Multimedia Systems manuscript No.
(will be inserted by the editor)

ALD: Adaptive Layer Distribution for Scalable Video

Jason J. Quinlan · Ahmed H. Zahran · Cormac J. Sreenan

Received: date / Accepted: date

Abstract Recent years have witnessed a rapid growth in the demand for streaming video over the Internet and mobile networks, exposes challenges in coping with heterogeneous devices and varying network throughput. Adaptive schemes, such as scalable video coding, are an attractive solution but fare badly in the presence of packet losses. Techniques that use description-based streaming models, such as Multiple Description Coding (*MDC*), are more suitable for lossy networks, and can mitigate the effects of packet loss by increasing the error resilience of the encoded stream, but with an increased transmission byte-cost.

In this paper, we present our adaptive scalable streaming technique Adaptive Layer Distribution (*ALD*). ALD is a novel scalable media delivery technique that optimises the tradeoff between streaming bandwidth and error resiliency. ALD is based on the principle of layer distribution, in which the critical stream data is spread amongst all packets thus lessening the impact on quality due to network losses. Additionally, ALD provides a parameterised mechanism for dynamic adaptation of the resiliency of the scalable video. The Subjective testing results illustrate that our techniques and models were able to provide levels of consistent high quality viewing, with lower transmission cost, relative to MDC, irrespective of clip type. This highlights the benefits of selective packetisation in addition to intuitive encoding and transmission.

Jason J. Quinlan · Cormac J. Sreenan
Department of Computer Science
University College Cork,
Ireland
E-mail: [j.quinlan,cjs]@cs.ucc.ie

Ahmed H. Zahran
Electronics and Electrical Communications Department
Cairo University,
Egypt
E-mail: azahran@eece.cu.edu.eg

Keywords Scalable Video · Lossy Networks · Layered Coding · Error Resilience · Layer Distribution

CR Subject Classification H.5.1 · C.2.5

1 Introduction

The popularity of media streaming, especially mobile video [6], is increasing the bandwidth crunch of network operators. This increase is enabled by new mobile devices that feature a huge diversity in their capabilities. However, the increase escalates many transmission issues faced by real-time applications, such as packet delay [33], buffering [32], bandwidth variation and congestion; both rate control [8] and packet dropping [39]. Hence, adaptive media streaming [7] represents a corner stone in the pervasiveness of mobile video by changing the streaming characteristics according to changes in the transmission context, e.g. device capabilities, service cost, and available resources. In this domain, scalable video encoding [14] is an important technique for streaming adaptability. Generally, a video is identified as a scalable stream when an original high quality version of the video can be encoded into a set of sub-streams such that a combination of one or more of these sub-streams can be used to replay the video.

Scalable Video Coding (*SVC*) [28], an extension to the H.264/MPEG-4 Part 10 or AVC (Advanced Video Coding) compression standard, represents the first standardised scalable video encoding scheme. In SVC, a high quality media clip is fragmented into N layers including a base *layer* and numerous enhancement layers as shown in Figure 1a in which $N = 6$.

The base layer provides a coarse minimal quality, the reception of subsequent enhancement layers increases the viewable quality by providing an increase in temporal, spatial or quality dimensionality. The temporal scalability is de-

Fig. 1 An example of (a) a six layered SVC stream encoded as (b) MDC-FEC (blue denotes original SVC data, green - additional FEC data)

termined through using different frame rates, spatial scalability is defined by changing the frame resolution, and quality scalability is achieved by scaling the amount of bits used to encode a picture without changing its resolution. As well as implementing the key concepts of layered coding, SVC also provides a mechanism for efficient scalable streaming. By gathering a number of continuous frames in to a collection known as a group of frames (*GOF*) or group of pictures (*GOP*). SVC provides an efficient mechanism for creating frame interdependency based on intra- and inter-frames. Intra-frames are fixed points in the stream, and are independent of other frames, while inter-frames provide a means of bitrate reduction by relying on adjacent frames for supplementary data prior to decoding.

A major limitation in layered coding is the prioritised encoding hierarchy by which the increase in quality provided by an enhancement layer is subject to the availability of lower layers that the enhancement layer is dependent upon. In this manner, the loss of a lower layer prohibits the receipt of a higher enhancement layer. More seriously, the loss of the base layer invalidates video decoding. The limitation is further exacerbated when the individual frame types i.e. I, P and B frames of a GOP, mandate inter-frame dependency such that the loss or a low quality decoding value of a frame can further limit the achievable quality of all dependent frames [36], thus mandating low quality decoding. This limitation makes SVC an unattractive approach for links featuring a high error probability such as wireless links, as it necessitates the overhead of retransmission schemes to recover lost packets.

To overcome the impact of packet losses without having to resort to retransmissions, Multiple Description Coding (*MDC*) [2,12] has been proposed. The key idea of MDC is introducing redundancy to the transmitted video to compensate for packet losses. MDC offers an encoding scheme where the original layered data is interspersed with error resilience, typically Forward Error Correction (*FEC*)) [23]. Each MDC description provides a low quality fidelity decoding, with the cumulative decodable quality based on the number of descriptions received by the device. In this regard, MDC provides a high level of consistency to stream quality by providing a high level of error correction to mitigate network transmission issues albeit at a much higher transmission cost in comparison to SVC. Recently, we presented preliminary results for Adaptive Layer Distribution (ALD) [25], a novel description-based encoding technique, that introduced several enhancements that significantly reduce the average transmission overhead while maintaining very close streaming quality to MDC.

In this paper, we evaluate how the consistency of viewable quality of SVC, MDC and ALD, and their respective variants, vary as the number of frames per GOP increases, known as GOP value, and we illustrate the results of Subjective testing undertaken with a GOP value of one. Our simulations show that as the number of frames per GOP increases, ALD, and its packetisation and transmission techniques, can offer consistency of viewable quality for longer periods of time, while the interdependence of layers and individual frame types, i.e. I, P and B frames, can further limit the achievable quality of existing scalable streaming models, i.e. SVC and MDC. Additionally, our results show that for larger GOP values, our models can increase the consistency and quality of scalable media for all users while leveraging the benefits of overall network transmission cost reduction offered by SVC with larger GOP values (approximately 90% reduction when comparing a GOP value of 1 to a GOP value of 32). Our single frame per GOP, GOP1, subjective testing results support our simulated experimentation and illustrate the benefits of selective packetisation and improved error resistance allocation.

The remainder of the paper is organised as follows. Section 2 presents relevant background and related work. Section 3 provides an in-depth explanation of ALD. Section 4 describes our evaluation framework. Section 5 presents the simulated results for consistency of viewable quality as the number of frames per GOP increases, while Section 6 summarises the results of our High Definition evaluation. Section 7 illustrate the results of our subjective testing undertaken with a GOP value of one, and is followed by our conclusions in Section 8.

2 Background and Related Work

Generally, transmission errors are handled by two mechanisms: FEC and automatic repeat request (ARQ). Transmission control protocol (TCP) is a key transport protocol that implements an ARQ scheme to achieve reliability. In [38], Wang et al. reveal that consistent media stream quality requires a TCP throughput twice the average media bit-rate. Additionally, the reliability and flow control mechanisms of TCP can further hinder delay sensitive real-time data [17]. These issues represent serious limiting factors when the user

has constrained bandwidth and lossy links, as it is the case for mobile video. Hence, schemes adopting FEC, such as description-based encoding, are a good alternative for media transmission over lossy links or where it is desirable to minimise latency.

Several variations of the MDC concept have been offered in the research literature and four of the pertinent implementations are Forward Error Correction (*FEC*), Sub-Sample, Quantisation and Transform. We focus on FEC as it provides a means of dynamic adaptive stream encoding, low computational complexity and has attracted considerable attention in the literature [20, 16, 31]. Typically FEC can provide either systematic or non-systematic encodings. Systematic schemes encode the original symbols as part of the transmitted stream, while non-systematic schemes encode and transmit the original symbols as new symbols. Raptor codes [30] proposes that a systematic encoding, with encoded symbols interspersed among the original symbols provides a greater level of decodability.

Protected layers are then subdivided into sections that are combined to create a number of equally important *descriptions*, each including one section from each layer as shown in Figure 1b. Note the blue (dark shade) sections denote the original SVC data, while the green (light shade) sections denote the additional FEC data, thus illustrating the incremental increase in levels of FEC and marked increase in transmission cost especially for lower, prioritised, layers. It is important to note that in reality all layer sections are either a combination of FEC and original data, assuming a systematic encoding, or all FEC data, assuming a non-systematic encoding.

Several description-based streaming models have been proposed to reduce the transmission byte-cost of MDC or increase the achievable quality. These include:

- adjusting the levels of FEC, such as Enhanced Adaptive FEC [19],
- modifying the layer allocation per MDC description, such as transmitting the base layer as a separate MDC description [5],
- modifying the base layer to create two individual descriptions [40],
- encoding one or more layers of an SVC stream into various bit rates, thus generating numerous descriptions composed of differing quality streams, such as Scalable Multiple Description Coding (*SMDC*) [41, 3], and
- increasing the number of descriptions while reducing the byte allocation per description section of the SVC layer data and FEC, coupled with application-layer packetisation, such as Adaptive Layer Distribution (*ALD*) [25]

As can be seen, most of the previous work has focused on either a specific issue, such as FEC or base layer quality, or mandated that overall transmission cost increase, as can occur by introducing additional bit-rates per description. Our approach focuses on identifying the interrelationship between the various elements so as to provide a heuristic solution using all the relevant elements at once, such as examining FEC allocation, reducing byte-cost per description and providing packetisation options that mandate consistency of quality over all GOP values. Thus providing a means of increasing achievable quality, while decrease overall transmission cost.

Table 1 Notation and Definitions

N	The number of SVC layers per Group of Frame (GOP)
$L_{l,x}$	Byte-size of SVC Layer l for frame x
$S_{l,x}$	Layer section byte-size of SVC Layer l for frame x
l	Integer value corresponding to the layer number of L_l
GOP	The number of frames per GOP
STF	Section Thinning Factor
D_c	A complete description, containing sections from layers 1 to N
q	Number of MDC descriptions required to decode Layer q
$q+STF$	Number of ALD descriptions required to decode Layer q
IER	Increased Error Resilience for a given layer

3 Adaptive Layer Distribution

In this section we introduce Adaptive Layer Distribution (ALD), a novel layered media technique that optimises the trade-off between streaming bandwidth demands and error resiliency. ALD is based on the principle of layer distribution, in which the critical stream data is spread amongst an increased number of descriptions as well as over all packets thus lessening the impact on quality due to network losses. ALD has been proven to reduce transmission cost relative to MDC and provide consistent high levels of play-out quality. The proposal of ALD is motivated by two main objectives: reducing the transmission byte-cost overhead and maintaining a consistent play-out quality over lossy networks. In this context, play-out consistency refers to reducing the frequency of transitions in play-out quality due to packet losses. In order to realise these goals, ALD leverages the benefits of reduced transmission costs provided by larger GOP values, and ALD introduces the concepts of section thinning, improved error resiliency, and section-based application packetisation as detailed in the following subsections.

3.1 Section Thinning

This component provides a means of reducing the byte allocation of each layer section per description, while increasing the number of descriptions being transmitted.

3.1.1 Layer Section Allocation

As illustrated in Figure 1b the level of additional FEC data in MDC is proportionally high compared to the initial level of SVC data, thus leading to a large increase in transmission byte-cost relative to SVC. An MDC description section from layer l from frame x, $S_{l,x}$, contains $\frac{L_{l,x}}{l}$ of the layer size, while a single *complete* MDC description from frame x, as shown in Equation (1), contains the transmission cost of one section from layers 1 to N:

$$\sum_{l=1}^{N} \frac{L_{l,x}}{l} \tag{1}$$

While we view the total transmission cost of one complete description from each frame per GOP as

$$MDC\_D_c = \sum_{frame=1}^{GOP} \left(\sum_{l=1}^{N} \frac{L_{l,frame}}{l} \right) \tag{2}$$

Note that it is not sufficient to multiply a single description by the number of frames per GOP, as each frame, as well as each layer, per GOP may have differing transmission cost, hence the requirement of the summation over all frames, using the $frame$ value, and the need to determine the layer cost per description section for each frame. Also note that the number of layers per frame, and number of frame rates per GOP depends on the underlying SVC encoding. In our equations for both MDC and ALD we determine the total transmission cost based on all layers required at the maximum frame rate. If a reduction in the frame rate is necessary, then a modified version of Equation (2) would mandate an additional variable, *frameStep*, which would increment over the frames not required. The following example illustrates a frameStep of 2 which would half the frame rate. Note that the frameStep value is dependent on the governing GOP value, such that the frameStep value can never be larger than the GOP value and that the frameStep value must always be a power of 2:

$$MDC\_D_c = \sum_{frame=1, frameStep=2}^{GOP} \left(\sum_{l=1}^{N} \frac{L_{l,frame}}{l} \right) \tag{3}$$

As can be seen from Figure 1b, the number of descriptions required to view a select layer can be defined by the layer value, e.g. using Equation 1 with $N = 3$, the section size of layer three allocated to each description is a third ($\frac{L_{3,x}}{3}$), which mandates three descriptions are required to decode layer 3. Note: while the maximum viewable quality from three descriptions is layer 3, three sections from layers 4 to 6 are also received. Hence the total transmission cost of three descriptions can be defined based on the number of initial SVC layers times the number of sections per layer required to decode the requested quality level. In our example this would be six layers times 3 sections for each layer. Equation 2 defines the transmission cost of one complete description, i.e. one section from all six layers. We can define the transmission cost to view layer 3 as $MDC\_D_c * 3$. Thus the total transmission byte cost of MDC per GOP and at the maximum frame rate required to decode quality layer q can be seen as

$$MDC\_D(q) = MDC\_D_c * q \tag{4}$$

While the total FEC transmission cost overhead for MDC quality layer q can be characterised as

$$MDC\_D(q) - \sum_{frame=1}^{GOP} \left(\sum_{l=1}^{q} L_{l,frame} \right) \tag{5}$$

Note that layer l defines a specific layer within the encoding and transmission of SVC, while quality, or layer quality, q defines the viewable quality achievable by decoding a number of descriptions.

ALD section thinning is motivated to reduce the percentage of FEC data per layer, thus leading to a significant reduction in transmission cost for ALD in comparison to MDC. Section thinning reduces the byte-size of each layer section by increasing the number of ALD descriptions. The formation of the ALD sections follows the same footsteps of MDC section formation, but the section size in each scheme corresponds to a different share of the original SVC layer. In ALD, each section layer-share is scaled by an additional *section thinning factor* (STF) such that an ALD description section from layer l from frame x, $S_{l,x}$, contains $\frac{L_{l,x}}{(l+STF)}$ of the layer size. Thus a single complete ALD description from frame x, as shown in Equation (6), contains the transmission cost of one section from layers 1 to N, but each section byte-size is smaller. Thus leading to a smaller transmission byte-cost per ALD description:

$$\sum_{l=1}^{N} \frac{L_{l,x}}{(l + STF)} \tag{6}$$

While we view the total transmission cost of one complete ALD description from each frame per GOP as

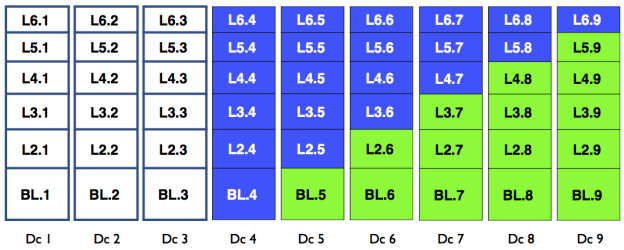


Fig. 2 ALD GOP for six-layers, with STF = 3

$$ALD\_D_c = \sum_{frame=1}^{GOP} \left(\sum_{l=1}^{N} \frac{L_{l,frame}}{(l + STF)} \right) \tag{7}$$

Similar to MDC, a frameStep variable can be used to increment over unwanted frame rates. As with MDC, the number of ALD descriptions required to view a specific quality level is based on the transmission cost of a complete ALD description, as per Equation 7, times the layer value plus the STF value. Using the same example of layer 3 as per MDC and assuming an STF of 3, we derive the following ALD_D_c* ((requested layer)+(STF)). Which equates to $ALD\_D_c * (3 + 3)$, or six complete ALD descriptions required to decode layer 3. This can be seen in Figure 2, where sections L3.1 to L3.6 are required to view layer 3. Thus the total transmission byte cost of ALD required to decode quality layer q can be written as

$$ALD\_D(q) = ALD\_D_c * (q + STF) \tag{8}$$

While the total FEC transmission cost overhead for ALD quality layer q can be characterised as

$$ALD\_D(q) - \sum_{frame=1}^{GOP} \left(\sum_{l=1}^{q} L_{l,frame} \right) \tag{9}$$

Thus, if STF > 0, the transmission cost of an ALD description is less than the cost of an MDC description (because ALD contains less FEC data), but more ALD descriptions are required to decode the same quality layer q.

It is important to note that ALD with an STF value of zero equates to the same layer section byte allocation, number of descriptions and transmission byte-cost as MDC. Thus ALD with an STF value equal to zero is exactly MDC. Figure 2 illustrates the representation of the six-layer SVC video from Figure 1a, using ALD with an STF value equal to three. As shown in the figure, each layer is further extended over the three additional descriptions in comparison to the original MDC.

There are a number of points to note when you compare MDC, Figure 1b, and ALD, Figure 2:

1. As previously highlighted, each MDC description is capable of providing base layer quality, thus mandating MDC to allocate the entire SVC base layer to each MDC description. This can be seen from Equation 1 when we specify $N = 1$, note: the base layer is the first layer and can be defined as layer 1. Hence the allocation cost of a base layer of any frame x to an MDC description is the total cost of the base layer divided by 1, i.e. the entire base layer. If we take the example in Figure 1b where 6 layers are transmitted, we can see that BL.1 from Dc 1 is the original (blue) SVC base layer, while BL.2 to BL.6, inclusive, are the additional FEC base layer sections. Thus, in this example, leading to six base layer sections being transmitted, or 600% of the original SVC base layer transmission cost. An alternative means of determining the total cost of the base layer in this example is to define the value of q in Equation 4 as 6, thus mandating the total cost of the base layer to be the cost of the original base layer $* 6$ or 600% of the original SVC base layer transmission cost.
While in Figure 2, by utilising *STF*, it can be seen that the original blue (dark) SVC base layer data is distributed over more ALD descriptions, BL.1 to BL.4 in our example, consequently reducing the byte cost of each ALD description base layer section to just 25% of the original SVC base layer. Again this can be determined for the base layer in ALD using Equation 6, where we define $N = 1$ and STF $= 3$. Hence the allocation cost of a base layer of any frame x to an ALD description is the total cost of the base layer divided by $1+3$, i.e. a quarter of the original base layer. It is important to note that the base layer section in all ALD descriptions in this example contain 25% of the base layer and not just the additional three ALD descriptions above the original quantity in MDC, i.e. in all 9 ALD descriptions and not just in the 3 additional ALD descriptions above the 6 descriptions in MDC. Finally by utilising Equation 8 we can determine the total transmission cost of the base layer using ALD. If we define q to be 6 (the maximum layer), STF to be 3 and multiply these by the percentage of the base layer in each description our answer is $(6 + 3) * 25\%$. Thus, in this example, leading to a transmission byte cost of 225% of the original SVC base layer transmission cost, or approximately 38% of the MDC base layer transmission cost. Note that the additional ALD descriptions are shown in white, to illustrate a visual comparison in number of descriptions required by ALD, nine, and MDC, six.
Once this mechanism for section thinning is applied to each layer in the transmitted stream, the transmission byte-cost of ALD is less than MDC. It can be seen that the original blue (dark) SVC data for each layer is shared over more ALD descriptions than MDC descriptions (ex-

cluding the highest layer in both schemes where no FEC occurs), thus leading to a reduction in transmission byte-cost, irrespective of encoding rate.
2. The number of FEC sections per layer is consistent between MDC and ALD, but the FEC section byte-allocation in ALD is smaller.
3. A greater number of ALD description, four from Figure 2, are required before base layer decoding is achievable. For a device that only needs to view at low-quality this has implications in terms of having to receive more descriptions than with MDC. This is discussed in the next section.

So clearly, the optimal choice of STF is an important design issue that will be introduced later in this paper.

3.1.2 Quality Transmission Cost

Generally, multiple users may be interested in viewing the same video at different qualities, depending on several factors such as the available resources and device capabilities. Section thinning realises significant savings for users interested in receiving high quality video. On the contrary, if a user is interested in receiving low video quality, ALD may result in a larger overhead in comparison to MDC, as additional (STF) ALD-descriptions have to be received in order to decode the base layer. As previously defined, only q MDC descriptions, Equation (4), are required to decode quality layer q in comparison to $(q + STF)$ ALD descriptions, Equation (8), to realise the same video quality.

Hence, the difference in the amount of transmitted data, or total relative overhead $D(q)$ per GOP, for a single user between ALD and MDC for video quality q, can be calculated as

$$D(q) = ALD\_D(q) - MDC\_D(q) \tag{10}$$

Note that negative total overhead implies that ALD is more bandwidth efficient than MDC for the selected quality level q. Future work will investigate mechanisms to reduce the transmission byte-cost increase for lower layer streaming.

3.1.3 Optimal STF Selection

As previously mentioned, multiple users may be interested in viewing the same video at different qualities, thus ALD provides a mechanism for optimal STF in streaming scenarios for both unicast, single user with one quality requirement, and multicast [9], numerous users with possibly differing requirements. Multicast provides two options for ALD transmission:

i) Each quality layer q is transmitted as a separate entity, thus implementing a multi-bitrate scheme (this option overly increases transmission cost)
ii) Each ALD description is transmitted as a single multicast stream, thus allowing users to subscribe to (q + STF) descriptions to receive the required q quality layer (this option reduces transmission cost, as only the maximum requested quality layer, (max[q] + STF) descriptions, are transmitted thus permitting multiple users access the same descriptions for their respective q' quality layers).

Let p_q denote the percentage of clients interested in viewing a video with quality level q. In a unicast scenario, this would be based on the requirements of a single user, while in multicast, would consider the needs of numerous users and their varied demands. Thus, the expected total overhead can be estimated as

$$E\{D(q)\} = \sum_{q=1}^{N} p_q D(q). \tag{11}$$

In our design, we choose an $\overline{STF_O}$ value that minimises the expected total overhead and can be expressed as

$$\overline{STF_O} = \arg\min_{STF} E\{D(q)\} \tag{12}$$

Note that the optimal $\overline{STF_O}$ would vary depending on different factors including the number of layers and the size of each layer.

3.2 Improved Error Resiliency (IER)

The main objective of the IER component is to enhance the streaming quality by ensuring a smooth play-out with fewer quality transitions. Clearly, the FEC overhead of higher layers in MDC is inversely proportional to the layer-level. For the top-most layer, no FEC is considered. Hence, the loss of any MDC description results in an immediate downgrading of the stream quality. Similarly, further proportional reductions in the stream quality for the same GOF is dependent on the cumulative loss of additional descriptions.

IER reduces the number of non-redundant sections of higher layers by distributing the higher layer data over a number of reduced sections allowing for one or more additional FEC sections. IER can be applied to any layer, or number of distinct layers, where additional error resiliency is required. However, it is typically applied to the top-most layers to reduce the incurred FEC overhead.

An example shall be used to illustrate the concept of IER. MDC in Figure 1b consists of 6 descriptions, where

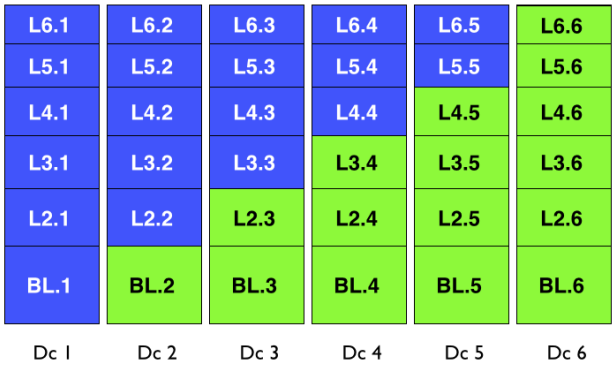


Fig. 3 One section of IER allocated to Layer 6 of MDC from Figure 1b

each description contains a segment from each SVC layer. Each SVC layer is distributed over the MDC descriptions, as per Equation 2, using $\frac{L_{l,x}}{l}$, where l denotes the SVC layer index and the remaining MDC sections are populated with FEC data for the respective layers. The SVC layer 6 is distributed over all six descriptions, $\frac{L_{6,x}}{6}$, and does not contain FEC, as such any packet loss will reduce viewable quality. To counteract this reduction in quality, we will improve the error resiliency of layer 6 by providing one section of IER, denoted as IER-1. This is accomplished by distributing the layer 6 data over five descriptions, $\frac{L_{6,x}}{5}$. Determining the reduced distribution of the SVC data provided by IER is undertaken during the initial SVC partitioning, prior to FEC allocation. The remaining layer 6 section is then populated with one FEC section. Thus IER mandates an increase in transmission cost as well as providing increased error resiliency. Figure 3 illustrates the final compositions of the modified MDC description structure.

Based on Equation 1 for SVC layer distribution to an MDC description structure, the reduction in divisor provided by IER for a specific layer in MDC can be generalised to

$$\frac{L_{l,x}}{l - IER} \tag{13}$$

While the allocation of IER to a specific layer in ALD can be generalised to

$$\frac{L_{l,x}}{l + STF - IER} \tag{14}$$

If IER is zero, then these equation defaults to their standard distribution. Assuming the description structure of MDC in Figure 1b, where the base layer is repeated in every description, IER can not be allocating to the base layer. The layer index must be larger than the level of IER being allocated to the specific layer, i.e. for layer 2, a maximum of one additional section of IER can be allocated, while for layer 6, a maximum of five additional section of IER can be allocated. IER must be a positive integer and IER must be smaller than l for any given layer. This does not mandate the maximum levels must be imposed, but that a maximum level exists that can not be exceeded. For ALD the level of IER available increases based on the level of STF. So for layer 2 in Figure 2, IER mandates that a maximum of four additional FEC sections can be allocated.

Finally, there is no optimal level of IER to implement by default. The choice of layer and the level of IER is user or provider specific and may reflect loss rates within the network or the prioritisation of a specific layer within the encoding hierarchy. Figure 3 can be viewed as an example where maintaining the quality of the maximum layer is important. As stated the level of IER required is dependent on the level of network loss and in this example layer 6 can incur approximately 16% packet loss prior to a degradation in viewable quality. The 16% threshold is determined based on the additional FEC section. Of the six sections of layer 6 transmitted only five sections are required, thus $\frac{1}{6}$, or 16%, of the transmitted data for layer 6 can be lost before layer 6 is undecodable. As each lower layer in Figure 3 contains either an equal amount, i.e. layer 5, or higher levels of resiliency, 16% of the transmitted stream can be lost before there is a reduction in viewable quality. This *loss rate* threshold over all transmitted data is achieved due to the packetisation options presented in the next section.

3.3 Section Packetisation

This component reduces the impact of packet loss on any description-based scalable video, such as MDC and ALD. The application transmission unit for MDC is its description. For purpose of illustration, we use a single GOP example from the widely-used video clip known as crew.yuv, encoded as a six-layer SVC stream. Table 2 shows the bytesize of each layer for the selected frame.

Table 2 GOP SVC Layer sizes

Layer	1	2	3	4	5	6
Layer Size	1440	1577	1601	1546	1255	3372

In today's Internet, the maximum packet size observed is usually limited by that of the Ethernet frame, which has a maximum payload of 1,500 bytes. We assume a packet payload of 1,440 bytes, allowing for overhead due to headers of network, transport and streaming media protocols. We assume that the GOP frames are transmitted over Ethernet packets. On transmitting this frame, eleven Ethernet packets would be required using SVC where the transmission unit is an individual layer. The same frame would require eighteen packets when encoded using MDC in which the description represents the application transmission unit. On losing any

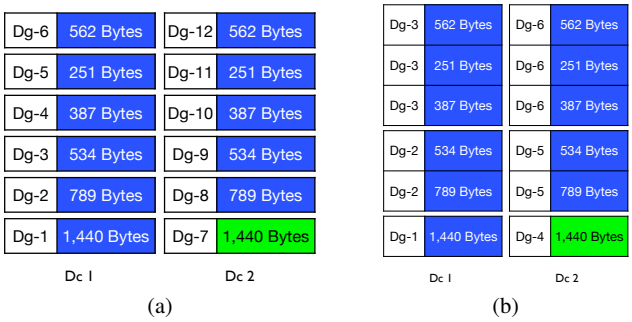


Fig. 4 An example of (a) MDC-SDP Option 1 - with two descriptions (D_c) consisting of six packets (D_g) (b) MDC-SDP Option 2 - with two descriptions (D_c) consisting of three packets (D_g). Note for each option only two of the six packetised descriptions are illustrated

of these packets, the application would not be able to decode the *entire* frame to the highest quality. In order to reduce the impact of losses on the stream quality, we propose to use two packetisation mechanisms, called section-based description packetisation and section distribution.

3.3.1 Section-Based Description Packetisation

With section-based description packetisation (SDP), we propose using sections as application transmission units instead of the entire description for description-based layered coding techniques such as MDC and ALD. As a consequence the description is decomposed of a number of sections, with each description section transmitted as a single unit, thus limiting the affects of packet loss to individual section while allowing partial description re-use. Partial description re-use in this instance means the availability at the device of one or more layer sections from a single description. The probability of loss affecting all sections from a single description, or all sections from a single layer, is low, while the probability of partial description re-use is high.

SDP improves the possibility of higher stream quality by mitigating lower layer loss thus increasing the availability of a sufficient number of lower layer sections.

SDP can be applied in several ways as follows:

- *Option 1 - Individual layer sections* - this option transmits each layer section as a separate group of one or more packets. This option may increase the number of packets being transmitted, depending on the original encoding but maximises the number of sections available during decoding. Using the example frame, it can be seen that for each MDC description six packets are required for transmission as shown in Figure 4a. This option increases the number of packets and in some instances creates packets not containing a full data payload. Consequently the overhead due to packet headers and processing is higher.
- *Option 2 - Minimising packets quantity* - this option groups layer sections together to fully occupy each transmitted packet, thus mitigating the problems with Option 1. Figure 4b illustrates this option for the example frame. This option reduces the number of transmitted packets. However, the loss of a packet may cause the loss of numerous layer sections. To reduce the probability of stream degradation due to packet loss, only one section for each layer should be included within a packet. If a packet were to contain numerous sections for one specific layer, then the loss of that specific packet may aversely affect the decoding of that layer and all enhanced layers that rely upon it. It can be seen that for each description, D_c, three packets, D_g, are required for transmission.

It is worth noting that the blue (dark) sections are the critical SVC data and the green (light) sections the FEC section allocation. It can be seen that in Figure 4a and 4b that the base layer consumes a single packet, D_g-1 in description one; in Figure 4a each section is allocated to an individual packet while in Figure 4b, a section from layer two and three are allocated to D_g-2 and a section from layer four, five and six are allocated to D_g-3. As six descriptions are transmitted, a total of thirty six packets are transmitted over the network with Option 1 in which only twenty one specific packets are required for maximum stream quality. In Option 2, eighteen packets are transmitted among which only ten specific packets are required for maximum stream quality. Thus Option 1 increases the probability of maximising stream quality in the presence of high levels of packet loss.

When the underlying bitrate of the GOP is low, which can occur with low quality clips or when there is a large number of frames per GOP (due to the increased frame interdependency) the total transmission cost of a single description can be less than the underlying packet payload. When this occurs the loss of a packet mandates the loss of a complete description which can lead to noticeable variation in the viewable quality. To this end, we also define an $\overline{STF_E}$ value that maintains a level of error resilience per description and can be expressed as a lower bound on the number of packets per description. For example in our evaluation results, $\overline{STF_E}$ is chosen such that a minimum of two packets are packetised for each description. Hence, the loss of one of these packets would not completely affect an entire description and as such would sustain high levels of video quality. Hence, the chosen $\overline{STF}$ value can be defined as

$$\overline{STF} = \min\{\overline{STF_O}, \overline{STF_E}\} \tag{15}$$

As with IER there is no default value for $\overline{STF_E}$, the granularity in the minimum number of packets mandated per description is dependent on the bitrate of the stream and

Fig. 5 SD packetisation of D_c 5 from ALD in Figure 2. It can be seen that each packet contains section segments from all layers (red denotes packet header)

the levels of network loss. However there is a trade off between the improvement in viewable quality mandated by the increase in the number of packets and the elevation in transmission cost due to the greater number of packet headers.

3.3.2 Section Distribution

Due to the transient nature of the Internet, network traffic can be affected by both *individual* and *burst loss* states corresponding to a single packet loss or numerous consecutive packet losses. Packet losses at lower layers have a negative impact on scalable video due to inter-layer dependency. As shown above, by manipulating the stream packetisation, we can increase stream quality and consistency. With this in mind, we propose Section Distribution (SD), a mechanism to further distribute the description sections over the packets used to transmit a description to further reduce the impact of losing critical sections. SD is beneficial for any description-based streaming model, such as MDC and ALD. SD extends the benefit provided by equally important descriptions to the packet level per frame. SD is utilised to distribute each section per description over a number of packets, thus limiting packet loss to only a segment of each section. SD is consistent with the well-known Interleaving [11] technique, which is widely used to combat the effect of burst loss.

We first determine the number of packets, denoted as R, required to transmit a single description for each frame per GOP by performing summation over the GOP value. This is achieved by dividing Equation (2), MDC, or Equation (7), ALD, by the data byte-size of a packet payload. We shall use ALD as an example

$$R = \frac{ALD\_D_c}{packet\ payload} \tag{16}$$

Each layer section per frame per GOP, denoted as $S_{l,frame}$, is spread over the R packets by allocating a subsection of each layer from each frame per GOP to a single packet, P_k, as per the following

$$P_k = \sum_{frame=1}^{GOP} \left(\sum_{l=1}^{N} \frac{S_{l,frame}}{R} \right) \tag{17}$$

Hence, a packet would carry subsections of different layers, as illustrated in Figure 5. In this manner, all packets per GOP are of *equal priority*, as each packet contains the same byte-size, i.e. quantity, of each layer per frame per GOP. Thus the loss of an individual packet, will result in a partial loss from each layer. Thus the quality of the packetised stream is limited only by the percentage of lost packets rather than the specific carried description or layer. Additionally, the probability of losing critical sections is reduced since lower layers enjoy greater redundancy.

Furthermore, on using section distribution, packets per frame would be identical in size and content, thus providing *packet equality*. This equality is provided in both packet byte-size and packet priority. Also as the GOP value increases, then SD will provide data equality for all frames within the GOP. In [22], the authors highlight that packets of dissimilar processing times, produce dissimilar transmission times. Such that by maintaining such packet byte-size equality, the order of packet delivery is improved. Thus SD packet equality results in a consistent delivery in network transmission. In our GOP evaluation, we combine both the SD and SP components with MDC to illustrate simple mechanisms to increase viewable quality, while not increasing transmission cost. It is important to note that packet equality may mandate a minor increase in transmission cost, as some byte rounding up may occur when subsections are divided by R.

Also as each packet now contains a subsection of each layer, i.e. subsections of NALs rather than a NAL for a specific layer as defined by SVC, the subdivision of each packet, i.e. the specific bytes for each layer subsection, per GOP must be identified to the receiving decoder. Possible options to provide this information are

i. For each GOP, provide a file which details the structure of each packet for each GOP or for all GOPs in the stream, similar in structure to an media streaming manifest file, e.g. DASH [34]. As each packet per GOP contains the same structure only one manifest file is required per GOP. An issue with this option, is that the GOP manifest file may be lost during transmission. One manner in which to reduce this issue is to provide the manifest file during stream setup, thus removing the issue of manifest loss during stream delivery.
ii. Include the packet structure as an additional header within each GOP packet. An issue with this option, is 1) an increase in overall transmission cost as each packet per GOP will contain the header information and 2) repetition, as the additional header is the same in each packet per GOP.
iii. By utilising R and the byte cost of decoding the base layer, $ALD_D(1)$, we can determine the minimum number of packets, $\min(P_k)$, required per GOP to decode the lowest layer. If we divide the byte size of a single instance of the additional header outlined in item ii. by $\min(P_k)$, we can determine the minimum additional byte allocation per packet required by SD so as to determine the structure of each packet per GOP once the minimum number of packets required by the base

layer have been received. FEC is utilised to extend this minimum byte size over all packets per GOP. The reason $ALD_D(1)$ is utilised to determine $\min(P_k)$, is that a lower number of packets will not permit decoding of the base layer, so the manifest file is not required, while an increase in packets may provide an increase in viewable quality and the structure of the layer subsections per packet is required for all layers, i.e. BL to N.

The SDP and SD claims made in this section are generally applicable to videos that have different number of SVC layers, as highlighted later in the evaluation section, where the chosen layer size has been increased to eight.

3.3.3 Transmission Unit Stream Quality Loss Rate

In Table 3, we show the transmission cost, in terms of bytes, for SVC, MDC, both options of MDC-SDP and by utilising Eq (15) ALD with an STF of 3. Thus increasing the number of ALD descriptions to nine. Table 3 also presents the number of packets per frame and highlights the best case (B-C) and worst case (W-C) maximum viewable layer based on the loss of a specific number of packets. It is worth noting that the SVC data transmission byte-cost for all versions of MDC are equal, but the total transmission byte-cost for each scheme will vary, dependent on the number of packets being transmitted and the increased byte-cost of packet headers. In this section to provide a simplified example, we evaluate the SVC data element only.

As the number of lost packets increases, and dependent on which packet is lost, the quality of the stream can remain high or degrade significantly. As can be seen, SVC is severely affected by packet loss. The worst case (W-C) for all four lost packets, highlights the loss of a packet from the base layer, while the best case (B-C) is based on consecutive losses from the highest quality layer down, i.e. layer six is composed of three packets, such that B-C will remain at quality level layer 5 until the fourth packet is lost, when the quality reduces to quality layer four.

MDC-FEC is similar in that each description is composed of three packets, such that the B-C remains consistent over three packet losses, and reduces quality to layer four when the fourth packet is lost. W-C is based on the loss of a single packet from distinct descriptions, thus incrementally reducing quality for each additional packet lost. The increase in viewable quality is consistent with the level of additional error resilience added to the original SVC data, but this increase in viewable quality requires an additional approximately 13,000 bytes of transmission bandwidth.

Consistent with MDC-FEC, both options of MDC-SDP achieving the same W-C viewable quality, again based on a single lost packet from distinct descriptions. Both options of MDC-SDP achieve the maximum B-C over all four lost packets, as loss can be confined to the green FEC section packets. Thus highlighting the benefits offered by section based description packetisation.

Table 4 QP value per layer for all clip types and GOP values

Resolution	QCIF		CIF			4CIF		
Layer	BL	2	3	4	5	6	7	8
QP Value	34	28	33	30	28	35	32	30

As previously stated, ALD employs the section distribution (*SD*) technique for packet packetisation, thus achieving packet equality. As highlighted in the Table 3, this equality produces a uniformity in the B-C and W-C achieved by ALD. As each of the nine ALD descriptions is composed of two packets, achievable viewable quality is incrementally reduced once two additional packets are lost.

A loss rate of six packet is illustrated to highlight that with the loss of six packets, the transmission cost of ALD over the network is less than the transmission cost of SVC with no packet loss. This offers a comparison of the B-C and W-C quality achieved by SVC and ALD for similar transmission byte-cost. It is important to note that while the B-C of ALD is less than SVC, the W-C of ALD is better, thus highlighting the balance offered by ALD between transmission cost and achievable consistent quality.

Note that by implementing the previously highlighted IER technique on the highest layer, layer six. The viewable quality layer value for both B-C and W-C achieved by ALD-IER for the loss of one or two packets is six. Thus maximum quality can be achieved for a very minor increase in transmission byte-cost, 47 bytes per ALD description.

4 Evaluation Framework

In this Section, we present our performance evaluation framework for our GOP evaluation and our subjective testing. Our GOP evaluation is based on the well-known 10 second *city* video, an aerial view of a building landscape, while our Subjective Testing utilises the *crew*, *city*, *harbour* and *soccer* videos, all obtained from the Leibniz Universität Hannover video library [37]. These videos are recorded at 30 frames per second totalling 300 frames per video. The videos are encoded using JSVM [26] to eight layers with spatial and quality scalability, using medium grain scalability (*mgs*) and quantizer parameter (*QP*) values as per Table 4.

As illustrated, we consider three resolutions (QCIF, CIF and 4CIF) with respective 2, 3, and 3 quality levels, i.e. two fidelity levels in the lowest resolution and three fidelity levels in each of the higher resolutions. For larger GOP values, we use the same *QP* values for all encodings and only vary the GOP value. The QP values in Table 4 provide a means of demonstrating how the bitrate of the encoded clip and associated GOP value can mandate variation in the maximum achievable quality of the individual SVC layers. Ta-

Table 3 Example transmission byte-costs for SVC, MDC, MDC-SDP (both options) with viewable quality as packet loss increases

Scheme	SVC	MDC-FEC	MDC-SDP opt1	MDC-SDP opt2	ALD
Transmission Cost	10,793	23,790	23,790	23,790	15,273
# of Packets	11	18	36	18	18
One Lost Pk (B-C / W-C)	5 / 0	5 / 5	6 / 5	6 / 5	5 / 5
Two Lost Pk (B-C / W-C)	5 / 0	5 / 4	6 / 4	6 / 4	5 / 5
Three Lost Pk (B-C / W-C)	5 / 0	5 / 3	6 / 3	6 / 3	4 / 4
Four Lost Pk (B-C / W-C)	4 / 0	4 / 2	6 / 2	6 / 2	4 / 4
...	...	...	...	...	...
Six Lost Pk (B-C / W-C)	3 / 0	4 / 0	6 / 0	6 / 0	3 / 3

Table 5 Maximum achievable PSNR value (dB) per clip type for layer 8 with a GOP value of one

Layer	City	Crew	Harbour	Soccer
PSNR	36.7	38.75	37.02	38.03

ble 5 highlights the changes in maximum achievable PSNR for layer eight for each of the clip types with a GOP value of one. In [13], the authors define that a typical choice of QP values in AVC and HEVC encodings to be 22, 27, 32 and 37 based on the software reference configuration specified by [4]. While these QP values are sufficient when comparing clips of a defined quality and a single resolution, i.e. clips containing only a single layer, for scalable video a separate QP value is required for each individual layer in the SVC encoding so as to determine a quality level for each layer. The QP values utilised in our encodings provide a means of mandating that the lower the layer value, the lower the underlying bitrate of that specific layer, i.e. the base layer will have the lowest bitrate, layer eight will have the highest bitrate and the layers in-between will have incrementally higher bitrates. This provides for a gradual increase in transmission cost as the viewable quality increases.

For GOP we evaluate six streaming models, SVC, MDC, MDC-SDP option 1, furthermore just referred to as MDC-SDP, MDC-SD (where MDC description data is packetised using section distribution), ALD and ALD-IER-2 (ALD with one additional FEC section for the two highest layers, L7 and L8, thus providing increased protection for the maximum viewable quality), over four distinct GOP values, i.e. number of frames per GOP, namely 1, 8, 16 and 32. While for our Subjective Testing, we also evaluate an additional model called Scalable Description Coding (*SDC*) [24], a previously published streaming model that we designed to mimic the benefits of both SVC and MDC. SDC modifies MDC by creating a low priority scalable description composed of a subset of only the higher layer sections. Thus achieving lower transmission costs, by removing the redundancy of lower layer FEC sections, and increased quality by adding an additional description either composed entirely of FEC sections or composed of XoR data from both the scalable and standard descriptions.

The transmission of the encoded videos is simulated in Network Simulator 2 (*ns-2*) [35] using myEvalSVC [15], an open source tool for evaluating JSVM video traces for SVC. myEvalSVC presents a means of dynamically determining bitrates, based on the JSVM trace data, and simulating real-time packetisation, over a lossy network, in ns-2. Modifications are made to myEvalSVC scripts to simulate MDC, ALD and their respective variants.

In our modified evaluation scripts, we packetise the various models based on their respective encapsulation unit, e.g. layer, description, section, thus providing clear distinction between the various units during transport. In this manner the loss in one unit will not effect the quality achievable from any other unit. For SVC, each individual layer per frame is partitioned in one or more packets. With MDC each description per frame is packetised separately. For MDC-SDP each section per layer is packetised individually. While in ALD, ALD-IER-2 and MDC-SD, each packet contains a segment of each layer per description (using SD). In ALD, ALD-IER-2 and MDC-SD, this would lead to a segment from every layer per packet. As can be seen once we begin to control the structure of the packetisation we can reduce the interdependence of the units right down to packet level. Thus lessing the effects of network loss on viewable quality.

For ALD and MDC-SD, as GOP value increases, SD mandates that each packet transmitted contains not only a segment of each layer per description per frame, but also a segment of each layer per description from each frame per GOP, thus mandating packet equality for all frames per GOP.

A two-node, server/client, model is utilised for the simulated topology in which we vary the packet error rate, μ, from 1% to 10% to test the streaming performance of different schemes over lossy links. We use an ns-2 Errormodel to define a total packet error rate with a uniform distribution. This defines that the total stream loss shall be equal to μ, but does not mandate that the individual frame loss rate shall also be equal to μ, thus permitting bursty loss during simulation.

Table 6 Transmission megabyte-cost for the city clip at quality layer 8 for each adaptive scheme, for each of GOP values and maximum achievable PSNR for Layer 8. All MDC variants, MDC, MDC-SDP and MDC-SD, have the same transmission cost, so only MDC is shown

Layer	PSNR	SVC	MDC	ALD	ALD-IER-2
GOP1	36.7	18.49	32.79	24.90	25.75
GOP8	35.5	4.56	7.79	6.05	6.25
GOP16	35.0	2.83	4.91	3.79	3.90
GOP32	34.5	1.93	3.41	2.61	2.68

So as to provide randomised loss rates per frame in our simulated experimentation, for each of the streaming models extensive simulations are run to create the ns-2 output traces, which are analysed to determine the average maximum stream quality per-frame at the client. Each trace is then saved as an achievable quality (*AQ*) trace file for each streaming scheme. The AQ trace files are utilised to 1) to provide a means of illustrating the transition in frame quality over time and 2) to create the modified YUV files, based on the maximum stream quality per frame, from the original YUV files. In our results, for each model we determine the maximum stream quality per frame based on the highest layer that contains no packet loss, thus containing no impairments that are visually observed. Some pixelation (upscaling of low quality resolutions thus creating noticeable square shaped single-colour display components on the screen) may occur when the resolution of the maximum achievable quality is less than the maximum viewable resolution. myEvalSVC and JSVM do not contain a reliable mechanism for this form of YUV modification, so a new program, modPSNR.exe, is created based on the original JSVM source code. modPSNR.exe supports basic error concealment by which non decodable frames are substituted by duplicating the previous frame. Finally JSVM is utilised to ascertain the PSNR [27] value of the modified YUV, in comparison to the original YUV file.

5 Evaluation Results for a varying number of frames per GOP

consistency

The purpose of evaluating an increase in the number of frames per GOP is to determine the effects of inter-frame dependency on viewable quality for scalable video. As the GOP value increases, the overall transmission cost is reduced but the inter-frame dependency increases. This increase typically affects the viewable quality. The developed techniques in ALD benefit from the reduced transmission cost of larger GOP values while maintaining consistent levels of viewable quality.

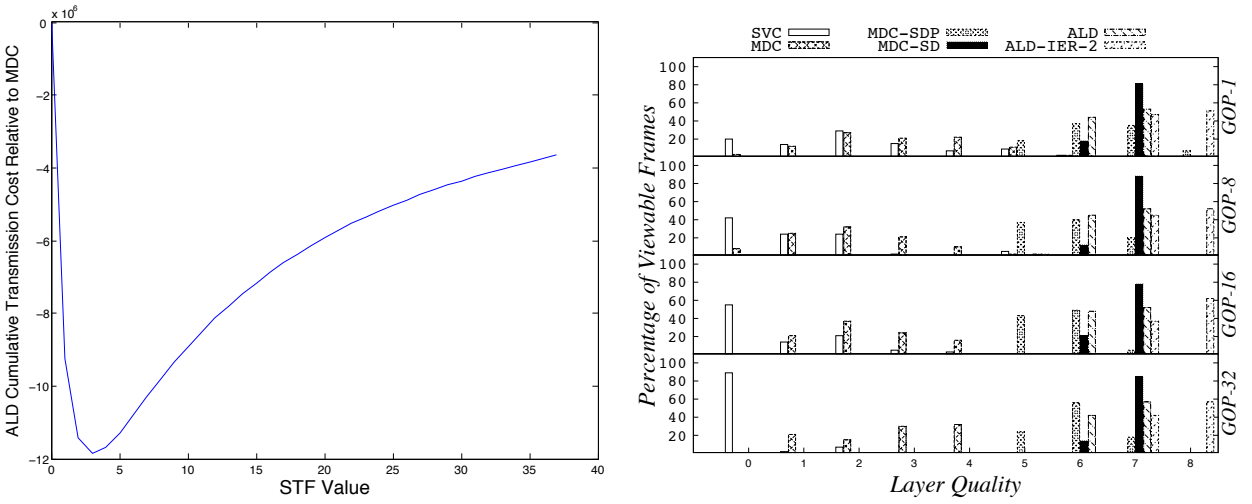
As GOP increases, as illustrated in Table 6, the transmission cost of MDC and ALD changes. As STF is based on the cumulative transportation cost of ALD relative to MDC, this has the potential to create different STF values for differing GOP values. For our evaluation, this created STF values of 3 for a GOP of one and a GOP of thirty two, and an STF value of 2 for a GOP of eight and a GOP of sixteen. To provide consistency of STF value used in the evaluation over all GOP values, we use the same value of STF, i.e. 3, as defined for a GOP of one, for all ALD and ALD-IER-2 simulations. The increase in STF value for a GOP of eight and a GOP of sixteen reduces their overall transmission cost by 294Kb and 189Kb respectively, by reducing their levels of FEC allocation. The STF is defined as per the developed optimisation framework shown in Section 3. Figure 6a provides the optimal STF value for the city clip type with a GOP of one. For each GOP value we evaluated packet loss rates from 1% to 10%. Due to page limitations, we only illustrate results for a 10% packet loss rate, but evaluation results for packet loss rates from 1% to 9% provided similar conclusions.

Table 6 displays the transmission megabyte-cost of layer 8 for each streaming model, for each of GOP values and maximum achievable PSNR for Layer 8. All MDC variants, MDC, MDC-SDP and MDC-SD, have the same transmission cost, so only MDC is shown. Note the approximate 90% decrease in transmission cost between GOP1 and GOP32. Thus illustrating the benefits provided by a higher number of frames per GOP, in scenarios where congestion and large burst loss may occur. Also note that mandating the same QP and encoding values, maximum achievable PSNR decreases, illustrating the link between encoding, transmission cost and viewable quality.

Figure 6b plots the percentage of viewable frames for each of the six streaming models for the city clip with a packet loss rate of 10%. Each plot illustrates a different GOP value. Higher quality is illustrated by larger percentage values in the higher layers. Note how

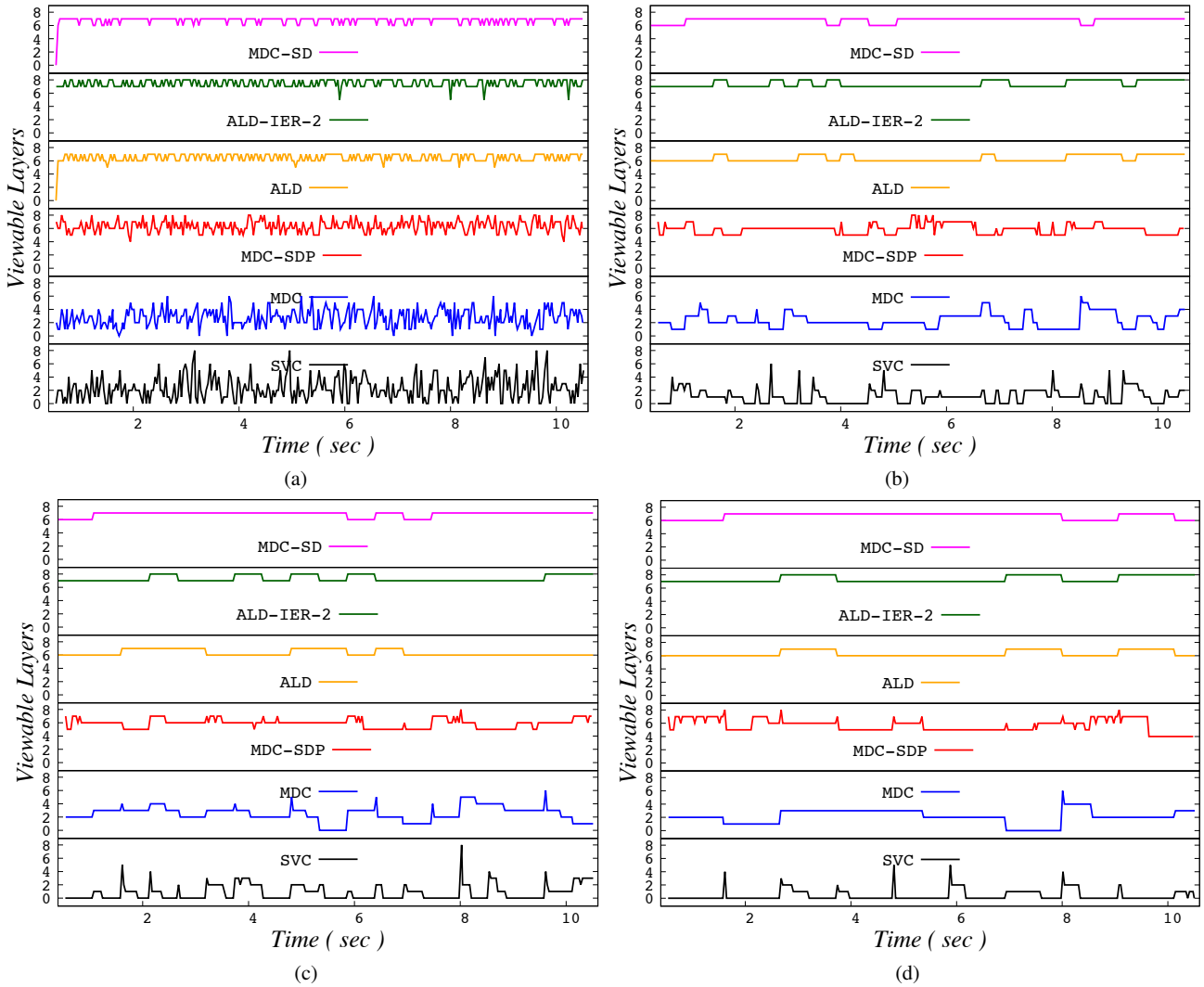
i. only MDC-SD, ALD and ALD-IER-2 provide the same approximate percentage rates for the higher layer values for each of the GOP values, thus providing consistency of higher quality decoding as GOP values increase.
ii. only ALD-IER-2 provides this consistency of higher quality decoding at the highest level, layer 8, as the GOP value increases.
iii. once SVC is encoded with 32 frames per GOP, over eighty percent of the frames are undecodable due to packet loss.
iv. the simple packetisation options of SD and SDP greatly increase the viewable quality of MDC, without increasing MDC data transmission cost.
v. the decodable quality of ALD-IER-2 never drops lower than layer 7.

Figure 7 plots ten-second examples of the stream quality transitions for each streaming model of the city clip with a GOP value of 1 (Figure 7a), 8 (Figure 7b), 16 (Figure 7c)



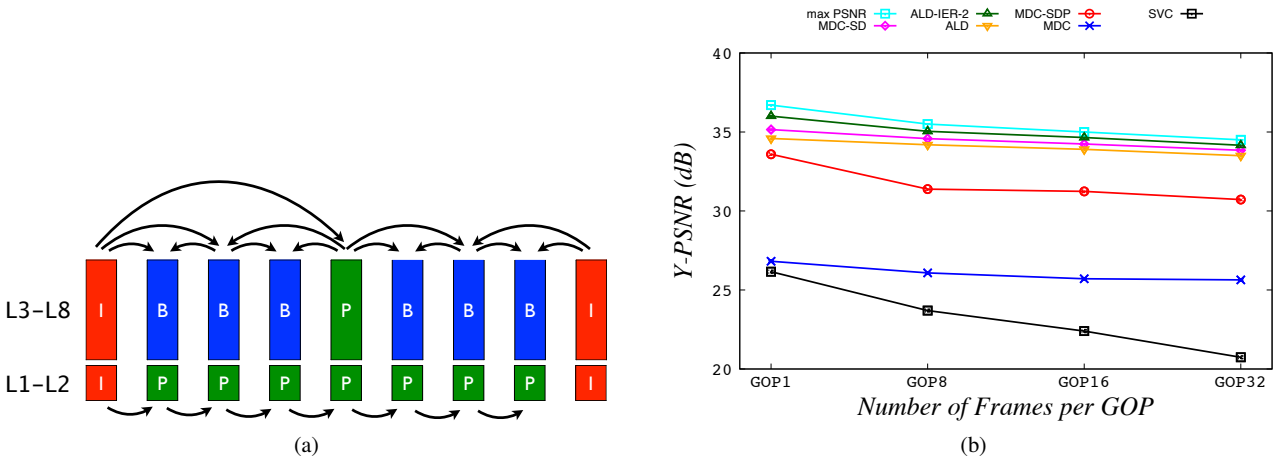
(a) (b)

Fig. 6 (a) City stream with a determined STF_o value of 3 for a GOP of one and (b) an example of the percentage of viewable frames, with a 10% packet loss rate, for each of the six streaming models for each of the four GOP values for the city clip



(a) (b) (c) (d)

Fig. 7 Ten second examples of the stream quality transitions for each streaming model of the city clip with a GOP value of 1 (a), 8 (b), 16 (c) and 32 (d) with a 10% packet loss rate



(a) (b)

Fig. 8 (a) Inter-frame dependency for our 8 frame GOP encoding and (b) a plot illustrates maximum achievable PSNR, shown as "max PSNR", and PSNR values for each of the streaming models with a 10% packet loss rate, as GOP increases, for the city clip

and 32 (Figure 7d) with a 10% packet loss rate. For each value of GOP it can be seen that SVC and MDC feature the highest frequency of variation and as such would provide a media stream with frequent variation in video quality. The other models contain less variation and more importantly these variations are limited to the higher quality layers, thus mandating higher achievable video quality. The plots also illustrate how a simple mechanism which re-packetises MDC data (MDC-SDP mandating section packetisation and MDC-SD where each packet contains a segment of each layer per description) can produce such considerable increases in viewable quality at no increase in transmission cost. Note how as GOP increases, the detrimental effects of inter-frame dependency decreases achievable stream quality for some of the streaming models, i.e. SVC, MDC and to some extent MDC-SDP. Furthermore for ALD, ALD-IER-2 and MDC-SD, the minimum transitions that can occur is consistent with the number of frames per GOP, e.g. for GOP32, the minimum number of frames for a given layer is 32. Finally, as ALD and MDC-SD do not contain FEC error resilience on the maximum layer, i.e. layer 8, only ALD-IER-2 can provide maximum achievable quality and in the plots for GOP16 to GOP32, ALD-IER-2 only varies between the highest two layer, i.e. Layer 7 and 8.

Figure 8a illustrated the frame interdependency of our 8 frame GOP encoding. For layers one and two, the JSVM encoding creates P frames for each frame, while for layers three to eight, the encoding implements a *IBBBPBBB* design. The arrows in Figure 8a present the frame interdependency, with the arrow point denoting the dependent frame. As can be seen, the loss or a low quality value of an I or P frame will mandate low quality streaming for all frames which are dependent on it. GOP16 and GOP32 contain the same structure of one I and one P frame per GOP.

Figure 8b illustrates the maximum achievable PSNR, shown as "max PSNR", and the changes in PSNR value for each of the streaming models with a 10% packet loss rate, as GOP increases for the *city* clip. PSNR provides a numerical representation of the achievable viewable quality of a model. As can be seen, the streaming models that deliver the highest layers from Figure 8a, achieve the highest PSNR values. SVC and MDC provide low quality overall. As was seen in Figure 6b, over 200 frames of SVC were undecodable with a GOP of 32, thus the evaluated PSNR value is primarily composed of duplicated frames with low fidelity. It is only the minor changes in background imagery, that mandate such a high PSNR value for SVC with a GOP of 32. MDC-SDP provides an increase in dB, relative to MDC and SVC, of between 6dB (GOP1) and 10dB (GOP32). ALD and MDC-SD provide a further noticeable increase in viewable quality, while ALD-IER-2 provides near maximum achievable PSNR for all GOP values. Thus the evaluation of a higher number of frames per GOP illustrates the benefits of selective packetisation, improved error resistance and adaptive FEC allocation provided by our techniques.

Three additional video streams, *crew*, *harbour* and *soccer*, were also assessed over all GOP values, using the same evaluation framework as *city*. To further confirms the ability of the ALD framework to realise similar gains for different videos, we present a sample of the results for the crew and soccer clips. Figure 9a presents an example of the percentage of viewable frames for each of the six streaming models for each of the four GOP values for the crew clip while Figure 9b illustrates a ten second example of the stream quality transitions for each streaming model of the crew clip with a GOP value of 32. Both figures with a 10% packet loss rate. Figure 10a and Figure 10b provides the same plots for the soccer clip. It can be seen that the results presented for crew in Figure 9b and for soccer Figure 10b are consistent with the results seen for city in Figure 7d. MDC-SD and ALD are viewable in layers 7, 6 and 5, ALD-IER-2 in layers 7 and 8, with the other streaming models containing large

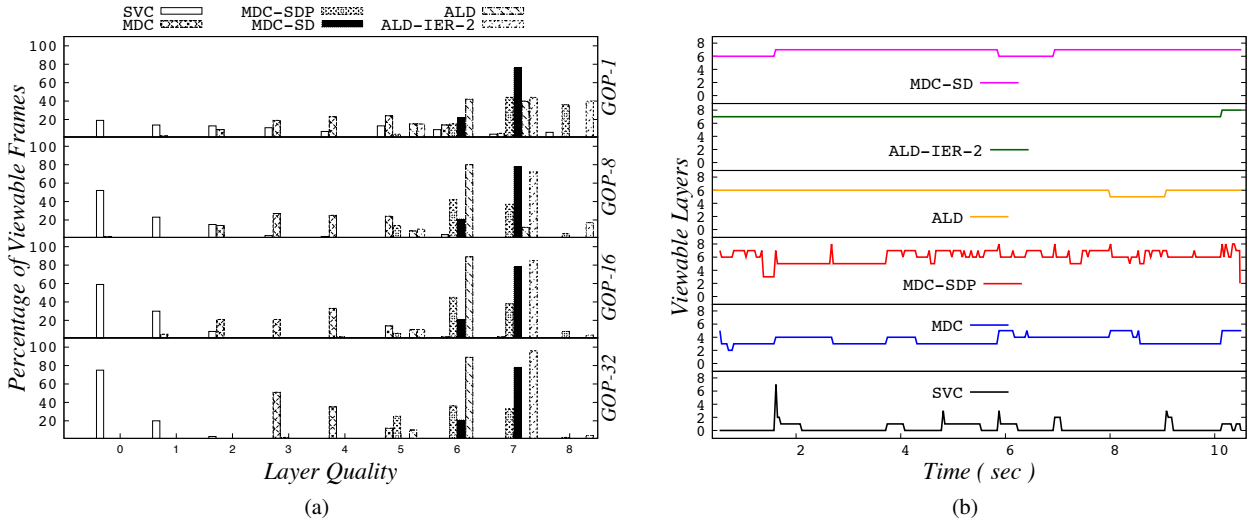


Fig. 9 (a) An example of the percentage of viewable frames for each of the six streaming models for each of the four GOP values for the crew clip and (b) Ten second example of the stream quality transitions for each streaming model of the crew clip with a GOP value of 32. Both with a 10% packet loss rate.

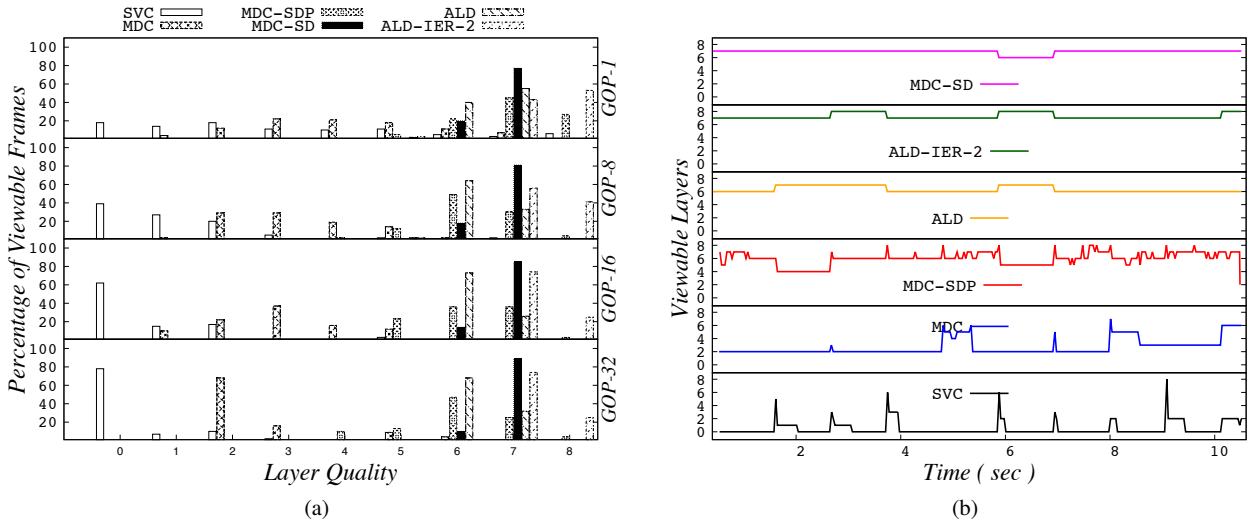


Fig. 10 (a) An example of the percentage of viewable frames for each of the six streaming models for each of the four GOP values for the soccer clip and (b) Ten second example of the stream quality transitions for each streaming model of the soccer clip with a GOP value of 32. Both with a 10% packet loss rate.

Table 7 STF for each of the clip types with a GOP of one

Layer	City	Crew	Harbour	Soccer
STF	3	6	7	3

variations in viewable layer value. One time to note in Figure 9b is that ALD-IER-2 is viewable primarily in layer 7, while ALD drops to layer 5 once during the duration of the stream. This reduction in viewable quality can also be seen in Figure 9a where there is a reduction in the viewable quality for the defined layer values of ALD and ALD-IER-2 for increasing values of GOP.

The reason for this reduction in quality is due to the selection of STF for crew. The STF values for each of the tested clips are illustrated in Table 7. It can be seen that city and soccer have the same STF value, thus would have similar levels of FEC resiliency. While crew and harbour have an increased STF which leads to an increase in the number of ALD descriptions required to decode the base layer and a decrease in the level of FEC resiliency and respective transmission cost. Even though the ALD variants in crew have a lower level of viewable quality in comparison to city and soccer, ALD-IER-2 still outperforms MDC-SD and has the highest levels of viewable quality of all streaming models for crew.

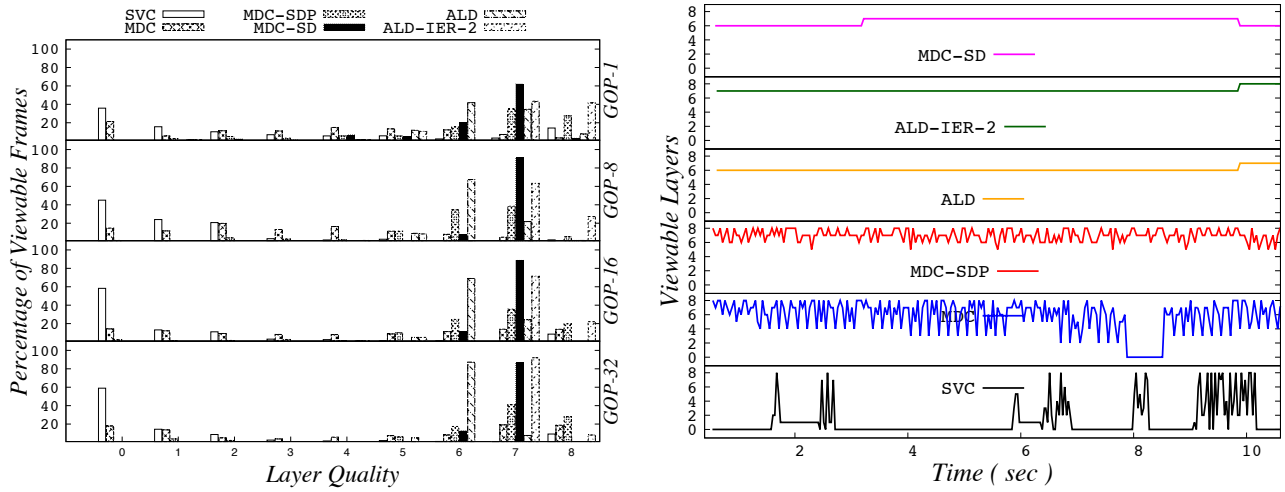


Fig. 11 (a) An example of the percentage of viewable frames for each of the six streaming models for each of the four GOP values for the Sintel HD clip and (b) Ten second example of the stream quality transitions for each streaming model of the Sintel HD clip with a GOP value of 32. Both with a 10% packet loss rate.

In our evaluation the percentage of viewable frames, quality transitions over time and maximum achievable quality per model for crew and harbour were consistent, while for soccer these results were similar to city. This highlights how the selection of STF and the underlying level of network loss mandates the maximum level of achievable viewable quality.

6 Evaluation Results for High Definition Content

For our high definition (*HD*) evaluation we use a trailer for the "Sintel" movie [29]. Sintel is an independently produced animated short film, initiated by the Blender Foundation, containing both slow and fast moving sequences. The Sintel trailer is 52 seconds in duration and contains 24 frames per second. Similar to our low resolution evaluation, for Sintel we encode an eight layer SVC stream with 1,253 frames in HD using three resolutions 854x480 (*480p*), 1280x720 (*720p*) and 1920x1080 (*1080p*); using a 2,3,3 quality to resolution ratio and QP values as per Table 4. The same streaming models, i.e. SVC, MDC, MDC-SDP, MDC-SD, ALD and ALD-IER-2, are simulated. ALD and ALD-IER-2 are allocated an STF value of 6 for each of the four GOP values. For each GOP values, Table 8 illustrates the maximum achievable PSNR and the transmission megabyte-cost for each adaptive scheme for quality layer 8.

Figure 11a presents an example of the percentage of viewable frames for each of the six streaming models for each of the four GOP values for the sintel HD clip while Figure 11b illustrates a ten second example of the stream quality transitions for each streaming model of the sintel HD clip with a GOP value of 32. Both figures with a 10% packet loss rate. Figure 11a provides similar results to our low resolution evaluations, where our techniques IER and SD mandate a greater percentage of viewable frames in the higher layers. For a GOP value of 32, 92% of the viewable frames of ALD-IER-2 are at layer 7, with the remaining 8% at maximum viewable quality (layer 8), consequently out performing all other models. Figure 11b shows the large variation in viewable quality provided by SVC and MDC, while illustrating the consistency of quality provided by ALD and SD.

Table 8 Transmission megabyte-cost for the Sintel HD clip at quality layer 8 for each adaptive scheme, for each of GOP values and maximum achievable PSNR for Layer 8 (in dB). All MDC variants, MDC, MDC-SDP and MDC-SD, have the same transmission cost, so only MDC is shown

GOP	PSNR	SVC	MDC	ALD	ALD-IER-2
1	49.6	49.9	137.3	69.8	70.8
8	49.1	16.9	44.0	23.3	23.7
16	48.3	12.5	32.3	17.1	17.4
32	47.4	10.4	26.8	14.2	14.4

It is important to note how the incremental increase in viewable quality provided by ALD-IER-2 mandates only a 2% increase in transmission cost as illustrated in Table 8. While ALD provides a reduction in transmission cost relative to MDC of approximately 46% for each of the respective GOP values. The variation in viewable quality for other GOP values, i.e GOP1, GOP8 and GOP16, are consistent with the plots previously shown for crew in Figure 7, while similar results were found for different loss rates. Thus validating the results seen for our low resolution evaluations and confirming that the benefits provided by ALD and our optimisation techniques are beneficial irrespective of clip type, encoding demands or underlying resolution requirements.

7 Subjective Testing Results

In this section, we present the results of scalable video subjective testing. The goal of our testing is to confirm the performance of the developed ideas with subjective evaluation. We utilised a packet loss rate of 10% and limited the frame interdependence of the model to one frame. Thus providing a means of illustrating the effects of packet loss rather than the effects of frame interdependence. The STF values for each of the tested clips are illustrated in Table 7. Note how the variation in STF values denotes a different optimal transmission cost for each clip.

7.1 Testing Setup

The test was implemented on a web server hosted locally on iMacs machines. Eighteen people participated in the subjective test. The test was performed in a well lit laboratory, and each clip type was shown seven times, once for each of the evaluated models. Each iteration of clips begins with the a viewing of the original clip with no packet loss, thus providing a base case on which the participants could rank/grade the streaming models. Each model per iteration was graded twice. Once immediately after viewing the model, thus providing the quality value for the individual model per iteration and a second time once all models had been viewed per iteration. As different models may have received the same quality value, the second grading is use to provide a means of ranking the models. For each streaming model, the achievable quality of the stream is based on the maximum layer per frame that contains no visual impairment when compared with the original clip with no packet loss.

In the Literature numerous references were found for scalable subjective testing, but these focused primarily on SVC only, examples of which include comparisons between SVC and AVC [21], different SVC codecs [18] and the effects of multi-dimensional scalability [10]. We are unaware of any subjective testing results which compare scalable and description-based coding.

Table 9 provides the streaming model PSNR values for each clip type, based on a simulated packet loss rate of 10%. Table 10 re-orders the streaming models and creates a streaming model ranking based on PSNR. Thus providing a means of comparing the streaming model ranking to the subjective testing ranking.

The ranking and grading of the streaming models per clip type is based on the following:

1. Per media clip, each iteration randomises the display allocation of the different streaming models. So that structure cannot be inferred, i.e. SVC is not always shown first, etc.

Table 11 Grading based on Impairment and Quality

Impairment	Quality	Grade
Imperceivable	Excellent	5.0
Perceptible, but not annoying	Good	4.0
Slightly annoying	Fair	3.0
Annoying	Poor	2.0
Very annoying	Bad	1.0

2. Per iteration, test subjects were asked to grade each clip. We implemented two grading schemes, based on test methods from the ITU-T Recommendation document: P.910 : Subjective video quality assessment methods for multimedia applications [1]. Both schemes are based on a rating of 1 to 5 inclusive, one being the worst, to five being the best. One scheme is based on Impairment, while the other is based on quality, both illustrated in Table 11. Impairment is based on how much variation or fragmentation a test subject can see, while quality is based on the tolerance in fidelity that a test subject can see.
   At the end of each iteration, we ask the test subjects to rate all seven clips in a scale of one to seven. With one being the best clip, or least annoying, and seven being the worst clip, or most annoying.
3. Finally, we ask the test subjects what annoyed them the most, what was best and what would improve their viewing pleasure.
4. The test subjects are not informed as to which clip belongs to which model, as this may have influenced them to try and choose specific streaming models in future tests cases.

7.2 Testing Results

Table 12 illustrates the streaming model ranking based on the mean grading value per clip type. The grading ranking would be very consistent to Table 10, once you take into consideration the very similar PSNR values for the models in Table 9.

In the test results, some subjects provided ranking based on the number of clips per iteration, i.e. 1 to 7, while others gave clips of similar quality the same ranking values. To maintain consistency of values over the entire subject base. In ranking scheme where similar values were given to multiple clips, higher values were changed to reflect actual ranking values, i.e. a ranking of 1, 2, 2, 2, 3, 3, 4 was changed to 1, 2, 2, 2, 5, 5, 7.

Table 13 displays the mean ranking values per iteration (clip type) for each of the streaming models. While Table 14 re-orders the streaming models and creates a streaming model ranking based on subjective testing ranking. Again the ranking is very consistent to Table 10, again taking into con-

Table 9 Streaming model PSNR dB values for each clip type, based on a 10% packet loss rate

Clip Type	SVC	MDC	MDC-SDP	MDC-SD	SDC-SDP	ALD	ALD-IER-2
City	26.13	26.71	33.72	35.13	34.07	34.56	36.00
Crew	31.40	33.47	37.82	37.57	37.41	36.70	37.56
Harbour	24.67	25.82	34.32	35.55	33.85	34.10	35.37
Soccer	29.32	32.21	36.61	36.57	36.31	36.19	37.34

Table 10 Ranking of streaming models per clip type based on PSNR values from Table 9

Clip Type	1	2	3	4	5	6	7
City	ALD-IER-2	MDC-SD	ALD	SDC-SDP	MDC-SDP	MDC	SVC
Crew	MDC-SDP	MDC-SD	ALD-IER-2	SDC-SDP	ALD	MDC	SVC
Harbour	MDC-SD	ALD-IER-2	MDC-SDP	ALD	SDC-SDP	MDC	SVC
Soccer	ALD-IER-2	MDC-SDP	MDC-SD	SDC-SDP	ALD	MDC	SVC

Table 12 Ranking based on mean grading results

Clip Type	1	2	3	4	5	6	7
City	ALD-IER-2	MDC-SD	SDC-SDP	ALD	MDC-SDP	MDC	SVC
Crew	MDC-SDP	SDC-SDP	ALD-IER-2	ALD	MDC-SD	MDC	SVC
Harbour	MDC-SD	ALD-IER-2	ALD	SDC-SDP	MDC-SDP	MDC	SVC
Soccer	MDC-SD	ALD-IER-2	MDC-SDP	ALD	SDC-SDP	MDC	SVC

sideration the very similar PSNR values for the models in Table 9.

7.3 Subjective Testing Conclusion

Each of the layered schemes contain known design issues which impede their respective deployment. While the adaptable quality is a benefit of SVC, the prioritised hierarchy and its dependency on the base layer is its greatest weakness. As we have highlighted, network transmission issues can affect all packets, and lower layer loss in SVC is detrimental to stream quality. While MDC offers consistent quality, the increased byte-cost of transmission is an inherent weakness. ALD provides the framework to achieve the high levels of adaptable stream quality promised by SVC, but the trans-

Table 13 Mean ranking value results

Clip Type	SVC	MDC	MDC-SDP	MDC-SD	SDC-SDP	ALD	ALD-IER-2
City	6.72	5.94	3.83	2.38	3.11	2.94	1.38
Crew	6.88	5.88	1.66	2.94	2.44	2.50	2.55
Harbour	6.61	6.05	4.16	1.50	3.00	3.16	1.88
Soccer	6.88	6.05	2.88	1.77	3.27	3.44	1.66

Table 14 Ranking based on mean ranking results

Clip Type	1	2	3	4	5	6	7
City	ALD-IER-2	MDC-SD	ALD	SDC-SDP	MDC-SDP	MDC	SVC
Crew	MDC-SDP	SDC-SDP	ALD	ALD-IER-2	MDC-SD	MDC	SVC
Harbour	MDC-SD	ALD-IER-2	SDC-SDP	ALD	MDC-SDP	MDC	SVC
Soccer	ALD-IER-2	MDC-SD	MDC-SDP	SDC-SDP	ALD	MDC	SVC

mission byte-cost of devices requesting lower layer decoding is dependent on stream encoding and the ALD selection value for STF, but mandates a high level of transmission cost for devices requiring lower layer streaming.

The results of our Scalable Video Subjective testing supported our simulated experimentation results. While SVC and MDC faired worst, our techniques and models were able to provide levels of consistent high quality viewing, with lower transmission cost irrespective of clip type. Our Subjective testing results highlight the benefits of not only intuitive encoding and transmission but also of selective packetisation. This can be seen in the increase in PSNR and ranking values attained by MDC when SDP and SD are utilised.

One item to note is that in some instances the grading results for the same clips per iteration were widely variable. So the quality of clip does not only depend on the layer quality achievable, but also on the person viewing the clip.

8 Conclusion

In this paper, Adaptive Layer Distribution (*ALD*) is proposed as a novel multifaceted approach to media streaming optimisation. ALD section thinning enables the reduction of the total streaming overhead while IER and section distribution improve ALD error resiliency to loss. Our simulation and subjective testing results show that the components of ALD achieve a superior performance to other scalable streaming frameworks, irrespective of video type and GOP size. Currently, we are working on improving the transmission efficiency of ALD for users interested in low quality video.

Acknowledgements The authors acknowledge the support of Science Foundation Ireland (SFI) under Research Grant 10RFP/CMS2952. The authors would like also to acknowledge the support of the National Telecommunication Regulation Authority (NTRA) of Egypt.

References

1. http://www.itu.int/rec/T-REC-P.910/en
2. Bai, H., Wang, A., Zhao, Y., Pan, J.S., Abraham, A.: Distributed Multiple Description Coding. Principles, Algorithms and Systems. Springer-Verlag New York Inc (2011)
3. Berkin Abanoz, T., Murat Tekalp, A.: SVC-based scalable multiple description video coding and optimization of encoding configuration. Signal Processing: Image Communication **24**(9), 691–701 (2009)
4. Bossen,F.:Common HM test conditions and software reference configurations JCTVC-L1100 of JCT-VC, Geneva, CH, Jan. 2013.
5. Chou, P., Wang, H.: Layered multiple description coding. IEEE International Packet Video Workshop (PV2003) (2003)
6. Cisco Visual Networking Index, 2012-2017.: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf
7. D Wu et al.: Streaming video over the Internet: approaches and directions. IEEE Trans. Circuits and Systems for Video Technology **11**(3), 282–300 (2001)
8. Dai, M., Loguinov, D., Radha, H.: Rate-Distortion Analysis and Quality Control in Scalable Internet Streaming. IEEE Transactions on Multimedia **8**(6), 1135–1146 (2006)
9. Dutta, A., Chennikara, J., Chen, W., Altintas, O., Schulzrinne, H.: Multicasting streaming media to mobile users. Communications Magazine, IEEE **41**(10), 81–89 (2003)
10. Eichhorn, A., Ni, P.: Pick your layers wisely - a quality assessment of h.264 scalable video coding for mobile devices. In: Communications, 2009. ICC '09. IEEE International Conference on, pp. 1–6 (2009). DOI 10.1109/ICC.2009.5305948
11. Gan, T., Gan, L., Ma, K.K.: Reducing video-quality fluctuations for streaming scalable video using unequal error protection, retransmission, and interleaving. IEEE Transactions on Image Processing **15**(4), 819–832 (2006)
12. Goyal, V.: Multiple description coding: Compression meets the network. Signal Processing Magazine **18**(5), 74–93 (2002)
13. Grois, D. and Marpe, D. and Mulayoff, A. and Itzhaky, B. and Hadar, O.: Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders In: Picture Coding Symposium (PCS), 2013, pp 394-397 (2013)
14. H. Schwarz et al.: Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Trans. Circuits and Systems for Video Technology **17**(9), 1103–1120 (2007)
15. Ke, C.H.: myEvalSVC: an Integrated Simulation Framework for Evaluation of H.264/SVC Transmission. KSII Transactions on Internet and Information Systems **6**(1), 379–394 (2012)
16. Kirihara, K., Masuyama, H., Kasahara, S., Takahashi, Y.: FEC Recovery Performance for Video Streaming Services Based on H. 264/SVC. Recent Advances on Video Coding, InTech (2011)
17. Kuschnig, R., Kofler, I., Hellwagner, H.: Evaluation of HTTP-based request-response streams for internet video streaming. Proc. of ACM Multimedia Systems Conference (MMSys '11) (2011)
18. Lee, J.S., De Simone, F., Ramzan, N., Zhao, Z., Kurutepe, E., Sikora, T., Ostermann, J., Izquierdo, E., Ebrahimi, T.: Subjective evaluation of scalable video coding for content distribution. In: Proceedings of the International Conference on Multimedia, MM '10, pp. 65–72. ACM, New York, NY, USA (2010). DOI 10.1145/1873951.1873981. URL http://doi.acm.org/10.1145/1873951.1873981
19. Li, L., Han, Q., Niu, X.: Enhanced Adaptive FEC Based Multiple Description Coding for Internet Video Streaming over Wireless Network. In: Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on, pp. 478–481 (2010)
20. Nafaa, A., Taleb, T., Murphy, L.: Forward error correction strategies for media streaming over wireless networks. IEEE Communications Magazine **46**(1), 72–79 (2008)
21. Oelbaum, T., Schwarz, H., Wien, M., Wiegand, T.: Subjective performance evaluation of the svc extension of h.264/avc. In: Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pp. 2772–2775 (2008). DOI 10.1109/ICIP.2008.4712369
22. Przybylski, M., Belter, B., Binczewski, A.: Shall we worry about Packet Reordering? Computational Methods in Science and Technology pp. 141–146 (2005)
23. Puri, R., Ramchandran, K., Lee, K.: Forward error correction (FEC) codes based multiple description coding for Internet video streaming and multicast. Signal Processing: Image Communication **16**(8), 745–762 (2001)
24. Quinlan, J.J., Zahran, A., Sreenan, C.J.: SDC: Scalable Description Coding for Adaptive Streaming Media. In: Proc. of 19th IEEE International Packet Video Workshop (PV2012) (2012)
25. Quinlan, J.J., Zahran, A., Sreenan, C.J.: ALD: Adaptive Layer Distribution for Scalable Video. Proc. of ACM Multimedia Systems Conference (MMSys '13) (2013)
26. Reichel, J., Schwarz, H., Wien, M.: Joint scalable video model 11 (JSVM 11). Joint Video Team, doc. (Jul. 2007)

27. Salomon, D.: Guide to Data Compression Methods. Springer (2002)
28. Schierl, T., Stockhammer, T., Wiegand, T.: Mobile Video Transmission Using Scalable Video Coding. IEEE Trans. Circuits and Systems for Video Technology **17**(9), 1204 – 1217 (2007)
29. Sintel, the Durian Open Movie Project:http://www.sintel.org/download
30. Shokrollahi, A.: Raptor codes. Information Theory, IEEE Trans. on **52**(6), 2551–2567 (2006)
31. Sorensen, J., Ostergaard, J., Popovski, P., Chakareski, J.: Multiple description coding with feedback based network compression. In: Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE, pp. 1–6 (2010). DOI 10.1109/GLOCOM.2010.5684254
32. Sreenan, C., Chen, J.C., Agrawal, P., Narendran, B.: Delay reduction techniques for playout buffering. IEEE Transactions on Multimedia **2**(2), 88–100 (2000)
33. Steinbach, E., Färber, N., Girod, B.: Adaptive playout for low latency video streaming. Proc. International Conference on Image Processing (2001)
34. Stockhammer, T.: Dynamic Adaptive Streaming over HTTP – Standards and Design Principles. In: MMSys '11 Proceedings of the second annual ACM conference on Multimedia systems, p. 133. ACM Press, New York, New York, USA (2011)
35. The Network Simulator - ns-2: http://nsnam.isi.edu/nsnam/index.php
36. Unanue, I., Urteaga, I., Husemann, R., Del Ser, J., Roesler, V., Rodriguez, A., Sanchez, P.: A Tutorial on H.264/SVC Scalable Video Coding and its Tradeoff between Quality, Coding Efficiency and Performance. In: Recent Advances on Video Coding, pp. 1–24. InTech (2011)
37. Video traces for network performance evaluation: ftp://ftp.tnt.uni-hannover.de/pub/svc/testsequences/
38. Wang, B., Kurose, J., Shenoy, P., Towsley, D.: Multimedia streaming via TCP: an analytic performance study. ACM Transactions on Multimedia Computing, Communications and Applications **4**(2) (2008)
39. Wang, Y., Lin, T.L., Cosman, P.: Packet dropping for H.264 videos considering both coding and packet-loss artifacts. Packet Video Workshop (PV), 2010 18th International pp. 165–172 (2010)
40. Z. Zhao et al.: Multiple description scalable coding for video streaming. 2009 10th Workshop Image Analysis for Multimedia Interactive Services. WIAMIS '09. pp. 21–24 (2009)
41. Zhao, Z., Ostermann, J.: Video streaming using standard-compatible scalable multiple description coding based on SVC. In: 17th IEEE International Conference on Image Processing (ICIP), pp. 1293–1296 (2010)