Source Descriptor Selection Schemes for Multiple Description Coded Services in 4G Wireless Communication Systems

F.H.P. Fitzek H. Yomo P. Popovski R. Prasad

Department of Communications Technology
Center for TeleInFrastruktur (CTIF)
Aalborg University, Denmark
[ff|yomo|petarp|prasad]@kom.aau.dk

M. Katz

Telecommunication R&D Center
Samsung Electronics Co Ltd.
Suwon, Korea
marcos.katz@samsung.com

Abstract

4G wireless communication networks are characterized by the need to support heterogenous terminals differing in size, display, battery, computational power, etc. For efficient usage of the wireless spectrum all devices should be served by the same spectrum instead of allocating spectra dedicated to the different terminal classes. This feature is naturally supported when the source coding of the traffic is done as a multiple description coding. The restoration quality of the information source is proportional to the quantity of the descriptors used in the restoration process. Hence, terminals with less capabilities may simply discard or not receive some of the descriptors, while high class terminals try to receive all information. In the case of partial reception of descriptors the performance can be improved by selection strategies for the descriptors. We advocate the usage of new descriptor selection schemes for multiple description coded services. The proposed schemes differ with respect to the availability of a feedback channel. All solutions are terminal oriented and are beneficial in the design of terminals that have robust and high quality services. Furthermore, our approaches inherently achieve fairness among different terminals. As an example of our results for video communication, we can show gains of 4 dB when the proposed approach is utilized in the system.

1 Introduction

The exponential growth of user demands and the limitations of the Third Generation of Mobile Communication Systems (3G) have brought researchers to start reflecting on the Fourth Generation (4G). The literature hosts many prophetic visions which present the future generation as the ultimate boundary of the wireless mobile communication without any limit in its potential. However, there are almost no practical design rules and thus a firm definition of 4G. The Second Generation of Mobile Communication Systems (2G) was a huge success story because of its revolutionary technology and the services brought to its customers. Besides the service of high quality speech, the global mobility was a strong reason for buying 2G terminals. The Third Generation (3G) has been started in some parts of the world, but the success story of 2G is hard to be repeated. One reason is that the evolution from 2G towards 3G has not brought any qualitatively new service for the customer, leaving the business model largely unchanged. The well known services plus some additional ones are provided, which may not be enough to encourage the customers to change their equipment.

The upcoming Fourth Generation (4G) is projected to solve still–remaining problems of the previous generations and to provide a convergence platform for a wide variety of new services, from high–quality voice to high–definition video, through high–data–rate wireless channels. Various visions of 4G have emerged recently among the telecommunication industries, the universities and the research institutes all over the world. In Europe, the European Commission envisions that 4G will ensure seamless service provisioning across a multitude of wireless systems and networks, from private to public, from indoor to wide area, and provide an optimum delivery via the most appropriate (i.e., efficient) network available. From the service point of view, it foresees that 4G will be mainly focused on personalized services. In Asia, the Japanese operator NTT DoCoMo has introduced the concept of MAGIC for defining 4G: Mobile multimedia; Anytime, anywhere, anyone; Global mobility support; Integrated wireless solution; and Customized personal service, which mostly focuses on public systems and treats 4G as the extension of 3G cellular service. Even if 4G is named as the successor of previous wireless communication generations, it is not limited to cellular systems. Therefore it should not be understood solely as a linear interpolation of 3G, made upon the observation of the 2G–3G development sequence. In order to be a step ahead of 3G technology, it is not enough that 4G provides only higher data rates, but it should also bring some clear and evident advantage and new quality in peoples everyday life.

The success of 4G should be sought out the combination of network and terminal heterogeneity. Network heterogeneity guarantees ubiquitous connection and provision of common services (e.g., voice telephony, etc.) to the user, ensuring at least the same level of Quality of Service (QoS) when passing from one networks support to another one. Moreover, due to the simultaneous availability of different networks, heterogeneous services are also provided to the user. Terminal heterogeneity refers to the support of different types of terminals in terms of display size, energy consumption, portability/weight, complexity, etc. as given in Figure 1 [4]. In contrast to 4G, 2G and 3G are characterized by homogeneous terminals. Since 4G will encompass various types of terminals that may have to provide common services independently of their capabilities, the service presentation is optimized by tailoring of the content to the end–user device. Furthermore, the provision of the upcoming new services will be accurately decided according to the capabilities of the terminal in use. This paves the way towards real service personalization.

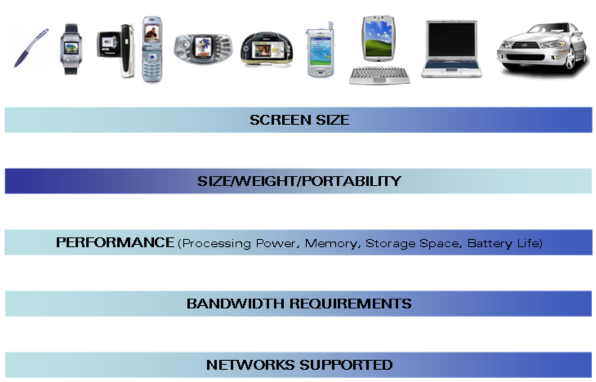


Figure 1: Heterogenous terminals with their characteristics.

The concept of multiple description coding (MDC) has recently been recognized as one of the key drivers to support heterogenous terminals in a communication system. MDC allows splitting of a single source (e.g. audio or video) into multiple descriptors, where each of the descriptors may be sent over different channels. At the receiver each descriptor can be decoded independently of the other descriptors. Here we consider *symmetric* or *balanced case*, where all descriptors have equal rates and thus carry the same amount of information. The perceived quality at the receiver is proportional to the number of correctly received descriptors. In contrast to multiple layered coding (MLC), where a strong dependency between base and enhanced layer exists, MDC has no prioritized layers. The use of MDC in future generation networks is motivated by three arguments. First, MDC reduces the end–to–end *delay* because the errors result only in degraded quality instead of frame retransmission. Second, MDC enables the support for *heterogenous network* devices, that is, the more advanced the device, the more descriptors it can use to reconstruct the source stream. Third, the fact that MDC splits the connection into multiple flows offers the possibility to obtain *diversity* gain by transmitting the flows via different paths (time, frequency, space, etc.) to the receiver. The effectiveness of the combination of MDC with multiple path transport (MPT) for video and image transmission over multi–hop wireless links is studied in [6].

In [5], an MDC technique called Multiple Description Scalable Coding (MDSC) was proposed. MDSC can simply be described as a combination of MDC and scalable coding. This technique addresses receiving device characteristics and bandwidth variations of the channel and also enables tradeoffs between throughput, redundancy and complexity which is not possible with non–scalable MDC schemes. In MDSC, the number and the composition of descriptions are changed dynamically to make the proposed system very robust to changing channel characteristics. This work is similar to one of our sub–modes. We will outline that this approach assumes to have the full channel information and is therefore more complex than the schemes that we will propose later in this document. Furthermore we will highlight even the support of multi–cast services, which is not addressed in [5].

This paper introduces descriptor selection schemes for services transported through MDC streams, as they are expected to be important part of the 4G wireless communication systems. While an information source is split up into multiple descriptors, terminals with less capabilities may simply discard or not receive some of the descriptors, while the high–class terminals try to receive all information. Observing the fact that descriptor selection strategies affect the resulting video quality in the case of partial reception of descriptors, we introduce novel descriptor selection schemes for MDC services. The proposed schemes differ with respect to the availability of a feedback channel. All solutions are terminal oriented, while the fairness among the terminals is inherently achieved.

2 Motivation

Our objective is to optimize the MDC–based services in a heterogeneous networking scenarios with unreliable wireless links. In order to optimize the perceived video quality, different mechanisms are presented to select the transmitted / received set of MDC descriptors. Given the number of received descriptors, there are optimal combinations of descriptors which maximize the video quality. Intuitively, in the set of decoded descriptors the information carried by each descriptor should have minimized correlation with the information in the other descriptors, such that the amount of pure information put in the reconstructed source frames is maximized. Clearly, when some descriptors are left out, the video quality proportionally degrades. Note that this degradation is particularly significant if a poor combination of descriptors is used in the source reconstruction. To investigate how much the performance degrades in case of loss of the MDC descriptors, we perform a set of initial measurements. For the quality measurements we focus on the H.26L encoder. The MDC is done by splitting the video stream on the basis of frames as explained in [3]. Video quality is calculated in terms of Picture to Signal Noise (PSNR) values. The calculation is done as given in [2]. The PSNR calculations were done using the videometer tool [1]. The videometer tool is additionally able to freeze video frames in case the following frames are lost. This is important to have some sort of error resilience.

In Figure 2(a) the PSNR investigations using the foreman video sequences in the Quarter Common Intermediate Format (QCIF) format versus percentage of successful received sub–streams are presented. We have obtained these results with different values of J, which is the total number of descriptors generated by the application. The successfully received sub–streams are picked randomly. The random approach is our reference scheme, as in this case the descriptors are chosen without any preferences. For each combination (number of sub–streams and percentage), we have repeated the simulations multiple times with a confidence interval of 99%.

Through these results we can observe that the PSNR values degrade for a larger number of J even if the same percentage of sub–streams is received. As an example consider the point where 50% of the descriptors are received for $J = 2$ and $J = 20$. With $J = 2$ the PSNR value is almost 31 dB, while $J = 20$ leads to a value slightly above 29 dB, respectively. The reason is that for $J = 2$ we alternatingly loose and receive a frame, while for $J = 20$, we might end up in not receiving any frame for a long time and then receiving ten frames in a row. The latter case we refer to as worst case, while the former one is referred to as best case. As a large number of J is expected in future 4G systems in order to support a large variety of terminal classes, we motivate the importance of the investigations that follow. If we now force the system to operate in the best case (only using the best possible sets of descriptors), results in Figure 2(b) can be obtained. Now the PSNR values are mainly the same if the percentage of received descriptors is the same independently of J. On the other hand, the results in Figure 2(c) represent the worst case. In case of $J = 20$ and 50% of the sub–streams are received the PSNR value is only 27 dB. By this initial investigation we see that by a sophisticated descriptor selection we can gain a lot of video quality even when a limited number of descriptors is used. This motivated us to develop different mechanisms for selecting the descriptors that the receiving terminal uses to reconstruct the source video stream.

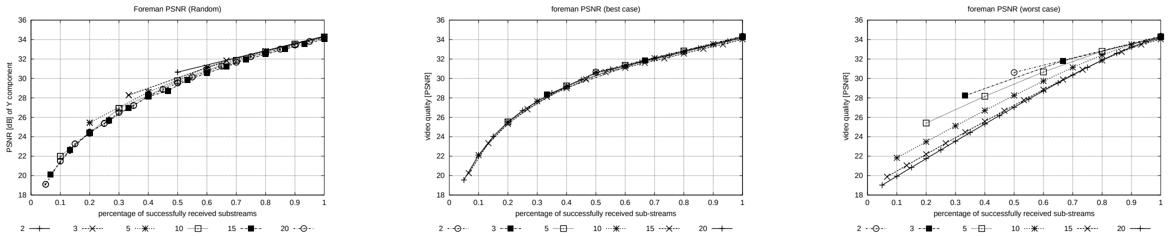
3 The New Approach

3.1 Terminology and Notation

A *channel* is the physical realization of resource allocation for communication between two entities. Channels are used to transmit packets between sender and receiver. We distinguish between good and bad channels. The communication is only possible over the good channels, while the bad channels always produce error. A *sender* sends packets to the receiver and it may optionally compress the headers, while a *receiver* receives the packets. Each channel transports a single descriptor, such that for J descriptors we assume existence of J parallel channels. The number of bad and good channels is denoted by J_{bad} and J_{good}, respectively, where $J_{bad} + J_{good} = J$.

3.2 The Proposed Descriptor Assignment Schemes

Motivated by our first results, we have developed optimized schemes to improve the video quality. We introduce two new entities which are in charge of assigning/selecting the best descriptors for highest possible video quality. Such entity can present either at the sender or at the receiver and it is called, respectively,



(a) random approach (b) best case (c) worst case

Figure 2: PSNR measurements using the foreman video sequences in the QCIF format versus percentage of successful received sub–streams.

Descriptor Assignment Entity for Multiple Description Coding (DAEMDC) and Descriptor Selection Entity for Multiple Description Coding (DSEMDC).

The descriptor assignment can be done either at the sender or at the receiver. For the sender–based assignment, there are two distinct modes: (1) A feedback mode (MODE–F), where the receiver informs the sender about the quality information, and (2) A random mode (MODE–R) where the used descriptors are chosen in a random manner. In the receiver–based approach there is only terminal mode (MODE–T), in which the best set of the received descriptors is selected. The transmitter based method is typically used in point–to–point scenarios (more spectrum efficient), whereas the receiver–based method is typically applied in multicast scenarios (more diversity).

3.2.1 Sender–based Descriptor Assignments

The sender's application conveys all descriptors to the DAEMDC. The DAEMDC can operate in two operation modes, namely MODE–R and MODE–F. The modes differ in the availability of a feedback channel from the receiver to the sender.

In MODE–F the receiver informs the sender about the set of J_{good} good channels. One possible solution for this status check would be to analyze the Real Time Control Protocol (RTCP) receiver reports. When the set of good channels is known, the DAEMDC on the sender side rearranges the streams in such a way that the good channels are equally distributed over the whole set of descriptors as given in Figure 3. For example, one possible solution would be the case if 6 out of 18 channels are available, descriptors 1, 4, 7, 10, 13, and 16 are transmitted. Other possibilities resulting in the same video quality can be found. While transmitting on the chosen channels the other channels are probed out. Probing can be done at higher layers using RTP/RTCP or at the link level using the given methods. In case one additional channel is available the descriptors are rearranged. In our example, if 7 out of 18 are available, one possible solution would be to use the descriptors 1, 3, 6, 8, 11, 13, and 16. A rearrangement is also necessary if the number of good channels decreases. The results obtained for MODE–F are given in Figure 2(b).

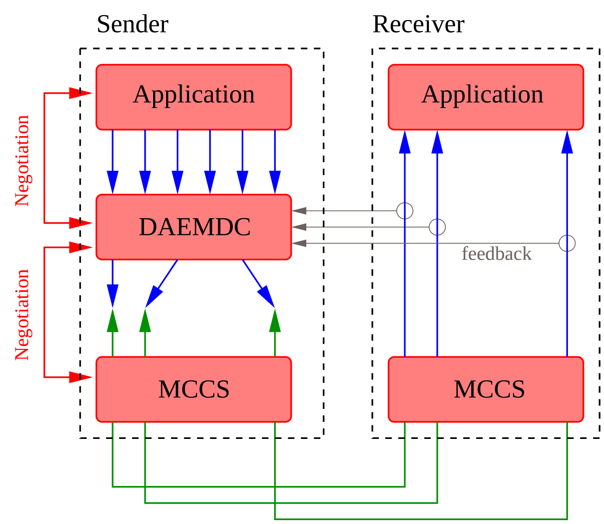


Figure 3: DAEMDC in MODE–F.

In MODE–R (referring to random) the sender has no feedback from the receiver on the quality information. Therefore, the sender mixes the streams randomly as given in Figure 4. This assignment is changed after a given time period and the instants of such change are discussed further in the text. This descriptor assignment statistically avoids the worse case and results in performance that is worse than MODE–F. The results for MODE–R are shown on Figure 2(a).

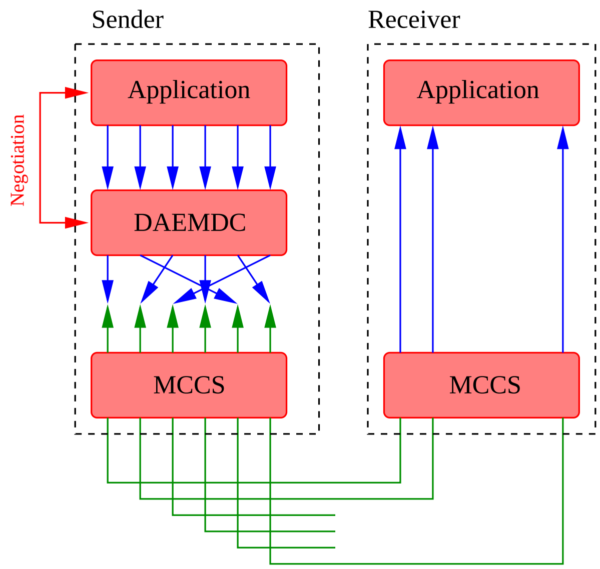


Figure 4: DAEMDC in MODE–R.

The advantage of MODE–R is that no feedback channel is required. Regarding Figures 2(a), 2(b), and 2(c), MODE–F achieves values as given in Figure 2(b), while MODE–R achieves results as given in Figure 2(a). It is important to note that due to the changing assignment of descriptors in MODE–R, the results do not vary anymore between those given in Figure 2(c) and those in Figure 2(b). A static assignment would make it difficult for a network provider to charge the users for such a service. To illustrate this, consider two customers paying the same fee for the same service sitting together and have different service qualities. The dynamic assignment solves such unfairness problem.

For both modes we assume that it is possible to change the assignments at any time. However, for some applications the *ad hoc* change in assignments may result in additional problems e. g. in the case when the encoded video streams are based on I, P, B frames. If a new channel is chosen, the video will be displayed after the receipt of an I frame, when an I frame is used depends on the Group of Picture (GoP) value. If the channel is changed before an I frame is received, then the information from the P and B frames is also lost. To prevent occurrence of such loss, a negotiation is introduced between the DAEMDC and the application.

It is beneficial to know the time point when a channel may be changed i. e. when an I frame is transmitted. This is especially valid for the MODE–R operation. The video application might signal this condition to the DAEMDC (negotiation). Another possibility would be that the DAEMDC is scanning each RTP packet and searches for the bit settings for an I frame. But as we want to support multiple applications, it would be a huge overhead to get all this knowledge to the DAEMDC. A further improvement would be if the DAEMDC and the application could negotiate on the GoP structure. In case the channel conditions change rapidly, the GoP value has to be small, while in the other case, the GoP value may be high. Also in the other modes negotiation could be introduced to further optimize the system performance.

3.2.2 Receiver–based Descriptor Selection

Consider now the multi–cast case when multiple terminals are requesting the MDC information from the same source. Clearly, the network operators are interested to support a large number of terminals in a multi–cast service and the terminals can be heterogeneous. In this case we advocate the use of a receiver–based approach. Some terminals attempt to receive as many descriptors as possible and their performance is limited by the channel conditions. On the other hand, the terminals with modest processing capabilities will discard some of the received descriptors since the number of descriptors they can handle is limited. In Figure 5 the sender conveys six descriptors towards multiple receiver. In our example the receiver may receive four descriptors correctly (two are lost due to channel errors), but only three can be used at the terminal (e.g. due to its hardware limitation).

The question arises once again which descriptors would be the best to choose among the total number of received ones. Analogously to the DAEMDC, we introduce a new entity called DSEMDC. Apparently, in DSEMDC a feedback information is not needed since the selection is performed at the receiver. The single mode in this case is referred to as MODE–T (referring to terminal). In case of multi–cast transmission, each DSEMDC will choose different descriptors as the channel conditions differ among the receiving terminals.

As for MODE–R and MODE–F, the problem when to switch between the channels has to be addressed also for MODE–T. This is especially important when more channels are available than the number of chan-

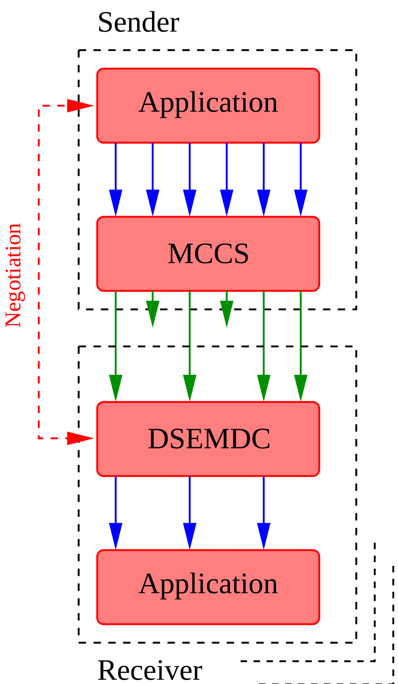


Figure 5: DSEMDC in MODE–T.

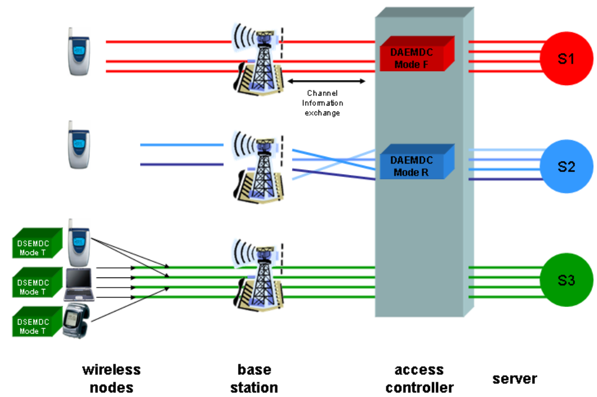


Figure 6: Application of our approach into the cellular concept.

nels which can be used for decoding (e. g. in low cost terminal). Here we can rely on the information of the GoP once it is available. If a negotiation is needed, this has to be done over the wireless link, but such operation becomes even more complicated since the same content is received by multiple receivers.

4 Network Architecture

In Figure 6 we give one possible and general architectural solution for the application of our idea for wireless networks. We illustrate it by three different services provided by the servers S1, S2, and S3. The DAEMDC entities can be placed in the access controller and are therefore part of the backbone. For the MODE–F, as it requires feedback information, it would be more efficient to place it directly into the base station. To lower the complexity at the base station, we place all DAEMDCs in the access controller for both MODE–F and MODE–R. For MODE–F, the access controller and the base station should exchange quality information.

For MODE–T the DSEMDC entity is placed in the terminal. No changes in the core backbone are required. Note that this is especially interesting for multi–cast services and therefore the network has to support this service type. The proposed network architecture can be applied to existing cellular networks as well as upcoming WLAN distribution networks. Even their combination, as specified by 3GPP [7] or the Unlicensed Mobile Access (UMA) [8] consortium, is possible and is supported by our approach.

5 Results

In this section we give results achieved for well known and accepted video sequences. The PSNR values versus the percentage of received sub–streams are given for the three modes and the worst case. Different scenarios were under investigation varying the number of channels/descriptors from 1 to 20. Due to space limitation we only depict those results for $J = 20$. The worst case scenario was chosen in order to bring forward the need for fairness. Without applying our approach, the video quality for the same video sequence (e.g. a multi–cast transmission) may vary between the best and the worst case.

Regarding Figure 12 PSNR values of 34 dB are achieved for all modes if 100% of the $J = 20$ descriptors are received. But for medium and low cost terminals the results differ dramatically among the modes. Assuming that a low cost terminal is only able to receive 25% of the sub–streams (thus $J = 5$), MODE–F and MODE–T achieve 1 dB higher results than MODE–R, which in turn is 3 dB better than the worst case. Regarding the fairness the quality span would be 4 dB in this case, which can be described as significant and customers would not accept this difference in quality with the same price for it.

In Figure 7, 8, 9, 10, and 11 the PSNR results for the video sequences Carphone, Claire, Container, Foreman, Highway, and Silent are given, respectively. Note, that these results are related to the specific video sequences and change among different sequences. MODE–F and MODE–T result in the same PSNR values as we assume a perfect feedback channel and that in MODE–T the number of received streams is much higher than the ones that can be operated on.

An advantage of the MODE–R usage is reflected in the following fact: Since the descriptor set is picked randomly for each user, then all users experience the same average performance. Thus, our approach prevents the customers from perceiving varying quality situations. As we motivated, to support heterogeneous terminals we should use large values for J, since our approach achieves the best performance in this case and, consequently, provides the best support for heterogeneous terminals. For high cost terminals (receiving a high percentage of the descriptors) the choice which mode to use is not that important. On the other hand, for medium and low cost terminals the choice of the appropriate descriptors becomes vital.

6 Conclusions

In this paper the use of Multiple Description Coding (MDC) is proposed to support heterogeneous terminals in future generation wireless networks such as envisioned in 4G. Based upon the capabilities of the various terminals, all the descriptors or only a subset of the descriptors can be used in the reception. This way the application can be kept unaware of the requirements of the individual terminals.

The data flow should be split into large number of descriptors, in order to support a large number of heterogeneous terminals within the system. Terminals with advanced capabilities will use many descriptors, while the terminals with limited capabilities will utilize proportionally smaller subset of descriptors.

When only a subset of the descriptors is used, the performance is optimized through the selection of an appropriate set of descriptors to be decoded. We have proposed two types of selection mechanisms: sender– and receiver–based. A sender–based mechanism selects a set of descriptors to be used and does not transmit data via the remaining ones. The choice of the subset can be random or greedy, and accordingly two modes are defined. In the greedy case the feedback is used to select only the best descriptors. Since only a limited set of descriptors is used, this method is more bandwidth efficient and more suitable for peer-to-peer connections. The receiver–based mechanism determines the best subset of the received descriptors and uses them to reconstruct the source stream. This method is suitable for multi–cast applications, where many terminals receive the data and they can each chose the best descriptors. To conclude, the advantages of our proposal are i.) support of heterogeneous terminals, ii.) improved video quality for medium– and low–cost terminals, iii.) the introduced solutions keep the terminal to be with low complexity and iv.) fairness among terminals can be assured.

In our future work we will investigate performance improvement due to the application of different, more advanced MDC techniques, such as unbalanced descriptors and quantizers. It has to be investigated if the overall performance using quantizers is different from our frame–based approach.

References

[1] F.H.P. Fitzek, P. Seeling, and M. Reisslein. VideoMeter tool for YUV bitstreams. Technical Report acticom-02-001, acticom – mobile networks, Germany, October 2002.

[2] F.H.P. Fitzek, P. Seeling, and M. Reisslein. *Wireless Internet*, chapter Video Streaming in Wireless Internet. Electrical Engineering & Applied Signal Processing Series. CRC Press, 2004. To be published.

[3] Frank H.P. Fitzek. Application and Network Measurements For Multiple Description Coded Video Sequences. Technical Report R-04-1005, Aalborg University, April 2004. ISNN 0908-1224.

[4] S. Frattasi, H. Fathi, F. Fitzek, M. Katz, and R. Prasad. A New Pragmatic Methodology to Design 4G. In *International Conference on Networking (ICN)*, Reunion Island, April 2005.

[5] M. Schaar and D. S. Turaga. Multiple Description Scalable Coding Using Wavelet-Based Motion Compensated Temporal Filtering. In *IEEE Int'l Conf. on Image Processing*, volume 3, pages 489–492, Barcelona, Spain, September 2003.

[6] N. Gogate, D. M. Chung, S. S. Panwar, and Y. 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*, 12(9):777–834, September 2002.

[7] Technical Specification Group Services and System Aspects. 3GPP system to Wireless Local Area Network (WLAN) Interworking; System Description (Release 6). Technical Report V2.4.0, 3rd Generation Partnership Project, Jan 2004.

[8] UMA. Unlicensed Mobile Access (UMA). Stage 1/2/3, UMA Consortium, October 2004.

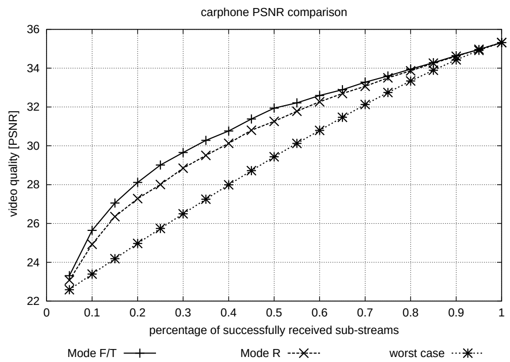


Figure 7: PSNR comparison for different modes and the worst case investigating carphone video sequences versus percentage of successful received sub–streams.

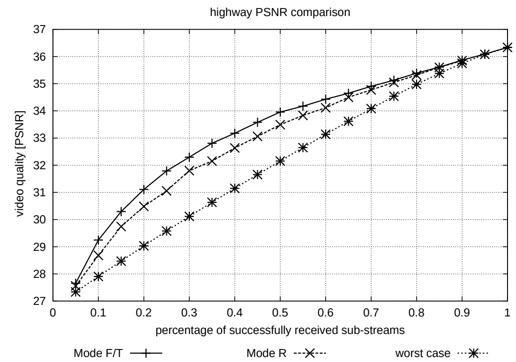


Figure 8: PSNR comparison for different modes and the worst case investigating highway video sequences versus percentage of successful received sub–streams.

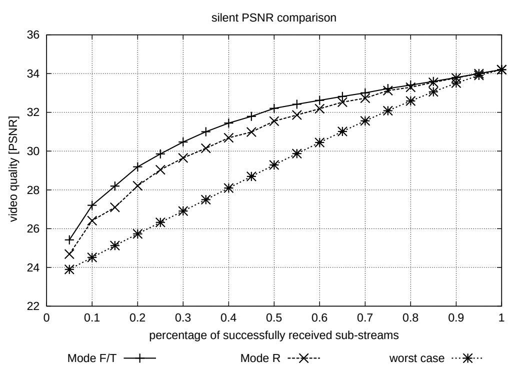


Figure 9: PSNR comparison for different modes and the worst case investigating silent video sequences versus percentage of successful received sub–streams.

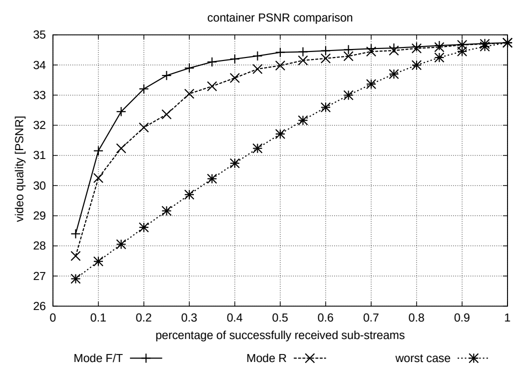


Figure 10: PSNR comparison for different modes and the worst case investigating container video sequences versus percentage of successful received sub–streams.

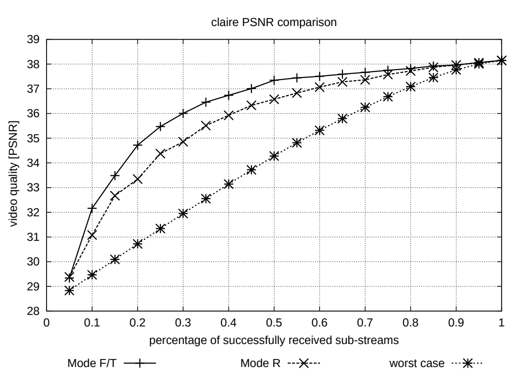


Figure 11: PSNR comparison for different modes and the worst case investigating claire video sequences versus percentage of successful received sub–streams.

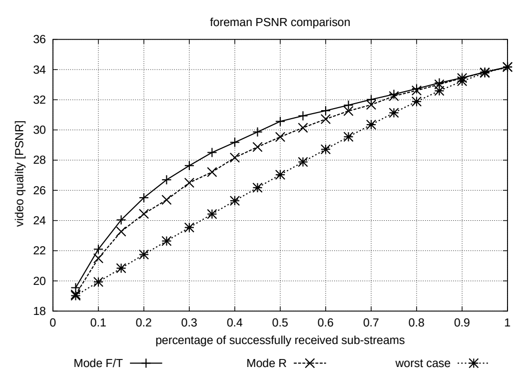


Figure 12: PSNR comparison for different modes and the worst case investigating foreman video sequences versus percentage of successful received sub–streams.