# Low-delay Peer-to-Peer Streaming using Scalable Video Coding

Pierpaolo Baccichet*, Thomas Schierl†, Thomas Wiegand†, Bernd Girod‡
*Max Planck Center for Visual Computing & Communication, 94305 Stanford CA, USA
†Fraunhofer Institute for Telecommunications - Heinrich-Hertz-Institut, 10587 Berlin, Germany
‡Information Systems Laboratory, 94305 Stanford CA, USA

*Abstract*— **Peer-to-peer (P2P) networks represent a valuable architecture for streaming video over the Internet. In these systems, users contribute their resources to relay the media to others and no dedicated infrastructure is required. In order to ensure a low end-to-end delay, P2P overlay networks are often organized as a set of complementary multicast trees. The source of the stream multiplexes the data on top of these trees and the routing of packets is statically defined. In this scenario, the reliability of the overlay links is critical for the performance of the system since temporary link failure or network congestion can cause a significant disruption of the end-user quality. The novel Scalable Video Coding (SVC) standard enables efficient usage of the network capacity by allowing intermediate high capacity nodes in the overlay network to dynamically extract layers from the scalable bit stream to serve less capable peers. On the other hand, SVC incurs a certain loss in terms of coding efficiency with respect to H.264/AVC single-layer coding. We propose a simple model that allows to evaluate the trade-off of using a scalable codec with respect to single-layer coding, given the distribution of the receivers' capacities in an error-free network. We also report experimental results obtained by using SVC on top of a real-time implementation of the Stanford Peer-to-Peer Multicast (SPPM) protocol that clearly show the benefits of a prioritization mechanism to react to network congestion.**

## I. Introduction

Video streaming over the Internet has become popular due to the increasing availability of network resources and recent advances in video coding technologies. In several application scenarios, such as IP-TV, the same video content has to be transmitted to a large population of users. Content Delivery Networks (CDN) are often used to support large numbers of users, but they require the deployment of special infrastructure. As an alternative, peer-to-peer (P2P) overlay networks have been considered for multimedia delivery. In fact, users who are interested in a video stream often have sufficient resources to act as relay points and forward the video to other users. This allows the system to scale with the number of nodes involved in the communication. P2P systems are widely used for file sharing applications. However, live multicasting represents a much more challenging problem, since the overlay network must guarantee high reliability as well as low startup latencies. Also, a very low end-to-end delay is desirable for interactive applications like online lectures, where some kind of feedback is provided by the users to the source of the video stream.

Several different architectures have been proposed in the literature. They can be grouped in two major categories. Mesh-based approaches aim to construct an overlay network whose connections are maintained through "gossip" messages. In this case, peers are self-organized into a mesh and independently request portions of the video from neighbors, with no particular emphasis on the structure of the distribution path [1], [2], [3]. Even though these solutions provide good error resilience and network utilization performance, they are usually characterized by high end-to-end delays, mostly due to the *push-pull* approach used for the dissemination of the video data that requires the receivers to coordinate the download of the data. On the other hand, Tree-based approaches simply *push* video packets along routes that are determined by constructing one [4] or many [5], [6] multicast trees rooted at the media source. Multiple trees increase the robustness of the system, since additional path diversity allows to avoid single point of failures. In this case, multiple description coding [7] (MDC) is often assumed to be used for error resilience of the source signal. In our recent work [8], [9], we introduce a content-aware prioritization algorithm that selectively drops unimportant packets and a local retransmission scheme that allows to sustain the end-user quality in case of errors, without introducing redundancy at the source. Even though tree-based approaches are suitable for low-delay video streaming, they require a constant throughput of the links in the overlay network to ensure the continuous flow of information from the source to the receivers. Network congestion occurring on the uplink of a node with many descendants in the tree can cause packet loss for a large number of peers. In order to serve all the peers with an acceptable quality, a content-aware rate adaptation mechanism is required to limit the effect of congestion.

The novel Scalable Video Coding (SVC) standard [10], [11] supports the encoding of a video signal at different qualities within the same layered bit stream. This allows for performing efficient on-the-fly rate adaptation, while achieving compression efficiency comparable to H.264/AVC single-layer coding. Furthermore, SVC allows a more efficient usage of the network bandwidth on a P2P network by enabling intermediate high capacity nodes in the overlay to dynamically extract layers from the scalable bit stream to serve less capable peers. We propose a mathematical framework that allows to quantify the advantage of using a scalable codec for P2P distribu-

1-4244-0981-0/07/$25.00 ©2007 IEEE.

tion from a resource constrained server in an error-free network, with respect to single-layer simulcasting. Also, we report experimental results obtained on a real-time system that show the advantage of using SVC to prioritize video packets in case of network congestion.

This paper is organized as follow. We report our model to estimate the expected distortion of H.264/AVC simulcast and SVC for P2P distribution in Sec. II. The *Stanford Peer-to-Peer Multicast* (SPPM) protocol, used for our streaming experiments, is described in Sec. III. In Sec. IV, we review the key features of the Scalable Video Coding standard. Experimental results obtained by streaming SVC over several SPPM clients are reported in Sec. V, while Sec. VI concludes the paper.

## II. Theoretical model of expected distortion

We consider a scenario where a throughput constrained source transmits video to a set of $N$ users with heterogeneous network connections over a P2P network. A sample applications may be an online lecture with a few hundred receivers. We assume that for every node, the uplink capacity is the same as the downlink capacity. For simplicity, it is also assumed that the source knows the distribution of the clients' throughput and the network is error free.

In order to deploy the video, the source of the stream can either use single-layer coding or a scalable codec. In the first case, multiple overlays may be constructed to support different groups of receivers. This scheme is often called *simulcast*. In the second case, only one overlay is required, since intermediate high capacity nodes are allowed to extract portions of the scalable bit stream to serve less powerful users.

When single-layer coding is in use, the source has to determine the optimal number of streams to be generated and the corresponding rate allocation, such that the expected distortion over the set of receivers is minimized. The optimization problem can be formulated as follows:

$$\text{Minimize} \quad \sum_{i=1}^{N} D(R_i) \tag{1}$$

$$\begin{aligned} s.t. \quad & R_i \leq C_i \\ & R_i \in \{S_1 \ldots S_K\} \\ & \sum_{j=1}^{K} S_j \leq C_0 \end{aligned}$$

where $K$ is the total number of single-layer streams and each stream $j$ is encoded at a rate $S_j$. $C_0$ is the uplink capacity of the source and $C_i$ is the available throughput for every peer $i$. The optimal bandwidth allocation for peer $i$ is then $R_i$ resulting in a distortion $D(R_i)$. Notice that this is a combinatorial problem over the number of encoded streams $K$ and the number of peers $N$. The optimal solution can only be determined for small values of $K$ and $N$ in a practical system. Furthermore, the optimal value depends on the distribution of the peers' capacities and the source's connection speed $C_0$, as well as the distortion-rate function of the video signal. The usage of scalable video coding greatly simplifies the problem at the source, since only one scalable bit stream will be encoded at the maximum allowed rate $R = \min(C_0, \max(C_i))$.

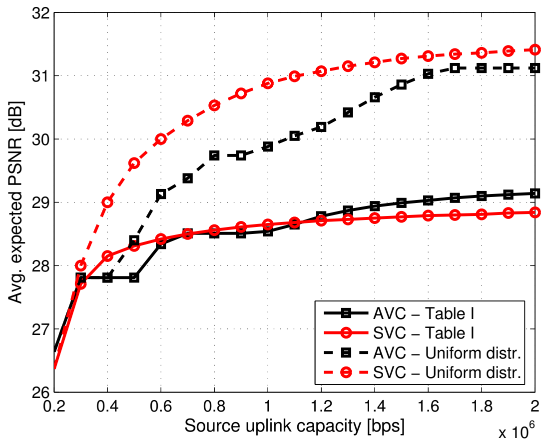


Fig. 1. Expected average quality over 1000 peers for *Bus* at CIF resolution. SVC provides a significant improvement with respect to H.264/AVC, especially when the source uplink speed is constrained.

TABLE I
DISTRIBUTION OF PEER'S BANDWIDTH [12].

| Bandwidth | Percentage |
|---|---|
| 256 kbps | 56% |
| 384 kbps | 21% |
| 896 kbps | 9% |
| 2 Mbps | 14% |

We illustrate the advantage of using a scalable codec by simulating 1000 nodes in a resource constrained scenario. We perform experiments by allocating the connection speeds as shown in Table I [12] and according to a uniform distribution for the same connection speeds. The distortion-rate function follows the parametric model proposed in [13]:

$$D = D_0 + \frac{\theta}{R - R_0}$$

We fit the three parameters $\theta$, $D_0$ and $R_0$ by performing trial encodings. The distortion-rate function for SVC is obtained by adding 10% of redundancy to the H.264/AVC rate-distortion points, as discussed in [14]. We assume that the video can be extracted at any possible rate point. We use the *Bus* sequence at CIF resolution at 30 frames per second. We report the optimal expected average distortion for all the peers in the system in Fig. 1. For the AVC curve, we run the optimization problem in (1) for each value of the source uplink capacity $C_0$. This allows to find the optimal number of streams to encode $K$ and the corresponding rate allocation. Then, every peer $i$ receives the bit stream $j$ with the best possible quality $S_j$ such that $S_j < C_i$. The SVC curve is obtained by averaging

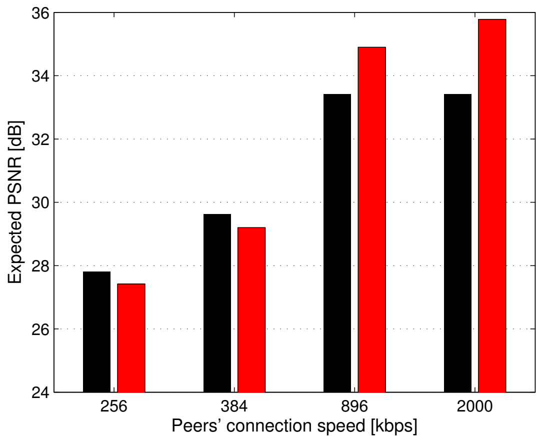


Fig. 2. Expected video quality as a function of the peers' connection speed for *Bus* at CIF resolution. The source uplink bandwidth is 1.5 Mbps.

the quality assuming that every peer $i$ receives a video at rate $R = \min(C_0, C_i)$.

We observe that SVC provides a quality improvement especially when the source capacity is limited and for uniform distribution of peers' capacities, while H.264/AVC performs similarly when the source throughput is significantly overprovisioned. In this case, the advantage of having a scalable coder does not fully compensate the loss in coding efficiency on average. It is important to note that, even if the gain in terms of average PSNR is quite limited, SVC allows a more fair distribution among the peers, providing a quality proportional to their bandwidth contribution, as reported in Fig. 2. In practice, H.264/AVC simulcasting penalizes high-quality peers in a constrained scenario because part of the source uplink capacity is reserved for the transmission of a reduced quality bit stream to serve peers with lower capacities. We repeated the experiment for several other sequences and different bandwidth distributions and found similar results for our model assumptions.

In the remaining part of the paper, we consider a resource constrained scenario where network congestion may occur and we show how SVC can be used to limit the quality degradation due to packet loss.

## III. Stanford Peer-to-Peer Multicast Protocol

The *Stanford Peer-to-Peer Multicast* (SPPM) protocol has been designed for video content delivery with very low end-to-end delay. In order to achieve this goal, peers are organized in an overlay network that consists of a set of complementary multicast trees. The source of the stream is the root of each distribution path and multiplexes video packets onto different trees to ensure load balancing and good error resilience. For instance, if the overlay consists of two complementary multicast trees, the source will schedule even packets on one tree and odd packets on the other tree. Fig. 3 shows a simple topology with eight peers and two distribution trees as an example of the

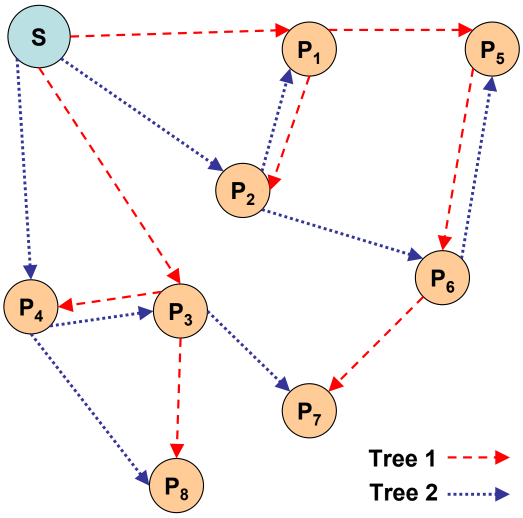


Fig. 3. An example of SPPM overlay network. Two complementary multicast trees are maintained to route the data from the source of the stream to the set of users.

overlay created by SPPM. In the following, we describe how SPPM creates and maintains the overlay network and guarantees good error resilience.

### A. Topology construction

A new node that is willing to join a multicast session is required to step through a bootstrap process in order to acquire some knowledge of the current status of the overlay network. The SPPM join process consists of a "six-way" handshake as described in the following and shown in Fig. 4.

- *JOIN*. The new peer contacts the source of the video stream to obtain some setup information, such as the number of multicast trees and the video bit-rate. The source provides a list of randomly selected peers that can be considered as "candidate parents" in the distribution trees.
- *PROBE*. For each tree, the new peer sends a packet to each candidate parent to obtain information about their current state, such as the available throughput on the uplink, the current distance from the source in terms of logical hops and the end-to-end delay from the source. These information is used to determine the best parent to which to connect.

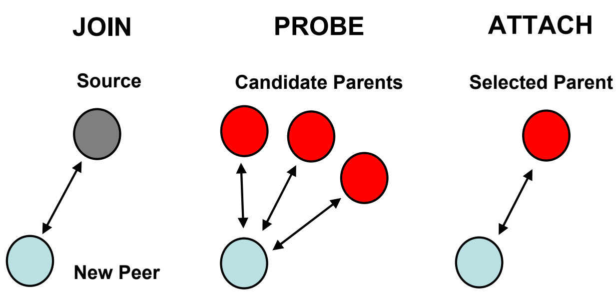


Fig. 4. SPPM implements a six-way handshake process to join peers to a multicast session. Each new peer 1) contacts the source, 2) probes a set of candidate parents and 3) attaches to a selected parent.

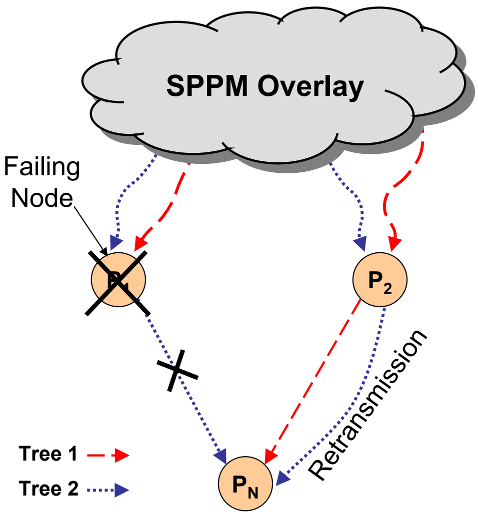


Fig. 5. Peer disconnection causes packet loss to occur for a distribution tree. SPPM employs local retransmission to limit quality degradation due to such failures. This may increase the congestion on the uplink of the nodes that are serving the retransmission requests.

- *ATTACH*. For each tree, one candidate parent is selected and an actual connection is established by sending an appropriate request. The parent is selected in a way to minimize the height of the distribution trees. Also, different parents are chosen as often as possible to increase path diversity.

### B. Overlay maintenance

Nodes can leave ungracefully causing temporary disconnection of a distribution path. In order to recover from such failures, each client monitors the state of the directly connected parents by sending periodic `HELLO` messages. Whenever a host leaves the overlay, it stops forwarding video packets and it is unresponsive to `HELLO` messages sent by the children. By setting appropriate timeouts, each child can promptly detect parent disconnection and consequently trigger the re-join procedure for the affected tree. During the reconnection process, a node may attach to one of its descendants, thus creating a loop that would eventually starve the whole subtree. To prevent this problem, SPPM maintains on each peer and for each tree an updated list of the ancestor nodes on the path from the source.

### C. Error resilience

In the peer-to-peer streaming scenario, packet loss can be caused either by congestion or by the disconnection of one node in the distribution path from the source of the stream. The loss statistics are highly time-variant as there might be some long periods during which a host is fully connected and experiences no loss, and other times when a large portion of packets are missing. SPPM employs retransmission to limit quality degradation due to packet loss. Path diversity allows retransmission requests to be limited to local parents, thus avoiding feedback implosion at the source of the stream. Fig. 5 shows an example of failure for one parent $P_1$. In this case the child $P_N$ will be able to request retransmission of missing packets to the parent in the second distribution tree $P_2$, while the re-join process is occurring.

It is important to notice that retransmission requests will place an additional burden on the uplink of peers already forwarding a portion of the video to one or possibly many children. This increase in congestion is more important when there are few multicast trees as a larger portion of the video will be requested from fewer parents. In order to relieve the congestion on the uplink, each peer has to perform rate adaptation by dropping less important packets whenever necessary.

### D. Content-aware prioritization

Content-aware prioritized scheduling can help maintain video quality whenever a peer has to drop some packets due to congestion to ensure timely delivery of the more significant portion of the streams. Setton et al., [8] presented a Congestion-Distortion optimized (*CoDiO*) packet scheduler. This prioritization algorithm bases its decisions on the “importance” of each enqueued packet. The metric used for prioritization reflects how decoding of a particular packet reduces distortion and it captures the dependencies among different packets. Such a prioritization scheme becomes particularly simple when the video source is encoded using a layered representation such as that supported by the SVC standard. Video packets can be prioritized according to the importance in contributing to the decodable quality, by transmitting the base layer first. The data for the enhancement layers will be transmitted only if the network conditions allow that. Additionally to the perceptual importance, SPPM considers the number of potential destinations for each packet while performing prioritization.

## IV. Scalable Video Coding

The SVC [10], [11] design, which is an extension of the H.264/MPEG-4 AVC [15] video coding standard, can be classified as layered video codec. SVC - layered video coding is suitable for different use-cases like, e.g., supporting heterogenous devices with a single, scalable video stream. Such as stream allows for delivering a decodeable and presentable quality of the video depending on the device’s capabilities. Here, quality refers to resolution, frame rate and bit-rate of the decoded operation point of the scalable video stream. Another use-case is the adaptation to varying network conditions, e.g, in case of congestion. Typically, Peer-to-peer streaming applications are lacking in Quality of Service (QoS) due to missing provisioning mechanisms by the underlying network, the Internet. Reasonable solutions for coping with congestion, e.g., on uplink connections, typically require rate adaptation of the data stream. SVC allows for removing packets from the bit stream, which implicitly results in bit-rate as well as quality reduction of the video.

Coder structure and coding efficiency of SVC depend on the scalability features required by an application. Fig. 6 shows a typical coder structure with two quality layers for SNR scalability. When the resolution between layers

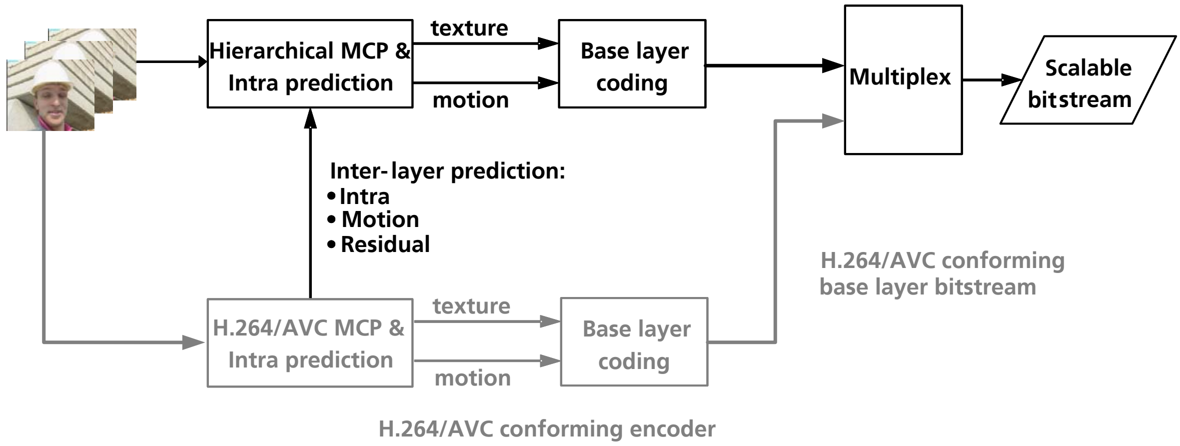


Fig. 6. Typical SVC coder structure for SNR scalability.

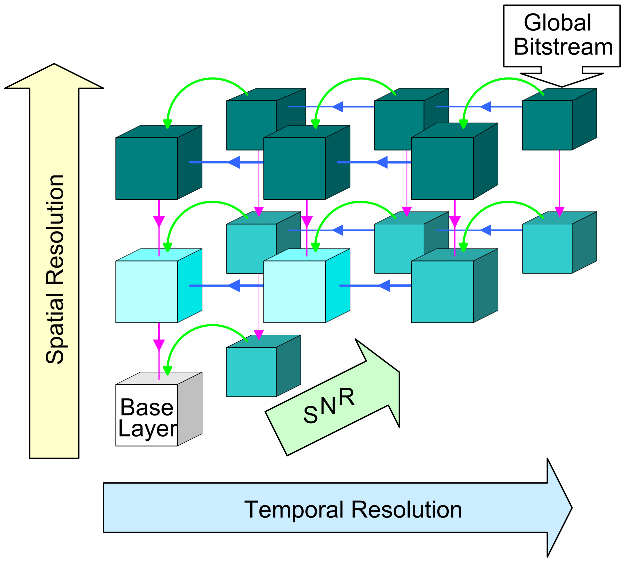


Fig. 8. High Level view of SVC in combined scalability mode.

changes, spatial scalability is used. Moreover, spatial and SNR scalability can be mixed. The enhancements are coded given the predictions from lower layers like the base layer. Temporal scalability is achieved by hierarchical B pictures as described later.

In SVC, the hybrid video coding approach of motion-compensated transform coding is extended in a way that a wide range of spatio-temporal and quality scalability is achieved. An SVC bit stream consists of a base layer and one or several nested enhancement layers. The removal of enhancement layers still leads to a reasonable quality of the decoded video at reduced temporal, SNR, or spatial resolution. The base layer is a H.264/MPEG-4 AVC [15] compliant bit stream that ensures backward-compatibility for existing receivers. The decoding process itself is still based on a single decoding loop for keeping processing overhead small.

The temporal scaling functionality of SVC for high-delay configurations is typically based on a temporal decomposition using hierarchical bi-predictive pictures. Fig. 7 shows hierarchical B pictures with two layers of SNR quality scalability: base layer and one quality enhancement layer. Pictures at time instants 0, 4, and 8 represent so-called *key pictures*. These pictures serve as synchronization points between encoder and decoder to allow recovering from drifts. These drifts are typically caused by dropping only some frames of the quality enhancements for bit-rate adaptation. The encoder-decoder synchronization is achieved at the cost of coding efficiency since also the enhancement layer pictures at time instants 4 and 8 in Fig. 7 are inter-predicted from base layer pictures at time instants 0 and 4, respectively. The B pictures between the key pictures are forming the temporal enhancement levels. Where picture 2 and 6 form the first temporal enhancement to the key pictures and picture 1, 3, 5, and 7 the second temporal enhancement respectively. The base layer pictures at time instants 1, 2, and 3 as well as 5, 6, and 7 are predicted from the highest available enhancement layer pictures. This approach provides high coding efficiency for the base layer in case the reference pictures are available and does not pose a problem when the reference pictures are not available, since only a small number of pictures depends on these through inter prediction. Each B picture of a higher temporal enhancement level is encoded with a higher Quantization Parameter QP (cascaded QP assignment), thus the fidelity per picture is decreasing with the decreasing importance in terms of the number of succeeding references by other pictures. The equation below provides a typical QP cascading with QP(x;y;z) being the QP value assigned to pictures at time instants x, y, and z. $QP(0;4;8) = QP(2;6) - 3 = QP(1;3;5;7) - 4$ As well, in Fig. 7 a typical structure of an SVC bit stream is shown, which comprises a group of pictures (GOP) of size four. GOPs can be independently decoded, if the corresponding key picture has random access properties and the preceding reference is available. For low-delay configurations, prediction dependencies can be selected in a way that no future dependencies are used. This allows for minimizing the structural decoding delay down to zero frames. Although this structure allows for the

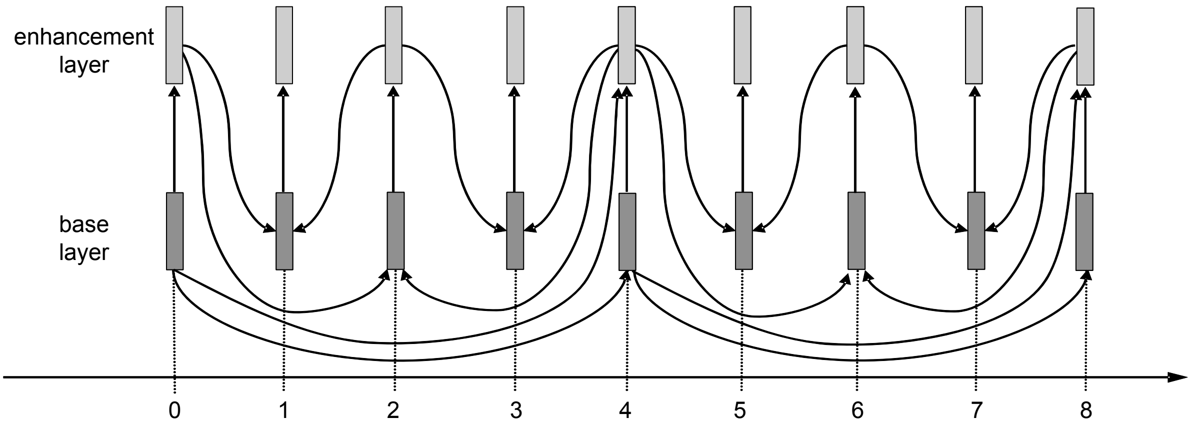


Fig. 7. SVC temporal prediction structure with key pictures, hierarchical B pictures and 2 quality layers with a group of picture (GOP) size of 4.

same temporal scalability functionality as that using future dependencies, it reduces the coding efficiency at the price of low-delay.

The spatial scalability of SVC is achieved by different encoder loops with an over-sampled pyramid for each resolution (e.g. QCIF, CIF, and 4CIF), including motion-compensated transform coding with independent prediction structures for each layer. In contrast to the encoder, the decoder can be operated in single loop, i.e., for decoding inter-layer dependencies it is not required to perform motion compensation in lower layers on which a layer depends. A combination of all three scalability functionalities within one bit stream is called combined scalability. Such a combined scalable bit stream allows for extraction of different operation points of the video, where each operation point is characterized by a certain level of SNR quality, temporal and spatial resolution. A global bit stream is entry point into the scalable bit stream at maximum scalability levels. Fig. 8 shows the general structure of a global bit stream from the high level point of view.

### A. *SVC for Network Transport*

One typical application for SVC is bit-rate adaptation for transport over packet-switched networks like the Internet. In the use-case presented in this paper, the media is not directly transferred from source to client, but is relayed via peers of an overlay network. For this use case signaling within the bit stream is an important feature for allowing the source node as well as relay nodes to apply bit stream adaptation. A relay or peer node may like to find out which parts of the bit stream can be dropped first, if adaptation is required. This may be the case when network congestion occurs. Since a SVC bit stream can support up to three dimensions of scalability, a peer node needs to get detailed information about the importance of the video data in terms of reconstructed quality. But a peer may also rely on absolute importance values for adaptation only. Therefore, the encoder must already have selected the adaptation path through a global bit stream. For the aforementioned reasons, the identification of video data belonging to different layers is achieved by an extended approach of the Network Abstraction Layer (NAL) concept of H.264/MPEG-4 AVC [15]. The NAL hides the detailed bit stream structure of H.264/MPEG-4 AVC and allows for high level (application layer) readability of the NAL packets. NAL packets typically represent a video frame, a part thereof, parameter sets (decoder initialization information), supplemental enhancement information (SEI - supplying additional bit stream information like time stamps not required for decoding) and bit stream organizing information like end-of-stream indication. For SVC, the NAL header syntax has been extended for allowing identification of temporal, spatial and quality scalability information per NAL packet. In order to give a pre-computed, single adaptation path through a bit stream, the NAL header also provides an one-dimensional priority indicator, which has to be set by the encoder. For more details on the NAL unit header and network adaptation, we refer to [16], [17]. Beside knowing which operation point a NAL packet belongs to, a peer furthermore needs information about the characteristics of such a point. For that a Scalability Information SEI message provides detailed information about values like average bit-rate and frame rate of operation points. This message can be transferred by in-band as well as out-of-band transport mechanisms.

If SVC may be supported by end user devices, still a huge number will rely on older H.264/MPEG-4 AVC profiles specified in [15], thus there may be still the need of adjustment of an SVC bit stream to an H.264/MPEG-4 AVC stream complying with [15]. Such an operation is principally realizable without re-encoding, but with a distinctly reduced overhead compared to re-encoding.

The scalability techniques used in this paper are based on bit streams with two SNR quality layers, as shown in Fig. 7. A GOP size of 16 is used instead of a size of four as shown in Fig. 7, which allows for creating five temporal levels of the bit stream. Although having only one explicit enhancement layer coded, multiple bit-rate operation points can be formed by keeping the base layer

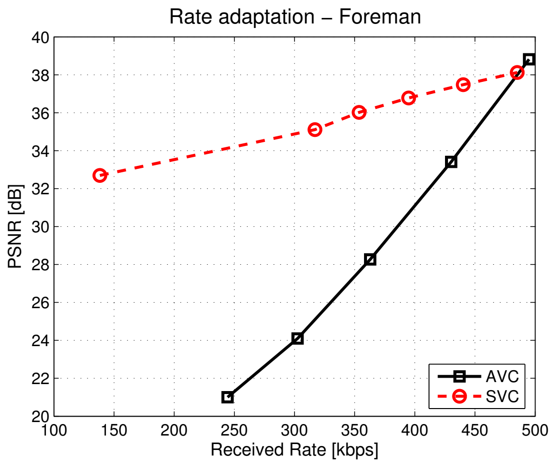


Fig. 9. PSNR versus bit-rate for different operation points of the SVC and AVC bit streams for *Foreman*. Rate adaptation for SVC is achieved using temporal scalability in the enhancement layer only, while keeping intact the base layer. Similarly, rate adaptation for AVC is achieved by temporal scalability and by employing the "copy" error concealment for missing frames.

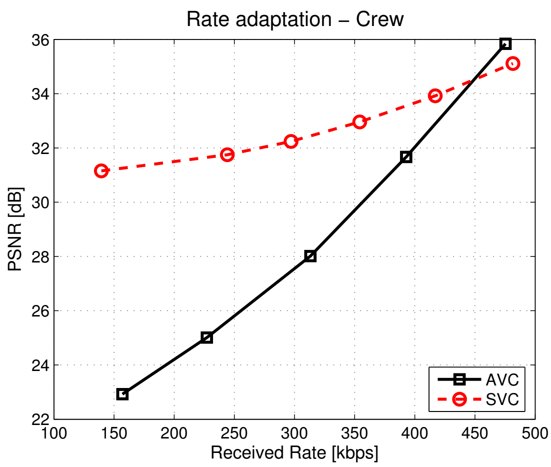


Fig. 10. PSNR versus bit-rate for different operation points of the SVC and AVC bit streams for *Crew*.

intact and dropping temporal levels of the enhancement layer only. By this a rate curve as shown in Fig. 9 and Fig. 10 can be achieved for the two layer SVC bit streams used in the experiments reported in the following section. These curves correspond to the H.264/MPEG-4 AVC single layer streams with temporal scalability in the same figures. Note, that the rate distortion curves for SVC can also be formed with a more concave shape by exhaustively searching the optimal dropping strategy for the enhancement layer pictures. Since we assume live input encoded at the source, a long-term precalculation as required by such an approach would not be feasible. Thus, we herein rely on the successively temporal dropping strategy of enhancement layer pictures only.

## V. Experimental Results

We performed experiments on a local network by deploying 100 SPPM clients on different physical machines. We encoded the Foreman and Crew sequences at CIF resolution, 30 frames per second, at 500 kbps by using the SVC reference software (JSVM 8.8). We fixed a GOP structure of 16 frames and used hierarchical B pictures for temporal scalability. One SNR enhancement layer is encoded on top of an approximately 150 kbps base layer. We used repeated parts of the sequences to create a 25 minutes long bit stream. Video frames are fragmented at the source in packets of 1400 bytes. The performance of SVC is then compared to single-layer AVC encoded at the same bit-rate and with the same temporal coding structure. Previous frame error concealment is assumed to be used at the decoder to limit quality degradation whenever a packet loss affects a base layer frame.

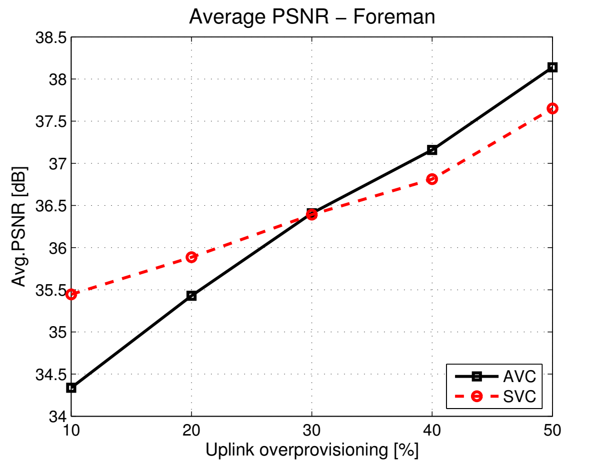


Fig. 11. Average PSNR computed over 100 peers for *Foreman*. SVC allows to limit the quality drop by providing a better rate adaptation and increasing the probability the base layer achieves the end user.

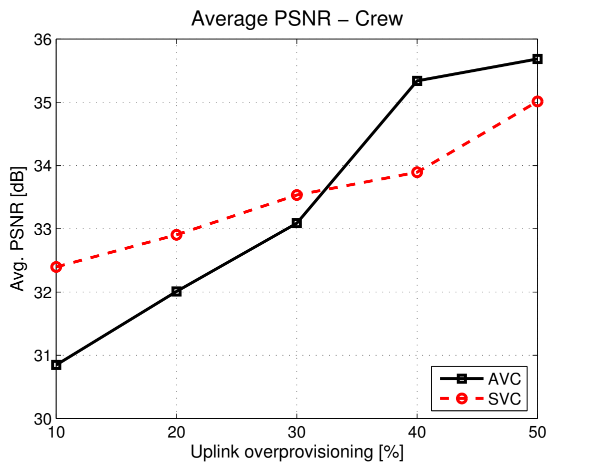


Fig. 12. Average PSNR computed over 100 peers for *Crew*.

In order to simulate peer churn, we scheduled each client to connect to the system for a random amount of time, varying from 1 up to 2 minutes, and then to stay disconnected for a random interval between 10 and 40 seconds. This highly dynamic system causes peers to require retransmission whenever they miss video packets from a disconnected parent, thus increasing network congestion. We constrained the available throughput on

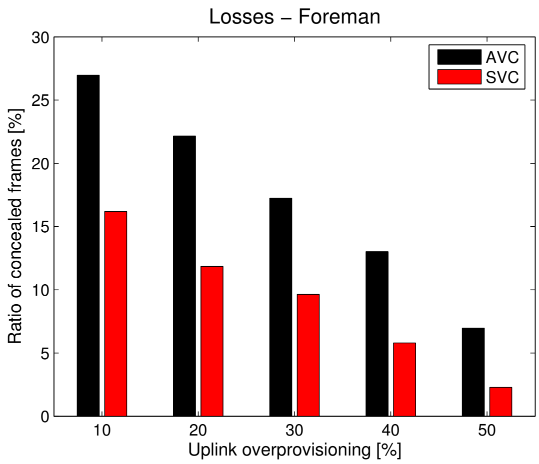


Fig. 13. Percentage of losses for *Foreman*. SVC consistently allows to reduce the number of losses with respect to AVC.

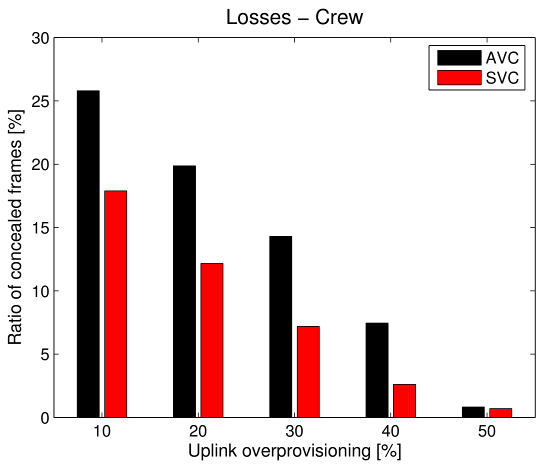


Fig. 14. Percentage of losses for *Crew*. SVC consistently allows to reduce the number of losses with respect to AVC.

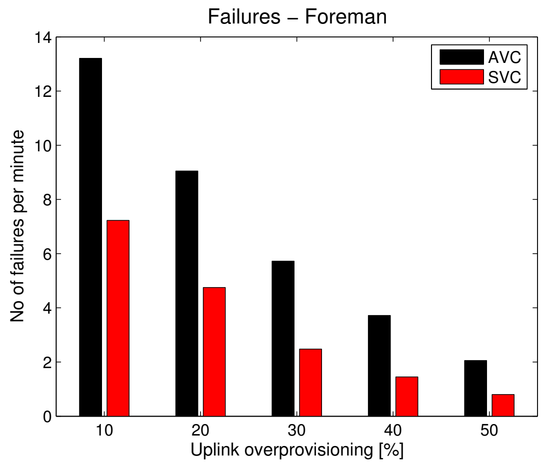


Fig. 15. Histogram of the average number of failures per peer per minute for *Foreman*. A failure is considered to be the loss of more than four consecutive frames.

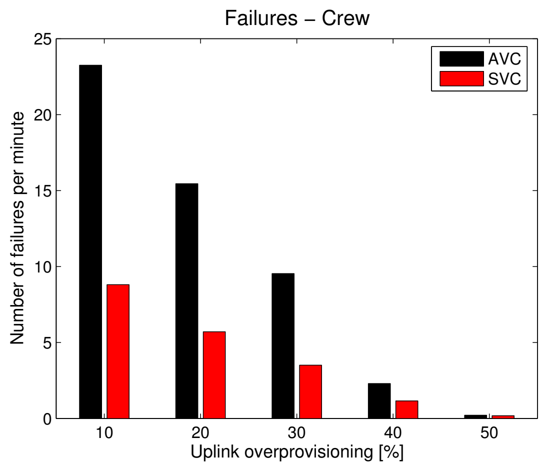


Fig. 16. Histogram of the average number of failures per peer per minute for *Crew*.

the uplink of each peer. Different experiments have been performed by varying the overprovisioning factor of the uplink from 10% to 50% of the rate required to serve the video. In practice, the available throughput for each peer varies in the experiments from 550 to 750 kbps.

We report in Fig. 11 and 12 the average PSNR as a function of the uplink rate, over all the 100 peers, for Foreman and Crew respectively. SVC allows to sustain a much better quality with respect to AVC when the uplink of the peers is particularly congested. A significant gain of 1 dB and 1.8 dB is observed for Foreman and Crew respectively, when the uplink is overprovisioned by only 10% of the rate of the source. This is because the prioritization mechanism integrated in SPPM schedules the base layer packets first, thus increasing the probability that at least a base quality video is received for all the peers even when the network is severely congested. We also observe that, by increasing the throughput to 750 kbps, it is possible to serve all the peers and achieve almost perfect quality. In this case, AVC outperforms SVC due to the better coding efficiency.

Fig. 13 and 14 show the histograms with the percentage of losses as a function of the uplink rate. These graphs confirm the results for the average PSNR, since the number of losses is consistently smaller when streaming SVC, especially when the network is heavily congested.

Finally, we report the average number of failures per minute observed at the peers in Fig. 15 and 16. We classify as a visible failure the loss 4 or more consecutive frames causing the video to freeze for more than one hundred milliseconds. Also in this case, SVC allows to guarantee a much better performance of the system by producing less than half of the failures observed by streaming AVC content.

## VI. Conclusions

We investigated the advantage of Scalable Video Coding for real-time P2P streaming. We reported an analytical model to quantify the expected performance of SVC and single-layer H.264/AVC-based simulcasting in an error-free network. Significant gains can be observed in case of heterogenous bandwidth distributions of the peers and limited uplink capacity of the source. It is shown how SVC allows a more fair delivery of the video quality to the peers corresponding to their connection speed.

Furthermore, we applied SVC on top of a real-time implementation of the *Stanford Peer-to-Peer Multicast* (SPPM) streaming system. Experimental results show that SVC improves the performance in case of congestion by providing better rate adaptation with respect to single-layer H.264/AVC coding. When the uplink of the peers is set close to the video source rate, significant gains are reported in terms of average PSNR over all the peers (1 dB for *Foreman* and 1.8 dB for *Crew*). Also, the frame loss rate and the total number of failures observed is significantly decreased.

## REFERENCES


[1] X. Zhang, J. Liu, B. Li, and T.-S. P. Yum, "DONet/CoolStreaming: A Data-driven Overlay Network for Live Media Streaming," *Proc. IEEE Infocom, Miami, USA*, Feb. 2005.

[2] M. Zhang, J.-G. Luo, L. Zhao, and S.-Q. Yang, "A Peer-to-Peer Network for Live Media Streaming using a Push-Pull Approach," *Proc. ACM Int. Conference on Multimedia*, pp. 287–290, Nov. 2005.

[3] ——, "Large-Scale Live Media Streaming over Peer-to-Peer Networks through Global Internet," *Proc. ACM Int. Conference on Multimedia, P2PMMS Workshop*, pp. 21–28, Nov. 2005.

[4] Y. Chu, S. Gao, and H. Zhang, "A case for end system multicast," *Proc. ACM Sigmetrics, Santa Clara USA*, June 2000.

[5] M. Castro, P. Druschel, A.-M. Kermarrec, A. Nandi, A. Rowstron, and A. Singh, "SplitStream: High-bandwidth content distribution in a cooperative environment," *Proc. IPTPS'03, Berkeley, CA*, Feb. 2003.

[6] V. N. Padmanabhan, H. J. Wang, P. A. Chou, and K. Sripanidkulchai, "Distributing Streaming Media Content Using Cooperative Networking," *Proc. ACM NOSSDAV, Miami Beach, FL*, May 2002.

[7] V. Goyal, "Multiple Description Coding: Compression Meets the Network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sept. 2001.

[8] E. Setton, J. Noh, and B. Girod, "Rate-Distortion Optimized Video Peer-to-Peer Multicast Streaming," *Workshop on Advances in Peer-to-Peer Multimedia Streaming at ACM Multimedia, Singapore*, pp. 39–48, Nov. 2005.

[9] P. Baccichet, J. Noh, E. Setton, and B. Girod, "Content-Aware P2P Video Streaming with Low Latency," *IEEE Int. Conference on Multimedia and Expo, ICME, Beijing, China*, July 2007.

[10] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. W. (Editors), "Joint Draft 11 of SVC Amendment, Joint Video Team (JVT)," *Doc. JVT-X201, Geneva, Switzerland*, June 2007.

[11] 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. 1103–1129, Sept. 2007.

[12] K. Sripanidkulchai, A. Ganjam, B. Maggs, and H. Zhang, "The feasibility of supporting largescale live streaming applications with dynamic application endpoints," *Proceedings SIGCOMM04, Portland, USA*, Aug. 2004.

[13] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of Video Transmission over Lossy Channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–32, June 2000.

[14] H. Schwarz and T. Wiegand, "Further results for an rd-optimized multi-loop SVC encoder, Doc. JVT-W071," Apr. 2007.

[15] *Advanced Video Coding for Generic Audiovisual services, ITU-T Recommendation H.264 - ISO/IEC 14496-10(AVC)*, ITU-T and ISO/IEC JTC 1, 2003.

[16] S. Wenger, Y.-K. Wang, and T. Schierl, "RTP Payload for SVC Video, draft-ietf-avt-svc-02," *Internet Engineering Task Force (IETF)*, July 2007.

[17] ——, "Transport and Signaling of SVC in IP Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1164–1173, Sept. 2007.