# A Quality-Driven Cross-Layer Solution for MPEG Video Streaming Over WiMAX Networks

Ehsan Haghani, *Student Member, IEEE*, Shyam Parekh, *Senior Member, IEEE*, Doru Calin, *Senior Member, IEEE*, Eunyoung Kim, and Nirwan Ansari, *Fellow, IEEE*

***Abstract*—Extensive efforts have been focused on deploying broadband wireless networks. Providing mobile users with high speed network connectivity will let them run various multimedia applications on their wireless devices. Satisfying users with different quality-of-service requirements while optimizing resource allocation is a challenging problem. In this paper, we discuss the challenges and possible solutions for transmitting MPEG video streams over WiMAX networks. We will briefly describe the MPEG traffic model suggested by the WiMAX Forum. A cross-layer solution for enhancing the performance of WiMAX networks with respect to MPEG video streaming applications is explained. Our solution uses the characteristics of MPEG traffic to give priority to the more important frames and protect them against dropping. Besides, it is simple and compatible with the IEEE 802.16 standards and thus easily deployable. It is shown that the proposed solutions will improve the video quality over WiMAX networks.**

***Index Terms*—Cross layer, MPEG, quality-of-service, video, wireless, WiMAX.**

## I. Introduction

THE excessive demand for ubiquitous broadband wireless access has attracted tremendous investment from the telecommunications industry in the development and deployment of WiMAX networks. The WiMAX technology is promising to provide broadband wireless access to mobile users in the near future. It is expected that video streaming will be a very attractive application for the rapid deployment of WiMAX networks. The stringent quality-of-service (QoS) requirements including high bitrate and low latency are some of the challenges the service providers and network designers are confronting. Furthermore, the popularity of many online video servers such as YouTube will encourage an increasing number of users to watch video clips on their mobile devices.

The scarcity of available bandwidth in wireless networks has called for efficient resource management. WiMAX networks are based on the IEEE 802.16 standards which have defined different QoS classes to support a broad range of applications with varying service requirements. The IEEE 802.16 standards provide true QoS classes for different types of applications. As a result, in WiMAX networks, each traffic flow is mapped into an appropriate service class based on its service requirements and the user's service level agreements (SLAs). Selecting appropriate service classes with proper parameters to support the required QoS while not wasting the scarce resources is the key challenge that we address in this paper. We study the traffic characteristics of video streaming applications and will show that an application driven, traffic aware service classification will provide the WiMAX subscriber stations (SSs) with better video quality.

The importance of efficient resource management has prompted a keen interest in the research community on supporting video streaming applications in wireless networks. Reference [1] has reviewed the challenges of video streaming in wireless networks. It has also proposed a network adaptive rate control and cross-layer design for enhancing the overall received video quality. Many other research works have also considered feedback based video rate control [2]–[6]. In most of the rate adaptive methods, the server receives some information such as the available bandwidth, loss rate, buffer size at the receiver, or the end-to-end delay to adapt to the optimum video coding rate. One of the main drawbacks for the rate adaptive methods is caused by the channel variations in the wireless networks. Owing to the fast variation in the wireless physical channels, the adaptation methods may not be able to track the fast changes in radio channel conditions and adapt to the optimum rate accordingly. Furthermore, selecting the appropriate rate increases the computational complexity at the video server which can result in overloading the video streaming servers. Moreover, sending feedback is not a feasible option in some multicasting applications such as IPTV or MobileTV. In such cases, a video server transmits the same content to multiple receivers with different physical channels. The heterogeneity of receivers in these applications makes it very complicated for the server to attain the flexibility and sustain the efficiency.

In order to reduce the complexity at the server side and support various types of clients, scalable video coding (SVC) has been introduced in [7]. The goal of this method is to encode high quality video streams into some groups of bit streams including one base sublayer and multiple enhancement sublayers. All clients register to receive the base sublayer. The addition of enhancement sublayers improves the video quality. Thus,

Manuscript received October 01, 2008; revised April 21, 2009. Current version published September 16, 2009. This work was conducted when E. Haghani was an intern at Bell Laboratories in 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aggelos K. Katsaggelos.

E. Haghani and N. Ansari are with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ, 07102 USA (e-mail: Ehsan@njit.edu; Nirwan.Ansari@njit.edu).

S. Parekh, D. Calin, and E. Kim are with Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ 07974 USA (Tel.: (e-mail: Sparekh@alcatel-lucent.com; Calin@alcatel-lucent.com; Ekim@alcatel-lucent.com).

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

Digital Object Identifier 10.1109/TMM.2009.2026099

clients select the number of enhancement sublayers to receive based on their network connection and the availability of resources [8].

The challenging and crucial problem of video streaming over WiMAX networks has attracted many researchers. Reference [9] presents an adaptive SVC approach for streaming video on demand to the subscriber stations. Different WiMAX network architectures and cross-layer solutions for supporting the broadcasting and multicasting applications have been introduced [10], [11]. They take advantage of the broadcasting capabilities of WiMAX wireless medium and leverage this feature to deliver video multicasting applications such as IPTV to WiMAX customers. A channel based, rate adaptive solution for video streaming in WiMAX network has been introduced in [12]. Reference [13] illustrates an active frame dropping approach for streaming real-time video over IEEE 802.16 networks. In this active dropping approach, the base station drops a frame if it does not have enough confidence about successfully delivering a video frame within the application delay limit. The concept of active dropping at the video server for video streaming has been employed in different research works. A frame discarding solution based on the packet lifetime is introduced in [14]. In this method, frames that cannot meet the deadline are dropped by the video server or the intermediate routers. Reference [15] explains a priority based frame dropping algorithm. In response to a temporary bandwidth reduction, the video server selectively drops least effective frames. In this work, the authors have considered an MPEG video streaming system in which the video server determines the priority of each frame based on the frame type.

In this paper, we propose a cross-layer design to enhance the quality of MPEG video streaming for the end users in WiMAX networks. Our solution uses the characteristics of MPEG traffic to give priority to the more important frames and protect them against dropping. Unlike the approach proposed in [15], we do not increase the complexity of video servers since the frames are dropped at the BS. Moreover, the complexity of the BS is not increased as well. In fact, our proposed scheme is capable of being used by any WiMAX certified BS. In our method, we do not send real-time feedback to the video server. Thus, a video server will be able to support multiple clients simultaneously, and this makes our solution flexible for multicasting applications as well.

The outline of the rest of the paper is as follows. Section II explains the video streaming traffic model used in our research. In Section III, we provide an overview of the quality-of-service support in WiMAX networks. Section IV explains our proposed solution for video streaming in WiMAX. We will present our simulation results in Section V. Concluding remarks are given in Section VI.

## II. Video Streaming Traffic Model

The transmission of digital video over broadband communication networks is an important service. However, the existence of different video generating applications along with the availability of numerous networking technologies with different service features have made this an extremely challenging problem. Providing the required quality-of-service to the end users is a difficult problem requiring in-depth understanding of the video traffic characteristics.

The variety of video applications and networking technologies has called for implementation of a variety of video coder and decoder (codec) standards. These standards may be deployed in a broad set of applications ranging from low bitrate video conferencing to high bitrate movie streaming. However, the efficiency of these codecs is different for different applications. Valid traffic models accounting for the key video characteristics are required to investigate how best to transport video from different applications and codecs over different networks.

Moving picture expert group (MPEG) has a series of advanced video compressing standards, of which MPEG-2 and MPEG-4 are the most pervasive ones, and the latter is the latest and the most advanced one. All these standards rely on removing the redundant information of each frame by predicting the changes between subsequent frames. The idea of prediction is based on the fact that consecutive scenes have few differences and the information in their pictures is highly correlated. By coding the small differences between the scenes, much less data needs to be transported and thus achieving data compression.

MPEG-4 encodes the input video into a sequence of frames called group of picture (GOP). The number of frames in each GOP is typically constant. It is possible that an MPEG frame is fragmented into multiple IP packets when transmitted over an IP network. The MPEG-4 encoder divides each scene of the video into a number of consecutive GOPs. The number of GOPs generating a scene is a function of the scene complexity and compression ratio. There has been extensive research work on modeling the MPEG video traffic. Reference [16] has separated the video traffic into I, P, and B frames. The authors have modeled each type of frame separately and have also provided a model for the combined traffic. Reference [17] has also suggested a traffic model by separating the MPEG frame into three different types. The WiMAX Forum has adopted the traffic model proposed in these papers and has given the parameters determined from empirical video traces for different video applications such as video conferencing and video streaming [18]. A rather comprehensive work on modeling MPEG video traffic is given in [19].

As mentioned above, the MPEG coded videos are composed of three different frame types, i.e., I, P, and B. I (intra coded) frames are single still images used as the reference frame in each GOP. I frames are used for synchronization of all frames in a GOP. If a GOP is lost or corrupted, the next GOP will be built based on its I frame which is not coded using any other frames. P (predicted) frames are built by predicting the changes from the closest match in the preceding I or P frames of their GOP. However, B (bi-predictive) frames use previous I or P frame and the next P frame to predict the changes in the picture. Thus, the B frames are used to predict both the backward and forward changes in the motion. Based on these definitions, it is understood that these frames are interrelated, and some P and B frames are derived from each I frame in a GOP. Similarly, some B frames are also derived from each P frame. Therefore, loss of I or P frames will affect some other frames in their corresponding GOP and this will degrade the perceived quality. We will elaborate on this problem in Section V.

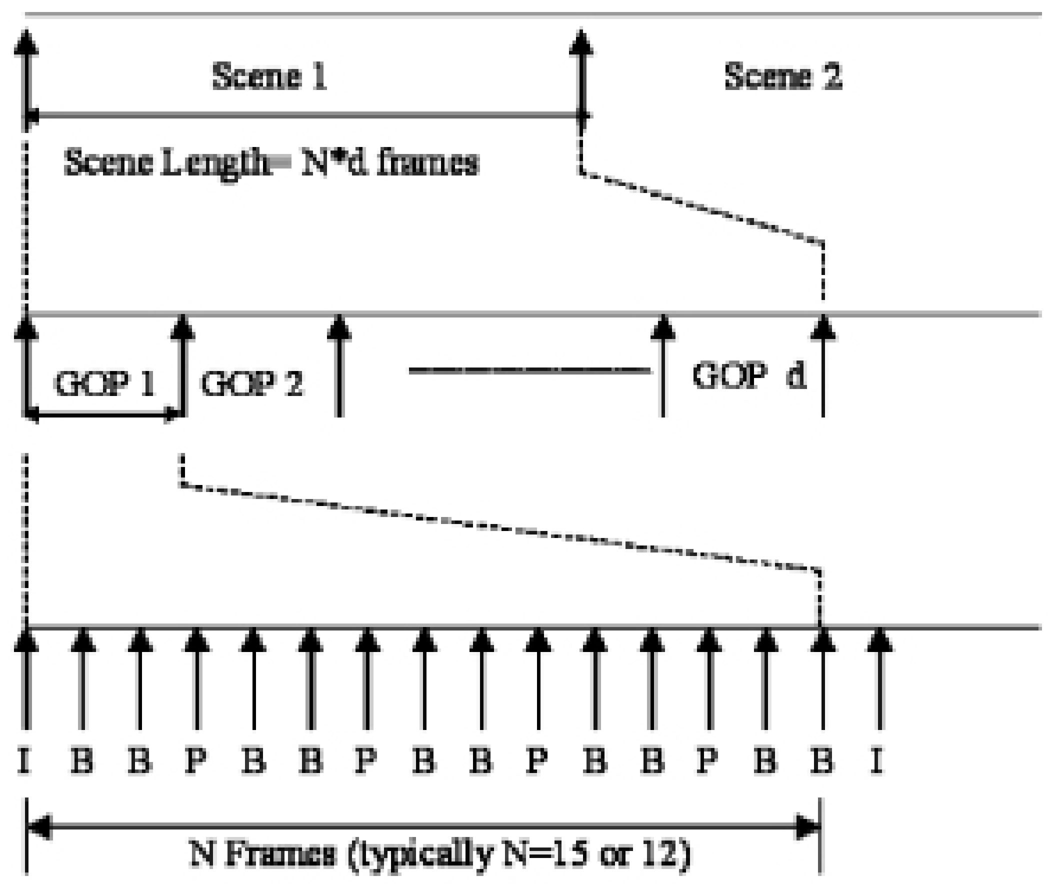


Fig. 1. Source model.

Although not required by the standard, MPEG encoders usually use a fixed pattern of frames in GOPs. The GOP pattern indicates the number of frames in each GOP, and their permutation order. In a regular GOP pattern, a GOP begins with an I frame and the number of B frames between I and P frames or between two P frames is constant. Such regular GOPs can be defined by two parameters: the I-to-I distance "$N$", and the I-to-P or P-to-P distance "$M$". A schematic illustrating the decomposition of video scenes into GOPs and formation of a GOP with $N = 15$ and $M = 3$ is depicted in Fig. 1.

The need for simulating the network performance has introduced different traffic models for different video applications such as video conferencing and video streaming. In this research, we use a traffic model for video streaming adopted by the WiMAX Forum [18]. In this model, I, P, and B frames are modeled separately, and a fixed pattern, similar to what is shown in Fig. 1, is used in building GOPs. As shown in Fig. 1, the number of GOPs in each scene is denoted by $d$, and it is modeled as a geometrically distributed random variable. We will explain the MPEG-4 traffic model next.

The I frames are modeled as variable bitrate (VBR) traffic. Based on the real MPEG traffic traces, the I frames have exhibited different behavior at different time scales. At the shorter time scales of a few seconds, the bitrate varies a little around a mean level. However, the mean level varies tremendously at larger time scales. The change of the mean levels in the large time scales is called the *scene* variation [20]. A scene is a short part of the movie that does not contain sudden changes in the view while it can possibly include some zooming or object movement. In our adopted traffic model, the concept of scene has been incorporated in the model, which results in more accurate performance prediction.

As explained above, the variations of the size of I frames have two scales: 1) the small variations within a scene period; 2) the large variations among different scenes. Thus, the model considers two independent components for defining the size of the $n$th I frame of the video stream, $X_I(n)$, located at the $k$th scene:

$$X_I(n) = \bar{X}_I(k) \quad \Delta_I(n). \tag{1}$$

TABLE I
MPEG MODEL PARAMETERS [18]

| Model Parameter | Value |
|---|---|
| Display size | 320x240 |
| GOP pattern | $N = 15$, $M = 3$ |
| Frame rate | 30 frames per second |
| Scene size parameter (d) | Exponential with p=0.1 |
| I frame size (bytes) | Log-normal ($\mu = 17068$, $\sigma = 7965$) |
| P frame size (bytes) | Log-normal ($\mu = 9190$, $\sigma = 7005$) |
| B frame size (bytes) | Log-normal ($\mu = 6839$, $\sigma = 5323$) |
| AR coefficients | $a_1 = 0.39$, $a_2 = 0.15$, $\sigma_\varepsilon = 4.36$ |

$\bar{X}_I(k)$ is the mean activity of scene $k$ and represents the large variations; thus, it may vary greatly from scene to scene. $\bar{X}_I(k)$ is constant for all I frames in scene $k$ while it will be different for other scenes. The $\bar{X}_I(k)$ is modeled by a log-normally distributed random variable [16]. $\Delta_I(n)$ represents the small variations of the I frames around the mean level of each scene. The $\Delta_I(n)$ is modeled by an order two autoregressive process, AR(2):

$$\Delta_I(n) = a_1\Delta_I(n-1) + a_2\Delta_I(n-2) - \varepsilon(n). \tag{2}$$

The $a_1$ and $a_2$ are assumed to be constant for each video stream and $\varepsilon(n)$ is a normal random variable with zero mean and constant variance for each stream [16]. The parameters defining the random variables depend on the content of the video; however, we will use a constant set of parameters in our simulations similar to what is adopted by the WiMAX Forum.

The sizes of P and B frames are modeled by log-normal distributions with parameters $(\mu_P, \sigma_P)$ and $(\mu_B, \sigma_B)$. The correlation between P frames (and similarly B frames) is negligible as compared to that of I frames, and thus the model considers them as independent random variables [16]. The MPEG model parameters used in this research are presented in Table I.

By considering the parameters displayed in Table I, we see that there are two GOPs per second. Thus, the model generates two I frames per second, eight P frames per second, and 20 B frames per second. Hence, the average bitrate for each frame type is as follows: $\bar{R}_I = 273$ Kbps, $\bar{R}_P = 588$ Kbps, and $\bar{R}_B = 1094$ Kbps. Thus the overall average bitrate for each video stream is $\bar{R}_{tot} = 1955$ Kbps. We observe that the average bitrate of B frames is higher than those of I and P frames. Although the average size of a B frame is less than that of other types, we have a higher bitrate associated with the B frames as there are more B frames in each GOP, as discussed earlier.

## III. QUALITY-OF-SERVICE IN WIMAX

WiMAX networks are planned to support traffic from many different applications. While the WiMAX technology can be designed to support the backhaul connectivity for broadband communications, it can also be tailored to provide wireless access to mobile users. Supporting different types of traffic requires flexibility in design and functionality. Owing to this requirement, there are many available options in the IEEE 802.16 standards that are to be chosen by vendors and service providers in their