Simulation Platform for Multimedia Broadcast over DVB-SH

Guenther Liebl
Nomor Research GmbH
Brecherspitzstr. 8
81541 Munich, Germany
liebl@nomor.de

Ktawut Tappayuthpijarn
Nomor Research GmbH
Brecherspitzstr. 8
81541 Munich, Germany
champ@nomor.de

Karsten Grüneberg
Fraunhofer HHI
Image Processing Dept.
Einsteinufer 37
10587 Berlin, Germany
grueneberg@hhi.fraunhofer.de

Thomas Schierl
Fraunhofer HHI
Image Processing Dept.
Einsteinufer 37
10587 Berlin, Germany
schierl@hhi.fraunhofer.de

Cedric Keip
Fraunhofer IIS
Communications Dept.
Am Wolfsmantel 33
91058 Erlangen, Germany
keip@iis.fraunhofer.de

Holger Stadali
Fraunhofer IIS
Communications Dept.
Am Wolfsmantel 33
91058 Erlangen, Germany
stadali@iis.fraunhofer.de

ABSTRACT

In this paper, we want to present a simulation platform for multimedia broadcast over DVB-SH. The platform has been designed and implemented within the context of the ESA study on "Scalable Video Coding Applications and Technologies for mobile satellite based hybrid networks". Its main features are realistic modeling of the multimedia encoding/decoding procedures done at the broadcasting headend and the mobile terminals, a detailed simulation of the DVB-SH link layer, and a representative emulation of the DVB-SH physical layer. With the help of this platform, important use cases for future DVB-SH systems, such as support of heterogeneous devices, graceful degradation of application quality across the coverage area, efficient statistical multiplexing of different services, etc. can be investigated in a straightforward and conclusive manner.

Categories and Subject Descriptors

J.2 [**Physical Sciences and Engineering**]: Engineering

General Terms

Experimentation, Measurement, Performance

Keywords

Simulation platform, multimedia broadcast, DVB-SH link layer simulation, DVB-SH physical layer emulation

1. INTRODUCTION

DVB-SH, short for Digital Video Broadcasting - Satellite services to Handhelds, is a standard for broadcasting IP-based media content and data to handheld terminals, like mobile phones or PDAs, via geostationary or quasi geostationary satellites, at frequencies below 3 GHz. Of particular interest are the S-UMTS bands which have been harmonized in Europe by the European Commission and worldwide by the ITU. Satellite use allows economical and fast deployment of nationwide coverage. A typical satellite coverage design using a 3-color frequency reuse pattern is shown in Figure 1.

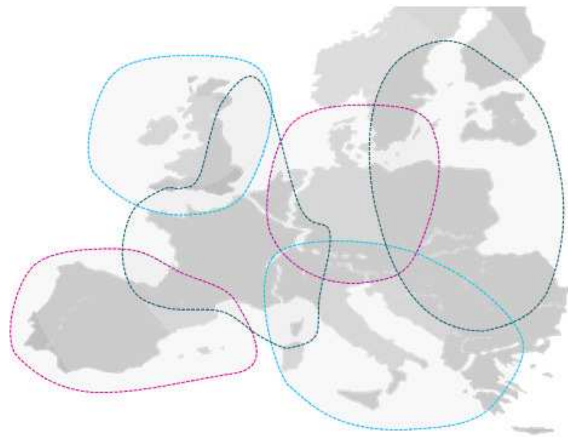

Figure 1: Example of satellite coverage design in S-UMTS band for Europe.

The satellite signal can be seamlessly complemented by a network of terrestrial gap-fillers, called Complementary Ground Component (CGC), in areas where the satellite signal would be severely blocked by obstacles such as urban areas. The use of CGC brings the ability of local content insertion, which may further enhance the variety of the service in "hot spots". The CGC re-uses the same frequency allocation as the satellite system it is associated with. In Figure 1 the total amount of frequency used by this hybrid Satellite/Terrestrial network planning is 3x5MHz.

While the DVB-SH standard itself already features a lot of options for tuning the system performance to meet specific environmental and commercial constraints, future broadcast applications used on top of it offer additional optimization potentials. For example, the operator has the choice between different encoding techniques that may or may not "scale"

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

with varying channel conditions and/or heterogeneous terminals in the coverage region.

For this reason, an ESA-funded study[1] on "Scalable Video Coding Applications and Technologies for mobile satellite based hybrid networks" is currently conducted by a consortium of four partners: Nomor Research, Fraunhofer HHI, Fraunhofer IIS, and Eutelsat. The activity shall investigate the viability of integrating Scalable Video Coding (SVC) technologies over DVB-SH networks and disseminate its findings to relevant fora (e.g. DVB TM SSP).

In this work, we want to present the simulation platform for multimedia broadcast over DVB-SH that has been designed and implemented during the first phase of this study. Its main features are realistic modeling of the multimedia encoding/decoding procedures done at the broadcasting head-end and the mobile terminals, a detailed simulation of the DVB-SH link layer, and a representative emulation of the DVB-SH physical layer. With the help of this platform, important use cases for future DVB-SH systems, such as support of heterogeneous devices, graceful degradation of application quality across the coverage area, efficient statistical multiplexing of different services, etc. can be investigated in a straightforward and conclusive manner.

The remainder of this paper will be organized as follows: We will start with a brief introduction to broadcasting via DVB-SH and the multimedia encoding strategies that can be used in this context. Next, we will list the main requirements on the simulation platform we had to face and present the architecture that has been chosen to address them. A large portion of the paper will be then reserved for the detailed description of the various components and models we have developed for and integrated into the simulation platform. The final part of the paper will provide a proof-of-concept, i.e., we will demonstrate that our platform allows to investigate one important reference use case for future DVB-SH systems: the support of heterogeneous terminal classes. The paper concludes with a summary of the major achievements and a brief outlook on future work in this area.

2. SYSTEM OVERVIEW

2.1 Broadcasting via DVB-SH

The DVB-SH standard was approved by the DVB Steering Board in February 2007 and published as European Norm by ETSI in July 2007 [2]. It can be considered as "best of breed" as it incorporates such advanced features as: Turbo codes (borrowed from 3GPP2), OFDM waveform (borrowed from DVB-T/H), and a satellite optimized single-carrier TDM waveform (borrowed from DVB-S2). Depending on the satellite amplification architecture, the available bandwidth, and the local content insertion need, the satellite signal can be transmitted with either OFDM waveform (SH-A mode) or TDM waveform (SH-B mode). The CGC uses only OFDM, as it is more resilient to multipath.

There exist situations where the satellite signal can suffer intermittent block-fading and the CGC can not be viably deployed, such as on rural roads with heavy vegetation and/or low elevation view to the satellite. For these cases, the standard provides mitigation tools in the form of: (i) a time interleaver at the physical layer and (ii) an UL-FEC, called MPE-IFEC, at the link layer. Both have programmable

[1]European Space Agency, ARTES1, CTR 21937, "SVConS"

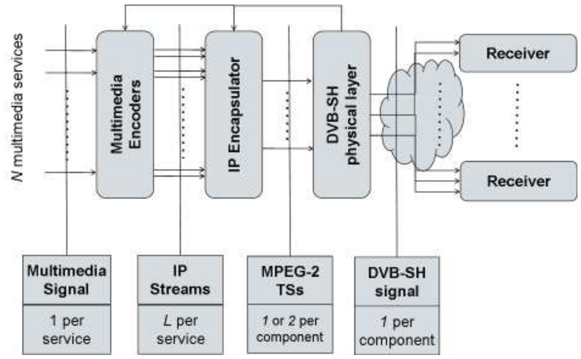

Figure 2: Typical end-to-end DVB-SH system architecture.

variable protection length and other parameters to adapt the error correction to various error patterns. Although the physical-layer time interleaver performs best for a given protection length and signal-to-noise ratio, it is quite memory demanding (it operates on soft bits and at the full multiplex rate). While the focus is on the battery-powered mobile phones, DVB-SH should also provide services to vehicular-installed terminals where power consumption and form factor are more relaxed, with on the other hand larger viewing screen size.

A typical end-to-end DVB-SH system architecture, as envisioned for most practical deployments [1], is shown in Figure 2. In general, the service offering consists of a total of N multimedia services, each of which is encoded in one or several sessions. In particular, if scalability is used for the encoding, the multiple sessions reflect the different layers (see section 2.2 for more details). This results in a total of L sessions per service, whereby a session is represented by a distinguishable IP stream (e.g. different ports, different IP address, different descriptors in the media header).

The IP encapsulator's main role is to map all incoming IP streams to one or several MPEG-2 transport streams (TS), depending on the number of transmit components and whether hierarchical modulation (HMOD) is available or not. This mapping includes the multiplexing of the different streams, time slicing, MPE encapsulation, and (optional) MPE IFEC generation. The IP encapsulator must also ensure that the "available" bit rate on the MPEG-2 transport stream(s) is not exceeded and that the time-slices are generated in such a way that channel switching times are not too excessive. For this reason, it may provide feedback on the current multiplex size to the multimedia encoders or perform subsequent rate adaptation (only possible if the media streams are scalable). Finally, the IP encapsulator maps the MPE sections into MPEG-2 TS packets for broadcasting over the satellite and/or terrestrial link to the receivers.

2.2 Multimedia Encoding for Broadcasting Applications

Todays digital video broadcast systems are using different video codecs. The first generation, introduced 1994, has been based on MPEG-2 video and systems specifications which are still widely used. Recent broadcast systems use the MPEG-4 AVC / ITU-T H.264 (AVC) video codec specification which roughly doubles the efficiency with respect to MPEG-2 video. Scalable extensions, in the latter referred

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

to as Scalable Video Coding (SVC), have been added to the AVC standard [5]. SVC enables novel applications, such as the support of heterogeneous terminal devices with the same bit stream or graceful degradation of the perceived video quality under varying reception conditions. An SVC bit stream consists of two or more layers which correspond to different levels of spatial or temporal resolution or fidelity in terms of signal-to-noise ratio (SNR). Upper layers depend on lower layers because the coding can imply prediction of texture or motion data from the same or lower layers. The lowest layer (called base layer) is a plain AVC bit stream which can also be decoded by legacy decoders. Scalability implies that part of the bit stream can be discarded, leaving a remaining stream which can still be decoded, resulting in a lower quality of the output video. The main feature of SVC is that the coding efficiency at each quality level is almost the same with respect to the reduced amount of data. In fact, an SVC encoded stream that includes two layers of spatial or SNR scalability yields the same quality in terms of peak signal-to-noise–ratio (PSNR) compared to a single layer AVC video stream typically at a data rate that is 10% higher than the AVC data rate. In most cases, simultaneous transmission (Simulcast) of both layers individually using an AVC encoder requires significantly more bit rate.

The scalable bit stream itself is structured according to the Network Abstraction Layer (NAL) syntax which had already been introduced with AVC. Adaptation of the scalable bit stream is done simply by dropping NAL units starting from the highest layer, thus keeping a decodable set of NAL units belonging to one or more lower layers. In a broadcast system, care has to be taken that a terminal receives at least a decodable subset of the original stream. Finally, if the broadcast system provides means for unequal error protection (UEP), the base layer should be best protected, while the protection may be decreased for each upper layer.

3. PROPOSED SIMULATION PLATFORM

3.1 Platform Requirements

The main requirements on the overall simulation platform architecture and interfaces are:

- The simulation architecture should closely reflect the functional split in the DVB-SH system architecture depicted in Figure 1.
- The design should be as modular as possible, taking into account the individual expertise of the different partners in the project.
- The architecture must be flexible, such that
  - extension to combinations of existing use cases or to newly defined use cases is straightforward,
  - replacement of certain modules with other reference implementations of the respective functionality is in principle possible.
- The design must clearly distinguish between static and dynamic parts of the simulator, i.e., between modules that only perform offline pre- or post-processing of data and those that are active during a simulation.
- The definition of interfaces between modules should be as generic as possible, i.e., cover all current and potential future functionalities of the respective modules.

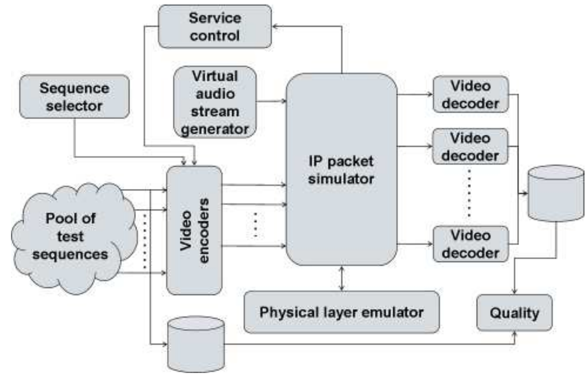

Figure 3: High-level simulation architecture.

To ensure that the simulation results are of relevance and can be disseminated to the scientific community and/or standardization bodies, the following is required:

- All modeling and abstraction processes applied during the design and implementation of the simulator parts must match current best practice in this field of research or technology.
- When designing the reference system for performance comparison, the focus should be on accurate and realistic modeling.
- All simulation results must be reproducible in multiple runs of the simulation platform and via potential other external reference implementations of the proposed simulation architecture.
- All simulation results must be statistically meaningful, i.e., require sufficient diversity in the simulation conditions during one or multiple combined runs.

3.2 Platform Architecture

The underlying assumption of the proposed simulation architecture is that some of the processes that happen in real-time in deployed systems are done offline such that both the simulation complexity and run time remain within reasonable bounds. Nevertheless, replacing some of the offline processes with a real-time emulation at a later stage is straightforward and already considered in the interface design.

A high-level overview of the simulation architecture is shown in Figure 3. It consists of four major parts:

- the multimedia encoding/decoding modules,
- the IP packet simulator,
- the physical layer emulation,
- the application-level quality evaluation module.

The main interactions between the different parts during a simulation run are as follows: For each of the N offered multimedia services, the video content to be encoded in each time slicing round is chosen from a pool of test sequences. This selection process is done in a reproducible and statistically relevant manner, and the generated video sequences are stored as reference for the later quality evaluation.

The bench of Video Encoders then generates the corresponding RTP/UDP/IP stream for each service using the

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

video coding format (e.g., AVC or SVC) specified at simulation startup. In addition, a (dummy) audio component is also inserted in each output stream. The overall generation process is controlled such that the resulting streams fulfill certain predefined bit rate and delay constraints, but also exhibit good quality across all services. The Service Control may also receive dynamic feedback from the IP packet simulator to update the bit rate target in each time slicing round, if required. For more details on the various multimedia encoding procedures, the reader is referred to section 4.

The generated test streams are fed to the IP packet simulator, whose task is to generate the delay and loss characteristics for each service based on the specific DVB-SH system configuration chosen at simulation startup. For this reason, the IP packet simulator contains a fully-featured model of the DVB-SH link layer processing done at the broadcast transmitter and the terminal receiver, which is described in more detail in section 5.

Emulation of the DVB-SH physical layer processing of the MPEG-2 transport stream(s) and the actual transmission process over the satellite and/or terrestrial link is handled by a separate module that communicates with the IP packet simulator in each time-slicing round. Besides realistic error flags at MPEG-2 TS level, this module supplies the necessary advances in transmitter and receiver timing due to the physical layer processing and propagation delay. The internal structure of this PHY emulator, including all relevant system and channel models, can be found in section 6.

Finally, the bench of Video Decoders processes the received media streams at the output of the IP packet simulator and stores them for subsequent quality evaluation. The latter is then performed taking into account the original and the decoded sequences, as described in section 4.4.

3.3 Scenario and Results Database

The behavior of the different platform components during a simulation run is controlled via a large set of parameters that can be chosen at startup. Each practical combination of parameter values constitutes a so-called "simulation scenario". The latter can be offline pre-configured and stored within a MySQL database, such that the user of the platform only has to specify a desired scenario and number of services to simulate at startup. All relevant parameters are then automatically read from the database. Thus, simulation runs can either be repeated or new scenarios can be added easily without having to change existing setups.

In addition, all relevant results that are produced during a simulation run are also stored in this database, together with a timestamp and the label of the chosen scenario. Thus, offline analysis of the results is feasible with a direct reference to the parameters used for the underlying simulation run. The platform also features scripts that automatically convert the stored results into the format defined for a certain interface, e.g., PCAP traces for the video decoder.

4. MULTIMEDIA ENCODING/DECODING PROCEDURES

4.1 Generation of Pre-Encoded Streaming Content

In order to allow reproducible simulation results, all video content is encoded off-line and stored in media files according to the MP4 file format standard which has recently been

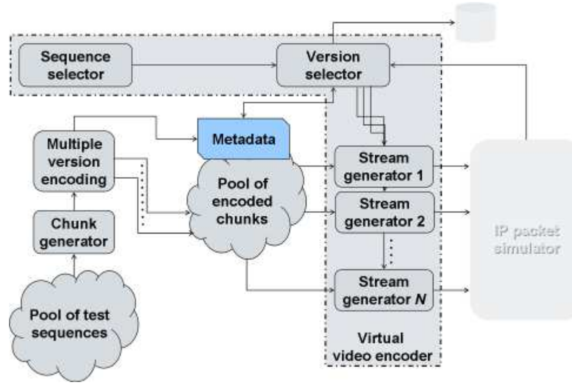

Figure 4: Encoding architecture

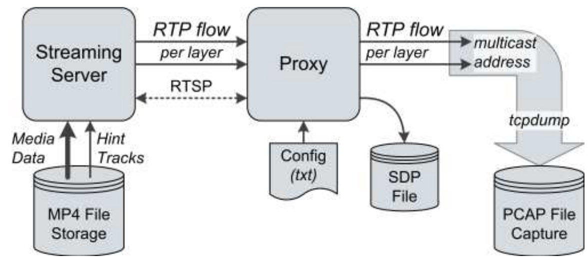

Figure 5: Capturing PCAP files

extended with features for SVC [3]. Instead of a continuous bit stream, short chunks starting with a random access point are generated which by concatenation allow for the construction of a long test sequence. Each chunk is encoded several times with different features so that the simulator can dynamically select versions that match certain data rates or quality parameters as shown on the left side of Figure 4. In order to enable this selection process, a data base is provided which indicates bit rate and resulting PSNR per layer for each version of a chunk (blue shaded block in Figure 4).

As the transmission includes the Real-time Transport Protocol (RTP), so-called server hint tracks are added to the MP4 file. These provide all necessary information for media encapsulation according to the appropriate RTP payload format [6] so that a media agnostic streaming server can handle the data correctly. For real-time experiments and demonstrations, a streaming server is used to generate an RTP/UDP/IP stream in a local network towards a multicast address. This stream is captured in PCAP file format as shown in Figure 5 and can be replayed at any time.

In addition to the MP4 file, a so-called trace file which contains a textual description stating timestamps, length, and other properties of the RTP packets is generated for each chunk. In the simulation environment, this information can be processed instead of the media data itself in order to speed up the evaluation. Furthermore, trace files are needed to keep track of the actual sequence of chunks which have been selected by the simulator. This plays an important role during the evaluation of the simulation results.

4.2 Simulation of Rate Control Functionality

Simulation of the rate control functionality in each round is done by the Virtual Video Encoder (VVE) shown in Figure 4. The latter aggregates the Seqence Selector, Version

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

Selector, and Stream Generator modules. Processing inside the VVE is done in three steps: First, the video content to be encoded for each service is chosen by the Stream Selector. The current implementation supports two options: The content can be chosen such that the final concatenated chunks resemble a continuous video sequence which is suitable for subjective viewing. Alternatively, the content can be randomly chosen such that the final concatenated chunks resemble a mixture of different video sequences, which provides a high degree of randomness and low temporal correlation both with itself and to other services.

Once the content of each service has been fixed, the Version Selector will choose a suitable encoded version for each chunk based on certain configurable criteria, such as desired bit rate and stream quality. The Version Selector can perform this task independently for the different services or jointly optimize the version selection amongst all services to maximize the average quality at the output.

Since the chosen chunks may be created from different original video sequences, they need further modifications to their packet headers so that they can be concatenated, decoded, and played successfully at the receiving end. Specifically, some information fields in the RTP and UDP headers, such as the sequence number, as well as the source and destination port, need to be modified to make them consistent and appear as they were from the same stream. This is done by the Stream Generators before the encoded chunks are passed on to the IP simulator for further processing.

4.3 Post-processing of Simulator Output

The simulator output is generated two-fold: First, the RTP packets that have been received correctly are written into another PCAP file. This file can be replayed and decoded by a client application, showing the actual video quality for the simulated transmission scenario. Second, a trace file is produced which consists of a modified concatenation of the input trace files, after removing those lines that correspond to RTP packets which have been lost during transmission. In a basic approach, this second output would be used to reconstruct the received video stream which would then be decoded and compared with the original sequence of raw video chunks. In our simulation system, we use a robust SVC decoder which can process bit streams that have suffered from transmission errors. If the base layer of an Access Unit is also corrupted, a simple "freeze frame" insertion will take place. When decoding spatial scalable streams, if the base layer is completely received, but the enhancement layer cannot be used due to transmission errors, the decoder can upscale the base layer or use a "freeze frame" of the enhancement layer. Similar decisions have to be made in case that subsequent frames depend on data that has been lost during transmission and hence is not available at the decoder. Regarding such decisions, the encoder behavior can be configured. Due to the large number of simulation runs, we use a more efficient approach described next.

4.4 Application-level Quality Evaluation

During the simulations, transmission parameters are varied and the same video chunks are transmitted multiple times. Each chunk may either be received completely, or one or more RTP packets may be missing due to transmission errors. Although it is not known beforehand which packets will be lost, there is a finite number of error patterns which can occur. Each error pattern can be identified by a binary vector of which the length corresponds to the number of RTP packets in that chunk. Although the number of error patterns is huge, many of them will result in the same output sequence because the number of variations is limited due to the fact that each frame can only be decoded at one of the scalability levels. Therefore, we generate a data base containing the pre-calculated PSNR for each frame on each scalability level and map all error patterns on the possible variations of the decoded output.

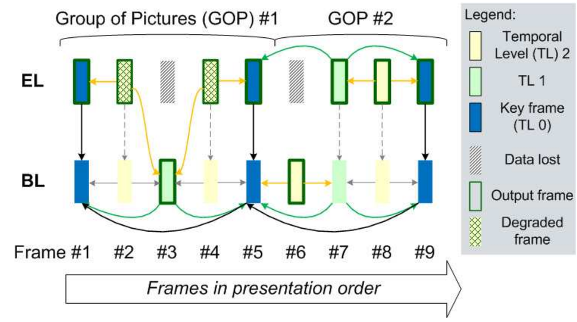

Figure 6: Decoding in presence of errors

Figure 6 shows an example of decoding dependencies between frames, indicated by arrows, for a bit stream with two layers of fidelity scalability, where BL and EL stand for base and enhancement layer, respectively. Due to the fact that inter frame prediction is structured hierarchically constituting so-called Groups of Pictures (GOP), the number of different error patterns for each GOP is very low and can be handled easily. Besides combinations that exist within each GOP, some additional cases have to be evaluated across GOP boundaries, namely if the previous frame at temporal level (TL) 0 has been decoded at a lower scalability level or been replaced by a freeze frame. In the data base, the chunks were encoded with an initial intra coded frame followed by three GOPs of size 8. In this constellation, we have to precalculate 3 times 278 PSNR values for each GOP. Once the PSNR data base is prepared, we can evaluate the outcome of all simulation runs by a GOP-wise mapping of the error patterns to the entries in our data base which speeds up the evaluation significantly so that it performs faster than real-time.

5. DVB-SH LINK LAYER SIMULATION

5.1 IP Packet Simulator Functionality

Simulation of the DVB-SH link layer functionality is handled by the so-called IP packet simulator. The latter represents the central module of the proposed simulation platform and thus interfaces to both the multimedia encoding/decoding modules and the physical layer emulator, as shown in Figures 3 and 4. In order to generate delay and loss characteristics for each service that are representative for a practical DVB-SH system, the following link-layer functionalities are modeled in detail: multi-protocol encapsulation (MPE) of IP packets, (optional) link-level error protection via IFEC, time-slicing, service multiplexing, and mapping to adequately generated MPEG-2 transport streams (TS). In addition, dynamic rate control with (optional) feedback to the multimedia encoding module is supported.

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

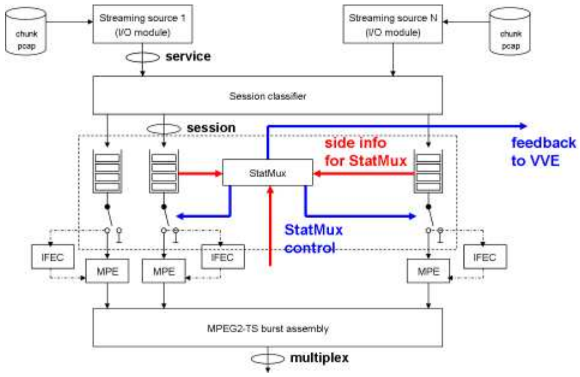

Figure 7: IP packet simulator architecture - Tx side.

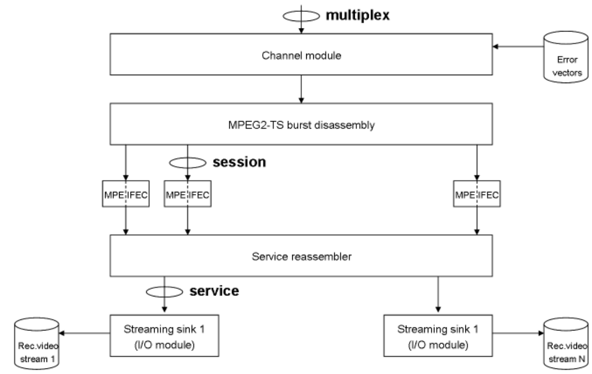

Figure 8: IP packet simulator architecture - Rx side.

The internal architecture of the transmitter and receiver part of the IP packet simulator is depicted in Figs. 7 and 8, respectively. In the following two subsections, the models used at the various processing stages will be explained in sufficient level of detail.

5.2 Model of the Transmitter-side Processing

The input to the IP packet simulator are N PCAP files per simulation round, i.e., one for each offered multimedia service as generated by the N Stream Generators depicted in Figure 4. These files contain the RTP/UDP/IP packets that correspond to the encoded media streams (including virtual audio) to be encapsulated at the DVB-SH link layer in the current simulation round.

For each service, the corresponding Streaming Source module located on the topmost part of Figure 7 reads the associated PCAP file and forwards the binary data to the Session Classifier. In a first step, the latter parses the data to extract all the information from the IP, UDP, RTP, and (optional) NAL headers that will be required at the various stages of the simulator. This meta-information is then stored, together with a pointer to the location of the actual packet in the system memory, in an internal format that is easier to handle and will be denoted as "IP packet" in the following. Once all information has been read, the PCAP file for the video stream is deleted such that the Stream Generator can overwrite it with new data for the next round.

In the second step, the Session Classifier assigns each new IP packet to one of the pre-defined sessions for each service based on a set of filter rules (e.g., by using the service id, IP address, port number, SVC layer, etc.) that match the current use case. For example, packets that contain SVC base layer data or audio information might be assigned to one session, while packets that contain SVC enhancement layer data might be assigned to a second one. All packets belonging to a single session are then stored in a separate FIFO queue to await further processing.

The StatMux module inside the IP packet simulator controls the drain rates from all the queues towards the multiplex. It has real-time knowledge on the available throughput of the physical layer, the current link layer overhead, and the input rates to the queues, shown as red arrows in Figure 7. Thus, it is able to adjust the drain rates appropriately in order not to exceed the available throughput at MPEG2-TS level. Three basic mechanisms for the StatMux operation will be considered: In the so-called "re-encoding" mode, which works for both AVC and SVC encoded streams, the internal StatMux module provides feedback to the external Virtual Video Encoder after each simulation round. Based on this feedback the latter then selects an appropriate version for the chunk of each service in the next round such that the bit rate constraint inside the IP packet simulator will be met by the aggregated input from all services. In the so-called "dropping" mode, which only works for SVC encoded streams, the internal StatMux module does not provide any feedback, but rather drops some packets from the enhancement layer of each service to reduce the overall bit rate, if required. The dropping rules, number of enhancement layer queues/sessions for each service, and the priority assignment among the queues and within the packets in each queue are configurable at simulation startup. Finally, in the so-called "hybrid" mode, which only works for SVC encoded streams, the internal StatMux module provides only a coarse feedback to the external Virtual Video Encoder, such that the bit rate of the next encoded chunks stays within a certain range. Final adaptation of the drain rates will then be handled as described for the dropping mode.

After the (surviving) packets have been released from the queues by the StatMux, they are passed on to the IFEC module in order to apply optional link-level error protection. Note that the classification into sessions allows for different levels of error protection between the base and the enhancement layer for SVC encoded streams. Next, both the data and the repair packets in each session are multiprotocol encapsulated into MPE sections. Time slicing is then applied to multiplex the MPE sections belonging to different services onto an MPEG-2 transport stream (TS). In case hierarchical modulation is not chosen in the physical layer configuration, only a single MPEG-2 TS exists and correspondingly all sessions that belong to a certain service must be mapped to the same time-slicing burst. However, if hierarchical modulation is chosen, the sessions belonging to a certain service may be mapped onto either one of the MPEG-2 TS depending on the required protection level.

5.3 Model of the Receiver-side Processing

The actual transmission and reception process is then covered by calling the PHY emulator with the current MPEG-2 TS parameters and system timing. As explained in section 6 below, the latter returns a realistic error vector for the chosen configuration, as well as info on the timing advances in both transmitter and receiver due to the PHY processing. The error vector can then be used inside the IP packet sim-

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

ulator to render those MPEG-2 TS packets as lost which were marked as erroneous at the physical layer.

The remaining receiver side processes mainly complement the aforementioned steps done at the transmitter: As shown in Figure 8, the received MPEG2-TS stream is de-multiplexed into different sessions. The received media stream for each service is then reconstructed from a number of sessions after optional IFEC decoding and packet reordering. Finally, the received streams are forwarded to the corresponding Streaming Sinks, which record them for later processing in the external Video Decoder and quality evaluation module.

Finally, we want to note that simulation of hybrid satellite and terrestrial use cases simply requires two separate PHY emulators that are run in parallel. For each service, the overall set of successfully received packets on both links is then determined inside the IP packet simulator.

6. DVB-SH PHYSICAL LAYER EMULATION

6.1 Emulator Overview

The DVB-SH physical layer emulation consists of a tool which is able to produce error flags (EFlags) at MPEG-2 TS level. In detail, this tool predicts the behavior of the physical layer in a given channel environment and for given transmission parameters, i.e., it indicates whether an MPEG-2 TS packet is erroneous or not. The calculations made are based on input channel time series, which represent a series of signal-to-noise–ratio values over time.

6.2 Concept

The concept of the EFlags generation is based on information theory and is performed by comparing the channel capacity at a given time (using the signal-to-noise–ratio) and the system capacity being determined by the forward error correction (FEC) scheme. The channel capacity is defined using the well known Shannon formula [4] :

$$C = B * \log_2 \left(1 + \frac{S}{N}\right), \tag{1}$$

where C is the capacity in bit/s/Hz, S/N the signal-to-noise–ratio, and B the channel bandwidth. Hence, in a time varying channel where S/N values change over time, the channel capacity is a function of time.

If R is the system capacity defined by the data rate of the receiver (measured at the output of the receiver) which is constant over time, by comparing R and C over time, errors can be detected using the following information theorem [4]: if $R < C$ there exists a forward error correction (FEC) scheme able to ensure a reliable (error-free) transmission. In other words, if $R < C$ we expect no error in the information transmission, while for $R > C$ (and $R = C$) we can assume that there will be a transmission error: an EFlag will be set. However, in a real transmission system, several limitations have to be taken into account:

- The FEC scheme defined by Shannon is an ideal FEC scheme. Real FEC schemes can only approach this limit. However, turbo codes as used in DVB-SH are a kind of error correction code which is able to get very close to Shannon's limit.
- The channel capacity is constrained by the modulation of the receiver, meaning that C can not be larger than a given value. Considering QPSK modulation, for example, each symbol carries 2 information bits, hence, the capacity can not be higher than 2 bit/s/Hz. This is called modulation constrained capacity.
- The formula is only valid in time-invariant additive white Gaussian noise (AWGN) channels. Assumptions have to be made for other kinds of channels.

Considering flat fading channels (meaning only the amplitude and phase of the channel is a function of time but without frequency selective components), such as the land mobile satellite (LMS) channel or the Rayleigh channel, the AWGN assumption is valid for a given time interval if the channel is invariant within this interval. This can be ensured if the sampling period of the channel is smaller than the coherence time of the channel. This assumption cannot be made in frequency selective channels, since in this case the energy is not uniformly distributed over frequency anymore. In addition, using real data gathered from field-tests, the capacity function has been refined to represent more accurately the behavior of a real system. This consists especially in including implementation losses for the FEC scheme and constraining the capacity for the modulation foreseen by the DVB-SH standard.

6.3 Principle of Operation

The FEC scheme of the DVB-SH standard is a block based FEC. Each FEC word is a block of 8 MPEG-2 TS packets coded at a given rate (the rate defines the ratio between information and code bits). We compute the capacity of an encoded FEC word and emulate the detection of errors by comparing the capacity of this FEC word to the system capacity. Before FEC decoding, the DVB-SH standard defines a convolutional interleaver used to spread data over time in order to overcome signal blockages. This interleaving process also has to be taken into account. The sequence of operations is as follows:

1. From the S/N series, the channel capacity over time is computed.
2. The time resolution of the interleaving process is defined by the duration of so called Interleaver Units (IU) which are the input of the interleaver. It has to be ensured that the coherence time of the channel is larger than the duration of an IU so that the AWGN assumption previously made is valid. A single capacity value is computed for each interleaver unit and this series is fed into the time deinterleaver to get a deinterleaved capacity series. The latter is divided into chunks corresponding to the number of IUs per FEC word.
3. The capacity of the FEC word is estimated by computing the mean capacity of IUs of the FEC word. Since turbo codes are using an intra word block interleaver, it can be assumed that the capacity of the whole code word is defined by this mean capacity.
4. The resulting capacity C is compared with the system capacity R (equal to the product of the FEC code rate and the modulation order).
5. An error flag is set if R is larger than C. The error flag is set on all the 8 MPEG-2 TS packets since the decision has been taken for the whole FEC word.

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

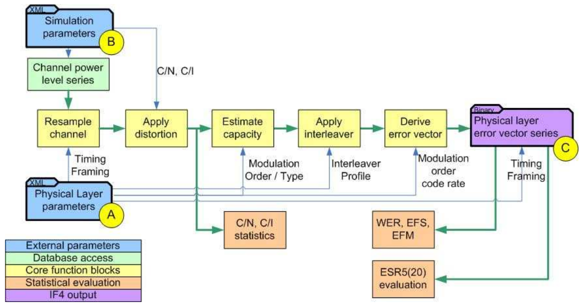

Figure 9: Physical layer emulator.

6.4 Implementation

In order to emulate a DVB-SH receiver and a transmission channel, several parameters have to be taken into account:

Transmission mode of DVB-SH can be set to TDM (used for satellite broadcast) or OFDM (can be used for both satellite and terrestrial broadcast). As each transmission mode has different implementation losses, they are handled separately. The modulation order is also needed in order to limit the channel capacity using modulation constrained capacity.

FEC parameters of DVB-SH consist of the rate of the turbo code used and the characteristics of the interleaver. This is shown as point (A) in Figure 9.

Channel parameters define the time series to use, the signal-to-noise–ratio, and the series resolution. It is also possible to set a signal-to-interference–ratio. The difference between noise and interference is that the interference power is proportional to the signal power. These parameters are marked as (B) in Figure 9.

The point marked as (C) represents the output of the the tool, i.e., the EFlags are handed over to the IP simulator. From these Eflags, common statistics, like frame error rate (FER) or error-second-free–ratio (ESR5), can be derived.

7. PROOF OF CONCEPT: SIMULATION CAMPAIGN FOR SELECTED USE CASE

In the following, we will demonstrate how our proposed simulation platform can be used to investigate the viability of integrating Scalable Video Coding (SVC) technologies over DVB-SH networks. From among the 13 reference use cases that have been identified during the first phase of the ESA study mentioned in section 1, we have selected the one dealing with “satellite-only transmission to heterogeneous terminals” to be used as proof of concept in this paper.

7.1 Reference Use Case Description

The objective behind the chosen reference use case can be described as follows: Consider an area where reception of the DVB-SH signal is only possible via the satellite, i.e., where there are no nearby CGCs. The mobile terminals within this area are assumed to belong to either one of two standard classes defined for DVB-SH. The latter will be simply referred to as “good” terminal (GT) class and “bad” terminal (BT) class here and usually differ in the following aspects:

- Reception capability: Assume that the GT class has a higher antenna gain-to-system noise temperature ratio (the so-called G/T figure) than the BT class. Hence, given a certain effective isotropic radiated power (EIRP) of the satellite transmitter, as well as fixed PHY code rate and modulation scheme, the coverage region of the GT class is much larger than the one of the BT class. This can be expressed via the carrier-to-noise–ratio (C/N) required for a certain target service quality, e.g., a certain error-second-free–ratio (ESR5)[2].
- Screen size: Assume that the GT class has a larger screen size than the BT class. This means that different spatial resolutions of the video stream are required for optimal presentation on both terminals. However, in a standard broadcast scenario, only a single resolution is produced by the video encoder and transmitted. Hence, either the GT class has to upscale from a lower resolution (leading to a loss in presentation quality[3] compared to the target resolution), or the BT has to downscale from a higher resolution (forcing these terminals to unnecessarily receive a too large data stream at almost no visual effect). Upscaling or downscaling could only be avoided by transmitting two versions of the same service in parallel (called Simulcast). However, for the same multiplex bit rates, this leads to a strong reduction in the number of services that can be offered with different content.

7.1.1 *Reference system design without scalability*

Without scalability in the video stream, the system operator has the following trade-offs to be made when determining the reference operation point of the system:

- Coverage vs. number of services: When choosing the PHY code rate and the modulation scheme, the operator must increase the protection up to a point where class BT achieves a decent overall service quality within the desired coverage region. Hence, the operation point of the system depends on the BT class, although the operator does not know how many of these terminals are actually present. In the worst case (only GT class terminals), a large portion of the bandwidth is wasted for unnecessarily high protection, which could otherwise be used to offer additional services.
- Presentation quality vs. number of services: When choosing the spatial resolution, the operator must ensure that the desired presentation quality is achieved at both terminal classes. Hence, either the higher one must be chosen, or the expected loss due to upscaling must be pre-compensated during the encoding. Note that choosing simulcast is only of interest, if for some reason the BT class terminals are not able to handle the larger resolution in the decoder. In all cases, this results in a significant increase in bit rate for each service, which directly affects the number of services that can be offered. In the worst case (only BT class terminals), this again results in a waste of bandwidth.

[2]In this work, we will use the ESR5(20) criterion, which requires that within a window of 20 seconds, at most 1 second is in error. Or in other words: Packet loss must be confined to a 1 second interval to achieve tolerable service quality.

[3]An objective measure for the video quality is the peak signal-to-noise–ratio (PSNR) of the luminance component.

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

Table 1: Terminal classes chosen for the analysis.

Class	Type	G/T figure	Screen size
BT	Handheld	-32 dB/K	2"
GT	Vehicular	-21 dB/K	7"

7.1.2 *Alternative system design with scalability*

With scalability in the video stream, the system operator may adapt the design as follows:

- To support both spatial resolutions without scaling or Simulcast, the encoded stream must contain two spatial layers: a base layer (BL) with the lower resolution for the BT class, and an enhancement layer (EL) with the higher resolution expansion for the GT class.
- The SVC encoding should be such that (without loss)
  - the BL almost achieves the same quality as with non-scalable encoding at lower resolution,
  - the BL+EL almost achieves the same quality as with non-scalable encoding at higher resolution.
- To support unequal protection (UEP) for the two embedded spatial resolutions, the hierarchical modulation (HMOD) option available in DVB-SH can be used. The design of the protection levels is as follows:
  - The BL is required by both classes and is thus mapped on the "High Priority" (HP) stream. The respective protection level must be tailored to the coverage and service requirements of the BT class.
  - The EL is only required by the GT class and is thus mapped on the "Low Priority" (LP) stream. The respective protection level is tailored to the coverage and service requirements of the GT class.

7.2 Simulation Scenario and Parameters

Due to the limited space available, we cannot treat all aspects of the reference use case described above in this work, but will rather focus on a representative subset of the results. Hence, among the many different combinations of two terminal classes, we have chosen a handheld (e.g., a mobile phone) and a vehicular terminal to represent the BT and GT class, respectively. The corresponding capabilities are listed in Table 1.

The two spatial resolutions that best match the different screen sizes and allow spatial scalability in the video encoding process are QVGA and VGA for the BT and GT class, respectively. Furthermore, for the chosen BT class, only QVGA decoding at rates below 384 kbps is possible.

The rate control applied in the StatMux operation will be uncoordinated VBR encoding, i.e., all services receive the same rate share in each time-slicing interval, whose duration is tightly controlled around 1 s. The DVB-SH system is of type SH-B for the reference system, with satellite-only transmission over a Ricean channel. Table 2 contains all relevant parameters for the reference system configuration, i.e., transmission of non-scalable video encoded with AVC.

If scalability is used for the encoding, some of the parameters have to be adapted[4], as shown in Table 3.

[4]Note that TDM is more power-efficient than OFDM. Hence, in a satellite-only scenario, the reference system may use TDM, while we have to resort to OFDM if HMOD is required to support scalability.

Table 2: Reference system configuration.

Parameter	Setting
System	SH-B
Bandwidth	5 MHz
Modulation type	TDM
Modulation order	QPSK
Roll-off factor	0.15%
PHY code rate	2/5
Time interleaver	Short
Channel	Rice: $K = 5$ (BT) or 10 (GT)
Speed	3 km/h
C/I	12 dB
Multiplex bit rate	3 Mbps
Virtual audio bit rate	32 kbps (per service)
Virtual control bit rate	100 kbit/s (per multiplex)
Time slice period	1 s
Section packing	Yes
MPE-IFEC	Off
Video coding	AVC
Spatial video resolution	QVGA (upscaled at GT) Simulcast: QVGA + VGA
IDR frame distance	1 s
StatMux rate control	Uncoordinated VBR encoding

Table 3: Change in configuration with scalability.

Parameter	Setting
System	SH-A
Modulation type	OFDM
Modulation order	16QAM
OFDM guard interval	1/8
Hierarchical modulation	alpha=1
PHY code rate	1/5 (HP), 1/5 (LP)
Video coding	SVC spatial scalability
Spatial video resolution	BL: QVGA; EL: VGA
Scalability ratio	Nominal rate split: BL:EL=1:1

7.3 Performance Results and Analysis

The simulation methodology applied for this use case can be split up into two main stages: During the first one, the reference system with non-scalable encoding is investigated. From the description in section 7.1.1, the following reference encoding strategies are possible:

- QVGA with no compensation for the upscaling (Ref_1): Encoding with AVC is done for QVGA resolution and a target minimum PNSR[5] of 32 dB. The latter value results in decent achievable quality, if none of the transmitted packets is lost and no upscaling is performed.
- Simulcast (Ref_4): Encoding with AVC is done for QVGA resolution as in Ref_1, and then repeated for VGA resolution and a target minimum PSNR of 31 dB. Both resolutions are then transmitted in parallel as two separate services with same content. Each terminal class then selects the appropriate spatial version for decoding and display.

The first two rows of Table 4 contain the achievable presentation quality in terms of the actual minimum PSNR at

[5]Minimum PSNR here means that the worst 5 % of the decoded chunks fall below this threshold.

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666

Table 4: Presentation quality and offered number of services for different encoding strategies.

		Ref_1	Ref_4	Sca_1
Achievable presentation quality	BT	33.0 dB	32.6 dB	31.9 dB
	GT	29.6 dB	32.1 dB	31.8 dB
Effective presentation quality	BT	32.4 dB	32.3 dB	31.7 dB
	GT	29.6 dB	31.9 dB	31.5 dB
Offered services		8	3	4

Table 5: Received signal strength and effective service quality at each terminal class.

		ESR5(20)		
	C/N	Single stream	HP stream	LP stream
BT	6.7 dB	94%	97%	0%
GT	17.7 dB	100%	100%	100%

each terminal class for the above encoding strategies. As can be observed, Ref_1 results in decent achievable quality for the BT class, but causes a significant quality drop for the GT class due to the upscaling. Hence, it may not be acceptable in a heterogeneous scenario. However, supporting both terminal classes with non-scalable encoding leads to a significant reduction in the number of offered services (with different content): From the last row of Table 4, we can see that Simulcast (Ref_4) is not a good choice, since the number of offered services drops from 8 to 3, which is not attractive anymore from a commercial point of view.

Next, the desired coverage region for the reference system must be determined. The criterion will be 90% ESR5(20) fulfillment for the BT class, which results in the required C/N at the edge of the coverage region listed in the second column of Table 5. Assuming that the GT class is located in the same spot, we can translate the reception quality via the larger G/T figure into a higher equivalent C/N for this class. As a direct consequence, the GT class would achieve a significant increase in effective service quality, i.e., better ESR5(20) fulfillment, as shown in the third column of Table 5. The corresponding results for the effective presentation quality after decoding at the terminal are contained in the third and fourth row of Table 4.

In the second stage of the analysis, the performance of the system with scalable encoding is now investigated. For this reason, a further encoding strategy (Sca_1) is introduced, where encoding is done with SVC and a rate split of 1:1 between base and enhancement layer. In order to allow fair comparison to AVC, the encoding is done such that the target minimum PSNR of the BL almost achieves 32 dB at QVGA resolution, while the target minimum PSNR of the BL+EL almost achieves 31 dB at VGA resolution. The rightmost column in Table 4 shows the respective achievable presentation quality at each terminal class.

In order to partially compensate for the bandwidth expansion of the scalable bit stream and allow UEP, the DVB-SH system now operates with hierarchical modulation. From the last row of Table 4 we can observe that the offered number of services is one larger than for the non-scalable strategy Ref_4. Assuming that the terminals are still located in the same spot (i.e., the reception conditions are maintained), the last two columns of Table 5 contain the effective service quality for the HP and LP stream. Obviously, the BL is available at both terminal classes, while only the GT class is able to receive the EL with decent quality.

Finally, we can compare the results for the effective presentation quality with the ones for the reference system. This proves that SVC allows to maintain a decent video representation at both classes, while giving the system operator the ability to offer a larger number of services.

8. CONCLUSIONS

In this paper, we have presented a simulation platform for multimedia broadcast over DVB-SH, which allows to investigate various practical use cases that have to be investigated before the deployment of future DVB-SH systems. The modular design with simple interfaces between the main components allows to further enhance the platform with potential extensions of the standard and/or the media services that are offered by the operators.

As proof-of-concept, we have presented selected results for one exemplary use case in the ESA study on scalable encoding for DVB-SH, which have been gained with the help of our proposed platform. Further analysis of this and all other use cases (like heterogeneous reception conditions, hybrid satellite/terrestrial operation, etc.) is currently in progress.

9. ACKNOWLEDGMENTS

Besides the authors of this paper, the following people have contributed to the successful completion of this work: Thomas Stockhammer, Thiago Martins de Moraes, Maria Alejandra Pimentel Nino (Nomor Research GmbH), Ernst Eberlein (Fraunhofer IIS), Robert Skupin (Fraunhofer HHI).

In addition, we also want to thank ESA for supporting this work.

10. ADDITIONAL AUTHORS

Additional authors: Nghia Pham (Eutelsat S.A., 70, rue Balard, 75015 Paris, France, email: `HPham@eutelsat.fr`).

11. REFERENCES

[1] Digital Video Broadcasting. DVB-SH implementation guidelines, May 2002. DVB BlueBook A120.

[2] Digital Video Broadcasting (DVB). System Specifications for Satellite services to Handheld devices (SH) below 3 GHz, July 2007. ETSI TS 102585 v1.1.1.

[3] ISO/IEC. Information technology – Coding of audio-visual objects – Part 15: Advanced Video Coding (AVC) file format AMENDMENT 2: File format support for Scalable Video Coding (SVC), 2008. 14496-15:2004/Amd2.

[4] T. S. Rappaport. *Wireless Communications, principle and practice*. Prentice Hall Communications Engineering and Emerging Technologies Series, Upper Saddle River, NJ, 2002.

[5] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the scalable video coding extension of the H.264/AVC standard. *IEEE Trans. on Circuits and Systems for Video Technology*, 17(9):1103–1120, Sept. 2007.

[6] S. Wenger, Y.-K. Wang, T. Schierl, and A. Eleftheriadis. RTP Payload Format for SVC Video, Sept. 2009. http://tools.ietf.org/html/draft-ietf-avt-rtp-svc-19.

Digital Object Identifier: 10.4108/ICST.SIMUTOOLS2010.8666
http://dx.doi.org/10.4108/ICST.SIMUTOOLS2010.8666