

Traffic Analysis of Multiple Description Coding of Video Services over IP Networks

Frank H.P. Fitzek^b, Başak Can^b, Ramjee Prasad^b, Marcos Katz[‡], DS Park[‡]

^bCenter for TeleInFrastructure (CTIF), Aalborg University

Niels Jernes Vej 12, 9220 Aalborg Øst, Denmark; ph: +45 9635 8661, e-mail: [ff|bc|prasad]@kom.aau.dk

[‡]SAMSUNG ELECTRONICS CO., LTD.; e-mail:[marcos.katz|dspark]@samsung.com

Abstract—For the performance evaluation of future wireless communication systems traffic traces are needed for simulation purpose. Multiple description coding is gaining a lot of interest lately and is a viable solution to provide robust video services over unreliable wireless links. By means of twelve well known video sequences, in different video formats, we generate the frame size traces and evaluate the multiple description coding characteristics. Furthermore we investigate the video quality in dependency of lost descriptors. This allows researcher using our traces to convert the losses in their simulation directly into video quality values. This makes this work unique as single layer coded video streams need further postprocessing to receive the video quality.

Keywords—MDC, video traces, PSNR, H.26L, CIF, QCIF

I. INTRODUCTION AND RELATED WORK

Future wireless communication systems will be dominated by multimedia applications such as video services. The unreliable wireless link makes it difficult to support the customers with high rate and stable video quality. Therefore multiple description coding (MDC) has gained a lot of interest to encode the information source. MDC has the capability to split the information stream in multiple sub-streams, where each of the sub-streams can be decoded without the information carried by the neighboring sub-streams and therefore has no dependencies to other sub-streams such as layered video coding. The advantages of MDC has been exploited for multi hop networks [1] [2], Orthogonal Frequency Division Multiplexing (OFDM) [3], Multiple Input Multiple Output (MIMO) systems [4], ad-hoc networks [5], Universal Mobile Telecommunications System (UMTS) [6], Transport Control Protocol (TCP) [7] and Content Delivery Networks (CDN) [8].

Unfortunately, the advantage of multiple description coding is achieved at the expense of higher bandwidth usage due to the smaller video compression of the encoding process. Therefore existing video traffic characterizations such as single and multiple layer coding, as presented in [9], can not be used for the evaluation of future wireless communication systems. In this paper we present traffic traces of MDC and derive their characteristics. The traffic traces can be used for simulations. Furthermore additional information is presented, that allow deriving the video quality after receiving a number of sub-streams at the receiver. The results are discussed in this paper and the traces are made publicly available on our web page [10]. Other researchers are encouraged to use these traces within their simulations. To our best knowledge no work in the field of MDC traffic trace were presented so far.

II. VIDEO BASICS AND MDC

Because of the large bit rates, digital video is almost always encoded (compressed) before transmission over a packet-oriented network. The limited bandwidths of wireless links make compression especially important. Video compression generally exploits three types of techniques. On a per-frame basis (i.e., single picture), neighboring pixels tend to be correlated and thus have spatial redundancy. Intra-frame encoding is employed to reduce the spatial redundancy in a given frame. In addition, consecutive frames have similarities and therefore temporal redundancy. These temporal

redundancies are reduced by inter-frame coding techniques. Inter-frame coding considers both prediction from past reference frames as well as future reference frames. The resulting frames compressed by encoding the original pictures in the temporal domain are named as: I (Intra), P (Inter), and B (Bi-directional) frames, as introduced in the MPEG-1 standard [11]. These encoding methods are applied on the frame or block level, depending on the codec. I-frames can be reconstructed without any reference to other frames. This leads to less compression gain. The P-frames are inter coded leading to a high compression gain and smaller frame size than I-frames. They are predicted from the last frame (I-frame or P-frame) i.e. it is not possible to reconstruct them without the data of the previous frame (I or P). Therefore, to prevent error propagation, I-frames are repeated periodically. The sequence of frames between two intra-coded frames is referred to as a Group of Picture (GoP). The relationship between these various encoding types and how frames rely on each other in a typical frame sequence for MDC with 3 descriptors is illustrated in Figure 1 for $GoP = 12$.

The raw data sequence is feeded into the splitter which is presented in the following section and the multiple descriptors are generated.

III. MEASUREMENT METHODOLOGY

MDC splits the video stream into multiple descriptors. Therefore the frame based approach using a splitter entity is exploited. The splitter, as given in Figure 2, takes the raw video sequence and splits it into J sub-sequences, ($J > 1$), such that the i -th sub-sequence contains picture i , $J + i$, $2J + i$, and so on. These J subsequences are feeded into the video encoder, i.e. H.26L or H.263.

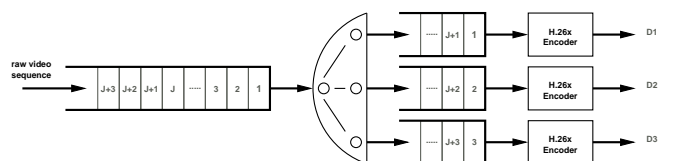


Fig. 2. Construction of multiple substreams and descriptions

In Figure 3, a basic diagram for the decoder and merger settings is given. Once the video is encoded with multiple descriptions we are measuring the quality in terms of PSNR in case of missing descriptors. A video frame is composed by $N \cdot M$ pixels (where M is the length and N the height of the frame). Each pixel is presented by one luminance value and a set of pixels by two chrominance values. Because the human eye is more sensible to the change in luminance we focus only on this parameter. The mean squared error (MSE) and the PSNR in decibels are computed by the following two equations [12]:

$$MSE = \frac{\sum_{\forall i,j} [f(i,j) - F(i,j)]^2}{N \cdot M} \quad (1)$$

$$PSNR = 20 \cdot \log_{10} \left(\frac{255}{\sqrt{MSE}} \right), \quad (2)$$

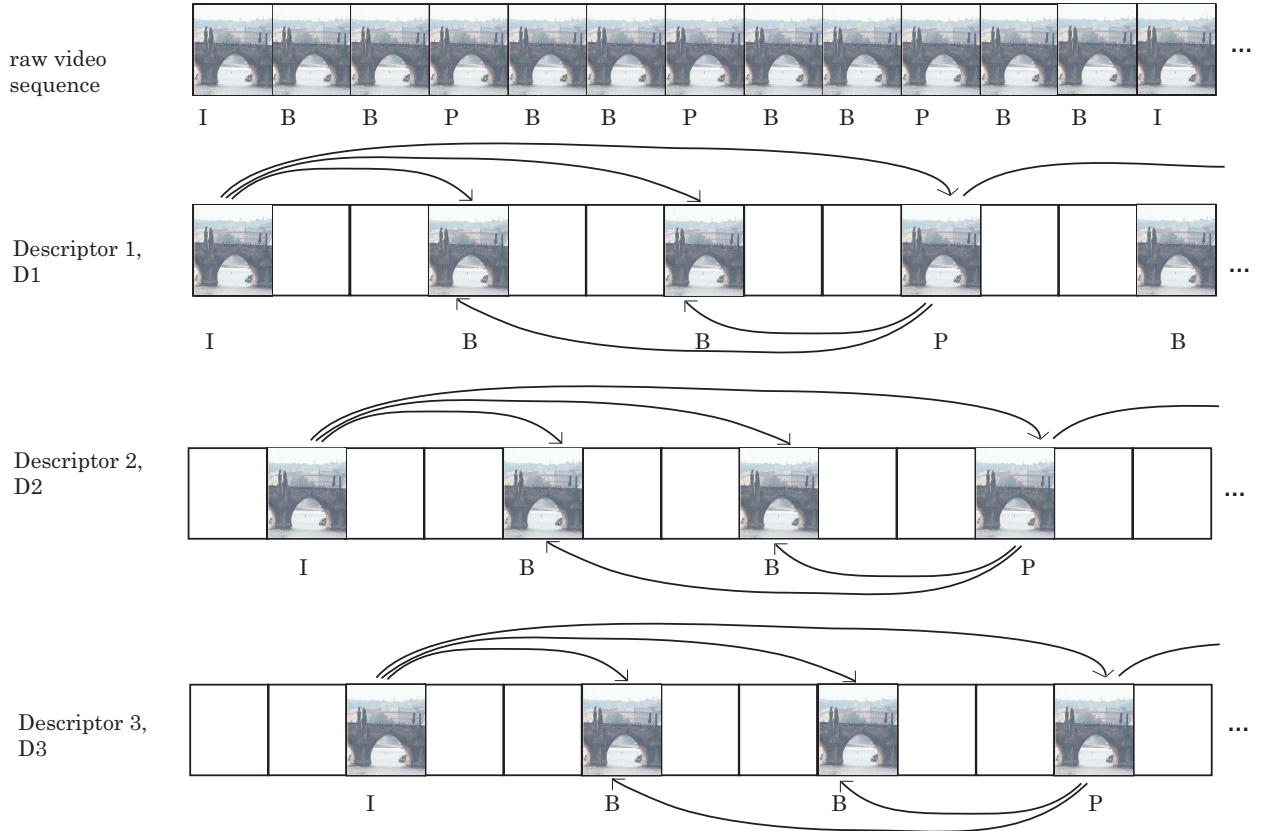


Fig. 1. Construction of multiple substreams and descriptions for GoP including 12 video frames

where $f(i, j)$ represents the original source frame and $F(i, j)$ the reconstructed possibly error-prone frame containing N by M pixels. If a descriptor is lost, we are assuming that all of its frames are completely lost. Each descriptor is decoded individually and then merged to a single stream. This stream will be conveyed towards the application.

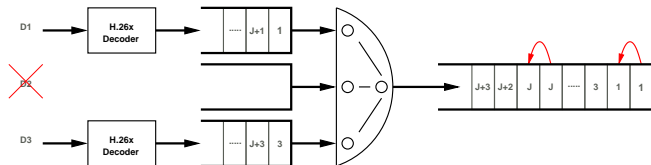


Fig. 3. Merger for $J=3$, where one descriptor is missing.

The calculation of PSNR values is done as given in [12] using the videometer tool [13]. The videometer tool calculates the PSNR values and is able to freeze video frames in case the following frames are lost. This is important to have some sort of error resilience. We have repeated the measurements multiple times with a confidence interval of 99% for the PSNR value.

The video sequences with frame length and a short information are given in Tables 1 and 2 for the QCIF and CIF formats.

IV. RESULTS

The results presented in this section concentrate on the frame size traces for MDC. In this work we use H.26L for video coding standard [14]. A sample of generated I and P-frames are given in Table 3.

Tab. 1. YUV QCIF

seq. name	frames	info
bridge--close	2000	Charles Bridge (Karlův most).
bridge--far	2101	Charles Bridge (Karlův most).
carphone	382	Man talking at the phone.
claire	494	Female talking to the camera.
container	300	Ship leaving the harbor.
foreman	400	Man speaking to the camera.
grandma	870	Grandma in front of the camera.
highway	2000	Driving over highway.
mthr.dotr	961	Mom and daughter.
news	300	News studio and two speakers.
salesman	449	Salesman in his office.
silent	300	Woman doing sign language.

In Figure 4, the frame size traces for single description and multiple description coding for the highway video sequence are given. This video sequence includes relatively slow motion than highway video sequence. The peaks occur when I-frames are generated by the encoder. The size of the I-frames are similar to each other in case of single, five and fifteen descriptors. However if we look at each descriptor individually, we can observe that the frame size increases with increasing number of descriptors. This is due to the increasing difference between inter-coded frames due to splitting process.

In Figure 5, the frame size traces for single description and multiple description coding for the highway video sequence are given. Similar to Figure 4, the generated frame sizes have peaks

Tab. 4. Mean frame size value in byte for the QCIF video sequences

QCIF	bridge-far	claire	grandma	highway	bridge-close	container	mthr_dotr	salesman	silent	news	carphone	foreman
1	658.5	1445.8	1763.5	2323.1	2763.3	2127.0	2234.5	2615.7	3314.1	3267.1	4628.1	4329.7
2	674.4	1748.2	2045.8	2781.3	3380.5	2476.2	2763.7	3111.5	4029.7	4125.1	5572.6	5358.6
3	687.2	1997.1	2272.1	3018.1	3697.1	2813.2	3155.4	3482.2	4517.8	4719.5	6198.2	6203.8
4	683.3	2201.3	2455.0	3261.7	3835.7	3033.8	3481.6	3809.9	4823.2	5101.1	6623.6	6869.7
5	709.2	2367.1	2564.5	3432.8	3965.2	3083.0	3698.4	3931.1	4857.2	5213.7	6957.1	7306.7
6	709.7	2384.8	2712.4	3590.8	4011.2	3531.3	4039.5	4217.9	5321.7	5754.7	7326.7	7679.5
7	711.6	2498.1	2800.3	3653.3	4065.6	3626.2	4248.8	4438.6	5448.0	5844.1	7585.9	8081.2
8	708.1	2658.1	2784.4	3826.4	4098.4	3944.6	4423.0	4518.0	5759.1	6158.9	7733.6	8404.0
9	730.8	2669.9	2833.0	3924.6	4195.3	3752.2	4602.1	4721.6	5642.6	6142.3	7980.7	8549.4
10	741.1	2918.2	3035.2	4061.7	4236.8	4059.7	4657.17	4634.3	5830.9	6329.0	8391.3	9165.0
11	765.7	2839.8	3073.1	4117.5	4328.4	4468.4	5027.1	4947.6	5996.7	6694.5	8473.1	9013.7
12	724.8	2932.8	3036.6	4170.1	4275.3	4729.9	5063.3	5258.3	6314.2	6887.3	8586.8	8936.2
13	740.7	2958.9	3207.3	4188.9	4340.7	4344.3	5334.2	4940.0	5908.1	6529.3	8879.7	9517.3
14	740.7	2918.8	3287.6	4217.1	4354.8	4498.6	5381.8	5117.2	6023.6	6816.3	8881.7	10057.1
15	780.4	3064.5	3215.0	4357.2	4458.1	4794.4	5425.8	5371.5	6279.6	6931.6	9040.9	10246.1
16	765.5	3025.8	3315.5	4510.9	4440.9	4968.8	5396.1	5442.5	6357.7	7173.7	8865.0	10592.9
17	754.1	3063.4	3490.3	4487.8	4431.5	5175.0	5685.5	5585.8	6354.8	7336.4	8911.2	9868.8
18	795.6	3133.0	3195.3	4550.1	4535.9	5366.5	5982.4	5090.3	6566.5	7612.5	9046.0	9917.4
19	760.2	3261.2	3248.5	4593.7	4496.3	5554.6	6013.2	5192.7	6847.4	7896.5	9422.8	10072.7
20	783.6	3062.6	3538.0	4645.4	4571.0	5628.2	5779.0	5412.0	6718.9	7711.4	9677.4	10582.8

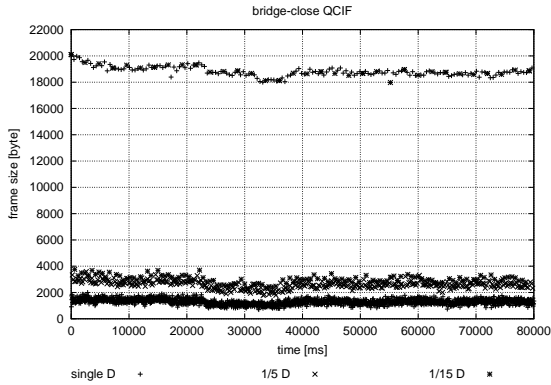


Fig. 4. Mean frame size values versus time for bridge--close video sequence in the QCIF format

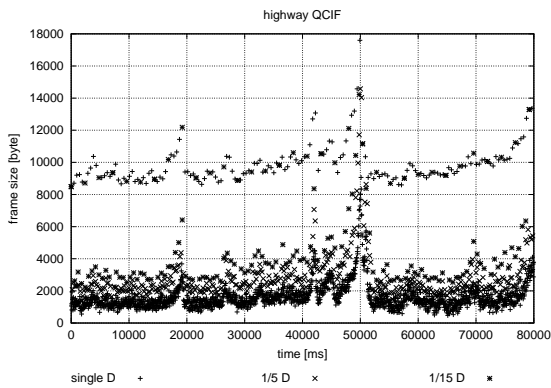


Fig. 5. Mean frame size values versus time for highway video sequence in the QCIF format

Tab. 2. YUV CIF

seq. name	frames	info
bridge--close	2000	Charles Bridge (KarlV most).
bridge--far	2101	Charles Bridge (KarlV most).
highway	2000	Driving over highway.
mobile	300	Train is moving
paris	1065	Two people talking to each other.
tempeste	260	Moving Cam

	Bit/pic	QP	SnrY
0 (I)	18112	31	34.8461
1 (P)	4000	30	34.6026
2 (P)	4088	30	34.3582
3 (P)	3376	30	34.3790
4 (P)	3144	30	34.3425
5 (P)	2648	30	34.1732
6 (P)	2944	30	34.2640
7 (P)	3968	30	34.3002
8 (P)	2792	30	34.2879
9 (P)	2864	30	34.3186
10 (P)	3496	30	34.3413
11 (P)	3704	30	34.3181
12 (I)	19984	31	34.5246
13 (P)	3008	30	34.3633
14 (P)	2832	30	34.5111
15 (P)	3216	30	34.5907
16 (P)	3584	30	34.7930
17 (P)	3712	30	34.6201
18 (P)	4112	30	34.6005

Tab. 3. I and P-frames generated by the H.26 L encoder for single stream

when I-frames are generated. highway video sequence includes relatively higher motion. Therefore the waves in the trace occur when the motion increases in the video. Obviously the inter-frame differences increase with larger J . This also means that compression gain of P-frames is less than single description case which leads to higher frame sizes with increasing J . However, the size of the I-frames are not affected by increasing J . In Figure 6 mean frame sizes versus number of descriptors for six QCIF video sequences are given. This figure is a plot of the measurements presented in Table 4. The carphone sequence has the highest

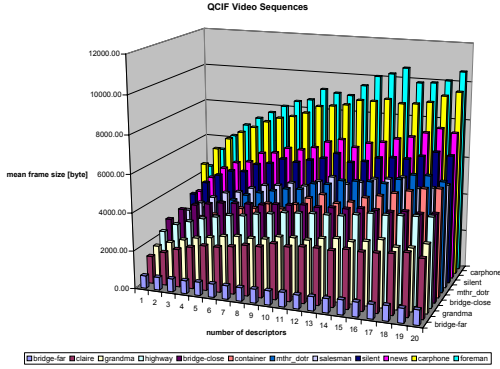


Fig. 6. Mean frame size versus number of descriptors for six different QCIF video sequences

motion activity. Whereas, the `bridge--far` sequence contains the lowest motion activity. This figure shows that the frame sizes strongly depend on the video content. The mean frame size trace of `bridge--far` sequence is not affected dramatically from splitting process. However, as the motion activity increases, the effect of splitting process on the mean frame size and the size of the frames increases as well.

V. VIDEO QUALITY

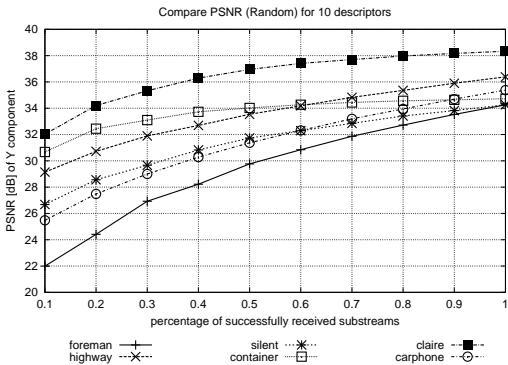


Fig. 7. PSNR values for different numbers of descriptors versus the percentage of received sub-streams for various video sequences.

For simulation purpose it is also interesting to map the channel losses to a video quality value. For this purpose we have chosen the Picture-signal-to-noise-ratio (PSNR) value. In Figure 7, PSNR measurements versus percentage of successfully received sub-streams for a total of $J = 10$ descriptors are given. The results are obtained using six different video sequences. The `foreman` video sequence includes the highest motion. Whereas `claire` and `container` video sequences include relatively lower motion than `foreman`. As we can observe in the figure, if the motion in a given video sequence is high, the slope of the PSNR degradation curve is also high. For example, for a loss of 60 % out of $J = 10$ descriptors, the PSNR degradation is 6 dB for `foreman`, 2 dB for `claire` and 1 dB for `container` video sequence. Thus, we conclude that the more the motion in a given video sequence, the more important is to receive as many descriptors as possible for a given J .

For the purpose of simulating erroneously received video sequences, the following steps are followed commonly in literature:

- Video traces are generated by encoding the raw video stream accordingly as given in Figure 2.
- The resulting traces are put into a simulator which simulates a given wireless scenario. At the output of the simulator, the erroneous video traces are obtained due to channel and/or link losses.
- To obtain the error-prone video sequence, the received erroneous video traces are decoded and merged accordingly as given in Figure 3.

However, with the PSNR curves for various video sequences which are presented in this study can be used to directly map the video quality onto the encoded video sequence for a given percentage of successfully received descriptors. In this way, the error-prone video sequence is obtained without the need of decoding and merging processes followed by the simulation of the channel and/or link losses.

VI. NETWORK OVERHEAD

Once the frame size traces are available, we have to take into account that the different descriptors are transmitted over IP based communication channels. This will lead to an additional overhead due to the IP header information (including higher layer protocols such as TCP or User Datagram Protocol (UDP)). Under the assumption that the RTP/UDP/IPv4 is used, each frame, carrying the full header information of each protocol layer (so no header compression is used), has an additional overhead of 40 bytes. For IPv6 the overhead equals 60 bytes. In case of fragmentation or framing the overhead has to be adjusted properly according to the chosen scenario. The effect of network overhead on MDC is presented in our study in [15].

VII. CONCLUSION AND OUTLOOK

The study in this paper presents for the first time a sophisticated library of frame size traces for MDC for commonly used video sequences, which has never been done before. Besides the frame size traces and their characteristics, we come up with a simplified mapping function to translate loss of descriptors of the multiple description process to a video quality metric. Whereby we are eliminating the need for decoding and merging processes.

ACKNOWLEDGMENT

This paper is based upon work supported by SAMSUNG, Korea. We would like to thank Patrick Seeling for his support in the methodology.

REFERENCES

- [1] Nitin Gogate, Doo-Man Chung, Shivendra S. Panwar, and Yao Wang, "Supporting Image and Video Applications in a Multihop Radio Environment Using Path Diversity and Multiple Description Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 9, pp. 777–792, September 2002.
- [2] N. Gogate and S.S. Panwar, "Supporting Video/Image Applications in a Mobile Multihop Radio Environment Using Route Diversity," in *IEEE International Conference on Communications*, vol. 3, Vancouver, Canada, June 1999, pp. 1701 – 1706.
- [3] Zhu Ji, Qian Zhang, Wenwu Zhu, Jianhua Lu, and Ya-Qin Zhang, "Video Broadcasting Over MIMO-OFDM Systems," *IEEE International Symposium on Circuits and Systems*, vol. 2, no. 25-28, pp. II844 – II847, May 2003.
- [4] Y. Altunbaşak, Nejat Kamaci, and Russel M. Mersereau, "Multiple Description Coding with Multiple Transmit and Receive Antennas for Wireless Channels: The Case of Digital Modulation," *IEEE Global Telecommunications Conference*, vol. 6, no. 25-29, pp. 3272–3276, November 2001.
- [5] Shunan Lin, Yao Wang, Shiwen Mao, and S. Panwar, "Video Transport Over Ad-Hoc Networks Using Multiple Paths," in *IEEE International Symposium on Circuits and Systems*, ser. ISCAS, vol. 1, Scottsdale, Arizona, May 2002, pp. 157–160.

- [6] M. Pereira, M. Antonini, and Michel Barlaud, "Multiple Description Video Coding for UMTS," in *Picture Coding Symposium*, Saint Malo, France, April 2003.
- [7] R. Puri, K. W. Lee, K. Ramchandran, and Vaduvur Bharghavan, "An Integrated Source Transcoding and Congestion Control Paradigm for Video Streaming in the Internet," *IEEE Transactions on Multimedia*, vol. 3, no. 1, pp. 18–32, April 2001.
- [8] J. G. Apostolopoulos, W. Tan, and S. J. Wee, "Performance of a Multiple Description Streaming Media Content Delivery Network," in *IEEE Int'l Conf. on Image Processing*, vol. 2, Rochester, New York, September 2002, pp. II–189 – II–192.
- [9] "Video traces for network performance evaluation: YUV 4:2:0 Video Sequences," <http://trace.eas.asu.edu/yuv/yuv.html>.
- [10] F. H. P. Fitzek, "Traffic Analysis for Multiple Description Coded Video Sequences," <http://kom.aau.dk/%7Eeff/MDC/>, June 2004.
- [11] J. S. 29, "MPEG-1, Coding of moving pictures and associated audio for digital storage media at up to 1.5 mbps," ISO/IEC, Tech. Rep. 11172, 1993.
- [12] Frank Fitzek, Patrick Seeling, and Martin Reisslein, *Wireless Internet*, March 2004, no. ISBN: 0849316316, Electrical Engineering & Applied Signal Processing Series, ch. Video Streaming in Wireless Internet.
- [13] P. Seeling, F. Fitzek, and M. Reisslein, "Videometer," *IEEE Network*, vol. 17, no. 1, p. 5, jan 2003.
- [14] Thomas Wiegand, "H.26L Standard," JVT, Tech. Rep., 2002.
- [15] Frank H. P. Fitzek, Başak Can, Ramjee Prasad, and Markos Katz, "Overhead and Quality Measurements for Multiple Description Coding for Video Services," *Wireless Personal Multimedia Communications (WPMC)*, September 2004.