

Optimal Rate Allocation for Scalable Video Multicast over WiMAX

Hsin-Yu Chi, Chia-Wen Lin and Yung-Chang Chen

Department of Electrical Engineering
National Tsing Hua University
Hsinchu, Taiwan
{cwlin,ycchen@ee.nthu.edu.tw}

Chih-Ming Chen

Chunghwa Telecom Co., Ltd
Telecommunication Laboratories
Taoyuan 326, Taiwan
cmchen2@cht.com.tw

Abstract—The IEEE 802.16 standard (commonly known as WiMAX), which has been proposed as a new wireless broadband standard, is capable of delivering very high data rate and covering wide area. Video multicast service would become one potential application over WiMAX with the popularity of streaming applications in the Internet. Our method mainly uses adaptive modulation to achieve the goal of rate-adaptive multicast, and combines with the concept of layered multicast. According to the size of video layer, SS distribution, and available symbols, our method adaptively changes the modulations of each video layer in each group of picture (GOP) time. We also propose a Genetic Algorithm (GA) to reduce computational complexity when finding optimal modulation. Experimental results show that the proposed method can achieve promising performance.

I. INTRODUCTION

With the popularity of streaming applications (such as audio and video) in the Internet, the demand for streaming device over wireless networks is increasing. Such demands are being addressed by variety of wireless access technologies such as, cellular, wireless LANs, and wireless MANs. Although the IEEE 802.11 (WiFi) standard has become the most prevailing technology for indoor access for mobile devices, the outdoor broadband wireless access is still being researched. Consequently, IEEE has defined a new standard for broadband wireless access, the 802.16 family of protocols.

There are two kinds of essential elements in a WiMAX system, one is BS (base station), and the other is SS (subscriber station) or CPE (customer premise equipment). The BS controls and manages all routine works in the network, including bandwidth resource allocation to provide QoS guarantees as well as fairness among the users. IEEE 802.16 provides flexible MAC with a variety of services having diverse quality requirements, including the real-time service with fixed bit rate (UGS), real-time service with variable bit rates and a bounded delay (rtPS), the non-real-time service with variable bit rates but insensitive to delay (nrtPS), and the best effort service (BE).

For video multicast, the key issue is to design an optimum bandwidth allocation scheme that considers the overall user satisfaction over an entire mechanism. The adaptive modulation and coding is one of the most effective techniques to improve the average throughput or to reduce wireless resource consumption. Kim and

Shin et al. [3] proposed an MP-AMC algorithm, which adjusts the number of time slots and modulation of each video layer of a scalable video. But this paper assumes that all slots are used in multicasting, without considering the condition of insufficient slots. In [3], it was shown that when the utility function is convex, the optimal utility problem can be formulated into a convex problem. However, it will be more complicated while more constraints are considered, such as the lower coding rate of modulation is necessary to be applied to the lower video layer. In [4], Hwang and Kim proposed an adaptive modulation and coding scheme in order to achieve better usage efficiency of spectrum. This method separates data into hierarchical layers including one base layer and multiple enhancement layers according to its importance.

In this paper, we propose an optimal rate allocation and symbols assignment scheme for video multicast over WiMAX. Our method mainly uses adaptive modulation to achieve the goal of rate-adaptive multicast, and combines with the concept of layered multicast based on scalable video coding (SVC) [6]. According to the size of video layer, SS distribution, and available symbols, our method adaptively changes the modulations of each video layer in each group of picture (GOP) time. We also propose a Genetic Algorithm (GA) [10] to reduce computational complexity when finding optimal modulation.

The remainder of this paper is organized as follows. The proposed optimal resource allocation scheme is presented in Section 2. Section 3 shows the simulation environment and the simulation results. Finally, conclusions are described in Section 4.

II. PROPOSED OPTIMAL RATE ALLOCATION SCHEME

The physical layer of IEEE 802.16 supports WirelessMAN-SC (10-66GHz), WirelessMAN-SCa (2-11GHz), WirelessMAN-OFDM (2-11GHz), and WirelessMAN-OFDMA (2-11GHz). This work is built on top of WirelessMAN-OFDMA system. The OFDMA PHY mode of IEEE 802.16 enables a BS to support multiple users at the same time. Furthermore, a BS utilizes the available channel by dividing the subcarriers into subchannels that can be assigned to multiple users in an adaptive and elaborate way. As a matter of fact, users can be assigned to different bandwidths and configured by different modulations based on various conditions of them. This allows WiMAX to support Adaptive Modulation and Coding (AMC) at the subchannel level for both the downlink and uplink. With the use of AMC, different modulations can be exercised independently in video multicasting. Individual bandwidth granularity described by bytes with different modulations applying at the subchannel level is summarized in TABLE 1.

TABLE I. BANDWIDTH GRANULARITY WITH AMC SUBCHANNELS [1]

Modulation Scheme	Bytes/subchannel
QPSK 1/2	6
QPSK 3/4	9
16-QAM 1/2	12
16-QAM 3/4	18
64-QAM 2/3	24
64-QAM 3/4	27

A. System Design

In WirelessMAN-OFDMA systems, a slot is defined as the minimum possible data allocation unit that requires both a time and subchannel dimension for completeness. The definition of an OFDMA slot depends on the OFDMA symbol structure. The type of symbol structure used in this paper is downlink partial usage of subchannel (PUSC). Accordingly, the relations between slots, subchannels, and symbols can be formulated as follows.

$$\text{Symbol} = \text{Slot} \times \text{Subchannel} \times 2 \quad (1)$$

In this paper, we only consider the channel propagation power that decreases with increase of the distance between BS and SS, and use it to represent the channel condition of each SS. The propagation model used in our experiment is the Two-Ray Ground Reflection model [5] as formulated below.

$$P_r = P_t \times G_t \times G_r \times \frac{h_t^2 \times h_r^2}{d^4} \quad (2)$$

where P_t and P_r represent the transmitter and receiver powers, respectively, G_t denotes the transmitter antenna gain, G_r is the receiver antenna gain, h_t and h_r respectively denote the heights of the transmitter and receiver, and d is the distance between the antenna of transmitter and the receiver in meters.

As for the video layers, we use the reference software (JSVM 2.0) [6] of the scalable extension of H.264 to generate $(L+1)$ video layers including one base layer and L enhancement layers according to the SNR scalability of SVC. We adopt the fine grain scalability (FGS) approach of SVC in which the transform coefficients are coded by using a progressive refinement mode to create a quality base layer and several quality enhancement layers. In order to maintain the essential video quality of all receivers, we need to ensure that all receivers can receive BL at least. Therefore, the modulation of BL should be chosen the most robust one for the entire multicast group in a video session. In other words, the modulation of BL is not necessarily QPSK but the modulation with lowest SNR in the entire multicast group. A receiver may receive one or more enhancement layers according to its channel condition.

B. Selection of Modulation Combinations

In the beginning, we consider the selection of modulations for a single video session provided that the number of available symbols is given. The network entry and initialization scheme in WiMAX MAC layer is required for a BS to scan a DL channel, gather the channel statistics of SSs, and acquire synchronization with SSs. The BS determines the maximum allowable receiving modulations of SSs according to their SNR that are obtained through the ranging processes. TABLE II shows the number of receivers in each modulation group, where m_i indicates the number of users in the i -th modulation group.

TABLE II. SS NUMBER OF EACH MODULATION GROUP

Modulation	QPSK	QPSK	QAM16	QAM16	QAM64	QAM64
Coding Rate	1/2	3/4	1/2	3/4	2/3	3/4
Index	0	1	2	3	4	5
Bytes/Subchannel	6	9	12	18	24	27
SS Number	m_0	m_1	m_2	m_3	m_4	m_5

To ensure each receiver could receive all data of BL, we choose the minimum one of the modulation groups whose SS number is not zero in a multicast session. As a result, the PER of BL for receiver c should be made zero. After allocating the available symbols to the data of BL, the remaining symbols are allocated to ELs according to the order of layer index. Furthermore, we assume the given number of available symbols must be larger than the symbols required for sending the BL data. Fig. 1 shows the allocation procedure if available symbol known in advance.

The demanded symbol number for the i -th layer, Symbol_{EL_i} , is determined by the layer size, the modulation scheme of layer, and the given number of available symbols. In addition, it is also affected by the characteristics of lower layers and the remaining number of symbols after transmitting the lower layers, as the symbols are allocated in order, i.e., BL, EL1, EL2..., EL(i). The symbol allocation process will be terminated until the symbol budget is used up. In order to accommodate the distribution of SSs, the size of each layer and the known number of available symbols, the modulation of each layer should be adaptively selected to utilize the given bandwidth most efficiently in each GOP time. Besides, the modulation selection scheme must meet some constraints, e.g.,

$$\text{Mod}_{BL} \leq \text{Mod}_{EL1} \dots \leq \text{Mod}_{EL(L)}.$$

To measure the performance of each combination of modulation schemes, we use the peak signal-to-noise ratio (PSNR) as the distortion metric in formulating the optimization problem. The symbol size allocated to each layer given a modulation combination $(\text{Mod}_{BL}, \text{Mod}_{EL1}, \dots, \text{Mod}_{EL(L)})$ is calculated as in (3), under the assumption that the available symbol number is more than that required for BL so that BL will be received completely.

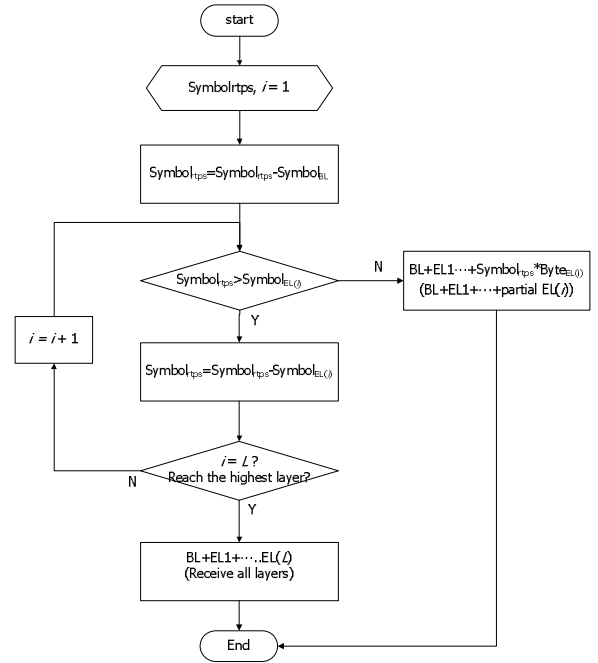


Fig. 1 Proposed symbol allocation procedure.

$$\text{Size}_{BL}(\text{Mod}_{BL}) = \text{BL}_{\text{GOP}} \quad (3)$$

$$\text{Size}_{EL_j}(\text{Mod}_{BL}, \text{Mod}_{EL1}, \dots, \text{Mod}_{EL_j}) = \begin{cases} EL_j \text{GOP} & , \text{if } \text{Symbol}_{\text{tps}} - \frac{\text{BL}_{\text{GOP}}}{\text{Byte}_{BL}} - \frac{\text{EL1}_{\text{GOP}}}{\text{Byte}_{EL1}} - \dots - \frac{\text{EL}_{j-1}\text{GOP}}{\text{Byte}_{EL_{j-1}}} \geq 0 \\ \left(\text{Symbol}_{\text{tps}} - \frac{\text{BL}_{\text{GOP}}}{\text{Byte}_{BL}} - \frac{\text{EL1}_{\text{GOP}}}{\text{Byte}_{EL1}} - \dots - \frac{\text{EL}_{(j-1)}\text{GOP}}{\text{Byte}_{EL_{(j-1)}}} \right) \times \text{Byte}_{EL_j} & , \text{else} \end{cases} \quad (4)$$

where $Symbol_{rtPS}$ is the available number of symbols reserved for the rtPS service. Then, we convert the allocated symbol size to the transmission rate for each layer.

$$Rate_{BL}(Mod_{BL}) = Size_{BL}(Mod_{BL}) \times \frac{R_{frame}}{N_{GOP}} \quad (5)$$

$$\begin{aligned} & Rate_{ELj}(Mod_{BL}, Mod_{EL1}, \dots, Mod_{ELj}) \\ &= Size_{BL}(Mod_{BL}, Mod_{EL1}, \dots, Mod_{ELj}) \times \frac{R_{frame}}{N_{GOP}} \end{aligned} \quad (6)$$

where R_{frame} presents the frame rate of sequence and N_{GOP} denotes the GOP size. The receiving rate of user c when the modulation combination is $(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)})$ can be expressed by (7) according to the transmission rate and PER of each layer, where $PER_{ELj,c}(Mod_{ELj})$ denotes the packet error rate of receiver c for the j -th EL which is modulated by Mod_{ELj} .

$$\begin{aligned} & Rate_c(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)}) \\ &= Rate_{BL}(Mod_{BL}) + [1 - PER_{EL1,c}(Mod_{EL1})] \times Rate_{EL1}(Mod_{BL}, Mod_{EL1}) + \\ & \dots + [1 - PER_{EL(L),c}(Mod_{EL(L)})] \times Rate_{EL(L)}(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)}) \quad (7) \\ &= Rate_{BL}(Mod_{BL}) + \\ & \sum_{j=1}^L [1 - PER_{ELj,c}(Mod_{ELj})] \times Rate_{ELj}(Mod_{BL}, Mod_{EL1}, \dots, Mod_{ELj}) \end{aligned}$$

As shown in (8), the PER of layers for receiver c is set either 0 or 1. We assume if the SNR of receiver c is larger or equal to the SNR of ELj determined by the modulation of ELj , the PER of ELj for receiver c is 0; otherwise, PER is 1.

$$PER_{ELj,c}(Mod_{ELj}) = \begin{cases} 0, & SNR_c \geq SNR_{ELj} \\ 1, & SNR_c < SNR_{ELj} \end{cases}, \text{ for each user } c \quad (8)$$

in a multicast session

Before calculating the PSNR value of each user, we need to find out the relation between the rate and distortion. We use the R-D function proposed in [7] to characterize R-D relation in each GOP as follows:

$$D_{GOP_k}(Rate_c) = \sigma_k^2 2^{-2\beta Rate_c} \quad (9)$$

where $Rate_c = Rate_c(Mod_{BL}, Mod_{EL1}, \dots, Mod_{ELj})$ in bit per pixel (bpp), σ_k^2 represents the variance of pixel value residual in GOP_k and β is a constant.

The PSNR of user c is then obtained by

$$\begin{aligned} & PSNR_{GOP_k,c}(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)}) \\ &= 20 \log \left(\frac{255}{\sqrt{D_{GOP_k}(Rate_c)}} \right) \end{aligned} \quad (10)$$

We define $Q(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)})$ as the average distortion of all users in one video multicast session.

$$Q(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)}) = \frac{1}{N} \sum_{j=0}^N D_{GOP_k,j} \quad (11)$$

The optimization problem is to find the optimal combination of modulations that minimizes $Q(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)})$ for each GOP under a set of constraints as formulated below:

$$\begin{aligned} & Min \{ Q(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)}) \} \\ & s.t. \begin{cases} Mod_{BL} \leq Mod_{EL1} \leq \dots \leq Mod_{EL(L)} \\ Mod_{BL} = \min \{ Mod_c \} \\ \text{Avalibale symbols for rtPS} = Symbol_{rtPS} \end{cases} \end{aligned} \quad (12)$$

In summary, we propose a scheme to find the optimal modulation combination under a given constraint on the number of available symbols in a video multicast session. The proposed algorithm is summarized as follows:

Algorithm: Exhaustive Search for Optimal Modulation Selection

```

 $N_m$ : the numbers of modulation scheme
for ( $Mod_{EL1}=Mod_{BL}; Mod_{EL1} < N_m; Mod_{EL1}++$ ) {
  if ( $Symbol_{rtPS} \leq Symbol_{EL1}$ ) {Opt;}
  else {
    ...
    for ( $Mod_{ELj}=Mod_{EL(j-1)}; Mod_{ELj} < N_m; Mod_{ELj}++$ ) {
      if ( $Symbol_{rtPS} \leq Symbol_{ELj}$ ) {Opt;}
      else {
        ...
        for ( $Mod_{EL(L)}=Mod_{EL(L-1)}; Mod_{EL(L)} < N_m; Mod_{EL(L)}++$ ) {Opt;}
      }
    }
  }
}
Opt: if video quality  $\geq$  (video quality) $_{max}$ , choose it as optimal modulation.

```

Using the above scheme, we can obtain optimal combinations of modulations according to a given rate budget, the distribution of SSs, and the size of each layer in each GOP period. The performance criterion is the average PSNR of all users in one video multicast system. Considering the different priority of users, the users with high priorities should be emphasized in modulation selection. By adding weight of each user in $Q(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)})$ as in (13), the characteristic of priority for different user can be revealed.

$$Q(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)}) = \frac{1}{N} \sum_{i=0}^N W_i D_{GOP_k,i} \quad (13)$$

where W_i denotes the weight of user i . By using weighted average method, the priority difference of user will be considered. But this method may sacrifice the users with low priority and cause unfairness condition. For this reason, the minimax optimization method [9] can be used for solving fairness problem as follows:

$$Q(Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)}) = \max_{i \in \{0,1,\dots,N\}} \{ D_{GOP_k,i} - D_{opt,c} \} \quad (14)$$

C. Complexity Reduction Using a Genetic Algorithm

Using the exhaustive search to find the optimal solution of (12) would result in high computational complexity, which is about $O(N^L)$, where L represents the number of layers. We propose to use a Genetic Algorithm (GA) to reduce the complexity while retaining the search accuracy. In the GA-based algorithm, a chromosome represents one modulation combination of video layers, and contains $L + 1$ genes: $Mod_{BL}, Mod_{EL1}, \dots, Mod_{EL(L)}$, each is the modulation type of video layers. At first, n chromosomes are generated randomly. So some chromosomes which may not obey constraints in (13) are called illegal chromosomes and should be corrected. For these illegal ones, we reassign the position of genes so that $Mod_{BL} \leq Mod_{EL1} \leq \dots \leq Mod_{EL(L)}$.

In our GA-based algorithm, we use the Roulette wheel selection algorithm [10] to perform fitness-proportional selection according to the fitness value of the chromosomes. The requirement is that the fittest individuals have a greater chance of survival than weaker ones. Besides, we use the one-cut-point crossover method [10] to obtain offspring with a crossover rate. Based on the mutation rate, some portion of genes will be changed. To ensure that the high-fitness chromosome is not lost from generation to generation, the best chromosome is copied into the new generation. Therefore, one

Elite chromosome is put into the next generation, yielding an offspring populations which have an improved average fitness. The fitness function in evaluation is defined as the video quality of system $Q(\text{Mod}_{BL}, \text{Mod}_{EL1}, \dots, \text{Mod}_{EL(L)})$. When the evolution is terminated, the chromosome with highest fitness value within all generations will be treated as a sub-optimal solution for optimization problem described in (12).

III. SIMULATION RESULTS

We adopt the ns-2 module for WiMAX developed by NDSL [8] as our simulation platform and implement the proposed modulation selection scheme as a part of the MAC layer protocol. The propagation model used in our experiment is the Two-Ray Ground Reflection model and the receiving power can be calculated by (2). The simulation parameters in (2) are $P_t = 25$ dBm, $G_t = 15$ dBi, $G_r = 1$ dBi, $h_t = 32$ m, $h_r = 1.5$ m, $T = 298$ K, and $NF = 7$ dB. We use the SVC JSVM 2.0 reference coder to generate the video layers. The QCIF (176x144) test sequences are pre-encoded at 15 fps. The GOP size is 16, and the number of FGS layers is 3. The total number of receivers is fixed as five. The coverage range of BS is a circle with a radius of approximate 39Km. The transmission from BS to SS is direct one-hop without any relay.

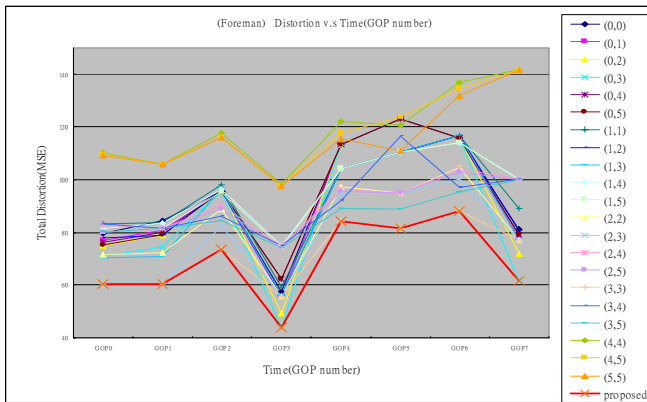


Fig. 2 Distortion of various EL modulations with time when SS distribution (0 1 2 0 1 1) and the number of available symbols = 900.

When the number of available symbols and the SS group number $(m_0 m_1 m_2 m_3 m_4 m_5) = (0 1 2 0 1 1)$, the variable factor in the optimal modulation selection scheme is the size of each video layer. Since the data sizes of GOPs are not the same, the sizes of layers are also different. We deal with this variable factor by performing the optimal modulation selection scheme for each GOP. Fig. 2 shows the optimal modulation is varying with the GOP number, where $(\text{Mod}_{EL1}, \text{Mod}_{EL2})$ denotes the selected modulation mode applied to EL1 and EL2, respectively. The impact on the size of each video layer in the optimal modulation selection scheme can also be observed. From another point of view, if we directly apply the optimal modulation for GOP #1 to the following GOPs, this may result in higher distortion in the following GOPs compared to their corresponding optimal results. Fig. 2 reveals the ability of adaptation of the proposed scheme as well as justifies that our scheme always maintains the lowest average distortion.

Fig. 3 shows the average distortion with varying available symbols for three optimization methods. According to this figure, the GA-based algorithm does not accuracy in the optimal modulation decision. Besides, since the method of minimax stresses the fairness problem, the performance is poorer than others.

The complexity of the exhaustive search lies mainly in the huge quantity of calculations in the symbols allocation procedure illustrated in Fig. 1. The run-time of the symbols allocation procedure in GA-based algorithm is significantly lower than that in the exhaustive search. Comparing with symbols allocation procedure, the

complexities of mutation and crossover could be neglected. For example, the exhaustive search method performs 216 (6^3) times of allocation procedure, whereas the GA-base algorithm only requires 49 times.

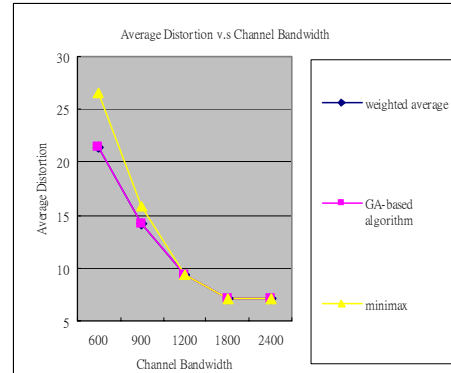


Fig. 3 Average distortion of different optimization methods in GOP #1 with different channel bandwidths.

IV. CONCLUSION

In this paper, we proposed an optimal rate allocation scheme for scalable video multicast over WiMAX. Our proposed method make best use of available symbols through adaptive modulation selection in each GOP time by taking into account the size of each video layer, SS group numbers and the available rate. The different performance metrics were exploited in the optimization. We have also proposed a GA-based scheme to reduce the computational complexity of the proposed rate allocation scheme. Simulation results show that the proposed method achieves lowest average distortion and the computational complexity can be significantly reduced with the GA-based approach. Additionally, other fast algorithms may be considered in the future to reduce computation superiorly.

REFERENCES

- [1] IEEE 802.16 std., *IEEE standard for local and metropolitan networks Part 16: Air interface for Fixed Broadband Wireless Access Systems*, Oct. 2004.
- [2] www.wimaxforum.org
- [3] J. Kim, J. Cho, and H. Shin "Resource allocation for scalable video broadcast in wireless cellular Networks," in *Proc. IEEE WiMob.*, vol.2, pp.174-180, Aug. 2005.
- [4] C. S. Hwang and Y. Kim, "An adaptive modulation method for multicast communications of hierarchical data in wireless networks," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, pp. 896-900, 2002.
- [5] T. S. Rappaport, *Wireless communications: Principles and practice*, Prentice Hall, Upper Saddle River, NJ, Oct. 1995.
- [6] ITU-T and ISO/IEC JTC1, JVT-Q202, *Joint Scalable Video Model JSVM-2*, Apr. 2005.
- [7] M. Dai and D. Loguinov, "Analysis of rate-Distortion Functions and Congestion Control in Scalable Internet Video Streaming," *Proceedings of ACM NOSSDAV.*, pp.60-69, June 2003.
- [8] J. Chen, C.C. Wang, Frank C.D. Tsai, and C.W. Chang, "The design and Implementation of WiMAX module for ns-2 simulator," in *Proc. ACM SESSION.*, vol. 202, no. 5, 2006.
- [9] C.-M. Chen, C.-W. Lin, and Y.-C. Chen, "Adaptive error resilience transcoding for video multicast over wireless networks," *Signal Processing: Image Communication*, vol. 22, no. 3, pp. 277-297, Mar. 2007.
- [10] M. Gen and R. Cheng, *Genetic Algorithms & Engineering Design*, John Wiley & Sons, Inc. 1996.