Hindawi Publishing Corporation
Advances in Multimedia
Volume 2008, Article ID 164027, 27 pages
doi:10.1155/2008/164027

Research Article

Traffic and Quality Characterization of the H.264/AVC Scalable Video Coding Extension

Geert Van der Auwera,[1] Prasanth T. David,[2] Martin Reisslein,[2] and Lina J. Karam[2]

[1] *Samsung Information Systems America, Digital Media Solutions Lab, 3345 Michelson Drive, Suite 250, Irvine, CA 92612, USA*
[2] *Department of Electrical Engineering, Arizona State University, Goldwater Center MC 5706, AZ 85287-5706, USA*

Correspondence should be addressed to Martin Reisslein, reisslein@asu.edu

Received 9 June 2008; Accepted 14 August 2008

Recommended by Shiwen Mao

The recent scalable video coding (SVC) extension to the H.264/AVC video coding standard has unprecedented compression efficiency while supporting a wide range of scalability modes, including temporal, spatial, and quality (SNR) scalability, as well as combined spatiotemporal SNR scalability. The traffic characteristics, especially the bit rate variabilities, of the individual layer streams critically affect their network transport. We study the SVC traffic statistics, including the bit rate distortion and bit rate variability distortion, with long CIF resolution video sequences and compare them with the corresponding MPEG-4 Part 2 traffic statistics. We consider (i) temporal scalability with three temporal layers, (ii) spatial scalability with a QCIF base layer and a CIF enhancement layer, as well as (iii) quality scalability modes FGS and MGS. We find that the significant improvement in RD efficiency of SVC is accompanied by substantially higher traffic variabilities as compared to the equivalent MPEG-4 Part 2 streams. We find that separately analyzing the traffic of temporal-scalability only encodings gives reasonable estimates of the traffic statistics of the temporal layers embedded in combined spatiotemporal encodings and in the base layer of combined FGS-temporal encodings. Overall, we find that SVC achieves significantly higher compression ratios than MPEG-4 Part 2, but produces unprecedented levels of traffic variability, thus presenting new challenges for the network transport of scalable video.

1. INTRODUCTION

We study the video traffic generated by the scalable video coding (SVC) extension [1, 2] of the H.264/MPEG-4 advanced video coding standard [3] (H.264 SVC for brevity). This extension is expected to have a broad application domain for heterogeneous wired and wireless video transmission to various terminals. Indications of the growing acceptance of H.264/AVC are its adoption in application standards and industry consortia specifications, such as DVB, ATSC, 3GPP, 3GPP2, MediaFLO, DMB, DVD Forum (HD-DVD), and Blu-Ray Disc Association (BD-ROM). At the same time, mobile TV technologies are made widely available. IPTV, mobile TV, satellite TV, and video surveillance are considered key applications that can make H.264/AVC and its SVC extension the dominant video encoder in the professional and consumer markets.

In order to examine the fundamental traffic characteristics of H.264 SVC's scalability modes, we focus on encodings with fixed quantization scales, that is, with variable bit rate (VBR). An additional motivation for the focus on VBR video is that the VBR streams allow for statistical multiplexing gains that have the potential to improve the efficiency of video transport over communication networks [4–9]. The development of video network transport mechanisms that meet the strict playout deadlines of the video frames and efficiently accommodate the variability of the video traffic is a challenging problem. Based primarily on the characteristics of MPEG-4 Part 2 single-layer and scalable video, transport mechanisms have been developed for a wide range of network transport scenarios, including video transport over the Internet (see, e.g., [10–16]) over wireless networks (see, e.g., [17–24]) over peer-to-peer networks (see, e.g., [25–32]) and over sensor networks [33–35]. The widespread adoption of the new H.264/AVC video standard necessitates the careful study of the traffic characteristics of video coded with the new H.264/AVC codec and its extensions. Recent traffic studies [36] indicate that despite the lower average bit rate of H.264/AVC and H.264 SVC single-layer video, elementary bufferless multiplexing of a small number of video streams can be more efficient with MPEG-4 Part 2 encoding than with H.264/AVC or H.264 SVC encoding due to the significantly higher traffic variability of H.264/AVC and H.264 SVC. Therefore, it is necessary to thoroughly examine the new SVC extension's statistical traffic characteristics from a communication network perspective.

The traffic characterizations and network transport mechanisms for scalable video encoded with MPEG-4 Part 2 and older codecs have received significant attention in the literature (see, e.g., [37–54]). The traffic characterization of H.264/AVC and H.264 SVC *nonscalable* (single-layer) traffic is studied in [36, 55, 56]. The study of network transport mechanisms in the context of H.264/AVC (see, e.g., [57–59]) and H.264 SVC (see, e.g., [60–64]) has begun to attract interest. To the best of our knowledge, the traffic of H.264 SVC-encoded *scalable* video is for the first time examined in the present study. Existing studies of the H.264/AVC codec and its SVC extension, such as [3, 65, 66], focus primarily on the bit rate-distortion (RD) performance, that is, the video quality (PSNR) as a function of the *average* bit rate, and typically consider only short video sequences up to a few hundred frames. In contrast, for the transport over communication networks, the traffic variability is also a key concern [5, 9, 14]. Therefore, we examine in the present study the *joint characterization* of bit rate-distortion and higher order bit rate statistics, such as the *variability of the bit rate*, as a function of the distortion. We perform a detailed analysis of elementary statistics of the scalable video traffic. We study statistics of frame sizes, group of picture (GoP) sizes, as well as frame and GoP qualities. We use bit rate-distortion (RD) and bit rate variability-distortion (VD) curves to compare the H.264 SVC-layered traffic to the equivalent traffic of MPEG-4 Part 2 [67], which is the predecessor of H.264/AVC and which supports temporal, spatial, and FGS scalability. In order to obtain reliable and meaningful statistical estimates of the traffic variability and other properties, it is necessary to examine *long* video sequences with several thousand frames, as we do in this study.

All encodings of this study are publicly available as video traces at http://trace.eas.asu.edu/. Video traces [47] are files mainly containing video frame time stamps, frame types (e.g., I, P, or B), encoded frame sizes (in bits), and frame qualities (PSNR). Video traces are employed in simulation studies of transport of scalable video over communication networks (see, e.g., [37–41, 44, 46, 52–54]). Key advantages of simulating with video traces over experiments with actual video are that only very basic knowledge of video encoding is required for simulations with video traces and that video traces are freely available without copyright protection. Also, network simulations with video traces can be conducted with standard network simulation programs and integrated in network simulation modules (see, e.g., [68]), whereas experiments with actual video require in-depth video coding expertise and large computational resources for the encoding of many long video sequences.

The paper is organized as follows. We provide a brief overview of the scalability modes of the H.264 SVC extension in Section 2. In Section 3, we describe the video test sequences, encoding tools, and video traffic metrics employed in our study. In Section 4, we analyze the traffic characteristics of the individual temporal scalability layers of long CIF videos. In Section 5, we study spatial scalability mode traffic with the same long CIF sequences and their QCIF subsampled versions. In Sections 6 and 7, we examine SVC's fine granularity scalability (FGS) traffic and medium granularity scalability (MGS) traffic, respectively. In Section 8, we consider the combined spatiotemporal and FGS-temporal scalabilities, which permit us to examine the separability of the combined scalability modes into the basic modes from a video traffic analysis perspective. We summarize our conclusions in Section 9.

2. OVERVIEW OF H.264 SCALABLE VIDEO CODING (SVC)

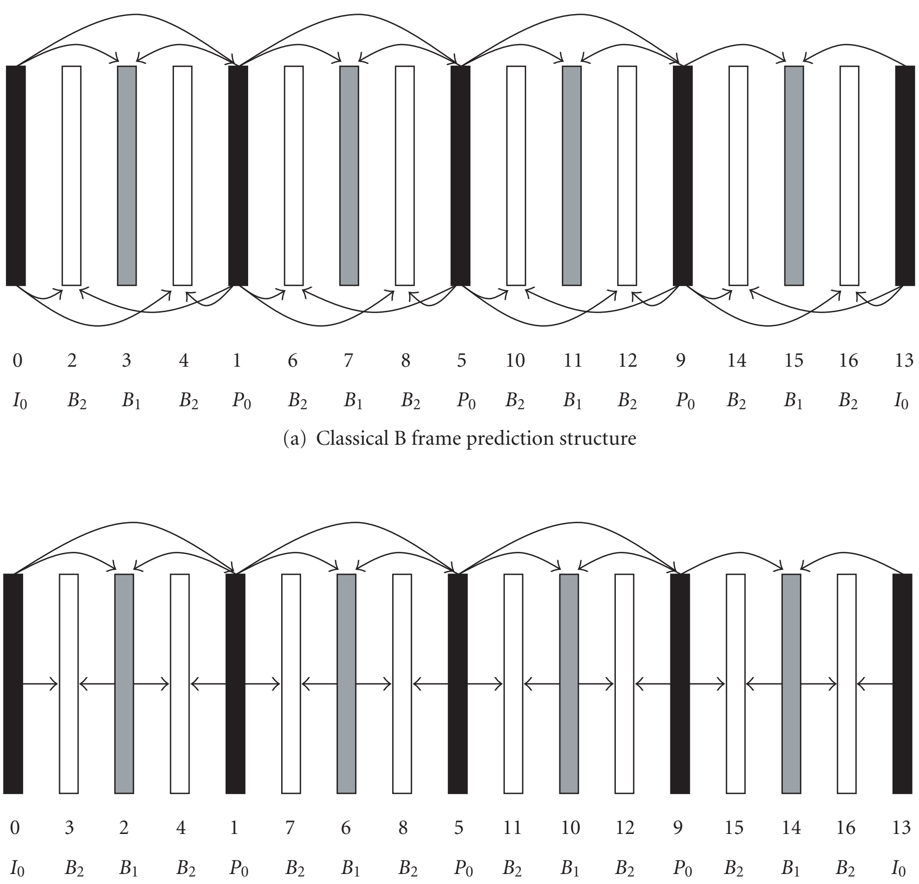
In this section, we briefly introduce the scalable video coding (SVC) extension of H.264/AVC. For a detailed discussion of the video technologies in the MPEG-4 family, such as MPEG-4 Part 2 [67] and H.264/AVC [3], we refer to [69]. At the end of 2007, the SVC scalability extension was added to the H.264/AVC standard. The SVC extension provides temporal scalability, spatial scalability, coarse (CGS) and medium (MGS) granularity scalability, as well as combined spatiotemporal SNR scalability (restricted set of spatiotemporal-SNR points can be extracted from a global scalable bit stream). The fine granularity scalability (FGS) mode was initially intended to be part of the SVC extension, however, FGS was not included in the initial version of SVC. Presently, investigations are ongoing to include FGS in a followup of the SVC extension.

While earlier scalable video encoders and receivers, such as MPEG-4 Part 2, did not gain wide market deployment, the H.264 SVC scalability extension is expected to play a major role in providing video services over heterogeneous networks due to the significantly improved rate-distortion efficiency of the H.264 SVC scalability encoding tools (with respect to MPEG-4 Part 2) and the growing industrial acceptance of H.264/AVC as the successor of the pervasive MPEG-2 standard.

In the following subsections, we briefly discuss the main scalability modes of this new H.264 SVC scalability amendment and refer to [2] for detailed information.

2.1. Temporal scalability with hierarchical B frames

The introduction of hierarchical B frames has allowed the H.264 SVC encoder to achieve temporal scalability while at the same time improving RD efficiency as compared to the classical B frame prediction method, employed by the older MPEG standards (MPEG-1/2/4 Part 2) and used by default in H.264/AVC. Figure 1(a) depicts the classical B frame prediction structure, where each B frame is predicted only from the preceding I or P frame and from the subsequent I or P frame. Figure 1(b) depicts the hierarchical B frame structure [70] which uses B frames to predict B frames. The illustrated case is the dyadic hierarchy of B frames, meaning


0 2 3 4 1 6 7 8 5 10 11 12 9 14 15 16 13
I_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 I_0
0 3 2 4 1 7 6 8 5 11 10 12 9 15 14 16 13
I_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 P_0 B_2 B_1 B_2 I_0

(a) Classical B frame prediction structure

(b) Hierarchical B frame prediction structure

Figure 1: B frame prediction structures.

that the number n of B frames in between the key pictures (I or P frames) must equal $n = 2^k - 1$. (We do not consider low-delay or constrained delay B frame prediction structures, for which we refer to [2].)

We depict the hierarchy with 3 B frames (I frame period is 16) in Figure 1(b). Temporal layer 0 consists of I and P key pictures, which are used to predict the B frames of temporal layer 1 (the temporal layer is indicated by the subscript of the I, P, and B symbols). The B frames of temporal layer 1 together with the key pictures predict the B frames of the second temporal layer. This halving of the prediction distance between frames in each prediction step is called *dyadic* hierarchy, with each splitting step resulting in one temporal layer, that is, the hierarchy with 15 B frames supports 5 temporal layers.

Underneath Figures 1(a) and 1(b), we provide for each frame the preferred encoding order with the smallest decoding delay. We observe that the encoding orders are identical for temporal layer 0, since the prediction dependencies of the key pictures are identical in both cases. With hierarchical B frames, the middle B frame is predicted first, while in the classical approach, the first B frame is predicted first.

The coding efficiency of hierarchical B frames depends on the choice of the quantization parameters for each B frame. H.264 SVC introduces cascading quantization scales which assign a higher quantization parameter value (lower quality) to B frames belonging to higher temporal layers.

2.2. Spatial scalability

A spatial scalable bit stream implies that streams with different frame resolutions, such as QCIF (176 × 144 pixels), CIF (352 × 288), and 4CIF (704 × 576), are extractable from a single bit stream. In this example, the QCIF layer would be the spatial base layer, and the CIF and 4CIF layers the spatial enhancement layers. An important new property of H.264 SVC is that a spatial layer is decodable with a single motion-compensation loop.

Besides the encoding mechanisms that we described in Section 2.1, the tools that exploit the interlayer redundancies between spatial layers are *interlayer motion prediction*, *interlayer residual prediction*, and *interlayer intra prediction* [2]. Figure 2 depicts the intra- and interlayer prediction dependencies for two spatial layers (base and enhancement), illustrating that the interlayer prediction mechanisms operate in a bottom-up fashion, that is, the base layer is used for the prediction of the spatial enhancement layer.

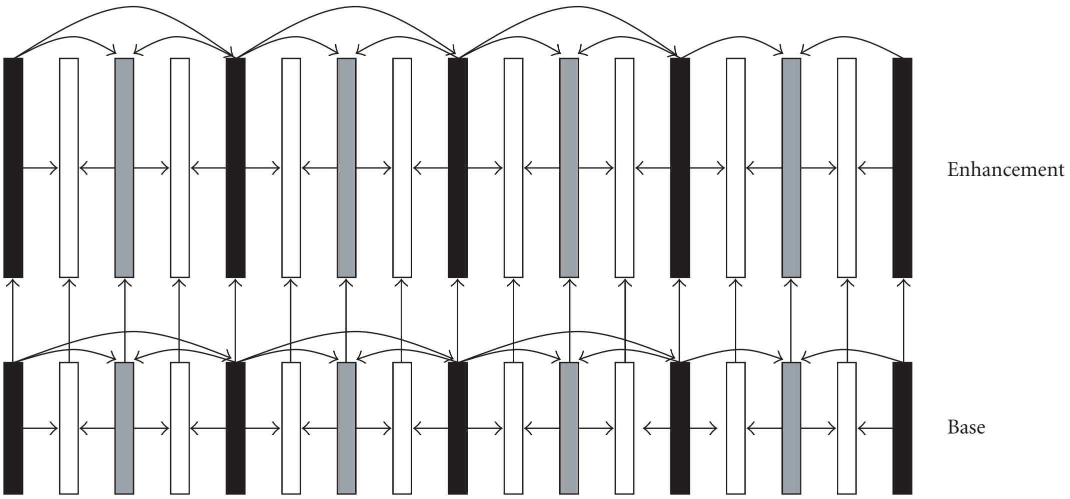


Figure 2: Two-layer spatial scalability intra- and interlayer prediction dependencies.

2.3. SNR scalability, including fine and medium granularity scalability

With SNR (quality) scalability, the quality of the video frames is improved for a given spatial resolution and frame rate. The main quality scalability modes, although not all are part of the SVC amendment, are coarse granularity scalability (CGS), medium granularity scalability (MGS), and fine granularity scalability (FGS). In our traffic study, we focus on MGS (included in first SVC) and FGS (not included in first SVC), which we now briefly review.

H.264 FGS supports single-loop decoding. The I/P key pictures of the quality base layer are predicted from one another as in Figure 1(b), but the B frames can be predicted using all quality refinements available in the higher quality layers, as illustrated in Figure 3(a). This prediction using the quality refinements of the enhancement layer improves the coding efficiency since the highest quality representation is used for prediction, but results in a decoding drift error, that is only stopped at the next I/P key picture [71]. Alternatively, the quality base layer prediction structure can be based on the hierarchical B frames of the quality base layer only, with identical dependencies in the quality refinement layer, as illustrated in Figure 3(b). This prediction structure is also known as closed-loop motion compensated prediction at low and high bit rates, and we consider this structure in our traffic study.

In MPEG-4 Part 2 FGS, closed-loop motion compensation is adopted only for the quality base layer while for the quality enhancement layer, a bit-plane technique is used to code the difference between the original picture and the picture reconstructed from the quality base layer, as illustrated in Figure 3(c). However, not exploiting the temporal redundancies between the adjacent pictures in the FGS enhancement layer incurs a considerable loss in coding efficiency, which schemes, such as PFGS [72], tried to alleviate.

In H.264 FGS, hierarchical B frames are used to efficiently exploit the temporal redundancy among adjacent pictures in the FGS enhancement layer. Using a different coding technique (requantization of quantization error) instead of bit-plane coding in MPEG-4 Part 2 FGS, H.264 FGS codes the enhancement layer information in progressive refinement (PR) slices that can be truncated with byte granularity. Furthermore, motion refinement is allowed in the FGS enhancement layer, as detailed in [1].

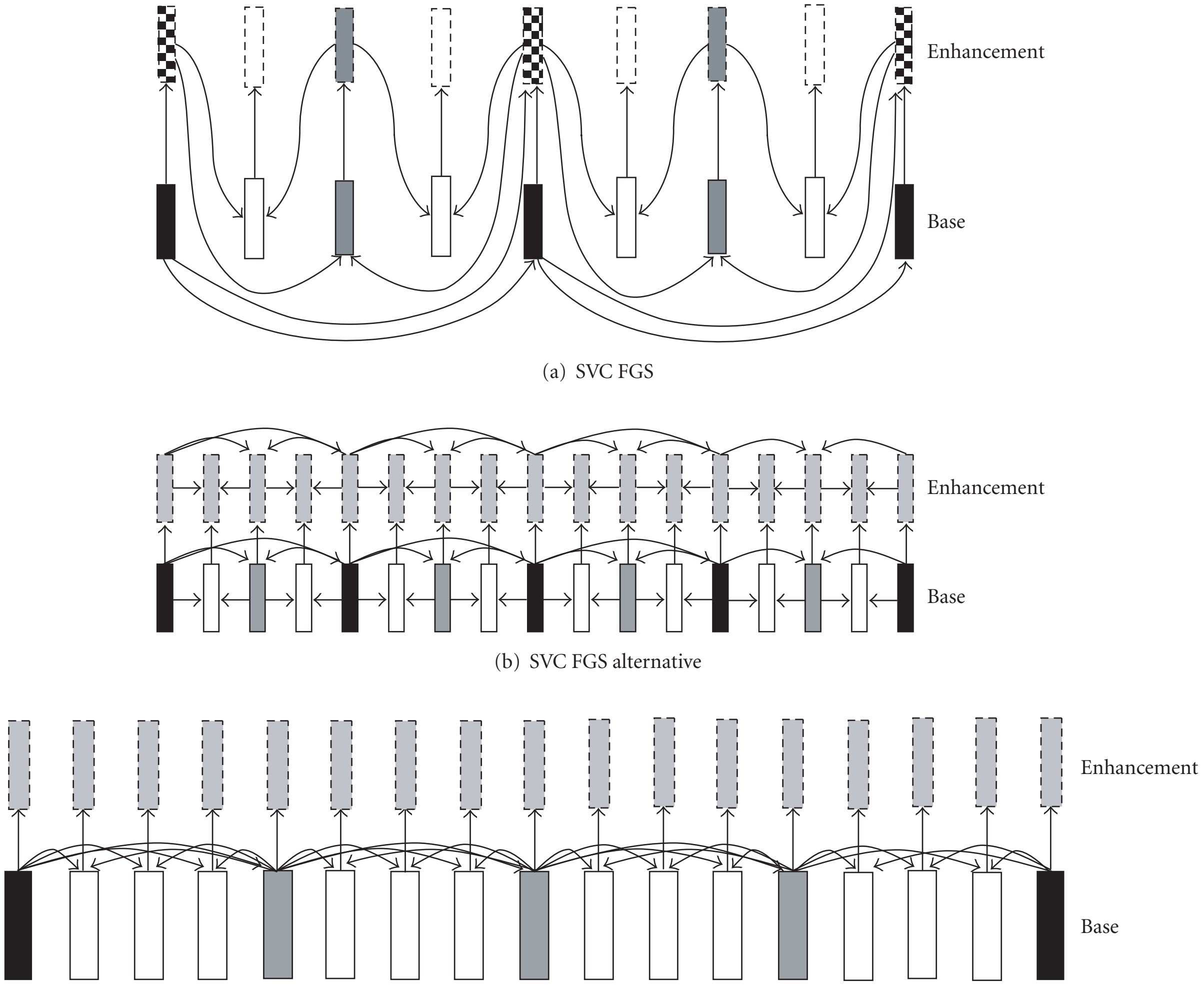
SVC MGS similarly encodes additional quality layers that each consist of disposable quantities that are coarser than the byte truncation offered by FGS. One MGS quality enhancement layer, for example, increases the base layer quality corresponding to quantization parameter QP to the quality of an encoding with parameter QP − 6. The information in each MGS enhancement layer can additionally be represented with a maximum granularity of 1/16 or equivalently up to 16 refinements included in the enhancement layer. This medium granularity enables network mechanisms to drop MGS enhancement packets in a simplified manner compared to FGS, which requires truncation.

2.4. Combined scalability

H.264 SVC supports spatiotemporal-SNR scalability, also referred to as *combined* scalability. This means that one global bit stream supports spatial, temporal, and SNR scalability. Depending on the encoding configuration, several individual bit streams with different spatial resolutions, frame rates, and SNR enhancement layers are extractable from the global bit stream. The SNR enhancement can be provided by CGS, MGS, or FGS. Note that not all scalability modes are necessarily supported by a combined scalable bit stream.

3. STUDY SETUP: VIDEO SEQUENCES, ENCODING TOOLS, AND VIDEO TRAFFIC METRICS

In this section, we introduce the setup used for obtaining the video traffic and quality characterizations presented in the subsequent sections.



(a) SVC FGS

(b) SVC FGS alternative

(c) MPEG-4 Part 2 FGS

Figure 3: Fine granularity scalability (FGS) prediction structures.

3.1. Video sequences

The Common Intermediate Format (CIF, 352 × 288 pixels) video sequences used for the statistics presented in this study are the ten-minute *Sony Digital Video Camera Recorder* demo sequence (17,682 frames at 30 frames/sec), which we refer to as *Sony Demo* sequence, the first half hour of the *Silence of the Lambs* movie (54 000 frames at 30 frames/sec), the *Star Wars IV* movie (54 000 frames at 30 frames/sec), and the first hour of the *Tokyo Olympics* video (133 128 frames at 30 frames/sec). We also use about 30 minutes of the *NBC 12 News* (49 523 frames at 30 frames/sec), including the commercials. The video sequences *Silence of the Lambs, Star Wars IV, Tokyo Olympics*, and *NBC 12 News* can, respectively, be described as drama/thriller, science fiction/action, sports, and news video. The *Sony Demo* sequence consists of 29 scenes with varying texture and motion complexities. Due to space constraints, we present in this paper only illustrative plots for encodings with *Silence of the Lambs* and *Star Wars IV*. The corresponding plots for the other video sequences are available in [73, 74].

3.2. Encoding tools

We used the MEncoder tool to decode the sequences into uncompressed YUV format and to subsample the originally higher resolution sequences to CIF resolution. We used the MPEG-4 Part 2 Microsoft v2.3.0 software, and the SVC reference software, named JSVM, version 5.9 for the temporal layer evaluations, and versions 7.10 and 7.13, respectively, for studying FGS and spatial scalability.

3.3. Encoding setup

We employ four GoP structures in our study of temporal scalability layers, namely, *IBPBPBPBPBPBPBPB* (16 frames, with 1 B frame per I/P frame), which we denote by *G16-B1*, *IBBBPBBBPBBBPBBB* (16 frames, with 3 B frames per I/P frame) denoted by *G16-B3*, *IBBBBBBBPBBBBBBB* (16 frames, with 7 B frames per I/P frame) denoted by *G16-B7*, and *IBBBBBBBBBBBBBBB* (16 frames, with 15 B frames per I frame) denoted by *G16-B15*. In the context of SVC, these four GoP structures are, respectively, designated by

Table 1: Video traffic and quality metrics for encoding with given quantization scale.

Metric	Definition
Frame size metrics	
M	Number of frames in video sequence
X_m	Size [bits] of encoded video frame m, $m = 1, \dots, M$
$\overline{X} = (1/M)\sum_{m=1}^{M} X_m$	Mean frame size of encoded video sequence
$S_X^2 = (1/(M-1))\sum_{m=1}^{M}(X_m - \overline{X})^2$	Variance of frame sizes (S_X is the standard deviation [bits])
$\text{CoV}_X = S_X/\overline{X}$	Coefficient of variation of frame sizes [unit free]
$X_{\max} = \max_{m=1,\dots,M} X_m$	Maximum frame size [bits]
$\text{PtM}_X = X_{\max}/\overline{X}$	Peak-to-mean frame size ratio [unit free]
Bit rate metrics	
T	Frame period [s]
$\overline{R} = \overline{X}/T$	Mean bit rate [bits/s]
$R_{\max} = X_{\max}/T$	Peak bit rate [bits/s]
GoP size metrics	
N	Number of frames in GoP
Y_n	Size of GoP n, $n = 1, \dots, M/N$ [bits]
$\overline{Y} = (M/N)\sum_{n=1}^{M/N} Y_n$	Mean GoP size [bits]
$\text{PtM}_Y = Y_{\max}/\overline{Y}$	Peak-to-mean GoP size ratio [unit free]
$\text{CoV}_Y = S_Y/\overline{Y}$	Coefficient of variation of GoP sizes [unit free]
Frame quality metrics	
Q_m	PSNR quality of frame m as defined in (2)
$\overline{Q} = (1/M)\sum_{m=1}^{M} Q_m$	Average PSNR quality of video sequence
$\text{CoQV} = S_Q/\overline{Q}$	Coefficient of quality variation

their "GoP size" which is the number of hierarchical B frames plus one key picture, either of type I or P. Hence, *G16-B1* has GoP size 2, *G16-B3* has GoP size 4, *G16-B7* has GoP size 8, and *G16-B15* has GoP size 16. In the following, we employ our own GoP structure notation to emphasize the repetitive I-P-B frame type patterns in the encodings. These four GoP structures are natural structures for hierarchical B frames and allow us to compare temporal layer statistics across encoders based on identical underlying GoP patterns.

Due to space constraints, we primarily focus in this paper on the temporal scalability layers for the *G16-B3* GoP structure, which supports three temporal layers. The other GoP structures are presented in [73]. In our study of the spatial and FGS scalability layers, we focus on the GoP structure *G16-B3* since the RD efficiency of MPEG-4 Part 2 deteriorates for more B frames, making a comparison across encoders less useful.

3.4. Video traffic metrics

We summarize the video traffic and quality metrics, which are all defined with respect to a given video sequence encoded with a fixed quantization scale, in Table 1. We remark that the coefficient of variation of the frame sizes CoV_X is widely employed as a measure of the variability of the frame sizes, that is, the bit rate variability of the encoded video. Plotting the CoV as a function of the quantization scale (or equivalently, the PSNR video quality) gives the rate variability-distortion (VD) curve [48]. Alternatively, the peak-to-mean (Peak/Mean or PtM) ratio of the frame sizes is commonly used to express the traffic variability.

Regarding the bit rate metrics, we note that if each video frame is transmitted during one frame period T (e.g., 33 milliseconds for 30 frames/s), then the bit rate R_m [bits/s] required to transmit frame X_m is $R_m = X_m/T$. The corresponding mean bit rate $\overline{R}$ and peak bit rate $R_{\max}$ [bits/s] are defined in Table 1.

We define a *Group of Pictures* (GoP) of an encoded video stream as one I frame and all subsequent P and B frames before the next I frame in the stream. The size Y_n [bits] of GoP n equals the sum of the sizes of the N frames that belong to the GoP.

We use the peak signal-to-noise ratio (PSNR) as the objective measure of the quality of a reconstructed video frame $R(x, y)$ with respect to the uncompressed video frame $F(x, y)$. The larger the difference between $R(x, y)$ and $F(x, y)$, or equivalently, the lower the quality of $R(x, y)$, the lower the PSNR value. The PSNR is expressed in decibels [dB] to accommodate the logarithmic sensitivity of the human visual system. The PSNR is typically obtained for the luminance video frame and in case of a $N_x \times N_y$ frame consisting of 8-bit pixel values, it is computed as a function of the mean squared error (MSE) as

$$\text{MSE} = \frac{1}{N_x \cdot N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \left[F(x, y) - R(x, y)\right]^2, \tag{1}$$

$$\text{PSNR} = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}}. \tag{2}$$

We denote the PSNR quality of a video frame m by Q_m. For a detailed definition of all statistics used in this study, we refer to [75].

4. TEMPORAL SCALABILITY TRAFFIC ANALYSIS

We examine the traffic characteristics of the temporal layers embedded in video streams encoded with H.264 SVC and MPEG-4 Part 2. We demonstrate that the traffic variability of H.264 SVC temporal layers is significantly higher than the variability of the corresponding MPEG-4 Part 2 temporal layers. For a fair comparison, we assume that the same temporal layers as for H.264 SVC can be extracted from the MPEG-4 Part 2 traffic. Although the bitstream syntax of the latter does not support this extraction, it is in principle feasible for an intelligent media gateway or decoder to drop the B frames belonging to the respective temporal layers according to the H.264 SVC dyadic layer principle.

4.1. *Temporal layer basics*

The *G16-B3* GoP structure implies the repetition of the frame type pattern $I_0B_2B_1B_2P_0B_2B_1B_2P_0B_2B_1B_2P_0B_2B_1B_2$, whereby the subscripts denote the temporal layers (0, 1, 2) to which a frame belongs. The temporal base layer (0) is therefore $I_0000P_0000P_0000P_0000$, with zeroes replacing the dropped B frames of temporal enhancement layers 1 and 2. The first temporal enhancement layer is $00B_1000B_1000B_1000B_10$ and the second enhancement layer is $0B_20B_20B_20B_20B_20B_20B_20B_2$.

In case of our CIF sequences at a frame rate of 30 frames per second (fps), the temporal base layer represents a stream with a frame rate of 7.5 fps, the combination (aggregation) of the base layer and the first enhancement layer increases the frame rate to 15 fps, and the reception of the second enhancement layer results in the full frame rate of 30 fps. We note that the temporal base layer frames are required for decoding enhancement layer 1 frames, and that enhancement layer 2 frames need both lower layers to be decoded.

Let us examine the video quality associated with receiving certain temporal layers. Clearly, the average PSNR video quality of the combination of all temporal layers, that is, of the aggregated traffic, is equal to the average quality of the single-layer video stream. However, if we would simply average the PSNR values (Q) of the base layer frames ($Q_{I_0}Q_{P_0}Q_{P_0}Q_{P_0}$), then this average would be unrealistically high compared to the average of the corresponding single-layer (30 fps) stream, since the subjective quality impression of human observers is much lower for the frame rate of 7.5 fps. In order to include this perceptual quality degradation in the PSNR measurement, we assume that the decoder duplicates a received base layer frame until the next frame is received and decoded. The result is the duplicated base layer frame sequence $I_0^0I_0^0I_0^0I_0^0P_0^4P_0^4P_0^4P_0^4P_0^8P_0^8P_0^8P_0^8P_0^{12}P_0^{12}P_0^{12}P_0^{12}$, where the upper index represents the duplicated frame number. This sequence has a frame rate of 30 fps.

The PSNR value of a duplicated frame n located at frame number i is calculated based on the MSE between this duplicated frame n and the original frame i from the original (uncompressed) sequence. This PSNR value reflects the subjective distortion that occurs when jerky sequences consisting of duplicated frames are viewed by human observers. In general, the perceived video quality of a sequence is high if the average PSNR is high and the quality variation is low [5]. When there is low motion activity in the successive frames, that is, when frames are alike (low MSE), then duplication of frames results in barely noticeable jerkiness. The variation of the PSNR values is therefore also small. On the other hand, when high motion activity is present, then successive frames differ substantially and the MSE between successive frames is large, as well as the quality variation. The computed overall PSNR average therefore sufficiently incorporates the perceptual video quality reduction due to the reduced frame rate (jerkiness).

We apply the same principle to the computation of the average quality when the temporal base and first enhancement layer are received and decoded. This means that the following frame sequence is displayed: $I_0^0I_0^0B_1^2B_1^2P_0^4P_0^4B_1^6B_1^6P_0^8P_0^8B_1^{10}B_1^{10}P_0^{12}P_0^{12}B_1^{14}B_1^{14}$. The combination of all temporal layers results in displaying the sequence $I_0^0B_2^1B_1^2B_2^3P_0^4B_2^5B_1^6B_2^7P_0^8B_2^9B_1^{10}B_2^{11}P_0^{12}B_2^{13}B_1^{14}B_2^{15}$, which is the single-layer sequence.

Before we analyze the temporal layer traffic statistics, we describe the simple smoothing that we apply to the temporal base and enhancement layers to decrease the traffic variability. Let X_i denote the frame size (bytes) of frame i. Since there are large transmission gaps between frames of the base layer, we can redistribute the frame sizes over these gaps by dividing the frame size by four, and hence sending a quarter of each base layer frame during one frame period: $X_0^{I_0}/4$, $X_0^{I_0}/4$, $X_0^{I_0}/4$, $X_0^{I_0}/4, X_4^{P_0}/4, X_4^{P_0}/4, X_4^{P_0}/4$, $X_4^{P_0}/4$, $X_8^{P_0}/4$, $X_8^{P_0}/4, X_8^{P_0}/4, X_8^{P_0}/4, \ldots$. Equivalently, we say that we have smoothed the temporal base layer traffic over $a = 4$ frames. Analogously, the first enhancement layer traffic is smoothed over $a = 4$ frames: $X_2^{B_1}/4$, $X_2^{B_1}/4$, $X_2^{B_1}/4$, $X_2^{B_1}/4$, $X_6^{B_1}/4, X_6^{B_1}/4$, $X_6^{B_1}/4, X_6^{B_1}/4, X_{10}^{B_1}/4, X_{10}^{B_1}/4, X_{10}^{B_1}/4, X_{10}^{B_1}/4, X_{14}^{B_1}/4$, $X_{14}^{B_1}/4, X_{14}^{B_1}/4$, $X_{14}^{B^1}/4$. The second layer is smoothed over $a = 2$ frames since only one frame is missing in between the B frames of this layer: $X_1^{B_2}/2, X_1^{B_2}/2, X_3^{B_2}/2$, $X_3^{B_2}/2, X_5^{B_2}/2, X_5^{B_2}/2, X_7^{B_2}/2, X_7^{B_2}/2$, $X_9^{B_2}/2, X_9^{B_2}/2, X_{11}^{B_2}/2$, $X_{11}^{B_2}/2, X_{13}^{B_2}/2, X_{13}^{B_2}/2, X_{15}^{B_2}/2, X_{15}^{B_2}/2$. This basic smoothing introduces extra decoding delays, but mitigates to some extent the high rate variability as we demonstrate in the next section.

4.2. *Results and discussion*

We treat each temporal layer separately in the following analysis, except for the layer quality where we assume the reception of all lower layers. The aggregation of all layers is equivalent to the single-layer case, which is analyzed in detail in [36]. The main reason for treating each layer separately is that streaming protocols, such as the Real Time Protocol [76, 77], typically packetize and stream each layer separately to allow for different treatment of the layers in the network.

In Table 2, we summarize traffic and quality statistics of the temporal base layer and the two temporal enhancement

Table 2: Traffic statistics for the layers of temporal scalability encodings using H.264 SVC and MPEG-4 Part 2 for selected quantization scales with GoP structure G16B3.

		Frame Size			Bit Rate		Smoothed (a)		GoP Size		Frame Quality	
Enc.		Mean	CoV_X	PtM_X	Mean	Peak	$\text{CoV}_X^{(a)}$	$\text{PtM}_X^{(a)}$	CoV_Y	PtM_Y	Mean	CoQV
Mode		$\overline{X}$	$\frac{S_X}{\overline{X}}$	$\frac{X_{\max}}{\overline{X}}$	$\frac{\overline{X}}{T}$	$\frac{X_{\max}}{T}$	$\frac{S_X^{(a)}}{\overline{X}}$	$\frac{X_{\max}^{(a)}}{\overline{X}}$	$\frac{S_Y}{\overline{Y}}$	$\frac{Y_{\max}}{\overline{Y}}$	$\overline{Q}$	$\frac{S_Q}{\overline{Q}}$
		[kbyte]			[Mbps]	[Mbps]					[dB]	
Temporal base layer with $a = 4$ for smoothing												
SV28	Min	0.654	2.134	15.982	0.157	3.207	0.623	3.996	0.365	2.494	29.307	0.256
SV28	Mean	1.269	2.546	21.473	0.305	6.200	0.925	5.369	0.647	3.614	33.932	0.286
SV28	Max	1.930	2.878	30.193	0.463	8.504	1.149	7.549	0.885	5.340	36.639	0.316
Mp04	Min	0.935	2.120	17.539	0.224	5.684	0.611	4.386	0.345	2.370	29.230	0.256
Mp04	Mean	1.673	2.458	21.646	0.401	8.272	0.866	5.412	0.611	3.475	33.834	0.282
Mp04	Max	2.468	2.697	26.344	0.592	10.979	1.034	6.587	0.812	4.958	36.489	0.307
SV42	Min	0.137	2.609	27.363	0.033	0.957	0.976	6.843	0.384	2.580	26.324	0.167
SV42	Mean	0.229	2.975	31.574	0.055	1.689	1.206	7.895	0.655	4.149	29.678	0.218
SV42	Max	0.327	3.276	39.933	0.078	2.230	1.390	9.985	0.906	6.587	32.023	0.268
Mp20	Min	0.267	2.354	20.636	0.064	1.529	0.797	5.160	0.359	2.253	26.344	0.169
Mp20	Mean	0.392	2.463	22.345	0.094	2.094	0.875	5.587	0.493	3.256	29.692	0.201
Mp20	Max	0.515	2.596	23.858	0.124	2.832	0.967	5.966	0.612	4.281	32.345	0.230
Temporal enhancement layer 1 with $a = 4$ for smoothing												
SV28	Min	0.111	2.569	36.256	0.027	1.449	0.948	9.062	0.852	6.609	33.074	0.196
SV28	Mean	0.173	3.087	50.352	0.042	1.979	1.270	12.587	1.200	9.651	37.616	0.215
SV28	Max	0.288	3.787	76.317	0.069	2.655	1.684	19.079	1.654	17.038	40.393	0.241
Mp04	Min	0.430	2.043	19.267	0.103	3.147	0.541	4.816	0.490	3.345	33.351	0.195
Mp04	Mean	0.738	2.481	28.568	0.177	4.554	0.874	7.141	0.824	5.626	37.645	0.208
Mp04	Max	1.151	2.850	42.022	0.276	5.858	1.132	10.504	1.111	9.433	40.193	0.222
SV42	Min	0.027	2.588	36.294	0.006	0.254	0.961	9.072	0.831	5.856	28.275	0.121
SV42	Mean	0.032	2.866	50.069	0.008	0.377	1.138	12.516	1.057	8.654	31.395	0.171
SV42	Max	0.047	3.226	70.348	0.011	0.479	1.361	17.586	1.329	15.193	33.646	0.229
Mp20	Min	0.141	2.211	18.437	0.034	0.719	0.687	4.609	0.594	3.545	28.473	0.126
Mp20	Mean	0.176	2.319	27.240	0.042	1.104	0.767	6.809	0.645	4.694	31.626	0.153
Mp20	Max	0.212	2.619	51.862	0.051	1.752	0.982	12.964	0.705	6.591	34.359	0.177
Temporal enhancement layer 2 with $a = 2$ for smoothing												
SV28	Min	0.117	1.785	28.638	0.028	0.990	1.046	14.319	0.860	7.010	38.062	0.040
SV28	Mean	0.168	2.445	41.446	0.040	1.565	1.569	20.722	1.223	10.894	41.254	0.110
SV28	Max	0.322	3.333	64.004	0.077	2.376	2.248	32.002	1.645	18.111	44.013	0.148
Mp04	Min	0.777	1.250	11.196	0.186	3.023	0.530	5.598	0.471	3.711	39.164	0.027
Mp04	Mean	1.331	1.664	16.860	0.319	4.868	0.926	8.429	0.801	5.560	41.604	0.078
Mp04	Max	2.161	2.038	23.837	0.519	5.808	1.255	11.918	1.076	9.377	43.814	0.113
SV42	Min	0.027	1.693	18.412	0.007	0.120	0.966	9.205	0.776	6.499	30.191	0.053
SV42	Mean	0.035	1.956	33.211	0.008	0.275	1.185	16.604	0.949	8.196	32.709	0.126
SV42	Max	0.055	2.259	46.077	0.013	0.366	1.432	23.037	1.086	12.315	34.984	0.198
Mp20	Min	0.264	1.377	11.069	0.063	0.799	0.669	5.534	0.565	3.272	30.550	0.055
Mp20	Mean	0.332	1.454	14.515	0.080	1.115	0.744	7.257	0.602	4.269	33.257	0.105
Mp20	Max	0.403	1.633	25.870	0.097	1.639	0.913	12.934	0.630	5.277	36.000	0.129

layers included in the *G16-B3* GoP structure. The table includes frame size, bit rate, smoothed frame size, GoP size, and video quality statistics. We estimate these statistics based on the five long CIF sequences that we encode with H.264 SVC and MPEG-4 Part 2. In the first column of Table 2, the encoding mode is specified by a code representing the encoder (*SV* for H.264 SVC and *Mp* for MPEG-4 Part 2) and the quantization scale. For each encoder, we present min/mean/max values (computed across the five sequences) for two selected quantization scales that result in approximately equal PSNR quality (max-to-min) ranges. For example, the base layer quantization scale 28 for H.264

Table 3: Maximum (over quantization scales) of maximum (over video sequences), and maximum of mean CoV and PtM values of H.264 SVC and MPEG-4 Part 2 temporal base and enhancement layers (unsmoothed and smoothed (a)).

Encoder	max CoV_{max}	max PtM_{max}	max $\overline{\text{CoV}}$	max $\overline{\text{PtM}}$	max $\text{CoV}^{(a)}_{\text{max}}$	max $\text{PtM}^{(a)}_{\text{max}}$	max $\overline{\text{CoV}^{(a)}}$	max $\overline{\text{PtM}^{(a)}}$
Temporal base layer with $a = 4$ for smoothing								
H.264 SVC	3.28	39.93	2.99	34.18	1.39	9.98	1.22	8.55
MPEG-4	2.81	29.88	2.61	25.21	1.11	7.47	0.97	6.3
Temporal enhancement layer 1 with $a = 4$ for smoothing								
H.264 SVC	3.79	79.70	3.09	55.41	1.68	19.92	1.27	13.85
MPEG-4	2.85	55.56	2.49	32.61	1.13	13.89	0.89	8.15
Temporal enhancement layer 2 with $a = 2$ for smoothing								
H.264 SVC	3.36	64.00	2.44	42.98	2.27	32.00	1.57	21.49
MPEG-4	2.04	26.84	1.66	18.28	1.25	13.42	0.93	9.14

SVC results in the PSNR quality range 29.3–36.6 dB, and the quantization scale 4 for MPEG-4 Part 2 results in the quality range 29.2–36.5 dB. We compare the various statistical quantities in Table 2 based on matching quality ranges between encoders. Detailed results for the full range of studied quantization scales, which gives the RD and VD curves presented in this paper, are available in [73, 74].

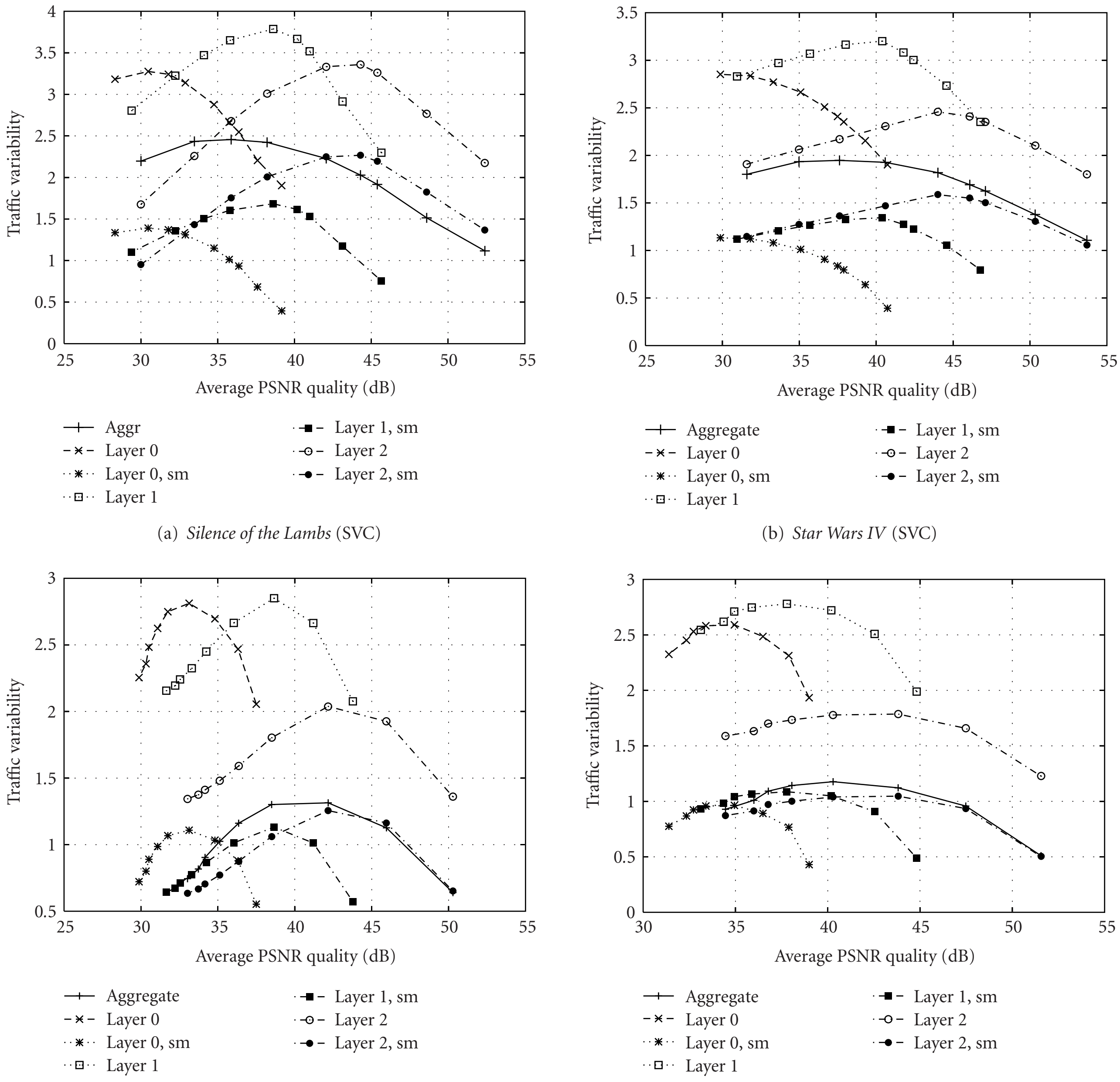
First, we can confirm the improved RD efficiency of the H.264 SVC temporal layers as compared to the MPEG-4 Part 2 layers based on the smaller mean frame sizes ranges (for corresponding quality ranges) or, equivalently, the lower mean bit rate ranges for H.264 SVC. Secondly, the mean bit rates are significantly lower for the H.264 SVC temporal enhancement layers as compared to the base layer rates. This is also the case for the MPEG-4 Part 2 enhancement layer rates as compared to the base layer, but to a lesser extent. The reason is that the base layer consists of large I and P frames (for both encoders). The assignment of cascading quantizers to the H.264 SVC B frames is responsible for the enhancement layer differences between the encoders. As opposed to MPEG-4 Part 2, H.264 SVC introduces cascading quantization scales that assign larger quantization parameters (lower quality and equivalently lower bit rate) to B frames belonging to higher temporal layers. This concept is based on the insight that the temporal base layer requires higher quality than the next temporal layer, since all other predictions depend on it. The quality (and bit rate) of each subsequent temporal layer can be gradually reduced, since fewer layers depend on it. The quality fluctuation that is introduced within a GoP is not subjectively noticeable by human observers, as studied in the standard committee. Hence, H.264 SVC temporal enhancement layers have significantly lower bit rates than the base layer as compared to MPEG-4 Part 2.

The quality analysis, and in particular the CoQV, demonstrates that the quality of the base layer is more variable than the quality when the first enhancement layer is additionally received by the decoder. When all layers are received, the CoQV is the lowest. We observe this quality variability decrease for H.264 SVC and MPEG-4 Part 2.

Next, we discuss the frame and GoP size coefficients of variation CoV and peak-to-mean ratios PtM of the temporal layers. Table 2 illustrates that the CoV and PtM of the unsmoothed frame size traffic, that is, traffic including zeroes (transmission gaps) for missing frames, are high for all temporal layers and all encoders. The zero frame sizes are the main reason behind this high variability. The H.264 SVC values are typically considerably higher than the MPEG-4 Part 2 values, for example, the H.264 SVC CoV values for the first enhancement layer are as high as 3.79 while the MPEG-4 Part 2 CoV reaches 2.85. With basic smoothing, the maximum CoV and PtM values decrease, for example, the maximum CoV of the first enhancement layer of H.264 SVC decreases to 1.68, while MPEG-4 Part 2's maximum CoV drops to 1.13. The GoP size CoV and PtM values exhibit similar trends, however, the differences between H.264 SVC and MPEG-4 Part 2 values are smaller, because the CoV and PtM of the GoP size are equal to the CoV and PtM of layers smoothed over the entire GoP ($a = 16$). Nevertheless, a fairly significant increase of H.264 SVC layer variability remains over MPEG-4 Part 2 with the mean CoVs of the H.264 SVC temporal enhancement layers being typically 1.5 times larger than the mean CoVs of the MPEG-4 Part 2 layers.

In Table 3, we provide an overview of the maximum CoV and PtM values for each temporal layer. The table includes the maximum of the maximum values, such as max CoV_{max}, and the maximum of the mean values, such as max $\overline{\text{PtM}}$. In every instance, the overall maximum is over all studied quantization scales (not only the selected quantization scales included in Table 2), while the inner maximum or mean is over all sequences for a given quantization scale.

Table 3 clearly demonstrates the higher CoV and PtM values of the H.264 SVC layer traffic as compared to the MPEG-4 Part 2 traffic. We observe that the first H.264 SVC enhancement layer has the highest CoV and PtM values among all unsmoothed layers. When smoothing is applied, the values of the second enhancement layer are highest, mainly because this layer is smoothed over two frames ($a = 2$) as compared to four frames ($a = 4$) for the other layers. Nevertheless, the advantage of traffic smoothing to reduce traffic variability is clear when comparing smoothed to unsmoothed values. After smoothing is applied, the H.264 SVC layers—especially enhancement layer 1 and even more so enhancement layer 2—still exhibit higher variability



(a) *Silence of the Lambs* (SVC)

(b) *Star Wars IV* (SVC)

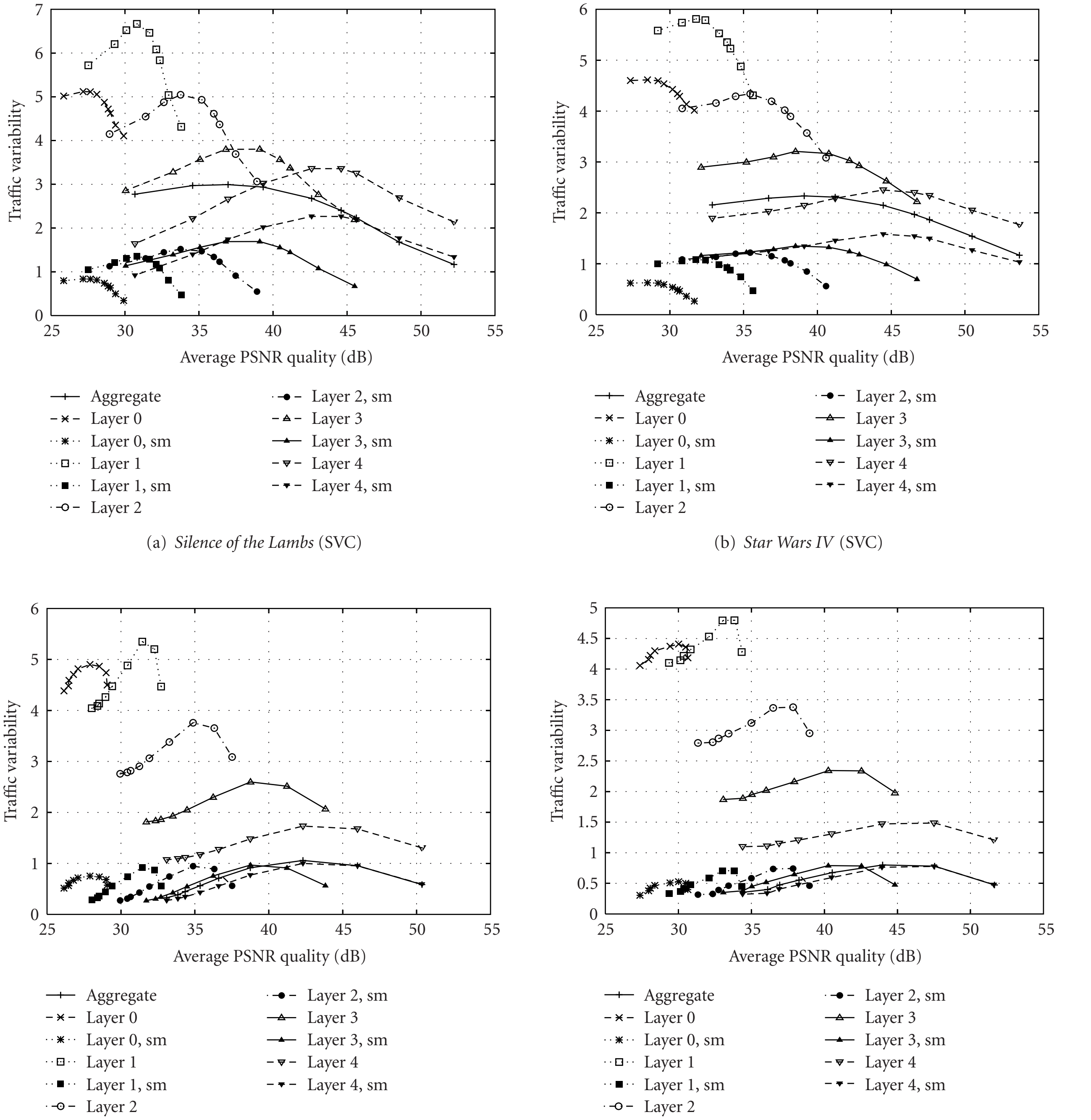
(c) *Silence of the Lambs* (MPEG-4)

(d) *Star Wars IV* (MPEG-4)

Figure 4: VD curves of three temporal layers (*G16-B3* GoP structure), unsmoothed, smoothed (sm), and aggregated (aggr), for the *Silence of the Lambs* and *Star Wars IV* sequences.

than MPEG-4 Part 2 layers, making network transport of H.264 SVC temporal layers more challenging. The main reason for the increased variability of H.264 SVC temporal layer traffic is attributable to the improved compression tools (e.g., motion compensated prediction) that manage to exploit redundancies more efficiently, and therefore are more amenable to frame content variations.

In Figure 4, VD curves are depicted for each temporal layer and the aggregated traffic (single-layer) of the *Silence of the Lambs* and *Star Wars IV* sequences encoded with H.264 SVC and MPEG-4 Part 2 (*G16-B3* GoP structure). We provide VD curves for unsmoothed and smoothed layer traffic. The VD curves for each temporal layer represent CoV values as a function of the average PSNR quality, obtained after decoding the current temporal layer and all lower layers, as we explained in Section 4.1. The average quality range increases from the temporal base layer VD curve with a quality range up to approximately 39 dB, to about 46 dB when the decoder additionally receives the first temporal enhancement layer, and to roughly 52 dB when the decoder receives all temporal layers. The figure also includes the VD curve of the aggregated traffic with values that lie between the individual unsmoothed and smoothed temporal layer VD curves. When comparing VD curves for the *Silence of*



(a) *Silence of the Lambs* (SVC)

(b) *Star Wars IV* (SVC)

(c) *Silence of the Lambs* (MPEG-4)

(d) *Star Wars IV* (MPEG-4)

Figure 5: VD curves of five temporal layers (*G16-B15* GoP structure), unsmoothed, smoothed (sm), and aggregated (aggr), for the *Silence of the Lambs* and *Star Wars IV* sequences.

the Lambs sequence in Figures 4(a) and 4(c), respectively, for H.264 SVC and MPEG-4 Part 2, the higher traffic variability (CoV) of H.264 SVC is pronounced. The same applies to the *Star Wars IV* sequence VD curves in Figures 4(b) and 4(d). Additionally, we depict the VD curves for five temporal layers of the *G16-B15* GoP structure in Figure 5. We observe even higher CoV values for the unsmoothed layers as compared to *G16-B3*.

5. SPATIAL SCALABILITY TRAFFIC ANALYSIS

In this section, we focus on the spatial scalability layers of H.264 SVC and MPEG-4 Part 2, employing GoP structure *G16-B3*. All five CIF sequences are downsampled to QCIF (176 × 144) resolution, which forms the spatial base layer of the encodings. The CIF layer forms the spatial enhancement layer. The statistical analysis treats each spatial layer separately, similar to the temporal layer analysis. We do not consider the temporal scalable layers that are present in each spatial layer, since they are the subject of the combined scalability analysis in Section 8. We compare the spatial layer traffic generated by both encoders, and we compare with single-layer QCIF and CIF traffic. The latter is warranted by the lower rate-distortion efficiency of spatial scalable encoding based on interlayer prediction, as compared to single-layer encoding, even though the H.264 SVC encoding tools represent an improvement over MPEG-4 Part 2.

5.1. Spatial layer basics

Since we do not consider temporal layer issues in this spatial layer analysis, the statistical processing of each spatial layer and the aggregated traffic follow the single-layer analysis approach. However, the average quality (PSNR) assigned to the QCIF layer does not represent the subjective quality perception, as compared to the CIF layer, if the lower resolution effect is not taken into account. Therefore, we upsample the decoded spatial QCIF base layer to CIF resolution and compute the average quality based on the MSE between the upsampled QCIF and the original (uncompressed) CIF sequence. The decoded CIF sequence is directly compared to the original sequence. This approach is warranted for receivers with CIF resolution displays, requiring upsampling of QCIF video streams to fit the display size. We realize that the applied upsampling technique plays a role in the subjective quality of the upsampled QCIF sequence. However for our practical traffic study, this is of a lesser concern. We also clarify that the quality that we associate with the spatial enhancement layer is identical to the quality of the aggregated traffic (base and enhancement layers), since the enhancement layer is only decodable if the spatial base layer has been received.

5.2. Results and discussion

In Table 4, we provide example H.264 SVC and MPEG-4 Part 2 traffic statistics (min/mean/max values across sequences as in Section 4) of the spatial base layer, spatial enhancement layer, the aggregated traffic, and single-layer QCIF and CIF traffic for comparison with the spatial layers. In the first column of the table, we specify the encoding mode by an encoder code (*SVS* for spatially scalable H.264 SVC and *Mp4S* for spatially scalable MPEG-4 Part 2) and the quantization scale.

We first analyze the spatial base layer traffic, comparing the mean frame sizes and mean bit rates of the H.264 SVC spatial base layer with the MPEG-4 Part 2 base layer for approximately the same quality ranges. We confirm the improved RD efficiency of H.264 SVC. The average qualities are overall quite low, since we used spatial upsampling to compute CIF resolution qualities, as explained in Section 5.1. The coefficient of quality variation CoQV is in the range of 0.11–0.19 for both encoders. For all spatial layers, Table 5 provides maximum-of-maximum and maximum-of-mean values for the CoV and PtM across all quantization scales and sequences. From the spatial base layer values, we observe overall significantly larger CoV and PtM values for H.264 SVC as compared to MPEG-4 Part 2, making the network transport of the H.264 SVC spatial base layer challenging.

In Table 4, we additionally summarize statistics of single-layer QCIF encodings for comparison with the spatial base layer statistics. Inspection of the values reveals that they are almost perfectly identical, which confirms that the spatial base layer is encoded independently from the spatial enhancement layers, and identical to single-layer encoding. The reason is that the interlayer tools predict the spatial enhancement layer employing the base layer and the latter is not predicted from the enhancement layer information. Therefore, the spatial base layer statistics follow single-layer trends that are extensively studied in [36].

Examples of H.264 SVC and MPEG-4 Part 2 traffic statistics of the spatial enhancement layer are summarized in Table 4. We also provide single-layer CIF statistics for comparison. For H.264 SVC, the average enhancement layer bit rate is more than twice the bit rate of the base layer for the highest qualities and converges to about the same bit rate for the lowest qualities, see [73]. For MPEG-4 Part 2, the enhancement layer bit rate is always significantly larger than the base layer rate. This is explained by the enhanced coding efficiency of H.264 SVC's interlayer prediction tools.

The enhancement layer average qualities extend to high qualities since the complete CIF resolution is decodable by receivers. The COQV values are about 0.04–0.13, which is lower than the base layer quality variability. Table 5 provides maximum-of-maximum and maximum-of-mean CoV and PtM enhancement layer values, which are typically twice as large or larger for H.264 SVC than for MPEG-4 Part 2. Furthermore, the H.264 SVC spatial enhancement layer has larger CoV and PtM values than the SVC base layer, while MPEG-4 Part 2 enhancement values are comparable to or lower than the base layer values. Secondly, the CoV and PtM enhancement layer values are only slightly larger than or comparable to single-layer CIF values in Table 5, for both H.264 SVC and MPEG-4 Part 2.

Next, we discuss the aggregated traffic statistics provided in Table 4, and compare with the enhancement layer and single-layer values. The mean frame sizes and bit rates are equal to the sum of the corresponding base and enhancement layer values. The quality statistics are identical to those of the enhancement layer, as discussed in Section 5.1. From Table 5, we again observe significantly larger maximum CoV and PtM values for the H.264 SVC aggregated traffic as compared to MPEG-4 Part 2. Compared to the SVC enhancement layer, the CoV and PtM values of the aggregated traffic are generally somewhat lower. Comparing the aggregated traffic statistics to the single-layer values reveals that the variabilities of the aggregate traffic are somewhat lower than the variabilities of the single-layer traffic.

In Figure 6, we depict VD curves of the spatial layers (QCIF and CIF) and the aggregated traffic, alongside the single-layer VD curves, for the *Silence of the Lambs* and *Star Wars IV* sequences encoded with H.264 SVC and MPEG-4 Part 2. We observe that the base layer and corresponding QCIF single-layer VD curves are identical for all sequences and encoders, as expected. Comparing Figures 6(a) and

Table 4: Traffic statistics for base (QCIF), and enhancement (CIF) layers, and aggregated traffic of spatial scalability encodings using H.264 SVC and MPEG-4 Part 2 for selected quantization scales.

		Frame size			Bit rate		GoP size		Frame quality	
Enc.		Mean	CoV$_X$	PtM$_X$	Mean	Peak	CoV$_Y$	PtM$_Y$	Mean	CoQV
Mode		$\overline{X}$	$\frac{S_X}{\overline{X}}$	$\frac{X_{\max}}{\overline{X}}$	$\frac{\overline{X}}{T}$	$\frac{X_{\max}}{T}$	$\frac{S_Y}{\overline{Y}}$	$\frac{Y_{\max}}{\overline{Y}}$	$\overline{Q}$	$\frac{S_Q}{\overline{Q}}$
		[kbyte]			[Mbps]	[Mbps]			[dB]	
Spatial base layer (QCIF)										
SVS28	Min	0.346	1.495	11.195	0.083	1.293	0.488	2.883	27.916	0.111
SVS28	Mean	0.632	1.757	15.603	0.152	2.163	0.664	6.344	31.930	0.146
SVS28	Max	0.890	2.033	25.207	0.214	2.886	0.901	11.235	36.493	0.161
Mp4S04	Min	0.639	0.951	7.610	0.153	2.444	0.557	3.560	27.940	0.111
Mp4S04	Mean	1.232	1.186	11.566	0.296	3.009	0.730	6.329	31.906	0.145
Mp4S04	Max	1.726	1.492	16.846	0.414	3.616	1.006	11.757	36.434	0.161
SVS42	Min	0.084	1.558	15.114	0.020	0.472	0.439	2.876	25.483	0.093
SVS42	Mean	0.135	1.756	20.831	0.032	0.645	0.535	6.241	28.516	0.132
SVS42	Max	0.186	2.201	26.993	0.045	0.835	0.630	9.837	31.685	0.165
Mp4S20	Min	0.168	1.028	9.527	0.040	0.598	0.486	2.776	25.486	0.104
Mp4S20	Mean	0.245	1.183	12.876	0.059	0.740	0.542	5.064	28.741	0.139
Mp4S20	Max	0.323	1.471	14.803	0.078	1.013	0.577	7.359	32.289	0.185
Single-layer for comparison with spatial base layer (QCIF)										
SV28	Mean	0.630	1.758	15.597	0.151	2.157	0.664	6.360	31.932	0.146
Mp04	Mean	1.230	1.188	11.543	0.295	2.994	0.731	6.342	31.909	0.145
SV42	Mean	0.134	1.758	20.931	0.032	0.644	0.531	6.198	28.532	0.132
Mp20	Mean	0.244	1.193	12.938	0.058	0.740	0.549	5.084	28.736	0.139
Spatial enhancement layer (CIF)										
SVS28	Min	0.427	1.612	14.171	0.103	2.316	0.539	2.805	37.068	0.047
SVS28	Mean	0.962	2.043	22.981	0.231	4.468	0.758	6.581	39.668	0.091
SVS28	Max	1.414	2.609	44.329	0.339	5.543	1.175	14.153	41.855	0.111
Mp4S04	Min	1.575	0.623	6.578	0.378	4.216	0.447	2.894	38.906	0.030
Mp4S04	Mean	3.578	0.987	10.719	0.859	7.629	0.727	5.137	40.856	0.079
Mp4S04	Max	5.372	1.422	19.985	1.289	9.477	1.048	10.109	42.717	0.105
SVS38	Min	0.131	1.861	21.410	0.032	0.830	0.473	2.894	31.365	0.062
SVS38	Mean	0.262	2.160	28.191	0.063	1.606	0.678	6.056	33.527	0.101
SVS38	Max	0.374	2.491	47.082	0.090	2.047	0.981	12.238	35.779	0.129
Mp4S16	Min	0.357	0.771	7.976	0.086	1.464	0.463	3.217	30.741	0.063
Mp4S16	Mean	0.667	0.931	12.594	0.160	1.787	0.641	5.056	33.247	0.095
Mp4S16	Max	0.954	1.148	18.205	0.229	2.262	0.807	9.413	35.976	0.124
Aggregated (base + enhancement) spatial traffic (CIF)										
SVS28	Min	0.773	1.545	12.888	0.185	3.374	0.510	2.773	37.068	0.047
SVS28	Mean	1.594	1.900	19.409	0.383	6.508	0.700	6.368	39.668	0.091
SVS28	Max	2.304	2.320	35.678	0.553	7.771	1.037	12.788	41.855	0.111
Mp4S04	Min	2.214	0.672	6.919	0.531	6.660	0.453	3.001	38.906	0.030
Mp4S04	Mean	4.810	0.989	10.787	1.154	10.462	0.710	5.344	40.856	0.079
Mp4S04	Max	6.983	1.380	19.069	1.676	12.019	1.021	10.585	42.717	0.105
SVS38	Min	0.252	1.761	19.558	0.061	1.461	0.460	2.735	31.365	0.062
SVS38	Mean	0.468	1.998	24.583	0.112	2.550	0.625	6.276	33.527	0.101
SVS38	Max	0.662	2.359	38.773	0.159	3.107	0.864	11.897	35.779	0.129
Mp4S16	Min	0.545	0.858	9.286	0.131	1.935	0.467	2.870	30.741	0.063
Mp4S16	Mean	0.953	0.964	12.334	0.229	2.627	0.605	5.154	33.247	0.095
Mp4S16	Max	1.343	1.084	16.330	0.322	3.489	0.744	9.256	35.976	0.124

Table 4: Continued.

		Frame size			Bit rate		GoP size		Frame quality	
Enc.		Mean	CoV$_X$	PtM$_X$	Mean	Peak	CoV$_Y$	PtM$_Y$	Mean	CoQV
Mode		$\overline{X}$	$\frac{S_X}{\overline{X}}$	$\frac{X_{\max}}{\overline{X}}$	$\frac{\overline{X}}{T}$	$\frac{X_{\max}}{T}$	$\frac{S_Y}{\overline{Y}}$	$\frac{Y_{\max}}{\overline{Y}}$	$\overline{Q}$	$\frac{S_Q}{\overline{Q}}$
		[kbyte]			[Mbps]	[Mbps]			[dB]	
Single-layer for comparison with aggregated traffic (CIF)										
SV28	Mean	1.616	1.922	19.987	0.388	6.780	0.680	6.289	40.979	0.082
Mp04	Mean	3.723	1.076	11.751	0.894	9.182	0.681	5.779	41.485	0.064
SV42	Mean	0.299	2.149	28.039	0.072	1.870	0.636	6.846	32.565	0.099
Mp20	Mean	0.922	0.944	10.687	0.221	2.339	0.485	4.127	33.377	0.094

Table 5: Maximum (across quantization scales) of maximum (across five video sequences), and maximum of mean CoV and PtM values of H.264 SVC and MPEG-4 Part 2 spatial base and enhancement layers, and aggregated traffic.

Encoder	max CoV$_{\max}$	max PtM$_{\max}$	max $\overline{\text{CoV}}$	max $\overline{\text{PtM}}$
Spatial base layer (QCIF)				
H.264 SVC	2.26	29.95	1.88	20.85
MPEG-4	1.57	20.27	1.33	15.00
Spatial enhancement layer 1 (CIF)				
H.264 SVC	2.63	49.05	2.17	28.19
MPEG-4	1.43	21.55	1.03	12.77
Aggregated (base + enhancement) traffic (CIF)				
H.264 SVC	2.36	39.95	2.01	24.58
MPEG-4	1.38	21.14	1.06	13.08
Single-layer (CIF)				
H.264 SVC	2.63	45.14	2.15	28.04
MPEG-4	1.41	19.21	1.15	13.56

6(c) for *Silence of the Lambs* encoded with H.264 SVC and MPEG-4 Part 2, clearly reveals the higher variability of the H.264 SVC base layer traffic. This is also observable in Figures 6(b) and 6(d) for the *Star Wars IV* sequence. The enhancement layer VD curves for H.264 SVC are above the MPEG-4 Part 2 curves in all cases. The VD curves of the aggregated traffic are the combined result of the base and enhancement layer variabilities, and as such, they are generally positioned between these two VD curves. An interesting distinction between H.264/SVC and MPEG-4 Part 2 is that the MPEG-4 layer 0 QCIF streams have higher traffic variabilities than the corresponding MPEG-4 layer 1 CIF streams. With H.264 SVC, this relationships is reversed, that is, the layer 1 CIF H.264 SVC streams have higher variability than the corresponding H.264 SVC layer 0 QCIF streams, further underscoring the high-traffic variability of the spatial enhancement layer of H.264 SVC.

6. FINE GRANULAR SCALABILITY TRAFFIC ANALYSIS

We compare H.264 SVC fine granularity scalability (SVC FGS) with MPEG-4 Part 2 FGS (MPEG-4 FGS) traffic based on GoP structure *G16-B3*. We analyze the base and enhancement layers separately and do not consider the temporal layers in this section, since they are the subject of our combined FGS-temporal analysis in Section 8.

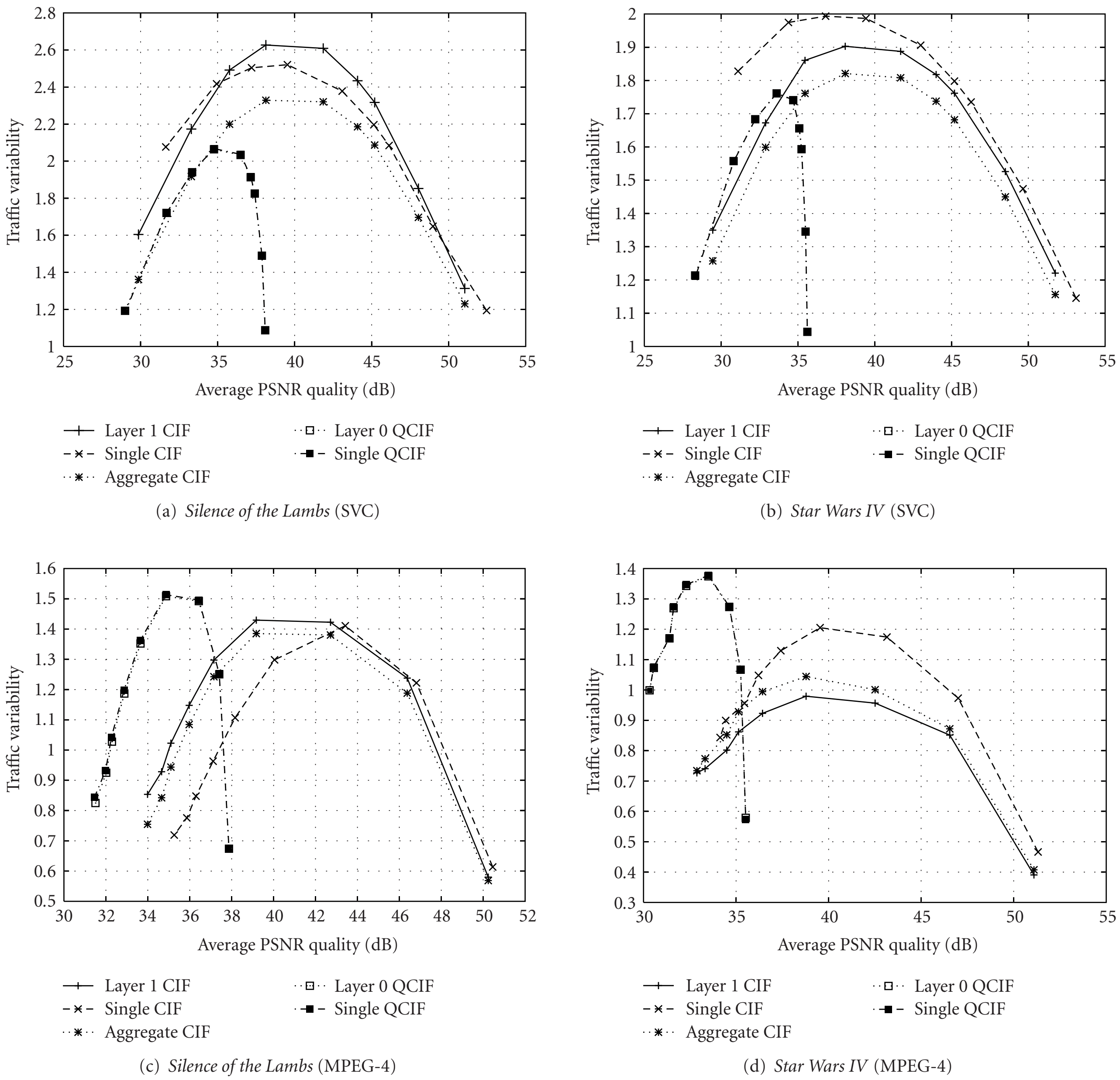
6.1. FGS layer basics

For MPEG-4 FGS, many possible FGS structures can be used such as basic FGS, FGS temporal (FGST), combined FGS-FGST, and multilayer FGST, which are detailed in [71]. In this study, we use the basic FGS structure, depicted in Figure 3(c), with one FGS enhancement layer frame for every base layer frame. We employ the H.264 FGS prediction loop illustrated in Figure 3(b), which is closed with respect to both the highest and lowest quality points.

The subsequent FGS analysis is based on the CIF video sequences *Silence of the Lambs*, *Star Wars IV*, *NBC 12 News*, and *Sony Demo*. We configured both encoders with one FGS enhancement layer and specified the base layer quantization scale. We study the traffic characteristics of the FGS base layer, the untruncated and the truncated enhancement layer, as well as the aggregated (base + enhancement) traffic.

6.2. Results and discussion

We analyze the statistics of base, enhancement, and aggregated traffic separately, in correspondence with the various



(a) *Silence of the Lambs* (SVC)

(b) *Star Wars IV* (SVC)

(c) *Silence of the Lambs* (MPEG-4)

(d) *Star Wars IV* (MPEG-4)

Figure 6: VD curves of two spatial layers (0 = QCIF, 1 = CIF), the aggregated traffic (aggr), and single-layer traffic (single), for the *Silence of the Lambs* and *Star Wars IV* sequences (*G16-B3* GoP structure).

possibilities of reception at the decoder. For selected base layer quantization scales, we present values of SVC FGS and MPEG-4 FGS traffic statistics for overlapping quality ranges in Table 6. We provide minimum, mean, and maximum (across the five video sequences) values of the traffic statistics. In the first column of the table, the encoder quantization scales are specified for MPEG-4 FGS (Mp4f) and SVC FGS (SVF). In Table 7, we present the maximum values across quantization scales and sequences. We observe from Table 6 a significant compression efficiency improvement in the base layer due to the improved tools in SVC FGS. These improved compression tools result in very high traffic variabilities for the SVC FGS base layer with maximum CoV and PtM values up to 2.5 and 39.9, as compared to up to 1.5 and 22.14 for MPEG-4 FGS, as observed in Table 7. The maximum of means values are similarly higher for SVC FGS. From these values, we conclude that significant traffic variability is introduced in the SVC FGS base layer as compared to MPEG-4 FGS. When comparing with single-layer H.264 SVC, see [74], we find that the base layer of SVC FGS (Table 6) is nearly identical since the prediction structure of both utilizes a closed loop.

Table 6 also gives selected examples to compare the untruncated FGS enhancement layers of both encoders. From Table 7, CoV and PtM have maxima up to 2.11 and 20.28, respectively, for SVC FGS, compared to up to 0.6

Table 6: Frame size, GoP size, bit rate, and quality statistics of FGS encodings using SVC FGS and MPEG-4 FGS for selected base layer quantization scales.

		Frame size			Bit rate		GoP size		Frame quality	
Enc.		Mean	CoV$_X$	PtM$_X$	Mean	Peak	CoV$_Y$	PtM$_Y$	Mean	CoQV
mode		$\overline{X}$	$\frac{S_X}{\overline{X}}$	$\frac{X_{\max}}{\overline{X}}$	$\frac{\overline{X}}{T}$	$\frac{X_{\max}}{T}$	$\frac{S_Y}{\overline{Y}}$	$\frac{Y_{\max}}{\overline{Y}}$	$\overline{Q}$	$\frac{S_Q}{\overline{Q}}$
		[kbyte]			[Mbps]	[Mbps]			[dB]	
FGS base layer										
SVF24	Min	1.384	1.451	11.885	0.332	4.227	0.439	2.617	38.194	0.041
SVF24	Mean	2.213	1.784	15.742	0.531	7.831	0.663	5.416	42.675	0.086
SVF24	Max	3.595	2.163	25.351	0.863	11.221	1.002	9.497	45.627	0.111
Mp4f04	Min	1.881	0.757	7.409	0.451	5.698	0.498	3.053	38.978	0.033
Mp4f04	Mean	3.741	1.116	11.915	0.898	8.976	0.705	6.014	41.303	0.080
Mp4f04	Max	5.884	1.470	18.992	1.412	11.170	1.055	11.442	43.332	0.103
SVF38	Min	0.271	1.787	18.807	0.065	1.380	0.485	2.759	30.353	0.066
SVF38	Mean	0.421	2.118	25.571	0.101	2.413	0.635	6.843	34.276	0.092
SVF38	Max	0.644	2.488	42.057	0.155	3.389	0.936	12.368	37.266	0.116
Mp4f16	Min	0.568	1.125	11.616	0.136	1.994	0.493	2.722	31.565	0.061
Mp4f16	Mean	0.881	1.257	13.975	0.211	2.821	0.593	5.506	34.144	0.087
Mp4f16	Max	1.284	1.431	17.229	0.308	3.581	0.720	8.633	37.045	0.115
FGS enhancement layer (untruncated)										
SVF24	Min	2.098	1.040	4.936	0.504	3.958	0.338	2.193	40.949	0.049
SVF24	Mean	3.230	1.320	6.905	0.775	4.847	0.422	3.079	45.829	0.072
SVF24	Max	4.641	1.572	10.007	1.114	5.635	0.578	4.557	48.452	0.088
Mp4f02	Min	2.404	0.328	1.708	0.577	1.738	0.199	1.527	44.064	0.014
Mp4f02	Mean	4.129	0.496	2.740	0.991	2.364	0.397	2.308	46.157	0.058
Mp4f02	Max	6.390	0.683	4.105	1.534	2.720	0.605	3.596	47.633	0.081
SVF28	Min	1.446	1.453	6.187	0.347	3.156	0.364	2.029	38.086	0.049
SVF28	Mean	2.089	1.591	8.804	0.501	4.062	0.440	3.179	43.481	0.077
SVF28	Max	3.381	1.788	13.013	0.811	5.020	0.615	4.903	46.279	0.096
Mp4f28	Min	12.573	0.274	1.920	3.017	7.530	0.266	1.865	42.052	0.093
Mp4f28	Mean	20.333	0.398	2.526	4.880	11.276	0.392	2.417	44.248	0.114
Mp4f28	Max	29.879	0.574	3.934	7.171	14.135	0.569	3.746	46.286	0.134
FGS aggregated (base + untruncated enhancement) traffic										
SVF24	Min	3.401	1.138	7.902	0.816	6.849	0.398	2.378	40.949	0.049
SVF24	Mean	5.128	1.398	9.632	1.231	11.325	0.518	3.940	45.860	0.079
SVF24	Max	8.236	1.659	13.883	1.977	15.619	0.724	6.055	48.801	0.098
Mp4f02	Min	6.695	0.379	3.508	1.607	9.912	0.300	2.119	44.064	0.014
Mp4f02	Mean	12.929	0.645	5.883	3.103	15.836	0.527	3.857	46.157	0.058
Mp4f02	Max	21.136	0.922	8.549	5.073	21.861	0.806	6.822	47.633	0.081
SVF28	Min	2.344	1.507	9.301	0.563	5.233	0.408	2.252	38.086	0.049
SVF28	Mean	3.329	1.601	11.470	0.799	8.811	0.503	4.022	43.433	0.083
SVF28	Max	5.644	1.777	16.390	1.354	12.682	0.688	6.445	46.565	0.102
Mp4f28	Min	13.149	0.268	2.045	3.156	7.975	0.259	1.847	42.052	0.093
Mp4f28	Mean	21.008	0.390	2.591	5.042	12.081	0.383	2.363	44.248	0.114
Mp4f28	Max	30.601	0.560	3.906	7.344	15.017	0.554	3.625	46.286	0.134

and 4.0 for MPEG-4 FGS. The SVC FGS enhancement layer has been subject to improved compression tools, resulting in increased variability at the frame level. Analogously, for the aggregated traffic with untruncated enhancement layer (Table 6), we have a CoV of 1.97 and a PtM of 25.5 for SVC FGS, as compared to 0.92 and 8.54 for MPEG-4 FGS.

Next, we examine the RD graphs of the SVC FGS and MPEG-4 FGS layers. Figure 7 depicts the base, untruncated enhancement, and aggregated traffic (base + untruncated enhancement) RD graphs for SVC FGS and MPEG-4 FGS encodings of the *Silence of the Lambs* sequence. The FGS base layer RD graphs are typical (quality increases monotonically

Table 7: Maximum-of-maximum, and maximum-of-mean CoV and PtM values of SVC FGS and MPEG-4 FGS base and enhancement layers, and aggregated traffic.

Encoder	max $\text{CoV}_{\max}$	max $\text{PtM}_{\max}$	max $\overline{\text{CoV}}$	max $\overline{\text{PtM}}$
FGS base layer				
H.264 SVC	2.5	39.9	2.13	25.9
MPEG-4	1.5	22.14	1.3	14.8
FGS enhancement layer (untruncated)				
H.264 SVC	2.11	20.28	1.87	11.9
MPEG-4	0.6	4.0	0.49	2.74
Aggregated FGS traffic (base + untruncated enhancement)				
H.264 SVC	1.97	25.5	1.76	15.5
MPEG-4	0.92	8.54	0.64	5.88

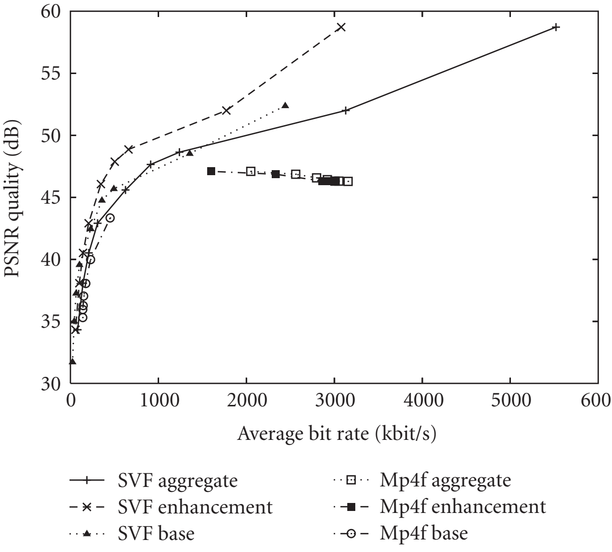


Figure 7: RD curves for SVC FGS and MPEG-4 FGS base and untruncated enhancement (`enh`) layers, and aggregated traffic (`aggr`) of *Silence of the Lambs* sequence (*G16-B3*).

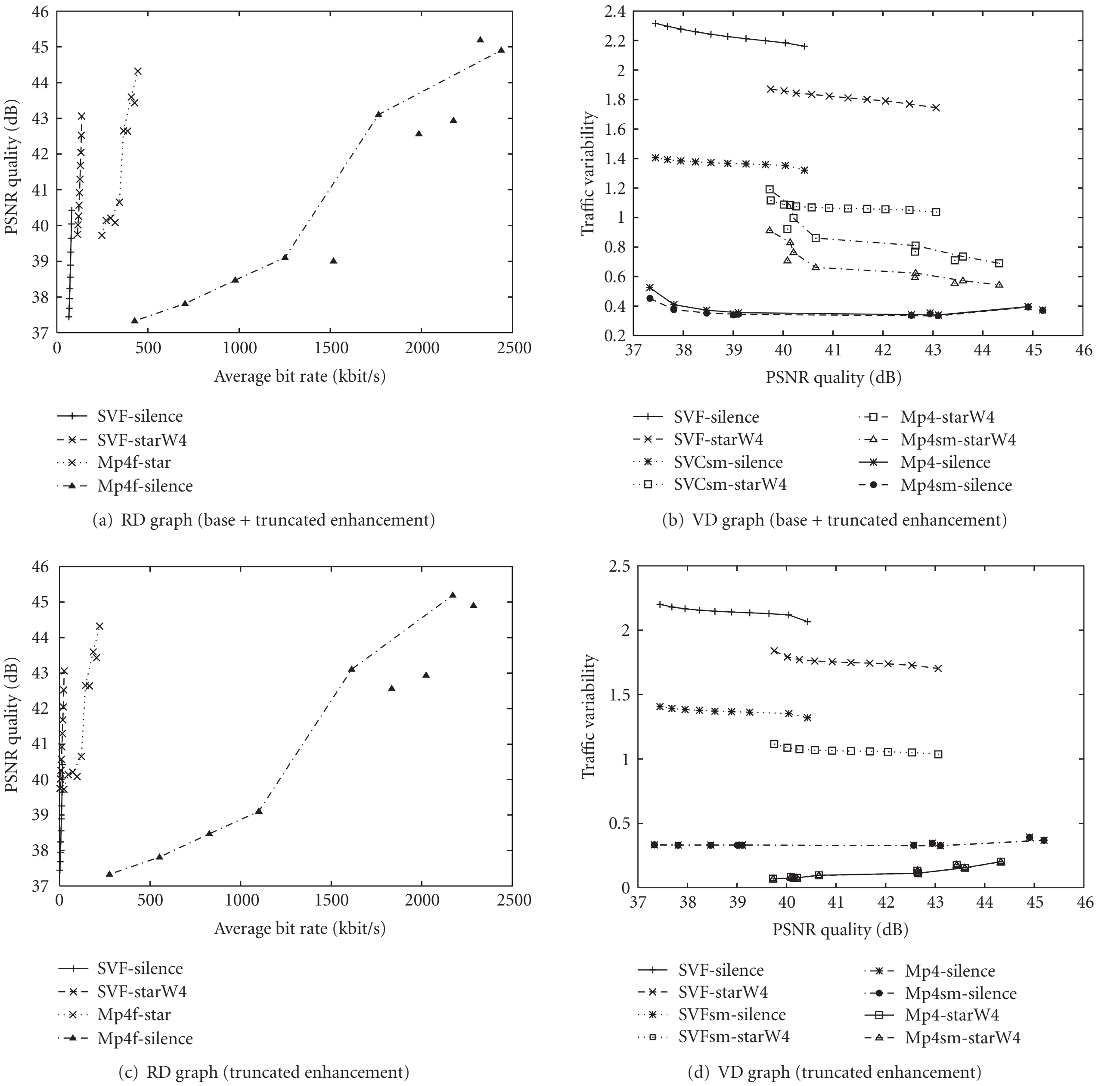
as a function of the bit rate) and demonstrate the improved RD efficiency of SVC FGS in the base layer. The untruncated enhancement layer for MPEG-4 FGS contains refinement information allowing high-quality reconstruction of the frames, resulting in the near-flat RD curve. The aggregated traffic RD graphs are the summation of the base and untruncated enhancement layer rates (per quality value).

To compare MPEG-4 FGS and SVC FGS with various truncations of the enhancement layer, we use average base layer PSNR qualities that are approximately equal. For *Star Wars IV*, we select quantization scales 34 and 8, respectively, for SVC FGS and MPEG-4 FGS, corresponding to an average base layer PSNR of approximately 34 dB. We further choose quantization scales 38 and 16 for SVC FGS and MPEG-4 FGS corresponding to a PSNR of approximately 37 dB for the *Silence of the Lambs* sequence. We truncate the enhancement layer progressively with 10% increments of the enhancement layer bit rate.

The RD graphs obtained for the aggregated (base + truncated enhancement) traffic for both sequences are depicted in Figure 8(a). The steep rise of the SVC FGS enhancement layer RD curve for every 10% increment in bit rate is in clear contrast to MPEG-4 FGS, which has a much lower RD performance with more gradual increments. This lower RD performance is explained by ignoring the enhancement layer in the prediction loop of MPEG-4 FGS. This also clearly demonstrates the substantial coding improvements made to the enhancement layer of SVC FGS, without significantly increasing the computational complexity (a major concern for portable devices). We also observe from Figure 8 that the *Star Wars IV* sequence has a better RD performance, which is consistent with earlier results. We note that the truncation of the MPEG-4 FGS enhancement layer resulted in outliers that are included in Figure 8 as disconnected tick marks.

The VD curves illustrate the significant contrast in variability between SVC FGS and MPEG-4 FGS. These VD curve points correspond to the RD curve points and represent the variability of the progressively truncated enhancement layer. For SVC FGS, we observe a marginal decrease in variability for increasing bit rate. The plots also include the smoothed traffic ($a = 4$, denoted by sm) VD curves, which show that the high variability of the SVC FGS stream can be significantly reduced by smoothing. However, the unsmoothed MPEG-4 FGS curves lie well below the smoothed SVC FGS stream curves, pointing to the inherently high variability introduced by the SVC FGS encoder. (The *Star Wars IV* VD curves for MPEG-4 FGS are above the *Silence of the Lambs* VD curves in Figure 8(b) due to the higher base layer CoV of *Star Wars IV* for the considered quantization scale 8.)

Although we consider a basic truncation strategy, which truncates each enhancement layer's progressive refinement (PR) slice by the same percentage, the traffic variability is still high. This is because the truncation of each PR slice results in widely variable truncated PR slice sizes (bytes). The SVC FGS traffic variability is consistently high across the range of percentages of enhancement layer added to the base layer; an important characteristic to take into account in the design of transport protocols as the enhancement layer is typically sent over a more error prone path with respect to the base layer.



(a) RD graph (base + truncated enhancement)

(b) VD graph (base + truncated enhancement)

(c) RD graph (truncated enhancement)

(d) VD graph (truncated enhancement)

Figure 8: RD and VD curves of MPEG-4 FGS and SVC FGS enhancement layers truncated progressively with 10% increments, for *Silence o/t Lambs* and *Star Wars IV* sequences.

7. MEDIUM GRAIN SCALABILITY TRAFFIC ANALYSIS

In this section, we focus on the medium grain scalability (MGS) mode of H.264 SVC, employing GoP structure *G16-B0*, which signifies 15 P frames in between I frames and no B frames. The resulting MGS base layer with CIF resolution conforms to the restricted Baseline profile of H.264/AVC. The MGS enhancement layer adds information that improves the quality of each video frame type up to the maximum quality encoded in the enhancement layer. Similar to the previous sections, the statistical analysis treats each layer separately and also aggregates the traffic in both layers. We compare the layer traffic generated by H.264 SVC MGS, however, we are not able to compare with equivalent traffic of MPEG-4 Part 2 since this standard does not include a similar quality scalability mode.

7.1. MGS layer basics

The statistical processing of the base layer, MGS enhancement layer, and the aggregated traffic follow the single-layer analysis approach. As for spatial scalability and FGS, the quality that we associate with the MGS enhancement layer is identical to the quality of the aggregated traffic (base and enhancement layers).

The MGS enhancement layer studied in this analysis supports one quality enhancement with a quantization parameter decrease of 6 (increased quality). We leave the study of multiple MGS quality extraction points (up to 16) within this enhancement layer for future research as well as the statistical analysis of the *G16-B3* GoP structure.

7.2. Results and discussion

Table 8 enumerates example H.264 SVC traffic statistics (min/mean/max) values across sequences of the base layer, MGS enhancement layer, and the aggregated traffic. In the first column of the table, we specify the encoding mode by the encoder code SVM followed by the quantization scale.

Comparing the mean bit rates between the base layer and corresponding MGS enhancement layer (same quantization scale), it is evident that the enhancement layer adds a large increase in bit rate to the base layer, and this for the entire range of studied quantization scales and sequences. The spanned decrease in quantization scale of 6, which halves the quantization step size, is encoded less efficiently by the MGS tools, resulting in the much larger required bit rates.

The CoV values of the MGS enhancement layer are considerably lower than the CoV values of the base layer (*G16-B0*). From Table 9, the maximum of the maximum CoV and PtM values are, respectively, 2.10 and 36.12 for the base layer while for the enhancement layer both values are 0.98 and 10.72. The differences are also this large for the maximum of the means of CoV and PtM. The aggregated traffic has maximum values that are comparable to or slightly larger than the values of the enhancement layer, hence the CoV and PtM values of the base layer are greatly reduced if transported in conjunction with the enhancement layer.

The statistics on the GoP level have similar trends, however the difference between the CoV values of the base; and enhancement layers is less pronounced while still significant differences exist between PtM values.

8. COMBINED SCALABILITY TRAFFIC ANALYSIS

The H.264 SVC encoder supports combined scalability that allows to extract temporal, spatial, and SNR layers from one bitstream. The result of this flexibility from a video traffic analysis viewpoint is that analyzing all possible temporal-spatial-SNR encoding combinations of layers is prohibitive. Therefore, we focus on two case studies: spatiotemporal and FGS-temporal scalability. We compare the base and enhancement layers to the traffic characteristics obtained and discussed in the preceding sections that analyzed each scalability mode in isolation.

First, we explore the combined spatiotemporal scalability case, which is based on the spatial scalable encodings used in Section 5, that is, we employ GoP structure *G16-B3* supporting three temporal layers in each spatial QCIF and CIF layer. Secondly, we analyze combined FGS-temporal scalability based on the encodings used in Section 6, supporting three temporal layers in the FGS base and enhancement layers.

8.1. Combined spatiotemporal scalability

Figures 9 and 10 depict the VD curves of the temporal layers in each spatial layer and in the aggregated traffic (base + enhancement) for the *Silence of the Lambs* and *Star Wars IV* sequences. Each complete spatial layer has been individually analyzed in Section 5. In the following, we focus on the temporal layers embedded in each spatial layer and compare with the corresponding single-layers.

First, we recall from Section 5 that the spatial base layer (QCIF) statistics are identical to the single-layer (QCIF) statistics, because these layers are identically encoded. Therefore, the temporal layer statistics of the spatial base layer are also identical to the statistics of the temporal layers embedded in the single-layer QCIF stream. Secondly, we compare the VD curves of the temporal layers embedded in the aggregated spatial stream in Figures 9(c) and 10(c) to the VD curves of the layers of the temporal-scalability only CIF encodings in Figures 4(a) and 4(b). Visual inspection reveals that these temporal layer VD curves have comparable values, however, with somewhat lower CoV values in the low-quality range of Figures 4(a) and 4(b). Thirdly, the spatial enhancement layer's temporal layers in Figures 9(b) and 10(b) cannot be directly compared with any prior results. However, visual inspection reveals that the VD curves in Figures 9(b) and 10(b) are very similar to the temporal layer VD curves of the aggregated spatial traffic in Figures 9(c) and 10(c). These VD curves have the same shapes, but the VD curves of the spatial enhancement layer have a slight vertical offset (somewhat higher CoV) than the VD curves of the aggregate streams. This indicates that the variability of the aggregated traffic is dominated by the spatial enhancement layer.

From (i) the similarity of the temporal layer VD curves of the spatial base and aggregate streams with the corresponding VD curves of the temporal-scalability only encodings, and (ii) the similarity of the temporal layers embedded in the spatial enhancement layer with the temporal layers in the aggregated spatial stream, we conclude that it suffices to separately analyze the layers of temporal-scalability only encodings at the individual spatial resolution (QCIF and CIF) to obtain good estimates of the traffic variabilities of the layers in the combined spatiotemporal encoding.

8.2. Combined FGS-temporal scalability

The SVC FGS encoder supports FGS-temporal scalability, which adds progressive refinement (PR) information to each temporal layer embedded in the base layer. This PR information is provided by the FGS enhancement layer. In this section, the three temporal layers included in the base and enhancement layer are separately examined. Figures 11 and 12 depict the temporal layers for base, untruncated enhancement, and aggregated (base + untruncated enhancement) traffic for the *Silence of the Lambs* and *Star Wars IV* sequences.

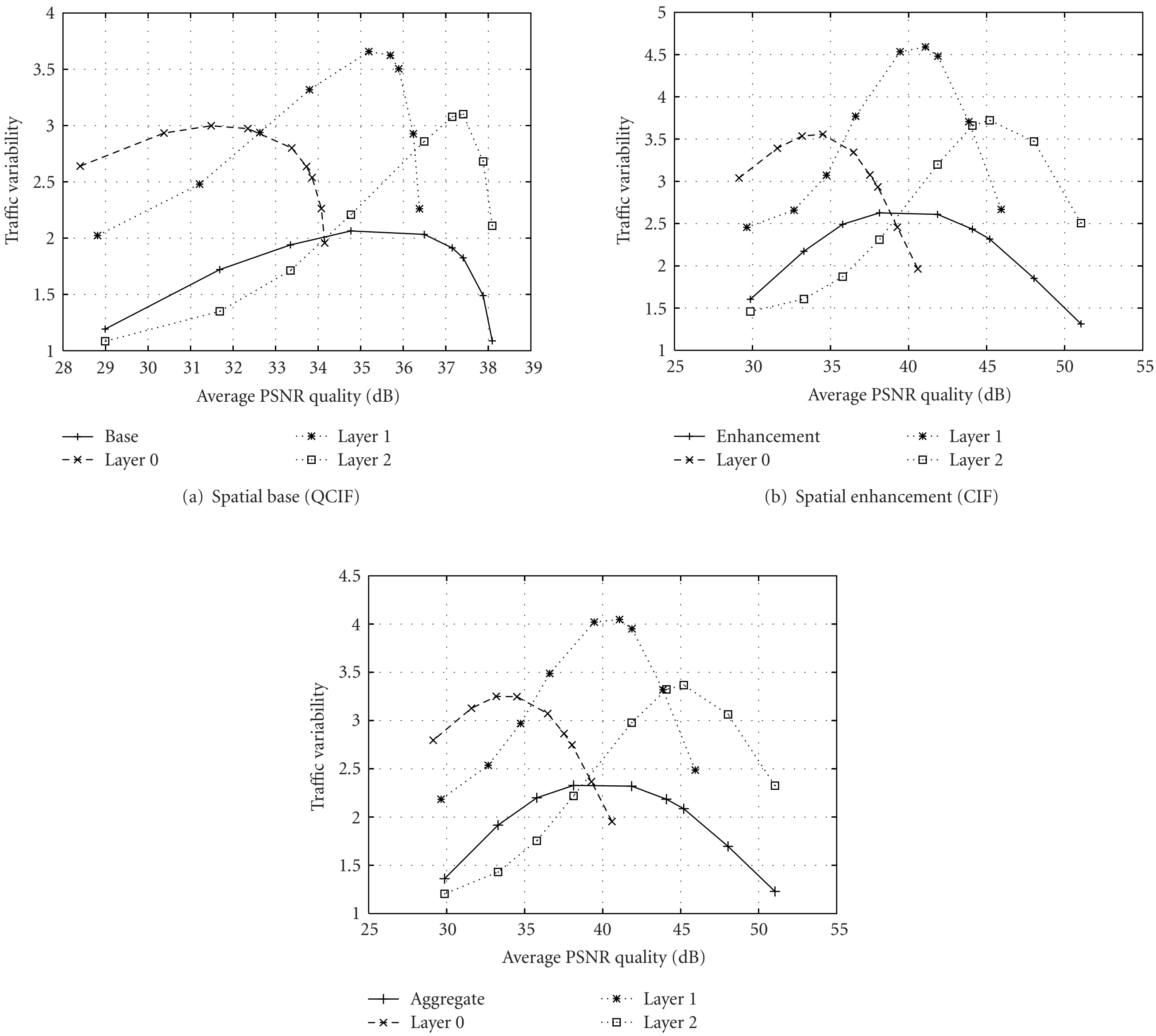
We compare the VD curves of the temporal layers embedded in the FGS base layers to the VD curves of the temporal-scalability only encodings in Figures 4(a) and 4(b).

Table 8: Traffic statistics for base, and enhancement layers, and aggregated traffic of medium grain scalability encodings using H.264 SVC for selected quantization scales.

		Frame size			Bit rate		GoP size		Frame quality	
Enc. mode		Mean $\overline{X}$ [kbyte]	CoV$_X$ $\frac{S_X}{\overline{X}}$	PtM$_X$ $\frac{X_{\max}}{\overline{X}}$	Mean $\frac{\overline{X}}{T}$ [Mbps]	Peak $\frac{X_{\max}}{T}$ [Mbps]	CoV$_Y$ $\frac{S_Y}{\overline{Y}}$	PtM$_Y$ $\frac{Y_{\max}}{\overline{Y}}$	Mean $\overline{Q}$ [dB]	CoQV $\frac{S_Q}{\overline{Q}}$
Base layer										
SVM28	Min	0.768	0.954	7.958	0.184	2.519	0.591	2.981	37.472	0.039
SVM28	Mean	1.725	1.383	15.551	0.414	5.347	0.790	7.395	39.682	0.083
SVM28	Max	2.843	1.950	29.873	0.682	6.646	1.195	14.523	41.519	0.109
SVM34	Min	0.363	1.248	11.808	0.087	1.542	0.567	3.177	33.324	0.054
SVM34	Mean	0.746	1.613	20.248	0.179	3.184	0.781	8.485	35.816	0.099
SVM34	Max	1.184	2.089	35.631	0.284	4.104	1.163	16.562	37.904	0.132
SVM38	Min	0.223	1.390	14.389	0.053	1.073	0.545	3.303	30.607	0.062
SVM38	Mean	0.430	1.695	22.234	0.103	2.083	0.742	8.808	33.216	0.098
SVM38	Max	0.675	2.051	36.118	0.162	2.623	1.073	16.600	35.411	0.147
SVM42	Min	0.141	1.403	16.700	0.034	0.771	0.522	3.236	28.229	0.069
SVM42	Mean	0.257	1.704	22.881	0.062	1.318	0.685	8.716	30.783	0.091
SVM42	Max	0.398	2.100	33.630	0.096	1.596	0.936	15.421	33.066	0.112
MGS enhancement layer										
SVM28	Min	3.653	0.328	1.975	0.877	2.616	0.319	1.854	41.648	0.020
SVM28	Mean	6.135	0.493	2.901	1.472	3.839	0.486	2.731	43.589	0.063
SVM28	Max	9.895	0.699	4.453	2.375	4.691	0.692	4.239	45.096	0.091
SVM34	Min	2.118	0.397	2.188	0.508	1.528	0.387	2.115	37.715	0.034
SVM34	Mean	3.774	0.579	3.531	0.906	2.809	0.570	3.395	39.566	0.077
SVM34	Max	6.513	0.833	5.662	1.563	3.420	0.823	5.517	41.211	0.110
SVM38	Min	1.409	0.416	2.383	0.338	1.040	0.404	2.348	34.438	0.043
SVM38	Mean	2.630	0.618	3.954	0.631	2.173	0.607	3.771	36.519	0.075
SVM38	Max	4.556	0.909	6.558	1.093	2.644	0.898	6.304	38.353	0.125
SVM42	Min	0.911	0.426	2.595	0.219	0.733	0.412	2.488	31.557	0.051
SVM42	Mean	1.735	0.648	4.409	0.416	1.595	0.634	4.140	33.772	0.069
SVM42	Max	2.936	0.960	7.282	0.705	2.042	0.946	6.858	35.523	0.087
Aggregated (base + enhancement) traffic										
SVM28	Min	4.421	0.396	3.378	1.061	4.798	0.341	1.988	41.648	0.020
SVM28	Mean	7.860	0.563	5.131	1.886	8.655	0.502	3.232	43.589	0.063
SVM28	Max	12.368	0.789	8.308	2.968	10.883	0.728	5.303	45.096	0.091
SVM34	Min	2.482	0.442	3.898	0.596	2.703	0.375	2.082	37.715	0.034
SVM34	Mean	4.520	0.622	5.869	1.085	5.634	0.561	3.544	39.566	0.077
SVM34	Max	7.514	0.882	9.785	1.803	7.031	0.824	5.798	41.211	0.110
SVM38	Min	1.632	0.452	4.072	0.392	2.028	0.382	2.267	34.438	0.043
SVM38	Mean	3.060	0.649	6.343	0.734	4.087	0.588	3.848	36.519	0.075
SVM38	Max	5.097	0.934	10.459	1.223	5.071	0.878	6.182	38.353	0.125
SVM42	Min	1.052	0.458	4.326	0.252	1.504	0.385	2.398	31.557	0.051
SVM42	Mean	1.992	0.670	6.741	0.478	2.834	0.607	4.110	33.772	0.069
SVM42	Max	3.244	0.962	10.715	0.779	3.478	0.907	6.486	35.523	0.087

Table 9: Maximum- (across quantization scales) of-maximum (across five video sequences), and maximum-of-mean CoV and PtM values of H.264 SVC MGS base and enhancement layers, and aggregated traffic.

Encoder	max CoV_{max}	max PtM_{max}	max $\overline{\text{CoV}}$	max $\overline{\text{PtM}}$
		MGS base layer		
H.264 SVC	2.10	36.12	1.70	22.88
		MGS enhancement layer		
H.264 SVC	0.98	7.68	0.66	4.84
		Aggregated (base + enhancement) traffic		
H.264 SVC	0.96	10.72	0.67	6.92

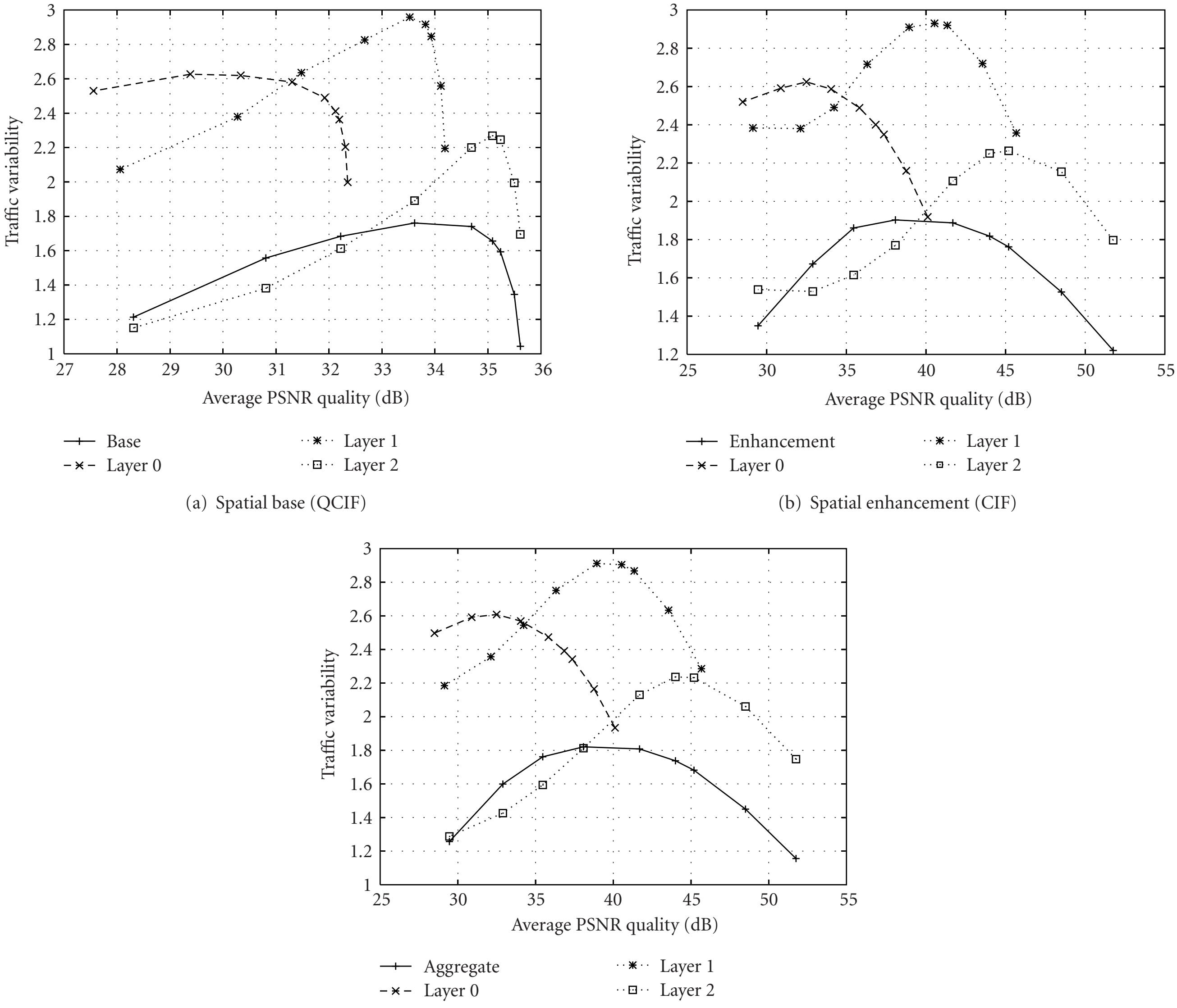


(a) Spatial base (QCIF)

(b) Spatial enhancement (CIF)

(c) Aggregated traffic (CIF)

Figure 9: VD curves of all temporal layers (0, 1, 2) embedded in each spatial layer and in the aggregated (base + enhancement) traffic stream, for the *Silence of the Lambs* sequence encoded with H.264 SVC (*G16-B3*).



(a) Spatial base (QCIF)

(b) Spatial enhancement (CIF)

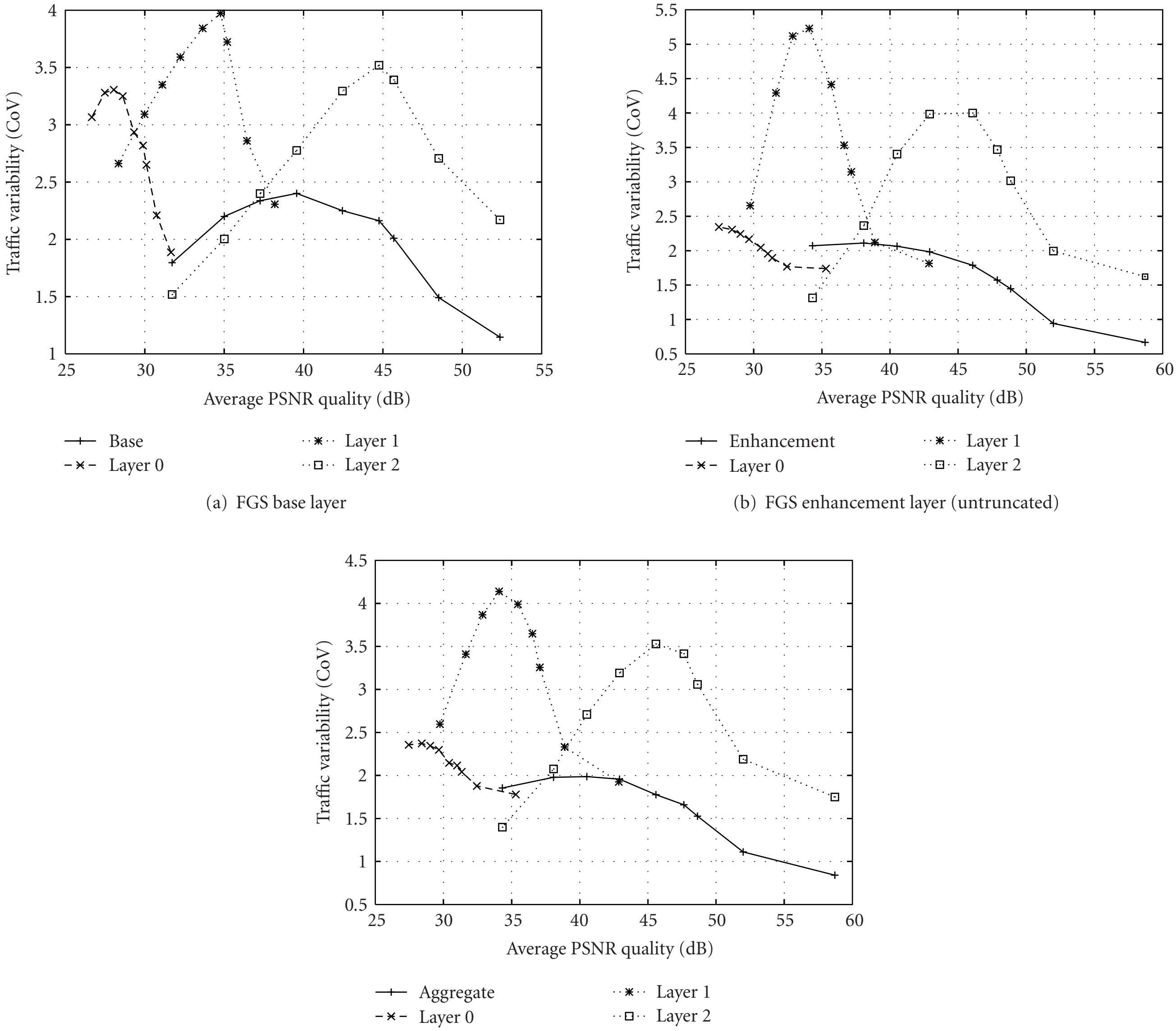
(c) Aggregated traffic (CIF)

Figure 10: VD curves of all temporal layers (0, 1, 2) embedded in each spatial layer and in the aggregated (base + enhancement) traffic stream, for the *Star Wars IV* sequence encoded with H.264 SVC (*G16-B3*).

First, we observe that the temporal layers embedded in the FGS base layers in Figures 11(a) and 12(a) have comparable variability to the layers of the temporal-scalability only encodings in Figures 4(a) and 4(b). Direct comparison of the VD curves in Figures 11(a) and 12(a) with the VD curves in Figures 4(a) and 4(b) is difficult, because the qualities associated with the temporal layers are computed differently (a constant low PSNR value is used for missing frames in Figures 11 and 12 versus the PSNR between duplicated, and original frame is used in Figures 4(a) and 4(b)). Nevertheless, the maximum CoV values and the CoV values at the low- and high-quality ends of corresponding curves are very close. Given this similarity between the VD curves of the temporal layers embedded in the FGS base layer and the VD curves of the layers in the temporal-scalability only streams in Figures 4(a) and 4(b), we conclude that it suffices to study the traffic statistics of the layers of temporal-scalability only encodings to obtain reasonable estimates of the traffic variabilities of the temporal layers embedded in the FGS base layer. On the other hand, the FGS enhancement layer traffic, the aggregated FGS traffic, and their embedded temporal layers cannot be meaningfully compared to any previously obtained results. However, the unprecedented high variabilities of these streams are indicative of the high variability the network path encounters when different layers are transmitted independently.

9. CONCLUSION

We examined the video traffic characteristics of the temporal, spatial, and FGS scalability modes of the scalable video coding (SVC) extension of the H.264/AVC standard and



(a) FGS base layer

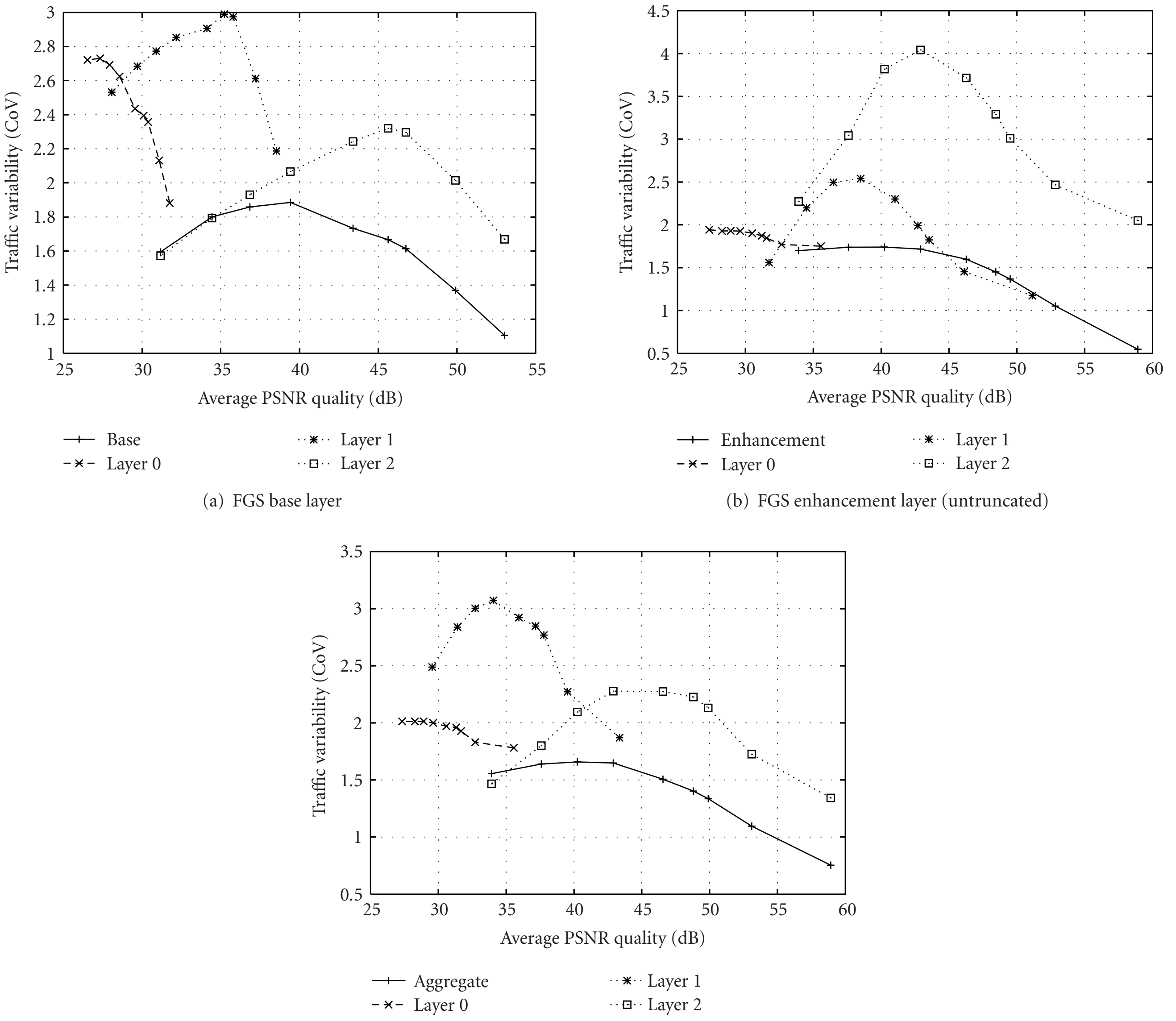
(b) FGS enhancement layer (untruncated)

(c) Aggregated FGS traffic (base + untruncated enhancement)

Figure 11: VD curves of all temporal layers (0, 1, 2) embedded in each FGS layer and in the aggregated (base + untruncated enhancement) traffic stream, for the *Silence of the Lambs* sequence encoded with SVC FGS (*G16-B3*).

compared with equivalent MPEG-4 Part 2 scalable video traffic. We also analyzed SVC's combined spatiotemporal and combined FGS-temporal scalability. Our traffic study focused on the joint characterization of the average bit rate and the bit rate variability as a function of the video quality. We employed long CIF resolution video sequences with a wide variety of texture and motion features. We summarize our findings for each scalability mode as follows.

(i) For the temporal scalability mode of SVC with three temporal layers, which we examined separately, we have found that the maximum coefficient of variation CoV of the frame sizes over all sequences and all unsmoothed SVC temporal layers is above 3.3, with the CoV of temporal layer 1 being as high as 3.8. For MPEG-4 Part 2, the maximum CoV stays below 2.9. Across temporal layers, we have found that temporal layer 1 has the highest variability. When basic smoothing is applied to SVC layers, we have found that the maximum CoV falls to 1.4 and 1.7 for the base layer and temporal layer 1, respectively, while the CoV of temporal layer 2 falls to 2.27. For MPEG-4 Part 2, the smoothed CoV does not exceed 1.25. These figures point to the significant increase in bit rate variability of temporal scalable SVC over MPEG-4 Part 2. From the bit rate and quality analysis, we find that the mean bit rates for the SVC temporal enhancement layers are significantly lower than for the base layer due to the presence of large I and P frames and the cascading quantizer assignment for SVC B frames. We also confirm that the coefficient of quality variation decreases as each layer is cumulatively added, thus increasing the subjective quality at the receiver.



(a) FGS base layer

(b) FGS enhancement layer (untruncated)

(c) FGS combination layer (base + untruncated enhancement)

Figure 12: VD curves of all temporal layers (0, 1, 2) embedded in each FGS layer and in the aggregated (base + untruncated enhancement) traffic stream, for the *Star Wars IV* sequence encoded with SVC FGS (*G16-B3*).

(ii) The spatial scalability traffic analysis first focused on the separate analysis of the QCIF base layer, the CIF enhancement layer, and the aggregated CIF stream, without considering the temporal scalability present in each spatial layer. We have found that SVC's spatial enhancement layer (CoV up to 2.6) has larger traffic variability than its base layer (CoV up to 2.3) contrary to MPEG-4 Part 2 enhancement layer's traffic variability (CoV up to 1.4) which is lower than or comparable to its base layer (CoV up to 1.6). We have also found that the spatial base layer statistics are perfectly identical to the single-layer QCIF statistics, confirming that the spatial base layer is encoded independently of the enhancement layer. The traffic variabilities of both SVC and MPEG-4 Part 2 for the enhancement layer (CIF) are comparable to or slightly higher than for single-layer CIF. For the aggregated traffic (CIF), we have found significantly higher traffic variability for SVC as compared to MPEG-4 Part 2. Comparing with the CIF enhancement layer, the CoV of the aggregated traffic is generally lower than that of the enhancement layer.

(iii) We analyzed FGS by treating base layer, enhancement layer, and aggregated traffic separately. There has been a significant effort in the SVC extension to improve the RD efficiency over MPEG-4 FGS, the success of which can be clearly seen in up to 50% improvement made in many cases. We have studied the simple truncation of the enhancement layer of both encoders in progressive steps of 10% of the total enhancement layer and have found that the variability of SVC for each point can be over 2.5 times that of MPEG-4 FGS, which has CoV values less than or equal to 1.

Smoothing the truncated bitstream lowers the SVC CoV to the range 1–1.5, while for MPEG-4 FGS, smoothing reduces the traffic variability to the range 0.4–0.6. Compared with single-layer encodings, we have found that the base layer statistics are quite similar, given that both use a closed loop prediction structure. We have observed that the untruncated enhancement layer of MPEG-4 FGS contains almost the full refinement information for the entire bit rate range (for all quantizers), resulting in an almost flat RD curve; in contrast, SVC provides significant quality increases for increases in the untruncated enhancement layer bit rate.

(iv) We examined combined spatiotemporal scalability by analyzing the temporal layers embedded in each spatial layer and compared with the layers in temporal-scalability only encodings. We have observed comparable values except in the low-quality range where somewhat lower traffic variability is exhibited by the temporal-scalability only encodings. We have also observed that the variability of the aggregated traffic is mainly determined by the spatial enhancement layer. From the fact that the VD curves of the temporal layers embedded in each spatial layer are similar to the VD curves of the corresponding temporal-scalability only encodings, and that the spatial enhancement layer is similar to that of the aggregated spatial traffic, we conclude that it suffices to analyze the video traffic of each resolution separately to obtain a good estimate of the traffic variabilities of all embedded layers. We also examined combined FGS-temporal scalability of SVC. Given the similarity of the temporal VD curves in the FGS base layer to the temporal layer curves embedded in the single layer, a reasonable estimate of the traffic variabilities of all layers embedded in the FGS base layer can be obtained from the single-layer equivalent.

Overall, these results clearly point to unprecedented levels of compression efficiency as well as traffic variability for SVC coding, a factor which should be taken into consideration for the design of efficient network transport protocols and mechanisms for H.264 SVC scalable-encoded video.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Patrick Seeling for contributing to the statistical processing. This work was supported in part by the National Science Foundation through Grants nos. ANI-0136774 and CRI-0750927.

REFERENCES

[1] H.-C. Huang, W.-H. Peng, T. Chiang, and H.-M. Hang, "Advances in the scalable amendment of H.264/AVC," *IEEE Communications Magazine*, vol. 45, no. 1, pp. 68–76, 2007.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[3] D. Marpe, T. Wiegand, and G. J. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *IEEE Communications Magazine*, vol. 44, no. 8, pp. 134–143, 2006.

[4] M. Krunz, "Bandwidth allocation strategies for transporting variable bit rate video traffic," *IEEE Communications Magazine*, vol. 37, no. 1, pp. 40–46, 1999.

[5] T. V. Lakshman, A. Ortega, and A. R. Reibman, "VBR video: tradeoffs and potentials," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 952–972, 1998.

[6] H. Radha, Y. Chen, K. Parthasarathy, and R. Cohen, "Scalable internet video using MPEG-4," *Signal Processing: Image Communication*, vol. 15, no. 1-2, pp. 95–126, 1999.

[7] D. E. Wrege, E. W. Knightly, H. Zhang, and J. Liebeherr, "Deterministic delay bounds for VBR video in packet-switching networks: fundamental limits and practical tradeoffs," *IEEE/ACM Transactions on Networking*, vol. 4, no. 3, pp. 352–362, 1996.

[8] Z.-L. Zhang, J. Kurose, J. D. Salehi, and D. Towsley, "Smoothing, statistical multiplexing, and call admission control for stored video," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 1148–1166, 1997.

[9] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Resource allocation for multimedia streaming over the internet," *IEEE Transactions on Multimedia*, vol. 3, no. 3, pp. 339–355, 2001.

[10] T. Ahmed, A. Mehaoua, R. Boutaba, and Y. Iraqi, "Adaptive packet video streaming over IP networks: a cross-layer approach," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 385–401, 2005.

[11] W.-C. Feng and J. Rexford, "Performance evaluation of smoothing algorithms for transmitting prerecorded variable-bit-rate video," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 302–313, 1999.

[12] M. Fidler, V. Sander, and W. Klimala, "Traffic shaping in aggregate-based networks: implementation and analysis," *Computer Communications*, vol. 28, no. 3, pp. 274–286, 2005.

[13] T. Kim and M. H. Ammar, "Optimal quality adaptation for scalable encoded video," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 344–356, 2005.

[14] A. R. Reibman and M. T. Sun, *Compressed Video over Networks*, Marcel Dekker, New York, NY, USA, 2000.

[15] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, and J. M. Peha, "Streaming video over the internet: approaches and directions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 282–300, 2001.

[16] M. Wu, S. S. Karande, and H. Radha, "Network-embedded FEC for optimum throughput of multicast packet video," *Signal Processing: Image Communication*, vol. 20, no. 8, pp. 728–742, 2005.

[17] S. Chatziperis, P. Koutsakis, and M. Paterakis, "A new call admission control mechanism for multimedia traffic over next-generation wireless cellular networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 95–112, 2008.

[18] C. Cicconetti, L. Lenzini, E. Mingozzi, and G. Stea, "Design and performance analysis of the real-time HCCA scheduler for IEEE 802.11e WLANs," *Computer Networks*, vol. 51, no. 9, pp. 2311–2325, 2007.

[19] M. Etoh and T. Yoshimura, "Advances in wireless video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 111–122, 2005.

[20] L. Haratcherev, J. Taal, K. Langendoen, R. Lagendijk, and H. Sips, "Optimized video streaming over 802.11 by cross-layer signaling," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 115–121, 2006.

[21] M. Hassan and M. Krunz, "Video streaming over wireless packet networks: an occupancy-based rate adaptation perspective," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 8, pp. 1017–1027, 2007.

[22] S. Khan, Y. Peng, E. Steinbach, M. Sgroi, and W. Kellerer, "Application-driven cross-layer optimization for video streaming over wireless networks," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 122–130, 2006.

[23] F. Yang, Q. Zhang, W. Zhu, and Y.-Q. Zhang, "Bit allocation for scalable video streaming over mobile wireless internet," in *Proceedings of the 23rd Annual Joint Conference of IEEE Computer and Communications Societies (INFOCOM '04)*, vol. 3, pp. 2142–2151, Hong Kong, March 2004.

[24] Q. Zhang, W. Zhu, and Y.-Q. Zhang, "End-to-end QoS for video delivery over wireless internet," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 123–134, 2005.

[25] Y. Cai, A. Natarajan, and J. Wong, "On scheduling of peer-to-peer video services," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 1, pp. 140–145, 2007.

[26] H.-Y. Hsieh and R. Sivakumar, "Accelerating peer-to-peer networks for video streaming using multipoint-to-point communication," *IEEE Communications Magazine*, vol. 42, no. 8, pp. 111–119, 2004.

[27] Y. Huang, Y.-F. Chen, R. Jana, et al., "Capacity analysis of MediaGrid: a P2P IPTV platform for fiber to the node (FTTN) networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 1, pp. 131–139, 2007.

[28] E. Kim and J. C. L. Liu, "Design of HD-quality streaming networks for real-time content distribution," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 2, pp. 392–401, 2006.

[29] B. Li and H. Yin, "Peer-to-peer live video streaming on the internet: issues, existing approaches, and challenges [Peer-to-peer multimedia streaming]," *IEEE Communications Magazine*, vol. 45, no. 6, pp. 94–99, 2007.

[30] J. Liang and K. Nahrstedt, "DagStream: locality aware and failure resilient peer-to-peer streaming," in *Multimedia Computing and Networking*, vol. 6071 of *Proceedings of SPIE*, pp. 1–15, San Jose, Calif, USA, January 2006.

[31] Y. Shen, Z. Liu, S. Panwar, K. Ross, and Y. Wang, "Streaming layered encoded video using peers," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '05)*, pp. 966–969, Amsterdam, The Netherlands, July 2005.

[32] K. Sripanidkulchai, A. Ganjam, B. Maggs, and H. Zhang, "The feasibility of peer-to-peer architectures for large-scale live streaming application," in *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM '04)*, pp. 107–120, Portland, Ore, USA, August-September 2004.

[33] E. Gurses and O. B. Akan, "Multimedia communication in wireless sensor networks," *Annals of Telecommunications*, vol. 60, no. 7-8, pp. 799–827, 2005.

[34] Z. He and D. Wu, "Resource allocation and performance analysis of wireless video sensors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 5, pp. 590–599, 2006.

[35] S. Misra, M. Reisslein, and G. Xue, "A survey of multimedia streaming in wireless sensor networks," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, 2008.

[36] G. Van der Auwera, P. T. David, and M. Reisslein, "Traffic and quality characterization of single-layer video streams encoded with the H.264/MPEG-4 advanced video coding standard and scalable video coding extension," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, part 2, pp. 698–718, 2008.

[37] M. Dai and D. Loguinov, "Wavelet and time-domain modeling of multi-layer VBR video traffic," in *Proceedings of Packet Video Workshop*, Irvine, Calif, USA, December 2004.

[38] M. Dai and D. Loguinov, "Analysis and modeling of MPEG-4 and H.264 multi-layer video traffic," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, vol. 4, pp. 2257–2267, Miami, Fla, USA, March 2005.

[39] T. Gan, K.-K. Ma, and L. Zhang, "Dual-plan bandwidth smoothing for layer-encoded video," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 379–392, 2005.

[40] R. Mangharam, S. Pollin, B. Bougard, et al., "Optimal fixed and scalable energy management for wireless networks," in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, vol. 1, pp. 114–125, Miami, Fla, USA, March 2005.

[41] S. H. Mian, "Analysis of MPEG-4 scalable encoded video," *IEE Proceedings: Communications*, vol. 151, no. 3, pp. 270–279, 2004.

[42] Z. Miao and A. Ortega, "Expected run-time distortion based scheduling for delivery of scalable media," in *Proceedings of the Packet Video Workshop (PVW '02)*, vol. 1, Pittsburg, Pa, USA, April 2002.

[43] S. Nelakuditi, R. R. Harinath, E. Kusmierek, and Z.-L. Zhang, "Providing smoother quality layered video stream," in *Proceedings of the 10th International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV '00)*, Chapel Hill, NC, USA, June 2000.

[44] A. Raghuveer, N. Kang, and D. Du, "Techniques for efficient stream of layered video in heterogeneous client environments," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 1, pp. 245–250, St. Louis, Mo, USA, November 2005.

[45] R. Rejaie, M. Handley, and D. Estrin, "Layered quality adaptation for Internet video streaming," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 12, pp. 2530–2543, 2000.

[46] D. Sarkar, U. K. Sarkar, and W. Zhou, "Bandwidth estimation for multiplexed videos using multinomial model," *Computer Communications*, vol. 30, no. 2, pp. 269–279, 2007.

[47] P. Seeling, M. Reisslein, and B. Kulapala, "Network performance evaluation using frame size and quality traces of singlelayer and two-layer video: a tutorial," *IEEE Communications Surveys and Tutorials*, vol. 6, no. 2, pp. 58–78, 2004.

[48] P. Seeling and M. Reisslein, "The rate variability-distortion (VD) curve of encoded video and its impact on statistical multiplexing," *IEEE Transactions on Broadcasting*, vol. 51, no. 4, pp. 473–492, 2005.

[49] G. Van der Auwera, M. Reisslein, and L. J. Karam, "Video texture and motion based modeling of rate variability-distortion (VD) curves," *IEEE Transactions on Broadcasting*, vol. 53, no. 3, pp. 637–648, 2007.

[50] X. M. Zhang, A. Vetro, Y. Q. Shi, and H. Sun, "Constant quality constrained rate allocation for FGS-coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 2, pp. 121–130, 2003.

[51] L. Zhao, J.-W. Kim, and C.-C. J. Kuo, "Constant quality rate control for streaming MPEG-4 FGS video," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '02)*, vol. 4, pp. 544–547, Scottsdale, Ariz, USA, May 2002.

[52] J.-A. Zhao, B. Li, and I. Ahmad, "Traffic model for layered video: an approach on markovian arrival process," in *Proceedings of the Packet Video Workshop*, Nantes, France, April 2003.

[53] F. Zhijun, Z. Yuanhua, and Z. Daowen, "Kalman optimized model for MPEG-4 VBR sources," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 2, pp. 688–690, 2004.

[54] W. Zhou, D. Sarkar, and S. Ramakrishnan, “Traffic models for MPEG-4 spatial scalable video,” in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM ’05)*, vol. 1, pp. 256–260, St. Louis, Mo, USA, November-December 2005.

[55] G. Van der Auwera, P. T. David, and M. Reisslein, “Traffic characteristics of H.264/AVC variable bit rate video,” *IEEE Communications Magazine*, vol. 46, no. 11, pp. 698–718, 2008.

[56] A. Undheim, Y. Lin, and P. J. Emstad, “Characterization of slice-based H.264/AVC encoded video traffic,” in *Proceedings of the 4th European Conference on Universal Multiservice Networks (ECUMN ’07)*, pp. 263–272, Toulouse, France, February 2007.

[57] P. Li, W. S. Lin, S. Rahardja, X. Lin, X. K. Yang, and Z. G. Li, “Geometrically determining the leaky bucket parameters for video streaming over constant bit-rate channels,” *Signal Processing: Image Communication*, vol. 20, no. 2, pp. 193–204, 2005.

[58] T. Ozcelebi, M. O. Sunay, A. M. Tekalp, and M. R. Civanlar, “Cross-layer optimized rate adaptation and scheduling for multiple-user wireless video streaming,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 760–769, 2007.

[59] T. Ozcelebi, A. M. Tekalp, and M. R. Civanlar, “Delay-distortion optimization for content-adaptive video streaming,” *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 826–836, 2007.

[60] H.-H. Juan, H.-C. Huang, C. Huang, and T. Chiang, “Scalable video streaming over mobile WiMAX,” in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS ’07)*, pp. 3463–3466, New Orleans, La, USA, May 2007.

[61] D. T. Nguyen and J. Ostermann, “Congestion control for scalable video streaming using the scalability extension of H.264/AVC,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 2, pp. 246–253, 2007.

[62] M. Van der Schaar, Y. Andreopoulos, and Z. Hu, “Optimized scalable video streaming over IEEE 802.11 a/e HCCA Wireless networks under delay constraints,” *IEEE Transactions on Mobile Computing*, vol. 5, no. 6, pp. 755–768, 2006.

[63] T. Schierl, K. Gänger, C. Hellge, T. Wiegand, and T. Stockhammer, “SVC-based multisource streaming for robust video transmission in mobile ad hoc networks,” *IEEE Wireless Communications*, vol. 13, no. 5, pp. 96–103, 2006.

[64] T. Schierl, C. Hellge, S. Mirta, K. Gruneberg, and T. Wiegand, “Using H.264/AVC-based scalable video coding (SVC) for real time streaming in wireless IP networks,” in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS ’07)*, pp. 3455–3458, New Orleans, La, USA, May 2007.

[65] D. Marpe, T. Wiegand, and S. Gordon, “H.264/MPEG4-AVC fidelity range extensions: tools, profiles, performance, and application areas,” in *Proceedings of IEEE International Conference on Image Processing (ICIP ’05)*, vol. 1, pp. 593–596, Genoa, Italy, September 2005.

[66] M. Wien, H. Schwarz, and T. Oelbaum, “Performance analysis of SVC,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1194–1203, 2007.

[67] ISO/IEC JTC 1/SC 29/WG 11 N2802, “Information technology-generic coding of audio-visual objects—part 2: visual, final proposed draft amendment 1,” Geneva, Switzerland, July 1999.

[68] O. Marques, P. Auger, and L. M. Mayron, “SimViKi: a tool for the simulation of secure video communication systems,” in *Proceedings of the 4th IASTED International Conference on Communications, Internet, and Information Technology (CIIT ’06)*, St. Thomas, Virgin Islands, USA, December 2006.

[69] M. Ghanbari, *Standard Codecs: Image Compression to Advanced Video Coding*, Institution of Electrical Engineers, London, UK, 2003.

[70] H. Schwarz, D. Marpe, and T. Wiegand, “Analysis of hierarchical B pictures and MCTF,” in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME ’06)*, pp. 1929–1932, Toronto, Canada, July 2006.

[71] H. Radha, M. Van der Schaar, and Y. Chen, “The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP,” *IEEE Transactions on Multimedia*, vol. 3, no. 1, pp. 53–68, 2001.

[72] F. Wu, S. Li, and Y.-Q. Zhang, “A framework for efficient progressive fine granularity scalable video coding,” *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 332–344, 2001.

[73] G. Van der Auwera, P. T. David, M. Reisslein, and L. J. Karam, “Video traffic analysis of H.264 SVC: temporal and spatial scalability,” Tech. Rep., Arizona State University, Tempe, Ariz, USA, September 2007.

[74] P. T. David, G. Van der Auwera, and M. Reisslein, “Video traffic analysis of H.264 SVC: fine granularity scalability,” Tech. Rep., Arizona State University, Tempe, Ariz, USA, September 2007.

[75] M. Reisslein, J. Lassetter, S. Ratman, O. Lotfallah, F. Fitzek, and S. Panchanathan, “Traffic and quality characterization of scalable encoded video: a large-scale trace-based study—part 1: overview and definitions,” Tech. Rep., Arizona State University, Tempe, Ariz, USA, December 2003.

[76] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, “RTP: a transport protocol for real-time applications, STD 64, RFC 3550,” July 2003.

[77] S. Wenger, Y.-K. Wang, and T. Schierl, “RTP payload format for SVC video,” July 2007, http://tools.ietf.org/html/draft-ietf-avt-rtp-svc-02.