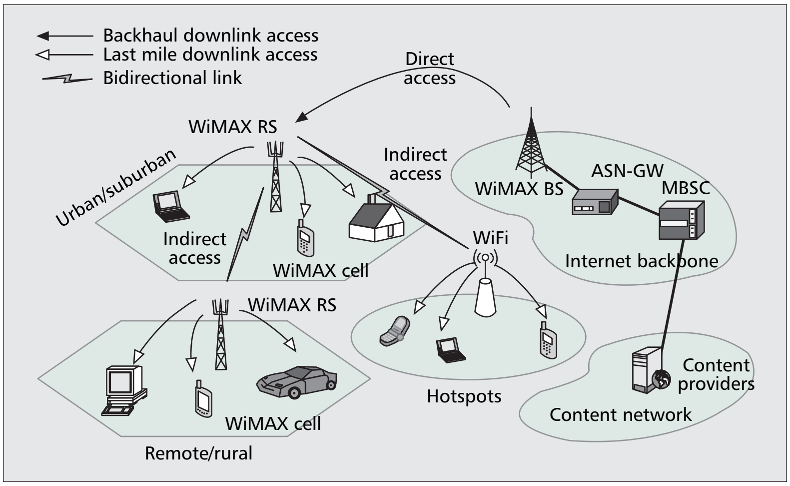
The rest of the article is organized as follows. We proceed with the preliminaries of MC/BC WiMAX and design challenges in downlink scheduling for MBS. We then survey various schemes studied in the literature. We then elaborate on the CMA framework by considering the downlink scheduling at the medium access control (MAC) layer for carrying heterogeneous MBS traffic types. We then discuss indicative performance results for our proposal, followed by our concluding remarks.

Background

WiMAX System Overview

As illustrated in Fig. 1, based on the MBS architecture in [4], MC/BC-based WiMAX defines a multicast/broadcast access network that offers ubiquitous access for heterogeneous MBS services. The MBS services are delivered from content providers to an MBS service controller (MBSC) before they are delivered to the subscriber station (SS) using either direct or indirect access. Being closely integrated into the Internet backbone, the system enjoys maximum reuse of existing technology and infrastructure, and therefore presents one of the most cost-efficient contenders among various last mile access technologies, such as IEEE 802.11-based WiFi and T1/E1 level services. The enhanced mobility support defined in 802.16e further gears the system to compete with 3G LTE cellular markets.

The 802.16e radio interface adopts orthogonal frequency-division multiple access (OFDMA), which is likely to emerge as the most preferred option due to the inherent features it



■ **Figure 1.** *Multicast/broadcast WiMAX system architecture for MBS service delivery.*

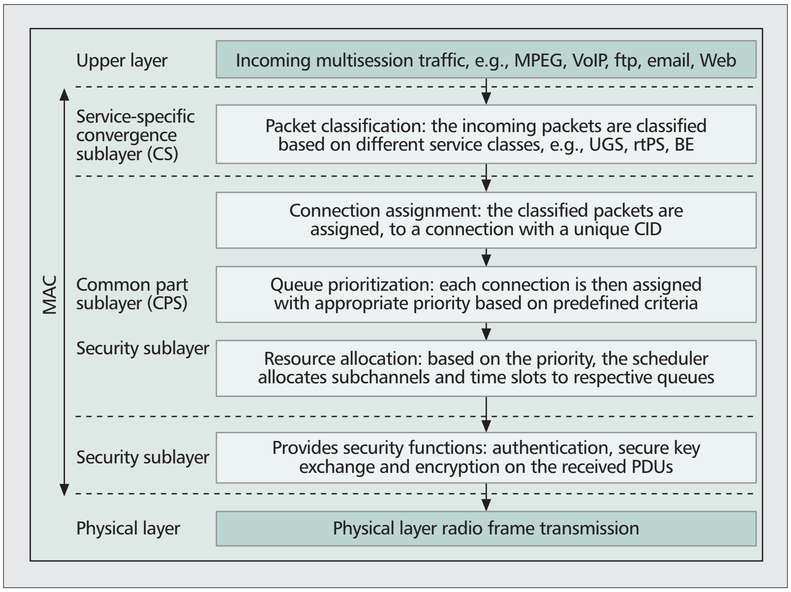
offers to tolerate multipath or selective fading in non-line-of-sight (NLoS) radio environments, to provide good bandwidth scalability, and to facilitate advanced antenna technology. We concentrate on the standard baseline WiMAX system in point-to-multipoint (PMP) mode, where the QoS is mainly enforced by the MAC and APP layers.

Medium Access Control

The MAC layer of MC/BC-based WiMAX is responsible for handling MBS sessions with diverse QoS requirements. It comprises three sublayers that interact with each other through the service access points (SAPs). Each service is mapped to one/multiple connection(s) through the convergence sublayer (CS) , where each connection is identified by a 16-bit connection identifier (CID), and service data units (SDUs) are classified to a connection based on specific QoS requirements. The queued data is then delivered to the common part sublayer (CPS), which is the core functional layer providing packing, concatenation/fragmentation, scheduling, and so on for transforming an SDU to a protocol data unit (PDU). The bottom sublayer in MAC provides security functions, independent of resource management functions and out of the scope of this article.

WiMAX supports two types of scheduling: *downlink scheduling* and *uplink request/grant scheduling*. The downlink scheduling in the base station (BS) determines the burst profile and transmission period for each connection for downlink traffic based on the QoS profile as well as channel/queuing related criteria. There is also a downlink scheduler at the SS for classifying the incoming packets into its subconnections. The uplink request/grant scheduling is performed by the BS with the intent of providing each subordinate SS with bandwidth for uplink transmission or opportunities to request bandwidth.

As illustrated in Fig. 2, during the session transmission period, the downlink scheduling is responsible for multiple consecutive tasks. Through the CS and CPS sublayers, the WiMAX BS performs centralized scheduling to direct users to transmit on and listen to a particular



■ **Figure 2.** *Downlink scheduling procedures at the BS.*

time slot and frequency channel. Given its connection-oriented per-flow transmission, rather than associating each connection with a basic CID, the bandwidth is associated with a MC/BC CID with a set of QoS parameters. Each service flow with a service flow identifier (SFID) is classified and uniquely identified by one/multiple connection(s) through the CS sublayer. The connections are served by a priority-based scheduler in a non-preemptive manner in accordance with their contracted QoS criteria. Being defined by OFDMA, a WiMAX scheduling slot is defined as a two-dimensional vector including both frequency (subchannel) and time axes [1].

QOS-BASED SCHEDULING IN WIMAX

QOS MECHANISM

The success of WiMAX lies in its comprehensive supports for a variety of dominant broadband services in a suite of QoS scheduling types [1]:

- **Unsolicited grant service (UGS)**: such as T1/E1 transport. It requires reserved traffic rate, maximum latency, and tolerated jitter.
- **Extended real-time polling service (ertPS)**: such as voice of IP (VoIP). It is built on the efficiency of both UGS and rtPS, reduces overhead and access delay of rtPS, and improves uplink resource utilization of the UGS. As a subcategory of rtPS, ertPS has the same QoS parameters as rtPS in the scheduler design.
- **Real-time polling service (rtPS)**: such as MPEG audio/video streaming and video conferencing. It supports variable bit rate (VBR) traffic via minimum reserved and maximum sustained traffic rates, and requires tolerable stringent latency constraints.
- **Non-real-time polling service (nrtPS)**: delay-tolerant streams with variable-sized packets, for which only minimum reserved and maximum sustained traffic rates are required, such as FTP.
- **Best effort (BE) service**: such as HTTP and email. BE services are handled on a space-available basis and do not require tight latency/jitter constraints, with upper limited bandwidth consumption via maximum sustained traffic rate.

For the MC/BC scenarios, the first three classes are designed for real-time applicationsm while the remaining services are designed for non-real-time applications. The goal of an efficient downlink scheduler in MC/BC WiMAX is to prioritize the incoming traffic flows and ensure the target QoS, in terms of the aforementioned performance profiles, to each WiMAX service class, while achieving fairness in the treatment of traffic flows with the same service class.

WIMAX SCHEDULING: STATE OF THE ART

It is noted that the 802.16e standard defines heterogeneous QoS service types; yet the scheduling algorithms to allocate the subchannel and time slots are not defined in the WiMAX specifications, but rather open for alternative implementations. Existing literature has witnessed the diverse effectiveness achieved by different scheduling algorithms, creating crucial hurdles for the design of a feasible and efficient scheduling scheme for mobile WiMAX. In the following we survey existing scheduling schemes that have been studied for WiMAX, and identify their key contributions and performance trade-offs.

To ensure a minimum bandwidth allocation and distribution, the Weighted Round Robin (WRR) and Deficit Round Robin (DRR) algorithms are evaluated in WiMAX in [5], for uplink traffic and downlink traffic respectively. The former allocates bandwidth to the SSs in proportion to the assigned weight to each class, e.g., QoS rank, which effectively performs QoS-differentiated resource allocation. DRR utilizes a deflect counter to better handle variable sized packet arrival pattern in the downlink traffic with low complexity. Authors in [11] address the scheduling issues at the BS of 802.16, where difference schemes are studied, e.g., Weighted Fair Queue (WFQ) and Earliest Deadline First (EDF). Each algorithm considers different design criteria; by regulating the weight associated to each session, it is possible to guarantee hybrid QoS metrics, and hence WFQ outperforms WRR in a VBR scenario. EDF allocates bandwidth based on a deadline associated with packets, and is best suited for delay-sensitive services (e.g., UGS and rtPS). By ranking mobile users in terms of their respective channel quality, Max C/I scheduling is capable of providing an upper bound of system capacity, but it encounters a starvation problem for users near the edge of a cell. In [12] key design issues for video broadcast over WiMAX for improving coverage, spectrum efficiency, and video quality are discussed. A queue-aware uplink bandwidth allocation and rate control mechanism is proposed in [6], where the bandwidth is adaptively allocated for polling service in the presence of higher-priority UGS service by exploiting the queue status information.

Recently, cross-layer scheduling optimizations have been extensively studied. The authors in [13] apply proportional fair scheduling in cross-layer adaptation, which is best suited for best effort traffic with no specific QoS requirements. Channel-aware and cross-layer adaptation between physical (PHY) and MAC layers have been extensively studied in wireless systems. Nevertheless, we notice that the majority of cross-layer methodologies for scheduling rely on either adaptive modulation and coding (AMC) at the PHY layer to combat the bad effects of wireless fading by utilizing the channel state information (CSI) [7], or dynamic resource assignment to compensate for the channel vibrations and errors [8]. However, this may not provide sufficient QoS differentiation and satisfy heterogeneous multimedia applications. To effectively exploit the service-oriented scheduling design, cross-layer interaction between the MAC layer and upper layers is crucial. Notable research in this category includes a cross-layer protection strategy introduced in [14] for scalable video transmission over WLANs, where a combination of MAC retransmission, application-layer forward error correction (FEC), bandwidth adaptive compression, and adaptive packetization strategies was used. However, it is not MAC-centric and thus cannot be applied directly to our system. Despite the above proposals for WiMAX uplink and downlink scheduling in literature, aspects related to the joint optimizations on source video coding and scheduling in WiMAX are not sufficiently stressed.

On the other hand, conventional downlink scheduling at the BS applies a simple and unified treatment to all traffic flows. The QoS differentiation is performed at SSs for the uplink via different scheduling algorithms, in a hybrid manner, tailored to their respective service types [8], such as EDF for rtPS/ertPS, WFQ/WRR for nrtPS, and first-in first-out (FIFO)/RR for BE. However, the operation at the SS introduces extra computational overhead and is not desired, especially for mobile users with power constraints. Therefore, we suggest performing the differentiated scheduling adaptation at the BS for downlink traffic, as the BS is normally considered to be without hard power constraints. We summarize important design issues pertinent to the scheduling problem in the MC/BC WiMAX as follows:

- Based on the WiMAX specifications, QoS differentiation is desired to be class-based, where heterogeneous service types defined in WiMAX are to be fully supported.
- The connection-oriented transmission entails flow-based resource assignment. This feature greatly facilitates the joint optimization between scheduling and flow differentiation.
- To guarantee the respective QoS bounds predefined in each service type, we suggest consistently monitor the instantaneous performance with target performance and effectively utilize it in the scheduling procedure.
- The scheduler should be aware of channel vibrations and queue fluctuations in order to combat the fast-varying wireless channels and queuing dynamics.
- To support multiple MBS sessions in MC/BC architecture, appropriate adaptations are needed for dynamically adjusting the transmitter side settings.
- Lastly, the scheduler should be designed with desired features in complexity, adaptability, flexibility, and scalability in order to be economically implemented in various large-scale systems.

Our work differs from existing work in that instead of developing a link layer technique, we investigate the joint cooperative optimizations of source coding, queue prioritization, and downlink scheduling in MC/BC-based WiMAX without violating the predefined specifications. Our aim is to provide a generic guideline to the design of a more efficient and practical downlink scheduler for the service-oriented MC/BC-based WiMAX network.

Our work differs from existing works in that, instead of developing a link-layer technique, we investigate the joint cooperative optimizations on source coding, queue prioritization, downlink scheduling in MC/BC-based WiMAX without violating the pre-defined specifications.

Connection-Oriented Multistate Optimization

Optimization Criteria

Efficient downlink scheduling for heterogeneous multimedia transmission entails simultaneous consideration among multidimensional profiles, such as metrics reporting bursty traffic, channel fluctuations, as well as receiver diversities. To measure those profiles, we define the service-oriented requirements designated for each session as a combination of multiple time-independent attributes, including traffic types, QoS targets, and performance thresholds. These constraints are determined during service establishment/renegotiation on a per-session scale, which remains constant for each connection during the session transmission. The joint optimization issue can be profiled into various dimensions:

- **QoS profile** ($\boldsymbol{P}_{QoS}$): It jointly considers end-to-end (E2E) delay factor ($\boldsymbol{F}_{ED}$), E2E packet loss rate (PLR) factor ($\boldsymbol{F}_{EP}$), and E2E throughput factor ($\boldsymbol{F}_{ET}$) for the ith session and is defined as $\boldsymbol{P}_{QoS}(i) = \boldsymbol{F}_{ED}(i) \cdot \boldsymbol{F}_{EP}(i) \cdot \boldsymbol{F}_{ET}(i)$, where each factor reflects the difference between the instantaneous performance and its application-specific target.
- **Queuing state profile** ($\boldsymbol{P}_{QS}$): It jointly considers multiple dynamic queuing metrics in terms of queuing delay, buffer occupancy, and overflow probability.
- **Channel state profile** ($\boldsymbol{P}_{CS}$): It reflects the overall reception condition for each session based on the reception conditions for all subscriber stations (SSs) within the MC/BC group.

Problem Formulation

Given the system constraints and the variety of services it supports, the downlink scheduling in MC/BC-based WiMAX is formulated in an unconventional manner, as shown in Fig. 3. We map the optimization problem onto an $m \times n$ multistate metric, where m is the number of functions in layers/sublayers considered, and n is the total number of active connections. Each element in the metric corresponds to a possible state for a specific function of a connection at the given scheduling slot. For each scheduling slot, we define the following metrics to derive the possible state element:

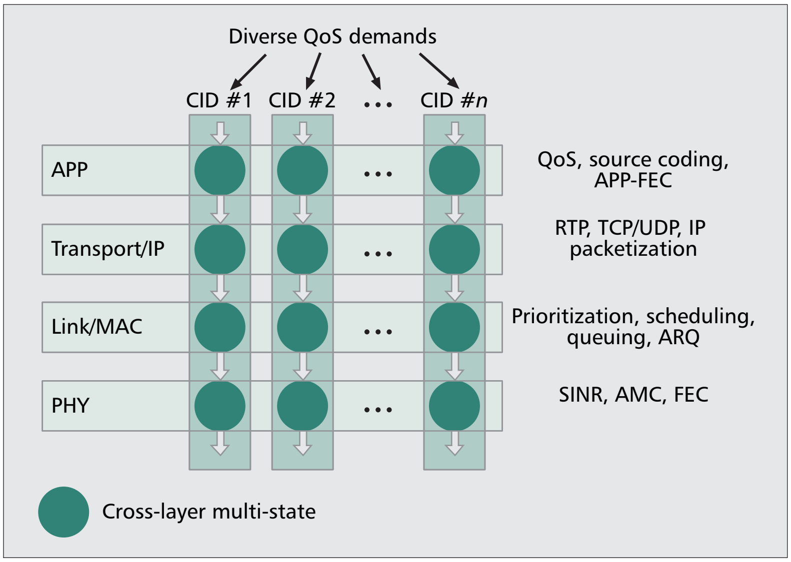


Figure 3. *Connection-oriented multistate adaptation.*

- ***Dynamic metric***: The instantaneous environmental dynamics, including queuing fluctuations, channel variations, and E2E performance dynamics.
- ***Threshold metric***: The predefined performance thresholds reflecting both service- and network-oriented QoS demands. It includes queuing thresholds, E2E thresholds, and minimum reception thresholds in terms of signal-to-interference-plus-noise ratio (SINR).
- ***Distortion metric***: The difference between the above two metrics, representing the current performance level with respect to its required level.
- ***Adaptation metric***: Based on the aforementioned metrics, the CMA scheme derives the most appropriate state of each function for each connection. Those functions include, for example, source coding, AP-FEC, scheduling, queuing, and modulation/coding scheme (MCS).

The dynamic metric is determined by aforementioned performance parameters: $\{P_{QoS}, P_{QS}, P_{CS}\}$. Each state in the adaptation metric is determined based on the current state as well as past historical states. Based on the derived states for the connection, the CMA scheme performs joint adaptation on functions across different layers. Such a mechanism can be applied to either PMP or mesh mode. In the former, optimizations across APP and MAC layers are most important, and also our focus in this article. In the latter, however, optimizations for the IP and transport layers become critical, because the channel/queuing states of the nodes along a multihop route have to be taken into account.

Queuing State

To maintain the queuing behavior and guarantee the service-oriented performance, a service flow has to take into account multiple performance metrics. The queuing behaviors in the queuing buffer at the BS are tracked, for each CID, against their target queuing thresholds in terms of:

- **Queuing delay factor** (F_{QD}): It includes the delay status for the packets that are currently queuing in the buffer, and for the packets that have been served and left the queue.
- **Buffer occupancy factor** (F_{BO}): It is defined as the ratio between the current queue length and its buffer length.
- **Overflow probability factor** (F_{OP}): It considers the packet drop rate due to buffer overflow.

These queuing factors keep tracking the current queuing dynamics and determine the overall queuing state P_{QS} for the ith session as $P_{QS}(i) = F_{QD}(i) \cdot F_{BO}(i) \cdot F_{OP}(i)$. It is worth noting that the threshold constraints can be managed in a more dynamic and flexible manner in order to incorporate bursty traffic loads induced by VBR video streams.

Effective Reception

The parameters abstracted at the SS side include two parts: channel state information (CSI) and E2E performance. We design an effective reception evaluation mechanism at the MAC layer of the SS that feeds back its current reception conditions to the WiMAX BS. The SS performs the following two important tasks:

- Monitors the CSI in terms of received SINR continuously, and sends this information to the BS
- Measures the instantaneous E2E performance in terms of delay, PLR, and throughput

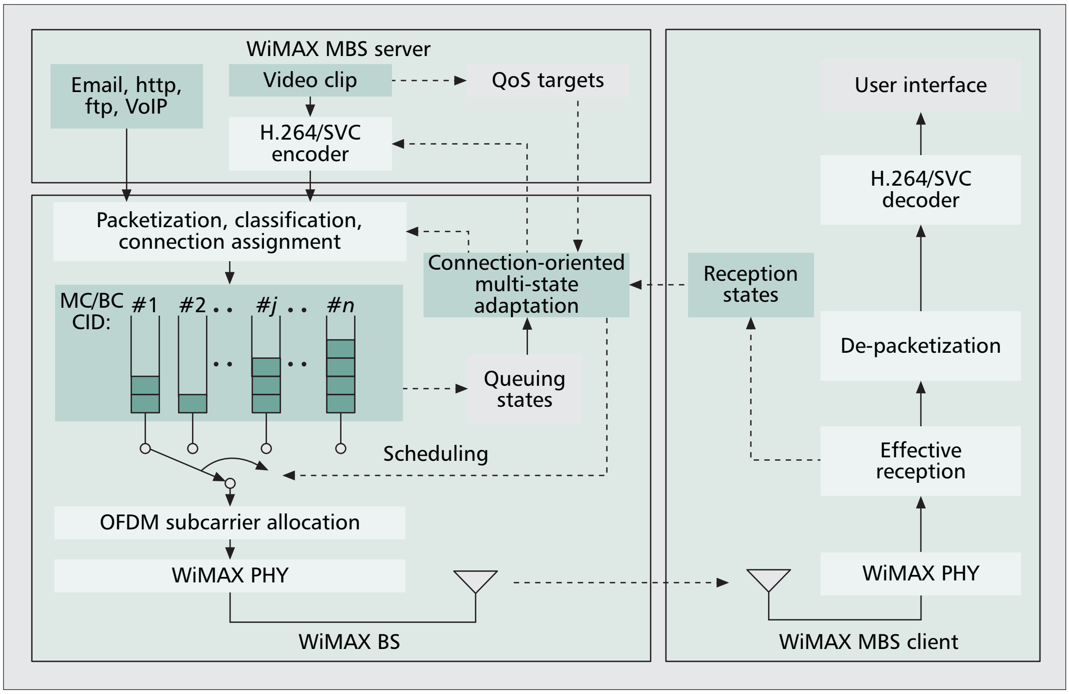
The instantaneous E2E performance is then compared with QoS constraints, and the following measures are determined:

- **E2E delay factor**: It is defined as the estimated historical mean delay performance over the maximum delay constraints. This factor is only effective, in a linear increasing manner, when the historical mean delay achieved so far is beyond its target.
- **E2E throughput factor**: This profile considers the E2E throughput, which is calculated as the ratio between the total bits released to the total bits that have been successfully scheduled and delivered to users up to the current scheduling slot.
- **E2E PLR factor**: This profile indicates the current packet loss performance from the transmission error in the link.

Both SINR and E2E factors at the SS are estimated and periodically fed back to the BS informing it on the current reception state of the MC/BC client group. Upon receiving the feedback information from the jth SS $P_{CS}(i,j)$ of the ith session with CID i, the BS subsequently derives the overall reception level $P_{CS}(i)$ for CID i associated with the MC/BC group as $P_{CS}(i) = Prob\{P_{CS}(i,j) \mid P_{CS}(i,j) > P_{CS}^*(i,j)\}$, where the $P_{CS}^*(i)$ is the minimum reception threshold specified for the ith session. This measure reflects the conditional probability of the number of SSs with a good reception condition over the total number of SSs in the MC/BC group, representing the instantaneous percentage of SSs whose reception levels are above the predefined thresholds.

Adaptive Queue Prioritization

As shown in Fig. 4, in MC/BC-based WiMAX the packets received by the BS and destined to the downlink are sorted by the packet classifica-



■ **Figure 4.** *The proposed CMA scheduling framework.*

To incorporate highly variable bit rates involved in link/queue dynamics, the CMA scheme adopts the Scalable Video Coding extension of H.264/AVC for cross-layer optimization between MAC and APP layers, so as to achieve graceful decoded video quality reduction.

tion before being buffered into one of the per class queues in the BS, according to the class of service (e.g., rtPS, nrtPS, and BE) to which they belong. Video streams are encoded via an H.264/SVC encoder before transmission together with other download services (email, ftp, etc.). Multiple connections/queues are then subject to scheduling functions based on the serving orders derived from the CMA module. The CMA considers multiple important metrics (i.e., the reception/queuing states, as well as the application QoS targets). Flows belonging to the UGS class are assumed to not be subject to MC/BC downlink scheduling operations, since their bandwidths are reserved consistently.

Upon receiving the instantaneous $\{\boldsymbol{P}_{QoS}, \boldsymbol{P}_{QS}, \boldsymbol{P}_{CS}\}$, the scheduler derives two important performance indices associated with each CID:

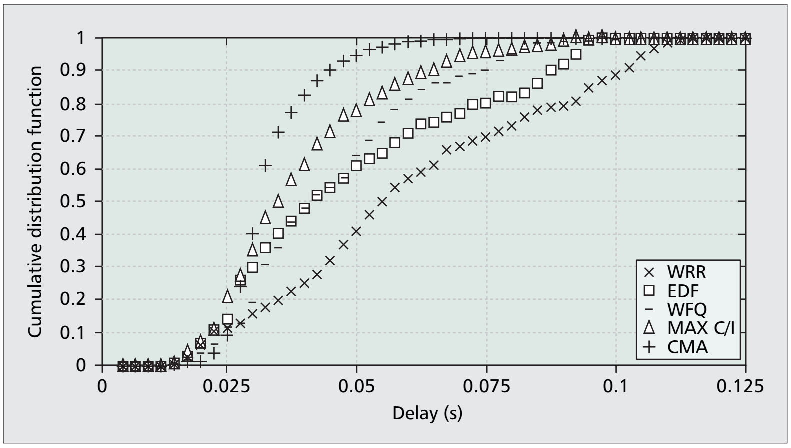
- **Demanding index** (**DI**): Depends on the factors currently under their performance targets, that is, how much resource the flow is demanding
- **Satisfaction index** (**SI**): Depends on the factors currently over their performance targets, that is, how much resource allowance the flow has for sharing with other flows

The serving orders of all competing flows are derived based on the above indices. It is noted that different weights on the DI and SI can be defined for each service class according to their QoS preferences.

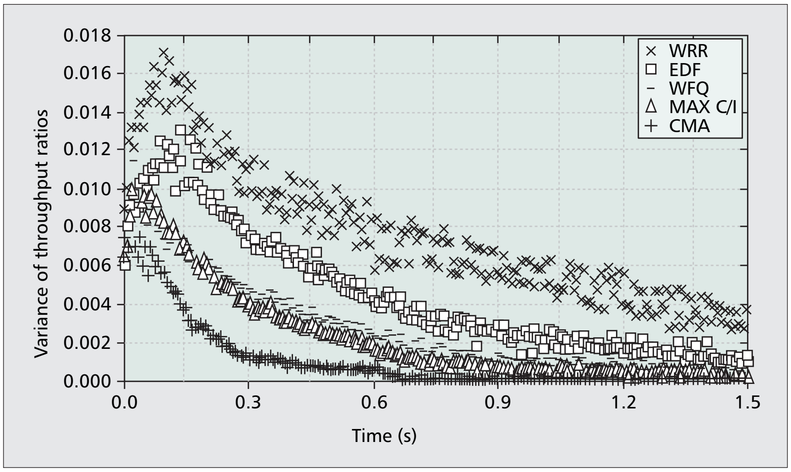
Source Coding Adaptation

To incorporate highly variable bit rates involved in link/queue dynamics, the CMA scheme adopts the Scalable Video Coding (SVC) extension of H.264/AVC for cross-layer optimization between MAC and APP layers so as to achieve graceful decoded video quality reduction. With SVC, a MBS video stream is split into a base layer (BL) and multiple enhancement layers (ELs) that progressively refine the video reconstruction quality. We consider all MC/BC connections to be originated from a single BS with no connection sharing or handover events. Given the native support of one-to-many mapping between SFID and CIDs, individual layered video streams from the same service flow are then carried in different connections. The BL for a group of frames is scheduled ahead of the next most important ELs for that group of frames. We assume the BL is successfully obtained by all users by applying robust modulation and coding or FEC/automatic repeat request (ARQ), while it also allocates additional resources for users with extra allowance to receive the ELs. To differentiate video sublayers for different CIDs, we associate different Real-Time Transport Protocol (RTP) payload types to the BL and the ELs. By inspecting the RTP payload type, various unequal error protection (UEP) strategies can be implemented for different video sublayers.

Notably, there are alternative ways to set the BL/EL rates for the MC/BC scenario. In general, the total rate should be capable of covering the full rate range of potential SSs. A simple setting on the BL rate would be the lowest decodable rate expected from the worst case SS; another may set the BL rate to the average expected rate of all SSs. We, however, find that those settings may lead to serious inefficiency and underutilization problems, especially when the instantaneous performance distribution in an MC/BC group is unpredictable and fast changing. We argue that the BL/EL stream rate setting should be designed closely in accordance with the overall reception condition for the specific MC/BC group. We suggest performing BL/EL rate control based on the QoS satisfaction of each SS, rather than unified treatment on all SSs; that is, the SSs with undersatisfied QoS will be assigned more important streams.



■ **Figure 5.** *The 95th percentile CDF of E2E delay for rtPS services under different downlink scheduling schemes.*



■ **Figure 6.** *Variance of overall throughput ratios under different downlink scheduling schemes.*

PERFORMANCE ANALYSIS

Our simulation methodology follows the recommendations of the WiMAX Application Working Group [15]. Our simulation consists of two parts: an H.264/MPEG-4 AVC JM reference software and a system-level simulation model in ns-2. Source coding functions including RTP are implemented in the JM reference software, where the video streaming trace is generated and converted to interpacket arrival time, and presents as the input traffic for the ns-2 simulator. RTP is implemented along with Real-Time Control Protocol (RTCP), which manages the traffic flow. RTCP provides feedback on the quality of the link, which can be used to modify encoding schemes if necessary. By using timing information, RTCP also facilitates synchronization of multiple streams, such as audio and video streams(BL/EL layers) associated with a session. BL/EL streams with different RTP payload types inspected from the JM model are treated independently as different traffic flows in the ns-2 system-level simulator.

We built a simplified realistic channel model where both slow fading and fast time-frequency fading components are modeled for the strongest interferer, while the remaining interferers are modeled as a classical Rayleigh fading process. To support high bandwidth multimedia contents, mobile WiMAX offers a maximum bit rate of 2 Mb/s [4] for an MBS channel rather than 384 kb/s for an MBMS channel in 3GPP. As we concentrate on the downlink scheduler at the BS for multisession MC/BC, an indicative simulation scenario is set to a total of 20 SSs located randomly within 5 × 5 km^2, with the BS located at the center of the area. We consider the following simple CID-SFID mapping settings to demonstrate our preliminary observations:

- CID 1-SFID 1: BL video stream at 360 kb/s, in VBR rtPS with 1.08 Mb/s sum stream rate
- CID 2-SFID 1: EL video stream at 720 kb/s, in VBR rtPS with 1.08 Mb/s sum stream rate
- CID 3-SFID 2: CBR nrtPS services at 360 kb/s
- CID 4-SFID 3: CBR BE services at 360 kb/s

We compare CMA with well-known algorithms WRR, WFQ, EDF, and Max C/I, and discuss performance gains achievable on different performance metrics.

We first measure the mean E2E delay statistics for rtPS service at the SSs, which are also obtainable at the BS via feedback channels. It is considered the sum of the propagation delay experienced by the wireless link, the processing and queuing delay, and the transmission delay over the return link. Figure 5 shows the 95th percentile cumulative distribution function (CDF) of packet delays. Compared to conventional schemes, it can be inferred that the rtPS service under CMA experiences much lower E2E delay and enjoys better delay variation (jitter) as it follows the steepest convergence with smoothest slope among all the schemes. Numerically, compared to other schemes, our simulation reports an average jitter reduction for rtPS services of 12.3 percent, while the average delay reduction for the nrtPS and BE services are 18.9 and 21.4 percent, respectively.

The fairness of the CMA scheme is studied by the throughput ratio, which is obtained by dividing the total bits successfully received by the SSs with the total bits released from the source. A lower variance indicates a fairer scheduling scheme. As shown in Fig. 6, different schemes are investigated during a sample simulation period. WRR and EDF achieve the worst performance of all, due to consideration of a single profile (i.e., either QoS or delay). WFQ and MAX C/I outperform the former schemes by taking into account the data rate and channel performance. Nevertheless, CMA achieves the lowest variance value with the fastest convergence curve and lowest max-min variations, which means it can provide SSs with better throughput equality in a short time.

CONCLUSION

In this article we survey the state-of-the-art scheduling schemes for provisioning quality services in WiMAX networks. We address the design problems encountered in conventional scheduling algorithms when applied to the multicast/broadcast scenario, and suggest a viable and effective cross-layer framework for downlink scheduling with multimedia traffic. The framework performs simultaneous adaptations across protocol stacks on source coding, queue prioritization, flow queuing, and scheduling. The heterogeneity in wireless link variations, queue fluctuations, and reception diversities are also incorporated. Simulation studies on the proposed framework showed improved performance on delay, throughput, and fairness.

References

[1] IEEE 802.16-2004, "IEEE Standard for Local and Metropolitan Area Networks — Part 16: Air Interface for Fixed Broadband Wireless Access Systems," Oct. 2004.

[2] IEEE 802.16e-2006, "IEEE Standard for Local and Metropolitan Area Networks — Part 16: Air Interface for Fixed Broadband Wireless Access Systems," Feb. 2006.

[3] 3GPP TS 23.246 v. 6.9.0, "Multimedia Broadcast/Multicast Service; Architecture and Functional Description (Release 6)," Dec. 2005.

[4] T. Jiang *et al.*, "Multicast Broadcast Services Support in OFDMA-based WiMAX Systems," *IEEE Commun. Mag.*, Aug. 2007, pp. 78–86.

[5] C. Cicconetti *et al.*, "Performance Evaluation of the IEEE 802.16 MAC for QoS Support," *IEEE Trans. Mobile Comp.*, vol. 6, no. 1, Jan. 2007, pp. 26–38.

[6] D. Niyato and E. Hossain, "Queue-Aware Uplink Bandwidth Allocation and Rate Control for Polling Service in IEEE 802.16 Broadband Wireless Networks," *IEEE Trans. Mobile Comp.*, vol. 5, no. 6, June 2006, pp. 668–79.

[7] Q. Liu, X. Wang, and G. B. Giannakis, "A Cross-Layer Scheduling Algorithm with QoS Support in Wireless Networks," *IEEE Trans. Vehic. Tech.*, vol. 55, no. 3, May 2006, pp. 839–47.

[8] A. Lera, A. Molinaro, and S. Pizzi, "Channel-Aware Scheduling for QoS and Fairness Provisioning in IEEE 802.16/WiMAX Broadband Wireless Access Systems," *IEEE Network*, vol. 21, no. 5, Sept.-Oct. 2007, pp. 34–41.

[9] K. Lu, Y. Qian, and H. H. Chen, "A Secure and Service-Oriented Network for WiMAX Networks," *IEEE Commun. Mag.*, 2007, pp.124–30.

[10] M. Chatterjee, S. Sengupta, and S. Ganguly, "Feedback-based Real-time Streaming over WiMAX," *IEEE Wireless Commun.*, Feb. 2007.

[11] K. Wongthavarawat and A. Ganz, "Scheduling for QoS Support in IEEE 802.16 Broadband Wireless Access Systems," *Int'l. J. Commun. Sys.*, vol. 16, 2003, pp. 81–96.

[12] J. Wang, M. Venkatachalam, and Y. Fan, "System Architecture and Cross-Layer Optimization of Video Broadcast over WMAX," *IEEE JSAC*, vol. 25, no. 4, May 2007, pp. 712–21.

[13] T. Kwon *et al.*, "Design and Implementation of a Simulator Based on a Cross-Layer Protocol between MAX and PHY Layers in a WiBro Compatible IEEE 802.16e OFDMA system," *IEEE Wireless Commun.*, vol. 43, Dec. 2005, pp. 136–46.

[14] M.van der Schaar *et al.*, "Adaptive Cross-layer Protection Strategies for Robust Scalable Video Transmission over 802.11 WLANs," *IEEE JSAC*, vol. 21, no. 10, Dec. 2003, pp. 1752–63.

[15] WiMAX Forum, "WiMAX System Evaluation Methodology V2.1," July 2008.

Biographies

HONGFEI DU [S'05, M'07] (hongfei.du@ieee.org) received a B.Eng. degree in electronic engineering from Beijing University of Aeronautics & Astronautics, China, in 2003. He received his M.Sc., M.Phil., and Ph.D. degrees in wireless communications from the University of Surrey, United Kingdom, in 2004, 2005, and 2007, respectively. From 2007 to 2008 he was with CREATE-NET International Research Institute, Italy, as a member of research staff, then project leader, coordinating and conducting EU research projects on middleware/software implementation, system architecture, and protocol design for the convergence between heterogeneous broadcast and mobile networks. Since 2008 he has been with the School of Computing Science and School of Engineering Science, Simon Fraser University, British Columbia, Canada, as a postdoctoral researcher and Ebco-Epic Fellow, working on adaptive video transmission over mobile WiMAX networks. His research interests lie in the area of mobile and satellite multimedia broadcasting, focusing on radio resource management, packet scheduling, QoS support, scalable video coding, and cross-layer design.

JIANGCHUAN LIU [S'01, M'03, SM'08] (jcliu@cs.sfu.ca) received his B.Eng. degree (cum laude) from Tsinghua University, Beijing, China, in 1999, and his Ph.D. degree from the Hong Kong University of Science and Technology in 2003, both in computer science. He was a recipient of a Microsoft Research Fellowship (2000), a recipient of a Hong Kong Young Scientist Award (2003), and a co-inventor of one European patent and two U.S. patents. He co-authored the Best Student Paper of IWQoS '08 and the Best Paper (2009) of the IEEE Multimedia Communications Technical Committee (MMTC). He is currently an assistant professor in the School of Computing Science, Simon Fraser University, and was an assistant professor in the Department of Computer Science and Engineering at the Chinese University of Hong Kong from 2003 to 2004. His research interests include multimedia systems and networks, wireless ad hoc and sensor networks, and peer-to-peer and overlay networks. He is an Associate Editor of *IEEE Transactions on Multimedia* and an editor of *IEEE Communications Surveys and Tutorials*. He is a member of Sigma Xi.

JIE LIANG [S'99, M'04] (jiel@sfu.ca) received his B.E. and M.E. degrees from Xi'an Jiaotong University, China, in 1992 and 1995, his M.E. degree from National University of Singapore in 1998, and his Ph.D. degree from Johns Hopkins University, Baltimore, Maryland, in 2003, respectively. Since May 2004 he has been an assistant professor at the School of Engineering Science, Simon Fraser University. From 2003 to 2004 he was with the Video Codec Group of Microsoft Digital Media Division, Redmond, Washington. His research interests include image and video coding, multirate signal processing, and joint source channel coding.

CMA achieves the lowest variance value with the fastest convergence curve and lowest max-min variations, which mean that it can provide SSs with better throughput equality in a short time.