Performance of Scalable Video Coding for a TV Broadcast Network with Constant Video Quality and Heterogeneous Receivers

Z. Avramova*, D. De Vleeschauwer**, P. Debevere***, S. Wittevrongel*,
P. Lambert***, R. Van de Walle***, H. Bruneel*
* SMACS Research Group, TELIN, Faculty of Engineering, Ghent University, Ghent, Belgium.
** Bell Labs, Alcatel-Lucent Bell, Antwerp, Belgium.
*** MMLab, ELIS, Faculty of Engineering, IBBT-Ghent University, Ghent, Belgium.
{kayzlat, sw, hb}@ telin.ugent.be, danny.de_vleeschauwer@alcatel-lucent.be, {pedro.debevere, peter.lambert,
rik.vandewalle}@ugent.be

***Abstract*—Video broadcasters currently envision to target a variety of receiving devices of different resolutions. In such a heterogeneous TV network the transport resource consumption is likely to increase as the types and number of receiving devices increase. We estimate the required capacity for an IPTV and mobile TV network taking into account parameters as currently proposed in standardization bodies. The consumed transport capacity fluctuates over time for two reasons. First, as we target constant video quality as much as possible, the video is encoded in variable bit rate. In order to characterize the fluctuations of the bit rate associated with one channel we have encoded a representative set of video clips. Second, the user behavior causes the transport capacity demand to fluctuate because we consider a multicast-based transport system where only the requested channels are transported. We compare two modes for transporting TV channels in multiple resolutions: a mode based on the Scalable Video Coding (SVC) where all video resolutions are embedded in one stream and a simulcast (SIM) mode where all resolutions are offered in parallel. We evaluate in some realistic examples which of the two effects is dominant, i.e., the bit rate penalty of SVC or the fact that SIM transports the video streams in parallel rather than embedding them in one stream as SVC does.**

I. INTRODUCTION

With the roll-out of High Definition TeleVision (HDTV), traditional broadcast TeleVision (TV) operators are facing growing bandwidth demand. Since there are still a lot of legacy TV sets (and, in future, several HDTV formats, e.g., 720p and 1080p), video will need to be delivered to a variety of receiving devices, hence in multiple resolutions. At the same time, mobile TV is gaining momentum (mainly through broadcast technologies), despite the problem of diversity of technologies through which it is offered and despite regulatory issues. Also in this context, the content provider will have to target a heterogeneity of devices and hence will have to offer the same channel in various resolutions (e.g., a VGA (Video Graphics Array) one for an in-car display and a QVGA (Quarter VGA) one for a mobile handheld device).

In such a heterogeneous environment, the limited available network resources on the part of the network shared by all users might become a bottleneck. It is still an open research area to find viable models to estimate the bandwidth demand required to deliver a set (often referred to as a *bouquet*) of TV channels in different resolutions or quality versions. To offer this bouquet of channels multicast is used to economize on the required transport capacity on the part of the network common for a group of users. So, all multicast trees (associated with all channels) are (dynamically) built such that a node only supports a channel that is requested by at least one of the users served by the node. Some mobile operators still offer TV content through unicast, but this is clearly inefficient leading to increase in required bandwidth proportional to the growing subscriber population.

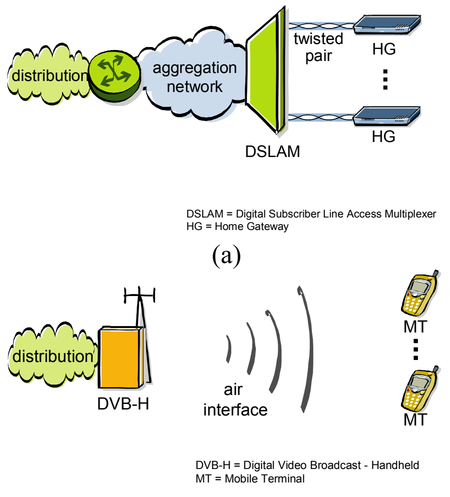
Most of the digital TV networks and mobile TV trials or commercially operating systems today deliver video encoded in H.264/AVC (MPEG-4 Part 10) [1] (e.g., those documented in [2]). To offer channels in multiple resolutions one method is to simulcast the different resolution versions in parallel. An alternative method is to use the scalable video coding (SVC) amendment to the H.264 codec [1]. Some consider that delivering the channels in multiple resolutions with SVC could help to alleviate the bandwidth limitation problem: the fact that with SVC all resolutions are embedded in one bit stream, rather than transported in parallel, tends to benefit SVC. However, encoding in SVC comes in general with a bit rate penalty, such that any resolution (except for the base resolution) requires a higher bit rate than if it were encoded in AVC. So, in a scenario where not all resolutions of all channels are requested all of the time (e.g., in an access network environment where many channels are served to a rather small community of users), simulcast might be favored, because SVC needs to deliver the highest resolution version that is requested (in effect wasting the overhead needed for all resolutions below it that are not requested), while with simulcast transport only the requested resolutions need to be provided. To assess in which direction the balance tilts in such scenarios, user models are needed to estimate the bandwidth demand in a TV broadcast (aggregation) network with heterogeneous receivers. In [5] a simulcast transport scheme was compared to a scalable video transport scheme when the video is encoded and streamed (via multicast) with constant bit rate (CBR) under various user models.

It is more realistic to assume that a broadcast operator will target to provide constant video quality implying non-constant video bit rate. In [7] capacity demand models were introduced which take into account the non-constant bit rate character of the multi-resolution video. These models require statistical description of the bit rate fluctuations expected when encoding video in various resolutions. It is precisely on this topic that the present paper focuses on, i.e., better characterization of these bit rate fluctuations and its implication to resource provisioning.

This paper is organized as follows. In the next section we discuss the two network scenarios introduced in [4] in which SVC was proposed as a viable technology. In Section III we show results for fourteen videos encoded in two resolutions in AVC and SVC when constant quality is aimed for and characterize the bit rate fluctuations and their statistical representation. Section IV discusses the methodology to assess the transport capacity in the part of the network shared by all users. In Section V we apply these models to estimate the capacity demand in a couple of realistic examples. Finally, Section VI draws the main conclusions.

II. NETWORK TOPOLOGIES

We consider the transport capacity on the part of the network shared by a user community. The document of [4] describes an evolution path of the traditional broadcast service to a service enabling support of HDTV (referred to as "fixed TV" below) and to mobile TV service in a scenario targeting a variety of devices. The fixed TV case is illustrated in a DSL (Digital Subscriber Line) context in Figure 1(a), while the mobile TV case is illustrated in Figure 1(b) in the mobile broadcast technology DVB-H (Digital Video Broadcasting-Handheld) context. In particular, in the fixed TV scenario we consider the required total transport capacity on the aggregation network (more precisely on the feeder links towards the DSLAM (DSL Access Multiplexer)), while in the mobile TV scenario we consider the transport capacity over the air interface. For the mobile TV scenario, [4] proposes that two resolutions are supported, QVGA (i.e., 240 lines per frame with 320 pixels per line) and VGA (i.e., 480 lines per frame with 640 pixels per line), while for the fixed TV scenario it proposes the resolutions SDTV (under the form of digital PAL (i.e., 576 lines per frame with 720 pixels per line) or VGA) and HDTV (under the form of 720p (i.e., 720 lines per frame with 1280 pixels per line) or 1080p (i.e., 1080 lines per frame with 1920 pixels per line)), all at 25 frames per second (in Europe).



(a)

(b)

Figure 1. Two network topologies: (a) an IPTV scenario, and (b) a mobile TV scenario

III. VIDEO BIT RATE FLUCTUATIONS

A. *Aiming for Constant Quality*

In [4] a constant quality is aimed for, but at the same time it is recognized the huge diversity in the complexity of the scenes to be encoded. Since targeting the same quality for all scenes would lead to a too large a bit rate fluctuation, it is proposed to set a maximum bit rate (taking the resulting quality degradation as an imminent burden). For practical reasons also a minimum bit rate is set (in which case the quality can be larger than the targeted one).

Similarly as in [6], we assume that the encoder has a bit rate shaper (often abbreviated to "rate shaper"). Such a rate shaper consists of a buffer and an algorithm to tune the quantization parameter Q_p and an algorithm to release the bits (under the form of packets) from the buffer on the network. This buffer is large enough to absorb and smoothen (e.g., over a Group of Pictures (GoP)) the peaks in the bit rate associated with the succession of I-, P- and B-slices, but is too small (in order to avoid a too large delay) to cope with the bit rate fluctuations due to changing scene complexity. The combination of the algorithm to set the quantization parameter Q_p and the algorithm to release the bits on the network keeps the buffer from under- and overflowing. If a complex scene persists, (the algorithms of) the rate shaper will increase the quantization parameter Q_p decreasing the bit rate and the video quality or send the bits at a higher rate on the network. Conversely, if an easy scene perseveres, the rate shaper will decrease the quantization parameter Q_p or send the bits at a lower rate on the network. If the buffer does not risk to under- or overflow, the quantization parameter Q_p is set such that a constant quality is achieved. Rate shapers for SVC have not had a lot of attention in the scientific literature. We assume that the rate shaper is such that the network solely sees the fluctuations due to the alteration of scenes. This has as consequence that the quantization parameter Q_p and the bit rate vary (slowly) over time at the pace with which scenes succeed one another.

B. *Encoding a Set of Representative Video Clips*

In order to assess the extent of the bit rate fluctuations due to the succession of different scene types, we have encoded a representative set of fourteen scenes, i.e., short clips of 257 frames each. Every sequence can be categorized into one of the following classes: "sports", "news", "action", "nature", "cartoon" and "general". The "general" class contains sequences that do not correspond to any of the previous mentioned classes. These categories are typical for the content that is commonly presented on broadcast TV.

Starting from the highest resolution (720p), the sequences having a lower resolution were generated using the "DownConvertStatic" tool which is part of the Joint Scalable Video Model (JSVM) software [3]. The default down-sampling method was used, which is based on the filter set proposed in [9]. Also cropping was applied in order to avoid deformation of the image because of the different aspect ratios of the selected resolutions.

To encode the fourteen video clips (in various resolutions), we took as guidelines the scenarios proposed in [4]. In the mobile TV scenario it is recommended to encode the two versions as follows: the QVGA resolution is to be encoded in the bit rate range [250 kb/s, 400 kb/s], while the VGA version in the range [750 kb/s, 1500 kb/s]. In the fixed TV distribution context, the two video resolutions are suggested to be encoded as follows: the SDTV (VGA) version in the range [1.5 Mb/s, 3 Mb/s] and the HDTV (720p) version in the range [6 Mb/s, 10 Mb/s]. In order to maintain the same aspect ratio in the base layer as in the total stream (720p) for the fixed TV scenario, a resolution of 640x360 pixels was selected. As the JSVM reference encoder only works with resolutions which are a multiple of 16, two black banners of 4 pixels were added, resulting in a sequence with a resolution of 640x368 pixels.

For the encoding of the streams, the reference encoder provided as part of (version 9.12 of) the JSVM software was used. A PSNR-Y (Peak Signal-to-Noise Ratio of the luminance component Y) of 40 dB is aimed for. In order to tune the quantization parameter Q_p such that an overall PSNR-Y of 40 dB was approximated as closely as possible without violating the ranges specified by [4], the “FixedQPEncoderStatic” tool was used which is also a part of the JSVM software. This tool iteratively encodes an entire sequence until a certain predefined PSNR-Y or bit rate is achieved (within a predefined tolerance interval).

The overall PSNR-Y associated with a clip is the average value of the PSNR-Y of all images in the clip. The bit rate associated with a clip corresponds to the number of bits in the encoded file divided by the clip duration of 10.28 s (257 frames each of 40ms duration).

The GoP size was set equal to 16 and the Instantaneous Decoding Refresh (IDR) period was set to 32. We used Context-Adaptive Binary Arithmetic Coding (CABAC) for the entropy coding, as CABAC has a higher compression efficiency than Content-Adaptive Variable Length Coding (CAVLC) [8]. In this way, the encoded sequences conform to the Scalable High Profile.

The bit rate ranges for the mobile TV scenario are such that sometimes the target PSNR of 40 dB cannot be achieved within this limit (see e.g., sequence “nature”), while for the HDTV scenario by encoding around the lower bit rate boundary of 1.5 Mb/s or 6 Mb/s, the obtained PSNR is even higher than 40 dB. Delivering a resolution in either of the transport modes provides comparable video quality.

As the PSNR-Y is used in the “FixedQPEncoderStatic” tool, this is currently the only viable method to encode video clips in constant quality. We could have used a method based on the (subjective) MOS (Mean Opinion Score), but this is clearly impractical: too many clips would have needed to be appraised by a test audience (fourteen video clips in four resolutions encoded with many quantization parameters). Objective tools that correlate slightly better to the MOS than the PSNR-Y does, are known, but are not part of the “FixedQPEncoderStatic” tool.

TABLE I.
BIT RATES FOR THE MOBILE TV SCENARIO (a) SIM, (b) SVC

(a)

	QVGA		VGA	
Sequence	**bit rate $R_{SIM,1}$ [kb/s]**	**PSNR-Y [dB]**	**bit rate $R_{SIM,2}$ [kb/s]**	**PSNR-Y [dB]**
action_a	399.23	38.04	1114.59	40.06
action_b	304.66	39.76	652.53	40.01
cartoon_a	397.52	40.11	923.74	40.01
cartoon_b	405.72	39.67	978.39	40.23
nature	400.09	33.46	1482.30	35.05
news	252.79	49.40	549.30	48.17
Sport_a	398.37	34.61	1387.95	38.61
Sport_b	250.27	40.47	619.83	40.30
Sport_c	361.08	40.12	785.22	40.05
Sport_d	247.29	42.73	610.75	42.11
General_a	400.66	36.44	1368.29	39.33
General_b	393.12	40.38	1122.96	40.12
General_c	253.00	41.40	645.69	41.06
General_d	252.07	47.65	564.50	45.64

(b)

	QVGA		VGA	
Sequence	**bit rate $R_{SVC,1}$ [kb/s]**	**PSNR-Y [dB]**	**bit rate $R_{SVC,2}$ [kb/s]**	**PSNR-Y [dB]**
action_a	399.23	38.04	1230.53	40.13
action_b	304.66	39.76	699.31	39.97
cartoon_a	397.52	40.11	1166.26	40.00
cartoon_b	405.72	39.67	1067.28	40.16
nature	400.09	33.46	1526.64	34.99
news	252.79	49.40	794.82	48.27
Sport_a	398.37	34.61	1512.90	38.70
Sport_b	250.27	40.47	760.60	40.28
Sport_c	361.08	40.12	879.35	40.04
Sport_d	247.29	42.73	755.57	42.19
General_a	400.66	36.44	1491.64	39.29
General_b	393.12	40.38	1294.10	40.18
General_c	253.00	41.40	757.19	41.14
General_d	252.07	47.65	747.34	45.54

Although a different quality metric could lead to different bit rate fluctuations, we assume that the fluctuations we observe are representative.

Table I(a) and Table I(b) give the bit rate vectors ($R_{SIM,1}$, $R_{SIM,2}$) and ($R_{SVC,1}$, $R_{SVC,2}$) in the mobile TV scenario for each of the fourteen clips (with the associated quality); the subscripts SIM and SVC stand for the simulcast and SVC transport scheme respectively, while the indices 1 and 2 pertain to the base and higher resolution respectively. Table II presents similar information for the fixed TV scenario.

Notice that $R_{SIM,1} = R_{SVC,1}$, because the base layer of SVC is AVC compatible. Remark as well that although the highest resolution in the mobile TV case and the lowest presented resolution in the fixed TV case are both VGA, none of the bit rates are the same. The reason for this is that for SVC, in the fixed TV case, it is just the base layer, while in the mobile case it embeds a QVGA resolution. Also, in the fixed TV case, only a resolution of 640x368 is encoded (in order to maintain the aspect ratio), the remaining part of the picture represents black banners.

TABLE II.
BIT RATES FOR THE FIXED TV SCENARIO (a) SIM, (b) SVC

(a)

	VGA		**720p**	
Sequence	**bit rate $R_{SIM,1}$ [kb/s]**	**PSNR-Y [dB]**	**bit rate $R_{SIM,2}$ [kb/s]**	**PSNR-Y [dB]**
action_a	1511.75	41.72	4673.66	43.44
action_b	1524.24	44.88	4800.63	44.99
cartoon_a	1519.43	44.73	5071.87	46.62
cartoon_b	1505.28	42.72	4787.83	43.49
nature	2958.20	39.17	7219.92	39.33
news	1532.97	50.18	4652.23	48.14
Sport_a	1908.32	40.42	4620.08	41.83
Sport_b	1514.66	43.33	4700.45	43.17
Sport_c	1525.32	43.70	4597.26	44.32
Sport_d	1516.11	46.58	4694.99	43.98
General_a	1504.02	39.76	4812.51	42.21
General_b	1512.69	42.92	4898.64	40.92
General_c	1500.29	44.38	4508.15	42.55
General_d	1504.87	48.22	4747.56	42.60

(b)

	VGA		**720p**	
Sequence	**bit rate $R_{SVC,1}$ [kb/s]**	**PSNR-Y [dB]**	**bit rate $R_{SVC,2}$ [kb/s]**	**PSNR-Y [dB]**
action_a	1511.75	41.72	6016.63	43.52
action_b	1524.24	44.88	6194.60	45.10
cartoon_a	1519.43	44.73	6095.76	46.53
cartoon_b	1505.28	42.72	6178.45	43.58
nature	2958.20	39.17	9905.84	39.39
news	1532.97	50.18	6002.40	48.31
Sport_a	1908.32	40.42	6011.37	41.75
Sport_b	1514.66	43.33	6050.91	43.14
Sport_c	1525.32	43.70	6105.16	44.38
Sport_d	1516.11	46.58	6124.58	43.97
General_a	1504.02	39.76	6001.92	42.12
General_b	1512.69	42.92	5969.71	40.87
General_c	1500.29	44.38	6002.59	42.65
General_d	1504.87	48.22	6123.55	42.66

Furthermore, the target bit rates are different. For SIM this is due to the fact that the bit rate ranges for the highest resolution in mobile TV and the lowest resolution in fixed TV ([750 kb/s, 1500 kb/s] and [1.5 Mb/s, 3 Mb/s] respectively) do not overlap.

Table I and II provides us fourteen instances of the random bit rate vectors for both the SVC and SIM transport scheme for the mobile and fixed scenario.

C. Statistical Description

Fourteen instances is too low a number to characterize the joint histogram of the two-dimensional bit rate vectors. Therefore we assume that the bit rate vector is governed by a bivariate Gaussian distribution. Consequently the (vector with) average bit rates and the covariance matrix associated with this bit rate vector suffice to capture the statistical properties of this vector. These two central moments can be easily determined based on these fourteen instances. The resulting vectors with average bit rates and the covariance matrices are given in Table III. Notice that in the highest resolution the average bit rate penalty of SVC in the mobile TV scenario is 15%, while in the fixed TV scenario, it is as high as 29% (to achieve the same PSNR-Y).

TABLE III.
VECTORS OF AVERAGE BIT RATES (IN kb/s) AND COVARIANCE MATRICES FOR SIM AND SVC

	mobile TV			
	SIM		SVC	
	QVGA	VGA	QVGA	VGA
	$R_{SIM,1}$	$R_{SIM,2}$	$R_{SVC,1}$	$R_{SVC,2}$
average	337	915	337	1049
covariance matrix	4697	18864	4697	17877
	18864	102336	17877	93392

	fixed TV			
	SIM		SVC	
	VGA	720p	VGA	720p
	$R_{SIM,1}$	$R_{SIM,2}$	$R_{SVC,1}$	$R_{SVC,2}$
average	1646	4913	1646	6342
covariance matrix	142849	229711	142849	358370
	229711	427177	358370	981846

IV. CAPACITY ESTIMATION

A. Required Capacity

We aim at estimating the required bandwidth on the shared part of the network in a TV broadcast network.

A bouquet of K TV channels is offered to an audience of N subscribers. The content of the channels is encoded in L=2 different resolutions (either as two versions in parallel (SIM) or embedded in one flow (SVC)).

Despite the occasionally large bit rate penalty for SVC, it is clear from the tables that in a situation where all resolutions need to be provided all of the time, the SVC transport mode is more economical than the SIM transport mode. Indeed, the highest resolution in SVC, embedding the lower resolution, is for all clips (and hence also for the average value) smaller than the sum of the bit rates associated with the lower and higher resolution encoded in AVC.

In an environment (typical for an access network) where a large bouquet of K channels is served to a relatively small audience of N users, it is wasteful to transport every channel in every resolution. The multicast technology allows (via the exchange of subscription and request messages) to prune the multicast tree such that each multicast node only needs to branch off a channel if it is requested by at least one user in the group of users it serves. In such a system (referred to below as the "pruned multicast tree") the aggregate capacity demand C_{SIM} or C_{SVC} for the SIM or SVC transport mode respectively at a randomly chosen moment in time is given by

$$C_{SIM} = \sum_{k=1}^{K} \sum_{l=1}^{2} r_{SIM,k,l} \cdot 1_{\{n_{k,l}>0\}} , \qquad (1)$$

$$C_{SVC} = \sum_{k=1}^{K} \left(r_{SVC,k,1} \cdot 1_{\{n_{k,2}=0, n_{k,1}>0\}} + r_{SVC,k,2} \cdot 1_{\{n_{k,2}>0\}} \right) . \qquad (2)$$

Here $n_{k,l}$ is the number of users that are tuned in to resolution l (assuming values 1 for the base or 2 for the higher resolution) of channel k, $r_{SIM,k,l}$ and $r_{SVC,k,l}$ are the bit rates associated with channel k in resolution l for the SIM and SVC transport mode respectively, and $1_{\{A\}}$ is the indicator function (that assumes the value 1 if the condition A is true and 0 otherwise). As $r_{SIM,k,l}$, $r_{SVC,k,l}$ and $n_{k,l}$ are subject to statistical fluctuations, C_{SIM} and C_{SVC} are fluctuating too.

The indicator function in eq. (1) points out that only those resolutions of the channels that are actually watched by at least one customer in the community, contribute to the aggregate bit rate demand. Similarly, the indicator function in eq. (2) designates that of each channel only the highest resolution that is requested contributes to the aggregate bit rate demand.

The vectors $(r_{SIM,k,1}, r_{SIM,k,2})$ and $(r_{SVC,k,1}, r_{SVC,k,2})$ are instances of the random vectors $(R_{SIM,1}, R_{SIM,2})$ and $(R_{SVC,1}, R_{SVC,2})$ respectively, the fluctuations of which were described in the previous section. The components of this vector are statistically dependent (as witnessed by the fact that the off-diagonal elements in the covariance matrix are non-zero). This is due to the fact that it is likely that for content that is difficult to compress, e.g., a scene with a lot of motion or with a complex texture (or both), all resolutions of that channel will have a higher bit rate than average and vice-versa for an easy scene. Additionally, we assume that the random vectors associated with the different channels (i.e., with different index k) are statistically independent but can be described by the same statistical model.

The statistical fluctuations of $n_{k,l}$ depend on the choices made by the N users of the served audience. Each user can be active or passive (with probability a and 1-a respectively, which we refer to as the "activity grade") and when active he or she can select any particular channel in any resolution (with a given probability). Notice that the variables $n_{k,l}$ are statistically dependent, due to the fact that the population is finite and a user can only watch one channel at any moment in time. Which channel an active user chooses is governed by a Zipf popularity probability distribution [10] (which is the most common model in these circumstances). For the choice of the resolutions two models were introduced in [5]. Here we just assume that irrespective of the choice of the channel, the lowest resolution is chosen with a probability p, while the highest resolution is chosen with a probability 1-p.

Finally we assume that the variables $r_{SIM,k,l}$, $r_{SVC,k,l}$ and $n_{k,l}$ are statistically independent. In layman's terms this means that the user's choice to select a channel (and a resolution) is independent of the broadcaster's choice to air content of a particular type.

Notice that in case all resolutions of all channels (also referred to as a "satiated multicast tree") were to be provisioned, the required capacity C'_{SIM} and C'_{SVC} for the SIM and SVC transport mode respectively would be

$$C'_{SIM} = \sum_{k=1}^{K} \sum_{l=1}^{2} r_{SIM,k,l} \quad , \tag{3}$$

$$C'_{SVC} = \sum_{k=1}^{K} r_{SVC,k,2} \quad . \tag{4}$$

In Section V we will compare this capacity when all channels in all resolutions are provisioned to the capacity required with pruned multicast trees given by eq. (1) and (2).

It is clear from eq. (3) and (4) (and the assumption that the bit rate vectors are governed by a bivariate Gaussian distribution) that C'_{SIM} and C'_{SVC} are Gaussian variables for which the average and the variance are straightforward to calculate.

The momentary aggregate capacity demands C_{SIM} and C_{SVC} of eq. (1) and (2) and C'_{SIM} and C'_{SVC} of eq. (3) and (4) are subject to fluctuations. The network will be designed such, i.e., a capacity C will be foreseen in the network, such that the actual capacity demand C_{SIM} and C_{SVC} will hardly ever exceed this provisioned capacity C, i.e., with a very small probability of unavailability, P_{unav}. Stated under this form it is obvious that the Complementary Cumulative Distribution Function (CCDF) (which is often also referred to as the Tail Distribution Function (TDF)) of the capacity (be it C_{SIM}, C_{SVC}, C'_{SIM} or C'_{SVC}) is needed to properly design the network.

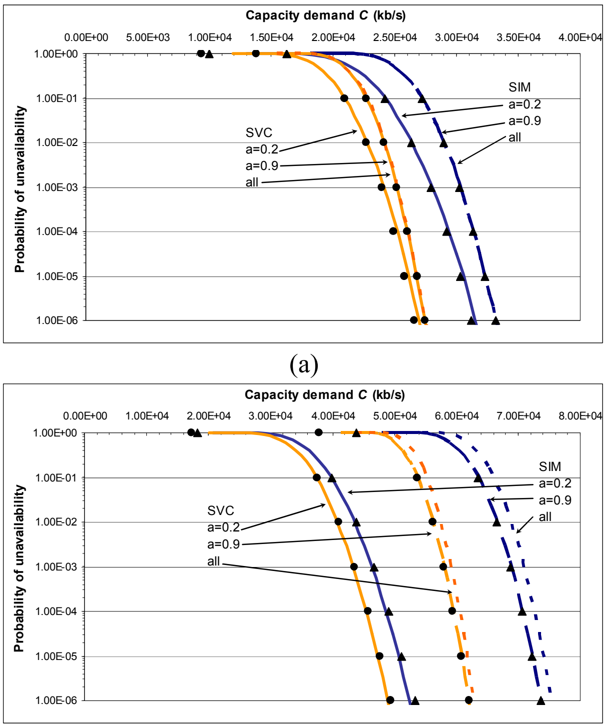
B. Theoretical Model

In the previous paragraph we argued that the momentary aggregate capacity demand in case of a satiated multicast tree (C'_{SIM} or C'_{SVC} of eq. (3) and (4) respectively) is easily characterized as it is a sum (of a fixed number of terms) of Gaussian variables. The momentary aggregate capacity demand (C_{SIM} or C_{SVC} of eq. (1) and (2) respectively) in case a pruned multicast tree is used, is less obvious to characterize. Strictly speaking, even though the aggregate capacity demand (C_{SIM} or C_{SVC}) is a sum of Gaussian variables, it is not a Gaussian variable itself, because the number of variables in the sum is subject to statistical fluctuations as well. Nevertheless we will *assume* that this capacity demand is Gaussian, calculate its average and variance and validate the assumption via simulations.

In [7] we derived formulae to calculate the average value and variance of C_{SIM} and C_{SVC} and we proved that these moments only depend on the average and the covariance matrix of the bit rate vector (both determined in the previous section). In the following Section V we use the tools of [7] to determine the CCDF and hence to design the network based on our own video encoding data.

C. Simulation Model

To validate this Gaussian assumption we wrote an event-driven C-based simulator, which can simulate both the SIM and SVC transport modes. This simulator generates a number of realizations of the random variable C_{SIM} or C_{SVC}. For every realization, a user is either tuned in to a given channel k and resolution l or is inactive (governed by the user model). This determines $n_{k,l}$. If the value of $n_{k,l}$ justifies it (i.e., if the indicator function in eq. (1) or (2) assumes the value 1), and according to the selected transport scheme, a bit rate ($r_{SIM,k,l}$ or $r_{SVC,k,l}$ respectively) is added to the aggregate capacity. The momentary streamed video bit rate of a given resolution is randomly selected, governed by a bivariate Gaussian law of which the statistics are described in Section III. This provides one instance of C_{SIM} or C_{SVC} respectively.



(a)

(b)

Figure 2. CCDF of the capacity demand for a mobile TV scenario for (a) K=20 channels and (b) K=50 channels. Lines pertain to theoretical results with full lines for a=0.2, dashed lines for a=0.9, and dotted lines (labeled "all") to the case with a satiated multicast tree; light colored lines pertain to SVC, dark colored lines - to SIM. Dot marks (for SVC) or triangle marks (for SIM) pertain to simulation results.

TABLE IV.
CAPACITY DEMAND AT $P_{UNAV}=10^{-4}$ FOR A MOBILE TV SCENARIO: COMPARISON OF THE SIM AND SVC TRANSPORT SCHEMES

Capacity demand in kb/s at $P_{unav}=10^{-4}$ with		mobile TV (N=500)	
		20 channels	50 channels
a=0.2	SIM	2.94E+04	4.87E+04
	SVC	2.52E+04	4.55E+04
a=0.9	SIM	3.14E+04	7.07E+04
	SVC	2.61E+04	5.95E+04
all	SIM	3.14E+04	7.26E+04
	SVC	2.61E+04	6.05E+04
relative gain with respect to "all" scenario	SIM, a=0.2	6.35%	32.88%
	SVC, a=0.2	3.29%	24.84%
	SIM, a=0.9	0.01%	2.61%
	SVC, a=0.9	0.01%	1.58%

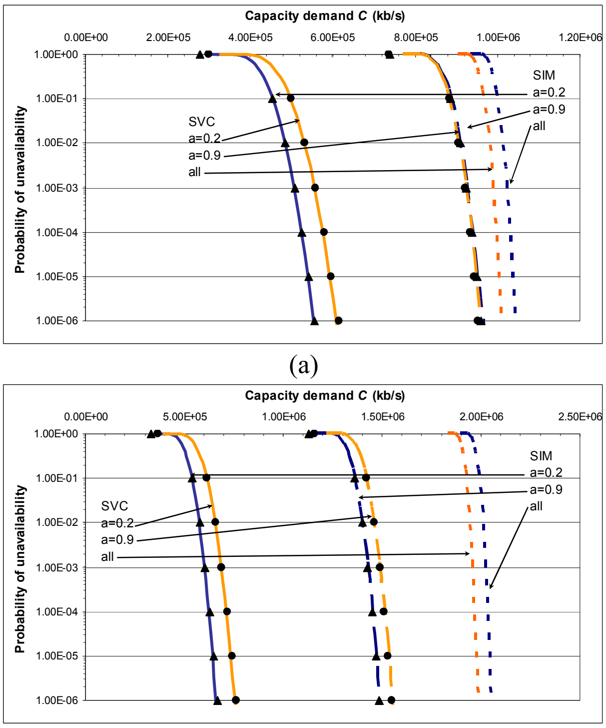
This process is repeated over a sufficiently large number of realizations, depending on the required accuracy to determine the CCDF of the variable C_{SIM} or C_{SVC}.

V. RESULTS

A. *Aggregate Capacity Demand Fluctuations*

1) *Mobile TV*

We design a mobile TV scenario with K=20 channels (representative of actual such systems [2]) and K=50 channels (representative of future upgrade of such systems). The number of users is N=500 with equal probability of choosing one of the two resolutions (p=0.5). Two values for the activity grade are considered: a low activity grade a=0.2 and a high one a=0.9. We take a Zipf distribution with parameter 0.6, as in [5] and [7], to model the popularity of the channels.



(a)

(b)

Figure 3. CCDF of the capacity demand for a fixed TV scenario for (a) K=150 channels and (b) K=300 channels. Lines pertain to theoretical results with full lines for a=0.2, dashed lines for a=0.9, and dotted lines (labeled "all") to the case with a satiated multicast tree; light colored lines pertain to SVC, dark colored lines - to SIM. Dot marks (for SVC) or triangle marks (for SIM) pertain to simulation results.

TABLE V.
CAPACITY DEMAND AT $P_{UNAV}=10^{-4}$ FOR A FIXED TV SCENARIO: COMPARISON OF THE SIM AND SVC TRANSPORT SCHEMES

Capacity demand in kb/s at $P_{unav}=10^{-4}$ with		fixed TV (N=1000)	
		150 channels	300 channels
a=0.2	SIM	5.24E+05	6.25E+05
	SVC	5.77E+05	7.13E+05
a=0.9	SIM	9.41E+05	1.45E+06
	SVC	9.37E+05	1.51E+06
all	SIM	1.03E+06	2.03E+06
	SVC	9.96E+05	1.97E+06
relative gain with respect to "all" scenario	SIM, a=0.2	49.10%	69.26%
	SVC, a=0.2	42.12%	63.73%
	SIM, a=0.9	8.68%	28.68%
	SVC, a=0.9	6.00%	23.12%

In Figure 2 we have plotted graphs of the CCDFs of the aggregate capacity demand obtained by the theoretical model (based on a Gaussian assumption) and the simulation approach. There is a good correspondence between theoretical results (lines) and simulation results (dots or triangle marks). Note that the lines for the case of an activity grade of a=0.9 in the case a pruned multicast tree is used and the case with the satiated multicast tree (labeled "all") overlap in Figure 2(a). The corresponding minimum bandwidth resources under the considered cases, providing probability of unavailability of $P_{unav}=10^{-4}$, are shown in Table IV. The relative capacity gain with respect to the "all" scenario is also displayed in the table.

2) *Fixed TV*

For a fixed TV broadcast system, we again take a Zipf popularity distribution with parameter 0.6 and again consider the same two activity grades, but the system is designed for *K*=150 and *K*=300 channels and for a user population of *N*=1000 users, again with equal probability of choosing one of the two resolutions (*p*=0.5).

Figure 3 shows the CCDFs of the aggregate capacity demand obtained by the theoretical model and simulations. Again there is a good correspondence between theoretical and simulation results. The corresponding minimum bandwidth resources under the considered cases, providing probability of unavailability of $P_{unav}=10^{-4}$, are shown in Table V. The relative capacity gain with respect to the "all" scenario is also displayed in the table.

B. *Required Capacity and Comparison of the Performance of SIM and SVC Transport Strategies*

A network operator wants to design its network for a given (sufficiently low) probability of unavailability, P_{unav}, the probability that the capacity demand exceeds the available bandwidth. In Table IV and Table V, a numerical comparison is made between the capacity demand in case the operator offers the channels in multiple resolutions via a SIM or an SVC transport scheme (with a pruned multicast tree) and the capacity needed for a satiated multicast tree case, at a probability of unavailability of $P_{unav}=10^{-4}$.

Considering Table IV pertaining to the mobile TV scenario, the SVC transport strategy is always more advantageous than SIM. Also, the capacity required for the satiated multicast tree is a lot higher than the one needed with the pruned multicast tree when the customers activity is low (*a*=0.2) and very close to it when *a* is high (*a*=0.9).

However, in the fixed TV scenario (see Table V), SVC only outperforms SIM in the case with the satiated multicast tree and in the case with the pruned multicast tree if the activity grade is high. If the activity grade is low (and if there are numerous channels with non-negligible popularity), it is likely that each user tunes in to a different channel. In that case SVC pays a bit rate penalty without benefit, because it is likely that the lower resolution is not consumed when the higher one is requested. Only at high activity grade SVC starts to outperform simulcast even if the penalty bit rate of SVC is 29%.

In Table IV and Table V, the relative capacity gain of using a pruned multicast tree rather than the satiated one (referred to with the label "all" in the table) is also shown. For any value of the activity grade the capacity required for the satiated multicast tree is higher than the one needed with the pruned multicast tree. With a tailored multicast tree according to the feedback from users, a considerable amount of transport capacity is saved in case there are many channels (and a limited amount of users or not very active customers).

VI. CONCLUSIONS

We considered the case of a fixed TV (IPTV) system and a mobile TV system where video content is offered (in streaming mode) in two resolutions. We assessed the performance of an SVC transport scheme in a multicast-enabled network and we compared its performance to a simulcast scheme with respect to resource demand. For a fair comparison, we used our own video encoding data, taking into account the proposed network scenarios in [4]. We demonstrated that there is no straightforward answer to the question which of both schemes is more efficient in terms of required capacity. In our mobile TV system, SVC is the winning strategy while in the fixed TV system, in general simulcast is the more efficient strategy, mainly because of the high bit rate penalty for SVC. However, even in the latter case, when the user population is very active, SVC starts to outperform simulcast since most of the channels are requested in all resolutions.

Whether or not SVC will yield resource savings depends on the concrete network parameters and no universal conclusions (e.g., for mobile and fixed TV networks) can be drawn. Moreover, other constraints need to be accounted for before the transport strategy is selected. For example, SVC may lead to bandwidth reduction, but incurs encoding complexity and faster battery drain on a mobile (thin) end device.

A more ambitious step would be to verify the proposed models in an experimental set-up with real videos, which we leave for future work.

ACKNOWLEDGMENT

This work was carried out in the framework of the Q-Match project sponsored by the Flemish Institute for the Promotion of Scientific and Technological Research in the Industry (IWT).

REFERENCES

[1] "Advanced Video Coding for Generic Audiovisual Services" (Vers.1: May 2003, Vers.2: May 2004, Vers.3: Mar. 2005, Vers.4: Sept. 2005, Vers.5 and Vers.6: June 2006, Vers.7: Apr. 2007, Vers.8 (including SVC extension): Nov. 2007), ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC).

[2] "Description of DVB-H mobile TV trials and services", http://www.dvb-h-online.org/services.htm.

[3] "Joint Scalable Video Model (JSVM) v9_12", available at http://gracon.ient.rwth-aachen.de/cvs/jvt.

[4] "Proposal to DVB for study work into Scalable Video Coding for IPTV", DVB-CM-AVC Document CM-AVC0128, November 2007.

[5] Z. Avramova, D. De Vleeschauwer, K. Spaey, S. Wittevrongel, H. Bruneel, C. Blondia, "Comparison of Simulcast and Scalable Video Coding in Terms of Required Capacity in IPTV Network", in Proceedings of the 16th International Packet Video Workshop, PV07, Lausanne, Switzerland, Nov. 2007.

[6] Z. Avramova, D. De Vleeschauwer, K. Laevens, S. Wittevrongel, H. Bruneel, "Modelling H.264/AVC VBR Video Traffic: Comparison of a Markov and a Self-Similar Source Model", Telecommunication Systems, Vol. 39, No. 2, pp. 91-102, June 2008.

[7] Z. Avramova, D. De Vleeschauwer, S. Wittevrongel, H. Bruneel, "Dimensioning of a Multi-rate Network Transporting Variable Bit Rate TV Channels", Accepted for the International Conference on Communications "QoS and Modelling Symposium" (ICC09), Dresden (Germany), June 14 - 18, 2009.

[8] D. Marpe, H. Schwarz, T. Wiegand, "Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, no. 7, pp. 620-636, 2003.

[9] S. Sun, J. Reichel. "AHG Report on Spatial Scalability Resampling", Report JVT-R006 of the 18th Meeting of the Joint Video Team, Bangkok, Thailand, 14-20 January, 2006.

[10] J. Zipf, Selective Studies and the Principle of Relative Frequency in Language, 1932.