Scalable Modulation for Video Transmission in Wireless Networks

Lin Cai, *Senior Member, IEEE*, Siyuan Xiang, *Student Member, IEEE*, Yuanqian Luo, *Student Member, IEEE*, and Jianping Pan, *Senior Member, IEEE*

***Abstract*—In conventional wireless systems with layered architectures, the physical (PHY) layer equally treats all data streams from the upper layers and applies the same modulation and coding schemes to them. Newer systems such as Digital Video Broadcast start to introduce hierarchical modulation schemes with SuperPosition Coding (SPC) and support data streams of different priorities. However, SPC requires specialized hardware and has high complexity, which is not desirable for handheld devices. In this paper, we propose scalable modulation (s-mod) by reusing the current mainstream modulation schemes with software-based bit remapping. The performance evaluation has shown that s-mod can achieve the same and, in some cases, even better performance than SPC with much lower complexity. We further propose how to optimize the configuration of the PHY-layer s-mod and coding schemes to maximize the utility of video streaming with scalable video coding (SVC). Simulation results demonstrate substantial performance gains using s-mod and cross-layer optimization, indicating that s-mod and SVC are a good combination for video transmission in wireless networks.**

***Index Terms*—Scalable modulation (s-mod) and coding, scalable video coding (SVC), SuperPosition Coding (SPC).**

I. INTRODUCTION

VIDEO services are anticipated to be a major revenue generator for next-generation wireless networks. How to efficiently support video streaming over wireless networks is a key issue. In current wireless networks, the physical (PHY) layer mainly focuses on how to efficiently transmit information bits over the time-varying channel to approach the channel capacity. With the layered network architecture, the PHY layer treats every bit from the upper layer with the same priority and tries its best to reliably deliver each bit with the smallest possible bandwidth and lowest energy cost. However, for video applications, bits from the same flow may have different importance and impact on the user-perceived quality of service (QoS). For instance, scalable video coding (SVC) schemes may encode the video streams into a base layer and several enhancement layers, where the packet losses in the base layer lead to a much severer QoS degradation than those in the enhancement layers.

Although differentiated services have been an active research topic in wireless networks, how to fine-tune the PHY layer, particularly the modulation scheme, to more efficiently support layered scalable video streaming remains an open issue. SuperPosition Coding (SPC) and its implementation (so-called hierarchical modulation or h-mod) have been considered as a promising candidate for video multicast. However, SPC (h-mod) requires more complicated hardware, and therefore, it is costly to be adopted in the current wireless mobile systems such as Third-Generation (3G)/Fourth-Generation (4G) and WiMAX. In addition, the cross-layer optimization, considering the configuration of the PHY-layer modulation and coding schemes, to maximize the utility of scalable video multicast beckons for more intensive research.

The main contributions of this paper are threefold. First, we propose and design scalable modulation (s-mod) schemes, which can provide differentiated services to upper layer data streams, so the minimum video quality can be ensured, even when the wireless channel condition deviates from the estimation. The s-mod schemes can be implemented using the existing quadrature amplitude modulation (QAM) modem. As a software-based approach, the s-mod schemes just redefine (or remap) the constellation points of the existing QAM to modulate and demodulate the layered bits with different bit error probabilities (or bit error rate, BER). Second, we formulate a cross-layer optimization problem, aiming to maximize the profit (which equals the utility minus the cost) of scalable video multicast and unicast, by optimizing the configuration of the PHY-layer s-mod and coding schemes. Finally, extensive simulations with real videos and realistic wireless channel profiles are conducted to evaluate the performance of different modulation schemes and demonstrate the advantage of the proposed cross-layer optimization framework and the s-mod schemes.

The rest of this paper is organized as follows. Section II introduces background information and related works about adaptive modulation and coding (AMC), SPC/h-mod, SVC, and cross-layer optimization for scalable video streaming. In Section III, we propose scalable modulation and demodulation schemes. The cross-layer optimization problem is formulated in Section IV. Performance evaluation by simulation is given in Section V, followed by concluding remarks in Section VI.

II. PRELIMINARIES AND RELATED WORK

A. AMC

Wireless channels suffer from time-varying impairments due to user mobility, fading, shadowing, etc.; therefore, the received

---

Manuscript received May 10, 2011; revised August 31, 2011; accepted October 5, 2011. Date of publication October 19, 2011; date of current version December 9, 2011. This paper was presented in part at the IEEE INFOCOM, San Diego, CA, March 2010. The review of this paper was coordinated by Dr. N.-D. Dao.

L. Cai, S. Xiang, and Y. Luo are with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC V8W 3P6, Canada (e-mail: cai@ece.uvic.ca; siyxiang@ece.uvic.ca; yqluo@ece.uvic.ca).

J. Pan is with the Department of Computer Science, University of Victoria, Victoria, BC V8W 3P6, Canada (e-mail: pan@uvic.ca).

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/TVT.2011.2172823

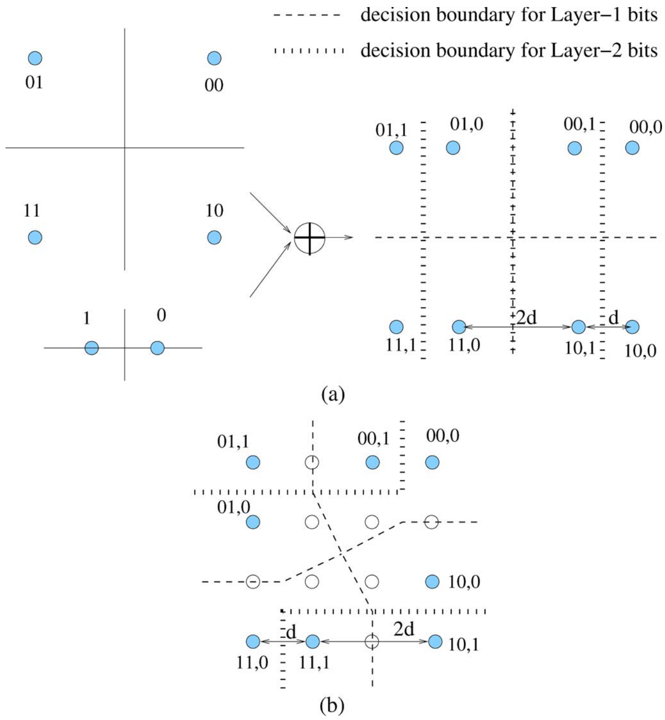


Fig. 1. S-mod versus SPC: an example. (a) SPC. (b) Scalable modulation.

signal-to-noise ratio (SNR) is changing. To maximize the application throughput with an acceptable BER, AMC schemes have been proposed and widely deployed in broadband wireless systems, e.g., the 3G/4G cellular systems [2] and WiMAX [3]. In short, instead of using fixed modulation and coding schemes, the receiver monitors the channel and predicts its condition. The receiver then feeds back the information to the transmitter, which will adjust the modulation and coding schemes accordingly for the next frame being transmitted. If the channel condition is relatively stable and the prediction is accurate, AMC can achieve high spectral efficiency on fading channels.

B. SPC and h-mod

Although AMC is preferred for unicast transmissions, it is not suitable for multicast or broadcast transmissions. First, getting real-time feedback from multiple receivers is nontrivial. Second, different users have different channel conditions, and they may prefer different modulation and coding schemes from the same transmitter from time to time.

Consider the scenario that a base station (BS) multicasts to a group of receivers in the downlink. To ensure reliable transmissions of all bits to all receivers, the BS has to choose the modulation and coding schemes that meet the BER requirement of the receiver with the worst channel condition at any time. On the other hand, for scalable video streams, receivers can enjoy the services, even if only part of the data is successfully received. For these applications, h-mod schemes based on SPC have been proposed [4].

The idea of h-mod (SPC) is to multiplex and modulate multiple layers of the data streams into a single symbol stream, where symbols for different layers are synchronously superimposed together before transmission [5]. For instance, as shown in Fig. 1(a), instead of separately transmitting two bits per symbol with quadrature phase-shift keying (QPSK) and one bit per symbol with binary phase-shift keying (BPSK), the transmitter can superimpose the QPSK symbol with the BPSK symbol and transmit the resultant symbol with three bits of information. On the receiver side, it first demodulates the bits of QPSK (the layer with a larger Euclidean distance) and then uses interference cancellation to decode BPSK (the layer with a smaller Euclidean distance). Because the minimum Euclidean distances (MEDs) of QPSK (Layer-1) and BPSK (Layer-2) are $2d$ and d, respectively, for receivers with good channel quality, they may successfully demodulate the bits from both layers; for those with bad channel quality, they may demodulate the Layer-1 bits only.

From an information-theoretical perspective, SPC can achieve a higher maximum sum rate of a Gaussian broadcast channel than the time-sharing schemes [4] (i.e., separately modulating and transmitting bits of different layers). However, existing h-mod schemes require specialized hardware for interference cancellation, and they suffer from interlayer interference. Specifically, the demodulation process for h-mod requires either hard- or soft-decision-based interference cancellation, which requires additional hardware support [6], and it will lead to a higher complexity and cost. Therefore, although h-mod has been adopted in the Digital Video Broadcast-Terrestrial (DVB-T) standard, it is more difficult to be used in wireless networks due to the hardware and performance limits. This fact motivates us to investigate this important area and propose the low-complexity scalable modulation schemes in Section III.

C. SVC

On the other hand, SVC [7] is an appealing coding technique for video/Internet Protocol television (IPTV) transmissions. It has been finalized as the ITU-T H.264 and ISO/IEC 14496-10 standards [8]. With a moderate increase in complexity and a slight decrease in efficiency, SVC provides spatial, temporal, and quality scalability. Sub-bitstreams with different types of scalability can be extracted from a single SVC-encoded bitstream. When different subsets of the bitstream are decoded, videos with different frame rates, resolutions, and quality can be reconstructed. This feature is naturally more efficient than simulcast or transcoding for heterogeneous network users. In particular, in a wireless multicast network, users can easily receive and decode the proper subbitstreams of an SVC bitstream due to the broadcast nature of wireless media.

A layered structure is essential for SVC. For each type of scalability, there is one base layer and several enhancement layers. Since there is a correlation between the lower layer (the base layer or the initial enhancement layers) and the higher layer, SVC exploits the information of the lower layer as much as possible to reduce the number of bits needed for the higher layer. For quality scalability, SVC provides medium- and coarse-grained quality scalability (CGS). Taking the two-layer CGS as an example, a video stream is sampled into a sequence of frames, and a number of (I, B, P) frames comprise a group of pictures. Encoded bitstreams can be divided into two layers: the base layer can provide the minimum satisfactory video quality; with the enhancement layer, better video quality can be achieved.

The reason SVC is more rate-distortion efficient than simulcast is that SVC utilizes interlayer prediction, which means that the information of the base layer is used to predict that of the enhancement layer; therefore, only the residual data of the enhancement layer and the base-layer prediction are encoded. However, when the base layer is corrupted or lost, the enhancement layer itself cannot improve the video quality or even decode the video bitstream.

Because of the importance of the base layer, more protection should be added to the base layer. Therefore, when we choose modulation and coding schemes, it is desirable to provide different services for different layers. This is another motivation for our proposed s-mod schemes, as well as existing work using h-mod for SVC [9].

D. Cross-Layer Optimization for Video Transmission in Wireless Networks

Resource allocation for supporting video transmission in wireless networks is a challenging issue, due to the time-varying wireless channels and the highly bursty video traffic. Given the wireless channel profile and video traffic model, we can use a fluid model to quantify the queue length distribution and packet loss rate, which can be used to determine the appropriate admission region for video/IPTV traffic [10]–[15]. In the literature, the theory of effective bandwidth and effective capacity has been proposed to obtain the probability of queuing delay exceeding a threshold, and thus, the statistical delay guarantee can be achieved for single-layer or layered video streams [16], [17]. These works mainly focused on the link layer and above, and did not consider how to manage the PHY-layer modulation schemes to support layered videos.

Recently, cross-layer optimization for multicast in wireless networks considering AMC began to attract attention. In [18]–[20], how to choose the right modulation and coding schemes to ensure efficient and reliable multicast was studied. How to optimize power allocation for SPC was proposed in [21]; however, the optimal configuration of s-mod cannot be obtained using the power allocation scheme for SPC; therefore, the solution in [21] is not applicable here.

Different from the previous approaches, in this paper, we first propose the s-mod schemes that can support differentiated services and layered videos without involving specialized hardware. Then, we propose a cross-layer optimization framework to appropriately configure PHY-layer transmission schemes, aiming to maximize the profit of videocast services.

III. SCALABLE MODULATION

A. An Example

The main objective for scalable modulation is to efficiently provide differentiated services to different classes of information bits. To achieve service differentiation, we can construct the signal constellation of the modulation scheme such that the MEDs of the bits in different layers are different.

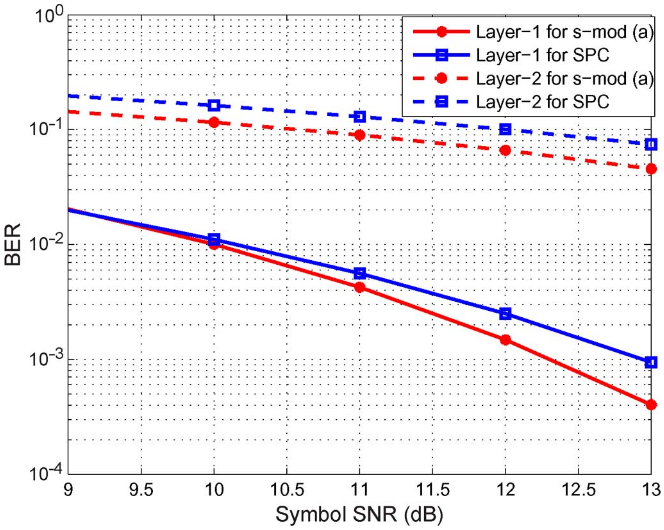


Fig. 2. BER comparison for s-mod and SPC.

Another design constraint for wireless mobile systems is the relatively simple transceiver design of handheld devices. Therefore, we aim to design the scalable modulation schemes based on the existing modulation hardware. As QAM schemes are typically supported in existing wireless systems, we can use a software-based approach to remap the constellations of the existing QAM modulation schemes to support differentiated services.

Fig. 1(b) gives an example of the proposed software-based s-mod scheme, which supports two layers of bits with the MED of $2d$ and d for layer-1 (L1), the high-priority layer, and layer-2 (L2), the low-priority layer, respectively. The demodulation decision regions can be obtained by finding the Voronoi diagram of the signal space for each layer. The probability of each layer's bits being successfully demodulated equals the probability that the received symbol is within their decision region.

The s-mod scheme is compared with the traditional SPC scheme shown in Fig. 1(a). In this paper, we use the same symbol SNR to compare different modulation schemes. For instance, with the received symbol SNR equal to 10 dB, error probability $e_1 = 0.010$ and $e_2 = 0.116$ for the s-mod scheme, and $e_1 = 0.011$ and $e_2 = 0.163$ for the SPC, where e_1 and e_2 are the BERs for L1 and L2 bits, respectively. As shown in Fig. 2, the s-mod scheme can outperform the traditional SPC (h-mod) because of its flexibility in arranging constellation points, so the minimum distance (normalized by the symbol energy) of the symbol to its decision boundaries for the bits in both layers can be larger.

B. S-mod for Single-Layer Bits

The main idea of s-mod is to remap the constellation points of existing modulation schemes to the bits in different layers. With more constellation points, there will be more design space for s-mod. In the current wireless systems such as WiMAX, up to 64-QAM is typically supported. Thus, in the following, we use the constellation of 64-QAM as an example to illustrate the s-mod design.

First, we investigate the case of transmitting single-layer bits using 64-QAM, where the MED of two adjacent constellation

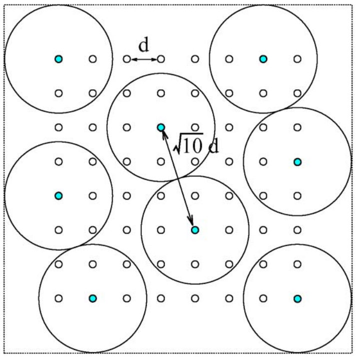


Fig. 3. S-mod for single-layer bits for the $k = 3$ case.

points in 64-QAM is d. Finding the maximum number of bits that can be transmitted with MED equal to αd can be formulated as a *circle-packing* problem: within the square area of $(7+\alpha)d \times (7+\alpha)d$, find how many circles (with the centers located at one of the constellation points in 64-QAM and radius equal to αd) can be packed. Solving the circle-packing problem, we obtain the following results:

Theorem 1: To transmit $k = 6, 5, \ldots, 1$ bits per symbol by mapping them to the conventional 64-QAM constellation, the maximum MEDs of the mapped bits are d, $\sqrt{2}d$, $2d$, $\sqrt{10}d$, $7d$, and $7\sqrt{2}d$, respectively.

The mappings to $k \in \{1, 2, 4, 5, 6\}$ single-layer bits are straightforward, and the case for $k = 3$ is shown in Fig. 3. The proof of the theorem is omitted.

C. S-Mod for Two-Layer Bits

The design for the two-layer s-mod is more complicated since there are many more choices to choose $2^{k_1+k_2}$ points from the 64 constellation points, where k_1 and k_2 are the number of L1 and L2 bits per symbol, respectively. Since L1 bits have a larger MED than that of L2 bits, we can use the single-layer s-mod design to divide the constellation point regions for L1 bits, which can significantly reduce the complexity of finding the appropriate two-layer s-mod schemes.

The procedures for finding the mapping for two-layer bits with MEDs equal to αd and βd, where $\beta < \alpha$, respectively, are listed here.

Step 1) Find the constellation points corresponding to transmitting k_1 single-layer bits (with the maximum MED according to Theorem 1).

Step 2) Find the Voronoi diagram of the aforementioned points.

Step 3) Construct circular rings with outer and inner radii equal to $\alpha/2d$ and $\beta/2d$, respectively, centered at one of the constellation points.

Step 4) For each Voronoi polygon, find 2^{k_2} circular rings such that the outer circle of the ring cannot exceed the Voronoi polygon, and the inner circle of the ring cannot overlap with each other.

Step 5) If the search in the fourth step fails, report failure.

The computational complexity of the search procedure is mainly due to the fourth step, which needs to search $k_1\binom{2^{6-k_1}}{2^{k_2}}$ choices at most. Since $k_1 + k_2 \le 6$, the computational complexity is tolerable, as the search procedure can be done offline.

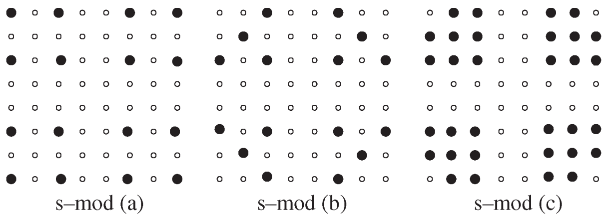


Fig. 4. S-mod schemes for two-layer bit mapping.

A failed search in the fourth step means that we can only support fewer bits for the two layers, or the MEDs of the two layers should be reduced.

Remarks: Both SPC/h-mod and s-mod can transmit bits with different BERs in the same symbol. By comparing SPC/h-mod and s-mod, we note, first, that h-mod can use fine-grained power allocation to change the Euclidean distances of constellation points to reach the L1 and L2 decision boundaries, whereas s-mod is confined to the limited number of constellation points in a QAM modulator. However, if we use a sufficiently dense QAM to design s-mod, we can find the equivalent s-mod design for any h-mod scheme. For example, the one shown in Fig. 4(a) is equivalent to an h-mod scheme. Thus, s-mod can be viewed as a general case of all h-mod schemes, and the performance of s-mod will be at least as good as that of h-mod. Second, s-mod schemes can be more flexible than h-mod, such as those in Fig. 1(b) or Fig. 4(b) and (c), so it can even outperform h-mod in some cases. Using the example in Fig. 1, given the flexibility in arranging the constellation points with s-mod, for the same SNR, both L1 and L2 bits can have lower BERs than those with h-mod, as shown in Fig. 2. Third, as discussed in Section III-A, s-mod is a software-based solution using the existing QAM modem, which is easier to implement, and therefore, it is more desirable for mobile devices with limited computation power and energy supply.

D. S-Mod Demodulation

Ideally, the receiver should construct the Voronoi diagram to define the decision region of each symbol for demodulation. This might require a more complicated demodulator. To simplify the implementation, it is possible to use the existing 64-QAM demodulator and map the demodulated 6 bits per symbol to the layered bits being modulated.

To minimize the mapping error, the mapping criteria are shown in Fig. 5. Using the 64-QAM demodulator, for the 64 constellation points, $2^{k_1+k_2}$ of them are corresponding to an s-mod symbol. For other 64-QAM constellation points, each point has one, two, or four nearest s-mod symbols, so it will be mapped to one of the nearest symbols with equal probability. For instance, as shown in Fig. 5, if the 64-QAM demodulates the symbol to be in V_5, it will be randomly mapped to s_1 (or s_2, s_3, and s_4) with an equal probability of 1/4. This simplified demodulation scheme unavoidably introduces more bit errors, which is the cost of using the existing demodulation hardware.

E. BER Estimation

If the Voronoi diagram is used to define the decision regions, the probability of demodulating bits in error is the probability

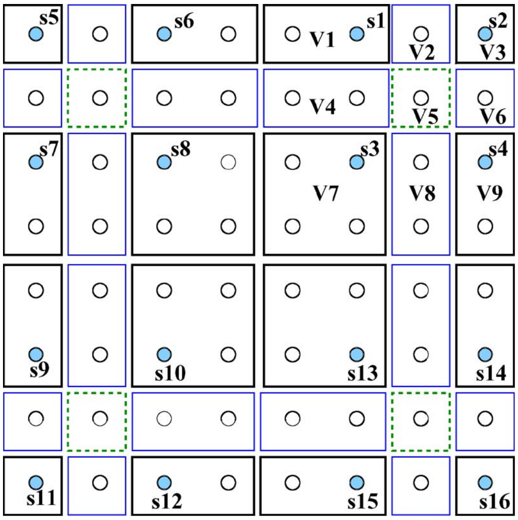


Fig. 5. Demodulation using reversed mapping.

that the received signal is located outside the decision region of the transmitted symbol.

Use the s-mod (a) in Fig. 4 as an example. Given an additive white Gaussian noise channel, the BER of the L1 bits for s-mod (a) in Fig. 4 is

$$\begin{aligned} e_1 &= \sum_{s_i} \Pr\left\{\text{received signal outside } S_i^{(1)} | s_i\right\} \Pr\{s_i\} \\ &= Q\left(3\sqrt{\frac{\gamma_s}{29}}\right) + Q\left(7\sqrt{\frac{\gamma_s}{29}}\right) - \frac{1}{4}Q^2\left(3\sqrt{\frac{\gamma_s}{29}}\right) \\ &\quad - \frac{1}{4}Q^2\left(7\sqrt{\frac{\gamma_s}{29}}\right) - \frac{1}{2}Q\left(3\sqrt{\frac{\gamma_s}{29}}\right) Q\left(7\sqrt{\frac{\gamma_s}{29}}\right) \end{aligned} \tag{1}$$

where $S_i^{(1)}$ is the decision region for the L1 bits corresponding to s_i, γ_s is the average received SNR, and $Q(\cdot)$ is the Q-function.

The BER of the L2 bits can be estimated by

$$\begin{aligned} e_2 &= \sum_{s_i} \Pr\left\{\text{received signal outside } S_i^{(2)} | s_i\right\} \Pr\{s_i\} \\ &= 2Q\left(2\sqrt{\frac{\gamma_s}{29}}\right) + Q\left(3\sqrt{\frac{\gamma_s}{29}}\right) - Q^2\left(2\sqrt{\frac{\gamma_s}{29}}\right) \\ &\quad - \frac{1}{4}Q^2\left(3\sqrt{\frac{\gamma_s}{29}}\right) - Q\left(2\sqrt{\frac{\gamma_s}{29}}\right) Q\left(3\sqrt{\frac{\gamma_s}{29}}\right) \end{aligned} \tag{2}$$

where $S_i^{(2)}$ is the decision region for the L2 bits corresponding to s_i.

If using the existing 64-QAM demodulator and mapping the demodulated bits into the layered bits, as shown in Fig. 5, the BER of L1 bits is the same as that in (1), and the BER of L2 bits is calculated as follows:

$$e_2 = 1 - (P_{s1} + P_{s2} + P_{s3} + P_{s4})/4 \tag{3}$$

where

$$P_{s1} = P_{V1} + (P_{V2} + P_{V4})/2 + P_{V5}/4 \tag{4}$$
$$P_{s2} = P_{V3} + (P_{V2} + P_{V6})/2 + P_{V5}/4 \tag{5}$$
$$P_{s3} = P_{V7} + (P_{V4} + P_{V8})/2 + P_{V5}/4 \tag{6}$$
$$P_{s3} = P_{V9} + (P_{V6} + P_{V8})/2 + P_{V5}/4 \tag{7}$$

and P_{Vi} is the probability that the received signal is located in region V_i.

For instance, when the received symbol SNR is 15 dB, the BERs of L2 bits using Voronoi diagram demodulation and the mapping from 64-QAM demodulated bits are 3.73×10^{-2} and 1.44×10^{-1}, respectively, and the BERs for L1 bits are 8.66×10^{-4} using both demodulation schemes. In the following, we use the demodulation by software mapping at the receiver side as it is easier to implement.

IV. CROSS-LAYER OPTIMIZATION FRAMEWORK

A. System Model

The network considered is an infrastructure-based wireless system. A BS covers the area of a cell, and it can multicast videos to the mobile stations (MSs) within the cell in the downlink. The video streams are coded with the layered SVC codec.

Concatenated Reed-Solomon (RS) and convolutional coding with interleaving has been used to enhance the BER performance in existing wireless systems, and it is also adopted in our system. As shown in Fig. 6, the layered video streams are encoded by the RS encoder and interleaved before being fed into the convolutional encoder. At the multicast receiver side, the procedures are simply reversed. The concatenated RS and convolutional code can effectively reduce the BER, which is essential for the QoS guarantee of video applications. Given a two-layer s-mod scheme, the ratio of the transmission data rates of the two layers is given, which might be different from that of the coded layered video streams. The leftover bits from one of the layers will be fed into the s-mod mapping for single-layer bits, as indicated by the dashed lines in Fig. 6. The s-mod scheme for the single-layer leftover bits is chosen such that the BER performance of these bits is close to the BER of the other bits from the same layer. The signaling procedure of which modulation scheme is used can be the same as that for AMC, i.e., in preambles.

In the proposed system, the s-mod and coding schemes can be changed according to the user requirements and wireless channel conditions. However, for multicast applications, it is very difficult, if not impossible, to track the real-time channel condition for all users. Practically, multicast receivers (the MSs) periodically report their channel conditions back to the sender (the BS). Thus, the sender needs to decide the best modulation and coding schemes based on the rough estimation of the channel conditions.

B. Optimal Configuration Problem Formulation

The main objective of the configuration problem is to maximize the utility of video multicast services and minimize the cost to provide the services. In other words, we aim to maximize the profit of such services (which is the utility minus the cost). Consider a multicast group with N users: $\mathbf{W} = \{w_1, w_2, \ldots, w_N\}$, and the average received SNR of the ith user w_i is γ_i. We assume $\gamma_i \in \mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is the range of the received SNR. Consider a two-layer video stream with data rates r_1 and r_2 for L1 (the base layer) and L2 (the enhancement

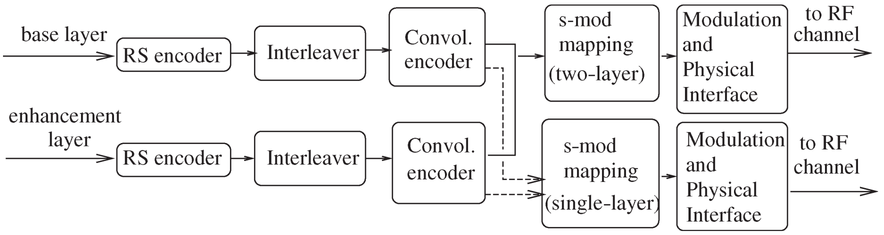


Fig. 6. System model at the multicast sender side.

layer). Let $m, m' \in \mathbf{M}$ be the s-mod schemes for the two-layer bits and the leftover single-layer bits and $c_1, c_2 \in \mathbf{C}$ be the coding schemes that are chosen for the L1 and L2 bits, respectively. Given the s-mod and coding schemes and the data rate of the video stream, we can estimate the bandwidth needed to deliver the two-layer video streams: $BW = f_1(m, m', c_1, c_2, r_1, r_2)$. BW is simple to calculate, and we omit the details here.

From Section III-E, for w_i, given γ_i and the channel fading model (e.g., Rayleigh fading), we can obtain the BERs of the two-layer bits and those of the leftover bits (from either L1 or L2), $e_1^i = f_2(m, c_1, \gamma_i)$, $e_2^i = f_3(m, c_2, \gamma_i)$, and $e_l^i(m', c', \gamma_i)$, respectively, where c' equals c_1 (or c_2), if the leftover bits are from L1 (or L2).

The utility describes the satisfaction level of the user, and it directly reflects the video QoS. For layered videos, different layers differently contribute to the QoS index. We can first map the PHY-layer performance index (BER of each layer) to the network performance index (packet loss rate of each layer). Then, we can map the network performance index to the utility. Denote f_4 as the mapping from the PHY performance index to the utility. Thus, the utility of w_i depends on its BER performance of the two layers: $U_i = f_4(e_1^i, e_2^i, e_l^i)$. A sample utility mapping of f_4 will be given in Section V.

The cost of the multicast service is proportional to the bandwidth needed; therefore, it is τBW, where τ is the unit cost of bandwidth, and it is determined by wireless service providers.

The optimization problem is then formulated as follows:

Problem 1 $(\mathbf{P1})$*:*

$$\max_{m,m',c_1,c_2} \left\{ \sum_{i=1}^{N} U_i - \tau BW \right\} \tag{8}$$

$$s.t.: \quad BW \le BW_{\max} \tag{9}$$

$$\min_{w_i \in \mathbf{W}} U_i \ge U_{\min}. \tag{10}$$

The objective is to find a modulation and coding configuration such that the profit of the video service can be maximized. The first constraint is that the total bandwidth allocated is bounded. The second constraint is that the user with the worst channel condition should obtain at least utility $U_{\min}$.

Since the unknown variables of the optimization problem $(\mathbf{P1})$ are all integers, it belongs to the traditional integer programming problems, which is NP-hard.

C. Algorithm for S-Mod Configuration

For the combinatorial optimization problem $(\mathbf{P1})$, there is no polynomial-time algorithm to obtain the optimal solution. To make it feasible for real-time optimal configuration, we divide the task into two parts to obtain the best solution. The first part is to create lookup tables offline, which is relatively time consuming. Given that we have a limited number of s-mod schemes, it is still feasible to use exhaustive search to compose the offline lookup table. The second part is to search the best solution using the lookup tables, which has low computational complexity and can be done in real time. Thus, we propose Algorithm 1 for creating the lookup table and Algorithm 2 for searching the optimal configuration.

Algorithm 1 Creating the BER table

```
find all single-layer s-mod schemes, and add them to $\mathbf{M}'$;
find all two-layer s-mod schemes using the procedure listed in Section III-C, and add them to $\mathbf{M}$;
for $m \in \mathbf{M}$ do
    find $m'_1 \in \mathbf{M}'$ with the BER closest to the L1 bits in $m$;
    find $m'_2 \in \mathbf{M}'$ with the BER closest to the L2 bits in $m$;
end for
create a BER table with columns of $\gamma$, $m$, $m'_1$, $m'_2$, $c_1$, $c_2$, $e_1$, $e_2$, $e_{l1}$, and $e_{l2}$;
for $\gamma \in \mathbf{\Gamma}$ do
    for $m \in \mathbf{M}$ do
        for $\{c_1, c_2\} \in \mathbf{C} \times \mathbf{C}$ do
            calculate $e_1$, $e_2$, $e_{l1}$, and $e_{l2}$;
            add the result to the BER table;
        end for
    end for
    sort the table w.r.t. $e_1$;
end for
```

Algorithm 2 Optimal transmission configuration

```
Require: $\vec{r} \vee \vec{\gamma}$
Ensure: optimal $m$, $m'$, $c_1$, $c_2$
for each row of the BER table do
    find which layer has leftover bits and the corresponding $m'$;
    calculate the BW cost for supporting $\vec{r}$;
    calculate the utility $U$;
    calculate the profit $U - \tau BW$;
    record the results in the utility table;
end for
$U\_MAX = 0$;
```

```
9: for each $m$, $c_1$, $c_2$ do
10:   $U\_TTL = 0$;
11:   for each item in $\gamma$ do
12:     search the utility table to find the utility;
13:     $U\_TTL += U$
14:   end for
15:   search the utility table to find $m'$ and the BW cost;
16:   if $(N\_TTL - \tau BW > U\_MAX)$ then
17:     if $(BW \le BW\_MAX)$ and $(NU \ge U\_MIN)$ then
18:       $U\_MAX = U\_TTL - \tau BW$;
19:       save $m$, $m'$, $c_1$, and $c_2$ as the optimal results;
20:     end if
21:   end if
22: end for
```

Algorithm 1 can be done offline, and the BER table created can be used for any two-layer video streams. For Algorithm 2, input variables $\vec{r} = [r_1 \ r_2]$ are the data rate of the layered video, and $\vec{\gamma}$ are the SNR of all users. The first seven lines of Algorithm 2 need to be executed for each video stream once it is admitted. Since the cardinalities of $\mathbf{M}$ and $\mathbf{C}$ are small, creating the utility table is manageable during the admission process. The optimal configuration needed to be executed in real time is lines 8–20 of Algorithm 2. Since this part involves looking up the sorted BER and utility tables once per configuration, the time needed to obtain the optimal solution is only $O(n \log(n))$, where n is the number of rows in the tables. As n is not a large number, the optimal configuration can be done in real time.

Claim 1 With more receivers of the multicast service, the optimal configuration tends to choose lower BER modulation and coding schemes.

Claim 2 With more low-channel-quality receivers of the multicast service, the optimal configuration tends to choose lower BER modulation and coding schemes.

The two aforementioned claims can be explained as follows: In general, the lower BER configuration results in higher utility per user and requires more bandwidth. For multicast services, once the transmission configuration is given, the bandwidth cost difference of two different schemes is fixed no matter how many receivers there are, but the difference of the utility increases w.r.t. the number of receivers, particularly when there are more low-channel-quality receivers. Thus, the number of receivers and their channel quality distribution have a great impact on the transmission configuration.

These two claims can help us to further accelerate the optimal configuration process: when the number of receivers increases or the channel qualities of receivers become worse, we can compare the current configuration with the schemes that lead to lower BERs and find the new optimal one, and vice versa.

Remarks: Similar to the existing AMC, the proposed s-mod configuration algorithm can also have the benefit if the real-time channel conditions are known and accurate. On the other hand, when the channel estimation is not accurate, the s-mod scheme can be robust against such estimation errors as the most important information can still be successfully received w.h.p. In addition, for multicast applications, given the different channel qualities of different users, real-time feedback of the channel information from all receivers will be more difficult. In this case, the s-mod can be even more desirable than the existing AMC solutions with the traditional QAM modem.

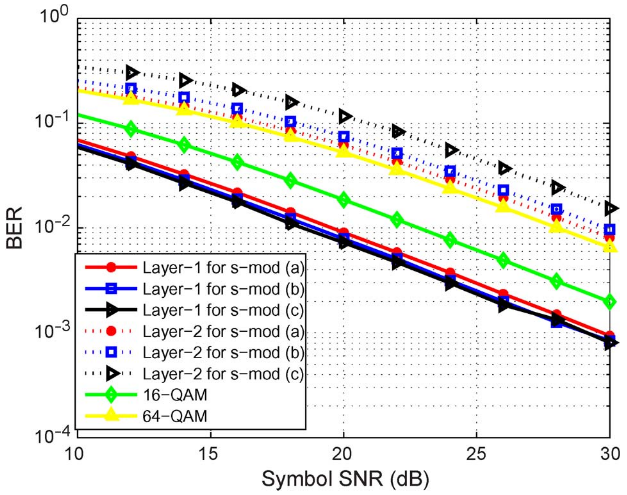


Fig. 7. BERs with different modulation schemes.

V. Performance Evaluation

A. S-Mod Performance Evaluation

We first evaluate the PHY-layer performance of the s-mod schemes. We simulate Rayleigh fast-fading channels with the received symbol SNR varying from 10 to 30 dB, and measure the BERs of different layers after demodulation, without using error coding. Fig. 7 shows the BERs of L1 and L2 bits for the s-mod schemes, 16-QAM and 64-QAM, respectively. The s-mod schemes (a), (b), and (c) are those shown in Fig. 4.

For 16- and 64-QAM, both layers have the same BERs. With the s-mod schemes, we can achieve a lower BER for L1 than that of 16-QAM, whereas the BER for L2 is higher than that of 64-QAM. In the succeeding sections, we will demonstrate the advantage of such service differentiation for layered video multicast. Among the three s-mod schemes, s-mod (c) has the lowest BER for L1, and it is most bandwidth efficient as it can transmit three L2 bits per symbol. From the figure, we also notice the tradeoff of the L1 and L2 BER performance. This is anticipated as the lower BER of L1 bits is achieved at the cost of a higher BER of L2 bits.

B. Utility of Video Services

Peak SNR (PSNR) reflects the distortion of the video and is frequently used as a video quality metric. Denote as F the total number of frames in a video sequence, M the number of pixels in a frame, and X_{ij} and Y_{ij} the values of the jth pixel in the ith original and reconstructed frames, respectively. PSNR is calculated by

$$\text{PSNR} = 10 \times \log_{10}(M \times 255^2/\overline{ssd}) \tag{11}$$

where $\overline{ssd} = (1/F)\sum_{i=1}^{F}\sum_{j=1}^{M}(X_{ij} - Y_{ij})^2$.

With layered video coding, the losses of L2 packets result in a higher ssd and degraded PSNR, and the losses of L1 packets result in not only PSNR degradation but also sometimes undecodable frames. To measure the user's satisfaction level, the utility function is defined as the product of PSNR (of the decoded frames) and the decodable rate. The decodable rate is the ratio of the number of successfully decoded frames over the total number of frames.

Two videos "Foreman" and "Soccer" with different video characteristics are selected. For "Foreman," receivers are more concerned about the detailed facial texture. For "Soccer," receivers are more sensitive to the smoothness of the moving players. These two complementary videos can represent a wide variety of videos and thus give a comprehensive evaluation on scalable modulation schemes. The resolution of both videos is 352 × 288. The video frame rate is 30 frame/s, and the total length of each video is 20 s.

The Joint Scalable Video Model (JSVM) is selected as the video codec. To simplify the simulation without loss of generality, the encoder is configured with two-layer CGS. The base-layer quality (in terms of PSNR) is around 30 dB, and the quality gap of the base layer with or without the enhancement layer for both videos is about 4 dB. For "Foreman," the bit rates of the base layer and the enhancement layer are 148.16 and 522.84 kb/s, respectively. For "Soccer," the bit rates are 182.76 and 811.39 kb/s, respectively. Packetization is performed according to the standard, i.e., using the Real-time Transport Protocol payload format for H.264/SVC SVC. If there are bit errors in any network abstraction layer (NAL) unit, the corrupted NAL unit is discarded. We rely on the JSVM codec to recover from the lost NAL unit. According to the publicly available open-source SVC codec,[1] other error concealment strategies have not been fully adopted yet, so we do not consider them in this paper.

Fig. 8 shows the utility function of "Foreman." Each solid line represents the utility with a fixed e_1, and the horizontal axis is e_2 for the s-mod schemes. When e_2 is larger than 5×10^{-4}, the utility function does not change since almost all L2 frames are lost due to the high BER of L2. The utility function of single-layer modulation (16-QAM or 64-QAM) is slightly different from that for the s-mod schemes since the bit errors of L1 and L2 are equal and less correlated, which is the dotted-line "single layer" in Fig. 8. As shown in Fig. 8, the correlation of bit errors in the two layers has negligible impact. Similar tendency is observed for the utility function of "Soccer," which is not presented due to the page limit.

For each video stream, the control or meta data information will be sent to the user before the video data frames, which can include the utility function of the video, so the BS can use it to select the right s-mod scheme in real time.

C. Video Performance of a Single Receiver

We first investigate the utility of a single user with a received SNR of 12 dB. First, we fix the modulation schemes 16-QAM, 64-QAM, or s-mod scheme (b), and use the optimal configuration algorithm to find the best RS and convolutional codes for it.

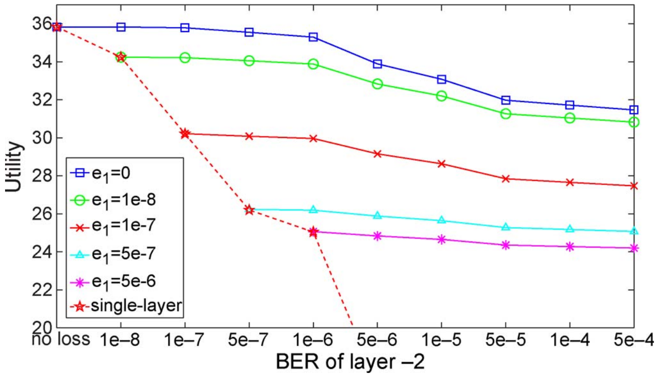
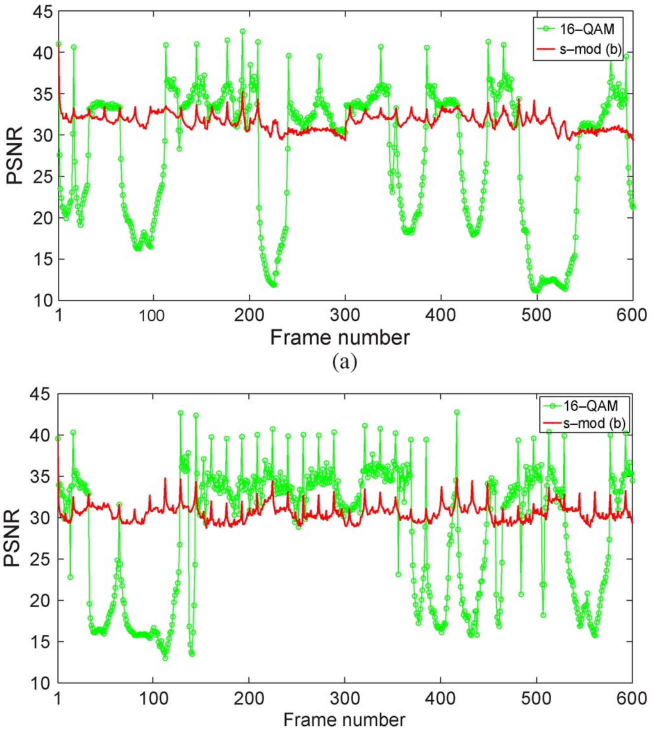


Fig. 8. Utility function of "Foreman."



(a)

(b)

Fig. 9. PSNR of 600 frames with 16-QAM and s-mod (b). (a) "Foreman." (b) "Soccer."

We use RS(255, 191) and convolutional code with rate 1/2 for 16-QAM and the L2 bits of the s-mod scheme, and use RS(255, 205) for the L1 bits of the s-mod scheme. With 16-QAM, e_1 and e_2 are on the order of 10^{-6}. With a nonnegligible loss rate of the base layer, the decodable rate is only 0.9, and the PSNR is very low when the base-layer frames are lost. With 64-QAM, as the BERs are too high, nothing can be decoded. With the s-mod scheme, the base-layer frames are all successfully received, so the decodable rate is 100%, and the PSNR is stable, although e_2 with s-mod is higher than that with 16-QAM.

Fig. 9 compares the PSNR with 16-QAM and the s-mod scheme (b) for the two videos. The performance of s-mod schemes (a) and (c) are close to that with scheme (b), and the results are omitted. The random screen shots of the videos are given in Fig. 10(a) and (b). As shown in the figures, with 16-QAM, the PSNR variation is significant, and the video quality may frequently degrade, which results in an annoying

---

[1] The open-source SVC codec can be downloaded from http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm.

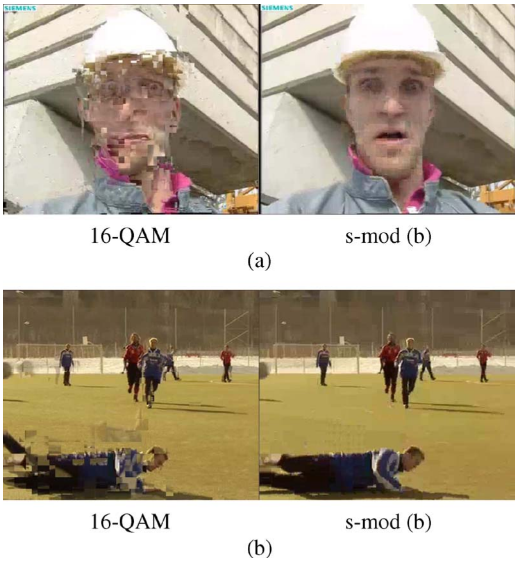


Fig. 10. Screenshot comparison for 16-QAM and s-mod (b). (a) "Foreman." (b) "Soccer."

experience for video users. The video quality is quite stable and acceptable with the proposed s-mod schemes. The simulation results demonstrate the importance of providing differentiated services for layered video streams and the advantage of the proposed s-mod schemes.

D. Video Utility of Multicast Receivers

Next, we investigate the total profits with a group of multicast receivers. We consider different scenarios: groups of 15 or 5 users with received SNR uniformly distributed in [10, 22] dB (`uni15` or `uni5`), and groups of five users with bad, good, or excellent channels, respectively (`bad5`, `good5`, or `exce5`). The `bad5`, `good5`, and `exce5` groups have their SNR distributed in the range of [10, 14], [14, 22], and [20, 22] dB, respectively. In Table I, we list the utility (U), bandwidth requirement $(BW$ in terms of kilo-symbol/sec), and profit (P) of the different modulation schemes with video "Foreman."

For the s-mod schemes, the leftover bits of the enhancement layer are modulated using 64-QAM, so the total bandwidth required using s-mod combined with 64-QAM is smaller than that using 16-QAM only and is also smaller than that using 16-QAM for L1 bits and 64-QAM for L2 bits. τ is set to 0.1 and 0.01 (dB $\cdot$ sec/kilosymbol), respectively. For each modulation scheme, the error codes leading to the highest profit are chosen.

As shown in Table I, for different scenarios, we should choose different modulation schemes to maximize the profit. In specific, when the users' channel quality is uniformly distributed or bad, we prefer the proposed s-mod schemes, which can substantially outperform the 16-QAM or 64-QAM. Comparing the `uni5` and `uni15` scenarios, although the channel quality distributions are the same, the optimal modulation scheme is different. This is because s-mod (a) needs to be coupled with strong error coding schemes and its bandwidth requirement is higher than that of s-mod (b), and the utility of s-mod (a) can be better than that of s-mod (b) when the SNR is about 18 dB. Below 18 dB, L2 frames cannot be successfully decoded with both s-mod (a) and (b); above 20 dB, L2 can be decoded with both schemes. Therefore, only when the number of users is higher and τ is not large (i.e., the bandwidth is not too expensive) is s-mod (a) (which is also an SPC scheme) preferable. This observation echoes our Claim 1 in the previous section.

TABLE I
UTILITY COMPARISON

cases		s-mod (a)	s-mod (b)	s-mod (c)	16-QAM	64-QAM
`uni5`	U	166	166	161.6	124.8	71.7
	BW	334.2	320.3	296.1	447.9	255.8
	$P(\tau : 0.01)$	162.7	**162.8**	158.7	120.3	69.1
	$P(\tau : 0.1)$	132.6	**134**	132	80	46.1
	$P(\tau : 0.2)$	99.2	102	**102.4**	35.2	20.5
`uni15`	U	492.6	489.3	480.5	428.7	215
	BW	343.78	320.3	296.1	447.9	298.6
	$P(\tau : 0.01)$	**489.1**	486.1	477.6	424.2	212
	$P(\tau : 0.1)$	**458.2**	457.3	450.9	383.9	185.1
	$P(\tau : 0.2)$	423.8	**425.2**	421.3	339.1	155.3
`bad5`	U	157.3	157.3	157.3	89	0
	BW	331.2	309.7	280.2	447.9	224.8
	$P(\tau : 0.01)$	153.9	154.2	**154.5**	84.5	-2
	$P(\tau : 0.1)$	124.1	126.3	**129.2**	44.2	-25
	$P(\tau : 0.2)$	91	95.3	**101.2**	-1	-49
`good5`	U	167.7	166	161.6	179.2	107.5
	BW	308.5	308.5	286.4	417.3	298.6
	$P(\tau : 0.01)$	164.6	162.9	158.8	**175**	104.5
	$P(\tau : 0.1)$	136.8	135.2	133	**137.4**	77.6
	$P(\tau : 0.2)$	**106**	104.3	104.3	95.7	47.8
`exce5`	U	179.2	179.2	170.4	179.2	179.2
	BW	301.1	308.5	286.4	367.2	255.8
	$P(\tau : 0.01)$	176.1	176.1	167.5	175.5	**176.6**
	$P(\tau : 0.1)$	149.0	148.3	141.8	142.4	**153.6**
	$P(\tau : 0.2)$	118.9	117.5	113.1	105.7	**128**

The results in Table I also verify our Claim 2: when all users have bad channel quality (`bad5`), the optimal solution is to use s-mod (c), which has the lowest BER for L1.

The optimal configuration will choose 16-QAM if all users have good channel quality and τ is not large and choose 64-QAM only if all users have a received SNR above 20 dB. In the latter, the BER of 64-QAM can ensure that both L1 and L2 video frames are successfully decoded, and since 64-QAM has the highest bandwidth efficiency, it will be the best candidate. Nevertheless, in these cases, the performance of the s-mod schemes is still very close to the optimal.

In summary, the s-mod schemes provide many more choices in the PHY layer to effectively support scalable video multicast. According to the bandwidth price τ, the optimal configurations can achieve much higher profit for wireless service providers and ensure user-perceived video quality.

Remarks: Note that, when the SNR is less than 14 dB, the performance of QPSK (or 4-QAM) for `bad5` can be probably better than that with 16-QAM. The reason that we did not include 4-QAM in the comparison is because the data rate of 4-QAM is much lower than the s-mod we used in this paper. When the target SNR range is lower, we can design other s-mod schemes with L1 performance close to 4-QAM and L2 performance worse than 16-QAM, which is anticipated to achieve better utility than 4-QAM; when the target SNR range is even higher than 22 dB, we can design s-mod schemes with L1 performance better than 64-QAM, etc. As the proposed s-mod is a generalization of traditional QAM modulation, with more choices using s-mod, the performance will be at least as good as that using traditional QAM. Due to the limited space, we did not include all these comparisons and designs in this paper.

VI. CONCLUSION

In this paper, we have proposed scalable modulation schemes that can provide differentiated services in the PHY layer using the current mainstream QAM modulation and demodulation hardware. We have further formulated a cross-layer optimization problem, aiming to maximize the profit of video services by selecting PHY-layer modulation and coding schemes. Extensive simulation results have demonstrated that, with the flexibility of s-mod, we can have better PHY-layer configurations to achieve substantial profit gains for wireless videocast. In this paper, if there is any bit error in a NAL unit, the NAL unit is discarded. The impact of other advanced error concealment strategies on the system performance with s-mod will be an important further research issue.

REFERENCES

[1] L. Cai, Y. Luo, S. Xiang, and J. Pan, “Scalable modulation for scalable wireless videocast,” in *Proc. IEEE Infocom*, San Diego, CA, Mar. 2010.

[2] Philips Semiconductors, 3GPP2, C50-200010611-013R1 1XTREME Physical Specification for Integrated Data and Voice Services in cdma-2000 Spread Spectrum Systems, Jun. 2001.

[3] IEEE standard for local and metropolitan area networks, part 16: Air interface for fixed broadband wireless access systems, IEEE Std. 802.16-2004.

[4] T. Cover, “Broadcast channels,” *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[5] S. Wang and B. K. Yi, “Optimizing enhanced hierarchical modulations,” in *Proc. IEEE GLOBECOM*, Nov./Dec. 2008.

[6] J. Long, H. Jin, D. Liang, X. Zhang, M. Peng, and W. Wang, “Comparison of several demodulation algorithms about hierarchical modulation in broadcast communications,” in *Proc. IEEE WiCOM*, Sep. 2010, pp. 1–4.

[7] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H. 264/AVC standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[8] T. Wiegand, G. Sllivan, J. Reichel, H. Schwarz, and M. Wien, *Joint Draft ITUT Rec. H. 264–ISO/IEC 14496-10/Amd. 3 Scalable Video Coding*. Geneva, Switzerland: Joint Video Team, 2007.

[9] C. Hellge, S. Mirta, T. Schierl, and T. Wiegand, “Mobile TV with SVC and hierarchical modulation for DVB-H broadcast services,” in *Proc. IEEE BMSB*, May 2009, pp. 1–5.

[10] D. Mitran, “Stochastic theory of a fluid model of producers and consumers coupled by a buffer,” *Adv. Appl. Probab.*, vol. 20, no. 3, pp. 646–676, Sep. 1988.

[11] E. Shihab, L. Cai, F. Wan, A. Gulliver, and N. Tin, “Wireless mesh networks for in-home IPTV distribution,” *IEEE Netw.*, vol. 22, no. 1, pp. 52–57, Jan./Feb. 2008.

[12] M. Kobayashi, H. Nakayama, N. Ansari, and N. Kato, “Robust and efficient stream delivery for application layer multicasting in heterogeneous networks,” *IEEE Trans. Multimedia*, vol. 11, no. 1, pp. 166–176, Jan. 2009.

[13] L. X. Cai, L. Cai, X. Shen, and J. W. Mark, “Resource management and QoS provisioning for IPTV over mmWave-based WPANs with directional antenna,” *ACM/Springer MONET*, vol. 14, no. 2, pp. 210–219, Apr. 2009.

[14] R. Zhang, R. Ruby, J. Pan, L. Cai, and X. Shen, “A hybrid reservation/contention-based MAC for video streaming over wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 28, no. 3, pp. 389–398, Apr. 2010.

[15] M. Kobayashi, H. Nakayama, N. Ansari, and N. Kato, “Reliable application layer multicast over combined wired and wireless networks,” *IEEE Trans. Multimedia*, vol. 11, no. 8, pp. 1466–1477, Dec. 2009.

[16] D. Wu and R. Negi, “Effective capacity: A wireless link model for support of quality of service,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, Jul. 2003.

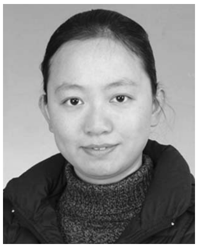
[17] Q. Du and X. Zhang, “Statistical QoS provisionings for wireless unicast/multicast of layered video streams,” in *Proc. IEEE Infocom*, Apr. 2009, pp. 477–485.

[18] R. Cohen, G. Grebla, and L. Katzir, “Cross-layer hybrid FEC/ARQ reliable multicast with adaptive modulation and coding in broadband wireless networks,” in *Proc. IEEE Infocom*, Apr. 2009, pp. 1917–1925.

[19] J. Kim, J. Cho, and H. Shin, “Resource allocation for scalable video broadcast in wireless cellular networks,” in *Proc. IEEE WiMob*, Aug. 2005, vol. 2, pp. 174–180.

[20] J. Xu, X. Shen, J. W. Mark, and J. Cai, “Adaptive transmission of multilayered video over wireless fading channels,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2305–2314, Jun. 2007.

[21] J. She, X. Yu, P. Ho, and E. Yang, “A cross-layer design framework for robust IPTV services over IEEE 802.16 networks,” *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 235–245, Feb. 2009.



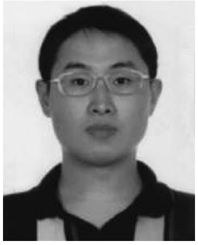
Lin Cai (S'00–M'06–SM'10) received the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2002 and 2005, respectively.

Since 2005, she has been an Assistant Professor and then an Associate Professor with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada. She has been an Associate Editor for *EURASIP Journal on Wireless Communications and Networking*, the *International Journal of Sensor Networks*, and the *Journal of Communications and Networks*. Her research interests include wireless communications and networking, with a focus on network protocol and architecture design supporting emerging multimedia traffic over wireless, mobile, ad hoc, and sensor networks.

Siyuan Xiang (S'10) received the M.Eng. degree from Tongji University, Shanghai, China, in 2008. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada.

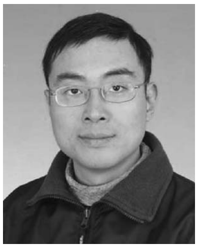
His research interest is multimedia communications.

Mr. Xiang is the recipient of a China Scholarship Council Scholarship.



Yuanqian Luo (S'09) received the B.S. and M.S. degrees in electrical engineering from Southeast University, Nanjing, China, in 2005 and 2008, respectively. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada.

His current research interests include cross-layer design and optimization for cooperative networks.



Jianping Pan (S'96–M'98–SM'08) received the B.S. and Ph.D. degrees in computer science from Southeast University, Nanjing, China.

He is currently an Associate Professor of computer science with the Department of Computer Science, University of Victoria, Victoria, BC, Canada. He did his postdoctoral research with the University of Waterloo, Waterloo, ON, Canada. He was also with Fujitsu Laboratories and NTT Laboratories. His research interests include computer networks and distributed systems. His current research interests include protocols for advanced networking, performance analysis of networked systems, and applied network security.