International Journal of Digital Multimedia Broadcasting

Convergence of Digital TV Systems and Services

Guest Editors: Maurizio Murroni, Sandro Scalise, Alessandro Vanelli-Coralli, Sooyoung Kim, and Robert Briskman

Convergence of Digital TV Systems and Services

International Journal of Digital Multimedia Broadcasting

Convergence of Digital TV Systems and Services

Guest Editors: Maurizio Murroni, Sandro Scalise, Alessandro Vanelli-Coralli, Sooyoung Kim, and Robert Briskman

Contents

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 817636, 3 pages
doi:10.1155/2009/817636

Editorial

Convergence of Digital TV Systems and Services

Maurizio Murroni,[1] Sandro Scalise,[2] Alessandro Vanelli-Coralli,[3] Sooyoung Kim,[4] and Robert Briskman[5]

[1] *DIEE, University of Cagliari, 09123 Cagliari, Italy*
[2] *German Aerospace Center (DLR), Institute of Communications and Navigation, 82234 Wessling, Germany*
[3] *DEIS/ARCES, University of Bologna, Viale Risorgimento 2, 40125 Bologna, Italy*
[4] *College of Engineering, Chonbuk National University, Jeonju 561-756, South Korea*
[5] *Sirius XM Radio Inc., USA*

Correspondence should be addressed to Maurizio Murroni, murroni@diee.unica.it

Received 24 December 2009; Accepted 24 December 2009

The migration from analog to digital of the broadcasting technologies, already well consolidated for satellite systems, is becoming a reality also for terrestrial transmission. Digital Terrestrial Television (DTT) is also evolving to offer interactive services and a degree of flexibility which can be exploited to offer tailored applications to users which include, for instance interactivity, different levels of personalization, and innovative location-based, as well as context-aware, services. A clear example of this trend is given by the rising success of the Internet Protocol Television (IPTV) which allows for a degree of flexibility on offered services unknown to the traditional broadcasting systems. In this framework, several satellite operators are starting to launch IPTV services using direct satellite links, as well as some terrestrial internet service providers are offering digital TV channels embedded in the IP streaming over XDSL. Furthermore, IPTV services are likely to be broadcast also wirelessly, exploiting advanced broadband access technologies such as WiMAX, LTE, or LTE-A. Last, but not least, TV and broadcast services for mobile users have also been deployed in many countries using DVB-H and will be soon available on an even broader scale thanks to its satellite counterpart, DVB-SH. In the near future, a set of different technologies able to offer personalized and customized services to different classes of users are expected in the area of wireless broadcasting and convergence of technologies is auspicious. This concept entails different levels of convergence, namely, at terminal level (one device fits all), at service level (convergence of traditional fixed, mobile, and broadcast services), and at transport and network level with a common and standardized set of protocols and at access layer thanks to the harmonic coexistence of different radio technologies. This special issue aims to capture the state-of-the-art research work concerning the integration of DTT/Satellite/IPTV systems for the broadcasting of multimedia and interactive services.

There are very different directions in which researchers investigate new solutions in order to improve the quality of the rich media content delivered over various network types. In this context, this special issue illustrates some of these avenues and presents some of the most significant and promising results.

Papers have been grouped into four thematic sections.

The first section is devoted to *Mobile Television and Interactive Multimedia Broadcasting Services*. The first paper entitled "Converged digital TV services: the role of middleware and future directions of interactive television" provides a brief overview of digital TV converged services, to present and categorize the digital Television middleware technologies that contributed to it as well as possible future trends and directions. A new Television era of converged wireless and mobile content delivery, user-authored content, multimodal interaction, intelligent personalization, smart space awareness and 3D content sensations is foreseen, creating ambient and immersive experiences. The following one entitled "Interactive digital terrestrial television, the interoperability challenge in Brazil" introduces different standards implemented in existing Digital Terrestrial Television Broadcasting systems to allow the fruition of interactive services and applications through digital Set Top Boxes. It focuses on the interoperability issue between the Brazilian and the European architectures. Efforts in this area are reported from a Brazilian perspective including work to harmonize Brazilian and European standards. A wide variety of architectures, methods, and systems are reviewed to provide such applications, particularly for learning uses. In the third paper "Sofa-TV: the new digital landscape" the authors address the current state of channel offerings for the so-called Sofa-TV, that is, all digital television offers viewed through the "traditional" television screen. In this context, three digital platforms, Sat TV, DTT, and IPTV, are considered. After a description of the commercial offer, the authors identify the principal players and their strategies, in order to provide a few predictions as to the possible future changes in the industry. The analysis, carried out in the Italian context, has a general applicability and the conclusions apply to the general market.

The second section deals with *Hybrid Cellular/Non-Cellular Broadcasting* and contains three papers. The first one entitled "MBMS—IP multicast/broadcast in 3G Networks" extensively describes MBMS for 3GPP network, including the main features, architectural overview, and capabilities of MBMS added on various protocol stacks. As a use case, Mobile TV over MBMS in UMTS is addressed by using simulation results on capacity improvement by MBMS. MBMS evolution issues are also addressed. The second entitled "Video streaming transfer in a smart satellite mobile environment" concentrates on video streaming provided on buses that move in urban, suburban, or highway environments. A contents' source utilizes a satellite DVB-S2 link for transmitting video streams to a bus, which, in its turn, relays it to its passengers' devices. A bus works in a smart mode taking advantage of the knowledge of the exact points where it will not receive the satellite signal for a certain time period, due to fixed obstacles. Finally, the third one entitled "Modeling and performance analyses of hybrid cellular and broadcasting networks" considers the combination of the cellular network UMTS and the mobile broadcast network DVB-H, which form a hybrid network. The authors investigate the performance of hybrid networks and develop a system model, which describes the hybrid network and the load switching between both networks.

The third section addresses *Source and Channel Coding Modulation and Signal Processing for Mobile Multimedia Broadcasting* and it comprises four papers. The first one "Multiple description coding using data hiding and regions of interest for broadcasting applications" presents an innovative scheme for multiple description coding (MDC) with regions of interest (ROI) support to be adopted in high quality television. The scheme proposes to split the stream into two separate descriptors and to preserve the quality of the region of interest, even in case one descriptor is completely lost. To demonstrate its effectiveness, the algorithm has been implemented in two different scenarios, using the reference H.264/AVC codec and an MPEG framework to evaluate the performance in absence of motion-compensated frames on 720p video sequences. The next paper, entitled "Slow motion and zoom in HD digital videos using fractals" presents an alternative technique combining fractals theory and wavelet decomposition to achieve spatial zoom and slow motion replay of HD digital color video sequences. Fast scene change detection, active scene detection, wavelet subband analysis and, color fractal coding based on Earth Mover's Distance (EMD) measure are used to reduce computational load and to improve visual quality. In "AL-FEC for Improved mobile reception of MPEG-2 DVB-T transport streams," the authors investigate the use of application layer FEC protection in DVB-T (Digital Video Broadcasting-Terrestrial) networks for the provision of mobile services. The paper discusses some implementation aspects of AL-FEC in real scenarios and proposes an implementation based on Raptor codes and hash sequences. In the last work, entitled "The implementation of a 2/4/8 antennas configurable diversity OFDM receiver for mobile HDTV applications" a mobile OFDM receiver LSI is developed which includes two pre-FFT adaptive array (AA) antenna combiners and a post-FFT carrier diversity (CD) combiner. Using the LSI, several combinations of AA and CD diversity system can be configured as a trade-off approach to improve the performance of mobile ISDB-T receiver with reasonable system cost. The paper deals with the use of antenna diversity to improve the service availability for HDTV. Both simulation and measurement results are presented.

The last section covers *System and Service for IPTV and Interactive TV*. In the first paper "Theoretical models for video on demand services on peer-to-peer networks," the authors study the scenario P2P-based VoD services and propose four models of the peer behavior to evaluate the impact of the peer churn on the system performance. The models are compared by computing the resource that system has to add on top of the P2P network to satisfy all the download requests. Simulations show important relationships between playback buffer length, peer request rate, peer average lifetime, and server upload rate. The second paper, entitled "A software and hardware IPTV architecture for scalable DVB distribution," deals with the problem of mapping popular public broadcast services such as DVB or ATSC to IPTV with acceptable effort to IP-based television (IPTV). The authors propose a mapping using a light weight framework as an important step towards all-IP multimedia. Then, they present the Net-Ceiver architecture based on well-known standards such as IPv6, which allows zero configuration and high scalability. They also describe a low-cost FPGA implementation of the proposed Net-Ceiver architecture, which can concurrently stream services from up to six full transponders. The last paper is entitled "Concept for mobility and interconnections aspects in converged NGN based IPTV architecture" and discusses several potential architectures implementing the converged NGN IPTV concept. In this direction, ETSI TISPAN and ITU-T architecture for IPTV (IMS and non-IMS) are considered. In particular, an evolution of IMS-based IPTV where a single platform can serve multiple access networks is described along with several application scenarios. Last but not least, the paper "P2P and MPEG FGS encoding: a good recipe for multipoint video transmission in the internet" proposes a multipoint video broadcast framework over a heterogeneous content distribution P2P network. In the proposed system the source generated the video flow by using an MPEG-4/FGS encoder, in such a way that no losses occur at the Base-layer stream even in the presence of short-term bandwidth fluctuations. Several solutions to this have been proposed but adoption has been hindered by many difficulties, particularly efficiency. Simulations of performance are presented which show that good efficiency would be achieved.

Maurizio Murroni
Sandro Scalise
Alessandro Vanelli-Coralli
Sooyoung Kim
Robert Briskman

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 643680, 19 pages
doi:10.1155/2009/643680

Review Article

Converged Digital TV Services: The Role of Middleware and Future Directions of Interactive Television

Emmanuel Tsekleves, John Cosmas, Amar Aggoun, and Jonathan Loo

School of Engineering and Design, Brunel University, Uxbridge, Middlesex UB8 3PH, UK

Correspondence should be addressed to Emmanuel Tsekleves, emmanuel.tsekleves@brunel.ac.uk

Received 30 January 2009; Accepted 25 June 2009

Recommended by Maurizio Murroni

The subject of the future of the interactive Television medium has become a topic of great interest to the academic and industrial communities particularly since in the recent years there has been a dramatic increase in the pace of innovation of convergence of digital TV systems and services. The purpose of this paper is to provide a brief overview of what we know as digital TV converged services, to present and categorise the digital Television middleware technologies that contributed to it, and to present possible future trends and directions. A new Television era of converged wireless and mobile content delivery, user-authored content, multimodal interaction, intelligent personalisation, smart space awareness, and 3D content sensations is foreseen, creating ambient and immersive experiences.

1. Introduction

Interactive Television (iTV) has been the subject of dramatic innovation in the recent years transforming a traditionally passive medium into a truly interactive experience.

Firstly there has been a trend towards more interaction giving the viewer control over video, audio, graphical, and text elements, by enabling him to consume simple games and quizzes and send simple communications back to the broadcaster. Secondly there has been a trend towards providing the viewer with a more enhanced Television (TV) experience by adding converged services such as Internet pages, video clips, 3D graphics, email, Internet blogs, and many other traditionally computer oriented features to the TV experience.

Thirdly there has been a trend towards providing TV programmes over a much wider range of TV screen sizes ranging from High Definition (HD) to Standard Definition (SD) to Mobile Definition (MD) TV with a trend for certain programmes to be targeted to particular screen sizes. The recent 2006 World Cup Football Championship has been targeted towards an HDTV experience whilst short fast-paced film entertainment has been targeted towards the mobile TV experience. Fourthly there has been a trend towards providing TV programmes on a number of different types of Televisions within different viewing contexts outside the sitting room and at home such as within cars, buses, trains and plains, outside in hotel lobbies and within portable consumer products such as within mobile phone, cameras or within pocket PC environments.

Finally the advent of new programming genres (e.g., soaps, current affairs, reality TV, radio, documentary, shopping, and arts programmes) combined with the multiplicity of channel choice brings personalised TV into a new era providing viewers with selective access to certain types of programmes and advertisements. Personalised TV includes the personal video recording function, which allows the viewer to pause live programming, fast-forward through commercials and record hours of programming without the use of videotape. This in turn is enabling viewers to skip over any 30-second advertising slot. Furthermore the plethora of choice and channels makes it increasingly difficult for advertisers to know where to place their adverts in order to achieve the biggest impact.

This paper categorises and describes the various software solutions that render the interactive part of digital Television possible in different converged platforms. Finally it presents and discusses future and emerging trends from technological, business and viewer perspectives that will shape the next pace of innovation in the iTV domain.

2. Digital TV Convergence Overview

This section offers a brief overview of the main milestones in the convergence of digital TV services.

2.1. First Converged Digital TV Service. The first digital interactive multimedia service integrating the transmission of voice, data, image, and video together, began in December 1994 in Orlando (Florida) with the Full Service Network (FSN) user trial. There were several companies behind the experiment including Time Warner Cable, Scientific-Atlanta, AT&T, SGI, and Toshiba. The trial consisted of a full-scale user experiment with 4000 subscribers by the end of the year 1995. The trial offered services such as Video On Demand (VOD), home shopping, interactive programme guide, US postal services, and games [1]. The FSN trial service ended in 1997 due to rising costs and lack of content and although the general perception outside the interactive Television industry was one of many failures, the companies behind the trial regarded it as pioneering.

Starting in 1990 and ending in 1993, AT&T conducted a use of interactive Television trial in 30 homes of its employees in Chicago. The services provided included home shopping, video games, education, news, and sport. AT&T concluded that there is no single irresistible consumer service [2].

Although these results were not representative, because of the size and nature of the user group involved (i.e., AT&T employees) the results were still very informative. Reactions to it were positive and especially indicated that interactive educational programmes caused greater and stronger interest for children. Also popular were programmes featuring sports, and games where households competed against each other, producing a strong family interest [3].

In 1996 a series of Digital Video Broadcasting (DVB) satellite services were introduced with the commercial availability of satellite receivers. DVB satellite is a suite of internationally accepted open standards for digital Television distribution and interactive content via satellite. These DVB satellite services had very different economic successes and failures. Canal+ in France became one of the largest satellite service and content providers in Europe, whereas the D-Box in Germany never became commercially successful. The only explanation as to why the same media technology and service was a success on one side of a border and a failure on the other are differences in programme content, cost, and the context of existing, and competing services. DVB platforms expanded in the late 1990s with the introduction of DVB digital cable and digital terrestrial in various parts in the U.S, Europe, and Asia [4].

2.2. Digital Television and Internet Convergence. With the Internet explosion during the 1990s broadcasters feared that the rapid growth of the Internet would draw audiences away from traditional broadcasting and lead to its eventual demise. This fear later subsided when many believed that the emergence of packaging and "channel" concepts on the Web (due to the ever increasing use of IP multicast) meant that broadcasters could use the Internet as a valuable extra resource through which to reach new consumers. The emergence of DVB and other digital TV technologies provided an obvious platform through which an Internet business could be fully integrated into a broadcasting strategy. The Internet supports a huge source of interinformation and content including tens of millions of Web sites. However the delivery of such Internet pages is not restricted to the Internet. The MultiProtocol Encapsulation (MPE) specification of DVB provides a mechanism for transporting data packets over MPEG-2 Transport Streams. MPE was optimised for the use of IP data. The data packets are encapsulated in datagram sections that are compliant with the Digital Storage Media Command and Control (DSM-CC) section format for private data. This allows both digital TV and Internet traffic to co-exist on the same system and be received by DVB set-topboxes [5].

During the late 1990s and early 2000 several experimental systems were developed to deliver Internet pages via the broadcast transmission. The selection of these Internet pages is generally made by an editor (service provider) or by the user and are displayed on either a computer or a Television set [6].

The new world of digital TV creates new opportunities since a digitised TV technology allows other digital technologies such as the Internet to be combined with it. In this respect, Television can be seen as the best way to bring the Internet to a mass market. As Martin Sims points out in Papathanassopoulos [7]:

> "It's the Internet on your Television with built-in modems to access websites, not...Television on the Internet".

People have embraced the Internet because of its interactive nature. With the introduction of digital TV and bearing in mind that digital TV is often seen as the main technology leading to interactive TV [8], a whole new world full of amazing possibilities is opening its doors to individual Internet and TV users. Digital Television viewers can now use their TV sets to gain access to activities more familiar to the Internet, such as browsing information on topics of interests, keeping up-to-date with their email communications, carrying out financial transactions (e-commerce) and several other applications and services existing in the domain of the World Wide Web. Therefore, the concept of convergence between Television and the Internet has been the dominant concept of interactive Television during the late 1990s and early 2000.

2.3. Interactive Television. The convergence of digital TV has made it possible to incorporate feedback into the traditionally one-way form of TV communication by combining video, audio, and data within the same signal, epitomizing the TV world. In a nutshell, iTV brings a range of new multimedia services that enables users to browse information on topics of interest, play interactive games, conduct e-commerce related activities such as shopping, banking, and personalize their viewing choices.

More precisely, interactive TV stands for the broadcasting of a digital transport stream of traditional audio-visual contents mixed with binary data, so making possible to deliver multimedia software applications to be executed in a digital TV (DTV) or a set-top box [9]. It must be noted that to fit the data signals in the transmitting channel, the actual transmission in DTV uses analogue signals by modulating it into an analogue waveform [5].

A set-topbox provides an interface between the TV set and the media received from Cable, Satellite, and Terrestrial delivery methods. On the hardware side, a typical set-top box incorporates a tuner, a demodulator for receiving and demodulating the TV signal, a de-multiplexer for demultiplexing the TV stream back into the TV programme and additional media, a descrambler to descramble the scrambled channels as well as a decoder for decoding the audio-visual content. Set-topboxes can also include a modem for access to interactive services connecting to Public Switch Telephone Network (PSTN), Integrated Services Digital Network (ISDN), or Asynchronous Digital Subscriber Line (ADSL) networks. Furthermore a set-top box also incorporates a microprocessor and memory for running digital local and interactive applications [10]. Second generation set-top boxes include also a hard disk for storing content (PVR functionality).

2.4. IP Television. Since the early 2000, there is a change in the convergence of TV and the Internet, with the TV becoming now part of the Internet thus forming Net Television, most commonly known as Internet Protocol Television (IPTV). IPTV differs from the traditional digital Television systems. Since its conception it has aimed to provide alternative view channels over the Internet instead of enhancing current view channels (enhanced Television) or providing users with additional control over the current viewing channels (personal Television) [11]. IP Television's main aim is to increase Internet access and services such as browsing and chat and therefore focuses more on the telecom networks rather than broadcast networks. The IPTV business model includes triple-play, Pay-TV, paid video-on-demand (VOD), advertisement-based TV, and so forth. The availability of content anywhere anytime is the main attraction of new customers and idealising existing ones.

The IPTV infrastructure can be deployed either with centralized or distributed video server architectures. The centralized IPTV is simply the content delivery network used in today's VOD service. However, the architecture is only good for relatively small network and requires adequate core and edge bandwidth. The distributed IPTV is more ideal for large network deployment by using P2P method. It is a scalable architecture that has advantage in bandwidth usages, but it requires content distribution system for effective delivery over scattered network nodes. IPTV employs IP multicasting for the delivery of digital TV services. IP Multicast is a method in which information can be sent to multiple computers at the same time. The playback of IPTV requires either a personal computer or a set-top box connected to a TV. Video content is typically compressed using either a MPEG-2 or a MPEG-4 codec.

One of the main issues of IPTV is the existence of several proprietary different IPTV standards (such as Microsoft TV [12] and Veoh TV [13]), creating a sense of confusion and interoperability issues in the market. The DVB consortium has attempted to promote standardization for IPTV and achieve interoperability by introducing the Digital Video Broadcasting-Internet Protocol (DVB-IP). specification [14]. DVB-IP includes two distinct phases. Phase I has the aim of DVB-IP Phase I was to build an IPTV system widely based on proven technologies from the broadcast world (Transport Stream layer and MPEG-2 A/V services), whilst Phase II aims to build on new technologies such as direct IP streaming, supporting the convergence of fixed, mobile TV networks, and web services. The key technologies specified by DVB-IP are service discovery and selection, a DVB Real-Time Streaming Protocol client, MPEG-2 transport over IP, IP address allocation and network time services, receiver identification, and a network provisioning option [15].

2.5. Mobile Television. Over the last few years, a number of studies and trials have shown that Digital Terrestrial technology (DVB-T) offers great potential for portable and mobile reception. Digital Video Broadcast-Handheld (DVB-H), which was developed in DVB Project, is one of the leading global technology standards for the transmission of digital TV to handheld receivers such as mobile telephones and Personal Digital Assistants (PDAs). DVB-H is a nonproprietary open standard physical layer specification designed to enable the efficient delivery of IP-encapsulated data over terrestrial networks. It is designed to accommodate the unique reception and power consumption requirements imposed by mobile users. The DVB-H standard is defined to transmit IP-based services to handheld terminals. IP Datacast (IPDC) is an end-to-end broadcast system for the delivery of any type of digital contents and services using IP-based networks. In particular, IPDC is designed to allow IPDC services reception on terminals, without having to connect to a cellular network.

On the transmission system link layer (implemented in the DVB-H IP encapsulation gateway) there are the following.

(i) Time-Slicing: this feature reduces average power consumption by sending the data in burst mode to the terminals. It also enables smooth and seamless frequency handover.

(ii) MPE-FEC: this feature gives additional robustness and mobility by improving C/N-performance and Doppler performance in mobile channels, and also by improving tolerance to impulse interference.

On the transmission system physical layer (implemented in the DVB-H modulator). there are the following

(i) DVB-H signaling in the TPS-bits to enhance and speed up service discovery.

(ii) 4K mode for trading off mobility and SFN cell size, allowing single antenna reception in medium SFNs at very high speed, thus adding flexibility in the network design.

(iii) In-depth symbol interleaver for the 2K and 4K-modes to improve robustness in a mobile environment and impulse noise conditions [16, 17].

There are also several other competing international mobile Television broadcasting standards. These are the DMB, Media Flo, and ISDB-T.

More precisely the Digital Multimedia Broadcasting (DMB) is a digital radio transmission system for sending multimedia (radio, TV, and datacasting) to mobile devices such as mobile phones. It is an extension to the Digital Audio Broadcasting (DAB) standard and this technology was first developed in South Korea under a national IT project. This standard is adopted in South Korea and DMB trials are currently taking place in several European countries [18]. Media FLO (Forward Link Only) is a Qualcomm-proposed technology for the broadcast of data to portable devices such as cell phones and PDAs. Media FLO is currently deployed in parts of the United States [19]. Finally, Integrated Services Digital Broadcasting-Terrestrial (ISDB-T) is a satellite-to-tower system used in Japan to provide digital service to TV sets and handheld mobile units [20]. China is also currently entering into the mobile broadcasting arena with its own local mobile Television and multimedia standard known as the China Multimedia Mobile Broadcasting (CMMB).

There currently a number of trials of all different mobile broadcast systems across the globe. Given the number of different technologies, the possibility exists to extend mobile TV reception to handheld devices—personal TVs, PDAs, even mobile phones. This can further be enhanced by using 3G networks such as General Packet Radio Service (GPRS) [21] and Universal Mobile Telecommunications System (UMTS) [22] to provide the user with converged services via an interaction channel (usually a network allowing the user to interact with the broadcast/network operator) while on the move [23].

A number of European Union funded research projects have been focusing on the convergence of Telecommunication and broadcast networks as well as the delivery of iTV services to a range of mobile terminals [24, 25].

3. Enabling ITV: ITV Middleware

Digital TV and Internet convergence, however, brings to the forefront several issues related to the distinct nature of the TV and PC display systems, altering in turn the way of viewing Television. This difference has been termed the “lean-forward” versus the “lean-back” experience of viewing. Where a PC user is seated in an upright position and interacts with the system, in contrast to the TV viewer who lies back comfortably.

A Television screen is essentially different from a computer. These differences can be grouped into interaction/input devices, display mechanisms, and user’s viewing styles.

To begin with, on a computer most of the interaction takes place by means of a moving pointer/cursor, controlled by a mouse, whereas on a TV set the interaction is performed via a selection point controlled by a remote control (using the “up,” “down,” “left,” “right,” “ok” and possibly other predefined buttons). There is clear limitation of the TV system and this is more evident when the viewer is being presented with more interactive content, which requires more precise, easy to handle, and sophisticated form of input. Thus it is believed that the sophistication of interaction will be limited by the remote control [8], unless the current remote control system is enhanced, such as including a “thumb navigation” function [26].

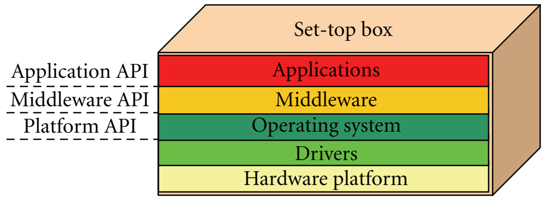


Figure 1: Set-top Box Software Layer Architecture [29].

With regards to display mechanisms, the computer can handle very high resolution (standard: 1024×768) and detailed images in contrast to the TV picture, which has relatively low pixel resolution (720×576) and is usually viewed from a distance [26]. Although this may substantially change if or when the Higher Definition TV (HDTV) becomes widely established (1080×1920 resolution) [27]. Apart from these there are other technical dissimilarities, such as colouring/monitor standards: RGB/CMY (PC) versus YUV/YIQ/YC_bC_r (TV); picture display mode: progressive scanning (PC) versus interlaced scanning (TV) [28].

3.1. The Role of Middleware. In the previous sections we have seen that the convergence of Internet and the TV is a reality in the sphere of iTV. However both mediums are considerably different so require significant differences in their underlying computing architecture models. This is where middleware comes in. Actually the term “middleware” is rather vague, denoting that it is situated between the hardware (the actual equipment) and software (anything that can be stored electronically, such as data and applications).

However in digital TV and set-top box technology, “middleware” stands for a software layer located between the classical operating system (software that provides access to resources and devices) and the applications [5]. Like every software layer, each middleware is characterised by a set of predefined functions that are available to each application and that are known as the Application Programming Interface (API). The task of such interfaces is to abstract components such as operating systems and hardware components thus making the application independent of such platform dependent components. To understand these notions better, Figure 1 illustrates the middleware sitting on top of the operating system as well as the drivers of the hardware platform of the set-top box. Following the discussion above, middleware can be functionally considered as a form of high-level operating system with its own Graphical User Interface (GUI) that defines its own UI look and feel which is the one presented to the viewer-user and not the UI of the low-level set-top box operating system (e.g., Windows, Linux) [29].

3.1.1. Digital TV Applications: EPG, ESG and the Navigator. Digital Television applications are equivalent to software

FIGURE 2: TV Guide from Sky Digital EDG.

programs in the PC domain. Since the word programme has already been reserved to refer to the TV content, a different term had to be adopted to avoid confusion. Hence in the context of Digital Television the term "application" stands for the interactive and stand-alone software programs that are executed in a set-top box or more generally, in a Digital Television environment forming the top of the software layer architecture, as is depicted in Figure 1. There are several types of applications that users come across in the digital TV world and it is therefore appropriate to provide a short description of these.

With so many Digital Television channels to choose from, new ways had to be developed to render the new form of TV more user friendly for viewers. This gave birth to a new series of applications known as Electronic Programme Guides (EPG). Typically, an EPG is a broadcasted application that guides the viewer through the maze of TV programmes [7]. The EPG describes in detail the audio-visual (A/V) content that are to be broadcasted (Names of Programmes, Titles of Programmes, Description of Programme, Schedules broadcasting times of Programmes, etc.). The EPG is described in a standardised XML-based format known as the TV-Anytime [30]. This is generated by each broadcaster/operator, that is, the BBC, Sky, and so forth, and then sent over to an independent third party that combines them into a single EPG that is broadcasted to the end-user terminal (usually set-top box) along with the A/V content/service. This on-screen guide is at the heart of the digital application functionality, enabling viewers to change channels and also see what programmes are on. With a press of a button a menu of channels comes up, allowing you to select and discover content by time, title, channel, and genre. Once the selection has been made the application calls the appropriate middleware and set-top box resources responsible for switching to the specified channel. Figure 2 illustrates a user interacting with Sky Digital's EPG, selecting the "All Channels" option, browsing the available programmes on each channel and finally tuning to the programme and channel of his/her choice.

Another important DTV application quite similar to EPG is the Electronic Service Guide (ESG). Since there are a lot of similarities between EPGs and ESGs there is a lot of confusion and both terms are often used interchangeably in literature. However the main difference between the two is that while an EPG is restricted in listing only the TV programmes available on each channel, an ESG lists all the available services and content a service provider (i.e. a broadcaster) is offering. This can include stand-alone and interactive applications, programme-related and independent content, games, VOD as well as a range of other services. An ESG can also incorporate an EPG, since an EPG is a type of service offered by the service provider. ESG is much more common in DVB-H systems as it is the key to accessing the IP Datacast services.

More precisely, the ESG enables the user of a new Mobile TV device to automatically discover all the service platforms and services available in the usage area, and it even prompts the user to make purchases. The ESG also provides a tool to strengthen customer loyalty to the services through brand imagery and various possibilities to interact with the broadcast service. In addition to the multiple audio and video streams, a service in the Broadcast ESG can include dynamic links and a whole dedicated data stream that populates the mobile terminal memory with files supporting the experience with the live stream.

The Service Guide comprises of data model that models the services, schedules, content, related purchase and provisioning data and the access data in terms of Service Guide fragments. Currently there are two ESG datamodels, the DVB-CBMS and the OMA-BCAST. DVB Convergence of Broadcast and Mobile Services (CBMS) has been specified by the DVB project as part of the DVB-IPDC standard and is mainly designed for DVB-H transmission of general content with bi-directional transmission over the mobile/cellular network (such as UMTS) for interactivity and dedicated content [31]. Open Mobile Alliance (OMA) Broadcast Services Enabler Suite (BCAST) is an open global specification for mobile TV and on-demand video services which can be adapted to any IP-based mobile and P2P content delivery technology [32]. OMA-BCAST is a broadcast bearer-agnostic (IP-based) enabler currently adapted to both broadcast-based systems, such as DVB-H and mobile-based systems, such as MBMS and BCMCS [33]. Comparison of the two standards is beyond the scope of this paper, however if readers are interested they can found more details about the two ESG standard differences in [34].

The Navigator, also known as the "application launcher," comprises a further distinction in the DTV application domain. The Navigator is the system software of the set-

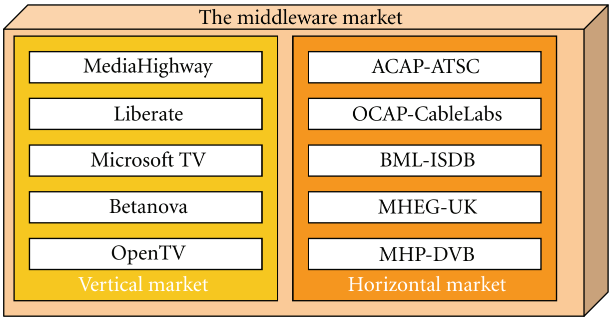


Figure 3: The Fragmented Proprietary versus Open DTV Middleware Market.

top box or mobile device that gives access to the EPG and ESG as well as launching iTV applications and tuning to channels. The Navigator is typically provided by the set-top box manufacturer with its look-and-feel usually determined by the manufacturer as well [35]. However many service providers choose to hide its implementation from the user, therefore substituting its pre-defined UI with their own custom look and feel and integrating its functionality within their own ESG or EPG.

3.2. The Middleware Market: Open versus Proprietary. Following our previous discussion, middleware was developed to enable viewers to use interactive applications, such as Electronic Programme Guides. The first interactive TV platforms using DVB transmission standards were all offered in a vertical market [5]. Typically, in vertical markets a single operator, usually the network/broadcast operator, controls the whole programme delivery chain, ranging from the set-top box specification to the applications and middleware that run on it. This leads to the development of numerous proprietary middleware solutions, which have now been available for several years, with MediaHighway, Liberate, Microsoft TV, NDS Core, Betanova, and OpenTV being the leading players. Obviously the services and applications running on proprietary middleware were tightly linked to these platforms, resulting in service providers and operators having to develop services and applications for all the fragmented middleware solutions.

With the growth iTV, Television standardisation bodies around the world came together to create open middleware standards. These solutions are said to be horizontal. Figure 3 illustrates both vertical and horizontal middleware markets and their standards. Since these were developed by the same bodies which shaped the other digital Television standards, Europe, Japan and the United States all produced different standards for middleware specially designed to work with their own data broadcasting standards. In particular, in the United States the ATSC standard developed the digital TV Applications Software Environment (DASE) middleware system, which formed the basis of the next generation Advanced Common Application Platform (ACAP) standard currently employed for Terrestrial transmissions. In the cable environment the CableLabs standardisation body developed the OpenCable Applications Platform (OCAP) middleware standard. The Japanese ISDB created the Broadcast Markup Language (BML) and the DVB project in Europe developed a common middleware standard for all three types of networks known as the Multimedia Home Platform (MHP) [36]. However because the MHP was only launched in 2001 in Finland several markets, which could not wait for that long, introduced other open standards such as the MHEG-5.

3.3. Proprietary Middleware Solutions. This section is by no means an exhaustive list of the proprietary middleware platforms but rather offers an overview of the most important and leading middleware solutions of the Digital TV vertical market.

3.3.1. MediaHighway. MediaHighway was one of the first proprietary middleware solutions developed by the Canal+ research and development department in 1994 for the launch of the first French Digital Satellite TV service in 1998. Since its launch MediaHighway has been mainly employed by the Satellite providers of the Canal+ Group, that is all, the national variations of Canal+ in Italy, Spain, Netherlands, Finland, Poland, and so forth [29]. MediaHighway's system architecture is not openly available, since it is a proprietary solution. It does however support a number of DTV applications, such as EPG, NVOD, and pay-per-view functionality. In its current version MediaHighway supports Sun Microsystems Java language as a programming language. The MediaHighway Virtual Machine is hardware independent and implements the MediaHighway API in compliance with the Canal+ Technologies (former Canal+ R&D) specifications. Towards the end of 2003, the MediaHighway company was acquired by the NDS middleware provider [37].

3.3.2. Liberate. Liberate Technologies is a provider of interactive TV software to Digital TV network operators. The Liberate middleware solution is based on the Java-based Liberate software engine which is called Navigator Standard. The TV Navigator is a customisable component that is used to match the individual needs of the network operator, supporting a limited number of interactive applications. Although Liberate is designed to work with both Satellite and Cable its main market lies with the Cable delivery of DTV [37].

3.3.3. Betanova. This middleware was developed by BetaResearch, the market leader in digital DVB set-top boxes in cable and satellite networks in German-speaking countries. The Betanova middleware is truly dependent on the D-Box set-top box platform [38]. Hence both Betanova and D-Box have found themselves limited to the German market. The D-Box is a set-top box platform being used for broadcasting TV services as well as interactive services in Germany and is based on the DVB and MPEG-2 standards [39]. The first version of the Betanova middleware was based on the C/C++ programming language. In 1999 BetaResearch deployed the world-wide first Java based middleware called Betanova 2.0. BetaResearch is committed to MHP and at the time of writing is currently migrating Betanova 2.0 to be fully compliant with MHP [29].

3.3.4. Microsoft TV. Microsoft TV is a middleware developed by the giant software company of Microsoft, with products aimed at both mid end and high end set-topboxes. Microsoft TV is a software platform designed specifically for today's cable architecture. It obviously runs on top of the Microsoft Windows operating system and can provide a number of DTV applications such as Video-On-Demand, Electronic Programme Guide, PVR functionality, and access to HDTV programmes [40].

3.3.5. OpenTV. Open TV Inc. is a middleware provider with the largest number of deployments worldwide, currently the leading middleware player in the vertical market, reaching over 27 million set-topboxes produced by more than 30 suppliers worldwide [29]. The core middleware architecture of OpenTV is said to be hardware independent, modular and extensible [5]. The core library, which is at the heart of the middleware is offering several basic functions. Optional functions can be found in the extensions library that allows service providers to personalise the middleware and extend its functionality by downloading several custom made plug-ins. Due to the great number of set-topbox manufacturers and service providers that employ OpenTV, it has to support many different conditional access systems and offer a range of interactive applications. In this respect, OpenTV supports Near-Video-On-Demand, pay-per-view, EPG, PVR functionality, and downloading of data and applications.

The OpenTV platform is based on a new opentv stream added to MPEG-2 audio and video. The opentv stream transmits OpenTV applications that are computer programs. The OpenTV applications are currently developed in ANSI C and compiled with a special development kit compiler. The output from the compiler is called O-code (also known as O-code Virtual Machine) and consists of a private byte code that is interpreted by the O-code interpreter and executed on the digital interactive decoder. The O-code Virtual Machine provides a layer of abstraction from the actual set-top box hardware and operating system beneath it, enabling compiled O-code applications to run on a common "virtual" set-top box that is implemented only in software.

OpenTV provides an object-oriented framework for defining classes of user interface elements called gadgets. A gadget class specifies the behaviour functions for all gadgets of the same class. Gadgets are created and combined by an OpenTV application to form its user interface. To support input processing, OpenTV has the notion of focus. Only one gadget in the tree is designated as having the focus. All input will be directed to this gadget. The gadget is notified of user input by receipt of messages of the appropriate types [41].

Apart from the C-code execution layer, OpenTV provides compatibility with applications authored in HTML and Java code and therefore extends OpenTV middleware to support DVB-MHP. Furthermore OpenTV offers a range of development tools for creating interactive Television applications for OpenTV middleware [42].

3.4. Open Middleware Solutions. The following sections provide a more comprehensive overview of the open middleware solutions developed and currently deployed worldwide, focusing particularly on the European MHP standard.

3.4.1. MHEG and MHEG-5. In 1997 the Multimedia and Hypermedia Experts Group was set up by ISO to create a standard method of storage, exchange and display of multimedia presentations. MHEG-5 is a standard devised for the middleware of digital Teletext services in the United Kingdom. It is an object orientated (scripting) language with predefined classes, objects, inheritance, links and programmes for the creation of Digital TV applications. The standard is also concerned with the interchange of these objects between storage devices and the various networks [43]. MHEG-5 uses the model of multimedia presentations, where a multimedia presentation is a group of scenes, which include a collection of objects such as buttons, graphics, text, and links that define the processes triggered by user interactions [44, 45]. To run MHEG-5 applications the set-top box must have a software component called the MHEG-5 engine, which performs the task of extracting the presentations and scenes to present to the user and handle user navigation and interaction between the different scenes. MHEG-5 was particularly designed to be supported by systems with minimal resources, rendering MHEG-5 ideal for low-end set-top boxes [5].

MHEG-5 applications are constructed from sets of scenes and objects that are common to all scenes. Scene composition consists of a group of objects used to present information, textual, graphical, and so forth and descriptions of those object behaviours based on events. Navigation in an MHEG-5 application is achieved by the transitioning between scenes. An MHEG application (MHEG script, and a collection of multimedia objects) is stored at the service provider end [46]. The MHEG application is then transported to the users/service subscriber's set-top box (terminal) in a bit stream format over the broadcast channel. At the user's terminal an MHEG-5 engine is responsible for extracting the multimedia objects, interpreting the MHEG script and thus displaying the extracted multimedia objects as instructed by the script.

MHEG comes in different versions:

(i) MHEG-1 to 4: the ancestors of MHEG-5, they are rarely used nowadays,

(ii) MHEG-5: that makes it the first horizontal market in Digital TV in the world, it is currently employed by UK digital Terrestrial broadcasters with the most prominent of these being the BBC, it supports applications such as EPGs, teletext, news tickers and interactive games,

(iii) MHEG-6: an extension to MHEG-5 allowing the creation of java-based applications;

(iv) MHEG-7: that defines test and conformance procedures of MHEG-5 applications;

(v) MHEG-8: an extension providing XML scripting for MHEG-5 [47].

3.4.2. The Multimedia Home Platform. In 1997, the DVB consortium decided to develop an open middleware system standard that would resolve the issues of software and hardware interoperability by hiding the specifics of hardware and the operating systems from the actual iTV applications. Apart from interoperability, other issues and aims on the DVB agenda included extensibility (being able to extend functionality), backwards compatibility, modularity, and robustness. The Internet had to be also taken into account and the new standard had to be based on "open" standards and technologies to guarantee a nondiscriminatory access to anybody desiring to use it [29]. The answer was the MHP or DVB-MHP as it is also referred to. The MHP is an open middleware system standard defining a generic software interface (API) between interactive digital applications and the terminals on which those applications reside and execute. It enables digital service and content providers to address many types of terminals, ranging from set-top boxes and integrated TV sets to multimedia PCs. Being an "offspring" of the DVB project, the MHP extends all the DVB open standards and all transmission networks [48].

MHP supports a vast range of iTV applications such as EPG, information services, pay-per-view, applications linked to the main programme, e-commerce, and interactive applications (games, e-betting, etc.). Developed through the open DVB process, MHP provides an essentially open standard and seeks to adopt a patent pooling approach to intellectual property. Every member of DVB is obliged to license on fair and reasonable terms and the major technology holder Sun Microsystems has essentially granted a free license to use its technology in core MHP implementations [36]. In order to provide interoperability in other markets, Globally Executable MHP or GEM specifies those elements of the DVB MHP standard that may be replaced by functional equivalents, thus defining a common core, forming in turn the basis of harmonisation of international standards such as the OCAP, ACAP.

3.4.3. Basic Architecture. The MHP architecture model consists of three layers (see Figure 4). Moving from bottom to top these include the following.

(i) Resources layer: this represents physical resources provided by the hosting terminal.

(ii) System software layer: this represents the MHP API implementation, transport protocols such as Digital Storage Media Control & Command (DSM-CC) and Java API.

(iii) Application layer: this layer represents DVB-J (Xlet) applications that are executed via the MHP

This is quite similar to the structure of the set-top Box software layer architecture of Figure 1. Since DVB MHP implementation is middleware software, it makes no assumption on the amount or the organisation of the hardware and software entities (resources). The resource model also considers that there may be more than one group of hardware/software entities. However, it is irrelevant to the resource model if the logical resources are mapped onto different hardware/software entities. The resource model must present the system resources of the terminal to the rest of the MHP DVB implementation transparently. In other words, the MHP applications (the controlling entity of a DVB services) should be able to have access to all locally connected resources as if they where elements of a single entity.

Applications are not allowed direct access to the system resources. Instead all requests for resources by applications must gain access through the system software layer. Providing an abstract view of resources to applications permits portability of DVB MHP applications. To achieve this, the system software includes the Application Manager function [49].

MHP supports several protocols for transmitting and accessing data through a broadcast as well as an interaction channel. Their definition and description though is beyond the scope of the paper. The MHP specification also defines three sets of profiles.

(i) *Enhanced Broadcast* that mixes digital broadcast of audio-visual data with transmitted applications. This profile does not support an interaction channel, consequently only local interaction is possible.

(ii) *Interactive Broadcast* that adds support for interactive applications. This profile though requires an interaction channel to send and receive data form the head-end. Typical interaction channels supported include PSTN, ISDN, DSL lines as well as 3G mobile networks.

(iii) *Internet Access Profile* that enhances the viewing experience with the addition of the custom functionalities brought by Internet access, such as web browsing, and emailing.

Figure 5 illustrates the functionalities supported by each profile. As we can see MHP version 1.0 includes the first two profiles, whilst MHP version 1.1 adds some further functionality to profiles one and two but deals mainly with the Internet Access Profile. This type of architecture allows the addition of more profiles in the future that can enhance even further MHP's versatility by providing new sets of features such as PVR functionality [49].

3.4.4. Application Model. MHP applications come in two flavours. The first type is DVB-HTML applications. These are not very popular, partly because the specification for DVB-HTML was only completed with MHP 1.1, and partly because many broadcasters, box manufacturers, and content developers find it too complex and difficult to implement [36]. DVB-HTML applications are basically a set of HTML pages that are broadcast as part of a service. Just as standard HTML supports JavaScript or VBscript, DVB-HTML supports a newly defined scripting language by the name ECMAScript.

The second and by far more popular type of MHP applications are DVB-J (DVB-Java) applications. The DVB-J platform includes a virtual machine as defined by the Java Virtual Machine specification from Sun Microsystems and is responsible for running the DVB-J applications. These are written in Java using the MHP API set and consist of a set of class files that are broadcast with a service. We have to remember that although MHP places a strong emphasis on Java it is not Java. In this respect DVB-J applications, also

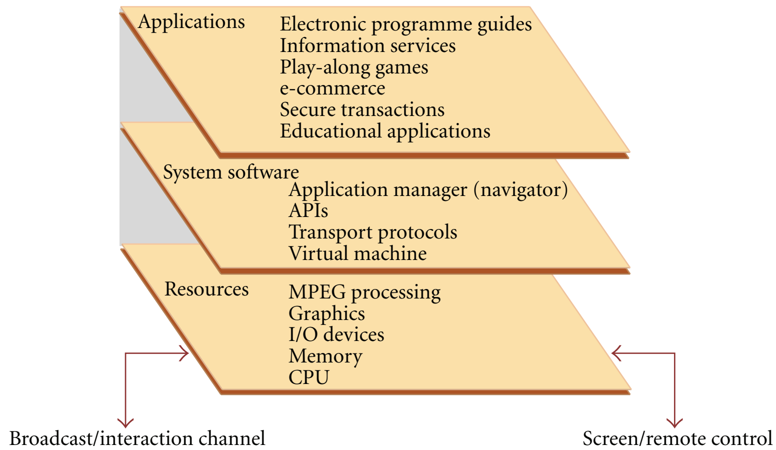


Figure 4: MHP Layer Architecture [48].

called Xlets, are not traditional Java applications, although they are quite similar to applets. Like applets, the Xlet interface allows an external source (the application manager in the case of an MHP receiver) to start and stop an application. There are some major differences, however, between Xlets and applets. The biggest of these is that an Xlet can also be paused and resumed. The rationale behind this feature is that in a Digital TV environment where several Xlets may be running simultaneously, hardware restrictions such as limited processing power in contrast to a standard PC mean that only one Xlet may be visible (playing) at any time and the others must be paused.

An Xlet takes a similar form as a Java Applet by defining certain execution states. The MHP application Manager and the Xlet itself can change its current state by use of the Xlet Context interface. An Xlet can be in one of the following states.

(i) Initialisation/Loaded: this state is reached when an Xlet has been loaded by the Application Manager and begins its execution cycle. Here the Xlet Context interface is obtained and required resources are allocated.

(ii) Started/Active: the Xlet is in an active state providing its intended service to the user.

(iii) Paused: Xlet applications may be paused for a number of reasons, for example, to wait for requested resources, to permit the execution of another Xlet application, and so forth. An Xlet in a paused state can be changed to a Started state. Xlet applications are paused after they have been initialised indicating that all required resources have been made available and the application is ready to be moved into the Start state.

(iv) Destroyed: in this state an Xlet application releases all of its resources and terminates. This state can only be entered once [36].

MHP defines one application model, whereas an application is associated or tied to a particular service. This means that the life cycle of an application is closely connected with its service. Therefore, when the viewer changes channel the Xlets that are associated with the previous channel will be "destroyed". There is actually a loophole to this that allows an application to run independently from a service, provided the Xlet has been previously downloaded onto the memory of the terminal.

3.4.5. Graphics Model, Graphical User Interface, and Applications. One of the main differences between developing applications for the PC and the TV is the way the platform handles graphics. Therefore, a new Graphics model had to be adopted. In the MHP graphics model the various graphical components are situated in three different graphic layers, which are from back to front, a *background* layer, a *video* layer, and lastly a *graphics* layer (see Figure 6). The background layer is used to display either a still image or filled to be a simple colour. The video layer, as its name suggests, displays video content such as the TV programme and/or any video clips. The graphics layer is the most important in terms of application, since it is the plane where all user interface components such as graphics and buttons are drawn. Typically MHP receivers are only required to support a resolution of 720×576 pixels [49]. Although the different layers are rendered independently, it is possible to draw all or parts of a layer transparent or semi-transparent, thus allowing the presentation of applications and video running in the background at the same time (see Figure 6).

Regarding the drawing of the GUI components in the graphics plane, MHP supports the so-called lightweight (platform independent) components of the Java AWT (Abstract Window Toolkit) graphics interface. However AWT has been specifically designed for the PC environment and it does not cope well with non-window-based systems as well as the constraints of a TV environment. As a substitute for the heavy-weight window-based components of AWT, the MHP employs a predefined standard known as HAVi [50]. The Home Audio Video interoperability or HAVi standard defines a set of Java GUI extensions known as the *HAVi Level 2 GUI* which include a new widget set that does not require

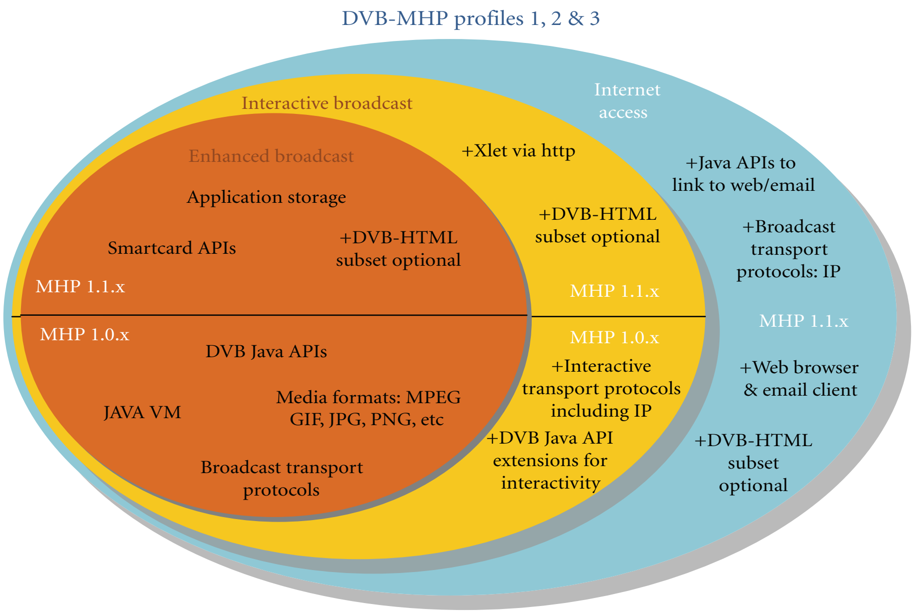


Figure 5: MHP Profiles [48].

a windowing system and a set of classes for managing scarce resources and allowing applications to share the screen when there is no window manager [36]. HAVi was also selected because it allows the control and navigation in an application via a remote control.

In order to speed up the process of creating interactive applications through the MHP, effort has been investited into developing authoring tools to semi-automate the process of creating MHP applications. In particular, Chiao et al. [51] have implemented a template-based MHP authoring tool. The temporal and spatial behavior of an MHP application can be authored and stored in an XML-based instance description file. The MHP authoring tool then generates the target MHP Java source codes. In addition, Hsu et al. [52] have created a layered scene-and-shot model to represent the interactive services content relationship and store in XML form. Individual object is mapped to the HAVi user interface component and associated action so that it could response to the TV user's remote control action. With online Java code generation and compilation, the interactive service can be easily transformed into MHP Xlet application.

Furthermore Alvarez et al. [53] taking into account the issue of manual content update cost, as the initial MHP trials have shown in Spain, developed a fully automated content update system for MHP applications, like news and weather forecasts. In a slightly different context, Cardoso et al. [54] developed a platform which allows the content providers to create enhanced audiovisual contents with a degree of interactivity at moving object level or shot changes in a video. The end user is then able to interact with moving objects from the video or individual shots allowing the enjoyment of additional contents associated to them (additional MHP applications, HTML pages, JPEG, MPEG-4 files, etc.).

3.4.6. OCAP and ACAP. Just as DVB developed a common platform for Digital TV, standards organisations in the United States decided to develop open middleware solutions as well. The ATSC group defined the ACAP middleware standard for Terrestrial and Satellite TV whilst the CableLabs consortium developed the OCAP middleware platform for Cable systems.

At the time, MHP was already well under development and rather than reinvent the wheel CableLabs decided to re-use elements of the MHP standard where it was appropriate. OpenCable Applications Platform provides a middleware software specification intended to enable the developers of interactive Television services and applications to design such products so that they will run successfully on any cable Television system in North America, independent of set-top box or Television receiver hardware or operating system. As with the MHP, OCAP applications come into two flavours; Java-based Applications also known as OCAP-J and HTML-based applications. OCAP also supports three different models of applications as follows.

(i) *Bound* applications are linked directly with the channel the user is currently tuned to and consequently terminate when the viewer selects another channel.

(ii) *Unbound* applications are independent from any particular channel and remain in operation even if a viewer selects another channel.

(iii) *Native* are applications written for a specific host and are not related to a specific broadcast. These may be stored in the firmware of the set-top box [55].

Advanced Common Application Platform is the result of collaboration of the CableLabs OCAP standard and the previous DTV application software environment DASE specification of the ATSC. Like OCAP it is a derivative of MHP. However there are some differences from MHP. These include a slightly modified version of the carousel system used by MHP, a mandatory return channel and support for

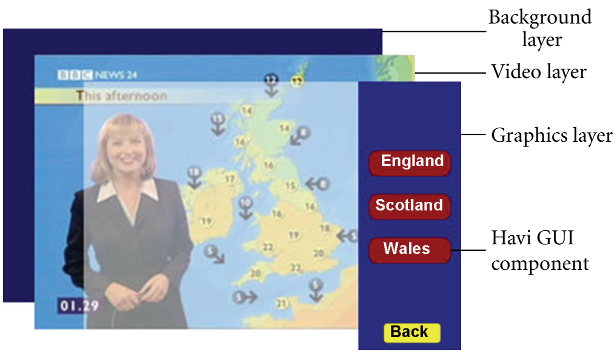


Figure 6: MHP Graphics Layer Model [5].

independent applications which can run at any time and are not tied to a particular channel. Like OCAP and MHP, ACAP Applications are also classified into categories depending upon whether the initial application content processed is based on a procedural or a declarative language. These categories of applications are referred to as procedural (ACAP-J) and declarative (ACAP-X) applications, respectively [56].

3.5. Portable TV and Middleware Platforms. As is clearly visible from the discussion earlier, MHP was designed to be specifically deployed in a stationary living-room environment. In such an environment power supply and processing power of the terminal is not an issue. However there is a trend and a demand today for multimedia services to be accessed in a mobile environment. Digital TV cannot constitute an exception to that. As we have seen the DVB project has defined a standard specially designed for the portable community of viewers, the DVB-H (described earlier). In terms of the software platform however the considerable amounts of hardware and software resources that MHP requires hinder its application on low processing power terminals. Furthermore the error prone nature of the radio-based network interfaces of the mobile world can further complicate matters by preventing an application from being executable if errors occur during the transmission.

The Mobile Information Device Profile or MIDP is a specification put out by Sun Microsystems for the use of Java on portable devices such as mobile phones and PDAs. MIDP sits on top of a configuration, known as the Connected Limited Device Configuration (CLDC) providing a standard Java runtime environment. The fact that the MIDP specification was defined through the Java Community Process by an expert group of more than 50 companies, including leading device manufacturers, wireless carriers, and vendors of mobile software [57], means that it is able to execute applications in a vast range of mobile terminals, which is very important in today's world where a new mobile phone set is introduced nearly every week. MIDP is in fact a cut-down version of Java Standard edition and although it is designed to take into account the restrictions caused by the limited hardware resources of embedded devices, specific DTV features, such as the presentation of video, audio, datagram services and TCP/IP are not included in the basic version of MIDP [5]. It is therefore the terminal manufacturer that has to implement these functions and add a new list of APIs that will interface MIDP with the DVB standards. As a result although MIDP is an open standard, the additional APIs specific to DTV constitute proprietary packages that are owned, for instance, by mobile phone manufacturers such as Nokia and Motorola.

3.5.1. Mobile TV Middleware: JSR 272. Digital broadcasting has recently emerged to bring live Television to cell phones, PDAs, and other mobile devices. Such broadcasts carry not only video and audio but also metadata, and even software applications, in a digital broadcast stream. The new JSR 272, Mobile Broadcast Service API for Handheld Terminals, aims to define a common Java API to control and access digital broadcast content from mobile devices. The JSR 272 is currently being specified via the Java Community Process and along with Motorola the JSR 272 initial expert group including Nokia, Vodafone, and Siemens.

The JSR 272 utilises existing Java Specification Requests (JSRs) of the Java Micro Edition platform for common mobile device-related applications and functions, such as application management and life cycle. The JSR 272 is an application programming interface (API) that allows the application to take control over the broadcast functionalities of a mobile device. It incorporates several distinct broadcast specific functionalities. These include the abilities of quering the electronic service guide, selecting a particular programme or service, presentation and recording of the media content, purchasing and access to broadcast files and objects such as additional content made available by the service provider for downloading [58].

4. A Vision of the Future of Interactive Television

As we complete fifty-five years since the broadcast of the first interactive TV service, it is worthwhile to offer a glimpse into the future of interactive Television. It has been argued that Television is not interactive enough and that it is primarily viewed as a household commodity. The former has dramatically changed since technological innovations in the networks domain have enabled the use of a return path in the broadcast experience and the later is about to change with the advent of mobile Television.

More precisely, the major drivers for the evolution of interactive Television are the latest technological advances in the digital and wireless networks domain, the explosion of nomadic and ubiquitous computing and also the way people consume interactive and new media applications and services today. In light of this, one can foresee five axes of development through which interactive Television will evolve over the next few years, some of which are already along their realisation and all of which are driven by the deployment of IP datacasting, 3D Imaging and user interaction and personalisation. These five axes are described in what follows.

4.1. IPTV—The Internet Revolution. The Internet Protocol (IP) is a data-oriented protocol used for communicating data across packet-switched networks. The Internet Protocol is revolutionalising and transforming Television into a new format for content that encapsulates TV signals within an IP packet data stream. Since IPTV or Internet Television employs the same protocol for the delivery of content such as the Internet, this indicates that IPTV content (realised as IPTV data packets) could potentially be distributed over any IP-capable network, such as ADSL over a cable or telephone broadband connection or as wireless (Wi-Fi, Wi-Max).

IPTV is currently the dominant medium of TV programme broadcasting in young age groups such as teenagers and students. It is expected that in the near future this would expand more to other age groups to include both young adults and professionals especially since more broadcasters across the world are expected to route their regular TV programmes online. In the UK the BBC as well as other service providers have pioneered by making their regular TV programmes available online for a limited period of time, typically one to four weeks [59]. IPTV's ease and cost-effectiveness of content distribution to a potentially wide audience is expected to "give birth" to a plethora of small private service providers that would offer short in length video episodes of niche content online. This new form of webisodes would be very popular to the "snacking" culture attributed to our current and future busy way of life.

The delivery of TV content over the Internet Protocol creates new opportunities for more advanced interactive applications that can be consumed on more powerful processing units other than the conventional set-top boxes and that can be controlled by more conventional PC input devices, such as a mouse and keyboard. Most IPTV users consume iTV services on their personal computers. Thus a new era in the field of iTV applications and services is expected were content providers of popular sci-fi TV series, such as Battlestar Galactica, Stargate Atlantis, and children TV cartoon series, would develop advanced iTV applications that would merge the viewing experience with highly interactive and rich-graphic games by recycling a lot of the TV content into the gaming experience and vice versa. Also several TV programmes of niche content are expected to include social networks, such as Facebook, MySpace, Bebo, and many more into their iTV services as a means of information and content sharing and exchange between members of common interests groups, such as history and travel. Recent research by Mantzari et al. [60] and Geerts and De Grooff [61] has provided an insight into the prospects and requirements for the development of the so-called Social TV. However IPTV has also the opportunity to deliver in the future an even more sociable experience where friends and family would watch virtually together favourite TV programmes, such as sport games, game shows, and movies despite being geographically several miles apart, by merging TV programming with real-time videoconferencing to create a new TV-telepresence service. Harboe et al. [62] have already initiated work in this area by developing a presence awareness platform, by allowing groups of users watching television at home to talk to each other over an audio link. Also Hemmeryckx-Deleersnijder and Thorne [63] have proposed a platform for Video-based awareness via domestic video-calling over the TV.

With TV being delivered via IP any network and any transmission mode be it cable, satellite, terrestrial, or wireless could be potentially employed. Currently, most IPTV services are reaching households via telephone lines and more commonly coaxial and fibre cables. However, the latest developments in the wireless networks domain, especially the introduction of Wi-Fi and Wi-Max, have provided the potential prospect, of what a decade ago seemed as a sci-fi scenario, of IP delivering TV to all households, replacing the conventional TV transmission to roof top antennas. Although this is currently not possible with the bitrates ADSL can achieve today further research investment into the optimisation of the wireless transmission of audiovisual data and potential release of UHF frequencies when all the broadcast to the home will be delivered by cable and wireless networks, would render this possible.

With the forecasted rise of the number of service providers and content multiplicity the current capacity of the Server Farms (a collection of servers employed for hosting and delivering content) will come to a point of no longer being able of hosting, archiving and distributing all available content to all the subscribed users of a service. Peer-to-Peer (P2P) Networks—a P2P network relies primarily on the computing power and bandwidth of the participants in the network rather than concentrating it in a relatively low number of servers—are expected to be employed in the near future to ease the issues arising in the hosting and delivery of IPTV content. The research interest in this area has already commenced several researchers measuring P2P IPTV traffic and developing new algorithms for P2P IPTV streaming [64–66]. In addition to this others are also investigating the use of mobile WiMAX as a potential candidate for the delivery of multimedia services to users [67, 68].

4.2. Mobile Television—The Next Trend. The proliferation of nomadic use of embedded computing devices along with the introduction of Universal Mobile Telecommunications System (UMTS) and 3rd generation (3G) mobile phones, offers for the first time the opportunity of interactive Television services for mobile terminals. This new opportunity for interactive Television on the move is further expanded with the creation of the DVB-H and DMB standards that are aimed at providing broadcast content to low power mobile devices.

Mobile Television is one of the most prominent areas in the context of iTV. Recent user trials of the DVB-H mobile TV system across Europe and the United States of America are reinforcing these predictions. More precisely network operators, broadcasters, and handset manufacturers have conducted a number of user-trials to ascertain the market value also the consumer viewing patterns and use of the emerging technology. The four main trials being in Oxford in the UK (involving 375 users), Helsinki in Finland (involving 500 users), Paris in France (involving 500 users), and New York in the USA (involving 200 users) have shown that there is a high user satisfaction of this new type of service; with 83% of users in the UK, 58% in Finland, 73% in France, and 87% is the USA providing a very positive response to mobile TV [69–72].

In addition 76% of them in the UK, 68% in France and 41% in Finland have expressed willingness to pay for mobile TV when launched. Very interesting are also the trial results which reveal the future viewing patterns of this technology. More precisely, the trials showed that the average British will spend approximately 23 minutes of mobile TV viewing per session, with one to two sessions a day. In France the average daily viewing session is 20 minutes and in Finland it ranges from 5 to 30 minutes of mobile TV per day and in the USA viewing sessions lasted 30–35 minutes with most viewing on weekdays. The most interesting and popular programmes as seen by viewers are new, music, sports and documentaries in the UK and France and local programmes available through Finish national Television and sports in Finland. Notably soap operas are the second most popular type of programme in the UK. Another exceptional finding of all user-trials is that despite specialists' expectations of use of mobile TV outdoors (in bus/train station, while on the move) the collected data clearly illustrates that 50% of the mobile TV use occurs at home and at work. It is shown that most viewers employ mobile TV handsets as an extra TV set that allows them to view Television in other rooms of the house and to resolve programming conflicts within the household. The next most favourite use of mobile TV is in buses, trains and cars. It is gathered from the above user-trials that mobile Television will affect users not only in the way content is viewed but also produced. More precisely mobile TV content will have to be suitable for a "snacking culture." That is to accommodate viewers' limited attention span, account for mobile devices small screens as well as the limited life of the battery. The average length of the average programme watched will be very short to approximately ten to fifteen minutes. In addition because of the small size of the mobile TV screen programme makers and directors will have to adopt and employ cinematic techniques that are more suitable to this new display medium, such as closeups, medium shots, and bigger fonts for titles and avoid the use of wide shots. Relevant work started in this area by looking into ways of automatically cropping the traditional broadcast programme to a chosen area (for mobile TV), based on the semantic attributes of the video content and artistic aspects of video productions [73]. In addition to this, others investigate the use of advanced automated zooming techniques for increasing the mobile TV audiovisual experience without having to crop the broadcast content [74].

Also given the short length of programmes narratives and dialogues would have to be cut down considerably to contain key catch lines for each character of a soap programme, for instance, and key and breaking news stories for news programmes. Advertising will also have to adapt to this medium by creating much shorter advertisements.

Also because of the personalised nature of mobile devices, users would be able to consume and view here-and-now services such as local-based news and information. This will be particularly popular in countries were more than one official languages exist in different regions such as Canada (English, French) and Spain (Spanish, Catalan, Basque), where viewers have a preference for local content broadcasted in their language. The nomadic use of mobile phones and devices also creates new prospects for delivering TV content across borders and across continents to mobile TV users. This would potentially create a new mobile TV roaming service that would enable a user to receive a TV programme(s) he has subscribed to, even when abroad and outside the broadcast region of the service provider. This could be achieved via a sophisticated synergy between IP networks that would route the content across various networks (satellite, wireless, cellular) in order to reach the subscribed user's mobile terminal. Park et al. [75] propose a multistandard global mobile TV system to resolve this, whereas audio, video, and data services provided by the different digital broadcasting mobile TV standards available today would be decoded on a single, unified platform.

The wide adoption and use of mobile TV indoors in contrast to the initially designed outdoor reception environment dictates that several modifications have to be applied to the transmission and reception mechanisms of the mobile TV standards for their successful and effective commercial launch. The PLUTO EU-funded project is investigating this area by researching and developing novel techniques for broadcast transmitter networks by using of transmitter diversity and low cost on-channel repeaters to improve reception in areas of poor coverage such as for mobile reception indoors as well as sparsely populated or obscured locations [76].

Although currently mobile phones and smart phones are not fully capable of PVR functionality, it is expected that the further increase of memory on embedded computing and the increased user demand for TV content recording on mobile phones would see PVRs being soon established as a common mobile TV feature where users would be able to catch up and watch at their own time their favourite shows and programmes while on the move. Currently, researchers are investigating ways of a remote virtual personal video recorder for mobile devices, where the audiovisual content can be selected by users from their mobile screen and recorded on a remote location/server [77, 78].

It is also expected to see changes in the area of graphic rendering and presentation, especially in the light of the range of terminals that TV services can be consumed nowadays. Thus far bitmap graphics, such as JPEG, PNG, GIF has been the popular choice of representing user interfaces elements and components of interactive services. However the demand for service scalability across networks and terminals dictates the use of a digital graphic structure that is able to handle scalability well. Therefore, it is expected to see the popularity of vector graphic formats that have been successful and effective in the Internet domain to dominate iTV user interface presentation systems too. Vector graphic formats such as the Scalable Vector Standard (SVG) specified by the W3C [79] work very well in mobile terminals because of the small file size and scaling enabling iTV services to be downloaded faster compared to conventional bitmap solutions. For this reason W3C has recently specified a separate vector-based standard for mobile terminals known as SVG Tiny [80]. The implication of this is that on one hand middleware developers would be expected to implement vector graphics in their new releases of future middleware solutions and on the other hand iTV service design communities would adopt vector graphic based tools such as Adobe Flash. In particular, as Adobe drops licensing fees and opens up parts of Flash technology [81], this may well trigger the adoption and use of Flash as a common set-top box and mobile user interface (front-end look and feel) iTV standard, dramatically increasing the participation of the creative community in the design of iTV services.

The predicted widespread adoption of mobile TV will very soon raise an issue of control over the look and feel of iTV services over the mobile terminals. On one side of the coin service providers wish to have some control over the look and feel of the services whilst terminal manufacturers wish to have control over the look and feel of their terminals and so there is a conflict here. How will the conflict be resolved is a question the future would decide. A potential solution would though be the introduction of the concept of downloadable user interfaces as opposed to the embedded user interfaces currently in use. In such a scenario the user interface of an iTV service would be designed by the service provider independently of the handset to be consumed onto and would be downloaded with each corresponding iTV service onto the mobile terminal. This is expected to provide both key players in the area of mobile Television with some control over their branding and customer satisfaction.

4.3. Personalised Television—The User-Authored Content Era. Personalised Television would become a very common trend in the near future. Personalisation would spread beyond interactive Television features and services to include TV programmes too. Television has been designed to accommodate single user interaction and selection of services. However in most households several users actually interact with their TV sets and set-top box and each one of them has a different set of preferences in terms of programmes and services.

It is, therefore, expected that future iTV services and applications would in a few-years-time incorporate customisation of several key iTV features via the concept of user profiling. Weiß et al. [82] and Harrison et al. [83] have demonstrated the potential and usefulness of incorporating user profiling in digital multimedia content and more specifically in EPG systems. For instance, each family member would have their own set of favourite channels stored into the set-top box, their own personalised TV guide that makes visible their own preferred channels, rendering the rest inactive. Also as video/movies on demand and Personal Video Recording (PVR) has become a popular feature, it is expected to see personalised PVRs being introduced soon, whereas different TV shows would be recorded for each member of the household according to their predefined profile of favourite shows. This will lead in viewers building their own library of TV shows and movies becoming as common as creating iTunes music playlists. In addition given the number of available channels and shows intelligent recommendation engines would suggest relevant content to viewers given not only their personalised profile and also by tracking the history of viewed content. Efforts in this area have been done by Fernandez et al. [84] in implementing in MHP an Avatar-based DTV recommendation system and by Vaguetti and Gondim [85] in developing a personal recommender prototype for mobile TV. User programme rating and recommendation is also a new feature that would be ported from the Internet domain to the TV environment making user rating a standard part of the TV guide information.

Also given the prospect of a networked Television (ADSL modem being part of the TV or set-top box routing users to the Internet), one can foresee TV programme user rating becoming part of a localised feature where the user can view how neighbours, people in the same city, region or entire country have rated a particular piece of TV content. In this context iTV services and applications can be envisaged as becoming more personalised too, whereas users of a knowledge quiz (e.g., test the nation) can compete against their neighbours or a family of another city and users of a voting or poll application can view at a more geographically divided area how people in their town, region, country registered their opinion on a specific issue.

Users would also be able to personalise the content to be viewed to adding their own personal flavour to the viewing experience. They would be able to modify and enrich broadcasted content at both the end-user terminal (household) and head-end (broadcasters) site. On the end-user terminal site they would be able to make creative alterations to a scene by modifying the background of a scene, perhaps replacing it with their personally acquired photo or video clip, adding a virtual actor or new props into the scene and then view the newly created content, store it, and share it within other household members or friends. They would also be able to dynamically change the storyline of a narrative piece of content or add their own video blog to a documentary by sending through the return path their own-authored content to the broadcaster. User-authored content would become a new content source for broadcasters and other service providers, enriching their regular programming and encouraging the creative and artistic aspirations of the new generation who wishes to share content with the rest of viewers. To achieve this great research investment would have to be made in the area of intelligent semantic annotation of metadata and intelligent extraction of semantic metadata from audiovisual scenes. Work in this area has already started by investigating the enrichment and editing of content by viewers [86]. Cattelan et al. [87] have proposed a watch-and-comment paradigm and Cesar et al. [88] are working on a prototype that would facilitate annotation, enrichment, and sharing of content. User-authored and local-based content are expected to play a vital role especially in mobile Television, as users will be able to easily create and upload their own mobile Television content such as videos and photos shot on the scene directly from their mobile phones.

The personalisation of Television creates opportunities and threats for the advertising revenue streams of several service providers, as with the proliferation of PVR use most advertisements can be fast-forwarded by viewers. The opportunities in this are to make advertising more personalised based on the viewers or households profile stored on the set-top box or mobile phone. These would have to include compelling content that would entice the viewers to interact. Advergames is becoming a very popular approach in the Internet which is expected to be extended in the iTV domain in the near future. The main concept of an Advergame is the implicit and subliminal advertising and awareness of a product via playing a highly interactive and engaging game. This is an entertaining way of implicit advertising where viewers often become both the users and distributors of Advergames, especially through online social networks, hence producing a virtual chain that results in increasing the awareness about a specific product, company or service.

4.4. Smart Space Television—The New Frontier. As wireless networks enter the household environment, it is anticipated that the house of the future would consist of a collection of networked devices and electronic appliances. Television has been and would continue to be at the epicentre of the household occupying the most prominent space in the household, such the living room, offering high quality audiovisual experience.

In such a networked household one can envisage a smart space Television, which apart from the prime purpose of watching TV programmes would be utilised as a media centre for sharing content amongst the household members. In such a scenario the TV would be aware of other networked devices of audiovisual content and would automatically either store them locally or create a link with the devices (iPod, Mp3 players, MP4 players, video cameras, photo cameras, mobile phone, etc.) where the content is hosted. Since TV forms the largest screen in the domestic environment, it is the natural medium for the consumption and sharing of audiovisual content to create a social experience within household.

The Television of the future would go beyond that to create a smart space environment. Using an embedded video camera the smart space Television would be aware of viewers presence in the room and would automatically adjust volume and initiate recording of currently watched content when the viewer is out of the room, so that the audiovisual content can be heard across other rooms and important scenes, such as sport game replays or live action are stored for later viewing. Smart space Television would also be able to control and adjust the lighting of the room to match the content's genre and environment's lighting condition. Efforts in this area have commenced looking in particular the interaction between digital TV receivers and home networks. These scenarios are based on free implementations of open interactive digital TV platforms (MHP) and home network platforms (OSGi) [89, 90]. Researchers are also investigating how can iTV be integrated for the Ambient Assisted Living (AAL) of elderly people [91].

A shift in the ways people interact with their TV services is also expected as new multimodal interaction devices would be developed to accommodate the new role of iTV and to ease the consumption of its services. The current interaction model of the remote control has not been designed for interacting with the vast amount of iTV services and channels people are confronted with. It could be thus forecasted that the drive for wider user adoption of iTV services would foster more research in the area of iTV control-related hardware devices. This, in the coming future, would lead to the use of mobile personal touch screen devices (such as PDAs, iPhone style devices) as a mechanism for accessing iTV services and controlling Television Electronic Programme and Service Guides. This concept of a dual screen interface is expected to become very popular as more and more users are accustomed in viewing and controlling multiple screens simultaneously (TV, PC, mobile phone, etc.). In this concept all graphics and user interface components would be removed from the TV, leaving the audiovisual experience uninterrupted and would be shifted to smaller handheld control units. Efforts in this area are being investigated by conducting user trials in controlling the TV using different remote secondary control devices [92], and some implementation efforts have been made by Cesar et al. [93].

Also advances in the area of multisensory interaction devices would encourage the adoption of a new type of Wii remote for the control of iTV services and applications. This would be also complemented by speech recognition interfaces that would assist the user in navigating through the maze of service providers and their content via registering ones voiced selection. In the years to come and given the dramatic and continuous growth of Television sizes in the household users would be able to use intelligent gesture recognition interfaces for the selection and consumption of content and iTV services, where cameras mounted on TV sets would capture user gestures and translate them into precise user input for the control of an iTV application.

4.5. 3D Television—Seeing Future in Depth. Content creators always look for new forms and ways for improving their content and adding new sensations to the viewer experience. High-Definition and Ultra High-Definition video have been the latest innovation in the area of content enrichment. 3D is the next single greatest innovation in programme-making. There has been a trend in cinema in producing films with 3D enriched content such the latest animated adventure film "Beowulf." These novel forms of 3D content, which is currently the prerogative of big Hollywood studios, would also find its way into small and medium size content creation companies, moving the experience from cinema halls and cinema projectors to the everyday household environments. Three-dimensional imaging and hence three-dimensional television (3DTV) are very promising approaches expected to satisfy these desires [94].

Many different approaches have been adopted in attempts to realise free viewing 3D displays. Several groups have demonstrated autostereoscopic 3D displays, which work on the principle of presenting multiple images to the viewer by use of temporal or spatial multiplexing of several discrete viewpoints to the eyes [95, 96]. However, these autostereoscopic 3D displays are not truly spatial displays since they exclude vertical parallax and rely upon the brain to fuse the two disparate images to create the 3D sensation. As a result stereo systems tend to cause eye strain, fatigue, and headaches after prolonged viewing as users are required to focus to the screen plane but converge their eyes to a point in space, producing unnatural viewing [97, 98]. With recent advances in digital technology, some human factors which result in eye fatigue have been eliminated. However, some intrinsic eye fatigue factors will always exist in stereoscopic 3D technology [99, 100].

Creating a truly realistic 3D real-time viewing experience in an ergonomic and cost effective manner is a fundamental engineering challenge. Holography is a technology that overcomes the shortcomings of stereoscopic imaging and offers the ultimate 3D viewing experience, but their adoptions for 3D TV and 3D cinema are still in its infancy. Holographic recording requires coherent light which makes holography, at least in the near future, unsuitable for live capture.

3D Holoscopic imaging is a technique that is capable of creating and encoding a true volume spatial optical model of the object scene in the form of a planar intensity distribution by using unique optical components [101, 102]. It is akin to holography in that 3D information recorded on a 2-D medium can be replayed as a full 3D optical model; however, in contrast to holography, coherent light sources are not required. This conveniently allows more conventional live capture and display procedures to be adopted. A 3D holoscopic image is recorded using a regularly spaced array of small lenslets closely packed together in contact with a recording device. Each lenslet views the scene at a slightly different angle to its neighbour, and therefore a scene is captured from many view points and parallax information is recorded. It is the integration of the pencil beams, which renders 3D holoscopic imaging unique and separates it from Gaussian imaging or holography. A 3D holoscopic image is represented entirely by a planar intensity distribution. A flat panel display ,for example, one using Liquid Crystal Display (LCD) technology, is used to reproduce the captured intensity modulated image, and a microlens array reintegrates the captured rays to replay the original scene in full colour and with continuous parallax in all directions (both horizontal and vertical). With recent progress in the theory and microlens manufacturing, holoscopic imaging is becoming a practical and prospective 3D display technology and is attracting much interest in the 3D area. It is now accepted as a strong candidate for next generation 3D TV [99].

This 3D Holoscopic content will be interactive and expressive allowing for new visual sensations, since it is inherently more interactive than other kinds of video because 3D objects can be extracted from the 3D Holoscopic video more easily and this will allow more efficiently objects segmentation in 3D space to make the objects in the video more "selectable" as 3D Holoscopic objects. This new 3D format would revolutionalise TV content production and interaction and will lead to a new form of storytelling and content manipulation.

Positioning of characters within a virtual scene at the right position without affecting the realism of the combined 3D scene is of great importance. The development of novel algorithms to accurately compute the depth maps from 3D video [103] will enable accurate positioning of real objects in synthetic scenes and allow the mixing of content and object extraction, where both real and particularly virtual 3D objects would be selectable and moveable across the 3D environment. For instance, in a 3D scene of a Theatre, the author would be able to move, add new, and rearrange virtual objects such as background scene or props but also add and move real 3D object such as actors. This facility is expected to create a very interesting and engaging form of storytelling that encourages content remixing and recycling to produce different narratives, where a new meaning is conveyed each time, based on the artistic touches of the author onto the scene.

Finally taking an even deeper look into the future one would gradually foresee 3D Television (based on holoscopic technology) being replaced by Holographic Television where the viewing experience becomes a real sensation and the viewer interacts with the holographic content in a truly ambient and immersive environment.

5. Conclusion

This survey paper has presented a brief overview of the evolutionary path to the convergence of iTV services. This paper has presented and categorised the key software technologies that enable the convergence of digital TV services and interaction of set-top boxes as well as mobile platforms. As the concept of Television services are amalgamated into other multimedia services, due to the convergence of networks, new service concepts are bound to redefine and add additional subcategories to those defined within this paper. The issue of spectrum allocation is also seen as one of the key thrusts for the future development of interactive Television especially as there are more ways to broadcast than ever before, namely, terrestrial, cable, satellite, mobile, and the Internet. Finally, a new vision for the future of interactive Television has been offered. As the principal drive and innovation of the 1990s was bringing the Internet in the TV environment, the 21st century would be about bringing the TV into the Internet and 3D into TV to create an ambient and immersive personalised user experience.

References

[1] K. Sanjay and S. Jadon, "Video on demand: an overview," in *Proceedings of the National Workshop on IT Services and Applications (WITSA '03)*, New Delhi, India, February 2003.

[2] D. S. Alberts and D. S. Papp, *Information Age: An Anthology on Its Impact and Consequences*, National Defense University, Washington, DC, USA, 1997.

[3] P. D'Agostino, *Transmission: Toward a Post-Television Culture*, Sage, New York, NY, USA, 2nd edition, 1995.

[4] D. Wood, "The options for Broadcast Interaction," in *Proceedings of the Interactive Services over Digital Broadcast and Cable Networks*, IBC Group, London, UK, October 1998.

[5] U. Reimers, *DVB: Family of International Standards for Digital Video Broadcasting*, Springer, Berlin, Germany, 2nd edition, 2005.

[6] J. F. Jensen, "A short history of ITV trials and ITV consumer studies," in *White Paper on Interactive TV*, M. Rose, et al., Ed., pp. 21–53, Technical University of Denmark, 1999.

[7] S. Papathanassopoulos, *European Television in the Digital Age*, Polity Press, London, UK, 2002.

[8] E. England and A. Finney, *Managing Multimedia: Project Management for Web and Convergent Media, Book 2: Technical Issues*, Addison-Wesley, London, UK, 2002.

[9] A. Gil, et al., "Surfing the Web on TV: the MHP approach," in *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME '02)*, vol. 2, pp. 285–288, Lausanne, Switzerland, August 2002.

[10] M. Milenkovic, "Delivering interactive services via a digital TV infrastructure," *IEEE Multimedia*, vol. 5, no. 4, pp. 34–43, 1998.

[11] D. Bradshaw, "Internet delivery of television material," in *Proceedings of the W3C Workshop on Television and the Web*, Sophia-Antipolis, France, June 1998.

[12] Microsoft, "Web TV products: MSN-TV," 2008, http://www.webtv.com/pc.

[13] VEOH, "Veoh TV," 2008, http://www.veoh.com/.

[14] H. Bürklin, R. Schäfer, and D. Westerkamp, "DVB: from broadcasting to IP delivery," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 1, pp. 65–67, 2007.

[15] A. J. Stienstra, "Technologies for DVB services on the internet," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 228–236, 2006.

[16] M. Kornfeld and G. May, "DVB-H and IP datacast—broadcast to handheld devices," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 161–170, 2007.

[17] Y.-K. Wang, I. Bouazizi, M. M. Hannuksela, and I. D. D. Curcio, "Mobile video applications and standards," in *Proceedings of the International Workshop on Mobile Video (MV'07)*, pp. 1–6, ACM, Augsburg, Germany, 2007.

[18] K.-T. Lee, Y.-S. Park, S.-H. Park, J.-H. Paik, and J.-S. Seo, "Development of portable T-DMB receiver for data services," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 1, pp. 17–22, 2007.

[19] M. R. Chari, F. Ling, A. Mantravadi, et al., "FLO physical layer: an overview," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 145–160, 2007.

[20] M. Takada and M. Saito, "Transmission system for ISDB-T," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 251–256, 2006.

[21] B. Ghribi and L. Logrippo, "Understanding GPRS: the GSM packet radio service," *Computer Networks*, vol. 34, no. 5, pp. 763–779, 2000.

[22] J. Soininen, "GPRS and UMTS release 2000 A11-IP option," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 4, no. 3, pp. 30–37, 2000.

[23] P. Leroux, V. Verstraete, F. De Turck, and P. Demeester, "Synchronized interactive services for mobile devices over IPDC/DVB-H and UMTS," in *Proceedings of the 2nd IEEE/IFIP International Workshop on Broadband Convergence Networks (BCN '07)*, pp. 1–12, Munich, Germany, May 2007.

[24] CISMUNDUS, "Convergence of IP-based services for mobile users and networks in DVB-T and UMTS systems," 2008, http://dea.brunel.ac.uk/instinct/.

[25] INSTINCT, "IP-based networks, services and terminals for convergence systems," 2008, http://www.ist-instinct.org.

[26] L. Eronen and P. Vuorimaa, "User interfaces for digital television: a navigator case study," in *Proceedings of the Workshop on Advanced Visual Interfaces (AVI '00)*, pp. 276–279, ACM, Palermo, Italy, May 2000.

[27] C. Sandback, "Digital TV in the convergent environment computer graphics and applications," *IEEE Computer Graphics and Applications*, vol. 21, no. 1, pp. 32–36, 2001.

[28] S. Othman, "Interactive data services for television: display and user interface issues," in *Proceedings of the Northeon Conference*, pp. 86–90, Seattle, Mass, USA, 1998.

[29] H. Benoit, *Digital Television: MPEG-1, MPEG-2 and Disciplines of the DVB System*, Focal Press, London, UK, 2nd edition, 2002.

[30] The TV-Anytime, "TV-Anytime Forum," 2009, http://www.tv-anytime.org.

[31] ETSI, Digital Video Broadcasting, IP Datacast over DVB-H: Electronic Service Guide (ESG), ETSI TS 102 471, V1.2.1, 2006.

[32] F. Hartung, M. Kampmann, and T. Rusert, "The OMA BCAST standard for bearer-independent mobile TV services," in *Proceedings of the International Workshop on Mobile Video (MV '07)*, pp. 13–18, ACM, Augsburg, Germany, September 2007.

[33] F. Hartung, U. Horn, J. Huschke, M. Kampmann, T. Lohmar, and M. Lundevall, "Delivery of broadcast services in 3G networks," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 188–198, 2007.

[34] H.-T. Chiao, "Comparison of the notification services between OMA BCAST 1.0 and DVB-IPDC phase 2," in *Proceedings of the 11th IEEE Singapore International Conference on Communication Systems (ICCS '08)*, pp. 327–331, November 2008.

[35] C. Peng and P. Vuorimaa, "A digital navigator," in *Proceedings of the 8th ACM International Conference on Multimedia (MULTIMEDIA '00)*, pp. 429–431, ACM, Los Angeles, Calif, USA, October 2000.

[36] S. Morris and A. Smith-Chaigneau, *Interactive TV Standards: A Guide to MHP, OCAP, and Java TV*, Focal Press, London, UK, 2005.

[37] LIBERATE, "The Liberate Middleware Product," 2008, http://products.liberate.com.

[38] F. Lonczewski and R. Jaeger, "An extensible set-top-box architecture for interactive and broadcast services offering sophisticated user guidance," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '00)*, vol. 3, pp. 1403–1406, New York, NY, USA, July-August 2000.

[39] S. Pekowsky and R. Jaeger, "The set-top box as multi-media terminal," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 833–840, 1998.

[40] Microsoft TV, "The Microsoft TV Platform," 2008, http://www.microsoft.com/mediaroom/.

[41] F. Bouilhaguet, J.-C. Dufourd, S. Boughoufalah, and C. Havet, "Interactive broadcast digital television. The OpenTV platform versus the MPEG-4 standard framework," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '00)*, vol. 3, pp. 626–629, Geneva, Switzerland, 2000.

[42] OpenTV, "OpenTV Middleware," 2008, http://www.opentv.com.

[43] R. Price, "MHEG: an introduction to the future international standard for hypermedia object interchange," in *Proceedings of the 1st ACM International Conference on Multimedia*, pp. 121–128, ACM, Anaheim, Calif, USA, 1993.

[44] ISO/IEC JTC1/SC29/WG12 DIS 13522-1, "Information Technology-Coding of Multimedia and Hypermedia information-Part 1: MHEG Object Representation and Base Notation (ASN.1)," August 1995.

[45] "Digital Television Group Digital Television MHEG-5 Specification," 2003, http://www.dtg.org.uk/publications/books_mheg.html.

[46] M. Echiffre, C. Marchisio, P. Marchisio, P. Panicciari, and S. Del Rossi, "MHEC-5-aims, concepts, and implementation issues," *IEEE MultiMedia*, vol. 5, no. 1, pp. 84–91, 1998.

[47] T. Mayer-Boudnik and W. Effelsberg, "MHEG explained," *IEEE MultiMedia*, vol. 2, no. 1, pp. 26–38, 1995.

[48] J. Jones, "DVB-MHP/Java TV data transport mechanisms," in *Proceedings of the 40th International Conference on Tools Pacific: Objects for Internet, Mobile and Embedded Applications*, vol. 21, pp. 115–121, ACM, Sydney, Australia, 2002.

[49] European Telecommunications Standards Institute (ETSI), Digital Video Broadcasting (DVB): Multimedia Home Platform (MHP) specification 1.1.1 ETSI TS 102 812 v1.2.1, 2003.

[50] HAVi, "The HAVi Specification: Specification of the Home AudioNideo Interoperability (HAVi)," Architecture.Sony, Philips, Hitachi, Sharp, Matsushita, Thomson, Toshiba, Grundig. Version 1.0 Final, 2008, http://www.havi.org.

[51] H.-T. Chiao, K.-S. Hsu, Y.-K. Chen, and S.-M. Yuan, "A template-based MHP authoring tool," in *Proceedings of the 6th IEEE International Conference on Computer and Information Technology (CIT '06)*, p. 138, September 2006.

[52] K.-S. Hsu, C. Yang, and T.-H. S Chao, "An authoring tool and extendable production architecture for interactive television services," in *Proceedings of the 10th IEEE International Symposium on Consumer Electronics (ISCE '06)*, pp. 1–6, 2006.

[53] R. P. M Alvarez, J. L. Rodriguez-Diez-Caballero, J. M. Pousada-Carballo, et al., "Automating content update for MHP applications: a practical experience," in *Proceedings of the 6th IEEE Consumer Communications and Networking Conference (CCNC '09)*, pp. 1–6, January 2009.

[54] B. Cardoso, F. de Carvalho, L. Carvalho, et al., "Hyperlinked video with moving objects in digital television," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '05)*, pp. 486–489, July 2005.

[55] CableLabs, "OpenCable Application Platform Specification," OCAP 1.0 profile OC-SP-OCAP1.0-I07-030522, 2004.

[56] ATSC, "ATSC Candidate Standard 101A," Advanced Common Application Platform (ACAP), August 2005.

[57] Java-Sun, "Mobile Information Device Profile (MIDP)," JSR 37, JSR 118, 2008, http://java.sun.com/products/midp/overview.html.

[58] JCP, "Mobile Broadcast Service API for Handheld Terminals," JSR 272, 2007, http://jcp.org/en/jsr/detail?id=272 .

[59] BBC, "BBC's iPlayer," 2008, http://www.bbc.co.uk/iplayer.

[60] E. Mantzari, G. Lekakos, and A. Vrechopoulos, "Social TV: introducing virtual socialization in the TV experience," in *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video (UXTV '08)*, vol. 291, pp. 81–84, ACM, Silicon Valley, Calif, USA, October 2008.

[61] D. Geerts and D. De Grooff, "Supporting the social uses of television: sociability heuristics for social TV," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*, pp. 595–604, ACM, Boston, Mass, USA, April 2009.

[62] G. Harboe, N. Massey, C. Metcalf, D. Wheatley, and G. Romano, "The uses of social television," *Computers in Entertainment*, vol. 6, no. 1, pp. 1–15, 2008.

[63] B. Hemmeryckx-Deleersnijder and J. M. Thorne, "Awareness and conversational context-sharing to enrich TV-based communication," *Computers in Entertainment*, vol. 6, no. 1, pp. 1–13, 2008.

[64] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross, "A measurement study of a large-scale P2P IPTV system," *IEEE Transactions on Multimedia*, vol. 9, no. 8, pp. 1672–1687, 2007.

[65] S. Liu, R. Zhang-Shen, W. Jiang, J. Rexford, and M. Chiang, "Performance bounds for Peer-assisted live streaming," in *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '08)*, vol. 36, pp. 313–324, Annapolis, Md, USA, June 2008.

[66] F. Liu and Z. Li, "A measurement and modeling study of P2P IPTV applications," in *Proceedings of the International Conference on Computational Intelligence and Security (CIS '08)*, vol. 1, pp. 114–119, December 2008.

[67] B. Sousa, K. Pentikousis, and M. Curado, "Evaluation of multimedia services in mobile WiMAX," in *Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia (MUM '08)*, pp. 64–70, ACM, Umeå, Sweden, December 2008.

[68] I. V. Uilecan, C. Zhou, and G. E. Atkin, "Framework for delivering IPTV services over WiMAX wireless networks," in *Proceedings of the IEEE International Conference on Electro/Information Technology (EIT '07)*, pp. 470–475, May 2007.

[69] S. Mason, "Mobile TV: results from the DVB-H trial in Oxford," EBU Technical Review, April 2006, http://www.ebu.ch/en/technical/trev/trev_306-mason.pdf.

[70] DVB-H, "Paris Mobile TV Customer Trial Results," DVB-H: Global Mobile TV, 2008, http://www.dvb-h.org/Services/services-paris-canalplus.htm.

[71] J. Mäki, "Finish Mobile TV Pilot. Research International Finland. Finish Mobile TV," 2005, http://www.finnishmobiletv.com/press/Final_RI_Press_300805_english.pdf.

[72] DVB-H, "NewYork (Modeo) DVB-H Trial Results," DVB-H: Global Mobile TV, 2008, http://www.dvb-h.org/Services/services-usa-modeo.htm.

[73] J. Deigmoeller, T. Itagaki, and G. Stoll, "An approach for an intelligent crop and scale application to adapt video for mobile TV," in *Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '08)*, pp. 1–7, March-April 2008.

[74] H. Knoche, M. Papaleo, M. A. Sasse, and A. Vanelli-Coralli, "The kindest cut: enhancing the user experience of mobile TV through adequate zooming," in *Proceedings of the 15th ACM International Multimedia Conference and Exhibition (MULTIMEDIA '07)*, pp. 87–96, Augsburg, Germany, September 2007.

[75] Y.-S. Park, S.-H. Park, K.-T. Lee, J.-H. Paik, and J.-M. Chung, "Implementation of a global mobile TV system," in *Proceedings of the International Conference on Consumer Electronics (ICCE '09)*, pp. 1–2, January 2009.

[76] Y. Zhang, C. Zhang, J. Cosmas, et al., "Analysis of DVB-H network coverage with the application of transmit diversity," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 568–577, 2008.

[77] C. Giordano, S. Modeo, G. Bernardi, and F. Ricchiuti, "Using mobile phones as remote control for ubiquitous video-recording," in *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia (MUM '07)*, vol. 284, pp. 125–130, ACM, Oulu, Finland, December 2007.

[78] H.-T. Chiao, F.-C. Chen, K.-S. Hsu, and S.-M. Yuan, "Video everywhere through a scalable IP-streaming service framework," in *Proceedings of the 3rd International Symposium on*

Wireless Communication Systems (ISWCS '06), pp. 190–194, September 2006.

[79] W3C, "Scalable Vector Graphics (SVG) 1.1 Specification," W3C Recommendation, January 2003.

[80] W3C, "Scalable Vector Graphics (SVG) Tiny 1.2 Specification," W3C Candidate Recommendation, August 2006.

[81] Adobe, "Open Screen Project," 2008, http://www.adobe.com/openscreenproject.

[82] D. Weiß, J. Scheuerer, M. Wenleder, A. Erk, M. Gülbahar, and C. Linnhoff-Popien, "A user profile-based personalization system for digital multimedia content," in *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts (DIMEA '08)*, vol. 349, pp. 281–288, Athens, Greece, September 2008.

[83] C. Harrison, B. Amento, and L. Stead, "iEPG: an ego-centric electronic program guide and recommendation interface," in *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video (UXTV '08)*, vol. 291, pp. 23–26, ACM, Silicon Valley, Calif, USA, October 2008.

[84] Y. Blanco Fernández, J. J. Pazos Arias, M. López Nores, A. Gil Solla, and M. Ramos Cabrer, "AVATAR: an improved solution for personalized TV based on semantic inference," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 1, pp. 223–231, 2006.

[85] L. Vaguetti and P. R. L. Gondim, "Hierarchical intra-context handoff awareness approach to personal recommender systems: a mobile DTV case study," in *Proceedings of the 1st International Conference on the Applications of Digital Information and Web Technologies (ICADIWT '08)*, pp. 168–173, August 2008.

[86] M. D. G. C. Pimentel, R. G. Cattelan, E. L. Melo, and C. A. C. Teixeira, "End-user editing of interactive multimedia documents," in *Proceedings of the 8th ACM Symposium on Document Engineering (DocEngN '08)*, pp. 298–301, ACM, Sao Paulo, Brazil, September 2008.

[87] R. G. Cattelan, C. Teixeira, R. Goularte, and M. D. G. C. Pimentel, "Watch-and-comment as a paradigm toward ubiquitous interactive video editing," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 4, pp. 1–24, 2008.

[88] P. Cesar, D. C. Bulterman, D. Geerts, J. Jansen, H. Knoche, and W. Seager, "Enhancing social sharing of videos: fragment, annotate, enrich, and share," in *Proceeding of the 16th ACM International Conference on Multimedia (MM '08)*, pp. 11–20, ACM, Vancouver, Canada, October 2008.

[89] M. R. Cabrer, R. P. Díaz Redondo, A. F. Vilas, J. J. Pazos Arias, and J. G. Duque, "Controlling the smart home from TV," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 2, pp. 421–429, 2006.

[90] M. R. Cabrer, R. P. D. Redondo, A. F. Vilas, and J. J. P. Arias, "Controlling the smart home from TV," in *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE '06)*, pp. 255–256, January 2006.

[91] V. Fuchsberger, "Ambient assisted living: elderly people's needs and how to face them," in *Proceeding of the 1st ACM International Workshop on Semantic Ambient Media Experiences (SAME '08)*, pp. 21–24, ACM, Vancouver, Canada, October 2008.

[92] E. Tsekleves, R. Whitham, K. Kondo, and A. Hill, "Bringing the television experience to other media in the home: an ethnographic study," in *Proceeding of the EuroITV*, D. Geerts, P. Cesar, and D. de Grooff, Eds., pp. 201–210, ACM, Leuven, Belgium, June 2009.

[93] P. Cesar, D. C. A. Bulterman, and A. J. Jansen, "Usages of the secondary screen in an interactive television environment: control, enrich, share, and transfer television content," in *Proceedings of the 6th European Conference on Changing Television Environments (EuroITV '08)*, M. Tscheligi, M. Obrist, and A. Lugmayr, Eds., vol. 5066 of *Lecture Notes in Computer Science*, pp. 168–177, Springer, July 2008.

[94] T. Motoki, H. Isono, and I. Yuyama, "Present status of three-dimensional television research," *Proceedings of the IEEE*, vol. 83, no. 7, pp. 1009–1021, 1995.

[95] B. Javidi and F. Okano, "Special issue on 3-D technologies for imaging and display," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 487–488, 2006.

[96] N. A. Dodgson, "Autostereoscopic 3D displays," *IEEE Computer*, vol. 38, no. 8, pp. 31–36, 2005.

[97] T. Yamazaki, K. Kamijo, and S. Fukuzumi, "Quantitative evaluation of visual fatigue encountered in viewing stereoscopic 3D displays," in *Proceedings of the Society for Information Display*, vol. 31, pp. 245–247, 1990.

[98] M. T. M. Lambooij, W. A. Ijsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: a review," in *Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490 of *Proceedings of SPIE*, pp. 1–13, 2007.

[99] L. Onural, T. Sikora, J. Ostermann, A. Smolic, R. Civanlar, and J. Watson, "An assessment of 3dtv technologies," in *Proceedings of the NAB Broadcast Engineering Conference*, pp. 456–467, Las Vegas, Claif, USA, June 2006.

[100] L. Onural, "Television in 3-D: what are the prospects?" *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1143–1145, 2007.

[101] S. Monaleche, A. Aggoun, A. McCormick, N. Davies, and Y. Kung, "Analytical model of a 3d recording camera system using circular and hexagonal based spherical microlenses," *Journal of the Optical Society of America A*, vol. 18, no. 8, pp. 1814–1821, 2001.

[102] A. Aggoun, "Pre-processing of integral images for 3-D displays," *IEEE Journal of Display Technology*, vol. 2, no. 4, pp. 393–400, 2006.

[103] C. Wu, A. Aggoun, J. McCormick, and Y. Kung, "Depth extraction from unidirectinal integral imaging using a modified multi-baseline technology," *Journal of Electronic Imaging*, vol. 14, no. 2, pp. 1–9, 2005.

[104] NDS, "MediaHighway Middleware," 2008, http://www.nds.com/solutions/mediahighway.php.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 579569, 17 pages
doi:10.1155/2009/579569

Research Article

Interactive Digital Terrestrial Television: The Interoperability Challenge in Brazil

Jordi Amatller Clarasó,[1] David Baldo,[2] Giuliano Benelli,[2] Giovanni Luca Daino,[2] and Riccardo Zambon[2]

[1] *Tmira Solutions, Camino Terminillo (Centro Empresas) KM 2, 16003 Cuenca, Spain*
[2] *Department of Information Engineering, University of Siena, Via Roma 56, 53100 Siena, Italy*

Correspondence should be addressed to Riccardo Zambon, zambon@unisi.it

Received 30 January 2009; Accepted 15 June 2009

Recommended by Robert Briskman

This paper introduces different standards implemented in existing Digital Terrestrial Television Broadcasting systems to allow the fruition of interactive services and applications through digital Set Top Boxes. It focuses on the interoperability issue between the Brazilian and the European architectures. In fact, despite in Brazil the GEM specification has been designed to foster wide content compatibility across a range of interactive platforms, it has never come to a final implementation and deployment. As a result the interoperability issue has been deeply explored in the BEACON project and an innovative system architecture has been developed to deploy t-learning services across Europe and Brazil, providing integration of those systems that were not able to interoperate until nowadays. This work is an important step in the direction of standards' interoperability. As a result, MHP and Ginga NCL-Lua implementation appeared to be the very best choice to deliver interactive services in an interoperable mode between European and Brazilian digital television.

1. Introduction

In the past, several DTTB (*Digital Television Terrestrial Broadcasting*) systems have been designed and developed, differentiating upon how each modulation system meets specific conditions such as the use of the spectrum resource, coverage requirements and transmission network structure, reception conditions, type of service required, policy, cost to the consumers and broadcasters, and so forth.

Those DTV (*Digital Television*) systems are now replacing old analog TV systems all over the world with the introduction of several standards (European, Japanese, North-American, Korean, Brazilian, etc.) to supply innovative features and services.

Moreover, a relevant added value for digital broadcasting comes from the chance to develop interactive applications through reference standards related to each DTV system implementations enumerated above.

As a matter of fact, the interactivity standardization has not been developed in an homogeneous way, because every DTV system needs its own *Application Programming Interface*s (APIs) to communicate with the interactivity middleware and every implementation is grown basing on its country technologies, needs and laws; so those standards are generally not compatible.

To allow broad content compatibility across a range of new interactive platforms, the GEM (*Globally Executable MHP*) [1] specification has been created. It defines a subset of APIs removing the transmission-related elements and retaining the application's ones. As a consequence, application developers can build applications that work on any GEM-compliant platform, and middleware developers can use GEM as the core of their products, by mean of some customization to provide different versions of middleware.

Unfortunately, the use of GEM specification in Brazil is unavailable due to some legal and royalties issues. As a result, it is actually not possible to adopt the GEM specification solution to create interactive applications that can work on different DTV systems also including the Brazilian one.

To solve this problem, going a step ahead in the interoperability between DTV standards, the BEACON project identified a different solution compliant with Brazilian middleware standards. Consequently, based on an innovative system for management and broadcasting, an integrated system architecture has been designed to deliver t-learning services across Europe and Brazil, providing integration of systems that until nowadays was not able to cooperate.

This goal is achieved providing methodologies, process schemas, and pilot applications granting new opportunities for all the players involved in the value chain.

A relevant improvement to the social inclusion is also provided because, as it will be explained in the next paragraphs, the t-learning services delivered with the BEACON platform are addressing the need to cover the Brazilian social-gap permitting (through t-learning) a profitable training for entering the university.

To help the reader going through the article, sections are organized as follows. Section 2 gives an overview of DTTB standards involved in the project and in its overall architecture. Section 3 describes the interactivity standard developed for those DTV systems, introducing a brief GEM description and an exhaustive paragraph regarding the Brazilian interactivity standardization issues and problems. Then Section 4 describes the BEACON project focusing on its objectives and the methodologies involved in the design of the platform. A deeper analysis of the system architecture implementation is provided in Section 5, where the whole architecture is described focusing on the t-learning service implementation, the technologies involved in the applications' deployment, and the needed upgrades to obtain the integration between the European and the Brazilian DTV frameworks. Finally, Section 6 present the project outcomes and Section 7 opens a window on the future works concerning these system architectures and how they can strongly improve the benefits brought from our works.

2. DTTB Standard Overview

In the next paragraphs we provide a brief description of the DTTB European standard, the Japanese standard, and the Brazilian standard (that is directly derived from the previous ones).

This overview is not exhaustive, because actually there are also a North-American terrestrial television implementation (named *Advanced Television Systems Committee*, ATSC) and a Korean one (*Digital Multimedia Broadcasting-Terrestrial*, DMB-T) with their related interactivity standards, but they are out of the scope of this work; so they are omitted.

As a matter of fact, this brief overview just provides a basic knowledge of the lower layers supporting the system architecture that we have deployed and that will be illustrated in this article.

2.1. DVB-T. The DVB-T (*Digital Video Broadcasting-Terrestrial*) specification refers to terrestrial broadcasting. The system has been designed to operate within the existing UHF spectrum allocated to analogue television transmissions. The system was developed for 8 MHz channels but it can be scaled to any channel bandwidth (8, 7, or 6 MHz) with corresponding scaling in the data capacity. The net bit rate available in 8 MHz channel is in the 4.98–31.67 Mbps range, depending on the choice of channel coding parameters, modulation types, and guard interval duration.

The system was essentially designed to be able to adapt to all types of channels. It is capable to cope not only with Gaussian channels, but also with Ricean and Rayleigh channels. It can withstand high-level (up to 0 dB) long delay static and dynamic multipath distortion.

The system is robust to interference from delayed signals, either echoes resulting from terrain or building reflections or signals from distant transmitters in a single frequency network arrangement [2, 3].

The system features a number of selectable parameters that accommodate a large range of carrier-to-noise ratios and channel behaviours. It supports fixed, portable, or mobile reception, with a consequential trade-off in the usable bit rate.

This range of parameters allows the broadcasters to select a mode appropriate to the application foreseen. For instance, a pretty robust mode (with a correspondingly lower data rate) is needed to ensure reliable portable reception with a simple set-top antenna. A less robust mode with a higher data rate could be used where the service planning involves frequency-interleaved channels [3, 4].

Less robust modes with the highest data rates can be used for fixed reception and for digital TV broadcasting.

2.2. ISDB-T. ISDB (*Integrated Services Digital Broadcasting*) is a new type of broadcasting intended to provide audio, video, and multimedia services. The system was developed for *terrestrial* (ISDB-T) and *satellite* (ISDB-S) broadcasting.

For terrestrial broadcasting, the system has been designed to have enough flexibility to deliver digital television and sound programs and offer multimedia services in which different types of digital information such as video, audio, text, and computer programs will be integrated. It also aims to provide stable reception through compact, light and inexpensive mobile receivers in addition to integrated receivers typically used in homes.

The system uses a modulation referred to as *Band-Segmented Transmission* (BST) OFDM, which consists of a set of common basic frequency blocks called BST-Segments. Each segment has a bandwidth corresponding to 1/14th of the terrestrial television channel spacing (6, 7, or 8 MHz depending on the region). For example, in a 6 MHz channel, one segment occupies 6/14 MHz = 428.6 kHz spectrum seven segments occupy 3 MHz.

It also provides hierarchical transmission capabilities by using different carrier modulation schemes and coding rates of the inner code on different BST-segments. Each data segment can have its own error protection scheme (coding rates of inner code, depth of the time interleaving) and type of modulation (QPSK, DQPSK, 16-QAM, or 64 QAM). Each segment can then meet different service requirements. A number of segments may be combined flexibly to provide a wideband service (e.g., High Definition TV). By transmitting OFDM segment groups with different transmission parameters, hierarchical transmission is achieved. Up to three services can be provided in one terrestrial channel and partial reception of services contained in the transmission channel can be obtained using a narrow-band receiver that has a bandwidth as low as one OFDM segment.

The system was developed and tested with 6 MHz channels, but it can be scaled to any channel bandwidth with corresponding variations of the data capacity. The net bit rate for one 428.6 kHz segment in a 6 MHz channel ranges 280.85–1787.28 kbps. The data throughput for a 5.57 MHz DTV channel ranges 3.65–23.23 Mbps [3, 5].

2.3. SBTVD-T. Brazil has chosen ISDB-T modulation for its Digital TV format, naming it SBTVD-T (*Sistema Brasileiro de Televisão Digital-Terrestre*).

This new standard has been developed by an association including Brazilian government, Brazilian universities and communication companies.

Basically, SBTVD-T differs from ISDB-T in that it uses the MPEG-4 Part 10 (H.264) video codec instead of ISDB-T's MPEG-2 [6].

As MPEG-4 video demands greater processing power, hardware designed for digital reception in Brazil has to include chips that are usually more expensive than those used in Japanese receivers, thus making compatibility between the two standards only through substantial software modification.

3. Interactivity Standards

In DTV broadcasting, a new interesting feature is the interactivity, that is the chance to deliver multimedia applications interacting with the user and giving a consistent added value to DTV services.

To implement interactivity in a correct way, several standards are available, and in this article will be given a description of the ones that are designed for the DTTB systems introduced in the previous section. Those interactivity standards involve the middleware that have been modified in the BEACON project in order to implement the interoperability system architecture.

3.1. MHP. The DVB-MHP (*Multimedia Home Platform*) stack defines a transport system, an execution environment and a set of API for the developer to provide a platform independent interface between applications from different providers and the manufacturer-specific hardware and software implementation. It enables any digital content provider to address all types of terminals ranging from low-end to high-end set-top boxes, integrated digital TV sets, or multimedia PCs.

Figure 1 shows a simple view of the architecture of an MHP receiver.

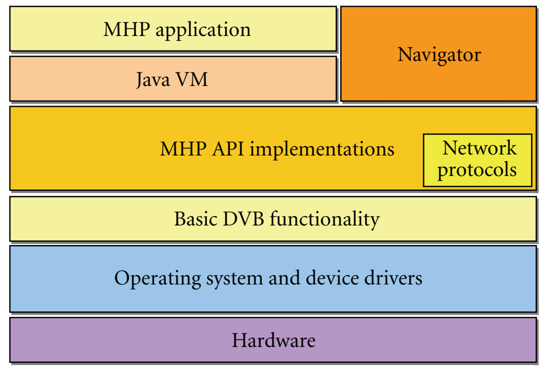


Figure 1: MHP Architecture.

Implementations of the various Java APIs included in the MHP specification need an implementation of basic DVB functionality to build upon.

Implementations of the network protocols are also needed which may not be part of a basic DVB receiver implementation. On top of these is the Java virtual machine that runs applications. The receiver manufacturer will include a resident application ("Navigator") providing basic TV functionality. This may fit above the MHP APIs and be implemented in Java or may fit alongside the MHP APIs and be implemented in some other technology.

The DVB system introduced the concept of profiles as a support to the implementation of standards. Every profile makes reference to a specific application area and defines the requirements of a *Set Top Box* (STB) necessary for its support. Three MHP profiles presently exist.

(1) *Enhanced Broadcast Profile.* It is defined in the specifications MHP 1.0 (BS 201 812). This profile requires a Set Top Box with no or limited capacity of management of the return channel.

(2) *Interactive TV Profile.* It is defined in the specifications MHP 1.0 [7]. It allows the use of the return channel (PSTN, ADSL, GPRS, Ethernet, etc.) for the implementation of interactive applications. This profile also supports the downloading of MHP applications through the return channel (only from the version 1.1), while in the previous profiles this is possible only through the broadcasting channel.

(3) *Internet Access Profile.* It is defined in the specifications MHP 1.1 [7]. This profile requires a more complex Set Top Box in terms of memory and computational power and allows complete interactivity and access to Internet content.

Actually, interactivity currently implemented refers to the first two profiles of the MHP specifications: the enhanced broadcasting profile and the interactive broadcast profile, which adds to the first profile the interaction with a back-end Service Centre through a bidirectional IP return channel. It allows the personalization of service content through the features introduced by the second MHP profile that guarantees the chance to customize service contents for each user through the return channel by [8]

(1) supporting HTTP protocol to get different types of data (text, images, and audio clip);

(2) supporting HTTPS protocol to grant security to the exchange of reserved personal data;

(3) using a Smart Card to support the storage of user related data and allow policies of strong authentication.

A full customisation should be possible with the exploitation of the MHP 1.1 specifications that will allow the download of whole applications (layout, programming logic, and contents) through the return channel with a huge saving of resources on the broadcast channel, but the first release of Set Top Boxes implementing MHP 1.1 has still to come in EU countries.

3.2. GEM. To allow the use of MHP in non-DVB networks, the GEM specification has been created. It defines a subset of MHP which removes the transmission-related elements of the MHP specification but retains the application API's, thus allowing broad content compatibility across a range of new delivery platforms developments.

The following platforms are defined, based on/extending GEM:

(1) *Multimedia Home Platform* (MHP), the open, multi-platform middleware specification developed by the DVB project;

(2) *OpenCable Application Platform* (OCAP), which is an ITV middleware software layer for US cable;

(3) ARIB B.23 specification from Japan's ARIB (*Association of Radio Industries and Businesses*);

(4) *Advanced Common Application Platform* (ACAP), the North American ATSC middleware standard for terrestrial broadcast;

(5) *BD-J* the Java-platform for the Blu-ray disc;

(6) *Ginga-J*, one of the middleware solution in the Brazilian framework.

Figure 2 outlines the concept of GEM [1].

As all these platforms are based on the common GEM-core, it is possible to write Java-application with interoperable operation on all these systems. So, interactive services developed following GEM specifications can be utilized in European, North-American, Japanese, and Brazilian TV networks.

3.3. ARIB. ARIB is the main Japanese association of industries and business operating for the development of Japanese DTV. It conducts research and development, establishes standards, provides consultation services for radio spectrum coordination, also cooperating with other foreign international organizations, and provids frequency change support services for the digital terrestrial television broadcasting in Japan. These activities are conducted in cooperation by telecommunication operators, broadcasters, radio equipment manufacturers, and related organizations [9].

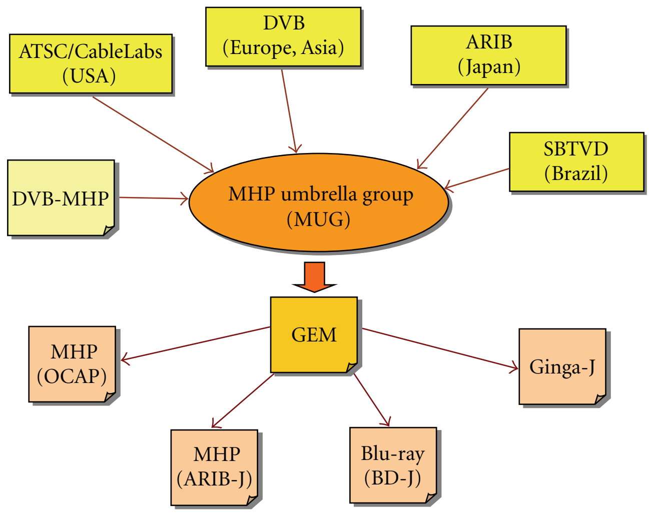


Figure 2: GEM Concept.

As regards the interactivity in Japanese DTV, ARIB has published *ARIB B23* [10]. The Standard comprises two parts, one concerning monomedia coding systems and another concerning application execution engine platforms. The standard embodies a system that is based on the MHP method of DVB specifications and GEM, with additional provision of the necessary implementations to work in the ISDB.

3.4. Ginga. Ginga is the middleware specification for Brazilian Digital TV System. This specification is made up by a set of standardized technologies and Brazilian innovations that make it the most advanced middleware specification and the best solution for the Brazilian requirements.

The Ginga Architecture can be divided into two major modules: Common Core and Specific Service.

The last one is also divided into two main integrated subsystems, which allow the development of applications following two different programming paradigms (as shown in Figure 3).

Depending on the required functionalities of an application project, one paradigm will be more suitable than the other one. Those subsystems are called *Ginga-J* (for Java applications) and *Ginga-NCL* (for declarative NCL applications).

Ginga-J is a logical subsystem of the Ginga System that processes Xlet object content. A key component of the procedural application environment is the procedural content execution engine, composed by a Java Virtual Machine.

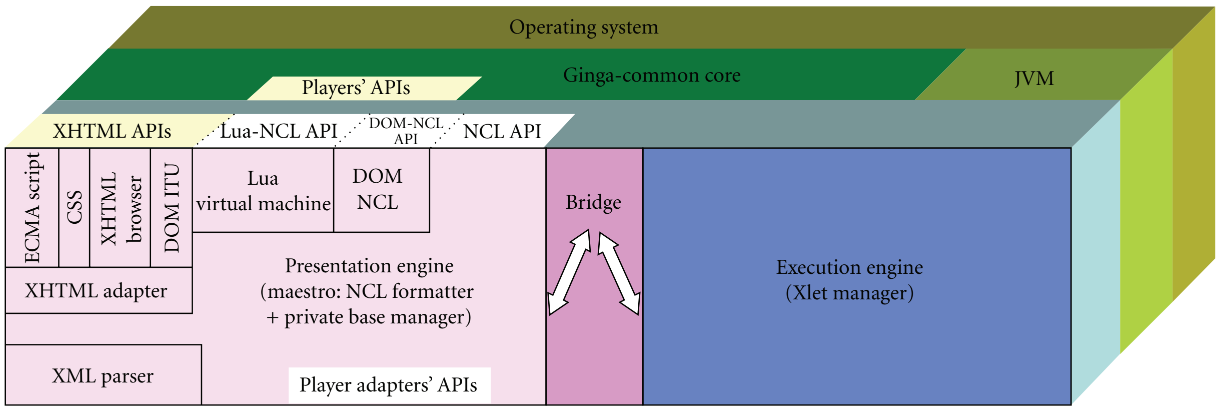


Figure 3: Ginga Architecture.

Ginga-NCL (NCL stands for *Nested Context Language*) is a logical subsystem of the Ginga System that processes NCL documents. A key component of Ginga-NCL is the declarative content decoding engine (NCL formatter named *Maestro*). Other important modules are the XHTML-based user agent, which includes a stylesheet (CSS) and ECMAScript interpreters, and the Lua engine, which is responsible for interpreting Lua scripts.

In fact, beside NCL declarative language, Brazilian DTV Forum has accepted Lua language as the scripting language for the platform, and it is the scripting language used to implement procedural objects embedded in NCL documents.

At this moment only Ginga-NCL has been accepted by the Brazilian Forum for Digital TV, and it is the only middleware mandatory for interactive Brazilian TV.

Ginga-NCL and Ginga-J are built over the services offered by the Ginga Common-Core module, whose organization is illustrated in Figure 4.

Common content decoders serve both procedural and declarative application needs for the decoding and presentation of common content types. The Ginga Common Core is composed of common content decoders and procedures to obtain contents, transported in MPEG-2 Transport Streams or via the return channel. The Ginga Common Core will also support the conceptual display model specified for the ISDTV-T [11].

The architecture and facilities of Ginga are intended to apply to broadcast systems and receivers for terrestrial broadcast. However, the same architecture and facilities may be applied to other transport systems (such as satellite, cable TV, etc.).

In general, Ginga is unaware of any native applications that also may choose to share the graphics plane. These include but are not limited to closed captioning, conditional access system messages, receiver menus, and native program guides.

3.5. NCL-Lua. A detailed explanation of NCL declarative language and then of Lua scripting language follows.

Unlike HTML or XHTML, NCL has a stricter separation between content and structure and it provides noninvasive control of presentation linking and layout. As such, NCL does not define any media itself. Instead, it defines the glue that holds media together in multimedia presentations.

NCL has been specified in a modular way since version 2.0, allowing the combination of its modules in language profiles. Each profile can group a subset of NCL modules, allowing the creation of languages according to the users' needs. Moreover, NCL modules and profiles can be combined with other language modules, allowing the incorporation of NCL features into those languages and vice-versa.

The current version of Ginga-NCL is 3.0, the version selected for the development of our work.

The NCL document only defines how media objects are structured and related in time and space. As a glue language, it does not restrict or prescribe the mediaobject content types. In this sense, we can have image objects, video objects, audio objects, text objects, execution objects (e.g., Xlet, Lua, etc.), and so forth as NCL media objects. Which media objects are supported depends on the media players that are embedded in the NCL formatter. In the Brazilian DTV system, one of these players is the MPEG-4 decoder/player, implemented in hardware in the DTV receptor. In this way, note that the main MPEG-4 video and audio is treated like all other media objects that can be related using NCL.

Another NCL media object required in ISDTV-T is the HTML-based media object. Therefore, NCL does not substitute but embeds HTML-based documents (or objects).

As with other media objects, what HTML-based language will support in an NCL formatter is an implementation choice and, therefore, will depend on which HTML browser will act as a media player embedded in the NCL formatter.

Although an XHTML-based browser must be supported, the use of XHTML elements to define relationships (including XHTML links and ECMAScripts) should be dissuaded when authoring NCL documents. Structure-based authoring should be emphasized.

During the exhibition of media-object contents, several events are generated. Examples of events are the presentation

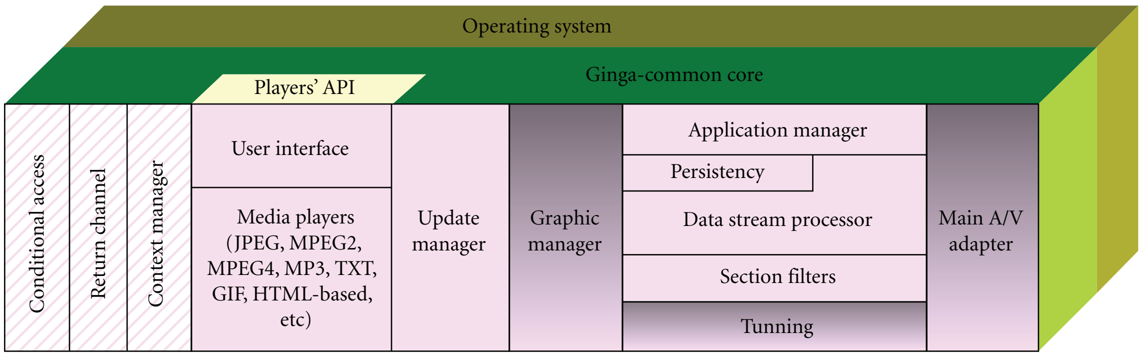


FIGURE 4: Ginga common core module.

of marked segments of a media-object content, the selection of a marked content segment, and so forth. Events may generate actions on other media objects, like to start, pause, or stop their presentations. Hence, events must be reported by media players to the NCL formatter that, in its turn, can generate actions to be applied to these or other players. Ginga-NCL defines an adapter API [12] to standardize the interface between the Ginga-NCL formatter and each specific player. Third part players, including XHTML-based browsers, usually need an adapter module to realize their integration.

Finally, for live editing, Ginga-NCL has also defined NCL stream events (DSM-CC events) in order to support live generated events in stream media, in particular the main program stream video. Again, the use of XHTML elements to define relationships (stream event elements in this case) should be dissuaded in authoring NCL documents emphasizing structure-based authoring.

Lua is the NCL Script Language, and it provides a library that offers a set of functions to support NCL's editing commands. All functions can receive a time reference as an optional parameter that can be used to specify the exact moment when the editing command should be executed. This parameter has the same scheduling function of the time reference present in stream events.

If the time-reference optional parameter is not provided in the function call, the editing command should be executed immediately. When provided, it defines the amount of time, in seconds, for how long the command must be postponed. However, this parameter can also specify the exact moment to execute the command in terms of absolute values.

NCL-Lua also provides an extended library that offers a set of functions to start, stop, pause or resume an interface defined in NCL for NCL-Lua media object executions. These function results can be used as conditions, evaluated by the NCL formatter, to trigger actions on other NCL objects of the same document. Besides commanding event state transitions, a procedural NCL-Lua code can also register itself as a listener of state transitions of any NCL event.

Depending on the middleware configuration, it is possible to have access in Lua to the same API provided by the Ginga-J, the Ginga procedural environment, in order to have access to some set-top box resources and Ginga facilities [13].

Finally, Lua has also a very particular set of characteristics that makes it unique in a DTV domain:

(1) it has a simple procedural syntax with powerful data description constructs;

(2) it is dynamically typed;

(3) it has automatic memory management with garbage collection, making it ideal for configuration, scripting, and rapid prototyping;

(4) it has also other features that make it a good option for embedded and soft real-time systems, as is the case of DTV systems: incremental garbage collection, weak references, and coroutines;

(5) it is interpreted from bytecodes;

(6) it is an "embeddable" language: the interpreter is a pure ANSI C library;

(7) it is small, and as consequence, easy to install, easy to adapt, and portable;

(8) it is efficient (faster than Perl and Python; much faster than JavaScript).

All these features made Lua very well accepted as a scripting language in the DTV domain and a key feature for the development of our system.

3.6. Brazilian Standardization. In this section some important arguments about the Brazilian standardization choices regarding the DTV interactive part are discussed, because those choices have an important role on the determination of the effective availability of the interactive standard that better fits system needs.

At the time of writing, for the implementation of interactive implementation, four alternatives have been taken into consideration for the Brazilian Forum for Digital TV.

(1) *Ginga NCL and Lua*. At the moment they are the only languages accepted by the Brazilian Forum for Digital TV, and this middleware is mandatory for interactive Brazilian TV.

(2) *Ginga-J* based on GEM. it is still not clear because the Brazilian Forum for Digital TV has concerns about MHP via licensing;

(3) *Java DTV.* is a specification developed recently by Sun Microsystems for interactive television, but at the moment it is only a specification, there is no set top box with this implementation (maybe in 1 or 2 years there will be set top boxes in the market but it is not approved for the Brazilian Forum and actually it remains only a specification);

(4) *Web browser based* on webkit and javascript. This is not a standard solution and it is not clear about this implementation due to hardware requirements.

It is clear that the only option for the BEACON project is Ginga NCL and Lua, because it has been accepted for the Brazilian Forum for Digital TV, and it is the only implementation available in real set top boxes from Brazil.

In fact, at the beginning of the project, it seemed that the set top boxes market was going in the Ginga-J direction (implementing Ginga-J as an exact map of GEM permits all interactive application GEM compliant to be interoperable with ISDTV and other standards, as noted before).

Technically it seems to be the optimum, but there are some legal problems for implementing GEM into Brazilian set top boxes: in Europe every MHP compliant STB producer must pay $1,75 per device for royalty on the MHP standard. GEM is a derived standard of MHP, but Brazilian interactive STBs producers do not want to pay this royalty and decided to leave the GEM implementation of Ginga-J.

Moreover, only Ginga-NCL implementation of Ginga Standard is mandatory at this moment, so the BEACON consortium decides to design the Ginga-NCL version of the player tool (developed in MHP for the European side). The Ginga-NCL choice guarantees the Consortium a real implementation of this standard.

4. Beacon Project

BEACON (*Brazilian-European Consortium for DTT Services*) is a specific targeted research project on Digital Terrestrial Television.

It will develop innovative t-learning pilot services related to social inclusion in the State of Sao Paulo (Brazil) on the basis of pioneering research on interoperability between the European (DVB) and the Brazilian (SBTVD) Digital Terrestrial Television standards and the definition of a pedagogically sound methodology for distance learning through digital television. Ultimately the project will result in the establishment of a Brazilian-European Consortium that will manage the exploitation of the assets and the services implemented by the project.

As regards the technological state of the art inside of which the BEACON project is supplying another step of innovation, it is important to understand that if several other systems are already developed to deliver t-learning services (with more or less interactivity) around the world, our system is the first one facing the challenge of t-learning interoperability between the well known DVB-MHP European stack and the Brazilian one.

In fact, in Brazil SBTVD-T is now defined and implemented, but regarding the interactivity part there are some important issues that, at the time of writing, still have to be solved and/or defined (as described in Section 3.6).

So, the innovation opportunity given by this work includes the following:

(1) to analyze the technologies available to find one that fits the needs of Brazilian platform integration;

(2) to develop an application running on Brazilian STBs that can be easily interfaced with the t-learning system;

(3) to integrate the European t-learning application and services involved in the BEACON project with the one that will be developed in Brazil;

(4) to create an important starting point for the development of several interactive applications delivered by SBTVD-T defining specific guidelines.

In this way, the project is addressed to build a concrete interaction between the European and Brazilian DTT interactive applications.

4.1. Objectives. BEACON aims to contribute for going a step ahead of the state of the art of Digital Terrestrial Television technology and services pursuing the following main objectives:

(1) to perform research on innovative and interoperable (between the European DVB and the Brazilian SBTVD standards) DTT applications and services addressed to the t-learning domain notably customized for the Brazilian specific needs related with social inclusion issues;

(2) to provide feedback data analysis of services pilot run to Brazilian Public Administrations in order to contribute to the relevant policy making processes regarding the DTT standard adopted and to its implementation;

(3) to establish a Brazilian-European Consortium which will manage the exploitation of the assets and the services implemented by the project activities.

In order to achieve the above stated objectives, BEACON will jointly consider technological solutions, pedagogical issues and sustainable models. The project's activities output the following:

(1) research on innovative DTT-based e-learning methodology and on the relevant technologies performed;

(2) t-learning definition availability;

(3) innovative interoperable (DVB, MHP-SBTVD, Ginga) t-learning DTT application development;

From the technological point of view, BEACON envisages to achieve the previously mentioned objectives by research activities that will lead to:

(1) develop a micro-XML browser, aiming to foster convergence towards IP service infrastructures, to enable innovative t-learning services fruition and to enhance their user-friendliness (usability);

(2) develop an enhanced interoperable service framework able to broadcast applications on different DTV platforms (DVB-SBTVD) and to set up a Service Center in charge of services delivery;

(3) identify and promote standard solutions for the development of interoperable t-learning applications —that is, methodological and technological guidelines to design, implement and test t-learning applications on a DTT platform;

(4) identify and integrate authoring tools to grant full compliancy to both t-learning and DTV standards.

Furthermore, BEACON will deal with pedagogical issues related to t-learning by mean of research activities aimed to:

(1) get a better understanding on how people learn in their home environments and how they may relate to learning through TV compared to other means, that is, when people have access and can take advantage of learning opportunities in their home;

(2) investigate sociological dynamics that operate in the home, how these relate to television and what impact this may make on creating learning opportunities in the home, that is, social barriers to preventing access to learning opportunities in the home;

(3) evaluate how informal learning, in the home, could draw people into formalised learning;

(4) define what aspects of interactivity are needed for home learning context and how it can motivate learners;

(5) investigate to what extent the current e-learning didactic strategies are heritable and achievable by mean of t-learning applications and services taking into account the innovative technological results of the project;

(6) deliver interoperable t-learning services aimed at fostering social inclusion in Brazil.

Those objectives are achieved addressing a particular context. In fact, in Brazil the academic reality shows that public universities are usually better than private ones. For a long time, governments were the only entities willing to invest the massive amounts of money required to create and maintain good universities, thus the majority of Brazilian scientific researchers are found in the public universities (very little research is funded by private colleges).

Also, public universities have no fees, differently from privates universities.

Because the number of candidates is usually superior to the number of vacancies offered, vestibular is adopted by practically all public universities. It is a test open to all students who have completed high school and the students with the best scores qualify for enrollment.

Obviously, a good preparation and a good training program can have a key role in the enrollment of students and in the improvement of the quality of their expectations (and then lives) given by studying in a good university.

In this scenario, the specific t-learning services delivered with the BEACON platform are focused on how to cover the social-gap in this country fostering a profitable training to access the university to all those peoples (students) that can not be reached by e-learning courses or other training opportunities.

Such a service has been chosen to deal with the inclusion matter described above, but every kind of interactive services (t-learning courses and so on) can be delivered through this platform between Europe and Brazil.

4.2. Methodologies. In this paragraph the methodologies used to realize the system will be discussed.

As for the learning scenario to be implemented in the BEACON test pilot, it is the result of an iterative design process based on the *User Centred Design* (UCD) approach. UCD approach will be used to define user needs and requirements and to analyse user activities. This methodology allows to design and to develop an artifact that effectively satisfy the user and supports him during the activity.

The UCD proposes an iterative cycle of some steps of the process that are repeated many times in order to refine the requirements before the implementation.

The methodology applied to the BEACON service model design is the result of an iterative process set of the following steps:

(1) User requirements analysis: user needs will be analyzed, along with service requirements (features, privacy, usability, ...) coming from the e-Health Service Provider;

(2) Feasibility studies, aiming to identify the most valuable state of the art technology to develop the service, for achieving the project goals;

(3) Definition of service features and specifications, aiming to identify the system architecture and related hardware/software modules, in order to guarantee service expected functionalities;

(4) DTT application testing in an emulated environment: the DTT application is implemented through software authoring tools in emulated environment,

which would be properly tested and debugged before deployment and broadcasting;

(5) Application deployment and testing with final users in a broadcasting environment: the DTT application is uploaded and managed by the Object Carousel Generator, multiplexed with audio/video contents, COFDM modulated and broadcasted to final users. The application is downloaded on the Set Top Box and users are involved in an iterative process of evaluation-redesign-testing (phases 3, 4 and 5) on the basis of "User Centred Design" approach.

5. DTT System Design

In this section the system architecture is described, with a deeper analysis of the key elements and their roles in the broadcast and presentation phases.

Then, the modification needed to the DVB-MHP-compliant system to become interoperable with SBTVD-T and Ginga NCL-Lua middleware are introduced, describing the involved technologies.

5.1. Architecture. The system architecture refers to the consolidated interactive DTV framework and comprises the following elements, as represented in Figure 5:

(1) Service and Content providers, providing content to be broadcasted on the communication channel;

(2) Broadcaster, in charge of multiplexing, modulating and broadcasting audio, video and applications over the air;

(3) Users, getting the DTT application (Java Xlet) over the air and downloading it on the STB through the interactive remote control;

(4) Service Centre, managing connectivity and user interaction through Internet protocols.

The BEACON proposed architecture, as described in the following sections, will not restrict the inter-working with other service providers and network operators.

Also, the adoption of a new framework, different from the European DVB-MHP platform formerly planned to be used, introduces the needs of facing the interoperability of different solutions as a key objective of the project.

As seen before, in a technological perspective the decision of the Brazilian Government to establish the SBTVD-T standard leads to the adoption of the following subsystems:

(1) *ISDB-T* as broadcasting standard. ISDB-T is maintained by the Japanese organization ARIB and is based on MPEG-2 video and audio coding as well as the transport stream described by the MPEG-2 System standard. ATSC and DVB also adopted the same standard for their transport system. DVB and ISDB also provide for other video compression methods to be used, including JPEG and MPEG-4, although JPEG is only a required part of the MPEG standard. As for the modulation, ISDB-T supports COFDM with QPSK/QAM in the VHF and/or UHF band.

(2) *H.264* (MPEG-4 Part 10) as the video and audio coding standard for supporting HDTV (*High definition TV*) 1080i video for fixed TV sets and LDTV (*Low definition TV*) for mobile terminals.

(3) *Ginga* as the middleware standard, based on an execution engine (Ginga-J) integrated with a presentation engine (Ginga-NCL).

As for the broadcasting, the interactive services that have been developed with this technology are only text (flat XML files) and resources (images, audios, ...). During the emission, with the set of interactive services, is transmitted the *tmPlayer* product, a micro browser which is located in the Set Top Box and is able to analyse and parse XML information for converting to MHP code.

Due to the Brazilian choice of SBTVD-T, the whole platform will be designed on the basis of a new architecture model.

In order to realize this solution some necessary adaptations have been identified: Audio/Video encoder: changed to MPEG4-H.264; a data channel server for the Ginga middleware; a multiplexer equipment ISDB-T compliant; a modulator ISDB-T compliant; Set Top Boxes compliant with ISDB-T standard and Brazilian Middleware software and user interfaces (see Figure 6 for European and Brazilian implementations).

The hardware/software modules used in the test bed platform are

(1) an *authoring tool*, with *tmManager* and *tmPlayer* (this two elements will be explained in the following paragraphs);

(2) an *object carousel* generator and content multiplexing using a playout system (tmCarousel, suitable for both DVB and ISDB);

(3) a *DVB-T modulator* and an *ISDB-T modulator*;

(4) an *MHP compliant STB* and *Ginga compliant STB*;

(5) a connection to PSTN as a return channel for interactive applications.

5.2. T-Learning Service. As regards t-learning, it is a subset of e-learning but the relevant access through a home-based TV could hugely enhance the learning opportunities in a way that Internet-based e-learning cannot currently do.

For BEACON, a t-learning service is developed using a *Learning Management System* (LMS).

This is the system that will be used in the project and will be correctly adapted to work on the system upgraded technology for interoperability on MHP and Ginga (NCL-Lua) standards.

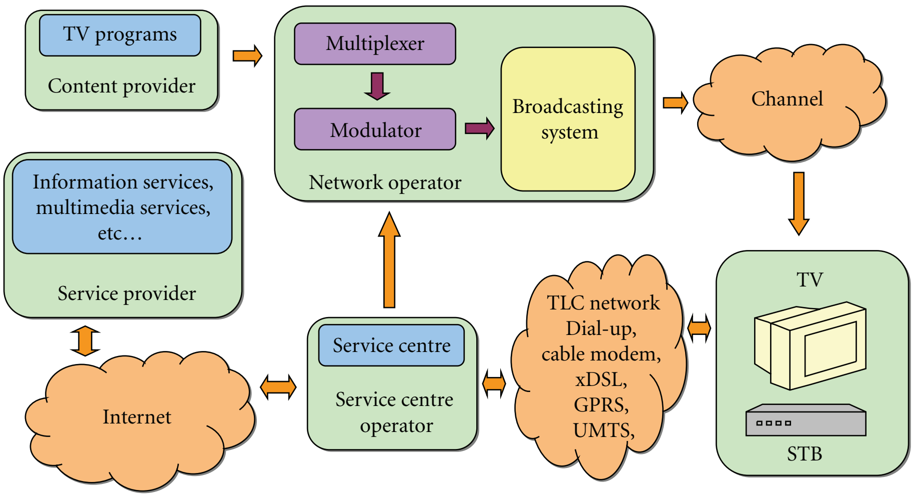


Figure 5: Interactive DTV Framework.

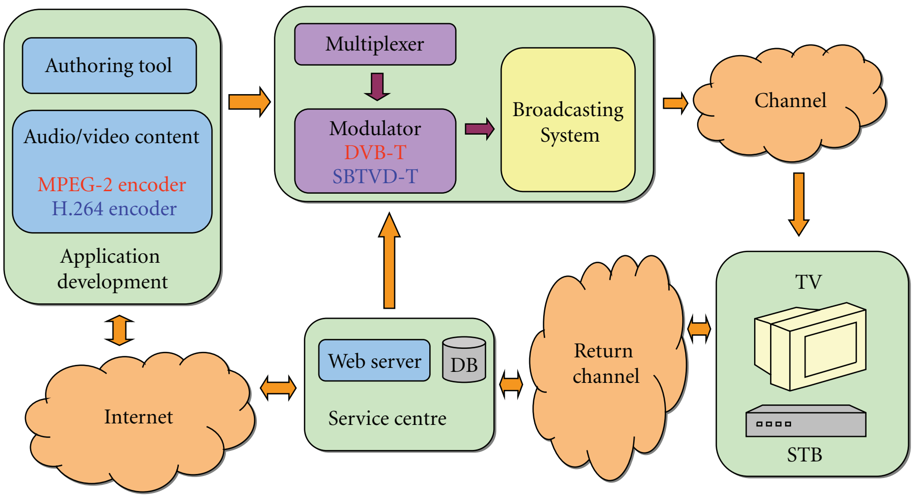


Figure 6: Interactive DTV Implementations in the Europe and Brazil.

An LMS is a web application for managing computer-mediated distance learning (also known as *e-learning*) via Internet.

An evolution of offline *Computer-Based Training* (CBT) is called *Web Based Training* (WBT), delivered by internet. Those courses need a more structured way of delivering e-learning for a more detailed tracking in real time of the user interactions. This led to the creation of LMS software.

In the DTT interactive domain, a t-learning session requires the following functionalities:

(1) *authentication*, to start the user session on the server;

(2) *authorization*, to receive a list of resources available to the user;

(3) *delivery*, to receive the course's content;

(4) *tracking*, to track the duration and the results of the user session;

(5) *communication*, to exchange short mail messages with the tutor or the teachers.

Apart from the delivery, that can take place using the broadcasted channel (i.e., the course contents can be included in the carousel), all the other functionalities require a connection to a service centre using the return channel.

The service centre role can be played by an LMS customized with APIs created to exchange data with the t-learning application. Then can take place the authoring process that can be streamlined according to the structure and layout of the courses, so that an automatic conversion of the content towards the format required for the DTT application could take place.

An authoring tool consists of

(1) a *text editor* with an automatic check of the maximum number of characters allowed per page;

(2) an *image selector* to choose the illustration for the page and check its format and dimensions;

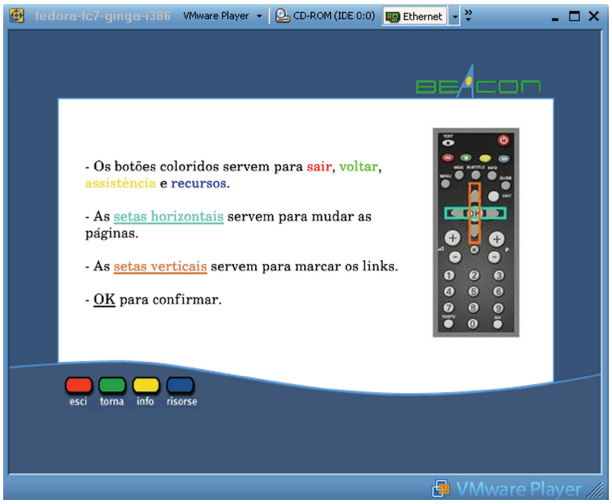


Figure 7: t-learning service implemented.

(3) a *property selector* to specify the page type: (i.e., tutorial, test, glossary, etc.);

(4) an *automatic ID generator* for each page or entry;

(5) a *function* to link an active word to an ID.

The output will be an XML file and a folder of images. This output can be reopened with the authoring tool for further editing, copied to a *SCORM* WBT [14] engine to test it in a web browser environment, or finally passed to the transformation engine to be converted to the specific XML format required by the DTT application.

In Figure 7 there is a screenshot from the t-learning service developed in the BEACON project.

5.3. System Technologies. In this section, an analysis of the existing technologies and upgrades needed to match the specifications given by the new libraries introduced (MHP and Ginga NCL-Lua) is shown, focusing on the architecture that will be the base for the provisioning of t-learning courses.

A system called *tmBroadcast* will be used. It is developed by tmira solutions, a Spanish partner of the BEACON project [15].

tmBroadcast is an integral system for management and broadcasting of *Electronic Program Guides* (EPG) and interactive MHP applications in DVB networks. It is developed according to DVB and World Wide Web Consortium standards [16–18].

The main features of the system are the following:

(1) *user management*: define users, assigning roles and services per user;

(2) *broadcast configuration*: all the options for configuring network parameters and modulation, maintenance services, generation of DVB/SI tables, bitrates configuration and selection of several outputs (ASI, RTP, UDP, modulator);

(3) *EPG*: the system has a complete tool fully integrated for management and emission of the Electronic Programming Guide and this part has all the functionality of the product *tmEPG*;

(4) *maintenance of applications*: management of interactive applications like creation, modifications, updates, editing contents, configuration of the return channel;

(5) *scheduling*: scheduling of applications at any time and for each service (Firing stream events);

(6) *system configuration*: Setting the parameters of the operating system (time, network);

(7) *system backup*: the system has the ability to perform backups with the configuration of the system, which may also include applications, EPG and log files;

(8) *automatic updates tasks*: includes the possibility of defining tasks for updating application contents in an automated manner according to a defined schedule.

tmBroadcast is composed of three subsystems, as shown in Figure 8

(1) *tmCarousel* is a carousel playout of interactive TV applications and DVB signalling. It works in DVB for cable, terrestrial and satellite networks.

(2) *tmManager* is a web platform that provides a web user interface and a set of tools to extract and transform contents in an automatic way, allowing to adapt any website or format for digital TV.

(3) *tmPlayer* is an application broadcasted that runs on STB and allows the viewer to browse the contents of an application. The browser is based in a custom XML language independent of the platform.

These three subsystems are the core of the project implementation and need a more detailed explanation to point out the key features of each component, in order to define the upgrades needed to integrate the t-learning services in Europe and Brazil.

5.3.1. tmCarousel. tmCarousel is a client server system that includes the following components:

(1) *Configuration Manager System*, in charge of receiving, processing and distributing system configuration to other system modules. It assures that configuration is always consistent. This module includes a timer that allows the insertion and removal of applications according to the broadcast programming.

(2) *DVB Service Information* (PSI/SI) *generator*. It generates DVB tables according to specific configurations. Tables generated are: PAT, PMT, NIT, SDT, DTT, TOT, AIT, EIT P/F, EIT S [16].

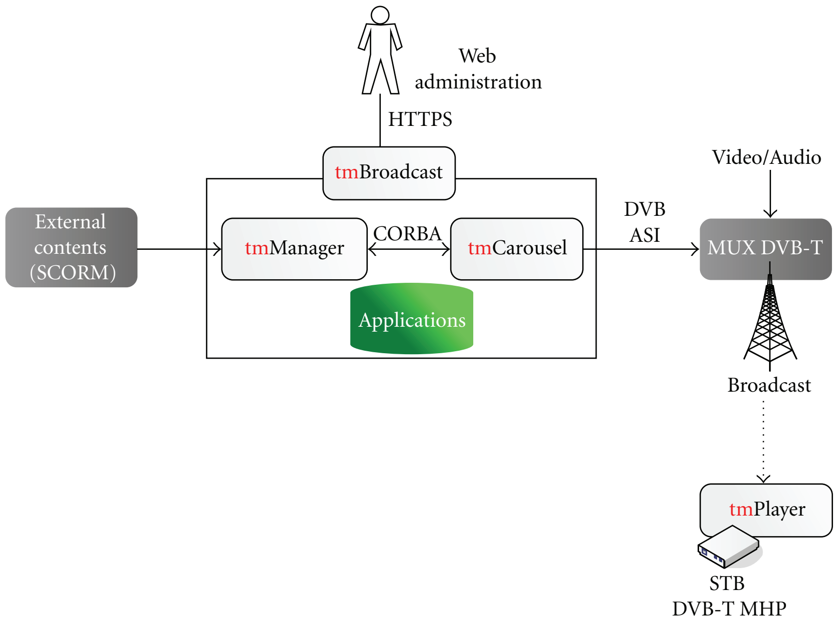


Figure 8: tmBroadcast and its subsystems

(3) *Data encoder*, responsible for encoding the interactive applications according to the standard identified for this purpose (DSM-CC) [19] and sending them to the multiplexer to be inserted into the output signal.

(4) *Software Multiplexer*, responsible for generating the output signal with DVB tables and applications. The output is transmitted through DVB-ASI, UDP or RTP.

tmCarousel works for DVB in cable, terrestrial and satellite (NIT configuration), and its development is based on the following standards:

(1) MPEG-2: TS construction (ISO/IEC DIS 13818-1), DSM-CC: (ISO/IEC IS 13818-6) [17];

(2) DVB Service Information (ETSI EN 300 468, ETSI ETR 211, ETSI ETR 162) [20];

(3) Data broadcast (ETSI EN 301 192, TR 101 202), MHP (ETSI TS 101 812 ver 1.3.1 (MHP 1.0.3), ETSI TS 101 812 ver. 1.2.1 (MHP 1.0.2), ETSI TS 102 812 ver. 1.2.1 (MHP 1.1.X), DVB BlueBook A068r3 (MHP 1.1.3), A107: DVB MHP Specification 1.2 (MHP 1.2)) [17, 20].

The system is highly reliable and robust. Several ASI or UDP Unicast/Multicast outputs can be set up.

5.3.2. tmManager. tmManager is the web platform for the secure control and management of tmBroadcast. It includes the definition of users and roles for the system. It also incorporates the functionality associated with the maintenance, scheduling and broadcasting of MHP applications, management and emission of EPG service and management of automatic updating tasks.

tmManager has the functionality associated with the multiplex operator that allows a complete network configuration, including services, components, EPG, interactive applications, bitrate, output type (DVB-ASI or IP), operating system time, backups, network interfaces, and so forth.

tmManager interface is multilingual, and it is currently available in English and Spanish, but can be easily translated to any other language like Portuguese.

tmManager is based on the following standards:

(1) DVB: Service Information (ETSI EN 300 468) [20];

(2) World Wide Web Consortium: XML, XSL, XPL, XSD, XPATH, XQUERY, XFORMS, XHTML, CSS [18].

One of the main advantages of tmManager is the ability to extract information from external data sources (web, database, content managers, RSS, Web services, etc.), and transform it into tmPlayer XML format to represent it in an interactive MHP STB.

tmManager has a core set of libraries to extract and transform contents to a desired format in an automatic way. This functionality is based on standard languages defined by World Wide Web Consortium [18]:

(1) XML (*eXtensible Markup Language*): for definition of contents;

(2) XSLT (*XML Stylesheet Transformations*): data transformation between formats HTML, CSV, XLS, XML, PDF, XML;

(3) XPL (*XML Processing Language*): XML processing language;

(4) XPath (*XML Path Language*): language to access data into an XML (nodes and attributes);

(5) XQuery (*XML Query Language*): XML language to query databases;

(6) XForms: XML language to define web forms and interfaces.

tmManager is able to transform any content in almost any format. In the system architecture developed for the BEACON project, tmManager will be in charge of transforming the course contents defined in SCORM language in tmPlayer XML for DVB and tmPlayer NCL for SBTVD.

5.3.3. tmPlayer. Is the technology that enables adoption of a Web, Database, CMS, and so forth, to bring their content directly to the television. Therefore all services available on Internet such as sending e-mail, online banking, games, information, and so forth, can be automatically reused for interactive TV.

tmPlayer is an XML-based micro-browser that is broadcasted permanently and is able to represent any type of interactive service starting from just a XML files. tmPlayer allows the definition, representation and navigation into a wide range of interactive services through XML files. The language of tmPlayer has been used in the MHP platform, but it can be easily adapted to other platforms such as GEM, IPTV, OpenTV, Ginga, and so forth.

Among its functionalities tmPlayer includes sending email, plugins, return channel access, use of smartcards. It also provides a development environment that enables the implementation of MHP interactive services, with no need for any skill in Java programming, or MHP itself. It only requires some expertise in XML technologies, much more widespread in the Internet community.

5.4. System Upgrades. The tmBroadcast system had been originally designed for DVB broadcasting in cable, terrestrial and satellite networks, but all technology involved in this system has been developed basing on standards and in modular components in order to be flexible and adaptable. As a result, it can be expanded to adopt any other standards of digital TV.

In the architecture developed for the BEACON purposes, tmBroadcast must be upgraded to be fully compatible with the Brazilian standard SBTVD. This upgrade is planned for each of the tmBroadcast components and it is described below.

First of all we have defined the sequence of processes to bring the course contents to the students through interactive Digital TV with tmBroadcast. In Figures 9 and 10 the steps of the whole process are shown, and steps that must be upgraded for SBTVD have been highlighted in orange colour. The other steps are the same for DVB and SBTVD and can be enumerated as:

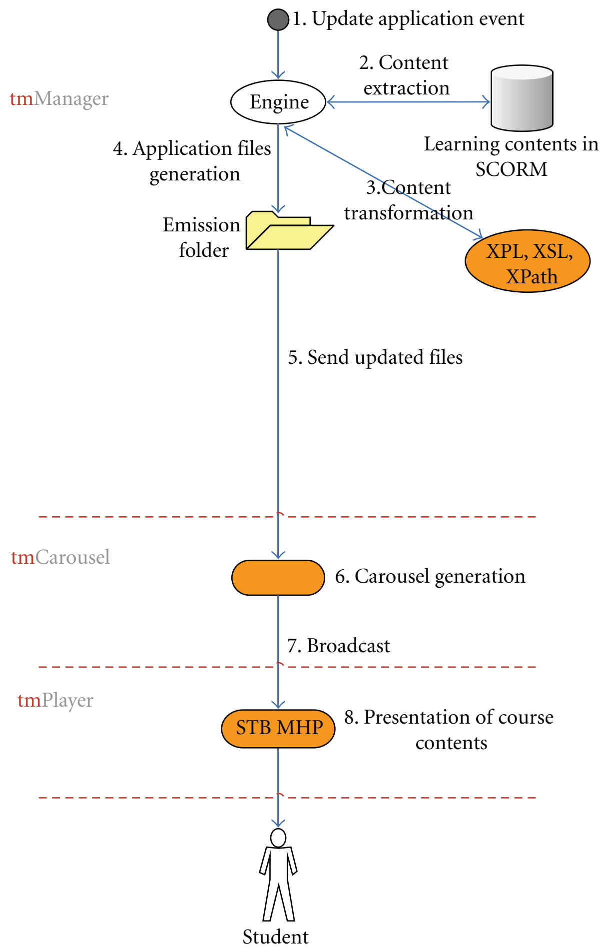


FIGURE 9: Upgrading Steps.

(1) *Update application event*: generated automatically or manually by the administrator user, this event produces the generation or update of an application, in our case a t-learning course;

(2) *Content extraction*: the contents in SCORM format are extracted from the Learning Management System;

(3) *Content transformation*: the contents are transformed using standard languages (XPL, XSL, XPath, etc.). Note that this process is different for DVB and SBTVD, because for the first one the result will be in tmPlayer (MHP) XML format and for the second one the result will be in tmPlayer NCL/Lua format;

(4) *Application files generation*: XML and image files are generated into the emission folder for each application or course;

(5) *Send updated files*: the new files are passed to the tmCarousel system;

(6) *Carousel generation*: tmCarousel generates the carousel and the signalling according to DVB or SBTVD standards for the application generated;

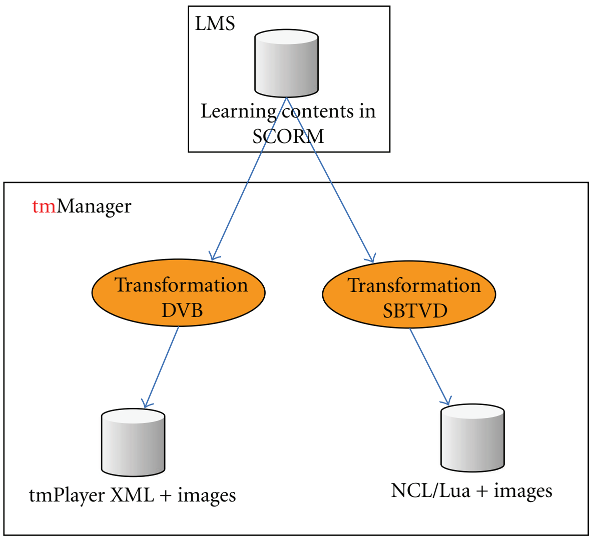


Figure 10: DVB and SBTVD Transformation.

(7) *Broadcast*: tmCarousel outputs through an ASI/IP PCI card and the output is connected to the multiplexer;

(8) *Presentation of course contents*: contents are presented to the student at home.

tmPlayer is broadcasted with the contents, and it is the application in charge of representing the contents into the student TV. In DVB tmPlayer is based on Java MHP whereas in SBTVD tmPlayer is defined by the Lua scripting language and the presentation is defined in NCL (XML format).

In the following paragraphs the steps to be accomplished to bring the course contents to the students will be introduced. To perform all these operations both in DVB and SBTVD, several upgrades have been necessary to tmCarousel, tmManager and tmPlayer.

5.4.1. tmCarousel Upgrade. DVB and SBTVD present several common features. As it can be seen, every table implemented in tmCarousel is included in the SBTVD standard, so tmCarousel is already implementing a subset of SBTVD which will be enough to generate interactive TV. It will not be necessary to make any modification to tmCarousel in order to generate interactivity in a SBTVD network.

The main tasks for tmCarousel update are testing the system with Brazilian equipments, due to the fact that the standard specification usually differs from the hardware and software implementation of the providers. The testing also provides the proper broadcasting parameters on an SBTVD network.

5.4.2. tmManager Upgrade. The tmManager role in the system is to generate the contents in the right format for each Digital TV platform. The input contents will be defined in SCORM format and tmManager will transform it automatically to tmPlayer XML (DVB) and tmPlayer NCL/Lua (SBTVD).

As seen before, tmManager transformations are defined using standard languages. In each case (DVB or SBTVD) the correct transformation is defined and then performed to transform correctly the same course contents defined in SCORM format to the corresponding output format.

The output files in both cases consist of XML files and images. In DVB these files are generated according to tmPlayer XML format, and for SBTVD the generation takes place according to the NCL/Lua format.

In this scenario any content can be easily delivered in European and Brazilian Digital TV stations. Once the appropriate tmManager transformations have been defined, it is possible to generate automatically both contents for both platforms. Since SCORM is a standard to define learning contents, any type of learning content will be interoperable in both platforms.

5.4.3. tmPlayer Upgrade. For the tmPlayer upgrade is necessary to distinguish between tmPlayer XML (for DVB-MHP STBs) and tmPlayer NCL/Lua (for SBTVD STBs).

The tmPlayer XML for DVB-MHP is a commercial product of tmira solutions that is on air in several TV stations. This player has almost all the functionalities required for t-learning applications of the BEACON project. As seen before, tmPlayer based on DVB-MHP is a microbrowser that represents interactive applications defined in XML language.

The mandatory middleware specified by the Sistema Brasileiro de TV Digital is based on the declarative language Ginga-NCL and scripting language Lua.

Moreover, the declarative pages are generated by tmManager from the course contents in SCORM format. There are parts of the applications that cannot be defined using a declarative language and therefore must be implemented using Lua scripting language. Lua is needed to be used in forms, user interactions and return channel management, all these functionalities need to be included in a common tmPlayer NCL/Lua application in order to be reusable in the future.

Briefly, the upgrades that have to be done in the system in order to assure the interoperability between European and Brazilian TV standards will follow the below steps:

(1) To develop a tmPlayer NCL/Lua prototype application with static content;

(2) To define structure and format of course contents (SCORM-BEACON format);

(3) To implement the transformation of SCORM-BEACON format to tmPlayer NCL/Lua, transformation processes defined in XPL and XSL language will be implemented in tmManager;

(4) To test courses created in the Ginga NCL/Lua Emulator;

(5) To test courses created in Ginga NCL/Lua real STBs, in European testbed at tmira solutions premises. To this extent, several Ginga STBs are available to test the applications created for the BEACON project, once they have been loaded using its USB port;

(6) To implement the transformation of SCORM-BEACON format to tmPlayer XML for DVB-T using tmManager transformation engine;

(7) To create real course contents according to the SCORM-BEACON format;

(8) To test courses created in Ginga NCL/Lua STBs and DVB-MHP STBs.

As seen before, the input format of the courses (BEACON-SCORM) is converted automatically to tmPlayer XML format using XSLT transformations. Then, the files generated are broadcast on a DVB-T/SBTVD-T network and tmPlayer is able to load, parse it and show the interactive content to TV students.

5.5. Alternative Approach. In this section, another approach is introduced to achieve interoperability between DVB standard (MHP or GEM) and ISDTV standard (Ginga) interactivity. This might be considered a new way to achieve interoperability, alternative to GEM, when a new implementation of the Brazilian television standard will be ready.

Ginga Ready is a Ginga implementation powered by MOPA Embedded Systems. Over this layer Brazilian partners are implementing an XHTML browser that support also ECMA Script. The idea will be to implement the applications for the tmira solutions player that will act as a software abstraction layer.

This solution will be optimal when implementations of standard MHP1.1 will run in European STBs: MHP1.1, with internet profile, has a full support of XHTML with ECMA scripts. So, any application that runs in a MHP1.1 STB can also run in others STBs that supports a XHTML (with ECMA scripts) browser, just with minimal modifications. Differences may be close to the ones experienced by users visiting a web site with two different web-browsers. Otherwise, web site developers need to optimize their sites to grant valuable navigation in different browsers. The same situation will take place once MHP1.1 will be implemented in European STBs.

Actually, with MHP1.0.3, all applications (included tmPlayer) are java ones. tmPlayer is an application that allows the reading of XML language: different applications running in tmira solutions player use the same player and different XML (an XML interpreter). Every application, for example different BEACON t-learning courses, can be presented by tmira solutions player, achieving interoperability.

Looking at the same thing in a different perspective, in this scenario another interoperability solution has been studied into the BEACON project, but, instead of the previous problem(the failed Brazilian introduction of GEM in Ginga) the core problem is the delay of real implementation on STBs of MHP1.1 in Europe.

Obviously, once future implementation of MHP (1.1 and so on) will be available in Europe, this solution can be developed with success.

As for the BEACON project, this way has been deeply explored and, although it is not the main implementation (focused on Ginga-NCL, as seen before), Brazilian partners of the project have yet implemented some modules of this solution to explore feasibility.

6. Conclusion

This article describes an innovative system architecture to deliver t-learning services related to social inclusion in Brazil, on the basis of pioneering research on interoperability between the European and the Brazilian Digital Terrestrial Television Broadcasting standards. Before of this, an introduction to the interactivity standards is given, focusing on the key differences between different options: an interoperability solution based on GEM implementation (at this moment not available in the Brazilian scenario but analyzed in this paper), the solution implemented in the BEACON project platform (involving Ginga NCL-Lua middleware for the Brazilian realization) and other choices based on the advent of MHP 1.1 in Europe.

This effort is the first important step in the direction of SBTVD interoperability with the other DTTB standards. It creates the basis for a wide range of interactive applications development, designing a common architecture involving MHP and Ginga without neglecting some other possible alternatives, but finding in the NCL-Lua implementation the actual best choice to deliver interactive services in an interoperable mode between European and Brazilian digital television.

7. Future Works

Future enhancements are strongly dependent on political choices such as GEM implementation in Ginga, delays of MHP 1.1 in European STBs and so on. During the design and the realization of BEACON system architecture several technologies and standard have been taken into account, so different ways could be covered if something changes.

Obviously, the solution that has been implemented faced the challenge of the creation of an interactive application that runs t-learning services on the SBTVD-T standard, but it permits interoperability only between Europe and Brazil.

The introduction of GEM in Ginga, instead, would have allowed interoperability with other standards (e.g., Japanese, North-American). As for the system described in this paper, it would be possible to perform the upgrade to a GEM middleware just with the creation of a GEM-compliant tmPlayer and some small modifications, because in our effort the broadcasting issues about tmBroadcast in a Brazilian (SBTVD) scenario have been already solved.

Another interesting evolution of the system regards the integration with HTML browsers. In fact, although it seems a step back from t-learning to e-learning services, it permits the fruition of the services on every HTML-enabled browser by only creating another one tmPlayer version able to translate the XML input in HTML pages (those pages will be adapted opportunely to be browsed easily as interactive DTV contents).

In this way, expanding the range of networks and devices, it will be possible to overcome the borders of the DTV broadcasting model to reach every internet-embedded device using different internet access networks (xDSL, Wi-Fi, WiMAX, UMTS, IMT2000 etc.), or by plugging small media centres (e.g., mini-PCs, Play Station, Wii or X-BOX) to a TV screen.

In this way, by yhe use of the familiar TV remote controller, it will be possible to draw benefits from new devices such as WiiMote and 3D mouse, making the interaction so much easier and intuitive, especially for the elderly and many other "infomarginated" peoples.

In this way it is possible to minimize the digital gap for those people and to increase the effective social inclusion and the digital inclusion.

Abbreviations

ADSL: *Asymmetric Digital Subscriber Line*
AIT: *Application Information Table*
ANSI: *American National Standards Institute*
ASI: *Asynchronous Serial Interface*
CMS: *Content Management System*
COFDM: *Coded Orthogonal Frequency-Division Multiplexing*
CSS: *Cascading Style Sheets*
CSV: *Comma Separated Value*
DQPSK: *Differential Quadrature Phase Shift Keying*
DSM-CC: *Digital Storage Media Command and Control*
DVB/SI: *Digital Video Broadcasting System Information*
ECMA: *European Computer Manufacturers Association*
EIT: *Event Information Table*
EIT P: *Event Information Table Present*
EIT F: *Event Information Table Following*
EIT S: Event *Information Table Scheduled*
ETSI: *European Telecommunications Standards Institute*
GPRS: *General Packet Radio Service*
HTML: *Hyper Text Mark-Up Language*
HTTP: *Hyper Text Transfer Protocol*
HTTPS: *Hyper Text Transport Protocol Secure*
IPTV: *Internet Protocol Television*
ISO/IEC: *International Organization for Standardization/International Electronic Committee*
NIT: *Network Information Table*
OFDM: *Orthogonal Frequency-Division Multiplexing*
PAT: *Program Association Table*
PCI: *Peripheral Component Interconnect*
PDF: *Portable Document Format*
PMT: *Program Map Table*
PSTN: *Public Switched Telephone Network*
QAM: *Quadrature Amplitude Modulation*
QPSK: *Quadrature Phase Shift Keying*
RSS: *Real Simple Syndication*
RTP: *Real-time Transport Protocol*
SCORM: *Shareable Content Object Reference Model*
SDT: *Service Description Table*
TOT: *Time Offset Table*
TS: *Transport Stream*
UDP: *User Datagram Protocol*
UHF: *Ultra High Frequency*
USB: *Universal Serial Bus*
VHF: *Very High Frequency*
XHTML: *eXtensible HyperText Markup Language*
XPL: *XML Pipeline Language*
XSD: *XML Schema Definition*
XSL: *eXtensible Stylesheet Language.*

References

[1] GEM specifications, http://www.mhp.org/specs/ts_102819v010401p.pdf.

[2] ETS 300 744, "Digital broadcasting systems for television, sound and data services; framing structure, channel coding and modulation for digital terrestrial television," ETSI Draft EN 300 744 V1.2.1, 1999-1.

[3] Y. Wu, E. Pliszka, B. Caron, P. Bouchard, and G. Chouinard, "Comparison of terrestrial DTV transmission systems: the ATSC 8-VSB, the DVB-T COFDM, and the ISDB-T BST-OFDM," *IEEE Transactions on Broadcasting*, vol. 46, no. 2, pp. 101–113, 2000.

[4] ETSI, "Implementation guidelines for DVB terrestrial services: transmission aspects," Tech. Rep. TR 101 190v1.1.1, 1997-12, ETSI, 1997.

[5] ARIB, "Terrestrial integrated services digital broadcasting (ISDB-T)—specifications of channel coding, framing structure, and modulation," September 1998.

[6] SBTVD Forum, http://www.forumsbtvd.org.br.

[7] MHP standard, http://www.standardsdirect.org/standards/ .

[8] J. Piesing, "The DVB Multimedia Home Platform (MHP) and related specifications," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 237–247, 2006.

[9] ARIB, http://www.arib.or.jp.

[10] ARIB standards, http://www.arib.or.jp/english/html/overview/sb_ej.html.

[11] Volume 1 of ISDTV-T Standard 06. ISDTV-T Fórum, December 2006.

[12] Volume 2 of ISDTV-T Standard 06. ISDTV-T Fórum, December 2006.

[13] Volume 3 of ISDTV-T Standard 06. ISDTV-T Fórum, December 2006.

[14] O. Bohl, J. Schellhase, R. Sengler, and U. Winand, "The sharable content object reference model (SCORM)—a critical review," in *Proceedings of the International Conference on Computers in Education (ICCE '02)*, vol. 2, pp. 950–951, Aukland, New Zealand, December 2002.

[15] tmira solutions, http://www.tmira.com.

[16] The Digital Video Broadcasting Project, http://www.dvb.org.

[17] ISO/IEC standards, http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html.

[18] The World Wide Web Consortium, http://www.w3.org.

[19] V. Balabanian, L. Casey, N. Greene, and C. Adams, "An introduction to digital storage media—command and control," *IEEE Communications Magazine*, vol. 34, no. 11, pp. 122–127, 1996.

[20] ETSI standards, http://www.etsi.org/WebSite/Standards/StandardsDownload.aspx.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 186281, 8 pages
doi:10.1155/2009/186281

Research Article

Sofa-TV: The New Digital Landscape

Giovanni Toletti and Luca Turba

Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy

Correspondence should be addressed to Luca Turba, luca.turba@polimi.it

Received 31 January 2009; Accepted 12 August 2009

Recommended by Alessandro Vanelli-Coralli

Television is attracting an enormous amount of attention from both researchers and managers, due to the profound changes that are taking place thanks to the diffusion of digital technology. The study of the digital landscape of television, including the players competing in its arena and their strategies, is well worth the effort. This paper, based on 32 case studies and the census of the Sofa-TV (Sat TV, DTT, and IPTV) offerings, aims at describing the current state of channel offerings, individualizing the principal players, and identifying their strategies, thus allowing us to give a few predictions as to the possible future changes in the industry. The analysis will have a general applicability, as the considerations made are not particularly country-specific, although performed within the Italian context, one of the most advanced in the development of digital television platforms.

1. Introduction

Television is attracting an enormous amount of attention from both researchers and managers, due to the profound changes that are taking place. The television business, traditionally static and conservative, is undergoing a radical transformation process in the multimedia age. The introduction of digital technology is indeed driving important changes in the market offerings and expanding the boundaries of the television business. While these boundaries are becoming fuzzier, opportunities for new players are largely increasing. Hence, the study of the digital landscape of television, including the players competing in its arena and their strategies, is well worth the effort.

2. The TV Industry

The television industry, for decades quite static, is currently undergoing important changes [1]: such business is in fact being shaped by a number of driving forces, such as the digitalization of the TV signal, the diffusion of new alternative access technologies, the development of broadband and streaming video technologies, the introduction of Web TV and new generation access terminals, and the progress of interactive and personal television solutions [2–7]. Past changes, as well as those to come, are so radical that some authors prefer to talk about TV (r)evolution rather than evolution, asserting that the very meaning of the term "television" needs to be revised [8].

The TV business is no longer restricted to few actors but is becoming territory to be conquered by new entrants from industries like Internet and telecommunications. The market offerings are reaching new access terminals (such as PC and handheld devices), meeting and integrating with the Internet and with other online multimedia services, and changing by creating new television formats, such as the personal and the interactive television.

The technological evolution has shifted the boundaries of the TV industry. Television platforms are in competition, not only among themselves, but also with "New Media" offering contents and services not specifically related to the television sector [9, 10]: (i) Internet and (ii) all the new ways of offline use of digital contents (e.g., podcasting, downloading of entertainment contents on the PC or mobile phone). There is strong competition between traditional media and new media for the *share of time* of users and the *share of advertising* of investors.

The evolution or (r)evolution of television has caused the emergence of various television typologies being all very different amongst themselves: ranging from the more "traditional" one, based on linear programming (whose main new features are its transmission platform, which is digital instead of analog, and its viewing device—not only the television screen but also the PC or telephone) to the more innovative on-demand based editorial contents or those generated by the user (for which the need for the television terminal itself could be challenged).

Six digital platforms can be identified on digital networks: Sat TV, DTT, IPTV, Web TV, Mobile TV on DVB-H network, and Mobile TV on cellular network. However, given the differences and similarities between the six platforms, a further clustering that would achieve a subdivision into only three TV macrocategories could be a possible scenario [11].

(1) *Sofa-TV*. It includs all digital television typically viewed through the "traditional" television screen. The "new" Sofa-TVs are based on three digital platforms: Sat TV, DTT and IPTV. The use of the expression, "Sofa-TV" aims to clearly describe the viewing opportunities and modalities of these televisions.

(2) *Desktop-TV*. It includs all the video channels that are viewable through the Web (and Internet in general). In this case, the distinguishing element is the proactive viewing ("elbows on the desk") of the contents.

(3) *Hand-TV*. It includs all the TV and video offerings available on the Mobile platform, based on both DVB-H networks and cellular networks. The use of the expression "Hand-TV" aims to focus on the concept of TV viewable in the palm of the hand, which frees this type of television from the underlying technologies, both at the network level (DVB-H, cellular networks, and—in the future—WiFi and its evolutionary products), and at the terminal level (not only cell phones, but possibly other small devices, like portable music readers, mobile game consoles, etc.).

While very different in format, modality, and viewing capabilities one from the other, these three macrotypologies of New TV, although not in tight competition with each other at this time, certainly do compete, together and, in a wider sense, with other nontelevision contents and services, for the users' share of time and investors' share of advertising.

Even if Desktop-TV and Hand-TV will be able to play a very important role in the future digital television arena and are already at present very promising platforms from the point of view of creation of new businesses, globally, the majority of sales presently continue to apply to Sofa TVs.

For this reason, this paper will focus on the area of Sofa TVs, in particular describing the current state of channel offerings, individualizing the principal players and identifying their strategies, thus allowing us to give a few predictions as to the possible future changes in the industry. The analysis will have a general applicability, as the considerations made are not particularly country-specific, although performed within the Italian context, one of the most advanced in the development of digital television platforms.

In order to achieve such results, the paper is further divided into four sections. The first section will present the empirical study on which our considerations are based. The second section will describe the offerings of the Sofa-TV sector. The third section will deal with the players of the Sofa-TV industry and their strategies. Finally, the fourth section will offer some elements for further discussion.

3. Empirical Study

The information needed for the analysis and evaluation of the Sofa-TV sector in Italy was mainly collected through a case study methodology, and an exhaustive census of the Sofa-TV offerings was also taken in order to complete the information base.

As far as the case studies are concerned, as Pettigrew noted [12], it makes sense to choose the cases of analysis as extreme situations and polar types in which the process of interest is "transparently observable." Hence, the sample has been selected considering companies that conformed to the main requirements of the study while presenting both similarities and differences considered important for the analysis. Thus the empirical study has been conducted on a sample of 32 companies that operate at different stages of the new digital television value chain and have different characteristics (e.g., size, revenue, business model).

The panel comprises the following:

(i) 13 cases from broadcasters and Telco operators (i.e., Elemedia, Fastweb, Gruppo Mediaset, Gruppo Telecom Italia, Infostrada, La7/TI Media, MTV Italia, Rai, Rete A, R.T.I. Interactive Media, SitCom, SKY Italia, Tiscali),

(ii) 6 cases from content providers (i.e., Digicast, Einstein Multimedia, Endemol, Turner Broadcasting System Italia, Walt Disney, Yam112003),

(iii) 8 cases from service providers (i.e., BIP, Cisco Italia, IBM, IconMedialab, Kora, Skylogic, TXT Polymedia, Xaltia),

(iv) 5 cases from media/advertising centers (i.e., Carat/Isobar, Digitalia '08, MediaCom Italia, Niumidia Adv, Sipra).

Information was gathered through semiopen interviews based on a common investigative framework. The use of semistructured interviews gives considerable freedom to the interviewer and interviewee, but at the same time, it ensures that all relevant subjects are discussed and that all the required information is collected. Our initial contact was made by email with the managers and/or those directly responsible for the media and platform strategy of each company, providing them with an explanation of the research objectives. We then contacted those managers directly by phone in order to check their availability and to schedule interviews. Finally, we had one or more face-to-face interviews with them.

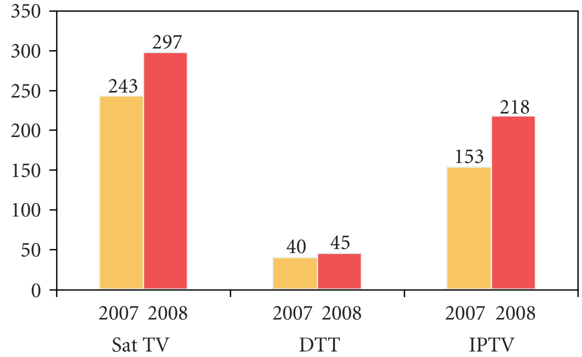


Figure 1: Sofa-TV: the number of channel offerings.

To help interviewers and ensure comparability, an investigative framework was created to cover the most important issues to be investigated in order to achieve the research results. In particular, four main areas of investigation were identified and analyzed during the interviews:

(i) Overall company strategy (e.g., television platforms portfolio, launch of new platforms),

(ii) Value chain analysis (e.g., make or buy decision, partnership relations),

(iii) Organizational structure (e.g., number, roles and structure of the people involved in the Sofa-TV organizational functions),

(iv) Television offerings (e.g., channel offerings on the specific platform, differences of own vs. competitor strategy).

Within the framework, we also indicated possible "prompts" to clarify questions, stimulate discussion, and help the interviewer in data-collecting activities.

For each case, we had single or multiple interviews (according to the relevance and the size of the firm) in order to talk with all the most important decision makers. After the interviews, a short questionnaire was sent out in order to clarify any unclear points and gather more quantitative information.

Moreover, an exhaustive census of the Sofa-TV offerings have been done aimed at mapping all the channels transmitted on the digital Sofa-TV platforms present in Italy today. More than 500 channels were individuated and analyzed both in 2007 and 2008, and for each, 23 important variables were investigated.

The interviews were carried out between April 2007 and October 2008. The census was first taken in October 2007 and then in October 2008, in order to highlight the changes. All the research on the New TV sector began in January 2007 and is still in progress.

4. The Sofa-TV Offerings in Italy

In this chapter, we will analyze in detail the television channel offerings in Italy relative to the three digital platforms included in the Sofa-TV category: Sat TV, DTT, and IPTV.

In Figure 1, we have reported the channels offered in Italy on the three platforms in 2007 and 2008: in all cases, we observe a significant amount of growth in the number of available channels.

(i) The offerings on Sat TV grow from 243 to 297 channels, thanks to SKY Italia's portfolio expansion.

(ii) The DTT offerings increase from 40 to 45 channels, thanks to the introduction of new channels on the part of, first, Mediaset (e.g., Joy, Mya, Steel, and Disney), then Rai (Rai4) and finally radio broadcasts from the Espresso Group (Radio Deejay, Radio Capital, m2o).

(iii) IPTV channel offerings rose notably, from 153 to 218, as the result of two phenomena: first, the transposition of channels from SKY Italia satellite platform to IPTV, and second, the significant increase both in the on demand channels, due to the entry of two new players (Tiscali and Infostrada) in this arena as well as in the channels of already existing operators (geared towards specific market niches that were, up until now, ignored).

The Sofa-TV offerings can be broken down referring to the means of distribution of the channels. There are essentially two: on demand channels, where contents are viewable upon viewer request, and linear channels, where there is a programmed schedule that is predefined and delivered in an ongoing manner. There are also certain channels that cannot be classified as either on demand or linear because they do not have sufficient ongoing programming nor are they viewable when the viewer chooses. All these channels provide valuable contents (particularly soccer, but also sports, movies, and TV series) but are available exclusively during certain predefined times and for only a few hours per day.

It should be noted that the on demand channels are only present on IPTV (see Figure 2), the only platform that, from the technological point of view, can support one-to-one delivery on request. The other platforms, nevertheless, are trying to increase the viewing flexibility of their own contents in order to more closely match the on demand rationale. On the one hand, both the satellite and the DTT platforms have introduced their +1 versions for certain pay channels in order to make them more easily viewable for their users (since they transmit the programming with a one-hour delay), and on the other hand, the monopolizer of the satellite platform, SKY Italia, counts mostly on its MySKY service which, in fact, allows the user to create his own on demand library through program recording (see the following box on "Time shifting and Catch-up TV").

To complete the analysis of the offerings, it is beneficial to distinguish between "native" channels on each platform (created specifically for that platform) and transposed channels (those replicated on the platform in question, but native of other platforms) (see Figure 3). Naturally, given the technological limitations that were previously mentioned, such a distinction makes sense only in regards to linear channels, since on demand channels are present exclusively

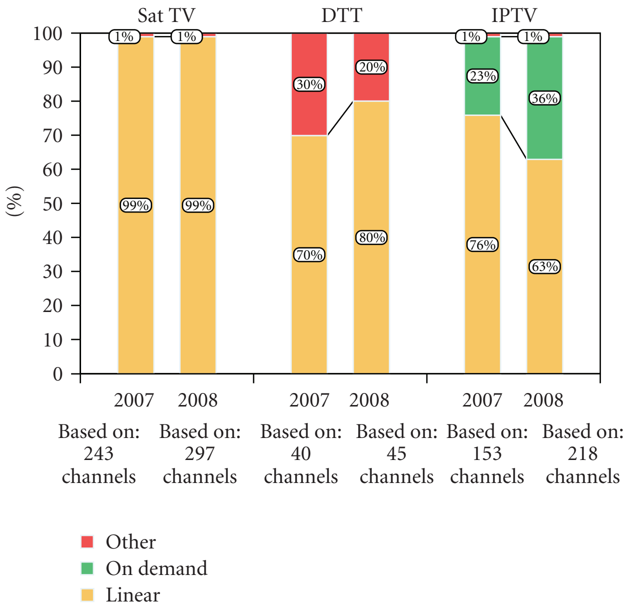


Figure 2: Sofa-TV: modalities of channel delivery.

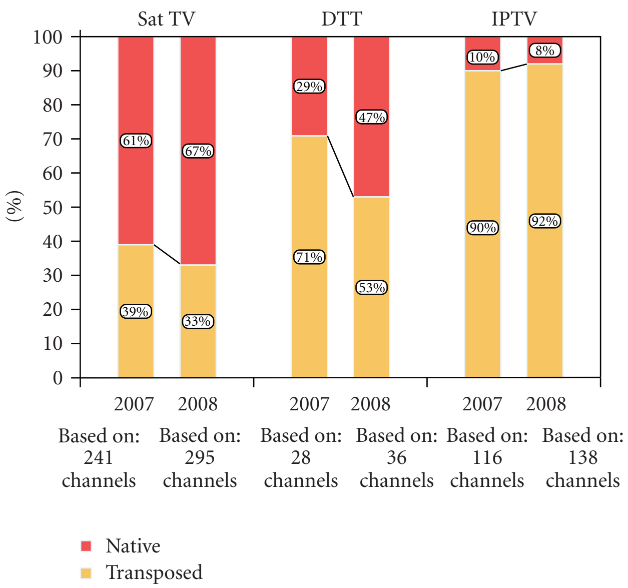


Figure 3: Sofa-TV: native and transposed linear channels.

on and were created solely for, IPTV. In the case of linear channels, we see how the previously indicated growth in the number of channels can be explained in two ways. On one hand in fact, it can depend on the diverse propensity for innovation (intended as the ability to create new channels) of the different platforms, and on the other, it can be tied to the diverse propensity for imitation (intended as the ability to replicate the offerings generated by other platforms). In the case of satellite and DTT, the propensity for innovation prevails, and the widest assortment of "native" channels continues to be that of satellite. However, in the last year, there has been a significant planning effort on the part of DTT as well, despite the intrinsic limitations of transmission capacity. In the case of IPTV, there is the prevalence of duplication of the SKY Italia channel offerings, since at the moment, such a platform is not able to create its own original offerings of linear channels.

4.1. Time Shifting and Catch-Up TV. The two most innovative services present on Sofa-TV platforms are Time shifting and Catch-up TV.

Time shifting is a term referring to the possibility of viewing a program the moment it is chosen by the user, at a time other than the actual streaming of the channel. This typically means that the program is recorded either on a decoder equipped with an integrated hard disk or remotely on a server. In the latter case, the service can be provided through subscription. Its functionality is very developed compared to traditional Video Recorders: for example, it is possible to begin viewing a film before recording is complete. Currently, Time shifting is present on Sat TV and IPTV platforms and can be activated remotely through the broadcaster website or user cellular phone. It is offered by most digital televisions in the world: in the United Kingdom and in Ireland, for example, Sky+ offers the option, whereas the US market is dominated by TiVo and DirectTv; in Italy the service is offered not only by Sky on satellite through MySky service but also by Fastweb and Infostrada through IPTV.

Catch-up TV is a service that allows through On Demand modality the possibility to view programs already transmitted previously during regularly scheduled programming. For their users, the service providers record on specific servers programs already aired on National channels. In this way, they create a temporary library available to the user after program airing or streaming has taken place. In Italy on Sofa-TV, this service is available only on Fasteweb's IPTV through the ReplyTV service, which records the previous three day's National channel programming (RAI, Mediaset, La7, and MTV). Orange, a French IPTV platform, offers a Catch-up TV service that allows repeated viewing of all programming transmitted on 5 national French channels on air from 18 to 24 hours prior to streaming and maintains it in a library for a minimum of 7 to 30 days.

5. The Principal Players and Their Strategies

The introduction of digital technology in the Sofa-TV arena has greatly changed the competition in this sector, allowing the entry of both new operators, born specifically to take advantage of the possibilities offered by digitalization as well operators coming from other industries. In particular, in the Italian Sofa-TV market, the categories of most important players refer to the "traditional" broadcasters of analog TV (i.e., RAI, Mediaset and La7), satellite TV broadcasters (i.e., SKY Italia—the result of the fusion of the first two Italian operators in the satellite TV arena—Stream and Telepiù), and a series of players who, with already consolidated businesses in sectors like telecommunications in particular, but also paper publishing, entered the Sofa-TV arena thanks to the opportunities introduced by the new digital technology (e.g., Fastweb, Telecom Italia, Infostrada, Tiscali, Gruppo L'Espresso, Class Italia).

Therefore, digital technology, while on one hand bringing a strong increase in the number of competitors in a sector characterized for years by the Rai-Mediaset duopoly, and on the other, putting in competition many different players, has decisively changed the strategies of the players and their ways of competing in the Sofa-TV arena.

Following this paragraph, we will try to highlight the main characteristics of the most important clusters of competitors discussing the strategies developed by these operators.

5.1. Traditional Broadcasters. Due to both recent EC legislation delineating the passage to digitalization and its analog-to-digital switch-over times, and to the effective strategy executed by SKY Italia on satellite, traditional channel operators have been pushed to a faster and more efficient capitalization of digital opportunities. However, in order for DTT to not represent merely a siphoning off of revenues from the analogical to the digital platform, they must maximize strategies and channel offerings that, on one hand, can offer a competitive alternative to SKY Italia's pay TV (and the IPTVs, as well), and on the other hand, can increase the volume of advertising on free-to-air channels either through an increase in the audience and/or its greater segmentation and profiling or through innovative advertising campaigns, that, thanks to the possibilities offered by digital technologies, could prove more effective.

Mediaset, in particular, completely changed its strategy in 2008, improving on an already extremely interesting offer, based no longer only on pay channels centered on soccer, cinema, and sports, but also on three new pay channels (which offer, in a premium rationale, the segmentation that characterized the historical free-to-air triad) as well as a children's channel (Disney). This offering primarily addresses a different target compared to the SKY Italia offer, with the clear intent of broadening its pay TV user market, relying on the ever-increasing penetration of integrated tuner televisions.

Rai introduced a new channel, Rai 4, and has continued experimenting on HD. There is a big question mark as to the role that Rai will be able to play, considering its limitations as public operator, which do not allow it to compete in the pay TV market.

La7 has worked on several interesting experiments tied to the assessment of the potential DTT offerings in terms of interactivity and advertising.

On the other platforms of Sofa-TV, traditional broadcasters have continued, up until 2008, to position themselves further upstream in the chain as content and programming suppliers with less concentration on the final client. Already from 2009 things should change, since a consortium of the three biggest traditional broadcasters is working to launch a free satellite platform as an alternative to that of SKY Italia.

5.2. Satellite Broadcasters. At the moment SKY Italia is the only important operator in the satellite platform arena, even if, as noted, things could change soon. In these last few years, it has been getting extremely positive results (it is in fact ranked the Number Two Italian television operator in terms of revenue, if we do not consider the national Rai-imposed tax that all Italians must pay in order to have television).

In the last year, SKY Italia has aimed for both a further broadening of its already ample and varied channel offerings, as well as a greater penetration of its most innovative services (MySKY, which allows an easy program recording and/or delayed viewing, and HD, which allows High Definition viewing of a great number of contents).

It is not present on DTT, whereas on IPTV it plays an indirectly key role, since its offerings represent a fundamental component of three of the four Telco operators' packages in this sphere.

5.3. The Telcos. The Telcos play their game in the field of television by pushing IPTV, even if, in absolute terms, the results are still marginal compared to other digital platforms. It is, however, necessary to highlight that the offerings in the Telco television sphere are more tied to the need to expand their service portfolio in order to reduce the churn rate and increase the ARPU than to their desire to develop a new business that can be profitable in and of itself.

Certain important events have characterized this division in the course of 2008: Telecom Italia changed its strategy, presenting its television offerings in a bundle with broadband, without extra cost to the customer (thanks to this new strategy and an intense promotional campaign, Telecom significantly increased its number of users, closing in on the leader in the sector, Fastweb); two new players, Infostrada and Tiscali, entered the field, and even if up until now they play an absolutely marginal role, they are interpreting the concept of IPTV in an original way, with a "hybrid" model that brings different characteristics, typical of the Web to Sofa-TV. Tiscali, in particular, has greatly invested in IPTV, creating a decisively innovative product in regards to user interface, graphic qualities, and on demand content offerings.

To complete the analysis of Sofa-TVs, besides the players in direct competition previously analysed, it is appropriate to make a few considerations regarding other players belonging to the television industry, which, from a strategic point of view, play an important role.

5.4. Content Providers. They have the task of creating and aggregating television contents, including shows, daily news, films, soap operas, fictions, cartoons, various series, and commercials. For this reason there is a wide range of players able to fill this role; amongst these we also include the user, the creator of UGC (User Generated Content), now very diffuse, especially in the world of the Web, but also on Sofa-TVs like Qoob TV, a channel transmitted through DTT, IPTV, and Web TV, where part of the programming, even if aggregated by the Broadcaster Telecom Italia Media, is formed by content totally generated by the user. At the moment, there is a simple transposition of contents from one platform to another, but there are some companies, such as YAM112003, that are realizing adhoc contents for specific digital television platforms.

5.5. Network Providers. These are those who deal with the distribution and diffusion of the channel; the main actors are Telcos that exploit their own infrastructures already diffuse on the territory, in order to transmit the television signal on various platforms, like Fastweb, which has created its own network. It has exploited the advantages of vertical integration, in the face of steep costs involved in sustaining in initial phases. There are then the broadcasters who operate as network providers as well, like Dfree, which, besides being a broadcaster, also transmits on its own MUX both its offer and a group of Mediaset channels.

5.6. Service Providers. They create services that are integrated and distributed in the broadcaster offers. The main services are, for example, Video On Demand, Time Shifting, Catch-up TV, T-Gaming, T-Learning, and all the services of Conditional Access. The actual creators of services are those like IconMediaLab, Enterprise Digital Architects, Kora, and Xaltia, which operate in close contact with all the main New TV players, creating applications, graphic interfaces, and interactive platforms inserted in contents and programming of different television offers.

In the following section the company profiles of the three main players in the Italian Sofa-TV sector (i.e Mediaset, Rai and SKY Italia) will be described.

5.7. Mediaset. Mediaset S.p.A is the main commercial Italian television group and leader, together with Rai television in free-to-air TV, and is present abroad as well (with 50.3% of Gestevisión Telecinco Group, a Spanish TV broadcaster) and operates on all the new digital television platforms with different roles in the value chain.

Given the increasingly greater strategic importance of controlling the creation of television contents, it has recently acquired 75% of Endemol, developing a twofold strategy: the in-house realization of contents transmitted on free-to-air channels and the acquisition from external providers (or production in partnership) of contents transmitted in the pay television area. Thanks to this strategy, the company has significantly reduced the lead of the first mover in the pay TV market, SKY Italia.

On Digital Terrestrial, the leading platform, it carries out the activities of Content Supplier, Channel Operator, and Service Supplier, setting itself up as a strongly integrated vertical operator. The offerings of the group can be divided into four typologies: the generalist channels present on analog TV as well, channels created ad hoc for the DTT platform, like Boing (children's channel) and Mediashopping (channel dedicated to telesales), Mediaset Premium channels (thematic channels with Pay content), and Premium Gallery channels (thematic channels viewable only by monthly subscription). Both Premium channel typologies are based on strong content appeal, like soccer, movies, TV series, and cartoons. For these contents the modality of acquisition does not require a contract but is based on the acquisition of a prepaid, rechargeable card. The company's entire array of DTT offerings aim to create channels that are complimentary to the generalist channels, with a well-defined viewer target (i.e., segmented channels) and a reduced membership cost for Premium packets compared to its competitor SKY Italia, although offering a narrower breadth of offerings. The company is integrated downstream as well with the control of two Multiplexes, through which it broadcasts its own offerings, and rents, in adherence to the norms in force, 40% of the transmission capacity to third-parties, setting itself up as a full-fledged network operator.

Finally on satellite and IPTV, the company is set up as a video content provider.

5.8. Rai. Rai, the Italian public broadcaster, controls all the New TV platforms using various strategies.

On DTT it maintains a vertically integrated control: from content production to the management of transmission infrastructure. At the level of national offerings, the Rai television channels can be grouped into three different typologies: the simulcast of three generalist channels (Rai 1, Rai 2, and Rai 3), channels initially broadcast via satellite and re-presented on DTT (Rai News 24 and Rai Edu 1), and finally ad hoc channels developed for DTT (Rai Sport Più, Rai Gulp, and Rai 4). The current business model of Rai's DTT offerings, in continuity with its generalist channels, is in better terms based prevalently on advertising and the national TV tax, as it has no pay offerings at present. Rai has done some experimenting with transmitting in HD for certain major sporting events like the Winter Olympics in Turin and the European Soccer Championship in 2008. Moreover, on DTT Rai acts as a network operator, managing two national Multiplexes and renting 40% of its transmission capacity to third-parties, setting itself up, in this case, as television content carrier for others.

On Satellite, Rai is present both as channel provider for the SKY Italia pay package as well as free public service broadcaster. Finally, through Rai Trade Company and in partnership with the Inter, Juventus, and Roma soccer clubs, it produces and commercializes channels dedicated to these teams and distributed by SKY Italia.

On IPTV, Rai has been present since 2001 with Rai Click, a joint venture between Rai (40%) and Fastweb (60%), created to broadcast the offerings originating from the current Rai1, Rai2, and Rai3 programming as well as the Rai archives and Rai cinema films in VOD modality through the Fastweb TV platform. Rai is also present on the IPTV platforms of Alice and, since 2008, those of Tiscali with offerings in thematic areas.

5.9. SKY Italia. SKY Italia is the only satellite pay operator in the Italian television market and boasts over 4.6 million subscribers (September 2008).

Its program and service offerings are qualitatively and quantitatively significant, with over 170 thematic and pay per view channels that cover all the different categories: film, entertainment, sports, news, documentaries, travel, music, and children's channels.

The company operates through partnerships with Italian and international content suppliers: thematic channel publishers hosted by the platform, and production companies involved in the creation of the programming and new formats for entertainment, cinema, sports, and news channels.

SKY Italia tries to make up for the reduced intrinsic interactive capacity of its satellite platform by increasing services embedded in its decoders: for example, it offers the possibility to record and save programs thanks to the hard disk included in the STB and offers its HD viewing service with 6 channel offerings, which at the moment is the only of its kind in Italy. Even within its programming there are certain interactive services available that allow the user to play games on TV and to have interactive mosaics during certain sporting events.

SKY Italia broadcasts its own channels on IPTV channels as well, on Fastweb, Alice Home TV of Telecom Italia, and Infostrada TV of Wind, through agreements with these carrier companies that allow it to increase its subscriber numbers and the penetration of its offerings within the Italian territory.

6. Conclusions: Current Trends and Future Turning Points

In conclusion, it is possible to make some consideration as to what might be the future evolution of the industry. Such considerations, although obtained from the analysis of the Italian reality, can certainly be applied to other international contexts as well, since the factors that could influence the development of the industry are, in large part, not country-specific.

The future turning points that we see on the horizon are (i) the competitive battle in the pay contents market, which is getting increasingly more interesting, with on one hand, SKY Italia's continued growth, increase, and enrichment of offerings, and on the other, Mediaset's new and improved, more aggressive game played by the field of DTT; (ii) Rai's ability to develop a position in the sector, considering its complete exclusion from the Pay TV arena, currently the most attractive attribute of digital television; (iii) IPTV operators' ability to carve out their own niche and exploit both the specificity of this platform and their key resources as operators of Telecommunication and Internet service providers.

In particular, the main factors that will be able to influence the development of the market, not only in Italy, but in general in every country characterized by the evolution of the TV towards digital platforms, are the following three:

(i) the normative and regulatory framework,

(ii) technological evolution of the networks,

(iii) the evolution of the systems and tools for measuring the audience and the users.

6.1. The Normative and Regulatory Framework. It is obvious that any evolution of the normative or regulatory framework could heavily influence the dynamics of this sector. Up until today in Italy the two most interesting phenomena are, on one hand, the definition of times and modalities (including the theme of public subsidies for the purchase of STBs) of the switch-off of analog TV, which have already begun to apply to certain regions of Italy and which, up until today, seem to be, in large part, fixed; and on the other hand, the theme of regulation of product placement in TV programs, which could have a strong impact on the roles and strategies of the operators.

6.2. The Technological Evolution of the Networks. The three categories of Sofa-TV are based on completely different channels from a technological point of view, due to both their current characteristics as well as the evolution that they will undergo in the years to come.

In an attempt to simplify and schematically present this theme, we can characterize the different digital platforms by three main performance characteristics:

(i) the bandwidth, which influences the number of deliverable channels and their quality (e.g., influencing the possibility of transmitting in HD);

(ii) the intrinsic interactivity, that depends on the presence of an embedded back hl channel;

(iii) The geographic coverage (which, in reality, is not an intrinsic characteristic of technology but depends on the investments made within the system of operators).

It is obvious that the positioning of the three digital platforms on these three parameters is varied and is destined, moreover, to undergo very different dynamics in time.

(i) Sat TV presents clear strengths in its territorial coverage and bandwidth; the weakness of this platform is, however, the absence of an intrinsic return signal, which limits its capabilities for certain formats (like on demand) and for certain more interactive services.

(ii) IPTV has, on the other hand, as its own strong point, the back channel intrinsic of its network (IP) and therefore capabilities on an interactive level; in terms of coverage of the territory and breath of range, it still suffers from problems tied to the digital divide, with the presence therefore of areas of the country not yet reached by a sufficient bandwidth.

(iii) DTT starts with good coverage of the territory, but suffers from a limited bandwidth and from a lack of intrinsic back channel.

With reference to the back channel it should be noted that the platforms that do not have embedded potential can still use external channels, with normal Internet connectivity at home or on the cellular network (as, i.e., has already been done for certain services, like the pay per view contents).

6.3. The Evolution of the Audience and User Measurement Systems and Tools. One of the great novelties brought by digital platforms is represented by the possibility of establishing a real relationship with users, to know their behaviors, and therefore, to profile and group them (even if this possibility is much greater in the presence of a back channel).

It is evident that, as these characteristics are very interesting for the advertisers, they have the ability to create new revenue opportunities for the broadcasters. However, in order that this truly become a reality, there is a need for adequate measurements which are both consistent with the new platforms, and trustworthy systems of measurement of the audience, ones accepted by the entire market.

The three factors analyzed above, the regulations, the technological evolution, and the achievement of audience measuring systems, will in coming years strongly influence the offering strategies and business models of the broadcasters on the different platforms and therefore their reciprocal competitiveness.

References

[1] J. Meisel, "The emergence of the internet to deliver video programming: economic and regulatory issues," *Info*, vol. 9, no. 1, pp. 52–64, 2007.

[2] A. Bria, P. Kärrberg, and P. Andersson, "TV in the mobile or TV for the mobile: challenges and changing value chains," in *Proceedings of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07)*, pp. 1–5, Athens, Greece, September 2007.

[3] A. Jirachaipravit and D. Probert, "Technology management and broadband internet regulation: the case of Thailand," *International Journal of Technology Management*, vol. 40, no. 1–3, pp. 157–175, 2007.

[4] D. H. Shin, "Potential user factors driving adoption of IPTV. What are customers expecting from IPTV?" *Technological Forecasting and Social Change*, vol. 74, no. 8, pp. 1446–1464, 2007.

[5] C. Burbridge, "IPTV: the dependencies for success," *Computer Law & Security Report*, vol. 22, no. 5, pp. 409–412, 2006.

[6] M. McGrail and B. Roberts, "Strategies in the broadband cable TV industry: the challenges for management and technology innovation," *Info*, vol. 7, no. 1, pp. 53–65, 2005.

[7] A. W.-L. Hu, "An empirical test of a use-diffusion model for Taiwan mobile digital TV," *International Journal of Technology Management*, vol. 39, no. 3-4, pp. 248–263, 2007.

[8] A. Rangone and A. Turconi, "The television (r)evolution within the multimedia convergence: a strategic reference framework," *Management Decision*, vol. 41, no. 1, pp. 48–71, 2003.

[9] M. Poel, A. Renda, and P. Ballon, "Business model analysis as a new tool for policy evaluation: policies for digital content platforms," *Info*, vol. 9, no. 5, pp. 86–100, 2007.

[10] E. Furregoni, A. Rangone, F. M. Renga, and M. Valsecchi, "The mobile digital contents distribution scenario," in *Proceedings of the 6th International Conference on the Management of Mobile Business (ICMB '07)*, Toronto, Canada, July 2007.

[11] G. Toletti and L. Turba, "Digital Televisions: the new Sofa-TV…and not only," in *Proceedings of the International Conference on Business and Economy*, Constanta, Romania, 2008.

[12] A. Pettigrew, *The Management of Strategic Change*, Blackwell, Oxford, UK, 1988.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 597848, 25 pages
doi:10.1155/2009/597848

Research Article

MBMS—IP Multicast/Broadcast in 3G Networks

Frank Hartung, Uwe Horn, Jörg Huschke, Markus Kampmann, and Thorsten Lohmar

Ericsson GmbH, Eurolab R&D, Ericsson Allee 1, 52134 Herzogenrath, Germany

Correspondence should be addressed to Markus Kampmann, markus.kampmann@ericsson.com

Received 1 March 2009; Accepted 1 December 2009

Recommended by Sooyoung Kim

In this article, the Multimedia Broadcast and Multicast Service (MBMS) as standardized in 3GPP is presented. With MBMS, multicast and broadcast capabilities are introduced into cellular networks. After an introduction into MBMS technology, MBMS radio bearer realizations are presented. Different MBMS bearer services like broadcast mode, enhanced broadcast mode and multicast mode are discussed. Streaming and download services over MBMS are presented and supported media codecs are listed. Service layer components as defined in Open Mobile Alliance (OMA) are introduced. For a Mobile TV use case capacity improvements achieved by MBMS are shown. Finally, evolution of MBMS as part of 3GPP standardization is presented.

1. Introduction

Mobile networks have emerged from voice telephony networks to multimedia delivery networks. Mobile TV services have become quite popular data services during the past two years. Apart from live TV channels, the offerings often include special mobile editions with highlights from the weekly TV program, such as series and comedies, delivered as looped channels.

The majority of today's mobile TV services are delivered over existing 3G networks [1, 2] since this is the fastest and easiest way to deploy mobile TV services. Existing 3G operators have enough capacity in 3G networks to scale up for a mass market of mobile TV services. Latest 3G technology such as HSDPA (High-speed data packet access) [3] gives room for several steps of capacity increases. This allows for more users while benefiting both the diversity and the quality of mobile TV services. The underlying technology is called packet-switched streaming (PSS) [4, 5]. PSS is nowadays supported by all UMTS terminal vendors and offers high-quality streaming services for live or on-demand services. Further quality improvements have been achieved by the introduction of the advanced H.264/AVC video codec [6, 7] and by the introduction of streaming bearers with specific Quality-of-Service (QoS) support.

Nevertheless, the increasing popularity of mobile TV and similar services may lead to situations in which many users want to watch the same content at the same time. Examples are live events of high interest like soccer matches, game shows, and so forth. For those cases, multicasting, known from the internet, or broadcasting is clearly more appropriate technologies.

The work on adding broadcast/multicast support to 3G networks started back in 2002 when both 3GPP and 3GPP2 created work items for broadcast/multicast services in GSM/UMTS and CDMA2000, respectively. In 3GPP the work item was called Multimedia Broadcast and Multicast Service (MBMS). In 3GPP2 it was named BroadCast and MultiCast Service (BCMCS). The specifications of 3G broadcast services were functionally frozen in 2005. Both MBMS and BCMCS introduce only small changes to the existing radio and core network protocols. Therefore, MBMS and BCMCS can be introduced by a pure software upgrade in general as long as the underlying hardware provides sufficient processing power which is the case for all state-of-the-art network nodes. This reduces the implementation costs both in terminals and in the network. It makes cellular broadcast a relatively cheap technology if compared to noncellular broadcast technologies [8–10] which require new receiver hardware in the terminal and significant investments into a new network infrastructure.

MBMS and BCMCS focus on the transport aspects of broadcast and multicast services. They provide transport bearers over which IP multicast [11] packets can be delivered.

They also provide protocols and codecs for the delivery of multimedia files and streams on top of the IP multicast bearers provided by MBMS [12].

Service layer functionality related to multicast/broadcast transport services has been defined by the Open Mobile Alliance (OMA), in the Mobile Broadcast Services 1.0 (BCAST 1.0) specification, which was finalized in 2007 [13, 14]. BCAST 1.0 addresses features like content protection, service and program guides, transmission scheduling, notifications, and service and terminal provisioning. It enables, besides linear TV services, also Podcast services. OMA BCAST 1.0 is agnostic with respect to the underlying broadcast/multicast distribution scheme and applies to MBMS, BCMCS and other noncellular broadcast systems like DVB-H.

Focus of this paper will be an introduction into MBMS as specified by 3GPP. We will start with highlighting the main features of MBMS. After that we give an architectural overview, followed by an explanation in which capabilities MBMS adds to the various protocol layers, starting with the physical layer and then moving up in the protocol stack. In addition to this, capacity and media quality issues will be addressed.

(a)

(b)

Figure 1: MBMS prototype terminals: (a) first prototype shown in public in February 2007, and (b) prototype terminals shown in public in February 2008.

2. MBMS Overview

This section starts with a history and a high-level summary of the key features of MBMS. After that we explain how MBMS fits into the 3G architecture. We will describe in a bit more detail the so-called BM-SC (Broadcast/Multicast Service Centre) which acts as a container for the new functions MBMS provides. At the end we will describe the phases in a typical MBMS service delivery scenario.

2.1. History and Key Features. Standardization of MBMS started in 3GPP and 3GPP2 back in 2002. The first version of the standard was functionally frozen in 2005 and finalized during 2006. Already beginning of 2007 the first MBMS prototype was demonstrated at the 3GSM World Congress in Barcelona (Figure 1(a)). A next version of an MBMS prototype terminal was presented at MWC 2008 (Figure 1(b)). Market introduction of MBMS in UMTS networks is expected during 2010/11

MBMS has been specified as an add-on feature to existing cellular network technologies, namely, GSM/EDGE [1] and UMTS [2]. Existing transport and physical radio channels of those systems were reused to a large extent in order to keep the implementation effort low. In the ongoing standardization of LTE [15–18] very efficient support of MBMS is taken into account from the start as a main requirement.

MBMS supports two basic transmission modes for delivering IP packets: broadcast and multicast.

The MBMS Broadcast mode can be used to deliver IP packets to all terminals in a certain area or the whole network. If the MBMS broadcast mode is used, a transmission bearer is setup for all cells in which the service should be available and is continuously transmitting as long as the service is up and running. In broadcast mode MBMS does not require an uplink connection and can thus be used like any other "downlink-only" broadcast technology such as DVB-H and DMB.

The MBMS Multicast mode works very similar to IP multicasting. A terminal which wants to receive information related to a particular multicast channels "joins" one or several content channels (e.g., expresses interest to receive content associated with this channel). This information is processed in the routing layer of the core network and is used for optimizing the data delivery path. "Optimizing" means that over connections shared by receivers of the same multicast channels, data is transmitted just once. The only drawback of multicasting is the additional delay when switching from one channel to another one. Therefore MBMS multicasting is less suitable for mobile TV services, which usually require a low TV channel switching delay. The main application of MBMS multicasting is for download or Podcast services.

MBMS was specified such that broadcast/multicast services can be used together with voice and data services within the same radio carrier. This gives the greatest flexibility to cellular operators.

In MBMS it is possible to define so-called MBMS service areas (MSAs). An MSA is defined as a collection of cells. In each of the defined MBMS service areas different services can be delivered. In this way MBMS services can be restricted to geographically limited areas.

The various network configurations MBMS supports range from multicast/broadcast transmission in a single cell, over locally restricted areas and up to a nationwide, single frequency network, broadcasting the same content (e.g., TV channels) across the whole country.

The specification of MBMS in the UMTS radio access network (UTRAN) contains an interesting mechanism called "Counting". "Counting" means that an uplink signaling channel is used to inform the radio access network when a terminal requests reception of a content channel. Based on this information, the radio access network can keep track of the number of terminals currently receiving a particular content channel. As we will explain later in more detail, this allows the radio access network to make an intelligent decision on whether a particular content channel needs to be transmitted in a cell or not and if yes which type of radio bearer should be used. "Counting" has originally been developed for MBMS multicast services in UMTS but it can also be combined with the MBMS broadcast services. We will come back to the counting mechanisms later in this Section.

The new bearer type which was introduced by MBMS is the so-called point-to-multipoint (P-t-M) radio bearer. While a point-to-point (P-t-P) bearer can only be received by one terminal, the new P-t-M bearer can be received by several terminals in a cell simultaneously.

In UMTS not only the new P-t-M bearer but also existing P-t-P bearers can be used for MBMS. A P-t-P bearer has the advantage that the network can adapt the radio resource usage to the terminal's reception conditions. In good reception conditions, less radio resources are allocated, whereas in bad conditions more might be needed. In contrast to this, a P-t-M bearer always requires a relatively high amount of radio resources to provide the wanted coverage also for terminals in bad reception conditions. Therefore under certain conditions one or even several P-t-P bearers (one for each receiving terminal) might consume less radio resources than a single P-t-M bearer. The decision which bearer type—P-t-P or P-t-M—uses for a particular UMTS MBMS transmission is made at the radio network controller (RNC). The information for making an optimal decision is obtained from the already mentioned "Counting" process which will be explained in more detail in Section 4.2.

MBMS provides support for both streaming and download services. Hence different types of services will benefit from the introduction of MBMS. One example is mobile TV services in which several operators have launched over existing UMTS networks and which will greatly benefit from the capacity enhancements MBMS brings. Besides mobile TV, other services will benefit, too. One example is audio or video podcasting where mobile users subscribe to a certain podcast provider. MBMS can then be used for the download delivery of the provider's audio and video clips to a large number of subscribers at the same time. Another example is public safety. Here, MBMS could be used for emergency alerts in certain areas, for example, in case of a tsunami or earth quake.

MBMS streaming and download services can be confidentiality and integrity protected, by using the Secure Real-Time Transport Protocol (SRTP [19]) for streaming and a special container format for downloaded data. The keys used for encrypting and decrypting the SRTP, called MBMS traffic keys (MTKs), change frequently and are transmitted in a key hierarchy where the MTKs are protected using MBMS service keys, MSKs. The MSKs are in turn protected using MBMS User Keys (MUKs) which are derived in a bootstrapping procedure between the smartcard (UMTS integrated circuit card, UICC, often also denoted as Subscriber Identity Module (SIM) card) and the device. Thus, the SIM card, more exactly its UMTS version, the UMTS SIM (USIM) card, is the basis for service protection in MBMS.

MBMS employs three packet error recovery schemes to cope with packet losses which could occur during bad radio conditions. The most important one is the use of Forward Error Correction (FEC) coding, which allows recovery of lost packets without any server interaction. FEC can be used both for MBMS download and streaming delivery. For the MBMS download service, two additional file repair mechanisms were defined. The first one uses a point-to-point connection to explicitly request the missing parts of a file. The second uses a point-to-multipoint bearer to deliver missing parts to several terminals at the same time.

The quality of a multimedia service is mainly determined by the compression efficiency of the used audio and video codec and the data rate at which audio and video data are encoded. MBMS reuses standardized, state-of-the-art audio and video codecs. The video codec is H.264/AVC [6, 7]. The latest MBMS release supports image resolutions up to QVGA (320×240 pixels) and frame rates up to 30 frames per second. For audio, the High-Efficiency AAC version 2 codec [20] and the various AMR codecs (narrow band [21], wideband [22], and enhanced wideband [23]) are supported. In the current MBMS release the transmission rate for streaming delivery is limited to 256 kbps. Higher bitrate bearers are under development.

MBMS defines basic components for the encryption of the delivered audiovisual content. However, MBMS does not specify additional service enablers like a program guide or support for interactive services such as voting delivered together with the audiovisual content. Instead those additional functionalities were specified by the Open Mobile Alliance (OMA) under the term "Mobile Broadcast Enabler", or abbreviated BCAST 1.0 [13, 14]. In Section 7 we will explain BCAST 1.0 and its application to MBMS in more detail.

If the same multimedia service shall be provided over geographical areas comprising several cells or even an entire nation, additional advantages can be taken from the inherent single frequency operation of the UMTS network. In single frequency network (SFN) mode all cells use the same carrier frequency and the terminal is therefore inherently able to receive signals from multiple adjacent cells simultaneously. This leads to a significant increase in spectral efficiency and thereby throughput on the used radio resources.

This efficient exploitation of the single frequency network nature for MBMS is called MBSFN in 3GPP. MBSFN operation is currently being specified by 3GPP as an enhancement for the WCDMA technology. MBSFN is also a main driver in the 3GPP work item on the long-term evolution (LTE) [15–18]) of UMTS. Apart from this, performance improvements of the P-t-P and P-t-M bearers used in MBMS are addressed as well.

2.2. Architecture. MBMS does not require architectural changes to existing 3G networks. The new functionalities which MBMS provides to operators and service providers are grouped in a new functional node called the Broadcast/Multicast-Service Centre (BM-SC). The BM-SC can be regarded as a functional interface between content delivery services and the MBMS service offered by a cellular network. Towards the core network the BM-SC controls the set-up and release of the MBMS transport bearers and the scheduling of MBMS transmissions.

Figure 2 shows how the Broadcast/Multicast-Service Centre (BM-SC) fits into the existing 3G architecture. The new functions introduced by the BM-SC require a few extensions to the existing Core Network and RAN nodes and their associated protocols. The GGSN (Gateway GPRS Support Node) is extended such that it can establish, manage, and release a point-to-multipoint distribution tree within the mobile network upon notification from the BM-SC. In case of the MBMS Multicast mode, the GGSN includes only those SGSNs (Serving GPRS Support Node) into the distribution tree, which serve subscribers for that bearer. In case of the MBMS Broadcast Mode, the BM-SC provides a list of SGSNs, which shall be included in the broadcast distribution tree. Furthermore, the GGSN is extended by the capability to receive IP multicast traffic from the BM-SC function and to forward it to the SGSNs participating in the multicast or broadcast session using the existing GPRS Tunneling Protocol (GTP). Currently, the multicast traffic is tunneled over unicast which minimizes the impacts onto the existing 3G architecture. The use of IP multicast on transport level between an evolved GGSN and the radio network nodes is currently discussed in LTE and will be discussed in more detail in Section 9.3.

To both radio access networks supported by 3G—UTRAN (UMTS Radio Access Network) and GERAN (GSM/EDGE Radio Access Network)—capabilities were added to efficiently deliver MBMS data to the designated MBMS service area; that is, the geographical area in which the related MBMS service is active.

In GERAN, MBMS may use up to 5 timeslots in downlink for a single transport channel. Depending on the modulation scheme and the network dimensioning, a transport channel capacity between 32 kbps and 128 kbps can be achieved. The total cell capacity depends on the number of supported frequencies in that cell.

MBMS in UTRAN may use up to 256 kbps per transport channel. Several MBMS bearers can be active at the same time in a single UTRAN carrier. Details of the radio bearer realizations in UTRAN will be discussed in Section 3.

2.3. The Broadcast/Multicast-Service Centre (BM-SC). The sub-functions of the BM-SC and their relations are depicted in Figure 3. It is not necessary to provide all BM-SC sub-functions via a single physical node. It is possible to host the Service Announcement, the Session and Transmission, and the Key-Management functions on different physical nodes.

2.3.1. Discovery and Announcement Functions. The User Service Discovery/Announcement function provides service announcements to end-devices. These Service Announcements contain all necessary information, such as a multicast service identifier, IP multicast address, time of transmission, and media descriptions which a terminal needs in order to join an MBMS service. The Service Announcement information is encoded in XML [24, 25] and SDP [26] format and is structured into information units, which are usually called fragments. MBMS supports several ways for delivering Service Announcements:

(i) Interactive retrieval of Service Announcements: HTTP [27] may be used to fetch Service Announcement fragments (XML encoded information element, see above) from the User Service Discovery/Announcement function.

(ii) Service Announcement using MBMS bearers: the Service Announcement fragments are distributed using the MBMS Download delivery method.

(iii) Push announcement using OMA PUSH [28]: the Service Announcement fragment is pushed using OTA-WSP or OTA-HTTP to the terminal. In the simplest way, that Service Announcement fragment may be encapsulated in one or more SMS messages.

2.3.2. Session and Transmission Functions. The Session and Transmission Functions include all data transmission related functions including data encryption. The function is further subdivided into the MBMS Delivery Function and the Associated Delivery Function. The MBMS Delivery Function basically includes the Multicast Sender, which either delivers files via the MBMS Download services (see Section 5.1) or streams via the MBMS Streaming service (see Section 5.2).

The Associated Delivery Function adds auxiliary procedures such as file repair or reception reporting. The purpose of the File Repair Procedure is to guarantee error-free reception of files delivered over MBMS. File Repair Procedure will be explained in more detail in Section 5. The Reception Reporting Procedure allows the BM-SC to collect reception statistics.

2.3.3. Membership Function. The Membership function is only necessary for MBMS Multicast Bearers. It is used to authenticate the “join” request from a mobile receiver. The authentication request is send by the GGSN using the Gmb interface (see Figure 3) to the BM-SC. The membership function may optionally be connected to the key management function.

2.3.4. MBMS Key Management. The MBMS Key Management function is used to provide authorized terminals (e.g., terminals subscribed to a particular MBMS service)

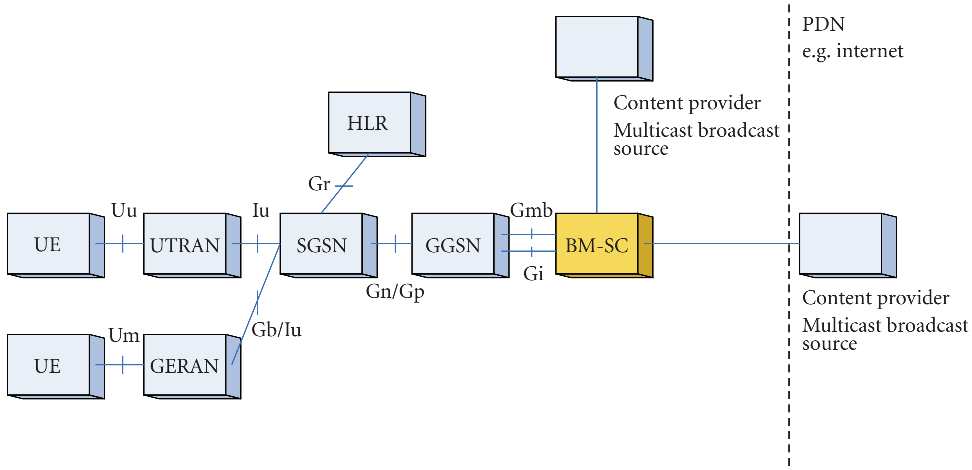


Figure 2: BM-SC in the 3G network architecture.

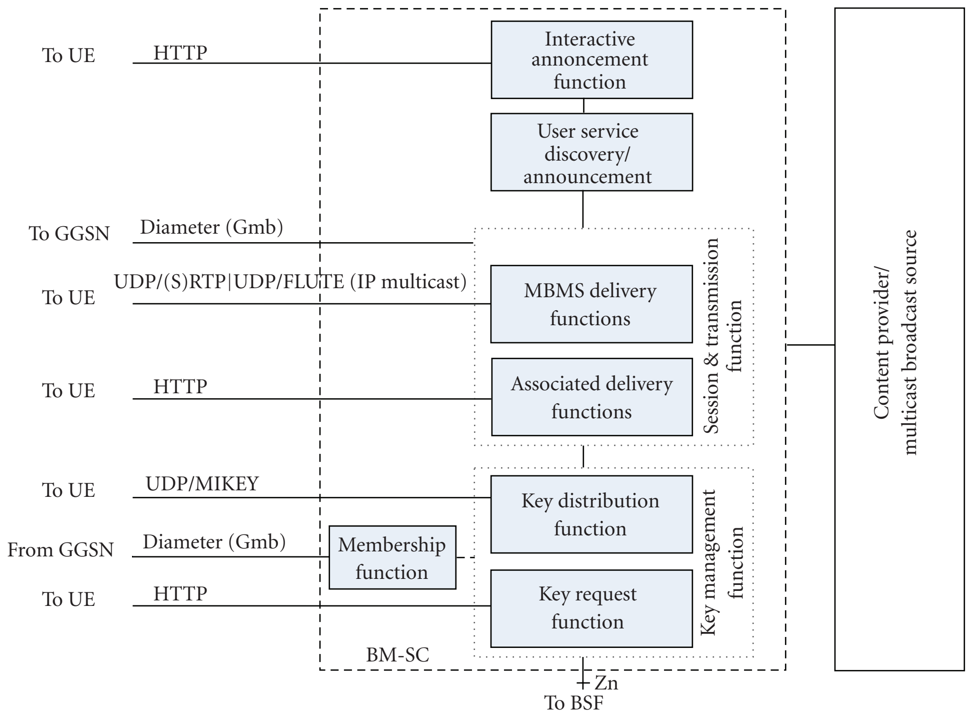


Figure 3: BM-SC Substructure.

with the necessary keys to decrypt the received files or streams.

Figure 4 shows an overview of the MBMS security functions and data flows. The BM-SC is responsible for the generation and distribution of the MBMS keys to terminals. A terminal requests a key when it needs to decrypt the data. This request may also be initiated by a message from the BM-SC to indicate a key update. The key management system is based on the use of SIM or USIM cards. SIM cards are commonly understood as the smartcard in 2G systems, and USIM as the smartcard in UMTS/3G systems. First, a bootstrapping procedure as defined by GBA (Generic Bootstrapping Architecture) [29] is executed. The bootstrapping procedure consists of several steps, during which the terminal and the BM-SC authenticate each other and agree on a shared secret, called MBMS User Key (MUK). Note that the MUK is never transmitted over the air. Two different variants of GBA exist, depending on the capabilities of the UMTS Integrated Circuit Card (UICC). The UICC is understood as the hardware smartcard containing either a SIM or a USIM application. In the UICC based version GBA_U, the MUKs are stored and further used for decryption of other keys within the UICC. In the terminal based version GBA_ME, the MUKs are stored and used in the device, but outside of the UICC. In both cases, the MUK is subsequently used to protect MBMS Service Keys (MSKs), which are transmitted

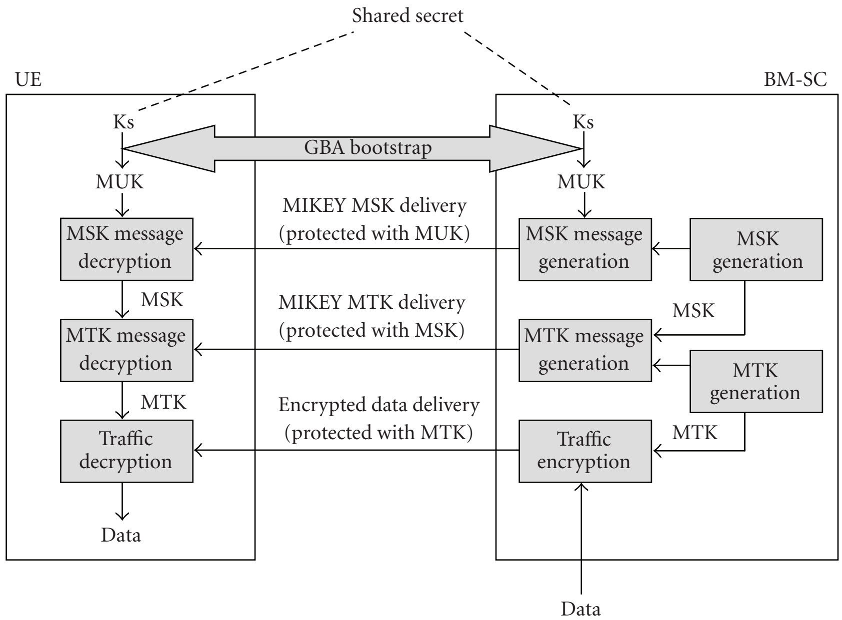


Figure 4: MBMS Security Overview.

in MIKEY (Multimedia Internet KEYing) messages [44]. MSKs can be either pushed to, or requested by the device. The MSKs are used to protect MBMS Traffic Keys (MTKs), which are also transmitted in MIKEY messages, interleaved with the actual content transmission on the MBMS bearer. The MTKs are finally the keys which protect the content streams and/or files. The keys that are used to protect the transmitted data in an MBMS User Service should be regularly changed to give potential hackers less time to extract and distribute them. This ensures that only legitimate users can decrypt the delivered data. In particular, frequent re-keying acts as a deterrent for an attacker to pass the MBMS keys to other users to allow them to illegitimately access the data in an MBMS User Service.

2.4. Phases in MBMS Service Delivery. In the following we will explain the main phases during MBMS service delivery. Figure 5 (taken from [30]) shows a timeline of the phases (from top to bottom), the entities (receiver, network, sender) which are involved in the phases, and how the phases are related to each other. Note that most of the network procedures are identical for both the broadcast mode and the multicast mode. Differences will be highlighted.

MBMS User Services (Sender and Receiver) use the MBMS Bearers (Network) in combination with unicast bearers, for example, for file repair or reception reporting procedures.

Subscription (Multicast Mode Only). Service subscription establishes the relationship between the user and the service provider, which allows the user to receive the related MBMS multicast service. Service subscription is the request of a user to receive the service(s) offered by the operator and that he agrees to a potential payment associated with the service. Subscription information is maintained in the BM-SC. The specification of how this agreement is obtained is out of the scope of the 3GPP specifications (it could be configured for instance on a web portal). The subscription should not confused with the actual registration of the end-users to a service. A service registration (e.g., security registration) is possible for both, the multicast and the broadcast mode.

Service Announcement. In this phase the terminal receives all information needed for the reception of a specific service, for example IP multicast addresses and associated port numbers, used codecs and codec configurations. Service announcements can be delivered over interactive (HTTP [27], MBMS and OMA-Push [28] bearers.

User Service Initiation. In this phase the receiver initiates the reception of a certain MBMS service. The MBMS service may use one or several MBMS Bearer Services. In case of an MBMS Multicast bearer, the initiation on the receiver triggers the "Joining" phase in the network. During this "Joining" phase the receiver becomes a member of the multicast group.

Session Start. During this phase, the sender requests the network to activate an MBMS data bearer. This happens independently of initiation or termination of the service by the receiver—for example, a receiver may activate the service before or after bearers for the MBMS service were activated. Furthermore, during this phase transmission resources for the upcoming MBMS data transfer are established.

MBMS Notification. In this phase the receiver is informed about upcoming and ongoing MBMS transmissions. In the radio access network a battery conserving signaling method is used for delivering those notifications to the receivers.

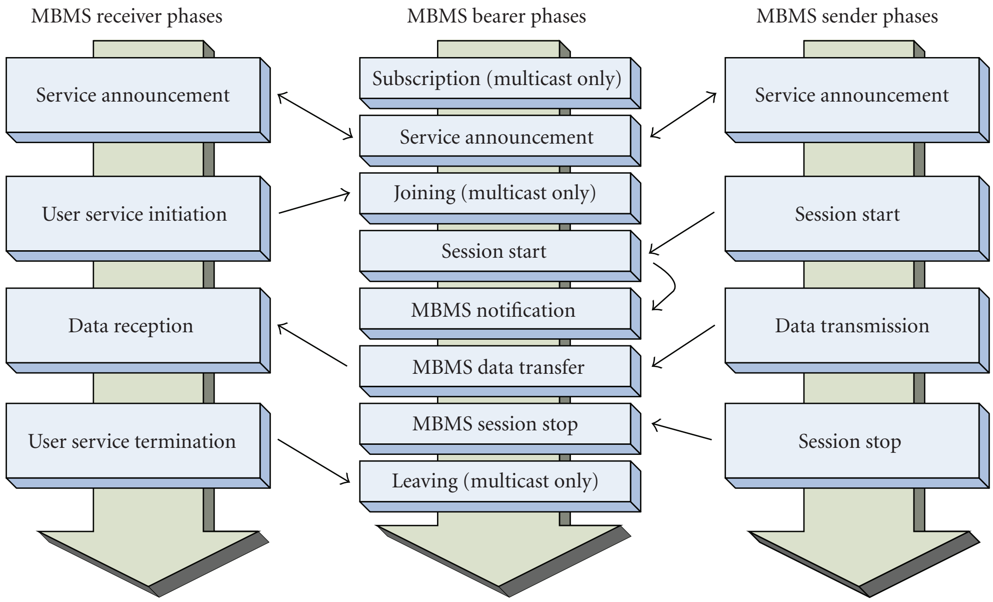


Figure 5: MBMS Phases.

Data Transfer. During the phase the actual data is transmitted to the terminal.

Session Stop. In this phase, the sender deactivates an MBMS data bearer. Consequently, the Radio Network may free the allocated transmission resources.

Service Termination. In this phase the receiver stops the reception of a certain MBMS service and its associated bearers. If reception of a multicast bearer is stopped the leaving phase in the network is trigged. During this phase the network removes the receiver from the multicast group.

This concludes the high level overview of MBMS. In the next section we will look at the various protocol layers in more detail. We will start with radio bearer realization and then move up in the protocol stack until we have reached the MBMS User Service layer. In addition we will also describe the complementing service enablers which were defined by OMA BCAST.

3. Radio Bearers for GSM/EDGE and UMTS/WCDMA

3GPP has defined MBMS radio bearer realizations for GSM/EDGE and UMTS/WCDMA. The MBMS radio bearer realization obviously has to differ for these radio access technologies. The present section provides an overview and also presents MBMS radio bearer performance results from simulations.

3.1. GSM/EDGE. In GSM systems, MBMS uses GPRS and EDGE modulation and coding schemes (that is, CS1-4 and MCS1-9 [1]). MBMS also uses the GPRS and EDGE packet data channel (PDCH) for MBMS P-t-M bearer, and the radio link control/medium access control (RLC/MAC) protocols on layer 2. MBMS bearers also support multi-slot operation. In this case, the radio network may use up to four timeslots per MBMS session.

Two enhancements have been introduced to further increase the performance of MBMS in GSM:

(i) RLC/MAC with automatic repeat request (ARQ)—also called Packet Downlink ACK/NACK (PDAN) mode. In this mode, ACK/NACK feedback is provided from up to 16 terminals in a given cell. This way, the RLC data blocks that a terminal did not receive correctly are re-broadcasted over the MBMS radio bearer so that the terminal can use incremental redundancy techniques to reconstruct an error-free data block from the rebroadcasts.

(ii) RLC/MAC without ARQ—also called blind repetition mode. In this mode, RLC blocks are repeated a pre-defined number of times, using an incremental redundancy technique, before the next RLC block is sent.

Simulations results [31] indicate that IP broadcasting at 40 kbps requires two to four timeslots with EDGE channel coding in PDAN mode depending on the number of users whereas four timeslots are needed in blind repetition mode. Note that a regular point-to-point EDGE channel can provide streaming at the same bit rate using two timeslots, but only to one user. For three users, six timeslots are required; for four users, eight, and so on. Therefore, if there are two users in the cell, the MBMS point-to-multipoint bearer is as efficient as a regular point-to-point connection and becomes much more efficient with an increasing number of users.

MBMS terminals will probably be based on existing EDGE hardware with a software update to support MBMS signaling procedures.

3.2. WCDMA. The radio bearers of UMTS are implemented using WCDMA technology. In WCDMA, MBMS reuses existing logical and physical channels to the greatest possible extent. In fact, the implementation in WCDMA requires only three new logical channels, called MBMS control channel (MCCH), MBMS traffic channel (MTCH), and MBMS scheduling channel (MSCH) and one new physical channel called MBMS notification indicator channel (MICH).

The MCCH carries details concerning ongoing and upcoming MBMS sessions. The MTCH carries the actual MBMS application data. The MSCH provides information on the data scheduled on MTCH. The MTCH can be configured with 40ms or 80ms interleaving depth (TTI). The selection of a longer TTI provides greater diversity in the time domain by spreading user data over the fading variations. This, in turn, yields improved MBMS capacity.

The MBMS notification indicator channel (MICH) is the new physical channel by which the network informs terminals of available MBMS information on MCCH. This approach was chosen in order to save battery power. A single bit contained in the MICH indicates the terminal whether there is new or changed information on the MCCH. As long as the information on the MCCH is not changed, the terminal does not need to receive the MCCH and can therefore shutdown the receiver to save battery power.

MCCH, MSCH and MTCH reuse the forward access channel (FACH) transport and secondary common control physical channel (S-CCPCH) in WCDMA.

P-t-P transmission within MBMS reuses the existing WCDMA shared or dedicated channel types and will be discussed in the following subsection.

3.2.1. P-t-P Realization Based on HSDPA. HSDPA [3] is designed to serve users in the most efficient way based on fast channel feedback from the terminal. Based on the feedback HSDPA avoids to schedule terminals when they are experiencing fading dips. Furthermore, the modulation, QPSK or 16QAM, and the channel code rate is dynamically selected to exploit to the greatest extend possible the theoretical instantaneous capacity of the radio channel perceived by the respective terminal. Finally, hybrid ARQ (HARQ) using incremental redundancy allows to limit the amount of transmitted channel coding redundancy to just the level that is instantly required.

For video streaming there is typically a small initial delay of about 1 second acceptable after the request for a particular stream has been send by the UE. This delay budget allows efficient channel dependent scheduling of the terminals and HARQ retransmissions; therefore a higher capacity can be achieved with HSDPA.

Performance simulations for a video streaming service have been reported in [32], for a typical network deployment with inter site distance of 1500 m in an urban environment with user speeds of 3 km/h. Table 1 summarizes those results and shows the Erlang capacity per cell sector for 95% satisfied users. A user is satisfied if in 99% of the cases a needed data packet is available from the streaming playout buffer. In other words, it means, a user is satisfied if a playout interruption occurs less or equal 1% of the time. Shown are results with and without receiver antenna diversity in the terminal ("RxDiv"). Terminal receive antenna diversity is not a requirement in the first MBMS release but is likely to be introduced during 2007/2008 in order to increase the availability of high data rates under locally varying coverage. It can be seen that receiver antenna diversity almost doubles the capacity. Apparently the capacity scales almost linearly with the data rate.

Table 1: P-t-P (HSDPA) capacity (for 95% satisfied users).

Data rate [kbps]	64	128	256
Capacity without RxDiv [Erlang]	24	10	4
Capacity with RxDiv [Erlang]	40	19	9

3.2.2. P-t-M Bearer Realization. In contrast to P-t-P radio bearers, a Point-to-Multipoint (P-t-M) bearer does not employ feedback. Since the enhancements provided by HSDPA rely on feedback from the terminal, HSDPA cannot be used for P-t-M bearers. The P-t-M transmission parameters need to be statically configured to provide the desired coverage in a given cell. The transmitted signal is lowest at the cell border and therefore the P-t-M bearer can greatly benefit from exploiting also the signals from adjacent cells transmitting the same service, that is, from soft-combining. For soft combining, it is required that transmissions are synchronized with an accuracy of 1 TTI plus 1 slot (= 0.66 msec).

Capacity results have been compiled by 3GPP [33] for radio bearers of 64 kbps for the case without and with soft combining assuming the terminal has capabilities to combine 2 or 3 radio links, one per nearby base station. The results are reproduced in Figure 6, which shows the percentage of the total transmit power (20 W) of a base station that is required to achieve a certain coverage percentage.

The results show that soft combining with 3 radio links significantly reduces the transmit power requirement by 6.5 dB. The capacity can be further increased by receiver antenna diversity in the terminals. Simulations for a different channel model (so called "3GPP case 3") have shown that terminals with antenna diversity require 3–5 dB lower signal to interference plus noise ratio.

From the fractional power requirements, the capacity in terms of the total number of simultaneous bearers can be calculated directly for a desired coverage percentage, taking into account that about 20% of the power needs to be reserved for control channels. For 95% coverage a total MBMS cell capacity of 1.7 Mbps can therefore be calculated assuming receivers with antenna diversity and soft combining with three radio links. Not shown in Figure 6 is that the required power scales approximately linearly with the bitrate up to the maximal rate of 256 kbps supported

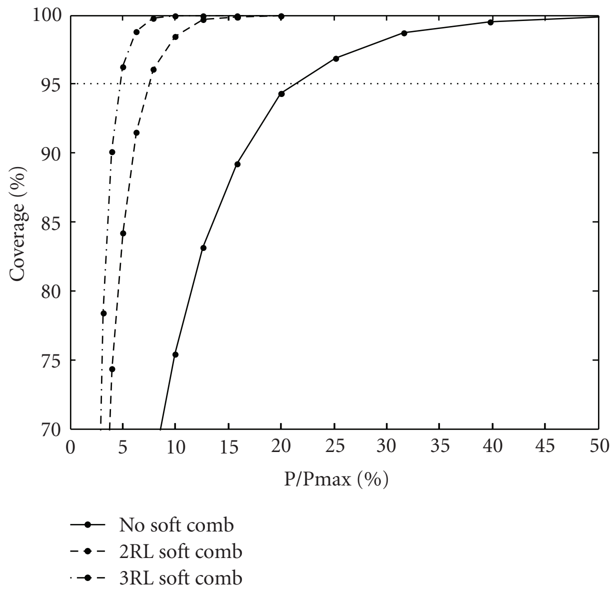


Figure 6: Estimated coverage versus fraction of total transmit power with Soft Combining (64 kbps, 80 ms TTI, 1% BLER).

per bearer. The total cell capacity can therefore be used for a flexible combination of 64, 128 and 256 kbps radio bearers.

4. MBMS Bearer Services in UMTS

MBMS in general provides two basic transmission modes for IP packets: broadcast mode and multicast mode (3GPP calls them "MBMS bearer services" [34]). In this section we will explain how those two modes are realized in UMTS. During this we will explain the "Counting" mechanism in UMTS and the possibility to adaptively select between the P-t-P or P-t-M WCDMA bearers mentioned in the previous section.

4.1. Broadcast Mode. A transmission in MBMS Broadcast mode can be received by all terminals who have activated the reception of the specific broadcast service locally and who are in the MBMS Services Area defined for that particular service. No information is sent in uplink direction. A terminal which likes to receive an MBMS broadcast service, just listens. The UMTS network has no information about active receivers in the MBMS Service Area. Reception of data sent in MBMS broadcast mode is not guaranteed. However, the receiver may be able to recognize data loss. The MBMS Broadcast Mode is very similar to existing Broadcast systems like DVB-T/DVB-H and can even be implemented using only downlink radio transmissions, that is, the uplink functionality of UMTS is not required.

4.2. User Counting for Enhanced Broadcast and Multicast Bearer Services. Within the UMTS Radio Access Network (UTRAN) the so-called "counting" or "re-counting" procedures can be executed to determine the number of terminals in each cell. "Counting" is initiated by the RNC as soon as the RNC needs to know the amount of active UEs that want to receive a specific MBMS service. This is used to determine the optimum transmission bearer, Point-to-Multipoint (P-t-M), Point-to-Point (P-t-P) or no transmission at all for a given MBMS service in the considered cell.

The need for counting is indicated on the MCCH channel per MBMS service. For the counting, the UEs need to reveal their interest to receive an MBMS transmission to the RNC. In case a large number of UEs wishes to receive the service, one has to avoid that all UEs start signaling at the same time, in order to prevent network overload. Therefore, a probabilistic response to the counting indication is used.

When the MBMS counting indication is received, the UE performs a random decision that with a probability equal to the probability factor contained in the counting indication message will generate a counting response from the UE, or with the complementary probability will not generate a counting response. In the latter case the UE continues monitoring the MCCH until the end of the MCCH modification period and if another counting indication is received repeats the random decision procedure. If the decision is to generate a counting response, the UE uses existing signaling procedures to indicate its presence (RRC connection establishment for idle UEs and Cell Update for UEs in URA-PCH state).

The MBMS counting procedure is terminated if the UE detects that the network has stopped sending the counting indication on the MCCH.

4.3. Broadcast Mode Enhanced with Counting of Users. A special mode in MBMS results from combining counting with the broadcast mode. In the following will use the term "Enhanced Broadcast Mode" (EBM) to refer to this mode. With EBM it is possible to switch to P-t-P bearers if appropriate or to completely stop the transmission in a certain cell if there are no (or not enough) terminals listening to the service. In this way EBM allows a more resource efficient delivery than the vanilla Broadcast Mode which would broadcast a service over a P-t-M bearer even if no terminal is listening.

Figure 7 shows the various situations which could occur if EBM is used. The content channels are delivered over the usual MBMS Broadcast bearers through the Core Network to all RNCs. Each RNC then makes for each cell it controls an intelligent decision, based on the outcome of the counting procedure, whether a particular content channel should be transmitted or not and if it should be transmitted whether a P-t-P or a P-t-M bearer is more efficient.

Note that EBM can be regarded as an optimization in the radio access network. It is fully transparent to higher protocol layers to which EBM appears just like the ordinary MBMS Broadcast mode.

4.4. Multicast Mode. In Multicast mode a terminals indicate interest to receive data of an MBMS Multicast service by "joining" the specific multicast group. The network keeps this "joining" state with the mobility management context in the Gateway GPRS Support Node (GGSN), Serving GPRS Support Node (SGSN) and Radio Network Controllers

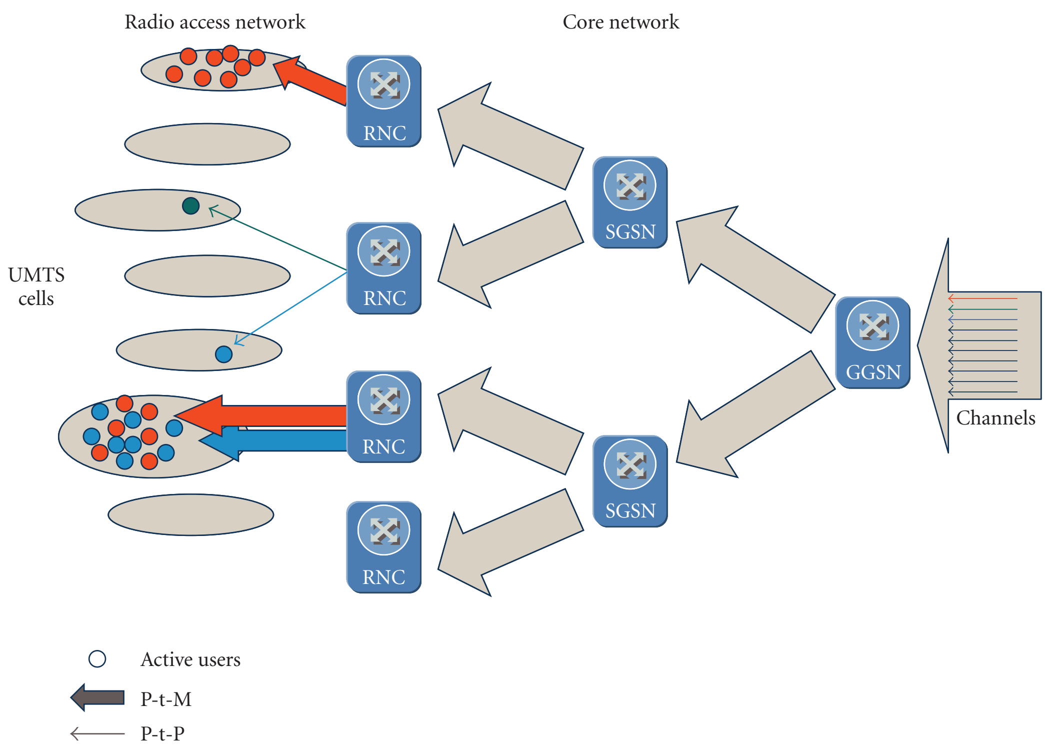


Figure 7: MBMS Broadcast Mode with "Counting" in UMTS.

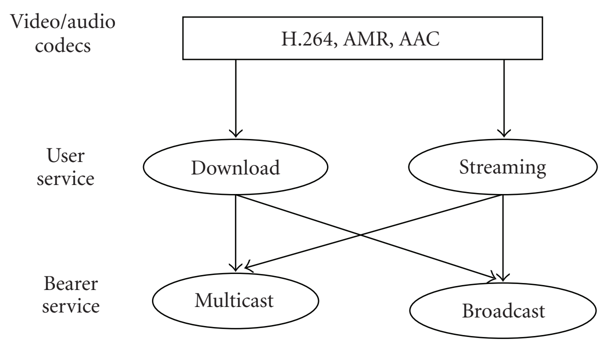


Figure 8: MBMS Bearer and User Services.

(RNCs). When the terminal moves from one area to another, the "joining" state for all joined services is also transferred to the new serving nodes.

5. Streaming and Download over MBMS

Whereas the MBMS bearer service addresses MBMS transmission procedures below the IP layer, the MBMS user services addresses service layer protocols and procedures above the IP layer.

The MBMS user service is a toolbox, which includes a streaming and a download delivery method. These delivery methods do not differ between or depend on the MBMS bearer services. Figure 8 shows that the MBMS User Services "Download" and "Streaming" sit on top of the MBMS Bearers Services which were explained in the previous section. Both the "Download" and the "Streaming" user service deliver media data encoded in various formats which will be explained in more detail in Section 6.

The MBMS download delivery method is intended to increase the efficiency of file distributions, including messaging services such as MMS. The download delivery method allows the error-free transmission of files via the unidirectional MBMS Bearer Services. The files are "downloaded" and stored in the local file system of the mobile phone. The streaming delivery method is intended for continuous receptions and play-out like in mobile TV applications.

Figure 9 shows the protocol stacks which are used for MBMS. The left hand side depicts that part of the protocol stack which requires an IP unicast bearer. The right hand side shows that part of the protocol stack which was designed for use over multicast/broadcast bearers and thus builds upon UDP. Since UDP packets can also be sent over unicast bearers, the right hand side of the protocol stack can also be implemented on top of a unicast bearer.

It can be seen that Service Announcements and other Metadata can be delivered both over unicast and broadcast/multicast connections. This means that a client can for instance download service announcement related information from a web page or it receives the information via a broadcast/multicast bearer similar to how ESG information today is delivered in broadcast-only services such as DVB-H. Unicast and broadcast/multicast delivery of Service

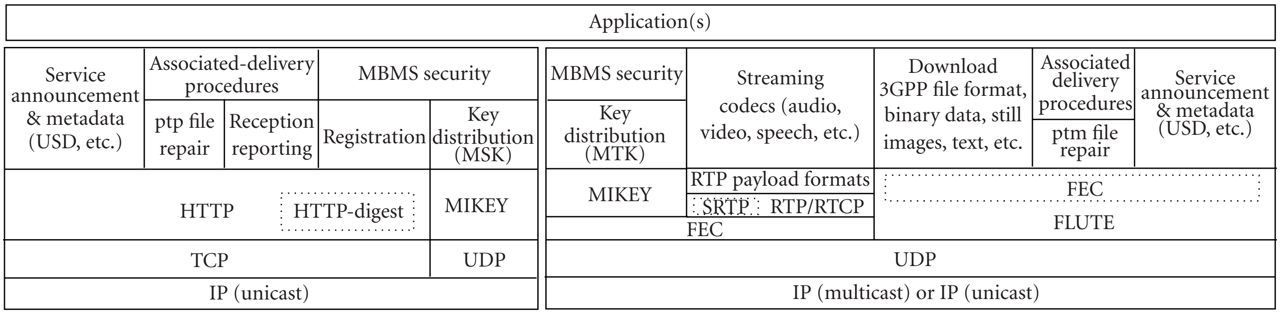


Figure 9: Protocol stacks used for MBMS User Services.

Announcement information can also be combined. For the Associated-Delivery procedures, certain procedures such as point-to-point file repair and reception reporting require a unicast connection whereas other procedures such as point-to-multipoint file repair (e.g., multicasting of missing packets) can be executed over a broadcast/multicast bearer. Similar holds for the Security functions. Some of them such as Registration and MSK distribution require a unicast connection whereas others such as MTK distribution work over broadcast/multicast bearers. The actual transfer of media data, that is, streaming and download, was designed for broadcast/multicast bearers but as it was said already above, it can also be done over a unicast bearer.

5.1. Download Delivery Method. The download delivery method is intended for file distribution services, which store the received data locally in a terminal. It can be used to deliver arbitrary files from one source to many receivers efficiently. Existing content-to-person MMS services, which deliver short video clips related to live events like a soccer match via MMS, will greatly benefit from this feature. Today, those services rely on point-to-point connections for MMS delivery. In the future the existing MMS sub-system can be easily interfaced with a BM-SC which then distributes the clip via MBMS download. The principle of an MBMS download delivery is shown in Figure 10.

The files are delivered during the MBMS data transfer phase (see Phase 1 in Figure 10). The MBMS transmission bearer is activated with the MBMS Session Start message. This message triggers the paging process in the radio access network, which inform the MBMS receivers about an upcoming transmission. After the MBMS bearer is successfully established the BM-SC starts sending the actual MBMS download data. Included in the download data is a file containing the File Delivery Table (FDT). The FLUTE [35] protocol is used to send the files via UDP. The FLUTE protocol allows the FEC protection of the files and it uses the IETF FEC framework [36]. After the MBMS data transmission, bearer resources are released via the MBMS Session stop message.

During the MBMS data transfer phase, certain mobile phones may experience packet losses due to fading conditions or handovers. Naturally, full reliability cannot be offered in a pure unidirectional distribution scheme because the packet loss rate can be excessive for some users.

Three packet error recovery schemes are foreseen for the download delivery method. The most important one is the use of FEC coding, which allows recovery of lost packets without any server interaction.

The Raptor forward error correcting code [37] was chosen as a basis for FEC scheme [38]. A broadcast of newly created FEC packets is of benefit for all receivers, which have not successfully reconstructed the original source block. Generally, Raptor codes can handle even large files as one source block. But since mobile phones have a limited amount of fast memory for decoding, a single source block for 3GPP release 6 receivers may only contain up to 4100 kByte of data. Thus, larger files are subdivided into a number of source blocks and the FEC repair symbols are generated for each source block. If an interactive bearer is used for the file repair procedure, the repair data is independently sent to different receivers and can even be tailored to the actual losses of that receiver. On the contrary, if the MBMS bearer is used, the same repair data is sent only once to multiple User Equipments (UEs) and the repair data should be useful for all receivers with losses. Therefore, the rateless property of the Raptor code, for example, the possibility to generate an arbitrary number of FEC redundant symbols out of one source block, is very beneficial for the PTM repair mechanism.

Beside FEC protection, MBMS offers two additional types of file repair procedures: one is using interactive bearers and another one uses MBMS bearers. In case of file repair, the MBMS client waits until the end of the transmission of files or sessions and identifies then the missing data from the MBMS download. Afterwards, it calculates a random *back-off time* and selects a file repair server randomly out of a list. Then, a *repair request* message is sent to the selected file repair server at the calculated time. The file repair server responds with a *repair response* message either containing the requested data, redirecting the client to an MBMS download session, redirecting the client to another server, or alternatively, describing an error case. The BM-SC may also send the repair data on a MBMS bearer (possibly the same MBMS bearer as the original download) as a function of the repair process. The performance of the post-delivery file repair procedures described above has been analyzed in [39, 40].

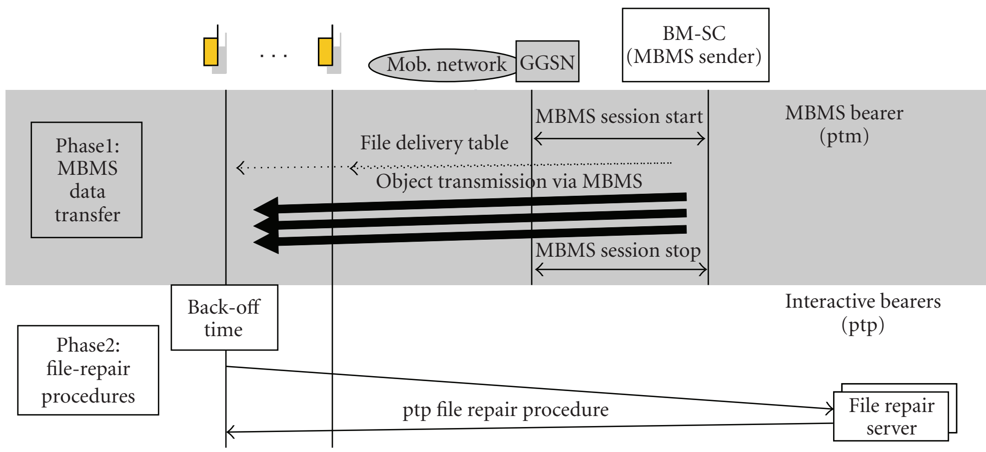


Figure 10: MBMS Download procedure principle with point-to-point file repair.

5.2. Streaming Delivery Method. The streaming delivery method is intended for the continuous reception and play-out of continuous media like video, audio or speech. A typical example of an application using the streaming delivery method is the transfer of live channels in a mobile TV service.

RTP [41] is used as transport protocol for MBMS streaming delivery. RTP provides means for sending real-time or streaming data over UDP (see Figure 9).

As previously mentioned packet losses could occur during the streaming data transfer that would results in distortions of the received video and audio quality at the receiver side. In case of streaming with its real-time constraints file repair procedures are not suitable. Instead, forward error correction (FEC) is used [42].

As for download delivery the Raptor forward error correcting code [37] may be used to increase bearer reliability for MBMS transmissions. The Raptor code belongs to the class of rateless codes and can thus generate an arbitrary number of FEC redundant symbols out of one source block. The FEC stream bundling concept shown in Figure 11 allows the protection of the actual audio/video data together with synchronization information (RTCP) and possibly decryption information. Packets of one or more UDP flows may be used to construct the source blocks for the FEC protection. The FEC redundancy information is transmitted in one separate traffic flow. Since the Raptor code is a systematic FEC code, the receiver can simply ignore the FEC flow, if no transmission errors occur.

The advantage of the FEC stream bundling concept, as shown in Figure 11, is that the FEC efficiency is increased when protecting several data flows together, because the FEC code works on a larger portion of data.

Figure 12 depicts how one or more out of several possible packet flows of different types (Audio, video, text RTP and RTCP flows, MIKEY flow) are sent to the FEC layer for protection.

The source packets are modified to carry the FEC payload ID and a new flow with repair data is generated. The receiver takes the source and repair packets and buffers them to perform, if necessary, the FEC decoding. After appropriate buffering received and recovered source packets are forwarded to the higher layers.

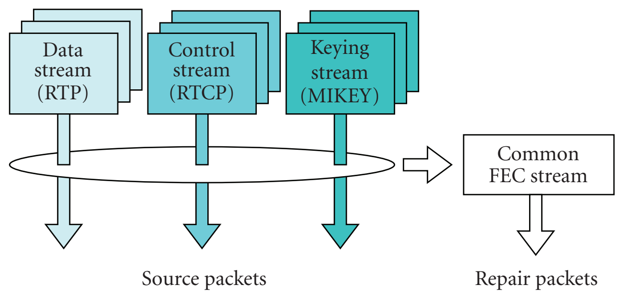


Figure 11: FEC Stream Bundling Concept.

MBMS streaming delivery may also use reception reporting, that is, reports about received parts of the service that are sent back to the server. The reception reporting procedure allows an operator to collect reception statistics and usage patters of the service from actual end-users.

The MBMS User Service framework is harmonized with OMA BCAST and DVB CBMS specifications. In particular the same set of audio/video codecs is supported. Therefore, for a given rate MBMS is able to deliver the same audiovisual quality as DVB-H.

6. Media Codecs

MBMS supports different media types, among which Speech, Audio, Video and Timed Text are ones supported both for MBMS Download and Streaming delivery. In addition, MBMS defines codecs for synthetic audio, still images, bitmap graphics, vector graphics, and text for use in MBMS Download services only.

In what follows we will summarize the codecs which have been specified for MBMS Download and Streaming services. The complete set of supported codecs can be found in [15].

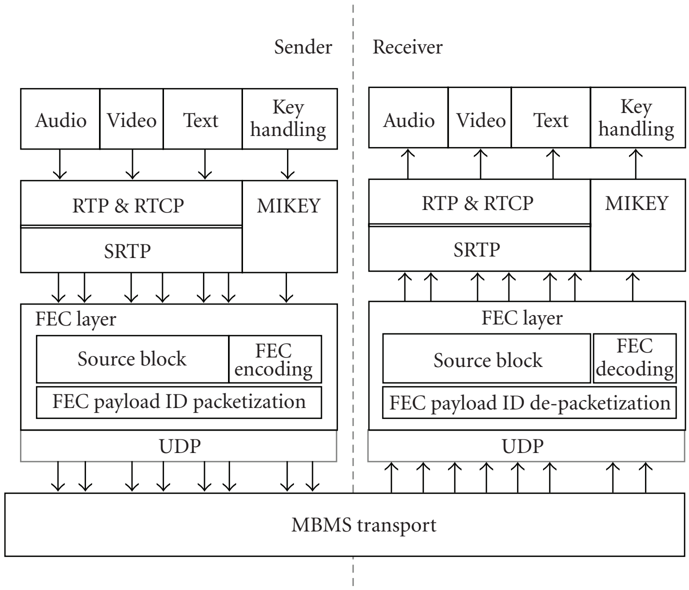


Figure 12: FEC mechanism for the streaming delivery method interaction diagram.

Table 2: Audio codec preference for different mono audio content types (Derived from characterization results in 3GPP TR 26.936. 24 and 32 kbps results from ITU-T (14 kHz band limited audio)).

Bit Rate (kbps)	Music	Mixed	Speech
~10	AMR-WB+	AMR-WB+	AMR-WB+
~16	AMR-WB+	AMR-WB+	AMR-WB+
~20	AMR-WB+	AMR-WB+	AMR-WB+
24	(AMR-WB+)		AMR-WB+
32	(AMR-WB+)		(AMR-WB+)

Table 3: Audio codec preference for different stereo audio content types (Derived from characterization results in 3GPP TR 26.936).

Bit Rate (kbps)	Music	Mixed	Speech
~16	(E-ACC+)	(E-ACC+)	AMR-WB+
~21	E-ACC+	(AMR-WB+)	(AMR-WB+)
~28	E-ACC+	(AMR-WB+)	AMR-WB+
~36	E-ACC+	(AMR-WB+)	(AMR-WB+)

For speech, one can choose between the AMR decoder for narrow-band speech and the AMR wideband speech decoder (AMR-WB) for wideband speech working at 16 kHz sampling frequency.

For audio coding, Enhanced AACPlus (E-AAC+) and AMR-WB+ codec can be used.

Within 3GPP the two supported audio codecs were tested for different type of audio content at different coding rates. These characterization results are summarized in Tables 2 and 3.

For video coding, MBMS uses H.264/AVC [6, 7]. Since dynamic negotiation of media codecs is not feasible in broadcast/multicast services, H.264/AVC is the only new video codec recommended for MBMS.

The current version of MBMS specifies the Baseline Profile at Level 1.2. In this profile/level combination, video content can be encoded up to QVGA resolution (320 × 240 pixels) and up to 30 fps.

In addition to speech, audio, and video, MBMS also supports "Timed Text". Timed text is text that is rendered at the terminal, in synchronization with other timed media such as video or audio. Timed text is used for such applications as closed captioning, titling, and other visual annotation of timed media. Timed text for MBMS reuses the Timed Text specification [43] which was specified originally for packet-switched streaming [4].

7. Service Layer Components

MBMS defines all components that are necessary to transmit audiovisual content to MBMS capable devices. Yet, MBMS does not include all service layer functionality known from cable or satellite TV networks. Most notably, MBMS does not specify a service guide or program guide that describes the available programs or channels in a user friendly way. Further, interactivity must be closely integrated into the service and its architecture, to allow easy and user-friendly access. Also, service access protection and content protection are favorably closely integrated into service guide and service purchase or subscription mechanisms. Thus, some additional functionality is needed in addition to MBMS for a complete and user friendly service offering.

The Open Mobile Alliance (OMA) has defined such a set of functionalities, called the "Mobile Broadcast Enabler", or abbreviated BCAST 1.0. In detail, the following functions are defined in BCAST 1.0: service guide, file and stream distribution protocols and associated mechanisms like fire repair and forward error correction (FEC), notification function, service and content protection, service interaction, service provisioning, terminal provisioning, roaming and mobility support, and specification on back-end interfaces for each function. BCAST 1.0 is a generic set of functions. However, in order to better integrate with broadcasting systems, OMA has defined adaptations to some specific broadcast systems. These include IPDC over DVB-H, 3GPP2 BCMCS and, notably here, 3GPP MBMS. Thus, BCAST 1.0 can be nicely combined with MBMS, giving a combined overall system for mobile TV services.

Two of the functions mentioned before are of special importance and are explained in some more detail: the service guide, and service protection. The BCAST service guide (SG) is a central component of the BCAST enabler. It consists of information describing the service and programs, and has some parts targeted to the end-user, and other parts targeted to the device. The SG is encoded in an XML-based format, and is structured into information units, so-called fragments. There are separate fragments that describe the service and program schedule, the technical details required to "tune in" and receive the service, interactivity, additional information for the user, and purchase and delivery related information. Figure 13 shows the logical structure of the BCAST SG. However, many of the fragments are not mandatory to be used for a specific service. A minimalistic SG that describes a continuously ongoing TV channel would for example only need a "service" fragment (describing the service, e.g., the channel name) and an "access" fragment (describing the technical parameters to receive the service, like codec and IP port information). In general however, more detailed information describing the sequence of programs is desired, making the used SG more complex. SG fragments are bundled and packaged into a specific container, called a service guide delivery unit (SGDU). Further, the totality of all current fragments and their grouping into SGDUs for a service is described in a so-called Service guide delivery descriptor (SGDD). SGDUs can be delivered over broadcast/multicast, where they are typically frequently repeated, or on demand over unicast. Delivery over broadcast takes away resources from the service delivery and is thus only preferred in systems that cannot rely on a unicast channel for interaction. In 3G systems however, it is advantageous to deliver the SG on demand, when a device or its user decides that it needs additional information, for example, in case the information present on the device contains only information for a short time of the future. The service guide is dynamic in nature; on-the-fly changes and updates of the SG and its fragments are possible.

Service protection is another important functionality. In fact, service protection is closely coupled to the overall control over service delivery and end user billing; the value chain participant that controls the service protection usually also sends the bill to the end user. Therefore, service protection does not only have a technical role, it is also closely connected to the business control and business model. In order to be more flexible, BCAST has defined two different solutions for service protection, called profiles: the Smartcard profile and the DRM profile. Both profiles follow however the same principle of a key hierarchy, as shown in Figure 14. Key hierarchy means that the audio and video media streams are not directly encrypted with just one key. Rather, the media streams are encrypted with a short term key, which changes frequently (e.g., every 5 minutes) and is broadcasted alongside with the media streams. The short-term key itself (corresponding to the previously mentioned MTK in MBMS) is encrypted with a long-term key. The long-term key (corresponding to MSK) changes less frequently (e.g., once a day or once a week) and is transported to the device in a point-to-point communication. The long-term key itself is encrypted with a shared secret (corresponding to MUK) that somehow has been agreed between sender and receiver beforehand.

The OMA BCAST Smartcard profile is derived from MBMS security and GBA as described above, but provides some extensions. In the Smartcard profile, the shared secret (Layer 1 (L1) as shown in the figure) is based on a key stored in a SIM or USIM or ISIM and also known in the network. From that stored key, the shared secret used for the communication is derived in a bootstrapping procedure. The long-term keys (L2) and short-term keys (L3) are sent in messages following the MIKEY [44] format. The media streams (L4) are streams transported using encrypted transport protocols. BCAST supports either SRTP [19] or ISMAcryp [45] or IPSec [46].

In the DRM profile, the shared secret (L1) is based on DRM public-key certificates issued by a central authority to both sender and receiver, and which are mutually authenticated in a DRM registration procedure. The long-term keys (L2) are sent in DRM Rights Objects, which is a known container format for keys (and permissions to use content). The short-term keys (L3) are sent in a special BCAST message format. As for the Smartcard profile, the media streams (L4) are streams transported using SRTP [19], ISMAcryp [45] or IPSec [46].

The specification of OMA BCAST was finalized in June 2007. First interoperability tests have been done by the companies participating in the bmcoforum [47] in the same month.

8. Use Case: Mobile TV over MBMS in UMTS

In this section we summarize simulation results from [48] showing the capacity improvements which MBMS will add to a UMTS network. We have assumed an MBMS configuration which uses broadcasting combined with "Counting", commonly referred to as "Enhanced Broadcast Mode" (see Section 4.3).

We assume a mobile TV service with 20 mobile TV channels, each is delivered at 128 kbps. The channel interest is not uniformly distributed. We assume that Channels 0 & 1 are selected each in 25% of the cases, Channel 2 in 15% and Channel 3 in 10% of the cases. The remaining 16 channels are selected with an equal probability of 1,56%. For simplicity we assume that the channel interest does not change during the simulation period.

The mobile TV usage model is derived from statistics on usage and viewing patterns obtained during the "Finpilot" study [49]. In our simulation we assume that each user

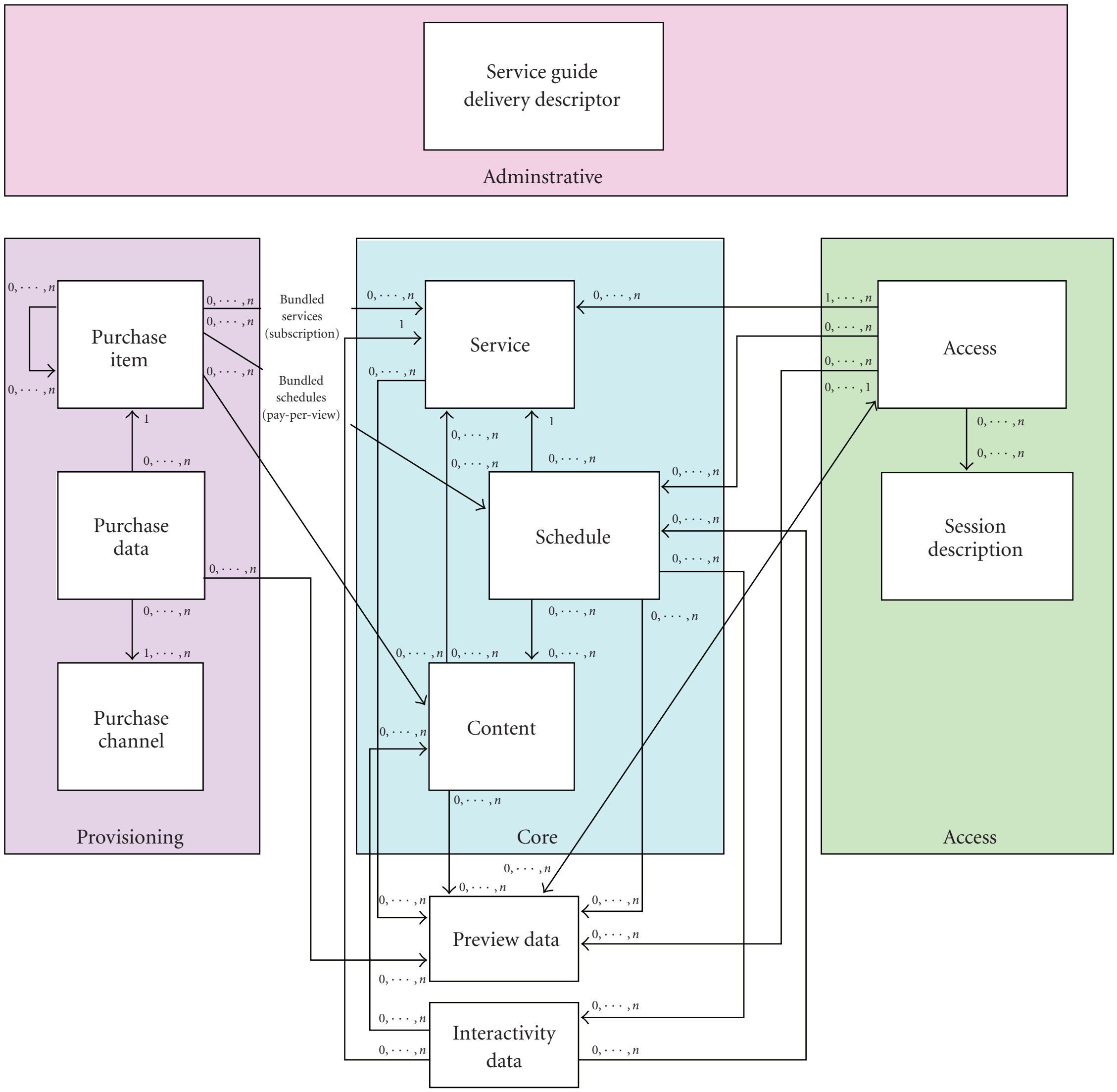


Figure 13: Logical structure of OMA BCAST service guide.

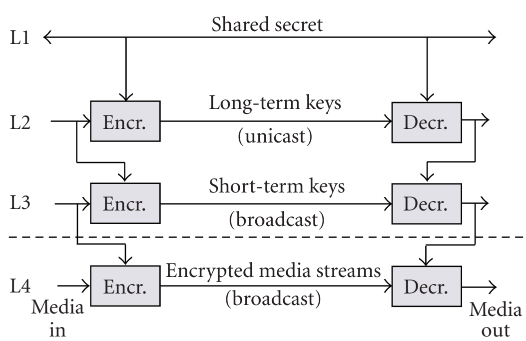


Figure 14: 4-layer key hierarchy.

is watching mobile TV for a total of 30 minutes within 12 hours. The session length is exponentially distributed with a mean of 12.6 minutes.

The user may switch between different channels during the mobile TV session. The user watches a channel for an exponentially distributed time with a mean of 3.25 minutes. The user switches off the mobile TV reception as soon as the end of the session is reached.

A cell cluster of 48 cells is used as cellular network simulation topology. We further assume that the MBMS Enhanced Broadcast Mode (see Section 4.3) is used. In this mode, an RNC can decide based on information obtained through counting whether a point-to-point or a point-to-multipoint bearer is more efficient. We further assume that the point-to-point bearer is realized as an HSPA bearer and that the point-to-multipoint broadcast bearers requires twice as much radio resources compared to an HSPA bearer. We further assume that each cell has capacity to serve up to 16 simultaneous users with point-to-point HSPA. When point-to-multipoint transmissions are used, then the available point-to-point capacity is reduced accordingly.

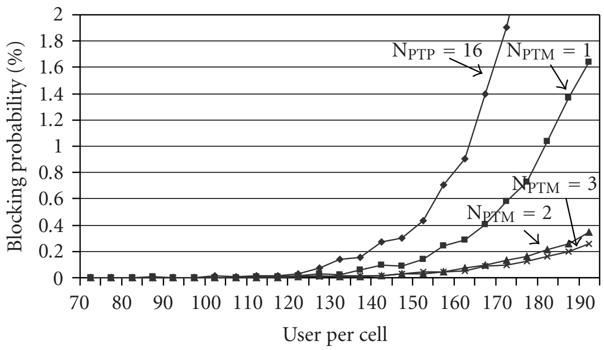


Figure 15: Blocking probability over user density for delivery of 20 TV channels over MBMS in a UMTS network.

The mobile TV subscribers are randomly distributed in the cell cluster. We varied the user density between 60 and 190 users per cell. Note that in urban areas the mobile subscriber density is typically around 300 users. However, the addressable market for mobile TV services today is much lower than 100% since not all users are interested in the service and not all users have TV enabled phones. Market research today indicates an addressable market of 10%–15%. By limiting ourselves to 190 mobile TV subscribers we can still capture a market take-off over the coming years.

The resulting channel usage per cell for different channel selection probabilities is depicted in Figure 15. It can be seen that for 140 users per cell the probability that one of the low interest channels 4 to 19 is watched is very low. Channel 0 (25% channel selection probability) is used on average by 1.45 users per cell, whereas channels 4 to 19 has 0.09 users per cell. The maximal number of simultaneous watcher per channel is of course higher.

In Figure 15 we show blocking probabilities for various network configurations. The case where all 20 channels are broadcasted is not shown since the blocking probability in that case is zero at the expense of excessive radio resource consumption: 20 always on broadcast bearers in every cell.

The case $N_{PTP} = 16$ corresponds to unicast only transmission, using HSDPA. From 120 mobile TV subscribers per cell onwards the blocking probability starts to increase and reaches 1% at 160 subscribers per cell. One can see that the blocking probability can be considerably decreased by introducing only a few broadcast bearers. With 2 broadcast bearers ($N_{PTM} = 2$) the blocking probability stays well below 0.5% even for 190 mobile TV subscribers per cell.

A user density of 190 users per cells corresponds to the case where roughly 30% of the total subscriber base is subscribed to the mobile TV service.

9. MBMS Evolution

MBMS evolution in 3GPP 3G networks from Release 6 is part of the two main tracks of standard developments, as depicted in Figure 16:

(i) evolution of WCDMA,

(ii) 3G long term evolution of UTRA, called 3G LTE (denoted by E-UTRA in the 3GPP specifications).

The evolution of WCDMA comprises improvements for MBMS for the P-t-M radio bearers, mainly the introduction of so called MBMS single frequency networks (MBSFN), and HSPA evolution that allows for higher capacity using P-t-P radio bearers. In 3GPP Release 7 also support for unicast bearers like Streaming [4, 5] or OMA Push [28] as part of an MBMS User Service has been added. The use case which has driven the integration of unicast bearers into MBMS is access to an MBMS service in the home network from a visited network which does not support the same MBMS services as in the home network or which does not support MBMS at all.

In a parallel track 3GPP has standardised the long term evolution of UTRA, called 3G LTE (denoted by E-UTRA in the 3GPP specifications). LTE is part of 3GPP Release 8 and defined in the 36-series of specifications. Originally Release 8 LTE was planned to include enhanced MBMS and significant standardization effort has been made in this area. However, beginning of 2008 MBMS was shifted out of Release 8 and into Release 9. By December 2009 the specification has been ranked 95% complete by 3GPP.

The LTE MBMS related aspects addressed here capture the status of 3GPP agreements at the time of the shift, as of [18].

The main goals for LTE are [50]:

(i) improved spectrum flexibility,

(ii) increased cell edge and cell average throughput,

(iii) efficient support of MBMS,

(iv) simplified architecture.

The most important innovations in LTE with respect to MBMS are the support of larger carrier bandwidths, use of OFDM modulation in the physical layer and efficient and flexible support for MBSFN.

The MBSFN principle is the same in both evolution tracks and is therefore first introduced hereafter for both in common.

9.1. MBMS Single Frequency Networks. If the same multimedia service shall be provided over geographical areas comprising several cells or even an entire nation, advantages can be taken from the inherent single frequency operation of the WCDMA and LTE, that is, all cells can use the same carrier frequency and the terminal is therefore inherently able to receive the signals from several cells simultaneously [51]. If different cells transmit different information, then the signals interfere in the terminal receiver. Whereas this cannot be avoided for user-individual services, it is possible to avoid intercell interference for broadcast services by sending the same signal from all cells in the service area at the same time. By the elimination of intercell interference, a dramatic increase in spectral efficiency and thereby throughput on the used radio resources is achieved.

This efficient exploitation of the single frequency network nature for MBMS is called MBSFN in 3GPP. MBSFN operation is currently being specified by 3GPP as an

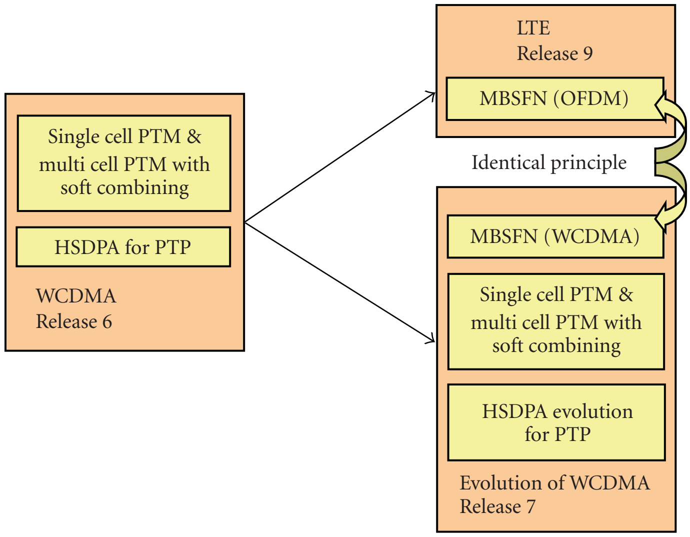


Figure 16: MBMS evolution tracks. Releases refer to 3GPP releases.

enhancement for WCDMA. MBSFN is also a main driver in the 3GPP work item on the long term evolution (LTE) (see next section).

If an entire carrier frequency is dedicated to broadcasting using MBSFN transmissions, such dedicated MBMS carrier obviously does not need to be matched with an associated uplink carrier. Therefore there is an opportunity to operate dedicated MBMS carriers also in unpaired frequency bands. Among potentially suitable frequency bands there are the following:

(i) UMTS core band: 1900–1920 MHz and 2010–2025 MHz—designated for UMTS-TDD operation that have been licensed but are currently unused in most countries.

(ii) UMTS extension band: the range 2570–2620 MHz is in between the UMTS FDD uplink and downlink. The decision on the usage is up to national administrations in Europe. The upper part of this range that is adjacent to the FDD downlink could be used for MBMS dedicated carriers, without coexistence problems between the MBMS dedicated carriers and FDD downlink.

(iii) Part of the terrestrial TV broadcasting band 470–862 MHz. Due to the transition from analog TV to the more spectrally efficient digital TV, it is widely accepted that part of this band can be vacated from terrestrial fixed TV broadcasting services. This so called “digital dividend” could be used to introduce mobile TV services based on MBMS.

9.2. Evolution of WCDMA

9.2.1. MBSFN. For WCDMA, 3GPP has introduced MBSFNs in Release 7 for FDD and TDD in 2007. In Release 8 this has been further developed into Integrated Mobile Broadcast (IMB) that is designed to operate in unpaired spectrum. IMB principles are identical to those for MBSFN FDD, except that a new synchronization code has been added to resolves potential cell search issues of legacy TDD terminals.

In the conventional WCDMA downlink, all the physical channels of a cell, except for the synchronization channels, are scrambled by cell-specific primary or secondary scrambling codes. While scrambling codes are used to differentiate cells, orthogonal spreading is used to separate multiple code-division multiplexed channels within a cell.

With orthogonal spreading, a receiver does not experience own-cell interference when the base station signal travels through a flat channel. State of the art receivers are capable of removing own-cell interference even in the case of non flat channel, using, for example, linear MMSE chip equalizer or G-Rake techniques. Due to lack of signal orthogonality; however, these receivers do not suppress other-cell interference as effectively.

In contrast, in the MBSFN transmission mode the same scrambling is used for all participating cells. When different base station uses the same waveform to simulataneously send a common set of MBMS content channels, the received signal is then the same as that for a single-source transmitted signal traveling through a heavily dispersive channel, where each path corresponds to a signal path between a base station transmitter and the receiver as illustrated in Figure 17. In this case, the other-cell signals share the same orthogonal properties as the own-cell signals. As a result, an advanced receiver (e.g., zero-forcing equalizer) not only collects the signal energy contributed by the multiple base station signals, but also gets rid of interference arising due to multipaths and transmission from multiple base stations. In order to exploit the resulting very high SINR, Release 7 adds support for 16QAM on the S-CCPCH on which an MTCH is mapped. This boosts the spectral efficiency of MBMS in a WCDMA network significantly.

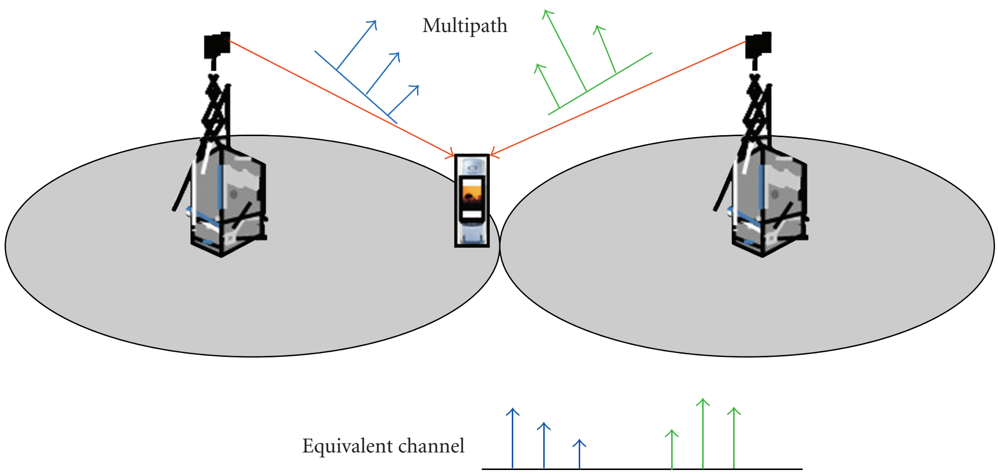


Figure 17: Principle of MBMS transmission with a common transmitted waveform.

In order to limit the receiver complexity, the delay spread needs to be limited and therefore the cells in the MBSFN need to transmit the same signal at the same time.

Synchronisation on the order of a few micro seconds is required. The synchronisation of NodeBs connected to one RNC is ensured by the RNC. As a consequence, synchronisation is not ensured accross RNCs and MBSFN across RNCs are not supported.

In a proposed application of the MBSFN mode, cells have multiple carriers of which some operate in the MBSFN mode using cell common scrambling and some operate in the normal 3GPP Release 6 mode using cell specific scrambling. Multiple MBMS physical channels can be code-division multiplexed on this dedicated carrier. However, the same channelization code is used in all the base stations to transmit the same MBMS content channel. Furthermore, of course the usual 3GPP Release 6 time multiplexing of MBMS radio bearers is possible.

Table 4 reports MBSFN spectral efficiency results from [52] obtained from radio network simulations with 2800 m inter site distance, 3 sectors per site, urban area propagation environment, with channel model "Vehicular A" and users moving at 3 km/h. The performance criterion is the spectral efficiency achievable with 95% coverage and 1% BLER.

Results are shown for receivers of 3GPP Type-2 (implementing G-RAKE or equivalent techniques) and 3GPP Type-3 (like Type-2 plus implementing antenna diversity). With Type-3 advanced receivers that can equalize, that is, take advantage of, the signals from the 7 closest NodeBs spectral efficiency in excess of 1 b/s/Hz can be achieved if 16QAM is employed. For less advanced Type-3 receivers or in case the inter site distance is significantly larger, so that only the signals from 3 NodeBs can be equalized, the spectral efficiency is still 0.6 b/s/Hz.

The higher cell capacity also allows for higher bearer data rates, thus the maximum supported bearer data rate has been increased to 512 kbps. As the cell capacity is several times higher than 512 kbps, time sliced transmission of services is enabled, where each services is mapped to one 2 ms subframe in a radio frame. This significantly improves the battery saving in the receiver.

9.2.2. HSPA Evolution. HSPA comprises HSUPA and HSDPA. As has been described in Section 3.2.1, HSDPA can be used for MBMS P-t-P radio bearers. For the HSDPA evolution, 3GPP is standardizing 2 × 2 MIMO and an increase in the modulation order from 16QAM to 64QAM for 3GPP Release 7, in order to drastically increase the per user data rates and system capacity.

The MIMO scheme chosen for HSPA evolution employs multiple codeword allowing for transmitting two separately encoded streams to a UE. Hence the data on each stream is separately encoded, modulated and spread. The up to 15 spreading codes (of spreading factor 16), available for HSDPA, are reused over both streams. Before transmitted on the antennas, the modulated and spread signal is spatially weighted (pre-coded) using a unitary transform. The weights are taken from the same codebook as used for closed-loop transmit diversity mode 1. For the link adaptation, the UE reports the number of streams, the spatial weight (pre-coding index) and transmission rate that it prefers. HARQ is operating independent between the streams.

Release 6 HSPA systems support the use of 16QAM in the downlink. For indoor or small-cell system deployments, higher SNRs and thereby 64QAM modulation can be supported. The maximum data rate per user achievable with 64QAM is 21.3 Mbps. While the combination of MIMO and 64QAM is not included in Release 7, it is being considered for future releases.

Section 9.4 presents some performance results for evolved HSDPA, together with results for P-t-P transmissions in LTE.

9.2.3. Flat Architecture. 3GPP has standardized an optional flat architecture for WCDMA in Release 8, as illustrated in Figure 18. The RNC functionality can be integrated into the Node-B and IP multicasting in the core network is supported down to the NodeBs. Functionality and signalling

Table 4: MBMS capacity and spectrum efficiency achieved by SFN operation (using 90% of the cell power) for receivers of Type-2 (implementing G-RAKE or equivalent techniques) and Type-3 (like Type-2 plus implementing antenna diversity).

Receiver	Receiver capable of equalizing signals from 3 NodeBs	Receiver capable of equalizing signals from 7 NodeBs
Type-2	1.54 Mbps 0.31 b/s/Hz	2.62 Mbps 0.53 b/s/Hz
Type-3	3.01 Mbps 0.60 b/s/Hz	5.38 Mbps 1.08 b/s/Hz

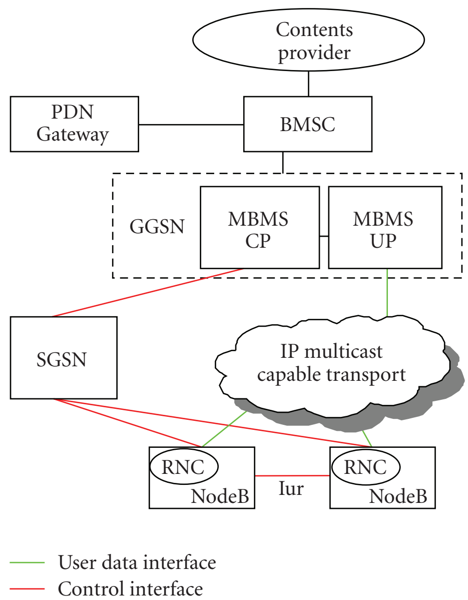


Figure 18: Architecture for MBMS in SAE/UTRA flat architecture.

via the Iur interface has been added to enable the necessary coordination between NodeBs, in particular synchronisation between NodeBs for MBSFN that in the classical architecture is performed by the RNC.

9.3. MBMS in LTE. MBMS in LTE uses an evolved architecture, in order to support MBSFNs with high flexibility, which was an important design goal of LTE from the start. An MBMS architecture is desired that supports the tight synchronisation of broadcast content transmit times across many cells in the network. Furthermore, for LTE it is desired to support MBSFN transmission and user individual services on the same carrier. The architecture needs to support the coordinated allocation of radio resources within the carrier for MBSFN transmission across all cells participating in the particular MBSFN

Figure 19 shows possible candidate architecture that has been discussed in 3GPP. The architecture is based on enhancements to the architecture of earlier releases.

The following logical entities are proposed:

(i) *E-MBMS GW*: entity that is located between the content provider and the eNode Bs. The MBMS GW will be involved in the MBMS session start/setup, and will also participate in the content synchronization for MBMS services using MBSFN (i.e., the mUPE functionality defined in 3GPP). This entity is expected to be part of the Evolved Packet Core. The detailed functionality of this logical entity is in the scope of SA2/CT work, for example, it is FFS if this entity would be split in a user plane and control plane part.

(ii) *MCE:* entity responsible for coordinating the usage of MBSFN transmission in the LTE RAN. This entity is expected to be part of the LTE RAN. The MCE will be responsible for all the eNode Bs in a given area.

Furthermore it is proposed to introduce the following logical interfaces:

(i) *M1* is a logical interface between the MBMS GW and the eNode Bs. The transport on this interface will be based on IP multicast. The MBMS content will most likely need to be transport in some framing or tunneling protocol, in order to support content synchronization and other functionalities. IP multicast signaling will be supported in the TNL layer in order to allow the eNode Bs to join an IP multicast group.

(ii) *M2* is a logical control interface between the MCE and the eNode Bs. This interface is used to coordinate the setting up of MBMS service in the eNode Bs for MBSFN operation. It is expected that the MCE will have a master role in the coordination, while the eNode Bs provide feedback to the MCE. It is expected that M3 will have the same signaling transport layer as S1/X2.
It is not precluded that M3 interface can be terminated in eNBs. In this case MCE is considered as being part of eNB. Therefore M2 does not exist in this scenario. This is depicted in Figure 19 which depicts two envisaged deployment alternatives. In the scenario depicted on the left MCE is deployed in a separate node. However, in this case the MBSFN controlled by the integrated MCE can only consist of cells belonging to the eNB hosting the MCE. The scenario on the right MCE is part of the eNBs

(iii) *M3* is a logical control interface between the MBMS GW and the MCE. This interface has similar MBMS functionality as currently specified over Iu (RANAP), meaning that the MCE is expected to take the role of the RNC with regards to setting up MBMS resources, while the MBMS GW takes the role of the SGSN. It is expected that M1 will have the same signaling transport layer as S1/X2.

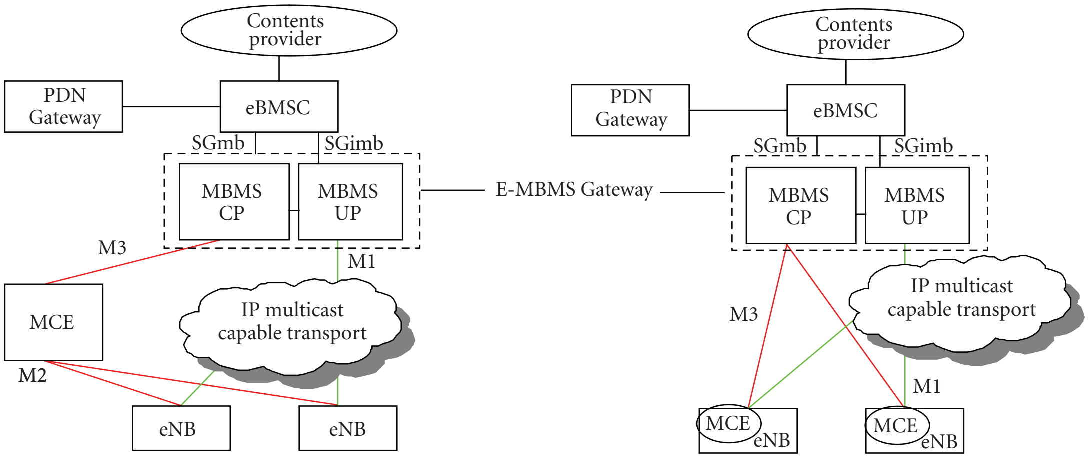


Figure 19: Overview of proposed architecture for MBMS in SAE/LTE (left) separate MCE, (right) MCE functionality integrated into eNB.

9.3.1. LTE Downlink Physical Layer. In the following we provide a brief introduction to the E-UTRA downlink physical layer [53]. The EUTRA downlink uses OFDM, because it efficiently supports flexible carrier bandwidth, allows frequency domain scheduling, is resilient to propagation delays which is particularly beneficial for SFN broadcasting configurations, and is well suited for multiple-input multiple-output (MIMO) processing.

The possibility to operate in vastly different spectrum allocations is essential. Different bandwidths are realized by varying the number of subcarriers used for transmission, while the subcarrier spacing remains unchanged. In this way operation in spectrum allocations of 1.25, 2.5, 5, 10, 15, and 20 MHz can be supported. Due to the fine frequency granularity offered by OFDM, a smooth migration of, for example, 2G spectrum is possible. Frequency-division duplex (FDD), time-division duplex (TDD), and combined FDD/TDD, are supported to allow for operation in paired as well as unpaired spectrum.

A subcarrier spacing of 15 kHz is adopted, allowing for simple implementation of dual mode Rel-6/LTE terminals as the same clock frequencies can be used. To minimize delays, the slot duration is selected as short as 0.5 ms, corresponding to seven OFDM symbols. The cyclic prefix length of 4.7 μs is sufficient for handling the delay spread for most unicast scenarios, while only adding modest overhead. Very large cells, up to and exceeding 120 km cell radius, with large amounts of time dispersion are handled by reducing the number of OFDM symbols in a slot by one in order to extend the cyclic prefix to 16.7 μs. Multi-cell broadcast services are supported by transmitting the same information from multiple (synchronized) base stations. To the terminal, the received signal from all base stations will appear as multipath propagation and thus implicitly be exploited by the OFDM receiver.

Figure 20 outlines the more detailed time-domain structure for LTE downlink transmission. Each 1 ms subframe consists of two slots of length $T_{\text{slot}} = 0.5$ ms. Each slot then consists of a number of OFDM symbols.

A subcarrier spacing $\Delta f = 15$ kHz corresponds to a useful symbol time $T_u = 1/\Delta f \approx 66.7\,\mu$s. The overall OFDM symbol time is then the sum of the useful symbol time and the cyclic-prefix length T_{CP}. As illustrated in Figure 20, LTE defines two cyclic-prefix lengths, the normal cyclic prefix and an extended cyclic prefix, corresponding to seven and six OFDM symbols per slot respectively. The exact cyclic-prefix lengths can be obtained from Figure 20. It should be noted that, in case of the normal cyclic prefix, the cyclic-prefix length for the first OFDM symbol of a slot is somewhat larger, compared to the remaining OFDM symbols.

The main use of the extended cyclic prefix is expected to be MBSFN-based multicast/broadcast transmission.

In addition to the 15 kHz subcarrier spacing, a reduced subcarrier spacing $\Delta f_{\text{low}} = 7.5$ kHz is also defined for LTE and specifically targeting MBSFN transmission. The use of the reduced subcarrier spacing also scales the OFDM symbol time by a factor of two, thus providing a twice as long cyclic prefix ($\approx 33.3\,\mu$s).

Taking into account also the downlink time-domain structure, the resource blocks mentioned above consist of 12 subcarriers during a 0.5 ms slot, as illustrated in Figure 21. Each resource block thus consists of $12 \cdot 7 = 84$ resource elements in case of normal cyclic prefix and $12 \cdot 6 = 72$ resource elements in case of extended cyclic prefix.

Each subcarrier of each OFDM symbol can be modulated by Quadrature phase shift keying (QPSK), 16-quadrature amplitude modulation (16-QAM), or 64-QAM modulation schemes.

The downlink share channel (DL-SCH) transport channel uses the physical layer to provide P-t-P radio bearers as well as P-t-M radio bearers with feedback from the UEs. The latter transmission mode can be used for MBMS if the number of users per cell is small, typically less than 10, so that the feedback does not cause a high signaling load. The feedback allows to adapt transmit parameters to the particular radio conditions of the few users and HARQ retransmission.

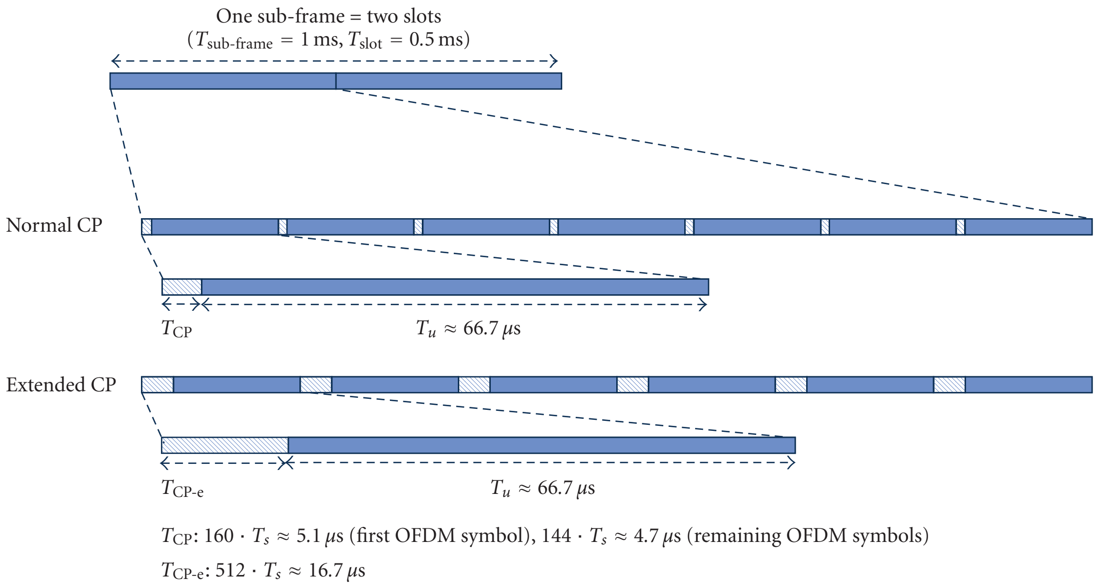


Figure 20: LTE downlink subframe and slot structure. One subframe consisting of two equal-size slots. Each slot consisting of 7/6 OFDM symbols (normal/extended cyclic prefix).

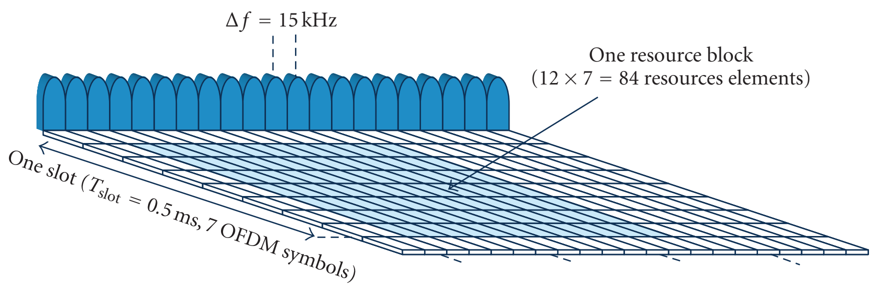


Figure 21: Downlink resource block assuming normal cyclic prefix. With extended cyclic prefix there are six OFDM symbols per slot and, consequently, 72 resource elements per resource block.

9.3.2. MBSFN. If a larger number of users of a particular MBMS are present in a cell, broadcast radio transmission in the cell are more suitable, which can be used either in single cell or multi cell transmission mode. In the single cell mode the MBMS service is transmitted to a particular user only from a single cell. Multi-cell transmission is supported by means of the MCH (*Multicast Channel*) transport channel. A multicell transmission essentially means that the cells transmitting the MBMS service are configured to form an MBSFN (see also Section 9.1). If an MBSFN is established using a particular MCH, then *the same* MCH information is transmitted *time aligned* from a group of cells using *identical transport formats and identical resource allocations* and with *identical scrambling* (cell-group-specific scrambling, see above). From a UE point-of-view, such multicell MCH transmission will appear as a *single* MCH transmission over a channel being the aggregation of the channels from the UE to each cell.

The MCH can time multiplexed on a subframe granularity of 1ms with other transport channels like the downlink shared channel (DL-SCH) that is used for single cell point to point transmission.

In order to be able to properly demodulate the multicell MCH transmission, the UE needs an estimate of the aggregated channel of all cells involved in the MBSFN. For this to be possible, specific reference signals ("*MCH-specific reference signals*") are needed. The MCH-specific reference signals are identical (identical time/frequency locations and identical reference-signal sequence) for all cells involved in the MBSFN. Based on such reference signals, the UE can directly estimate the aggregated channel of all cells involved in the MBSFN.

The aggregated channel of the cells involved in the MBSFN will typically have a large delay spread due to the differences in the propagation delay as well as residual transmit-timing differences. To handle this, LTE defines 3 different values for the OFDM cyclic prefix: 4.67 μs, 16.7 μs and 33 μs. Signals from eNBs arriving within the CP duration of the UE synchronization point contribute useful signal energy and thereby improve the coverage. Signals arriving outside the CP contribute interference. Since the CP does not contains user data, its length is a tradeoff between the time fraction available for user data and the CINR value achievable with the desired probability in the MBSFN area. If MBSFN is not used, the shortest CP will be sufficient in almost all propagation environments. If MBSFN consisting of more than a few cells are used, the CP of 16.7 μs should be used. In the case "high power high tower" sites are available or deployments in low frequency bands, good coverage can be achieved with even higher distance between sites and in this case the *extended long* CP of 33 μs should be used. In order to limit the relative overhead imposed by the CP, the OFDM core symbol duration is also doubled for the configuration with CP = 33 μs, and the subcarrier spacing is cut by half to keep the carrier bandwidth unchanged.

Based on the assumptions made by 3GPP, initial radio network simulations have been carried out. Results are presented in [54]. The key parameters are shown in Table 5.

Figure 22 shows the achieved broadcast spectrum efficiency versus inter-site distance for the case of multi-cell broadcasting. The spectral efficiency is valid for subframes allocated for MBSFN transmission. In Release 9 MBMS at most 6 out of every 10 subframes can be allocated to MBSFN transmission. A large cluster of cells transmits the same broadcast content synchronously thereby achieving signal aggregation gains and avoiding strong intercell interference. It can be seen that the spectrum efficiency requirement of 1 bps/Hz, is achieved for inter-site distances of up to approximately 1850 m inter site distance (ISD) for the 3GPP simulation scenario "Case 1", that is, in the 2 GHz band and with 20 dB indoor penetration loss, and up to 4700 m for the "Case 4", that is, in the 900 MHz band and with 10 dB indoor loss [53]. Naturally, the capacity decreases with increasing separation between transmitters as the power per transmitter is assumed to be fixed and the proportion of cells that are so far away that they cause interference rather than contribute useful signal increases.

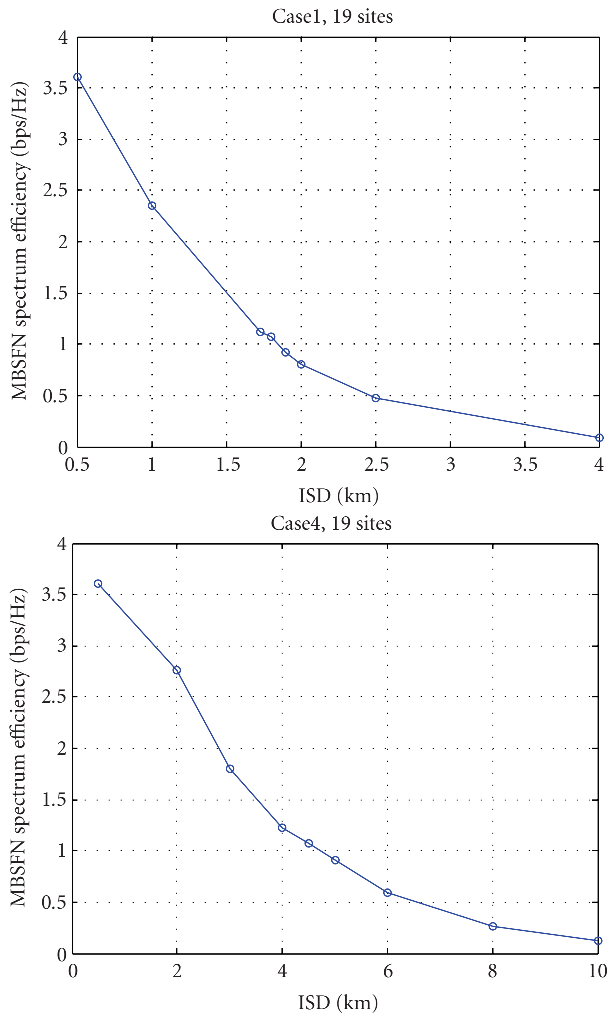


Figure 22: Broadcast capacity versus intersite distance.

9.4. PTP Bearer Performance of LTE and HSPA Evolution. In [15] throughput results for E-UTRA based on the agreed 3GPP current assumptions have been published where the radio resource is shared equally amongst all users. Here we are more interested in the capacity for streaming where all users need approximately equal throughput. The raw data that was used for [15] has been processed here to get an estimate of the streaming capacity per cell and 5 MHz as shown in Table 6, where the radio resource split between the users follows from the per user channel quality. The table is based on 95% satisfied users. This estimate is intended to show the technology potential, given, for example, perfect channel estimation, error-free feedback and disregarding transmitter impairments.

From these rates we calculate the number of channels for a given per channel rate. From this and 5% blocking we calculate the corresponding load, and from this and the load per subscriber follows the capacity in numbers of subscribers. For the assumption of 600 potential subscribers per cell (active and inactive) the results are plotted in Figure 23.

It can be seen that with LTE, even for 256 kbps and even if all potential subscribers actually subscribe to the service, each user can use the service on average for 25 minutes per 12 hours.

10. Conclusions

Mobile networks have evolved from voice telephony networks to multimedia delivery networks. Mobile TV services have become quite popular during the past two years. Apart from popular TV channels, the offerings often include special mobile editions with highlights from the weekly TV program, such as series and comedies, delivered as looped channels.

Although today's technologies such as HSPA provide enough capacity for the introduction phase of new mobile TV services, it is foreseeable that the increasing popularity of

Table 5: Radio network simulation parameters.

Spectrum allocation	10 MHz
Base station power	40 W
Propagation	$L = 35.3 + 37.6 * \log(d)$, d = distance in meters (includes $P_L = 20$ dB penetration loss) 14 dB_i transmitter antenna gain: 8 dB log-normal shadowing
Inter-site distance	500 to 4000 m (varied)
Modulation and coding schemes	Depending on intersite distance, QPSK, 16QAM, or 64QAM, and turbo coding

Table 6: Streaming capacity per cell and 5 MHz.

HSPA R6 1-Rx	HSPA R6 2-Rx	HSPA-evolution	LTE
1.9 Mbps	2.9 Mbps	5.3 Mbps	6.7 Mbps

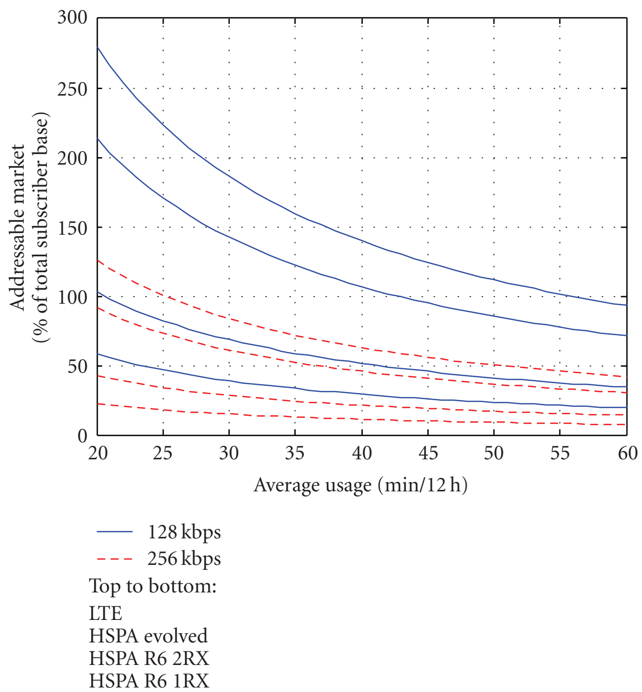


Figure 23: Estimate of addressable market for future UMTS evolution tracks.

those services might sooner or later lead to capacity bottlenecks. Therefore, 3GPP back in 2002 started the specification of broadcast/multicast services for GSM/UMTS.

MBMS adds broadcast and multicast capabilities to existing GSM/UMTS networks. The MBMS functionalities can be accessed through functions provided by the so-called Broadcast/Multicast Service Centre (BM-SC), a new logical node in the 3GPP architecture. Together with new Point-to-Multipoint radio bearers MBMS enables efficient stream and file delivery to large user groups with various error protection and recovery options. Related application and service layer functionality has been defined by the Open Mobile Alliance (OMA), in the Mobile Broadcast Services 1.0 (BCAST 1.0) specification, which was finalized in 2007.

MBMS has been specified such that broadcast/multicast services can be used together with voice and data services within the same radio carrier. This gives greatest flexibility to cellular operators. The various network configurations MBMS supports range from single multicast/broadcast transmission in a single cell up to a nationwide, single frequency network, broadcasting the same content (e.g., TV channels) across the whole country.

The specification of MBMS in the UMTS radio access network (UTRAN) enables the estimation of the approximate or exact number of UEs in a cell that are currently interested in a particular MBMS content channel. This allows the radio access network to make an optimal decision about if and how a broadcast service should be delivered in a particular cell: via P-t-P or P-t-M radio bearers, or not at all. Thus often the term “Enhanced Broadcast Mode” is used to refer to this particular MBMS configuration.

MBMS supports standardized, state-of-the-art audio and video codecs. The used video codec is H.264. Image resolutions up to QVGA (320×240 pixel) and frame rates up to 30 fps are supported. For audio, the High-Efficiency AAC version 2 codec, and the various AMR codecs (narrow band, wideband, and enhanced wideband) are supported.

Simulation results have shown, that an MBMS enabled UMTS 3GPP Release 6 network is able to deliver 20 mobile TV channels to a mobile TV subscriber base which corresponds to roughly 30% of all subscribers.

In 3GPP Release 7 and 8 the MBMS broadcasting capacity when delivering the same service over larger geographical areas (MBSFN) is drastically improved. If all cells use the same carrier frequency, the terminal is able to receive the signals from several cells simultaneously. By the elimination of intercell interference, a dramatic increase in spectral efficiency and thereby throughput on the used radio resources is achieved. MBSFN is an enhancement available for both, networks based on WCDMA technology and LTE. With LTE it is even possible to time multiplex MBSFN and unicast transmission on the same carrier frequency, thereby providing the best possible integration of high performance broadcasting and individual user service delivery.

References

[1] M. Eriksson, A. Furuskar, M. Johansson, et al., “The GSM/EDGE radio access network—GERAN; system overview and performance evaluation,” in *Proceedings of the 51st IEEE Vehicular Technology Conference (VTC ’00)*, vol. 3, pp. 2305–2309, Tokyo, Japan, May 2000.

[2] H. Holma and A. Toskala, *WCDMA for UMTS—Radio Access for Third Generation Mobile Communications*, Wiley, New York, NY, USA, 2004.

[3] J. Sköld, M. Lundevall, S. Parkvall, and M. Sundelin, "Broadband data performance of third-generation mobile systems," *Ericsson Review*, vol. 82, no. 1, pp. 14–23, 2005.

[4] 3GPP TS 26.234, "Transparent end-to-end Packet-switched Streaming Service (PSS)".

[5] I. Elsen, F. Hartung, U. Horn, M. Kampmann, and L. Peters, "Streaming technology in 3G mobile communication systems," *Computer*, vol. 34, no. 9, pp. 46–52, 2001.

[6] ITU-T Recommendation H.264 (03/05), "Advanced video coding for generic audiovisual services".

[7] ISO/IEC 14496-10:2005, "Information technology – Coding of audio-visual objects—Part 10: Advanced Video Coding".

[8] M. Kornfeld and G. May, "DVB-H and IP datacast—broadcast to handheld devices," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, part 2, pp. 161–170, 2007.

[9] S. Cho, G. Lee, B. Bae, et al., "System and services of terrestrial digital multimedia broadcasting (T-DMB)," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, part 2, pp. 171–177, 2007.

[10] M.R. Chari, F. Ling, A. Mantravadi, et al., "FLO physical layer: an overview," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, part 2, pp. 145–160, 2007.

[11] S. McCanne, "Scalable multimedia communication using IP multicast and lightweight sessions," *IEEE Internet Computing*, vol. 3, no. 2, pp. 33–45, 1999.

[12] 3GPP TS 26.346, "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs".

[13] "Mobile Broadcast Services," Open Mobile Alliance^{TM}, OMA-TS-BCAST_Services-V1_0, http://www.openmobilealliance.org/.

[14] "Service Guide for Mobile Broadcast Services," Open Mobile Alliance^{TM}, OMA-TS-BCAST_ServiceGuide-V1_0, http://www.openmobilealliance.org/.

[15] H. Ekström, A. Furuskär, J. Karlsson, et al., "Technical solutions for the 3G long-term evolution," *IEEE Communications Magazine*, vol. 44, no. 3, pp. 38–45, 2006.

[16] E. Dahlman, H. Ekström, A. Furuskär, et al., "The 3G long-term evolution—radio interface concepts and performance evaluation," in *Proceedings of the 63rd IEEE Vehicular Technology Conference (VTC '06)*, vol. 1, pp. 137–141, Melbourne, Canada, 2006.

[17] A. Toskala, H. Holma, K. Pajukoski, and E. Tiirola, "Utran long term evolution in 3GPP," in *Proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, pp. 1–5, Helsinki, Finland, September 2006.

[18] 3GPP, TS 36.300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2," [v8.4.0 check if newer avail].

[19] M. Baugher, D. McGrew, M. Naslund, E. Carrara, and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)," RFC 3711, March 2004, http://www.ietf.org/rfc/rfc3711.txt.

[20] 3GPP TS 26.401, "General audio codec audio processing functions; Enhanced aacPlus general audio codec; General description".

[21] 3GPP TS 26.071, "AMR speech codec; General description".

[22] 3GPP TS 26.171, "AMR speech codec, wideband; General description".

[23] 3GPP TS 26.290, "Audio codec processing functions; Extended Adaptive Multi-Rate—Wideband (AMR-WB+) codec; Transcoding functions".

[24] H. Thompson, D. Beech, M. Maloney, and N. Mendelsohn, "XML Schema Part 1: Structures," W3C Recommendation, May 2001.

[25] P. Biron and A. Malhotra, "XML Schema Part 2: Datatypes," W3C Recommendation, May 2001.

[26] M. Handley and V. Jacobson, "SDP: Session Description Protocol," IETF RFC 2327, April 1998.

[27] IETF RFC 2616, "Hypertext Transfer Protocol—HTTP/1.1," June 1999.

[28] OMA Push OTA Protocol, "WAP-235-PushOTA-20010425-a," April 2001, http://www.openmobilealliance.org/tech/affiliates/LicenseAgreement.asp?DocName=/wap/wap-235-pushota-20010425-a.pdf.

[29] 3GPP TS 33.220, "Generic Authentication Architecture (GAA); Generic Bootstrapping Architecture".

[30] 3GPP TR 26.946, "Multimedia Broadcast/Multicast Service (MBMS)—User service guidelines, v 6.1.0".

[31] M. Bakhuizen and U. Horn, "Mobile broadcast/multicast in mobile networks," *Ericsson Review*, no. 1, pp. 6–13, 2005.

[32] F. Hartung, U. Horn, J. Huschke, M. Kampmann, T. Lohmar, and M. Lundevall, "Delivery of broadcast services in 3G networks," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 188–198, 2007.

[33] 3GPP TR 25.803, "S-CCPCH performance for MBMS," V6.0.0, September 2005.

[34] 3GPP TS 23.246, "Multimedia Broadcast/Multicast Service (MBMS); Architecture and functional description".

[35] T. Paila, M. Luby, R. Lehtonen, V. Roca, and R. Walsh, "FLUTE—File Delivery over Unidirectional Transport," RFC 3926.

[36] M. Watson, "FECFRAME requirements," IETF draft-ietf-fecframe-req-00, work in progress.

[37] A. Shokrollahi, "Raptor codes," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[38] M. Luby, T. Gasiba, T. Stockhammer, and M. Watson, "Reliable multimedia download delivery in cellular broadcast networks," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, part 2, pp. 235–245, 2007.

[39] T. Lohmar and M. Elisova, "Evaluation of the file repair operations for multicast/broadcast download deliveries," in *Proceedings of European Wireless Conference*, Nicosia, Cyprus, April 2005.

[40] T. Lohmar, P. Zhaoyi, and P. Mähönen, "Performance evaluation of a file repair procedure based on a combination of MBMS and unicast bearers," in *Proceedings of International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM '06)*, pp. 349–357, Niagara Falls, NY, USA, June 2006.

[41] H. Schulzrinne, et al., "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 3550, July 2003.

[42] T. Stockhammer, A. Shokrollahi, M. Watson, M. Luby, and T. Gasiba, "Application layer forward error correction for mobile multimedia broadcasting," in *Handbook of Mobile Broadcasting: DVB-H, DMB, ISDB-T, and MEDIAFLO*, B. Fuhrt and S. Ahson, Eds., CRC Press, Boca Raton, Fla, USA, 2008.

[43] 3GPP TS 26.245, "Transparent end-to-end Packet switched Streaming Service (PSS); Timed text format".

[44] J. Arkko, E. Carrara, F. Lindholm, M. Naslund, and K. Norrman, "MIKEY: Multimedia Internet KEYing," RFC 3830, August 2004, http://www.ietf.org/rfc/rfc3830.txt.

[45] "ISMA 1.0 Encryption and Authentication, Version 1.1," release version, http://www.isma.tv/.

[46] S. Kent and R. Atkinson, "IP Encapsulating Security Payload (ESP)," RFC 2406, November 1998, http://www.ietf.org/rfc/rfc2406.txt.

[47] bmcoforum, http://www.bmcoforum.org/.

[48] T. Lohmar and U. Horn, “Hybrid broadcast-unicast distribution of mobile TV over 3G networks,” in *Proceedings the 31st IEEE Conference on Local Computer Networks (LCN ’06)*, pp. 850–851, Tampa, Fla, USA, November 2006.
[49] C. Södergard, Ed., “Mobile Television—Technology and user experiences. Report on the Mobile-TV project,” Report, VTT Information Technology, Espoo, Finland, 2003.
[50] 3GPP TR 25.913, “Requirements for Evolved UTRA and Evolved UTRAN”.
[51] Ericsson, “Dedicated MBMS Carrier Using Common Transmitted Waveforms,” contribution R1-062268 to 3GPP RAN1, August 2006.
[52] 3GPP, TR 25.905, “Improvement of the Multimedia Broadcast Multicast Service (MBMS) in UTRAN (Release 7), v2.0.0,” September 2006.
[53] 3GPP TR 25.814, “Physical Layer Aspects for Evolved UTRA”.
[54] Ericsson, “E-UTRA Performance Checkpoint: MBSFN,” contribution R1-071958 to 3GPP RAN1, April 2007.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 369216, 12 pages
doi:10.1155/2009/369216

Research Article

Video Streaming Transfer in a Smart Satellite Mobile Environment

Nedo Celandroni,[1] Franco Davoli,[2] Erina Ferro,[1] and Alberto Gotta[1]

[1] *CNR-ISTI, Area della Ricerca del C.N.R., Via Moruzzi 1, 56124 Pisa, Italy*
[2] *CNIT-University of Genoa Research Unit and Department of Communications, Computer and Systems Science (DIST), University of Genoa, Via Opera Pia 13, 16145 Genova, Italy*

Correspondence should be addressed to Franco Davoli, franco@dist.unige.it

Received 27 February 2009; Accepted 10 July 2009

Recommended by Sandro Scalise

Transportation media are becoming "smart spaces", where sophisticated services are offered to the passengers. We concentrate on video streaming provided on buses that move in urban, suburban, or highway environments. A content provider utilizes a DVB-S2 satellite link for transmitting video streams to a bus, where they are relayed to passengers' devices. We say that a bus works in smart mode if it takes advantage of the knowledge of the exact points where fixed obstacles will prevent receiving the satellite signal for a certain time period. This information is sent to the hub via a return channel. The hub, in its turn, suspends the transmissions to that specific bus for the given time interval, thus avoiding information losses and unnecessary bandwidth occupation. Buffering video packets, without any quality of service (QoS) degradation, seamlessly compensates channel blockages up to a given duration. We determine the most appropriate transmission parameters for video streaming with good video QoS in a mobile satellite environment; moreover, we evaluate how "smart" the system can be in terms of bandwidth saving, by comparing it with the situation where the bus does not exploit the description of its route, still maintaining the same QoS requirements.

1. Introduction

We consider a new utilization of buses, transforming them from pure travel media for transporting passengers to a mobile communication system, able both to contribute services for the public administrations (environmental and traffic monitoring, urban road surveillance, sensing data collection while passing, etc.) and to offer services to passengers, such as news, advertisements, trains/flights timetables, city and tourist info, bus/train routes and timetables, Internet access, music, instant messaging, video, TV on demand, and so on. In this paper, we concentrate on this second aspect, that is, the multimedia streaming services for passengers who utilize long-haul buses.

The system assumptions are the following (Figure 1). We consider a geostationary satellite system with a terrestrial hub station acting as a network control center (NCC). The hub transmits a large amount of multimedia data (video streams and the other information flows, previously mentioned) in time division multiplexing (TDM) mode to a fleet of vehicles, using the DVB-S2 standard [1]. The return link (from the vehicles to the hub) is assigned to a mobile network link, (from 2.5 G up to 3.5 G service provider), which guarantees a widespread coverage and sufficient bandwidth for signaling toward the hub. We assume that the hub is able to change the average speed of each streaming connection by changing the transmission frequency of the relevant packets. In Section 3, we will address two different rates for each video streaming transfer, r_n (normal rate) and r_h (higher rate), respectively.

The transmission of the video streams to a fleet of buses is done in a "personalized" mode: a multimedia flow is addressed to a specific bus of the fleet either in response to an explicit request, or simply to disseminate information. The Relay Media Proxy allocates a separate portion of memory (i.e., buffers) for each flow, in order both to serve the flows and to manage the store and forward mechanisms that will be detailed in the following. In this paper, we consider two types of video streaming transmissions, which are real-time Video on Demand (VoD), based on the Real-time Streaming Protocol (RTSP) [2], and Live Video Streaming

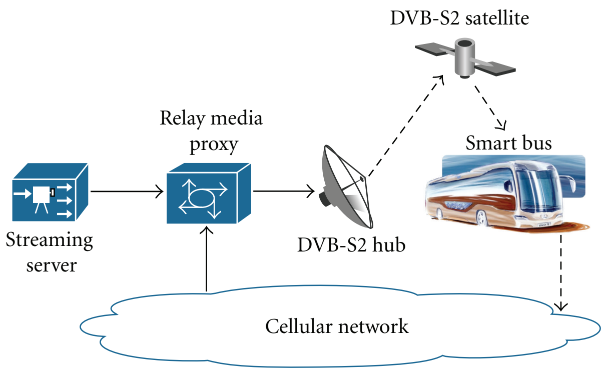


FIGURE 1: The video streaming transmission in the "smart" bus environment.

(LVS), such as IPTV [3]. Each bus relays the received video to its passengers as an on-board service.

We also assume that each bus is able to know and exploit some pieces of information relevant to its route. Bus routes can traverse different environmental settings, each one characterized by its typical obstacles: (i) urban, with tall buildings; (ii) suburban (small villages and towns, as well as the suburbs of big cities); (iii) rural (open areas with tree alleys and forests); (iv) highway (open areas with the presence of relatively frequent bridges and tunnels). In any of the mentioned environments, the buses' paths can be well traced by any type of GPS device available on the market; each environment can be characterized by a specific channel model, which takes into account multipath fading and shadowing phenomena in a static seasonal environment [4]. The communication from the hub toward buses is therefore adapted to the channel type, by using the most appropriate transmission parameters, as detailed in the following sections. The authors are well conscious that for some trips the usage of the train is surely more convenient than the bus; nevertheless, the connection between distant cities by means of extraurban buses is a reality both in Europe and in the USA, and this is the case we address in our paper. In this scenario, the only possible means for connecting a bus with the hub is the satellite. As far as the authors know, this research topic has not yet been addressed in the literature.

The behavior of the mobile satellite channel, as described by the Markov models in [5], is such that relatively long periods of signal blockage are possible. Such blockage events can be counteracted—in addition to the DVB-S2 adaptive coding and modulation (ACM) applied at the physical layer—by adopting Upper-Layer Forward Error Correction (UL-FEC) and interleaving at the packet level. A possible drawback of these techniques is the introduction of a sensible delay in the playback of the stream. In any case, even in the presence of such strong countermeasure actions, the occurrence of particularly long blockage periods may lead to a complete channel outage, since the loss of all packets transmitted during a long period may overcome the correcting capacity of the upper-layer code. When the routes are well known, as when traced by means of GPS, these periods can be predictable with good precision. Therefore, in unicast transmission mode, our "smart bus" can alert the hub in advance, thus making it stop sending packets, which would otherwise be lost, for the duration of the blocking time interval. We denote this operational condition as *Smart Mode* (SM); we say that a bus works in *Normal Mode* (NM) if it does not take advantage of the route information. However, UL-FEC and packet interleaving are applied in both modes.

Thus, our smart bus interacts with both the GPS and the hub. The bus may reach the latter either via a cellular network or via a satellite return channel. When a bus, by means of the GPS, detects the presence of a bulky obstacle along the route (tunnel, long/tall buildings, long foliage lines, etc.), it advises the hub, via the return channel, that it would loose packets, starting from an estimated time instant, which accounts for uncertainties in the position estimation, after a round-trip delay time; such an estimation depends on the precision of both the current vehicle's position and speed. Thus, the hub can temporarily stop sending packets to this bus. At the end of the "blockage-time", the smart bus will again advise the hub to resume the transmission. Of course, many factors contribute to the estimation of the transmission suspending time: bus speed, environment type, constraints imposed by the traffic class, and so forth. The error made in predicting such an instant is due to the transmission delay between the bus and the satellite. We compensate for such an error by introducing a suitable guard time, which slightly worsens the system performance.

The aim of this paper is determining the most appropriate UL-FEC parameters, together with the interleaver depth, for video streaming transmissions in a satellite mobile channel, and selecting the buffering strategies to take advantage of the knowledge of no-reception time periods. Furthermore, this paper highlights "how smart" the system is in terms of saved bandwidth, by comparing the smart behavior with that of a bus that does not take advantage of the route information. The determination of the above quantities only involves the study of a single channel between a generic bus and the hub. Therefore, we concentrate on this, and do not deal with the multiplexing and multicast transmission problems connected with the delivery of multiple video streams to a multitude of buses.

To reach our goals, we first model the satellite mobile channel for different environments where the bus can move (Section 2); then, we study the mechanisms for video streaming transmission (Section 3), and we make some considerations on the application of the interleaving technique (Section 4). How smart the system is, in terms of saved bandwidth, is shown in Section 5. Our considerations on the results obtained are summarized in Section 6.

2. The Mobile Channel Model

One of the goals of this paper is to quantify packet error rates experienced in the nonblocked states of the channel, in order to make possible choosing the best operating conditions for the video streaming transmission. This means choosing the modulation and coding scheme (MODCOD), and the Reed-Solomon Erasure (RSE) [6] block length and code rate, for the E_s/N_0 (symbol energy over two-sided noise power

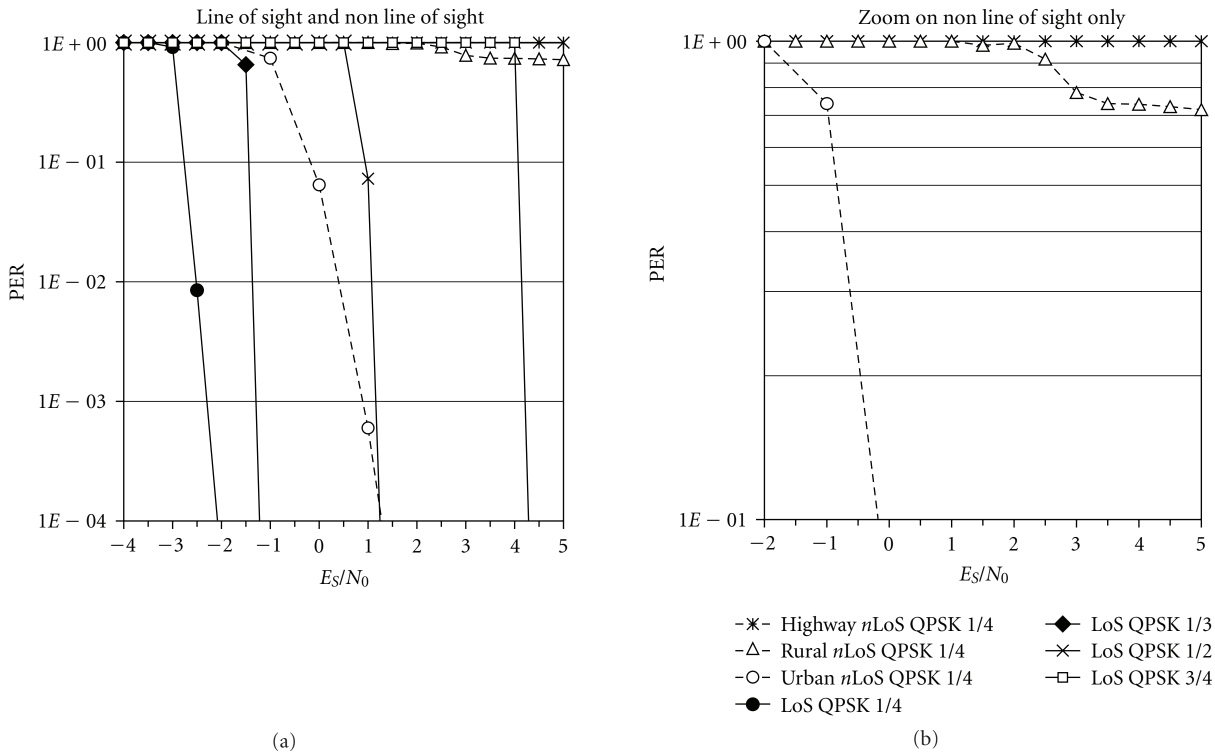


Figure 2: Packet error rate for mobile DVB-S2 in LoS and shadowed states.

spectral density) ratio available at the receiver. To model the mobile satellite channel at the packet level, we adopt the 3-state Discrete-Time Markov Chain propagation channel model provided by the European Space Agency (ESA) for the Ku-band satellite channel availability to mobile users in four environments [4, 5]: urban, suburban, rural, or highway. Suburban and rural environments present quite similar characteristics, as it results from the transition matrix in Table 1, which shows that the transition probabilities—and, thus, the steady-state probabilities—are almost the same for these two environments. For this reason, between the two, we chose considering the rural case only. Urban and highway have completely different characteristics, as they differ in the average vehicle speeds (higher in the highway, lower inside towns), other than in the horizon profile, mainly owing to buildings; these concepts are confirmed by the transition probabilities in Table 1.

For each environment, Table 1 reports the state transition matrix of the ESA mobile channel model [4]; the state transition period T_s is equal to 41.7 ms, which corresponds to blocks of 1000 samples at the sampling frequency of 24 kHz, at the average speeds of the vehicle indicated in the table. In each environment, the three states of the Markov chain are named Line of Sight (LoS), Shadowed (Shd), and Blocked (Blk), respectively. The parameters in the four environments essentially depend on the morphology of the landscape and the average speed of the vehicle.

In order to evaluate the performance of our proposed strategies in the mobile DVB-S2 environment, we developed a simulation model in Simulink starting from the Communication Demos DVB-S2 [7] and according to the parameters reported in Table 1. The simulation model takes into account a standard DVB-S2 transmission/reception chain [1], thus including LDPC [8] and BCH [9] FEC codes in order to simulate the ESA channel model. The DVB-S2 FECFRAME length was set to 64 800 bits, which is the normal frame size in DVB-S2, and includes the BBFRAME (BBheader + Data Field), the LDPC FEC, and the BCH FEC. Figure 2 reports simulation results of the Packet Error Rate (PER) versus E_s/N_0, relative to a set of DVB-S2 MODCODs in LoS and Shd conditions. We assume to operate with MPEG-2 transport streams (TS), which we will refer to as “packets” throughout the paper. The DVB-S2 BBFRAME can contain more than one of such packets. However, owing to the multiplexing capability allowed by the DVB-S2 standard (Merging/Slicing policy), it is reasonable to assume that a BBFRAME contains no more than one packet belonging to a single stream flow. In this case, the PER relative to the stream coincides with the BBFRAME Error Rate. The range of the x-axis accounts for a link budget that is prevalently limited by the antenna size and by regulatory issues [4], which impose a maximum power flux density on the earth’s surface and, consequently, a maximum effective isotropic radiation power (EIRP) from the satellite. The authors carried out this simulation study because, to the best of their knowledge, no PER evaluation in a mobile environment was available in the literature at the time of writing; in fact, the novelty consists in implementing the DVB-S2 standard in a mobile environment.

Table 1: Ku-band land vehicular channel model parameters [4, 5].

Environment	State transition matrix	Steady state probability	Rice factor (K)[1]	Average speed	Doppler shift
	$\begin{pmatrix} p_{\text{LoS,LoS}} & p_{\text{LoS,Shd}} & p_{\text{LoS,Blk}} \\ p_{\text{Shd,LoS}} & p_{\text{Shd,Shd}} & p_{\text{Shd,Blk}} \\ p_{\text{Blk,LoS}} & p_{\text{Blk,Shd}} & p_{\text{Blk,Blk}} \end{pmatrix}$	$\begin{pmatrix} \pi_{\text{LoS}} \\ \pi_{\text{Shd}} \\ \pi_{\text{Blk}} \end{pmatrix}$	$\begin{pmatrix} \text{LoS} \\ \text{Shd} \\ \text{Blk} \end{pmatrix}$		
Highway	$\begin{pmatrix} 0.9862 & 0.0138 & 0.0000 \\ 0.1499 & 0.8378 & 0.0123 \\ 0.0008 & 0.0396 & 0.9596 \end{pmatrix}$	$\begin{pmatrix} 0.8929 \\ 0.0821 \\ 0.0250 \end{pmatrix}$	50.12 0.69 0.00	100 km/h	1342.87 Hz
Rural	$\begin{pmatrix} 0.9795 & 0.0204 & 0.0001 \\ 0.1007 & 0.8277 & 0.0716 \\ 0.0010 & 0.1813 & 0.8177 \end{pmatrix}$	$\begin{pmatrix} 0.7794 \\ 0.1581 \\ 0.0625 \end{pmatrix}$	50.12 1.32 0.00	70 km/h	940.45 Hz
Suburban	$\begin{pmatrix} 0.9796 & 0.0204 & 0.0000 \\ 0.0929 & 0.0929 & 0.0500 \\ 0.0015 & 0.1876 & 0.8109 \end{pmatrix}$	$\begin{pmatrix} 0.7834 \\ 0.1713 \\ 0.0453 \end{pmatrix}$	63.10 1.82 0.00	40 km/h	537.6 Hz
Urban	$\begin{pmatrix} 0.9902 & 0.0097 & 0.0001 \\ 0.0714 & 0.8756 & 0.0529 \\ 0.0000 & 0.0140 & 0.9860 \end{pmatrix}$	$\begin{pmatrix} 0.6014 \\ 0.0825 \\ 0.3161 \end{pmatrix}$	50.12 2.30 0.00	30 km/h	403.2 Hz

[1]The Rice factor K is defined as the ratio of signal power in the dominant component over the scattered reflected power. It determines the distribution of the received signal amplitude. The knowledge of the Rician K factor allows determining the bit error rate of a channel, among other metrics.

In LoS, curves are quasivertical, and are not affected by a change in the environment. The operating MODCOD is determined by the available E_s/N_0, which depends on the symbol rate, given the C/N_0 (carrier power over two-sided noise power spectral density) ratio that results from the link budget. In the Shd state, for all cases, except the urban one, the PER does not sufficiently decrease to acceptable values, for obtainable levels of E_s/N_0, even for the most robust MODCOD of DVB-S2, that is, QPSK 1/4. In all such cases, then, the Shd state is equivalent to Blk.

Hence, by choosing the appropriate MODCOD for the available value of E_s/N_0, the channel behavior at packet level can be seen as a 2-state Discrete-Time process. It is worth noting that, in reducing the Markov chain from 3 to 2 states, by merging Shd and Blk states, the Markovian property is generally lost; we will examine this in detail further on in the section. The two states are a "Good" one (LoS) and a "Bad" macro-state (which includes Shd + Blk). The only exception to this assumption is the urban case, in which the Rice Factor K is sufficiently high. In this case, we note in Figure 2 that, when E_s/N_0 is about 1 dB, the PER in the Shd state is close to 10^{-4} (which is equivalent to that of a Good state) for QPSK 1/4; the same PER is obtained in LoS with QPSK 1/2, which allows operating at double speed with respect to QPSK 1/4.

Note that, in DVB-S2, QPSK is the maximum energy efficient modulation scheme, and LDPC = 1/4 is the maximum coding protection.

Then, there are two possible choices. At QPSK 1/4, the LoS and Shd states can be merged into one "Good" state (which we may call BG, for "Big Good"); at QPSK 1/2, the Shd state becomes practically indistinguishable from the Blk one, as in the previous situations. It is worth noting that considering the Shd state as a Good one (which implies higher physical layer redundancy) allows using a lower UL-FEC (RSE) redundancy, with respect to considering the Shd state as a bad one (BB, for "Big Bad"), to the end of obtaining the same final PER. As we mentioned earlier, aggregating two or more states of a Markov chain into a macro-state generally does not preserve the Markovian property in the new reduced process.

Figure 3 highlights the probability mass function (PMF) of the sojourn time in the new macro state in both BB and BG cases (jittered line, obtained by simulation), in comparison with an exponential distribution with the same average sojourn time τ_{BB} or τ_{BG} (straight line). The PMF gives the probability that the discrete random variable is exactly equal to some value.

In addition, we provide further *hypoexponential* fittings of the PMF in the two cases, which are plotted in the two graphs by means of a white interpolating line; the expression of the fitting curve is

$$F(t) = \frac{a}{\tau_1} e^{-t/\tau_1} + \frac{b}{\tau_2} e^{-t/\tau_2}. \quad (1)$$

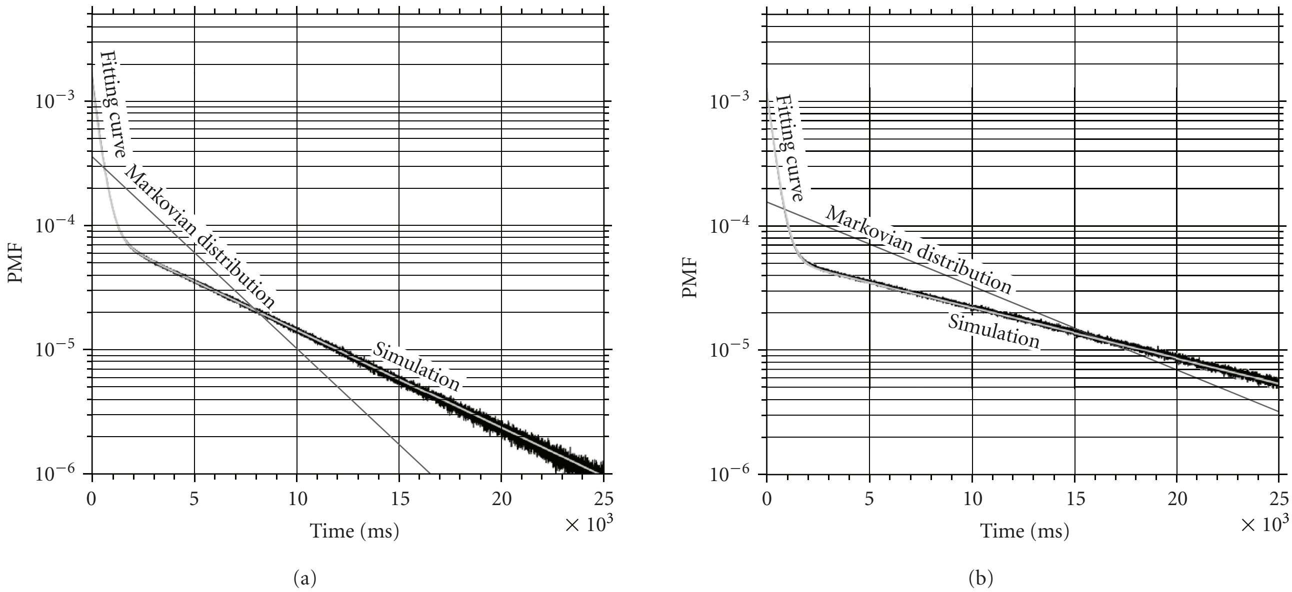


Figure 3: Probability mass function of the sojourn time in BB (a) and BG (b) macro-states in the urban environment (the PMF of a Markovian state with the same mean value is also shown; the white line represents the interpolation fitting curve of simulation results).

Table 2: Parameters of the hypoexponential distribution.

BB parameters	BG parameters
$\tau_1 = 2.979$ s	$\tau_1 = 2.979$ s
$\tau_2 = 5.495$ s	$\tau_2 = 10.700$ s
$a = 0.463$	$a = 0.604$
$b = 0.529$	$b = 0.399$

Table 2 provides the relative parameters that define the hypoexponential curve for each distribution.

3. Mechanisms for Video Streaming Transfer

Video on Demand is a typical unicast streaming service, similar to a file transfer, but concerning video or audio (such as YouTube), where a user downloads contents while watching them. Generally, this kind of service is quite delay-tolerant; short periods of congestion or link failures may be compensated for with opportune playout buffering, which filters the delay jitter caused by the variability of both source rate and channel bandwidth [10]. Live Video Streaming (LVS) refers to those broadcasting streaming services that produce data that are not prerecorded, and cannot be delayed so much, such as newscasts, alerts, last minute offers, and so on.

Figure 4 shows the Tx/Rx chain of the DVB transport streams. In case of LVS, a video camera, which can be deployed anywhere, transmits a video to the hub; in the VoD case, data can reside in the hub or in a remote streaming server. The Tx and Rx buffers shown in the figure are needed to store packets whenever necessary during the smart mode operations. The buffering of the k packets of each codeword is handled within the RSE encoder and decoder. We do not deal with the playout buffer, which is contained inside the playout device; we treat the Tx and Rx buffers only.

The smart mode behavior mentioned previously just refers to the set of features driven by feedback commands that allows for the hub stopping packet transmissions during long periods of channel blockage of the bus, and resuming them at normal speed (r_n) or at a higher speed (r_h), when the buffer of the receiving bus has to be refilled. The reason for the two speeds is explained in the following.

At the receiver on the bus, correctly demodulated DVB packets fill the deinterleaver, which flags as *erasures* all packets between two valid ones; that is, when a Transport Error Indicator (TEI) is set to 1 by the demodulator (as it cannot correct errors in the stream), all packets in-between this packet and the next one containing a correct sync byte are flagged as *erasures*. Then, blocks that contain both valid packets and erasures are passed to the RSE decoder. The RSE (n, k) decoder, which is located at the output of the deinterleaver, is able to rebuild k information packets if any k, out of a block of n packets (codeword), are correctly received. The coder first sends k information packets, followed by n-k redundant ones. Thus, if the first k packets are correctly received, no decoding actions are needed. If k or less than k out of n valid packets are received, the decoder delivers a block of k or fewer packets to the playout device (normal mode) or to the Rx buffer (smart mode).

The different characteristics of the LBVS and VoD traffic impose a distinction. We note that, given the presence of the functional blocks shown in Figure 4, the latency in the beginning of video playback (start-up delay) results in the sum of three components: the interleaver filling time, the RSE coding/decoding time, and the Rx buffer filling time. We denote by *stdelay* the value of the sum of these three time components. In other words, *stdelay* is the end-to-end delay

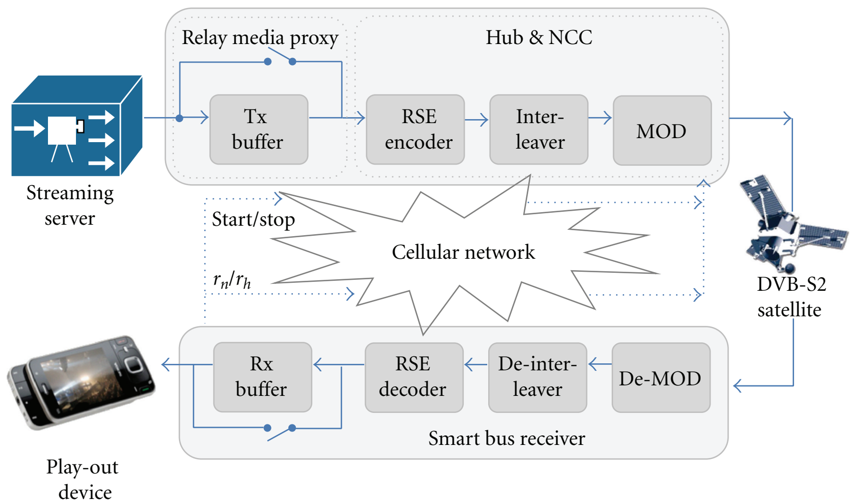


Figure 4: Tx/Rx chain for DVB communication to a vehicular device.

in the video playback that we pay to reduce the overall packet loss within the acceptable limit of 2% [11]; *stdelay* is suitably "*spent*" in the three components in different proportions, according to the various cases. We call *outage* the set of time periods in which certainly there are no good packets for video playback; this occurs at all times in which the bad channel duration exceeds the *stdelay* interval. In fact, in this case, invalid packets fill the whole deinterleaver and no good ones get out of the RSE decoder.

By considering a continuous-time approximation of our channel model, we can calculate the outage probabilities, given the threshold *stdelay* and the probability density functions of the sojourn times in the blocking states associated with the BB and BG cases (t_{BB} and t_{B}, respectively); we also indicate with τ_{B} and τ_{G} the respective average durations of the bad and good intervals:

$$P^{\mathrm{BG}}_{\mathrm{Outage}} = \Pr\{t_{\mathrm{B}} > stdelay\} = \frac{1}{\tau_{\mathrm{BG}} + \tau_{\mathrm{B}}} \times \int_{stdelay}^{\infty} (\vartheta - stdelay)\frac{1}{\tau_{\mathrm{B}}} e^{-\vartheta/\tau_{\mathrm{B}}} d\vartheta, \tag{2}$$

$$P^{\mathrm{BB}}_{\mathrm{Outage}} = \Pr\{t_{\mathrm{BB}} > stdelay\} = \frac{1}{\tau_{\mathrm{BB}} + \tau_{\mathrm{G}}} \times \int_{stdelay}^{\infty} (\vartheta - stdelay) F(\vartheta) d\vartheta, \tag{3}$$

where

$$\begin{aligned}
\tau_B &= \frac{T_s}{\left(1 - a^{(u)}_{33}\right)}, \\
\tau_G &= \frac{T_s}{\left(1 - a^{(u)}_{11}\right)}, \\
\tau_{\mathrm{BG}} &= \tau_B \frac{(\pi_1 + \pi_2)}{\pi_3}, \\
\tau_{\mathrm{BB}} &= \tau_G \frac{(\pi_2 + \pi_3)}{\pi_1},
\end{aligned} \tag{4}$$

with $\{\pi_1, \pi_2, \pi_3\}$ being the stationary probability distribution of the 3-state Markov chain in LoS, Shd and Blk states, respectively; $a^{(u)}_{ii}$ indicates the ith diagonal element of the transition probability matrix of the discrete-time chain, in the urban environment. By using relations (2) and (3), for the urban environment, we get outage percentages 0.29% and 3.4%, for the BG and BB cases, respectively.

3.1. LBVS Case. In normal mode, the transmission buffer is not present in the hub. In this case, *stdelay* corresponds to the interleaver filling time at speed r_n plus the RSE coding/decoding time. The packet loss is the one resulting by adopting an RSE (n, k) coder plus an $n \times m$ interleaver of depth m.

In smart mode, the time needed to fill the Rx buffer is equal to *stdelay*, decremented by the interleaver filling time and the RSE coding/decoding time. The interleaver depth is kept smaller than in normal mode, because it must only compensate for short channel interruptions up to a certain threshold (denoted by *maxtb*); emptying the Rx buffer, while the transmission is suspended, compensates for longer interruptions. At the beginning of the process, the Rx buffer filling is done by asking the hub to transmit at speed r_n (and not r_h), because the streaming flow is available at speed r_n. When the Rx buffer is completely filled, the Tx buffer is still empty and the video playback can begin. When, along the route, the bus predicts a channel blockage longer in time than *maxtb*, it sends a stop signal to the modulator in the hub in suitable advance; the hub suspends the transmission and the Tx buffer begins to be filled. At the bus receiver side, the Rx buffer empties out during the whole duration of the blockage period, and the video playback seamlessly continues. When the blockage period has ended, the receiver requires the hub to resume the transmission at speed r_h, up to the complete refilling of the Rx buffer; this time instant coincides with the Tx buffer emptying out, apart from the satellite propagation delay.

3.2. VoD Case. In normal mode, all works exactly as in the LBVS case. In smart mode, there is no need of the Tx buffer, because the data are already available on mass storage and they can be passed to the RSE coder at higher speed, whenever needed. The video playback delay can be reduced with respect to the one required in the LBVS case, because the Rx buffer can be filled at speed r_h, even at the start-up time.

4. Some Considerations about Interleaving

In addition to the ACM technique (e.g., MODCOD adaptation), we used UL-FEC and interleaving to improve the transmission efficiency; these techniques are particularly effective in mobile channels, which are characterized by blockage periods. These operations are performed both by adding redundant packets to produce a codeword of suitable length and by scrambling the packet order. Indeed, combining UL-FEC and interleaving allows reducing the codeword size with respect to using UL-FEC alone, just keeping the packet error rate within a target value [12]. In any case, the presence of both interleaving and UL-FEC functionalities introduces a sensible delay in the playback of the stream. In our smart mode, these functionalities are employed to compensate for channel blockages shorter than *maxtb*, while longer blockages are absorbed by the Rx buffer, which is filled before starting the playback. Therefore, there is a compromise between the interleaver depth (which, multiplied by the size of the corresponding codeword, produces the interleaving storage) and the buffer length; both these two items contribute to the maximum start-up delay (*stdelay*).

The effect of the interleaver is the spreading of a train of erasures, or incorrectly received packets, over more codewords, thus improving the error correcting capacity of the RSE code. The higher the interleaver depth, the closer the independent error characteristics are to those of bursty error channels, like the ones we consider in this paper. In fact, upon increasing the interleaver depth, the error characteristics of the environments here considered tend to be as good as in the case of independent errors (or erasures).

Under independent errors, the packet error rate for an RSE (n, k) is given by

$$p^{(\text{ind})} = \sum_{i=n-k+1}^{n} \sum_{j=0}^{n-k} \left(\frac{i-n+k+j}{k} \binom{k}{i-n+k+j} \times \binom{n-k}{n-k-j} p_e^i (1-p_e)^{n-i} \right), \tag{5}$$

where p_e is the raw packet error rate. In fact, in a block of n packets, a number of errors i that exceeds the threshold $n-k$ of possible recoverable errors, yields a residual

$$\text{PER} = \frac{i-n+k+j}{k}, \tag{6}$$

where j $(0 \leq j \leq n-k)$ expresses all possible divisions of i errors (j and $i-j$) in the redundancy and information fields, respectively.

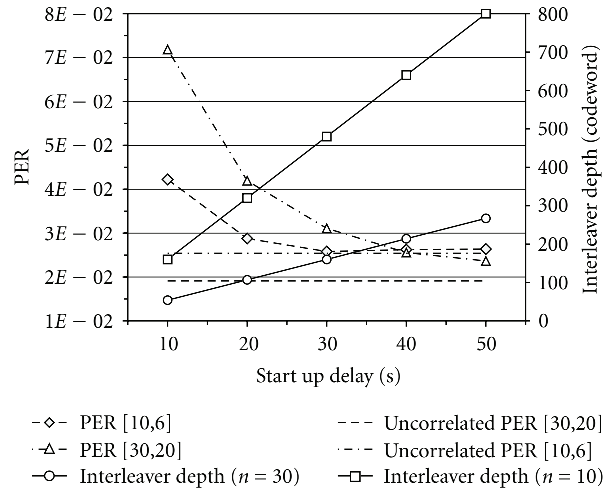


Figure 5: Residual PER and interleaver depth versus start-up delay.

In Figure 5, the residual PER characteristics are reported as a function of the start-up delay, which corresponds to a certain interleaver depth (right axis); the figure shows the highway case for different values of the codeword length and coding rate. The comparison is made between sources of independent errors that have the same average PER as the ones of the bursty error channel.

In the case of LBVS, we point out the suitability of using convolutional interleavers, which reduce the overall playout delay with respect to our block-interleaver. Convolutional (or periodic) interleavers were introduced by Forney [13] and similar techniques were proposed, independently, by Cowell-Burton and Ramsey [14, 15]. More recently, they have been investigated in [16, 17] (in the latter, also in relation with turbo codes). As shown in Figure 6, in convolutional interleavers packets from blocks are moved into shift registers before their transmission (interleaver) and before their reblocking at the Rx site (deinterleaver).

The stage of the shift registers goes from $0, 1, \ldots,$ up to $n-1$ units of m/n packets, in a triangular-wise fashion, and in such a way that the sum of the stages of the Tx and Rx registers that belong to the same transmission chain line is equal to $n-1$ units of m/n packets for all chain lines. The four deviators are simultaneously switched at each packet transmission time in a cyclic fashion and following the ascending order of the chain lines. The time interval between the channel crossings of two consecutive packets belonging to the original block is equal to m packet sending times; thus, an error train of m packets affects one packet of each of m blocks of n packets. All coded blocks of packets are delayed by $n(m-1)$ packet sending times, plus the channel latency. In block interleavers, packets are laid down in the columns of an n row by m column matrix, and sent out from rows. A block is available at the receiver side, in which another $n \times m$ matrix is read from columns, after a delay that is equal to that introduced by a convolutional interleaver. Also the time interval between two consecutive packets is the same in both cases.

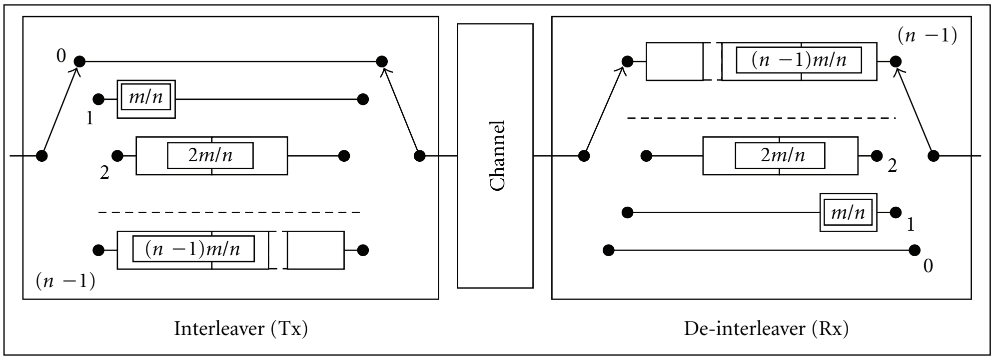


Figure 6: $n \times m$ convolutional interleaver/deinterleaver.

The difference in the overall playout delay, between the two types of interleavers, arises in the case of LBVS. Here, in fact, the transmission of the Tx matrix of packets of a block interleaver can begin only after that the first k rows of information packets have been collected by the real-time streaming flow. In the convolutional interleaver case, instead, it is sufficient to wait the time relative to n packets sending before beginning the transmission. The rationale behind the last assertion is that the block encoding time is lower than k packet transmission times.

In the VoD case, the video stream comes from a file, stored somewhere, that may be available in the memory at the Tx site in a negligible time; thus, the transmission can start almost immediately in both block and convolutional cases.

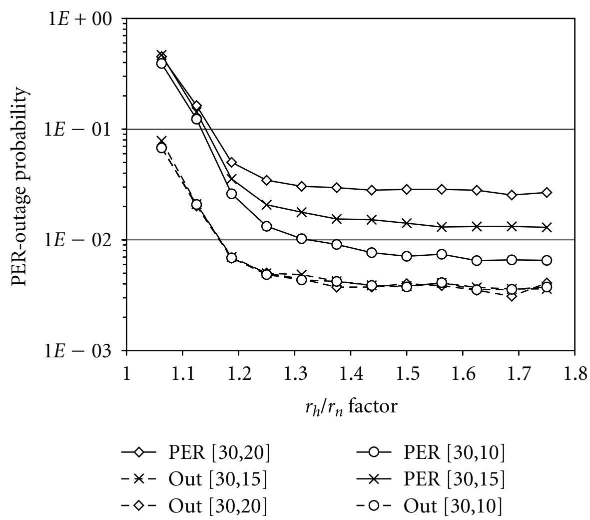


Figure 7: PER and outage probability (Out) versus r_h filling rate, for a 30-packet codeword in urban Big Good state; in the legenda, the couple (n, k) indicates the codeword length (n) and the information block length (k), respectively.

5. Benefits of Working in Smart Mode

In order to evaluate the performance of our smart system and choose the best transmission parameters, we carried out a set of simulations. Even if the simulation campaign is not exhaustive, we considered three different environments, which are the most significant ones. In fact, they differ in the landscape profile and in the average speed of the vehicle, specifically: highway, rural (generally suburban cases), and urban. In particular, we divided the urban case in two different subcases, both operating at E_s/N_0 higher than 1 dB:

(1) *Big Bad* (BB). We assume working with code rate 1/2 ($E_b/N_0 = 1$ dB, E_b/N_0 being the bit energy over two-sided noise power spectral density); in this case, we cannot distinguish the *Shadowed* state from the *Blocked* one, as they result merged together into a big bad macro state where packets are not received.

(2) *Big Good* (BG). We assume working with code rate 1/4 ($E_b/N_0 = 4$ dB); the information bit rate results 1/2 the BB one, because the redundancy is doubled, but it allows for merging *Line of Sight* and *Shadowed* state into a big good macro states where all packets are correctly received.

As far as the playout parameters are concerned, we chose the start-up delay (*stdelay*) as the maximum time the user can accept to delay the playout. This delay is composed by three items: (i) *tstore*, that is, the time used by the smart system for prebuffering the video content; (ii) the interleaving time, that is, the delay introduced by the interleaver; (iii) the UL-FEC coding/decoding time.

The value of *stdelay* strictly depends on the duration of the channel blockage periods. We chose the value of 9 seconds for both the rural and highway environments, whose value is below the key performance parameters suggested by ITU G.1010 [11]. Nevertheless, we were obliged to raise this value to 14 seconds for the urban environment, in order to get an acceptable PER level, which in [11] has been fixed below 2% for a video with data rate between 20 and 384 kps. When the blockage period overcomes s*tdelay,* we consider the residual blockage time as an outage.

The value of *tstore* results from the estimation of the best interleaver depth, which is the number of registers that are filled with a codeword each. Indeed, the best choices for both the interleaver depth and *tstore* generally depend on the applied redundancy, given a certain codeword length.

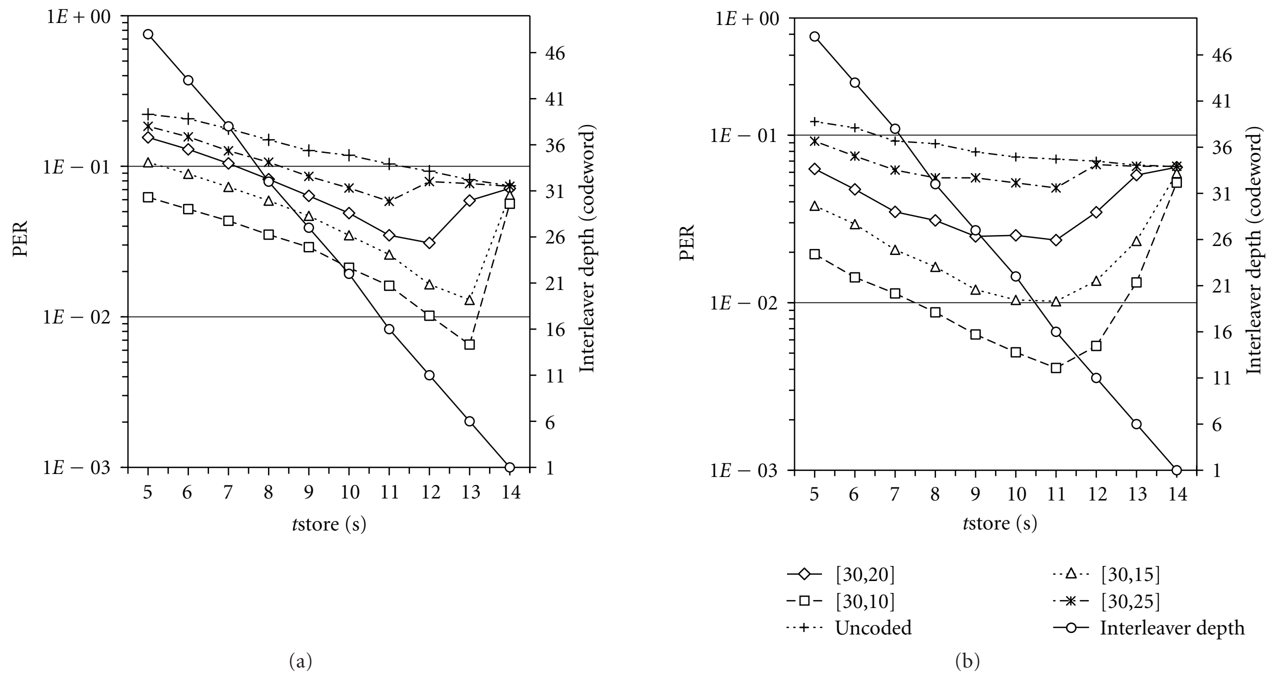


Figure 8: Packet error rate and interleaver depth versus *tstore*, for a 30-packet codeword in urban Big Bad state (a) and Big Good state (b). In the legenda, the couple (n, k) indicates the codeword length (n) and the information block length (k), respectively.

Table 3: Simulation parameters.

Parameter	Value
Packet rate	160 MPEG2-Pkt/s (241 kbps)
Modulation	QPSK
DVB-S2 LDPC coder identifier	1/2 (suburban, urban BB) 1/4 (urban BG)
DVB-S2 BBFRAME Data Field	16008 bits (urban BG) 32208 bits (suburban, urban BB)
stdelay	9 s (suburban) 14 s (urban)
gaptb	0.5 s
UL-FEC *codeword*	(30, 10) MPEG2-Pkt

Our simulation parameters are given in Table 3. We used a guard time *gaptb* in order to manage the reestablishment of the communication between the bus and the satellite after a blockage period longer than 0.5 seconds. We set *gaptb* to a value of 0.5 seconds, which seems to be reasonable; however, simulations with *gaptb* set to 1 second did not produce significant changes in the results.

The packet rate reported in Table 3 is relevant to the normal speed (r_n). The assigned r_h should be sufficient to fill the Rx buffer as fast as possible in good periods. The lower r_h, the higher the probability is that the buffer is not fully filled before a blocking period; thus, the system could not manage a blocking period as long as *tstore,* with a resulting increase in both mean PER and outage probability. Figure 7 shows PER and outage probability as functions of the r_h/r_n ratio. It is worth noting that increasing r_h over certain values does not produce any further benefit in terms of PER and outage; therefore, it results in a waste of resources.

Figure 8 shows that the best choice of *tstore* for a 30-packet codeword results to be about 11 seconds in BG, and ranges between 11 and 13 seconds in BB. By now, *tstore* is optimized for each RSE redundancy level, in order to get always the minimum PER. Once chosen *tstore*, the interleaver depth is directly determined, for a given *stdelay*.

Figures 9–12 summarize the performance of the analyzed architecture. All simulation results are obtained with a 95% confidence interval, whose width never exceeds 5% of the estimated value.

We chose two different codeword sizes: 30 and 10 packets, respectively. As resulting from the figures, the PER does not substantially differ, under the same coding rates, between the two values; however, the choice is not straightforward. In fact, while the smaller codeword is less costly in terms of coding/decoding time—and hence in power efficiency—the larger one permits choosing the redundancy with a smaller granularity. For this reason, we report results with both codeword lengths.

Figures 9 and 10 refer to the suburban case, while Figures 11 and 12 to the urban one. Each figure compares PER, computed in non-outage periods, Packet Loss Rate (PLR), which accounts for losses due to both corruption and outage periods, and bandwidth occupancy, as functions of the applied redundancy. In highway and rural environments, outage periods, that is, periods longer than the assessed 9 s of *stdelay*, are really unlikely; hence, PLR is not reported in this case, as it practically coincides with PER.

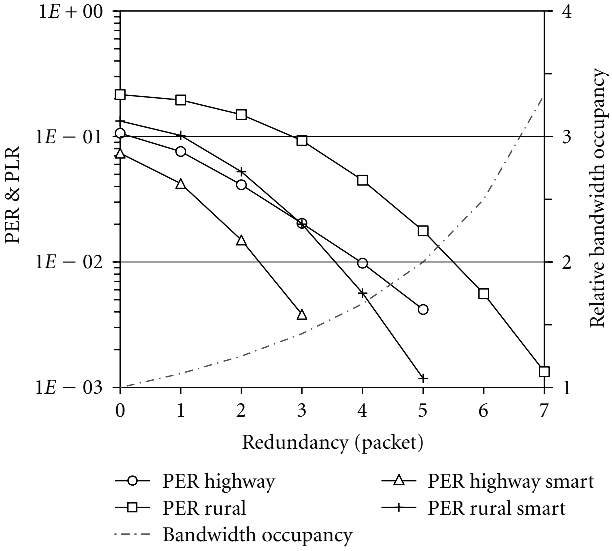


FIGURE 9: PER, PLR, and bandwidth occupancy versus redundancy with RSE $(10, k)$ in highway and rural environments.

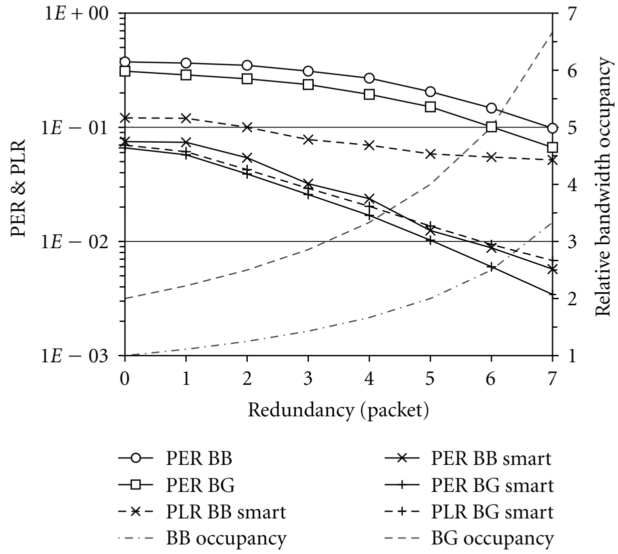


FIGURE 11: PER, PLR, and bandwidth occupancy versus redundancy with RSE $(10, k)$ in urban environment.

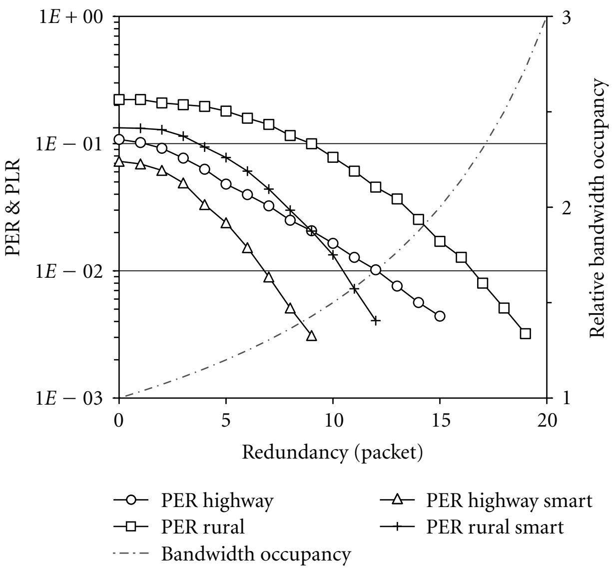


FIGURE 10: PER, PLR, and bandwidth occupancy versus redundancy with RSE $(30, k)$ in highway and rural environments.

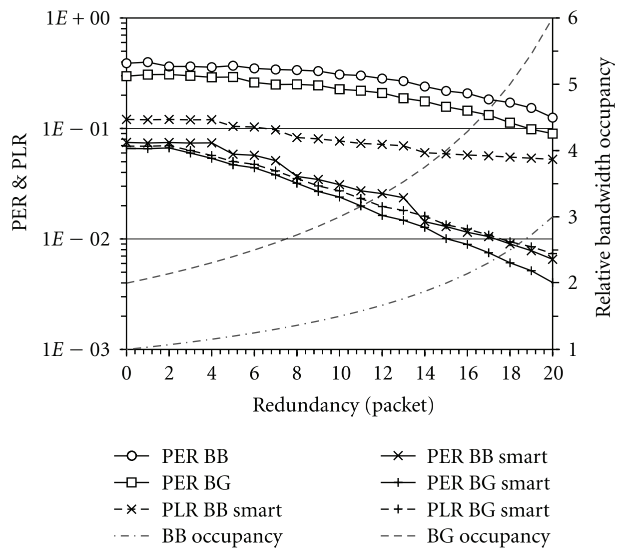


FIGURE 12: PER, PLR, and bandwidth occupancy versus redundancy with RSE $(30, k)$ in urban environment.

In the urban case, instead, very long blockage periods are possible and they impose choosing a longer *stdelay* time, in order to yield an acceptable QoS. The urban case is more complex to be illustrated and more challenging. As previously stated, we can adopt two different MODCODs—1/2 and 1/4, respectively—which produce different overheads. Hence, the relative bandwidth occupancy is very different in the two cases. As a matter of fact, Figures 11 and 12 report two different curves relevant to bandwidth occupancy in BB and BG states. However, the most relevant result is that the smart technique outperforms the sole adoption of UL-FEC and interleaving in all environments, thus consisting in a notable gain in terms of PLR.

Without the smart technique, it is impossible to get a PER below the acceptable threshold of 2%. There is, instead, a tradeoff between bandwidth occupancy and PLR in choosing between BG and BB. In fact, although the acceptable threshold is reached with 4 and 5 redundancy packets, in the BG and BB cases, respectively, the bandwidth occupancy is much higher in the former case; this is because of the adoption of 1/4 coding rate, which halves the net bandwidth with respect to the BB case. Regarding the PLR,

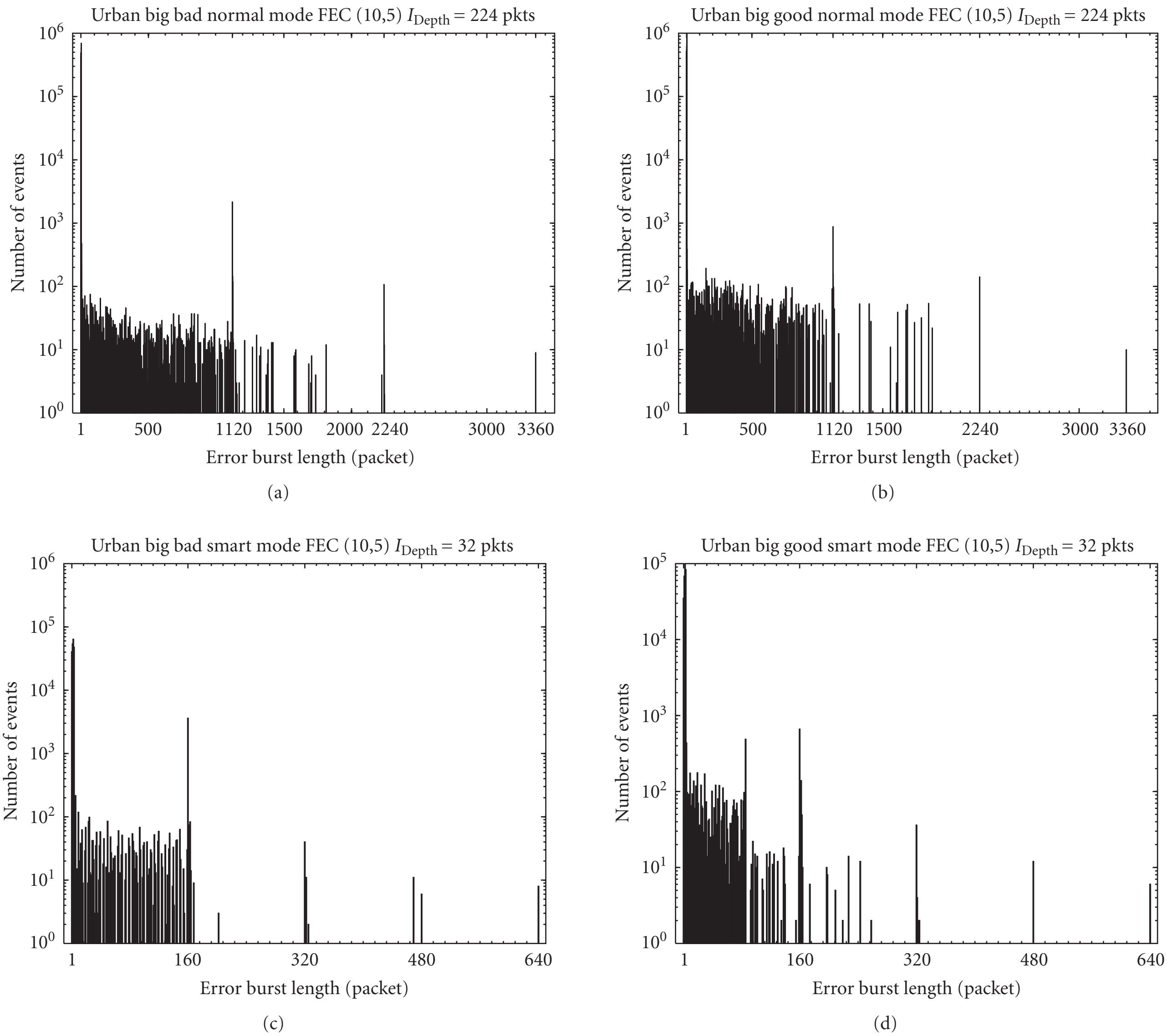


Figure 13: Error burst length distributions with FEC (10, 5) in BB and BG and operating in smart mode and normal mode, respectively.

it results much higher than the PER in the BB case, while it results very close to the PER in the BG case. This means that outage periods have a stronger influence in the BB case, due to the heavy tail distribution of blockage times in the bad state of this case (see Figure 3). Very similar results are obtained by using a 30-packet codeword (Figure 12). Here, the acceptable threshold is reached with 10 and 13 redundancy packets in the BG and BB cases, respectively, and, therefore, with coding rates close to the 10-packet codeword case.

Figure 13 reports the distribution of the error burst lengths both in BB and BG cases, when operating in normal and smart modes, respectively. Figure 14 shows the normalized distributions reported in Figure 13 in the most significant range of the x-axis; it highlights that the average error burst length is about four packets with such parameters. Figure 13 shows some periodic peaks of burst lengths, which are multiples of half of the interleaver size (note that the coding rate is 1/2); this phenomenon occurs when a burst longer than the interleaver size fills the interleaver. In fact, due to the redundancy 1/2 applied in Figure 13, only half of the packets (the information ones) are actually lost. This means that when the interleaver is filled (for example the interleaver of 2240 packets in NM) with erasures, 1120 packets of information are lost (the peak in Figure 13). This event may occur with a probability that can be derived in the PMF of Figure 3, just passing from packet to time domain.

In smart mode, the interleaver is sized in such a way to absorb bursts of errors long up to *maxtb*; in normal mode, *stdelay* determines the size of the interleaver. Therefore, *maxtb* and *stdelay* determine the positions of the peaks in

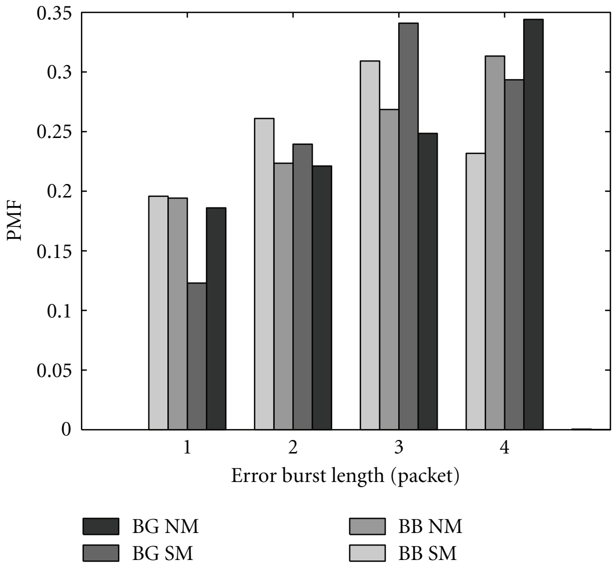


Figure 14: Zoom on the distributions of Figure 13 in the x-range of 1–5 packets with FEC (10, 5) and interleaver depth of 32 (SM) and 224 (NM) packets, respectively.

smart and normal mode, respectively. In smart mode, the error burst peaks are more than one order of magnitude shorter than in normal mode, under the same number of events.

6. Conclusions

We proposed a smart architecture for DVB-S2 vehicular connectivity, which allows delivering multimedia contents to fleets of buses. In particular, we studied the video streaming transmissions from a hub station, anywhere located, towards a specific bus reached via a satellite link. The system takes advantage of feedback and prediction of obstacles on the buses' routes provided by GPS navigation devices, installed on board the buses. In addition to classical error recovery techniques, such as forward error correction and interleaving, the proposed smart system enhances the quality of service, by reducing packet error rate and outage periods in both urban and extraurban environments. The proposed technique may be employed as an alternative to gap-fillers, whenever the gap is sufficiently limited in time. This study aims at fostering the development of new communication and entertainment services, as well as monitoring and security issues on future public transportation vehicles, by adapting the most suitable technologies to existing telecommunication infrastructures.

Acknowledgment

This Work was funded by the European Commission in the framework of the SatNEx II Network of Excellence (NoE) FP6 project (Contract no. 027393).

References

[1] EN 302 307, "Digital video broadcasting: second generation framing structure, coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications," ETSI, DVBS2-74r15.

[2] "Real Time Streaming Protocol," http://tools.ietf.org/html/rfc2326.

[3] http://www.openiptvforum.org.

[4] ESA Project, "Feasibility study of a mobile Ku-band terminal," Contract No. 15593/01/NL/DS, Final Report, http://telecom.esa.int/feasibility_ndsatcom.

[5] S. Scalise, H. Ernst, and G. Harles, "Measurement and modeling of the land mobile satellite channel at Ku-band," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 693–703, 2008.

[6] J. Nonnenmacher, E. W. Biersack, and D. Towsley, "Parity-based loss recovery for reliable multicast transmission," *IEEE/ACM Transactions on Networking*, vol. 6, no. 4, pp. 349–361, 1998.

[7] http://www.mathworks.com/products/commblockset/demos.html?file=/products/demos/shipping/commblks/commdvbs2.html.

[8] R. Gallager, "Low density parity check codes," *IRE Transactions on Information Theory*, no. 1, pp. 21–28, 1962.

[9] I. S. Reed and X. Chen, *Error-Control Coding for Data Networks*, Kluwer Academic Publishers, Boston, Mass, USA, 1999.

[10] K. Fujimoto, S. Ata, and M. Murata, "Adaptive playout buffer algorithm for enhancing perceived quality of streaming applications," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '02)*, vol. 3, pp. 2451–2457, Taipei, Taiwan, November 2002.

[11] ITU-T Recommendation G.1010, "End-user multimedia QoS categories,"http://www.itu.int/ITU-T/.

[12] P. Barsocchi and A. Gotta, "Experimental video broadcasting in DVB-RCS/S2 with land mobile satellite channel: a reliability issue," in *Proceedings of the IEEE International Workshop on Satellite and Space Communications (IWSSC '08)*, pp. 234–238, Toulouse, France, October 2008.

[13] G. D. Forney Jr., "Burst-correcting codes for the classic bursty channel," *IEEE Transactions on Communications*, vol. 19, pp. 772–781, 1971.

[14] W. R. Cowell and H. O. Burton, "Computer simulation of the use of group codes with retransmission on a Gilbert burst channel," *AIEE Transactions on Communications and Electronics*, vol. 80, pp. 577–585, 1962.

[15] J. L. Ramsey, "Realization of optimum interleavers," *IEEE Transactions on Information Theory*, vol. 15, no. 3, pp. 338–345, 1970.

[16] S. A. Hanna, "Convolutional interleaving for digital radio communications," in *Proceedings of the 2nd International Conference on Universal Personal Communications (ICUPC '93)*, pp. 443–447, Ottawa, ON, Canada, October 1993.

[17] S. Vafi, T. Wysocki, and I. Burnett, "Convolutional interleaver for unequal error protection of turbo codes," in *Proceedings of the DSPCS & WITSP*, pp. 485–491, Coolangatta, Gold Coast, Australia, December 2003.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 329073, 9 pages
doi:10.1155/2009/329073

Research Article

Modeling and Performance Analyses of Hybrid Cellular and Broadcasting Networks

Peter Unger and Thomas Kürner

Institute for Communications Technology, Braunschweig Technical University, Schleinitzstrasse 22, 38106 Braunschweig, Germany

Correspondence should be addressed to Thomas Kürner, t.kuerner@tu-bs.de

Received 1 March 2009; Revised 28 October 2009; Accepted 30 November 2009

Recommended by Sandro Scalise

Mobile communication services are getting more and more important and, in particular, multimedia services have attracted the interest of the users. Mobile TV is one of the most demanded candidates. Powerful and efficient communication systems are needed, which provide high capacities, especially at the downlink. Furthermore, interactivity is essential for supporting the user needs and to extend the service offering. As one possible solution to meet the mentioned requirements, we consider the combination of the cellular network UMTS and the mobile broadcast network DVB-H, which form a hybrid network. We investigate the performance of hybrid networks and develop a system model, which describes the hybrid network and the load switching between both networks. One of the contributions is the definition of the switching bound concept, which represents an efficient tool to assess the necessity and the feasibility of hybrid networks and the amount of load switching. The performance indicators cell load and grade of service are analyzed by using theoretical and realistic scenarios.

1. Introduction

With the introduction of the 3rd Generation cellular network Universal Mobile Telecommunications System (UMTS), higher data rates have been made possible and the variety of services has been increased, compared to the previous cellular networks such as General Packet Radio Service (GPRS). In addition to the traditional services, such as voice telephony and messaging, the multimedia services and high data rates have been made possible and accepted by the users. The new offered services, for example, video telephony, Internet-based data communications, and streaming services, make high downlink capacities and efficient networks necessary.

The currently available terminals are equipped with several receivers, which allow for the reception of data from different networks, such as UMTS and DVB-H. Hybrid networks are possible for these types of terminals and the broadcast downlink channel provided by DVB-H can be used very efficiently to serve many users at the same time with high data rate services.

The DVB-H technology is based on the terrestrial digital TV system DVB-T, which has been successfully introduced in many countries [1, 2]. The signal has been made more robust in order to enable mobile reception with high velocities and at mobile environments, for example, to cope with multipath reception. An additional error protection scheme has been included at the link layer, called Multiprotocol Encapsulation Forward Error Correction (MPE-FEC) [3]. Furthermore, the time slicing concept has been introduced for power saving at the terminal side, while services are transmitted in time slots and the receiver front end can be switched off in the meantime. Furthermore, the time slicing approach enables soft handover and optimized handover algorithms [4], and local content areas in single frequency networks [5]. Depending on the modulation and coding scheme, and the used channel bandwidth, a total downlink data rate of about 6–10 Mbps can be achieved. The individual user data rate depends on the time slicing setup and is typically set to 100–400 kbps.

The currently deployed cellular system UMTS is standardized by the 3rd Generation Partnership Project (3GPP) and makes use of the Wideband Code Division Multiple Access (W-CDMA) technology. All users are served at a common frequency with a 5 MHz bandwidth and are separated by orthogonal codes. The network capacity and coverage suffer from interference and have to be planned carefully. The network is deployed in a cellular structure, which is dense in urban environments and larger cells occur in less populated areas. Mobile terminals are assigned to one or more cells, which are established by the Node-Bs. Different bearer services can be used in a circuit or packet switched mode and thus different types of user services are available, such as voice telephony, web browsing, streaming multimedia, and file download.

The combination and the cooperation of both network types has been enabled by the IP Datacast standard, which has been specified within the International DVB Project [6]. It defines the higher layers and form an Internet Protocol- (IP-) based end-to-end system, together with the DVB-H specification as the physical layer [7]. In addition to the broadcast downstream, the IP Datacast reference architecture defines an optional link with an interactive cellular network, which is used, for example, for service requests. In terms of hybrid networks, the connection to a cellular network is defined as a mandatory feature.

The combination of both network types has several advantages. On the one hand, popular content, which is requested by several users at the same time, can be switched from UMTS to a broadcast transmission for a more efficient delivery. The UMTS network can be unloaded in order to avoid a cell overload or to reduce the overall cell power and the intercell interference. On the other hand, the broadcast services are enhanced by an uplink channel and new features are possible, such as on-demand services.

Hybrid networks do not exist nowadays. But at this stage, the deployment and operation are possible, since the fully standardized and currently deployed networks UMTS and DVB-H, the IP Datacast standard, and load balancing algorithms as, for instance, proposed in [8] exist. Nevertheless, efficient models are still needed for hybrid network planning and network performance analysis.

We propose a performance model, performance criteria, and constraints for load switching in hybrid network, which is called the switching bound concept. The performance is shown by applying different extents of load switching for a theoretical and a realistic scenario.

This paper is structured as follows. In Section 2 the services and the scenario are defined, which have been considered for our analyses. Section 3 describes models for the individual networks UMTS and DVB-H. In Section 4, the model for hybrid networks is derived, the network performance is evaluated, and the switching bounds are defined. In Section 5, the simulation results are shown by means of a realistic network scenario.

2. Definition of Services and Scenario

As mentioned above, several types of services are enabled by UMTS, and eight different types have been summarized in [9]. The defined service set S includes voice and video telephony, web browsing and email, location-based services, short/multimedia massaging service, file download, and streaming multimedia. It is assumed, that some of these services, namely, *file download* and *streaming multimedia*, are suitable to be transmitted by both network types, UMTS and DVB-H. For this subset of services, denoted as $S^{(H)}$, load switching can be applied by transferring the transmission from UMTS to the broadcast network. In this paper, we focus on the streaming multimedia service with an average stream data rate of 128 kbps. It has been turned out that this service generates a much higher contribution to the cell load, compared to the file download service. The other six services are called cellular services and have always to be delivered by the UMTS network.

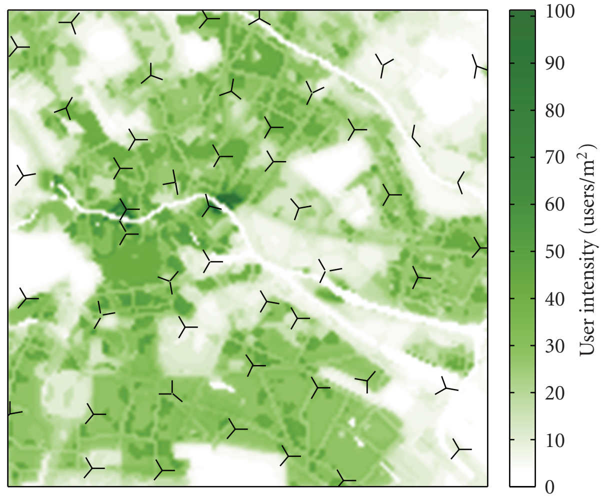


Figure 1: UMTS network structure and user intensity for the streaming multimedia service in a 7.5 km ×7.5 km scenario with a resolution of 50 m. The network is based on the public reference scenario of Berlin available at [10].

In addition to the definition of the service characteristics and the requirements, several scenarios have been proposed in [10], which contain reference network architectures with reference sites, antenna configurations, and predictions of the signal propagation loss. Furthermore, realistic traffic demands have been provided in the form of user intensity maps, which describe the spatial distribution of the service request rate.

We select the public reference scenario of Berlin for our investigation, which is publicly available at [11]. Figure 1 shows the selected UMTS sites and the user intensity of the streaming multimedia service for the entire scenario area. A subset of 48 sites has been selected from the reference scenario, in which the angles of azimuth and elevation have not been modified. However, new signal prediction maps have been calculated with a more sophisticated propagation model, based on [12].

3. Modeling the UMTS and the DVB-H Network

The UMTS network is a bidirectional network, which has to be planned for both uplink and downlink. In general, the downlink part is the more critical one, due to a higher amount of traffic demand, caused by asymmetric services. The hybrid network can be beneficial for the downlink part, if the additional capacity and the combing capability of the broadcast network are used efficiently. It is assumed that all users are connected to the UMTS network at the uplink, regardless of the way of downlink delivery. Thus, load switching does not effect the uplink and traditional models can be applied. Therefore, we focus on the modeling of the downlink part.

3.1. Modeling the UMTS Downlink. An efficient model for UMTS network evaluation has been proposed in [13] and described in detail in [14]. We review this model in this section, since it is the basis for developing the model for hybrid networks. The UMTS model describes the service requirements, the generated user load, the intracell interference and the interference coupling between cells, which is caused by intercell interference. It provides an efficient estimation of the necessary transmit power for the mobile terminals and the Node-Bs by solving a linear equation system, which is called interference coupling system.

The model can be used in conjunction with snapshot analyses using Monte-Carlo simulations. Several details can be implemented with this approach and thus, it generates accurate results for a high number of snapshots. But it is computational expensive. A more efficient approach has been proposed in [15] and is described more in detail in [14], which uses expected values for the user demand, instead of individual users. Thus, it replaces the computational expensive snapshot approach with an average view on the current network load. The drawback is that the accuracy of the results is somewhat decreased and some details are not modeled, for example, different velocities.

In the context of this work, both methods have been considered. First, the downlink of hybrid networks is modeled and the principal behavior of load switching is shown by means of the expected approach. For network simulations performed in Section 5, a realistic network and the snapshot approach are applied.

For determining the cell load and the coupling between the cells, mobile terminals are assigned to the Node-Bs by the best server approach, whereas that cell is selected, which provides the strongest pilot (CPICH) signal. The best server area A_i of cell i is determined by using the pilot channel power $p_i^{(\mathrm{CPICH})}$ and the end-to-end channel gain $\gamma_i(x)$ from antenna i to location x, which basically includes the propagation loss and antenna gains.

Each service requires a specific carrier-to-interference-and-noise ratio (CINR) target, denoted as μ_s, which includes the processing gain of the selected bearer service and the required E_b/N_0 target [13]. The activity of the bearer usage is denoted by α_s and the orthogonality loss factor of a location $x \in A_i$ is denoted by $\omega(x)$. The orthogonality of the signals is partly lost due to multipath propagation effects and thus, intracell interference is caused. The caused load of service s at location x is calculated by

$$l_s(x) := \frac{\alpha_s \mu_s}{1 + \omega(x)\alpha_s \mu_s}. \tag{1}$$

The interference coupling approach defines the coupling factors for cell i (2) as expected values, which form the interference-coupling downlink matrix $\mathbf{C} = (c_{ij})_{i,j\in\{1,\ldots,N_{\mathrm{cell}}\}}$. The parameter $T_s(x)$ represents the user intensity of service s at location x:

$$\begin{aligned} c_{ii} &= \int_{x\in A_i} \omega(x) \sum_{s\in S} l_s(x) T_s(x)\,dx, \\ c_{ij} &= \int_{x\in A_i} \frac{\gamma_j(x)}{\gamma_i(x)} \sum_{s\in S} l_s(x) T_s(x)\,dx. \end{aligned} \tag{2}$$

The coupling system for all cells i can be written as

$$\overline{p}_i = c_{ii}\overline{p}_i + \sum_{j\neq i} c_{ij}\overline{p}_j + \overline{p}_i^{(\eta)} + p_i^{(CC)} \tag{3}$$

with $\overline{p}_i^{(\eta)} = \int_{x\in A_i} (\eta(x)/\gamma_i(x)) \sum_{s\in S} l_s(x) T_s(x)dx$ being the average required power to overcome the noise in a non interfered cell. The constant power values of all common channels including the pilot channel are summed up in $p_i^{(CC)}$. The parameter $\overline{p}_i$ denotes the average transmit power of cell i in order to serve all users and to overcome noise and interference.

The resulting power values $\overline{p}_i$ must not exceed the maximum feasible power $p_{\max}$. In order to avoid cell overloading caused by high traffic demand and high interference, the scaling factor $\lambda_i \in [0, 1]$ has been proposed in [14]. It is used to scale the coupling matrix with $\mathrm{diag}(\lambda)\mathbf{C}$ in order to reduce the traffic demand per cell and the coupling to the other cells. This approach assumes perfect load control with applying an exact downgrade of cell load. The scaling factor λ_i represents the *Grade of Service* (GoS); that is, the fraction of served traffic compared to the offered traffic in cell i, and $(1 - \lambda_i)$ defines the blocking probability of cell i. In [14] an iterative process has been defined to estimate the parameters λ_i and $\overline{p}_i$ for each cell. The interference coupling concept assumes perfect power control, which enables to precisely assign the necessary power to each user.

Another measure of the quality in network planning is the average other-to-own interference $\overline{\iota}_i$, which describes the total intercell interference to cell i (4). Note that this definition has to be divided by the average orthogonality loss (e.g., $\overline{\omega} = 0.673$ for urban environments) in order to be comparable to traditional definitions, for example, in [13]:

$$\overline{\iota}_i = \frac{\sum_{j\neq i} c_{ij}\overline{p}_j}{c_{ii}\overline{p}_i}. \tag{4}$$

In [14], the so-called pole equations have been defined to calculate the necessary average transmit power $\overline{p}_i$ (5) and the grade of service $\lambda_i^{(C)}$ (6) per cell with a given other-to-own interference. We have used the notation$^{(C)}$ to denote the cellular network, since we have also defined the GoS criteria for the broadcast$^{(B)}$ and the hybrid$^{(H)}$ network:

$$\overline{p}_i = \begin{cases} \dfrac{p^{(CC)} + \overline{p}_i^{(\eta)}}{1-(1+\bar{\iota}_i)c_{ii}} & \text{if } (1+\bar{\iota}_i)c_{ii} < \dfrac{p_{\max} - p^{(CC)} - \overline{p}_i^{(\eta)}}{p_{\max}}, \\ p_{\max}, & \text{otherwise,} \end{cases} \tag{5}$$

$$\lambda_i^{(C)} = \begin{cases} 1 & \text{if } (1+\bar{\iota}_i)c_{ii} < \dfrac{p_{\max} - p^{(CC)} - \overline{p}_i^{(\eta)}}{p_{\max}}, \\ \dfrac{p_{\max} - p^{(CC)}}{p_{\max}(1+\bar{\iota}_i)c_{ii} + \overline{p}_i^{(\eta)}}, & \text{otherwise.} \end{cases} \tag{6}$$

The transmit power $\overline{p}_i$ and the GoS λ_i^C form a complementary system. The transmit power increases until the offered traffic load can be served or the maximum power is reached. Additional traffic load is blocked, which reduces the GoS of the cell. In this work, alternative concepts of solving a network overload as described in [16] are not considered.

The GoS of the entire UMTS network $\overline{\lambda}^{(C)}$ is estimated by a weighting sum, considering the offered traffic load $\tau_i = \int_{x\in A_i} \sum_{s\in S} l_s(x) T_s(x) dx$:

$$\overline{\lambda}^{(C)} = \frac{\sum_i \tau_i \lambda_i^{(C)}}{\sum_i \tau_i}. \tag{7}$$

The utilization of the transmit power in each cell is measured with the downlink cell load. It is defined by the ratio of the necessary downlink transmit power compared to the total available maximum power: $L_i^{(C)} = \overline{p}_i / p_{\max}$.

3.2. Modeling the Broadcast Network. In contrast to the UMTS network, the required system capacity of broadcast network is independent on the actual number of users. The necessary amount of capacity depends on the number of different content items, that are requested by the users [17].

The requests for different content have different probabilities; that is, some items are more popular compared to others. A common description of the popularity is the Zipf-distribution defined in (8). Detailed analyses and measurements have been performed and described in [18].

The probability that a user request corresponds to item i is defined as

$$P(X = i) = \frac{1}{\Omega} i^{-\kappa} \quad \text{with } \Omega = \sum_{i=1}^{N_{\text{item}}} i^{-\kappa}. \tag{8}$$

The parameter N_{item} represents the number of items offered by the content provider and $\kappa \geq 0$ is the shape parameter of the popularity distribution. For a value of $\kappa = 0$, the requests are uniformly distributed to all items. For an increasing value of κ, the popularity concentrates more and more on specific items.

As mentioned above, we consider the streaming multimedia service for hybrid delivery, whereas several requested streams are switched to the broadcast network. Since streaming is a real-time service, the broadcast capacity is shared among all switched streams in a parallel transmission. Each stream requires a constant data rate d_i in order to guarantee sufficient Quality of Service (QoS).

We define the broadcast cell load $L^{(B)}$ in (9), which depends on the number of delivered streams and the capacity $D^{(B)}$, reserved for switchable services. If $L^{(B)}$ exceed one, the broadcast cell is overloaded, some streams have to be blocked and the corresponding users cannot be served:

$$L^{(B)} = \sum_{i=1}^{n} \frac{d_i}{D^{(B)}}. \tag{9}$$

The maximum number of streams, which can be delivered in parallel, is calculated by $n_{\max} = \lfloor D^{(B)}/\overline{d} \rfloor$ with $\overline{d}$ being the average required data rate of all streaming services.

4. Modeling Hybrid Networks

In a hybrid system, the delivery of services can be switched between the downlinks of both networks, UMTS and DVB-H. The overall network performance can be optimized with properly applying load switching.

The load switching in hybrid networks is modeled by specifying the amount of traffic, which is switched from the UMTS network to the broadcast network. In contrast to GoS scaling, the switched traffic is not lost, but served by the broadcast system. Thus, the UMTS network is unloaded, while improving the overall grade of service.

First of all, the offered traffic has to be classified into the cellular service demand and switchable services (see Section 2). Since the coupling matrix $\mathbf{C}$ is generated by summing up the user demand for each service, this matrix can be split into the cellular $\mathbf{C}^{(C)}$ and the hybrid part $\mathbf{C}^{(H)}$. All elements in the matrices can be calculated by using (2) and the appropriate set of services $S/S^{(H)}$ and $S^{(H)}$.

We define the load share parameter β, which specifies the fraction of cellular (nonswitchable) traffic load as $\beta = c^{(C)}/(c^{(C)} + c^{(H)})$. For $\beta = 0$, it is possible to switch the entire traffic load to the broadcast network and for $\beta = 1$, no switchable traffic load exists at all. The ratio parameter β is dependent on the location. An average value of about 43% occurs for evaluating the area of the Berlin reference scenario by considering the streaming multimedia service for load switching. That means, about 57% of the traffic can be used for unloading the UMTS network.

4.1. Definition of Load Switching. The fraction of traffic load is estimated, which is actually needed to be switched to the broadcast network for a successful unloading of the UMTS network. For example, a typical target criteria for network planning is to achieve a cell blocking rate of 2% or less, which corresponds to $\lambda_i \geq 0.98$.

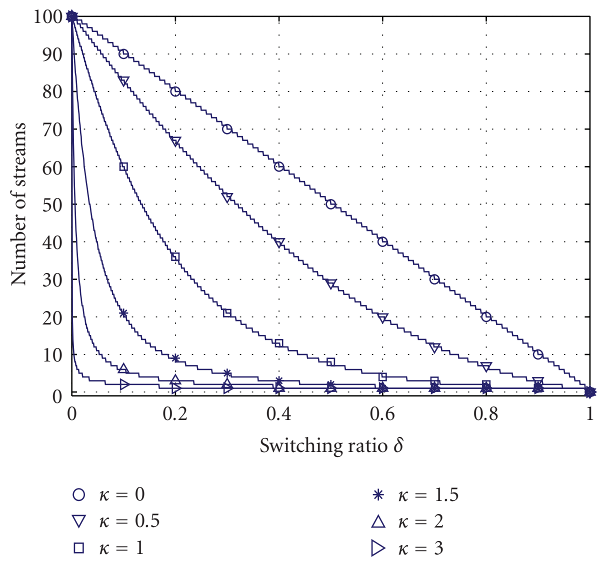


Figure 2: Comparing the switching parameter δ with the number of streams n, which are offered to the broadcast system, depending on the popularity distribution.

We define the switching parameter $\delta_i \in [0,1]$, which is used to control the amount of traffic load, switched between both networks. A value of $\delta_i = 0$ indicates that all requests for the switchable service are switched to the broadcast network. For example, all streaming services are served by the DVB-H network. Therewith, we define $\tilde{c}_{ij}$ as the new coupling factor for the UMTS network with

$$\tilde{c}_{ij} = c_{ij}^{(C)} + \delta_i c_{ij}^{(H)}. \quad (10)$$

It has been shown in [8] that those service items are beneficial for optimizing hybrid networks, which causes a high load in the UMTS network. The product of arrival rate and required resources (i.e., data rate) has been taken as a measure to define the switching order. In this work, we assume equal data rates for all streams and therefore, the item request rate is directly proportional to its popularity. Thus, we assume that the most popular items are switched first.

As mentioned before, both networks can be described by different schemes, the user and content dependence. In order to estimate the broadcast cell load, we have to transform the applied load switching with the switching parameter δ into the corresponding number of streams n_δ, which have to be delivered by the broadcast network. The fraction $(1-\delta)$ corresponds to the n most popular items. Thus, we look for the minimum n, which fulfils $1-\delta \le P(X \le n)$. In other words, the $(N_{\text{item}} - n)$ low popular items have to be served by the UMTS network. Thus, the relation between n and δ can be defined as

$$n_\delta = \min_n \{ n \mid P(X > n) \le \delta \}. \quad (11)$$

On the other hand, the estimation of δ for a given number of streams n, which have to be switched, can be described by $\delta_n = P(X > n)$.

In Figure 2, the number of streams is shown, which is offered to the broadcast system, depending on δ and the popularity shape parameter κ. The total number of items is set to $N_{\text{item}} = 100$. It can be seen that, with $\kappa = 0$, all items have the same popularity and thus, n increases linearly for more broadcast traffic. For an increasing κ, a higher fraction of traffic corresponds to the more popular items. The broadcast technology allows for combining user requests to the corresponding high popular items. Thus, for the same δ, the broadcast network is less loaded for higher values of κ.

4.2. Definition of Switching Bounds. Depending on the system parameters and the traffic demand, the possible range of switching may be limited, if cell load values of $L^{(B)} \le 1$ and $L^{(C)} \le 1$ have to be guaranteed. We introduce the switching bound concept, which defines a bound for δ for each network separately. In order to apply reasonable load switching in a hybrid network, the broadcast network should define the lower and the UMTS network the upper bound of δ. The switching bounds indicate the necessity and feasibility of applying a hybrid network. It is obvious, if no bound has been set by the UMTS network, no unloading is required.

4.2.1. Broadcast Bound. The bound $\delta^{(B)} \in [0,1]$ is defined by the parameters of the broadcast network and the streaming service. Since the broadcast capacity is limited, the number of parallel transmittable streams is limited too. If the broadcast cell load defined in (9) equals one, the maximum number of streams $n_{\max}$ is reached. The broadcast bound $\delta^{(B)}$ is calculated by

$$\delta^{(B)} = P(X > n_{\max}) = P\left(X > \left\lfloor \frac{D^{(B)}}{\bar{d}} \right\rfloor \right). \quad (12)$$

Therefore, $\delta^{(B)}$ depends on the broadcast capacity $D^{(B)}$, the service data rate d_i, and the popularity distribution, that is, N_{item} and κ.

Figure 3 shows $\delta^{(B)}$ depending on the popularity shape parameter κ and the broadcast cell capacity $D^{(B)}$. The number of items is set to $N_{\text{item}} = 100$ and the average streaming data rate is $\bar{d} = 128$ kbps. It is obvious that for higher broadcast data rates, more traffic can be switched to the broadcast network and $\delta^{(B)}$ is decreased and less restrictive. For a more even popularity distribution (smaller values of κ), the bound increases due to a higher variety of the requested streams and thus, a higher broadcast capacity demand. For a very large values of κ, the popularity corresponds to a very small number of items, in extremum to one single item. Thus, all requests can be delivered by the broadcast network. For $D^{(B)} = N\bar{d} = 12.8$ Mbps, it is possible to serve all items by the broadcast network, thus $\delta^{(B)} = 0$ and no restrictions exist from the broadcast network side. For a too small capacity, that no stream can be transmitted, the broadcast bound equals one for any value of κ.

4.2.2. Cellular Bound. The cellular bound $\delta_i^{(C)} \in [0,1]$ describes the minimum necessary unloading of cell i by switching streaming services to the broadcast network, until

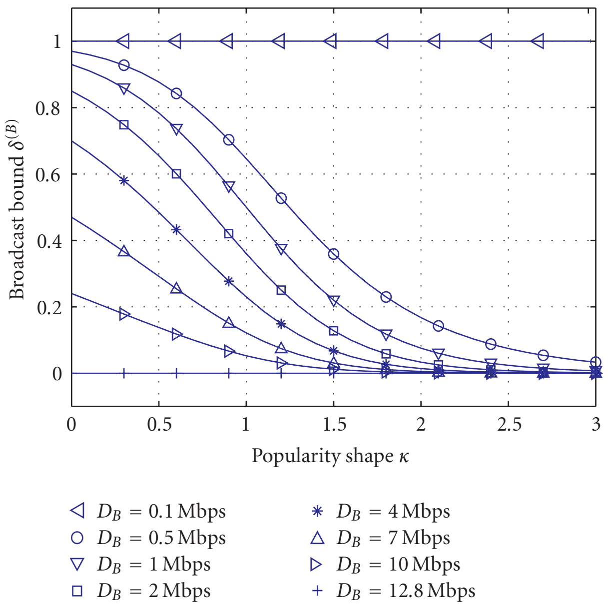


FIGURE 3: Broadcast bound $\delta^{(B)}$ depending on the popularity shape κ and the total broadcast capacity.

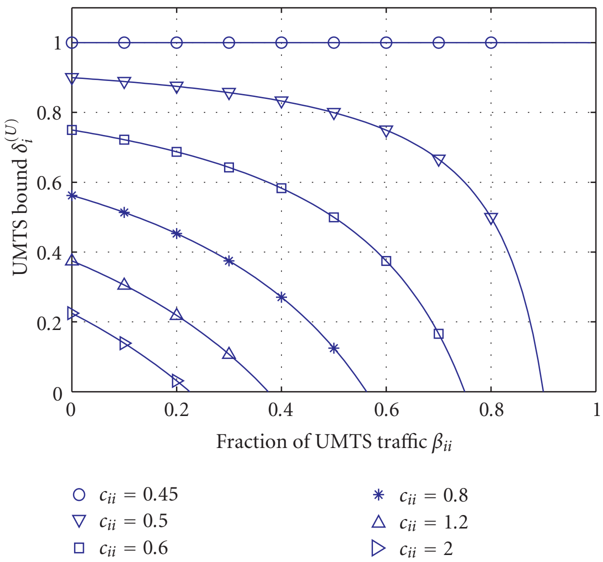


FIGURE 4: Cellular bound $\delta_i^{(C)}$ for cell i depending on the load share parameter β and the total offered traffic load.

the GoS target $\hat{\lambda}$ is reached. In this work, we use $\hat{\lambda} = 0.98$. Once the cell is able to handle the amount of offered traffic, a further unloading yields to a reduction of the required transmit power $\tilde{p}_i$. In order to estimate $\delta^{(C)}$, we use (6) with $\tilde{c}_{ii} = c_{ii}^{(C)} + \delta_i c_{ii}^{(H)}$ and $\tilde{p}_i^{(\eta)} = p_i^{(\eta,C)} + \delta_i p_i^{(\eta,H)}$, and transform it to δ_i:

$$\delta_i^{(C)} = \begin{cases} \dfrac{p_{\max} - p^{(CC)} - \hat{\lambda} p_{\max}(1+\bar{\imath}_i)c_{ii}^{(C)} - p_i^{(\eta,C)}}{\hat{\lambda} p_{\max}(1+\bar{\imath}_i)c_{ii}^{(H)} + p_i^{(\eta,H)}}, & \lambda_i^{(C)} < \hat{\lambda}, \\ 1, & \hat{\lambda} \le \lambda_i^{(C)} \le 1. \end{cases} \quad (13)$$

It is obvious that a solution is only possible, if switchable traffic exists in cell i, so that $c_{ii}^{(H)} > 0$. If all services are switchable ($\beta = 0$) and the GoS target is set to one, the bound $\delta_i^{(C)}$ becomes equal to the GoS $\lambda_i^{(C)}$.

In the following, we investigate a theoretical scenario of a single cell, in which the influence from the surrounding cells, that is, the average other-to-own interference $\bar{\imath}_i$, is assumed to be constant. The curves in Figure 4 depict the cellular bound $\delta^{(C)}$. The bound is shown for different cell loads c_{ii} and depending on load share parameter β_{ii}. For $\beta_{ii} = 1$, no hybrid traffic exists and no load switching can be applied. For $\beta_{ii} = 0$, all services can be transmitted by both network types. This case shows the less restrictive bound values, since no cellular background traffic exists.

In this example, the following parameters have been assumed: $p_{\max} = 14\,\mathrm{W}$, $p^{(CC)} = 4\,\mathrm{W}$, $\bar{\imath}_i = 0.6$, and $p^{(\eta)} = 0.96\,\mathrm{mW}$. If $\lambda_i^{(C)} \ge \hat{\lambda}$, no traffic needs to be switched, since the QoS target is already achieved without applying unloading. The threshold in our case is $c_{ii} = 0.45$. This threshold can be calculated by using (6) with $\lambda_i^{(C)} = \hat{\lambda}$. Thus, $\delta_i^{(C)}$ is set to one for all values of c_{ii}, which are lower or equal as this threshold.

In general, it can be seen that $\delta_i^{(C)}$ decreases for higher cell load, since the cell needs to be unloaded more and more. For high traffic loads, the proportion of switchable traffic needs to be high, that is, low values of β, in order to apply sufficient load switching for successfully unload the cells.

4.2.3. Discussion on Switching Bounds. We have defined two bounds for the load switching, which depend on the offered traffic, the load share, and on the parameters of the broadcast and UMTS network, respectively. The broadcast bound is valid for the entire broadcast cell. Each UMTS cell defines an individual bound, depending on the current loading, the grade of service, and the cell parameters.

In order to successfully apply unloading in a hybrid network the broadcast bound needs to be lower than the cellular bound, $\delta^{(B)} \le \delta^{(C)}$. All UMTS cells, which are covered by a broadcast cell and which are considered for a hybrid mode, need to be evaluated. Therefore the minimum of all $\delta_i^{(C)}$ has to be considered, if the quality target has to be reached. Nevertheless, strategies for network planning and estimating the appropriate cellular bound are beyond the scope of the paper.

If no broadcast bound is defined, the necessary broadcast data rate can be determined. The estimated switching parameter δ is based on the evaluation of the UMTS cells. It defines the necessary capacity of the broadcast network, which can be estimated by the number of streams to be transmitted (11). By summing up the required data rates of all items switched to broadcast the necessary broadcast data rate can be calculated.

4.3. Grade of Service in Hybrid Networks. The grade of service of a broadcast network can be defined in two ways, a content- or a user-based description. The user-based definition has been selected, due to compatibility reasons to the cellular network and for a common GoS estimation of hybrid networks. It is estimated by evaluating the probability of the blocked streams.

If $\delta < \delta^{(B)}$, the capacity of the broadcast network is not sufficient for transmitting the switched content and users have to be blocked. The probability of not served streams is

$$\begin{aligned} P(X \le n_\delta) - P(X \le n_{\max}) &= (1-\delta) - \left(1-\delta^{(B)}\right) \\ &= \delta^{(B)} - \delta. \end{aligned} \tag{14}$$

Thus the grade of service of the broadcast network is calculated by $\lambda^{(B)} = \min\{1, 1-(\delta^{(B)}-\delta)\}$. It is obvious that for $\delta \ge \delta^{(B)}$, no blocking occurs and $\lambda^{(B)} = 1$.

In order to estimate the grade of service $\lambda_i^{(H)}$ for the hybrid network, the GoS parameters $\lambda_i^{(C)}$ and $\lambda^{(B)}$ have to be combined. Equation (15) describes the ratio of the traffic served by the hybrid network compared to the total offered traffic. This ratio is transformed into a weighted term; whereas the weights describe the fraction of cellular and switchable traffic:

$$\begin{aligned} \lambda_{i,\delta}^{(H)} &= \frac{\widetilde{\lambda}_{i,\delta}^{(C)}\left(c_{ii}^{(C)} + \delta c_{ii}^{(H)}\right) + \lambda_\delta^{(B)}(1-\delta)c_{ii}^{(H)}}{c_{ii}} \\ &= \left[\delta(1-\beta_{ii}) + \beta_{ii}\right]\widetilde{\lambda}_{i,\delta}^{(C)} + (1-\delta)(1-\beta_{ii})\lambda_\delta^{(B)}. \end{aligned} \tag{15}$$

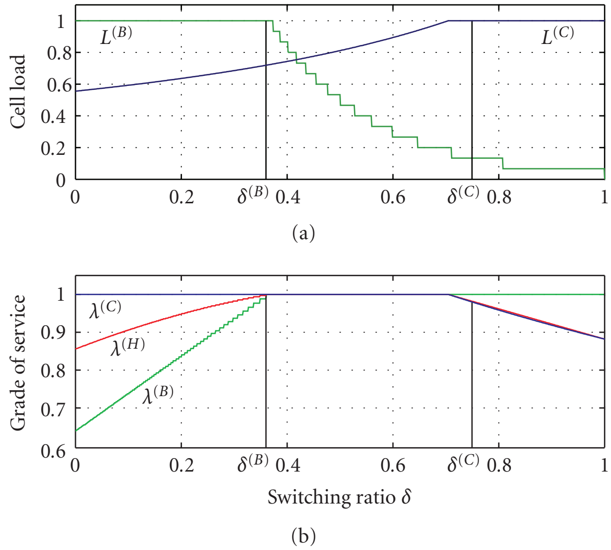
For $\delta = 1$, the total traffic has to be transmitted by the UMTS network, thus, $\lambda^{(H)} = \lambda^{(C)}$. In the case that $\delta = 0$, the UMTS network is unloaded by all switchable traffic, but this traffic is partly lost due to an overloaded broadcast network. The GoS of the broadcast network results in $\lambda^{(B)} = 1-\delta^{(B)}$. The GoS of the UMTS network can be calculated by (6) using $\widetilde{c}_{ii} = c_{ii}^{(C)}$ and $\widetilde{p}_i^{(\eta)} = p_i^{(U,\eta)}$.

4.4. Performance Analysis. A single cell is considered with fixed interference from the surrounding cells. This simplified scenario is used in order to show the basic behavior of the network performance indicators cell load ($L^{(C)}$ and $L^{(B)}$) and grade of service ($\lambda^{(C)}$, $\lambda^{(B)}$, and $\lambda^{(H)}$).

Figure 5 shows these performance indicators. In this example, the parameters of Section 4.2 and the following parameters have been used: $D^{(B)} = 2$ Mbps, $d_i = 128$ kbps, $N_{\text{item}} = 100$, $\kappa = 1$, $c_{ii} = 0.5$, and $\beta = 0.6$.

The used parameter values result in the bound values of $\delta^{(C)} = 0.78$ and $\delta^{(B)} = 0.36$. The cell load of the broadcast cell increases with a decreasing δ until the broadcast bound is reached. The nonuniform steps are caused by the different content popularity values. The UMTS cell load decreases while unloading beyond the cellular bound, caused by a lower necessary transmit power $\widetilde{p}_i$.

The grade of service is shown for both network types and the hybrid combination. For high values of δ the curves of $\lambda^{(C)}$ and $\lambda^{(H)}$ are similar due to a small contribution of the broadcast part with a small amount of traffic switched to the broadcast network. A significant gain of the GoS exists, if the initial case ($\lambda_{\delta=1}^{(C)}$) is compared to compared to the hybrid network with applied unloading ($\lambda_{\delta=\delta^{(B)}}^{(H)}$).



(a)

(b)

FIGURE 5: Performance indicators cell load and grade of service in a hybrid network for a single cell, depending on the switching ratio δ.

In the depicted case, a wide operating range is shown for achieving sufficient GoS. From a network planning point of view, applying an $\delta < \delta^{(B)}$ is not preferable, since the overall GoS $\lambda^{(H)}$ decreases very fast, due to the high influence of the overloaded broadcast network (see (15)).

5. Simulation of a Realistic Scenario

In this section, the behavior of the performance indicators GoS and the cell load is shown by means of a realistic scenario. In the shown example, it is assumed that the broadcast cell covers the entire scenario area. This can be assured by, for example, reusing a DVB-T transmitter for DVB-H, a real transmitter is available at the Berlin scenario.

As mentioned before, the snapshot approach has been used in order to achieve accurate results of the performance indicators. The snapshot method has been implement, considering indoor and outdoor users, and signal shadowing with spatial correlation. We have performed a sufficiently large number of snapshots and tested *t confidence interval* [19, Ch. 4] in order to achieve average values of the transmit power and the GoS, that are within a range of ±1% around the expected value with a 99% confidence.

5.1. Simulation Results. For the network scenario, the parameters have been used as described in Section 4.4. The cells of the UMTS network are unloaded, if content is switched from the UMTS network to the broadcast network. The grade of service λ_i is increased, and the transmit power p_i is decreased significantly.

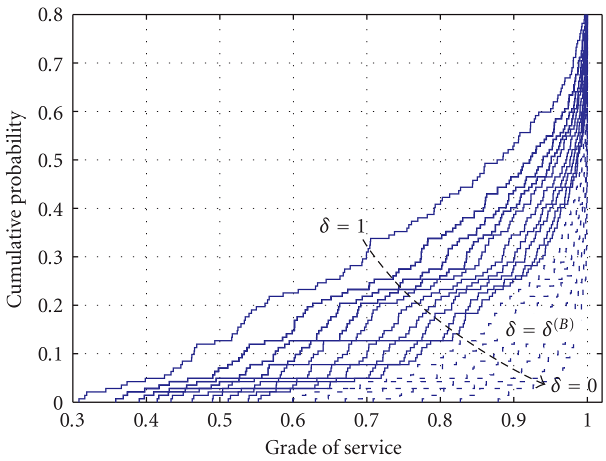


Figure 6: Empirical CDF for grade of service depending on the switching parameter δ. Dotted curves represent the cases with the broadcast network in overload.

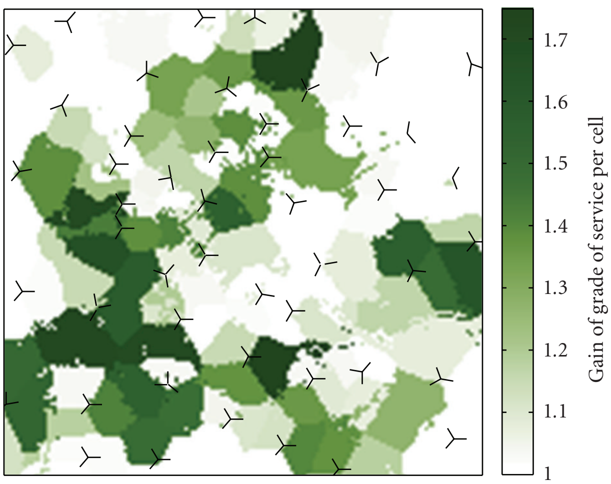


Figure 7: Gain of grade of service for unloading the UMTS cells by $\delta = \delta^{(B)}$.

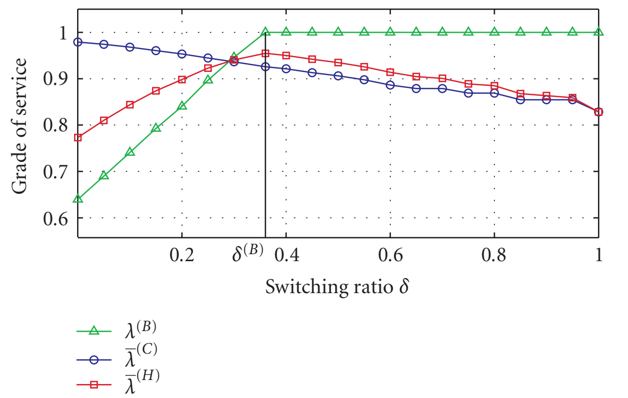


Figure 8: Grade of service of the entire network depending on the amount of load switching.

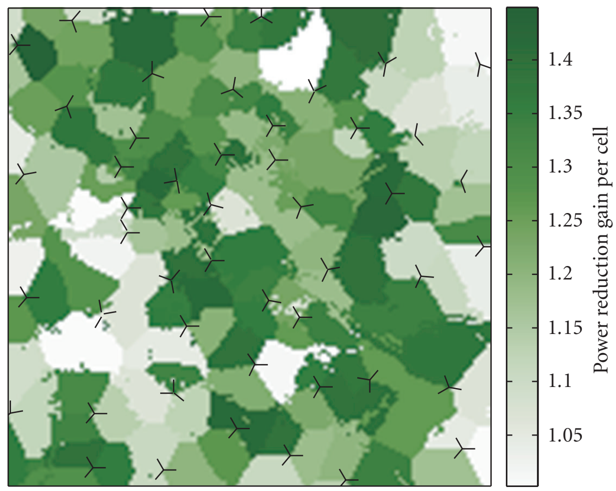


Figure 9: Power gain per cell resulting from cell unloading with $\delta = \delta^{(B)}$.

Figure 6 shows the empirical cumulative distribution of the GoS of all cells in the network. Each curve represents a switching value for δ from zero to one in 0.1 steps and the broadcast bound $\delta^{(B)} = 0.36$. It can be seen that the GoS increases for lower values of δ. For the full-loaded network ($\delta = 1$) about 31% of the cells are above the GoS target $\hat{\lambda} = 0.98$. By unloading the cells with $\delta = \delta^{(B)} = 0.36$ about 55% and for $\delta = 0$ about 89% of the cells meet the target threshold. The dotted lines represent the cases with $\delta < \delta^{(B)}$, in which the broadcast network is overloaded.

In order to estimate, which UMTS cells benefit from load switching, the GoS gain is defined by the ratio of the unloaded cells ($\lambda_{i,\delta=\delta^{(B)}}^{(C)}$) to the fully loaded case ($\lambda_{i,\delta=1}^{(C)}$). Figure 7 shows the gain values, based on the best server cell structure. A high gain occurs for those cells, in which a high traffic demand exists (compare to Figure 1). For the cells with a gain equal to one, unloading has not been necessary due to a low traffic load. A maximum gain of about 75% increase in GoS can be seen in some selected cells, which benefit most of using a hybrid mode.

We use (7) in order to calculate the grade of service of the entire network. For network planning the grade of service should be $\hat{\lambda} = 98\%$ or higher. In Figure 8 the performance of the hybrid network is shown. The GoS of the UMTS network $\overline{\lambda}^{(C)}$ increases by unloading but does not meet $\hat{\lambda}$ for an unloading till $\delta^{(B)}(\lambda_{\delta=\delta^{(B)}}^{(C)} = 92.6\%)$. The UMTS bound $\delta^{(C)}$ is not depicted since for $\delta = 0$ a GoS of $\overline{\lambda}^{(C)} = 97.9\% < \hat{\lambda}$ is achieved. Since the $\delta^{(C)} < \delta^{(B)}$ the maximum total GoS occurs for $\delta = \delta^{(B)}$ with $\overline{\lambda}^{(H)} = 95.5\%$.

If cells are unloaded and the GoS target is reached, a further unloading leads to a reduction of the necessary transmit cell power. In Figure 9, the power reduction gain is shown per cell. A maximum reduction gain of 1.45 is achieved. It can be seen that cells with high GoS gain do not show a power gain, since the GoS target has not been reached and thus the power is still at its maximum $p_{\max}$. The complementarity to the GoS can also be seen in this case.

6. Conclusion

We have proposed a system and performance model for hybrid networks, which are composed of the cellular network UMTS and the mobile broadcast network DVB-H. The developed model is based on the existing interference coupling model of the UMTS network, which has been enhanced by the load switching concept. Furthermore, a broadcast system model has been developed. The two network performance indicators cell load and grade of service (GoS) have been defined; whereas the GoS parameter is used as the primary performance criteria for hybrid networks.

Based on the developed models and the GoS definition, the switching bound concept has been been proposed, which define bounds for the load switching for each network separately. They determine thresholds for the necessary unloading of the UMTS network and for the maximum loading of the broadcast network in order to achieve the GoS target. These bounds are used to evaluate the conditions for reasonable load switching and define the necessity and feasibility of hybrid networks. The impact of load switching for the streaming multimedia service has been analyzed by a theoretical and a realistic scenario. We have shown the behavior of the cell load, the grade of service, and the appropriate gain values on a cell basis and for the entire network. The results have shown that a hybrid network can be applied to unload the UMTS network and to serve the multimedia services very efficiently.

References

[1] U. Reimers, *DVB—The Family of International Standards for Digital Video Broadcasting*, Springer, Berlin, Germany, 2nd edition, 2005.

[2] ETSI, "Digital Video Broadcasting (DVB); transmission system for handheld terminals (DVB-H)," Tech. Rep. ETSI Std. EN 302 304 V1.1.1, ETSI, Sophia Antipolis, France, 2004.

[3] G. Faria, J. A. Henriksson, E. Stare, and P. Talmola, "DVB-H: digital broadcast services to handheld devices," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 194–209, 2006.

[4] G. May, "Loss-free handover for IP datacast over DVB-H networks," in *Proceedings of the International Symposium on Consumer Electronics (ISCE '05)*, pp. 203–208, June 2005.

[5] G. May and P. Unger, "A new approach for transmitting localized content within digital single frequency broadcast networks," *IEEE Transactions on Broadcasting*, vol. 4, no. 53, pp. 732–737, 2007.

[6] ETSI, "Digital Video Broadcasting (DVB); IP datacast over DVB-H: architecture," Tech. Rep. TR 102 469 V1.1.1, ETSI, Sophia Antipolis, France, May 2006.

[7] M. Kornfeld and G. May, "DVB-H and IP datacast—broadcast to handheld devices," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 161–170, 2007.

[8] C. Heuck, *Optimierung hybrider (Rundfunk-/Mobilfunk-) Netze durch Steuerung der Lastverteilung*, Ph.D. dissertation, Institut für Nachrichtentechnik, TU Braunschweig, Braunschweig, Germany, 2009.

[9] L. Ferreira, L. M. Correia, D. Xavier, I. Vasconcelos, and E. R. Fledderus, "Final report on traffic estimation and services characterisation," Tech. Rep. IST-2000-28088 MOMENTUM, D1.4, IST, 2003, http://momentum.zib.de/paper/momentum-d14.pdf.

[10] B. Rakoczi, E. R. Fledderus, B. Heideck, P. Lourenço, and T. Kürner, "Reference scenarios," Tech. Rep. IST-2000-28088 MOMENTUM, D5.2, IST, 2003, http://momentum.zib.de/paper/momentum-d52.pdf.

[11] Momentum Project, IST-2000-28088, Momentum public UMTS planning scenarios, 2003, http://momentum.zib.de/.

[12] T. Kürner, "Propagation models for macro-cells," in *Digital Mobile Radio Towards Future Generation Systems (COST 231 Final Report)*, COST Telecom Secretariat, pp. 134–148, CEC, Brussels, Belgium, 1999, http://www.lx.it.pt/cost231/.

[13] M. Nawrocki, H. Aghvami, and M. Dohler, Eds., *Understanding UMTS Radio Network Modelling, Planning and Automated Optimisation: Theory and Practice*, John Wiley & Sons, New York, NY, USA, 2006.

[14] H.-F. Geerdes, *UMTS radio network planning: mastering cell coupling for capacity optimization*, Ph.D. dissertation, Vieweg and Teubner, Wiesbaden, Germany, 2008.

[15] A. Eisenblätter, H.-F. Geerdes, T. Koch, A. Martin, and R. Wessäly, "UMTS radio network evaluation and optimization beyond snapshots," *Mathematical Methods of Operations Research*, vol. 63, no. 1, pp. 1–29, 2006.

[16] J. Laiho, A. Wacker, and T. Novosad, *Radio Network Planning and Optimisation for UMTS*, John Wiley & Sons, New York, NY, USA, 1st edition, 2002.

[17] C. Heuck and P. Unger, "Load balancing and network planning for DVBH/UMTS hybrid networks," ITS-Europe 2005, Hanover, Germany, June 2005.

[18] J. Aaltonen, *Content distribution using wireless broadcast and multicast communication networks*, Ph.D. thesis, Tampere University of Technology, Tampere, Finland, 2003.

[19] A. M. Law and D. W. Kelton, *Simulation Modelling and Analysis*, McGraw-Hill, New York, NY, USA, 2000.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 234360, 9 pages
doi:10.1155/2009/234360

Research Article

Multiple Description Coding Using Data Hiding and Regions of Interest for Broadcasting Applications

M. Verza,[1] N. Conci,[2] G. Boato,[1] and F. G. B. De Natale[1]

[1] *Department of Information Engineering and Computer Science(DISI), University of Trento, Via Sommarive 14, 38100 Trento, Italy*
[2] *Multimedia and Vision Research Group, School of Electronic Engineering and Computer Science, Queen Mary, University of London, London E1 4NS, UK*

Correspondence should be addressed to N. Conci, nicola.conci@elec.qmul.ac.uk

Received 27 February 2009; Accepted 1 June 2009

Recommended by Maurizio Murroni

We propose an innovative scheme for multiple description coding (MDC) with regions of interest (ROI) support to be adopted in high-quality television. The scheme proposes to split the stream into two separate descriptors and to preserve the quality of the region of interest, even in case one descriptor is completely lost. The residual part of the frame (the background) is instead modeled through a checkerboard pattern, alternating the strength of the quantization. The decoder is provided with the necessary side-information to reconstruct the frame properly, namely, the ROI parameters and location, via a suitable data hiding procedure. Using data hiding, reconstruction parameters are embedded in the transform coefficients, thus allowing an improvement in PSNR of the single descriptions at the cost of a negligible overhead. To demonstrate its effectiveness, the algorithm has been implemented in two different scenarios, using the reference H.264/AVC codec and an MJPEG framework to evaluate the performance in absence of motion-compensated frames on 720p video sequences.

1. Introduction

The widespread of digital TV and high-quality broadcasting requires the definition of innovative schemes to guarantee improved robustness and error resilience to video streams. The robustness requirement is particularly necessary in the areas where the digital signal is not strong enough and frequent errors that damage slices or even entire frames are common. In fact, one of the most significant drawbacks of current technologies for terrestrial video broadcasting is that if the signal-to-noise ratio drops under a certain threshold, the program is obscured until the quality of the signal returns back to the standard operational range. This is actually one of the main concerns that arose with Digital Video Broadcasting (DVB), and it is one of the pending issues that need to be addressed. With this respect, scalable coding can be regarded as a potential solution to the problem, allowing the split of the stream into a number of substreams. The substreams can provide reduced quality at lower bit rate, and the full quality can be achieved by suitably combining all layers. In such a framework, either layered coding or MDC could be adopted as a viable approach [1].

MDC [2] has emerged as a promising approach and a valid alternative to more common layered coding standards, to improve error resilience in a video. Starting from a video source, the core idea consists of splitting the original stream in a predefined number of substreams, called descriptors. Descriptors are typically balanced, meaning that each of them has comparable bit-rates and provides similar quality when individually decoded. This results in a big advantage if compared to layered coding (from MPEG-2 to H.264/SVC), where the hierarchical organization of the stream strongly relies on the correct decoding of the base layer. In MDC, instead, each descriptor can be independently decoded at low but acceptable quality. The higher the number of received descriptors, the better the quality. If all substreams are received and properly merged, the original quality can be guaranteed.

In our work, we propose to spatially differentiate the quality of each description, defining a suitable region of interest (ROI) where the quantization strength is reduced in order to improve the perceived quality within the desired area. Since this information (location, dimension, quality factor, etc.) must be shared with the decoder to correctly display the visual content, we exploit a watermarking algorithm, which allows improving the PSNR of each single description without significantly affecting the resulting overhead. The algorithm is tested on both MJPEG and H.264/AVC. The choice of performing the validation on two different coding schemes is motivated by the recent developments in HDTV broadcasting research. In this area, bandwidth is not a critical issue anymore as it used to be, and, therefore, the transmission of static images instead of motion compensated videos could provide better perceptual quality, and higher robustness to error propagation. The validation phase has been carried out taking into account standard metrics like PSNR, together with VQM, in order to determine the perceptual impact of the ROI and the watermarking scheme.

The paper is structured as follows. Section 2 reports a review of the most relevant works in the field of MDC. In the same section, also the adoption of data hiding tools for nonconventional applications is discussed. The proposed MDC scheme is described in detail in Section 3, while ROI embedding is presented in Section 4. Experimental results are reported in Section 5, and Section 6 presents some concluding remarks.

2. Related Work

2.1. Multiple Description Coding. The generation of multiple streams imposes the introduction of a variable amount of overhead (also referred to as redundancy $\rho \in [0,1]$), which is related to the number of descriptors and to the requirements of robustness of the specific application. Given N descriptors, the overhead is calculated in general as in (1), where size_{SDC} refers to the video encoding in single descriptor mode, traditionally:

$$\rho = \frac{\sum_{n=1}^{N} \text{size}(n) - \text{size}_{\text{SDC}}}{\text{size}_{\text{SDC}}}. \quad (1)$$

The overhead is caused by the need of storing in each descriptor some auxiliary information like

(i) headers, trailers, synchronization markers, and presentation data that must be replicated in order to make each single stream visible;

(ii) side information to facilitate the reconstruction in presence of losses.

One of the first effective applications of MDC to image coding that is still a reference method was proposed by Vaishampayan in [3]. His idea of a Multiple Description Scalar Quantizer was demonstrated, for the sake of generality, with application to a memoryless Gaussian source; however, it turns out to be effective also in image and video coding applications.

In literature we find a large amount of algorithms based on MDC, and one of the expected results has always been the ability to deploy effective implementations capable of maximizing quality in the single descriptor reconstruction while minimizing overhead. This can be achieved through different techniques. The most common approaches are on the one hand related to the improvement of the single descriptor quality, by sharing relevant data among descriptors, and on the other hand focused on the maximization of the number of substreams to provide diversity. In fact, the intrinsic nature of MDC does not impose any limitation in the number of descriptors to be considered and it has been demonstrated that this goal can be achieved by applying, for example, wavelet-based approaches. There are several implementations dealing with these problems as it can be seen in [4–6] where the different strategies have been tested considering up to sixteen descriptors. The methods proposed in these articles mostly deal with static images, since wavelet-based tools for video coding have not received great attention yet. However, the implementation of a high number of descriptors introduces often a non-negligible redundancy.

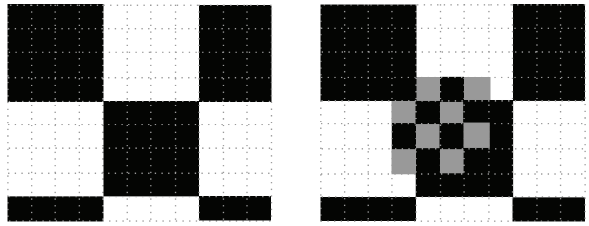


Figure 1: Macroblock grid without (left) and with the ROI (right).

If not so, the achieved quality of the single descriptor reconstruction is very poor and the streaming provider usually prefers other alternatives such as FEC as discussed in [7]. In addition, a large number of descriptors would introduce a significant increase in the complexity both at the encoder and at the decoder, making it almost impossible (with the current technologies) to perform live video transmission. Therefore, the most likely number of descriptors to deal with can be limited to a few units, according to the channel capacity and the network setup, even though most of the works in literature adopt two descriptors only.

Even though several MDC techniques have been proposed in the video framework in both the form of MJPEG contents and motion-compensated coding standards, still little research has been conducted on the combination of MDC and H.264. In [8] the authors focus on the implementation of a coding scheme to be implemented in a MIMO architecture. In [9] an innovative standard compliant approach is presented: descriptors are composed by alternating primary and redundant slices in a complementary manner, so that losses can be mitigated by replacing each missing portion of data with its redundant slice from the complementary descriptor. More recently, another MDC system was proposed in [10]. Here, the smoothness and edge features of DCT blocks are modeled in such a way that their perceptual tolerance against visual distortion is measured; the key components identified via this method are duplicated, while the remaining ones are effectively split among descriptors.

2.2. Data Hiding. As far as data hiding is concerned, we have applied the concept in a slightly different manner with

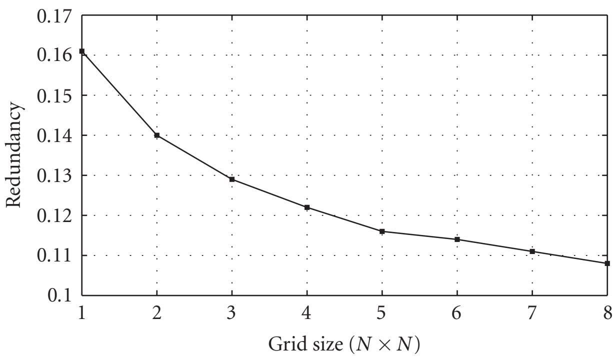


FIGURE 2: Resulting overhead by varying the gridsize in the H.264/AVC configuration.

respect to the standard applications. Data hiding schemes are usually applied to digital rights management [11]. Such methods invisibly embed a watermark into a cover data (a portion of the image in which the mark is inserted), allowing the identification of copyright violation, presence of illegal copies or illegal distributors, by recovering the embedded information. In particular, classical watermarking applications consist of copy control, broadcast monitoring, fingerprinting, authentication, copyright protection, and access control [12]. Since early 1990s, digital watermarking plays a key role in security of multimedia communication and digital right management. In the last few years, the watermarking schemes have evolved significantly and the classical Spread Spectrum techniques have been outperformed by the so-called informed watermarking, the Quantization Index Modulation (QIM) algorithms [13, 14].

As far as video watermarking techniques are concerned, H.264 requires the definition of new appropriate tools for both copyright protection and authentication. Few works have addressed this problem so far. A hybrid approach is presented in [15] that faces both features by embedding a robust mark in the transform coefficients and a fragile mark in the motion vectors. Authentication is instead achieved in [16] via a fragile watermark insertion using skipped macroblocks of H.264 compressed sequences. Recently, the work in [17] proposed an H.264 video watermarking method which exploits a perceptual model to select the coefficients for the mark insertion. The goal consists, as usual, in increasing the size of the payload, guaranteeing a reduced impact on the visual distortion, and therefore focusing on the standard problem of robustness to common signal processing attacks. The same issue is faced in [18].

In the last few years, data hiding has received considerable attention also in different research areas and has been adopted also to nonsecurity oriented applications. As far as error concealment is concerned, there are some works exploiting watermarking, where embedded information is extracted to detect and conceal errors. In [19] important data for each macroblock (MB) are extracted and inserted into the next frame with suitable MB-interleaving slice-based data hiding techniques for I and P frames, while in the decoding phase, they can be exploited to conceal the corrupted MBs. Hidden information about the original (high-quality) data can be also used to estimate the quality degradation of the received data. Indeed, the extraction of the embedded information can support the reduced-reference methods for quality assessment [20]. Recently, a simple data hiding technique has been adopted also in video surveillance applications to enhance the visual quality of faces [21].

A first proof of concept that demonstrates the suitability of data hiding tools also in the MDC has been proposed in our previous work [22]. Here we underline the advantages of data hiding in improving MDC schemes while reducing the overhead. The approach embeds the DC coefficient of each block of the JPEG frame into the AC coefficients, therefore, avoiding its transmission.

3. The Proposed MDC Coding Scheme

According to the typical constraints of MDC, each descriptor should be

(i) independent,

(ii) nonhierarchical,

(iii) complementary,

(iv) balanced.

Following these principles, we propose to modify the quality of each MB on the basis of a predefined checkerboard pattern of $N \times M$ MBs, in which high- and low-quality MBs are alternated. In case only one descriptor out of two is correctly received and decoded, the achievable perceptual quality is reduced, due to the checkerboard pattern. In Figure 1 (left) an example of a 4×4 checkerboard pattern is shown. White MBs are coarsely quantized being the least expensive in terms of bit allocation, while a finer quantization is applied to black MBs.

Being the amount of overhead and, therefore, the final bitrate strongly related to the checkerboard structure and the quantization factors of the MBs, we show in Figure 2 the relationship between the redundancy and the size of the grid when adopting H.264/AVC as the coding algorithm. In H.264/AVC, the quality factor is given by the quantization parameter (QP), which values span in the range $[1, 51]$, being 1 the best achievable quality. In the presented configuration, we have chosen $QP_H = 20$ for high-quality macroblocks, while the coarsely quantized MBs have $QP_L = 40$. The same evaluation could be carried out using the MJPEG encoder. However, since only H.264 has predictive features also in the intra frames, the MJPEG curve would be almost flat, since the number of MBs with high and low quality is constant regardless of the grid size and in particular, each block is independently coded and has no relationship with the adjacent ones. For the MJPEG configuration, we have instead set the quality factor (QF) to $QF_H = 92$ and $QF_L = 22$.

The choice of a simple checkerboard pattern and QP (QF) alternation would be reasonable in presence of limited packet losses, and it may generate annoying artefacts in presence of significant losses. However, according to the human visual system, we can say that users typically focus

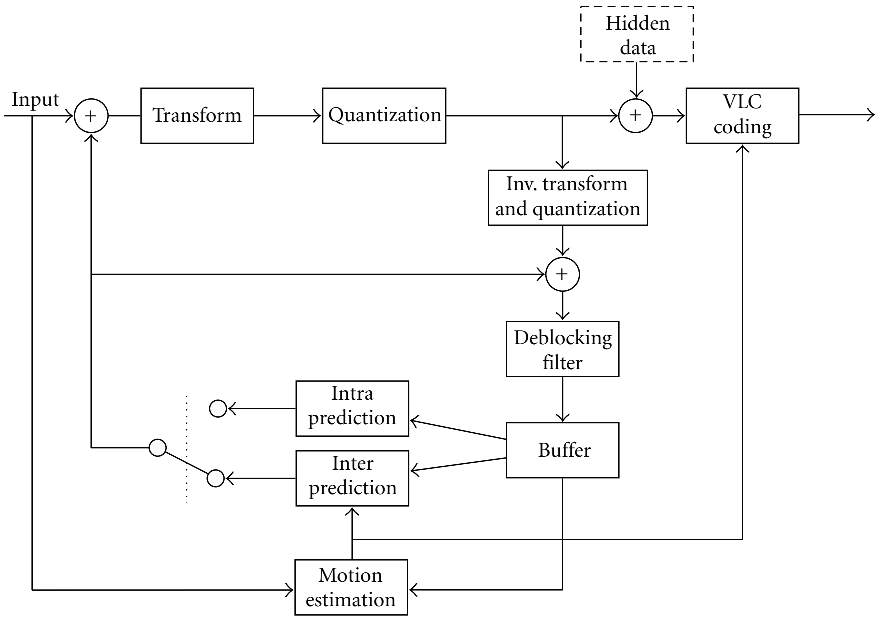


FIGURE 3: The proposed ROI data hiding scheme.

TABLE 1: Set of parameters for ROI coding.

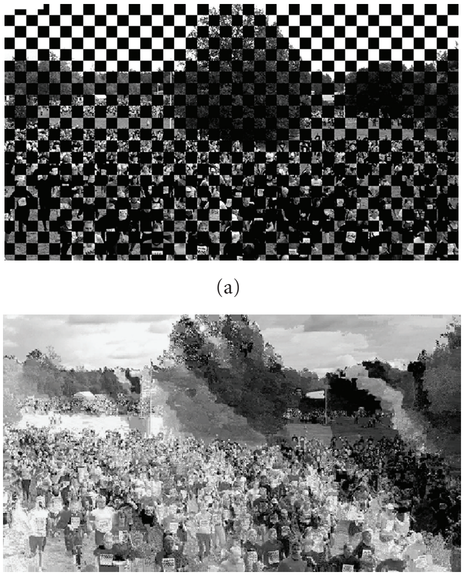
Param	Description	Range	n.bits
$select_D$	Identifies D1 or D2	BOOL	1
N	MBs in the horizontal grid	[1–8]	3
M	MBs in the vertical grid	[1–8]	3
ROI	Activate ROI	BOOL	1
QP_L or QF_L	Set low quality quantizer	[1–max]	6
X_{st}	Start of ROI in X (%)	[1–100]	7
Y_{st}	Start of ROI in Y (%)	[1–100]	7
X_{end}	End of ROI in X (%)	[1–100]	7
Y_{end}	End of ROI in Y (%)	[1–100]	7
ΔQP (ΔQF)	$QP_R - QP_L$ ($QF_R - QF_L$)	[0–max]	6
	Total required bits		48

the attention on a specific area of the image when watching a video [23]; therefore, we have introduced a mechanism to provide better perceptual quality of the single descriptions reconstruction in a ROI, by (i) reducing the gap between the two quantization levels in the ROI and (ii) decreasing the size of the grid in that area to the minimum size of 1×1 MBs. Figure 1 (right) illustrates an example where the grid of the background is set to 4×4 MBs and the ROI grid is set to 1×1 MBs. A new parameter QP_R ($QP_L < QP_R < QP_H$) for the H.264/AVC configuration and QF_R ($QF_L < QF_R < QF_H$) for the MJPEG configuration is introduced to replace QP_L and QF_L within the ROI and it is represented by the gray MBs in the figure. This parameter allows reducing the gap between the two category of blocks, smoothing the quality differences. In order to make the scheme more flexible, we keep all parameters (Table 1) configurable every intra-frame, so to adapt the ROI to the visual content. Furthermore, the size of the checkerboard can also be chosen a priori and changed every GOP (in H.264) or every frame (in MJPEG), making it possible to adjust the quality and the corresponding overhead to the video source.

In order to allow the correct decoding of the video sequence, the ROI parameters necessarily require to be transmitted to the receiver/decoder. This is usually achieved by either delivering the data out-of-band or by introducing new syntax elements in the stream. In our approach we have preferred to exploit data hiding techniques due to (i) the negligible impact on the resulting bitrate and (ii) the intrinsic nature of data hiding in terms of data protection, allowing only authorized users to recover the embedded information. The procedure of data embedding and recovery is explicitly described in the next section.

4. ROI Parameters Embedding

The choice of a suitable data hiding scheme requires the estimation of the payload to be transmitted together with the video. For the sake of demonstration, we have roughly calculated a reasonable number of bits that need to be embedded as watermark. According to our design we have considered that 48 bits are sufficient to encode the ROI information, including the size of the grid, the quantization parameters for the background and the ROI, and its coordinates in the image plane (see Table 1). In order to allow a one-to-one comparison between the two coding schemes, ΔQP and ΔQF values are represented with 6 bits, meaning that in the JPEG case, only every second value can be chosen, since $1 < QF < 100$, and this would require instead 7 bits. This approximation is reasonable and does not affect the achieved results significantly, since the visual appearance variation is negligible.



(a)

(b)

Figure 4: Decoding without knowing the watermark extraction key on the *crowd* sequence in case of MJPEG (a) and H.264 (b).

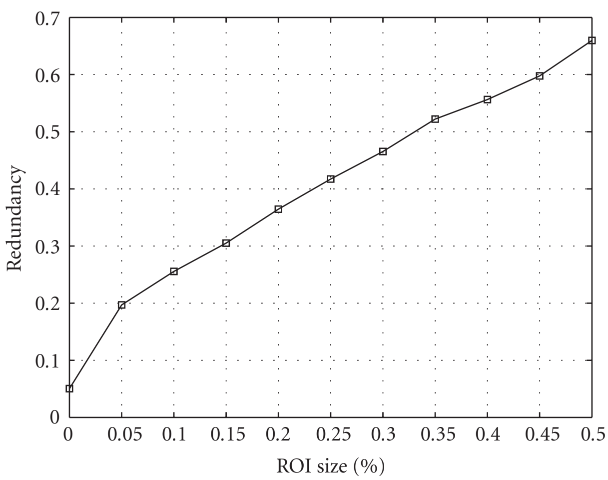


Figure 5: Relationship between ROI size and resulting overhead.

In order to limit the additional computational burden of the algorithm, the complexity of the adopted data hiding technique must be kept as low as possible. Therefore, we have chosen a simple yet flexible watermarking scheme in the class of quantization index modulation (QIM) techniques [13]. The adopted embedding technique is the Binary Scalar Costa Scheme (BSCS) [24], and the formula is as follows:

$$q_n = Q_\Delta\left\{x_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right\} - \left(x_n - \Delta\left(\frac{d_n}{D} + k_n\right)\right), \qquad (2)$$

where d_n is the watermark symbol, x_n is the coefficient to be marked, D is the cardinality of the alphabet of the watermark (2 in our case), k_n is the watermarking key, and $Q_\Delta()$ represents scalar uniform quantization with step Δ. We selected $\Delta = 2$, $k_n = 1$ to minimize the impact of the inserted information thus leading to perfect mark invisibility and very low complexity. Since the embedding of the ROI parameters has a negligible impact on the redundancy, it provides a powerful tool to improve the perceptual quality of single description, as experimentally verified in Section 5.

In our embedding scheme, we have decided to replicate the mark in three different locations to improve the robustness in case a portion of the image is damaged and we have chosen the first seven MBs of the first, middle, and last row as a cover for the watermark insertion. In each MB, we embed 8 payload bits by modifying the first two AC coefficients of each 8×8 block composing the MB after quantization. At the decoder side, the knowledge of the watermark extraction key (i.e., its location) allows the extraction of the embedded mark (i.e., ROI parameters). Notice that the first parameters of Table 1 (select_D, N, M, ROI, and QP_H—or QF_H) are fundamental to correctly start the decoding process. Therefore, to ensure their reconstruction, we set QP (QF) of MB_0 and MB_1 to $QP_H(QF_H)$ in both descriptions D_1 and D_2. The proposed data hiding exploitation adds a significant level of protection to the MDC scheme. As a matter of fact, it is impossible to reconstruct an acceptable version of the video without knowing the watermark extraction key. If the embedded ROI parameters cannot be extracted (in case for example of nonauthorized users) neither descriptor can be correctly decoded. Details about this concept are provided in Section 5. The overall encoding scheme for H.264/AVC is shown in Figure 3. As it can be noticed, the proposed architecture does not alter the coding process, returning a completely standard-compliant stream, resulting in another advantage of this architecture with respect to other solutions. Indeed, the information related to the ROI is included in the transform coefficients and no additional syntax elements need to be defined. The corresponding scheme for the MJPEG implementation is similar to the one in Figure 3, but clearly without the motion compensation loop.

5. Experimental Results

This section refers to the validation of the system using three different test sequences retrieved from the European Broadcasting Union (EBU) website [25] in the 720 p format, namely, 1280×720. We will first review the results obtained in the MJPEG framework and successively using H.264. In both cases, the encoders have been configured in order to achieve a PSNR higher than 40 dB on the luminance component in case both descriptors have been correctly received.

As a common achievement in both implementations, we demonstrate that the data hiding procedure serves also as a protection tool, which does not allow enjoying the visual content properly; if the user does not know how to interpret the embedded key, the exact quantization values are known only for the black MBs in Figure 1, while QP_L/QF_L are hidden in the cover. The effects on the resulting image depend on the adopted codec, but in general the visual content is not preserved. An example is presented in Figure 4

(a)

(b)

Figure 6: Full reconstruction (a) and single descriptor reconstruction (b) for *Crowd Run* sequence using MJPEG.

(a) (b) (c)

Figure 7: ROI disabled (a), ROI enabled with $\Delta QF = 35$ (b), and ROI enabled with $\Delta QF = 70$ (c) for *Crowd Run* sequence using MJPEG.

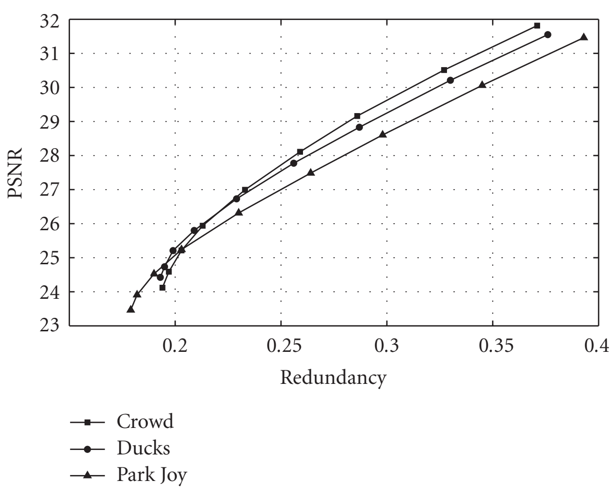


Figure 8: Redundancy versus PSNR in case of one descriptor is completely lost for different values of QP_H and for the *Park Joy*, *Ducks take off*, and *Crowd Run* video sequences.

Table 2: Location of ROIs in the test sequences.

Test sequence	X_{start}	Y_{start}	X_{end}	Y_{end}
Park Joy	40	40	80	70
Crowd Run	10	60	50	90
Ducks take off	30	40	70	70

Table 3: Overhead and average PSNR using MJPEG on the *Crowd Run* sequence.

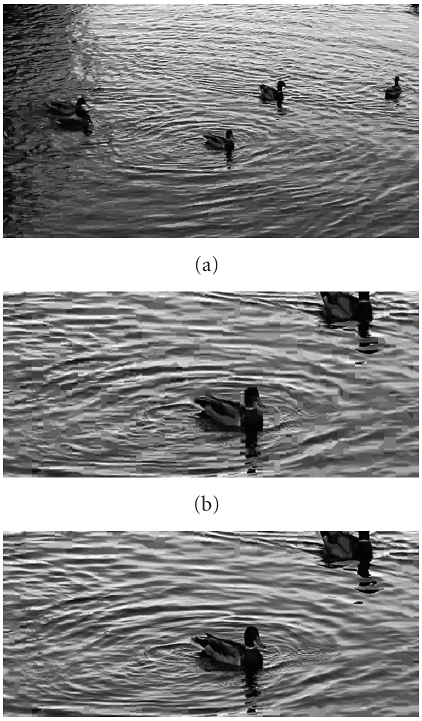
ΔQF	ρ	PSNR (dB)
0 (no ROI)	0.23	29.8
35	0.25	30.69
70 (high quality ROI)	0.34	30.91

without the ROI, for MJPEG and H.264, respectively. The MJPEG implementation shows an evident grid. In fact, during the process of inverse quantization, values are not reported at the original magnitude, and the details of the picture are lost. In case of H.264, instead, the content appears to be more *scrambled*. This is due to the intra prediction algorithm of the codec that tends to compensate possible errors that may occur in intrablocks as well. For the tests, we have set a grid of 2×2 MBs, in order to visually reduce the blocking artifacts due to the high- and low-quality block alternation. In the ROI, the grid is set to 1×1 MB. The location of the ROI for each test sequence is set in terms of percentage as shown in Table 2.

It is clear that the ROI size has an impact on the total overhead. In our system, the increase of the redundancy is proportional to the ROI size, as demonstrated in Figure 5. Here we have chosen the central portion of the *Ducks take off* video sequence and we have progressively increased the area occupied by the ROI at regular intervals.

5.1. MJPEG Experiments. This first set of results aims at demonstrating the viability of the proposed approach in absence of motion compensation videos. It is worth noting that although we have chosen a simple MJPEG codec, there is no practical constraint in terms of compressing standard, since the data hiding procedure can be applied to any transform-based coding scheme. Focusing on the target quality of 40 dB in case of full reconstruction, we have set $QF_H = 92$ and $QF_L = 22$ in order to achieve a reasonable compromise between perceived quality and overhead. This configuration returns an average PSNR of the video sequence around 30 dB. In Figure 6, a comparison between the full reconstruction and the single descriptor decoding is shown for the *Crowd Run* sequence. In Figure 7 a detail of the picture is shown, in which the improvements given by the ROI are highlighted at different values for ΔQF. The corresponding numerical values are reported in Table 3.

5.2. H.264 Experiments. Also in this case, the encoder has been configured in order to ensure a PSNR in the range of 40dB in case both descriptors are received. In our experiments we have set $QP_H = 20$ and $QP_L = 40$. All



(a)

(b)

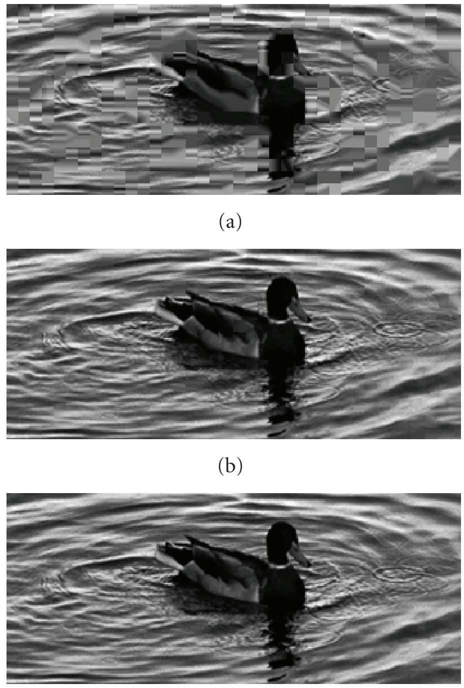
(c)

Figure 9: *Ducks take off* sequence: entire frame and ROI disabled (a); detail without (b), and with ROI (c).

(a)

(b)

Figure 10: *Crowd Run* sequence detail with disabled (a) and enabled (b) ROI.



(a)

(b)

(c)

Figure 11: Detail of the ROI area for the *Ducks take off* sequence with different ΔQP: 0 (a), 15 (b), and 30 (c).

the tests presented in this section have been performed using the JM reference software (v12.4) [26]. In Figure 8, the relationship between redundancy and distortion for different values of QP_L is reported, which demonstrates the versatility of the proposed method in modulating the overhead. We can notice that, on one hand, very low redundancy values can be achieved at the cost of low single description quality, while, on the other hand, it is possible to obtain higher quality reconstruction by increasing the bitrate. In any case, even increasing QP_L significantly, the overhead is still constrained below 40% with an average distortion of the whole frame around 30 dB.

Figure 9(a) shows a single description reconstruction ($QP_L = 50$) with disabled ROI, with Figure 9(b) being its detail, while in Figure 9(c), the ROI is inserted. The same experiment is presented in Figure 10 for the *Crowd Run* sequence. In Figure 11, we illustrate the details of the ROI by applying different values of ΔQP. As it can be noticed from the picture, the quality of the ROI can be significantly improved by filling the gap in terms of quantization parameter.

Figures 12 and 13 highlight the performance improvement due to the ROI introduction: although this results in an unavoidable bitrate increase, the PSNR in the ROI can be significantly improved by varying the QP_R, letting the user perceive a better quality in the significant area of the video. Both Figures 12 and 13 refer to a configuration in which QP_H is set to 20, QP_L is set to 50, and QP_R is between 45 and 25.

To assess the perceived quality improvement introduced by the ROI we employed the NTIA General Video Quality Metric (VQM) [27], which returns values in the 0–1 range, with 1 representing worst quality. This metric has been proved to be highly correlated to subjective ratings of processed HD video sequences [28]. The results are reported in Table 4.

As a final comment on the achieved results, we underline that the proposed method embeds the ROI data within a reduced number of MBs, altering a maximum of only

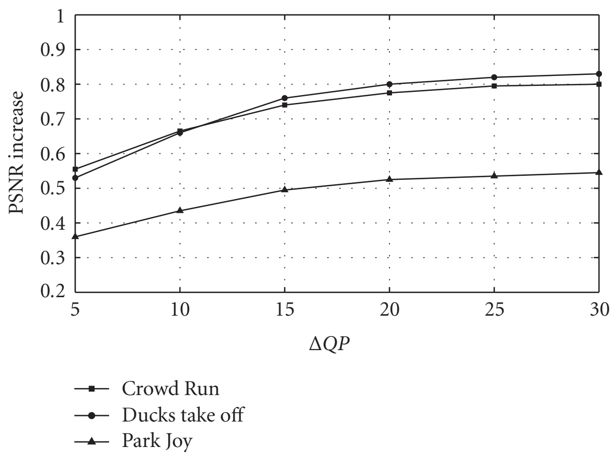


Figure 12: Quality improvement due to ROI introduction for the *Park Joy*, *Ducks take off,* and *Crowd Run* video sequences.

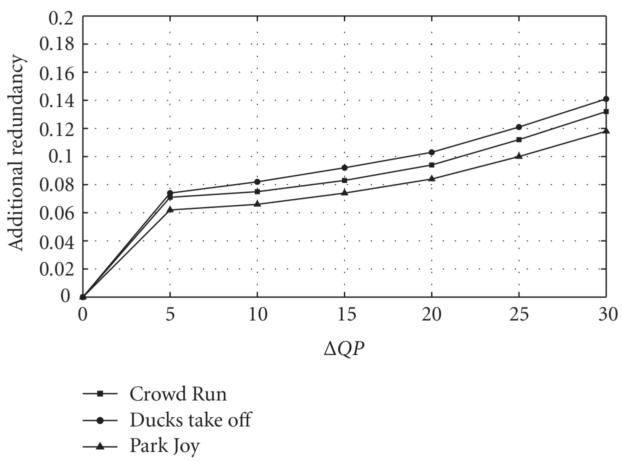


Figure 13: Overhead increase due to ROI introduction for the *Park Joy*, *Ducks take off,* and *Crowd Run* video sequences.

Table 4: Comparison of VQM results between enabled and disabled ROI for the *Park Joy*, *Ducks take off,* and *Crowd Run* video sequences.

Sequence	VQM (ROI Off)	VQM (ROI On)
Park Joy	0.5955	0.5689
Crowd Run	0.6011	0.5773
Ducks take off	0.7061	0.6848

two transform coefficients per block. Therefore, the quality degradation due to data hiding is completely negligible. Simulation results refer that the average loss in a single watermarked macroblock can be assessed around 0.45 dB in both MJPEG and H.264/AVC, therefore, minimally affecting the visual appearance of the image. We have replicated the watermark three times in different areas of the frame, thus reducing the probability of ROI details being lost due to channel errors. Even in this case, the impact on stream quality and bitrate is minimal: our experiments refer negligible additional distortion (<0.005 dB) and increased redundancy (0.004%) when applying three fully redundant watermarks on the *Ducks take off* sequence.

6. Conclusions

In this paper, we presented a novel MDC technique for HDTV video sequences, using MJPEG and H.264/AVC. The proposed method exploits data hiding to improve the perceptual quality of each single description within specific regions of interest. ROIs are defined a priori and coded with better quantization factors, and the corresponding information required at the decoder side is transmitted via a suitable watermarking algorithm. We reported extensive experimental results demonstrating the effectiveness of the proposed method, which allow increasing the PSNR of single descriptions at the cost of a slightly higher redundancy.

Acknowledgment

This work is partially funded by the Province of Trento (Italy) in the framework of the TRITON Project.

References

[1] Y.-C. Lee, J. Kim, Y. Altunbasak, and R. M. Mersereau, "Layered coded vs. multiple description coded video over error-prone networks," *Signal Processing: Image Communication*, vol. 18, no. 5, pp. 337–356, 2003.

[2] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, 2005.

[3] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 821–834, 1993.

[4] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Proceedings of the Visual Communications and Image Processing*, vol. 3653 of *Proceedings of SPIE*, pp. 998–1008, San Jose, Calif, USA, January 1999.

[5] A. C. Miguel, A. E. Mohr, and E. A. Riskin, "SPIHT for generalized multiple description coding," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 842–846, Kobe, Japan, October 1999.

[6] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 813–826, 2000.

[7] C. Soldani, G. Leduc, F. Verdicchio, and A. Munteanu, "Multiple description coding versus transport layer FEC for resilient video transmission," in *Proceedings of the International Conference on Digital Telecommunications (ICDT '06)*, 2006.

[8] M. Tesanovic, D. R. Bull, A. Doufexi, and A. R. Nix, "H.264-based multiple description coding for robust video transmission over MIMO systems," *Electronics Letters*, vol. 42, no. 18, pp. 1028–1029, 2006.

[9] T. Tillo, M. Grangetto, and G. Olmo, "Redundant slice optimal allocation for H.264 multiple description coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 59–70, 2008.

[10] S. Moradi, S. Gazor, and T. Linder, "A multiple description video coding motivated by human visual perception," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 689–692, Las Vegas, Nev, USA, March 2008.

[11] E. T. Lin, A. M. Eskicioglu, R. L. Lagendijk, and E. J. Delp, "Advances in digital video content protection," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 171–182, 2005.

[12] G. Doerr and J.-L. Dugelay, "A guide tour of video watermarking," *Signal Processing: Image Communication*, vol. 18, no. 4, pp. 263–282, 2003.

[13] B. Chen and G. W. Wornell, "Quantization index modulation methods for digital watermarking and information embedding of multimedia," *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 27, no. 1-2, pp. 7–33, 2001.

[14] L. Pérez-Freire, F. Pérez-González, and S. Voloshynovskiy, "Revealing the true achievable rates of Scalar Costa scheme," in *Proceedings of the 6th IEEE Workshop on Multimedia Signal Processing*, pp. 203–206, September-October 2004.

[15] J. Zhang, A. T. S. Ho, G. Qiu, and P. Marziliano, "Robust video watermarking of H.264/AVC," *IEEE Transactions on Circuits and Systems II*, vol. 54, no. 2, pp. 205–209, 2007.

[16] D. Pröfrock, H. Richter, M. Schlauweg, and E. Müller, "H.264/AVC video authentication using slapped macroblocks for an erasable watermark," in *Visual Communications and Image Processing*, vol. 5960 of *Proceedings of SPIE*, pp. 1480–1489, Beijing, China, July 2005.

[17] M. Noorkami and R. M. Mersereau, "A framework for robust watermarking of H.264-encoded video with controllable detection performance," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 14–23, 2007.

[18] D. Profrock, M. Schlauweg, and E. Müller, "A new uncompressed-domain video watermarking approach robust to H.264/AVC compression," in *Proceedings of the 3rd IASTED International Conference on Signal Processing, Pattern Recognition, and Applications*, pp. 99–104, 2006.

[19] L.-W. Kang and J.-J. Leou, "An error resilient coding scheme for H.264/AVC video transmission based on data embedding," *Journal of Visual Communication and Image Representation*, vol. 16, no. 1, pp. 93–114, 2005.

[20] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1680–1689, 2006.

[21] P. Meuel, M. Chaumont, and W. Puech, "Data hiding in H.264 video for lossless reconstruction of region of interest," in *Proceedings of European Signal Processing Conference*, 2007.

[22] N. Conci, M. Carli, and G. Boato, "Data hiding tools for multiple description video coding," *Electronics Letters*, vol. 45, no. 2, pp. 98–99, 2009.

[23] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, 2004.

[24] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.

[25] European Broadcasting Union, http://www.ebu.ch.

[26] I. 14496-10 and I. R. H.264, Advanced Video Coding 2003.

[27] VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II," August 2003.

[28] S. Wolf and M. H. Pinson, "Application of the NTIA general video quality metric (VQM) to HDTV quality monitoring," in *Proceedings of the 3rd International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM '07)*, Scottsdale, Ariz, USA, January 2007.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 496934, 12 pages
doi:10.1155/2009/496934

Research Article

Slow Motion and Zoom in HD Digital Videos Using Fractals

Maurizio Murroni, Cristian Perra, and Daniele D. Giusto

DIEE, University of Cagliari, Piazza D'Armi, 09123 Cagliari, Italy

Correspondence should be addressed to Cristian Perra, cperra@diee.unica.it

Received 1 March 2009; Accepted 19 October 2009

Recommended by Sandro Scalise

Slow motion replay and spatial zooming are special effects used in digital video rendering. At present, most techniques to perform digital spatial zoom and slow motion are based on interpolation for both enlarging the size of the original pictures and generating additional intermediate frames. Mainly, interpolation is done either by linear or cubic spline functions or by motion estimation/compensation which both can be applied pixel by pixel, or by partitioning frames into blocks. Purpose of this paper is to present an alternative technique combining fractals theory and wavelet decomposition to achieve spatial zoom and slow motion replay of HD digital color video sequences. Fast scene change detection, active scene detection, wavelet subband analysis, and color fractal coding based on Earth Mover's Distance (EMD) measure are used to reduce computational load and to improve visual quality. Experiments show that the proposed scheme achieves better results in terms of overall visual quality compared to the state-of-the-art techniques.

1. Introduction

Today's TV broadcasting industry is rapidly facing new challenges to chase the technological progress if compared to the previous fifty years of its existence. The migration from analog to digital systems which has begun in the early days of the last decade with the satellite broadcasting is almost completed also for its terrestrial counterpart. Furthermore, the DVB family of standards has recently extended its arena with the release, alongside traditional DVB-S, DVB-T, and DVB-C, of the new DVB-H and DVB-SH to cope with mobile applications for handheld terminals, while several new means to service delivering are arising beside traditional terrestrial and satellite systems, such as TV video streaming over IP (IPTV) based either on XDSL or in the forthcoming near future on WiMAX access. A common characteristic of this new access technologies is the ability to provide broadband services allowing High Digital TV (HDTV) to become now a reality. Furthermore, new generation set-top boxes provided with the multiple access feature are able to decode heterogeneous TV input signals (e.g., DVB-T, DVB-S, IPTV). Within this framework, regardless the broadband access technology deployed, more and more new common features and services are been developed to enlarge the quality of the video at user side. Video rendering refers to all the techniques able to add flexibility to the end user by modifying somehow the view of a video sequence. With the birth of the new LCD or Plasma screen fully supporting the Full HD technology special effectS like image zoom and resize as well as slow motion are likely to be integrated in the new generation Full HDTV set-top boxes.

Slow motion replay is another special effect used in video rendering. It consists of a presentation of video scenes at rates lower than originals. Already consolidated as a commercial feature for analog video players, today slow motion is ready to be extended to the digital formats. Within an analog framework, given a video sequence at a certain frame rate, the classical slow motion effect is obtained, at the display, by reducing the frame rate to a certain amount, so that a frame is frozen and remains visible for a time proportional to the slow motion factor. On the other hand, analog slow motion is realized in preproduction by means of fast-shuttered cameras able to capture the scene at a frame rate higher than the standard rate (i.e., 25 frame/sec for PAL/SECAM, 30 frame/sec for NTSC). Slow motion is achieved by filming at a speed faster than the standard rate and then projecting the film at the standard speed. In this case, the slow motion factor achievable is limited to shutter speed and is fixed at the preproduction stage.

In a digital environment, these limits for fast-shuttered cameras can be overcome through processing techniques.

At present, commercial digital video players allow users to browse a video sequence frame by frame, or by chapter selection with prefixed indexes. Slow motion replay is achieved by reducing the frame rate display or keeping it constant [1, 2] and inserting within the sequence additional intermediate frames generated by interpolation. Interpolation can be applied at either at pixel or grouping pixels into blocks. Data replication, linear or cubic spline can be used at first sight. A major drawback of these approaches is that yield to a marked degradation of the video quality which can be noticed in terms of "fading" effect (spline) and "jerky" distortion (data replication), both resulting in low motion quality for the human visual system.

Similar issues arise in image plane if interpolation is used to perform spatial zoom. Block distortion and/or blurring effect can be experienced in the enlarged frames.

In the recent years, motion compensated frame interpolation (MCFI) techniques were proposed to improve performance in case of slow motion. Although these techniques were formerly used to convert frame rate between PAL, NTSC, and HDTV, MCFI methods are also used in video streaming and conferencing applications [3–5]. MCFI idea is to implement motion estimation on the previous and current frame, and then generate the corresponding interpolated frame by averaging pixels in the previous and current frame that are pointed by the half of motion vectors. Motion estimation can be achieved by block-based or pixelwise methods. In general, pixelwise motion estimation can attain more accurate motion fields, but needs a huge amount of computations. Thus, it is often used in off-line MCFI rather than real-time processing. In contrast, block matching algorithms (BMAs) can be efficiently implemented and provide good performance (most MCFI methods are based on BMA). A comparison between pixelwise and block-based motion estimation for MCFI is discussed in [6].

In [7], the joint use of fractal coding and wavelet subband analysis to obtain spatial zoom and slow motion replay of luminance video sequences is proposed to avoid the above mentioned distortion effects.

In digital pictures, fractals are mainly used to achieve data compression by exploiting self-similarity of natural images [8, 9], but their potentials are not limited to compression. The properties of fractal coding allow expanding a multidimensional signal (e.g., image and video) along any of its dimensions. A major weakness of fractal representation is the high computational complexity in searching similarities among blocks using affine transformations; therefore, a "best match" algorithm is very time-consuming for multidimensional data sets.

Several methods have been proposed to speed up fractal coding [10]. A class of proposed solutions is based on wavelet subband analysis [11]. Due to their orthogonal and localization properties, wavelets are well suited (and extensively adopted) for subband data analysis and processing.

Our algorithm exploits these features by performing the fractal coding of each subband with particular attention to the frequency distribution of the coefficients. To further reduce the high computational cost of fractal encoding, we use active scene detection so as to perform fractal coding in high information (moving) areas only. Furthermore to improve overall visual quality, overlapped block coding and postfiltering are used, as suggested in [12] but extended to the three-dimensional case. Experimental results presented in the following show that our approach achieves higher subjective and objective quality, with respect to state-of-the-art techniques.

Conventional fractal coding schemes can easily be extended to color image (video) as represented in multichannels such as Red, Green, and Blue (RGB) components. Thus each channel in color image can be compressed as a grey-level image. Hurtgen, Mols, and Simon proposed a fractal transform coding of color images in [13]. To exploit the spectral redundancy in RGB components, the root mean square (RMS) error measure in gray-scale space is extended to 3-dimensional color space for fractal-based color image coding [14]. Experimental results show that a 1.5 compression ratio improvement can be obtained using vector distortion measure in fractal coding with fixed image partition as compared to separate fractal coding in RGB images. However, RGB space is not perceptually uniform. A system is not uniform if a little perturbation of a value is perceived linearly along the possible variation of that value. This means that a color space is perceptually uniform if a distance from a color a and another color $b = a + \Delta c$ will be perceived as constant independently from a or b. Using a nonperceptually uniform space as RGB has the drawback that the Human Vision System (HVS) will be affected by computer measures for digital video processing, since the distance from RGB value will not be uniform in respect of the HVS. Starting from these considerations, the Commission Internationale d'Eclairage (CIE) defined a uniform color model, called L*a*b* that represents all the color humans being able to resolve. Danciu and Hart [15] presented a comparative study of fractal color image compression in the L*a*b* color space with that of Jacquin's iterated transform technique for 3-dimensional color. It has been shown that the use of uniform color space has yielded compressed images less noticeable color distortion than other methods. In this paper we will propose a novel approach for coding color images based on the joint use of the L*a*b* color space and Earth Mover's Distance (EMD) measure [16]. EMD has been suitably deployed for color image retrieval applications [17]. It is a vector metric that combines spatial and color information to resolve similarities among color images. In this work we implement a fractal coding approach that relies on EMD for finding self-similarities within color images represented in the L*a*b* color space. The proposed approach has achieved better results in terms of objective quality assessment compared a classic codec based on RMS measure.

The present paper is the natural extension of [7] to the case of HD color video sequences and it copes with the further issue distinctive of the fractal coding of color video scenes. The EMD measure is used during the coding. This measure has proven to be suitable for detecting similarities between color multimedia contents [18]. To reduce the high computational cost of fractal coding, an active scene detector is used, so as to perform full three-dimensional coding only in high information areas (moving areas), whereas static zones are coded using a two-dimensional coder. To further speed up the coding process a wavelet subband analysis is performed whereas postprocessing techniques are used to improve visual quality. In addition, a fast scene change detection algorithm [17] is exploited for determining the optimal temporal window for maximizing the video quality of the zoom and slow motion processing.

The paper is organized as follows. In Section 2, a description of fractal theory applied to color video processing is given. Section 3 details the proposed method. Experimental results are provided in Section 4. Conclusions are finally drawn in Section 5.

2. Fractal Theory Applied to Color Video Processing

The fractal representation of natural signals is possible irrespective of the signal dimensions and can be used in applications concerning voice/sound, images, video sequences, and so on. Fractal representation/coding is based on the Iterated Function System (IFS). The basic idea of IFS approach is to exploit the redundancy given by the self-similarity always contained in natural signals. For instance, a "fractal image" can be seen as a collage composed by copies of parts of an original image that have been transformed through opportune geometric and "massive" transformations (i.e., luminance or contrast shift).

The mathematical foundation of this technique is the General Theory of Contractive Iterated Transformations [8, 9]. Basically, fractal coding of an image consists in building a code τ (i.e., a particular transformation) such that, if μ_{orig} is the original image, then $\mu_{\text{orig}} \approx \tau(\mu_{\text{orig}})$, that is, μ_{orig} is approximately self-transforming under τ. If τ is a contractive transformation, μ_{orig} is approximately the attractor of τ, that is, $\mu_{\text{orig}} \approx \lim_{k\to\infty}\tau^k(\mu_0)$ for the some initial image μ_0. The code τ is built on a partition of the original image. Each block R_i of this partition is called Range Block and is coded independently of the others by a matching (local code τ_i) with another block D_i in the image, called a Domain Block. If R and D are the range and domain block's sizes (in case of squared blocks), respectively, then $D = p \cdot R$ with $p > 1$ scaled factor used for the local self-similarity search.

Classical τ_i transforms are isometries (i.e., rotations, flip, etc.) and massive transform (i.e., contrast scaling and grey shifting). If L is the number of range blocks, the fractal code of the initial image is then $\tau(\mu_{\text{origin}}) = \bigcup_{i=1}^{L} \tau_i$ where $\tau_i : D_i \to R_i$ and $\tau_i = M_i \circ I_i \circ r_{i,p}$ with $M_i(x) = a_i \cdot x + b_i$ an affine operator with a scale a_i and a shift b_i on the luminance pixel, I_i a transformation selected from eight discrete isometries, and $r_{i,p}$ a reduction by a factor p using an averaging. In other words, the fractal encoder has to find, for each range block, a larger domain block that constitutes, after an appropriate transformation, a good approximation of the present range block.

The fractal code for the original image is a collection of so extracted local codes. This approach, implemented by Jacquin [9], gives a representation of an image as composed by copies of parts of itself. The classical fractal decoding stage consists in an iterated process starting from an arbitrary initial image μ_0. In fact, if τ is a contractive transformation, the τ's attractor $\tau^{\infty}(\mu_0)$ gives an approximation of the original image μ_{orig} independently from the initial image. Essentially, the fractal code τ is a collection of linear affine transforms τ_i, and it has no intrinsic size. Hence, we can assume that self-similarities, that is, matching between areas with different sizes in the original image are scale independent. As to this, the decoding process results "resolution independent," that is, at decoding stage the fractal code enables zoom [19]. Practically, this operation consists in increasing, during the decoding stage, the range block's size R, and therefore the domain block's size D (being $D = p \cdot R$). For a zoom by a factor z, the new sizes will be $R' = z \cdot R$ and $D' = z \cdot D$, but all the local codes τ_i and consequently the fractal code τ will be unchanged.

In this work we propose a novel solution to the self-similarity search for IFS fractal coding of HD color video sequences. At first we transform the image mapping of the colors in the uniform L*a*b* space. Then, by means of a clustering process, for each range and domain block, we extract a block *signature* which is a summary of the spatial and color information of the image blocks. To obtain the fractal code as previously described, we compare the range and domain blocks signatures by means of the EMD measure. Here, is the novelty of our approach. In fact, we perform the comparison between summaries of the space and colors information contained within image blocks, in opposition to classic IFS schemes that compare them at pixel level by means of RMS measures. A sketch of the algorithm is shown in Figure 1. In the following we give the details of the proposed scheme.

2.1. Image Blocks Signature Extraction. An image block signature is a set of features extracted by means of a clustering process. Clustering is a technique aiming at partitioning the image block in a set of subregions, (i.e., clusters) formed by pixels gathered according to some distance rule. To each cluster is associated a feature representative of the cluster. Formally, given an image block C of size n, its signature $S(C) = \{(c_j, w_j)\}_{j=1}^{T}$, with T number of clusters, w_j weight, and c_j centroid (i.e., the representative element) of the cluster j. For the clustering process we use the classic k-mean algorithm [20]. We measure the distance among pixels both in the spatial and color domains. As to the spatial domain, for every pixel $\{y_i\}_{i=1}^{n}$ we limit the search area to a circle centered in y_i with radius r. The length of r is computed considering the medium spatial distance between y_i and the initial distribution of centroids. The color distance is also upper bounded by the resolution of the HVS in the uniform

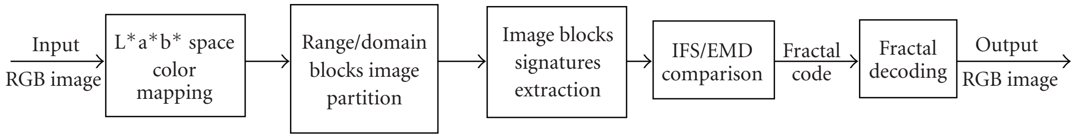


Figure 1: Fractal coding of colour images.

color L*a*b* space (HVS_{res}), that is, the minimum distance in the L*a*b* color space that allows the HVS discriminating two different colors. Formally, we define the distance between the generic pixel y_i and a centroid c_j as

$$\mathbf{d}\left(y_i, c_j\right) = \sqrt{\mathbf{dis}_s^2\left(y_i, c_j\right) + \mathbf{dis}_c^2\left(y_i, c_j\right)},$$

$$\mathbf{dis}_s\left(y_i, c_j\right) = \frac{\left\|y_i - c_j\right\|_{\text{spatial}}}{r} < 1, \quad r = \frac{1}{T}\sum_{j=1}^{T}\left\|y_i - c_j\right\|_{\text{spatial}},$$

$$\mathbf{dis}_c\left(y_i, c_j\right) = \frac{\left\|y_i - c_j\right\|_{\text{L*a*b*}}}{\text{HVS}_{\text{res}}} < 1, \tag{1}$$

where $\mathbf{dis}_s(y_i, c_j)$ and $\mathbf{dis}_c(y_i, c_j)$ are their normalized Euclidean distances in the spatial domain and in the L*a*b* color space, respectively. It is worth noticing that $\mathbf{d}(\cdot,\cdot)$ is nonnegative, symmetric and satisfies the triangle inequality—thus we really work with a metric space. The clustering process associates y_i to c_j according to

$$\mathbf{d}\left(y_i, c_j\right) = \min_j \mathbf{d}\left(y_i, c_j\right). \tag{2}$$

The initial position of the centroids is chosen to be invariant to the possible affine transformation τ performed by the fractal coding. This assures that, given a block signature $S(C)$ and a transform τ, $\tau[S(C)] = S[\tau(Q)]$ [16].

The number of centroids T is chosen as to satisfy two constraints: maximum uniformity in the distance among centroids and invariance to the geometrical affine transformations (i.e., isometries).

The initial set of 12 centroids for an 8×8 size image block is shown in Figure 2. This positioning is spatially homogeneous while the distance between centroids as well as the distance between pixel and surrounding centroids is essentially constant. This displacement is invariant as to the 8 possible isometries. At the end of the clustering process a signature is assigned to each range and domain block.

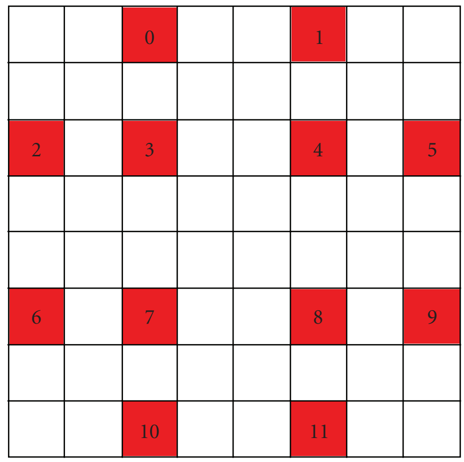


Figure 2: Initial displacement of centroids.

2.2. Earth Mover's Distance for IFS. The self-similarity search within the color image is performed by IFS comparing the signatures of the domain and range blocks as defined in Section 2.1. The matching process relies on the Earth mover's distance (EMD). EMD is a useful and extendible metric distance, developed by the Stanford Vision Laboratory (SLV), based on the minimal cost that must be paid to transform one signature into another. The EMD is based on the *transportation problem* from linear optimization, also known as the Monge-Kantorovich problem [21]. Suppose that several *suppliers*, each with a given amount of goods, are required to supply several *consumers*, each with a given limited capacity. For each supplier-consumer pair, the cost of transporting a single unit of goods is given. The transportation problem is then to find a least expensive flow of goods from the suppliers to the consumers that satisfies the consumers demand. Signature matching can be naturally cast as a transportation problem by defining one signature as the supplier and the other as the consumer, and by setting the cost for a supplier-consumer pair to equal the *ground distance* between an element in the first signature and an element in the second. The ground distance is defined as the distance between the basic features that are aggregated into the signatures. Intuitively, the solution is then the minimum amount of "work" required to transform one signature into the other. Formally the EMD is defined as a linear programming problem. Let P, Q be two image blocks and $S(P) = \{(p_h, w_h)\}_{h=1}^{N}$, $S(Q) = \{(q_k, w_k)\}_{k=1}^{M}$ their signatures with N and M clusters, respectively; let d_{hk} be the ground distance between two centroids p_h and q_k, and let f_{hk} be the flow between p_h and q_k, defined as the amount of weight of p_h matched to q_k, we want to find a flow that minimizes the overall cost:

$$\text{WORK}\left[S(P), S(Q), f_{hk}\right] = \sum_{h=1}^{N}\sum_{k=1}^{M} d_{hk} \cdot f_{hk} \tag{3}$$

with the following constraints:

$$\begin{aligned} & f_{hk} \geq 0, \quad 1 \leq h \leq N, \ 1 \leq k \leq M, \\ & \sum_{k=1}^{M} f_{hk} \leq w_h, \quad 1 \leq h \leq N, \\ & \sum_{h=1}^{N} f_{hk} \leq w_k, \quad 1 \leq k \leq M, \\ & \sum_{h=1}^{N}\sum_{k=1}^{M} f_{hk} = \min\left(\sum_{h=1}^{N} w_h, \sum_{k=1}^{M} w_k\right). \end{aligned} \tag{4}$$

The first constraint assures for unidirectional supplies transportation from $S(P)$ to $S(Q)$. With the second we limit the amount of supplies that can be sent by the clusters in $S(P)$ to their weights. The third constraint allows the clusters in $S(Q)$ to receive no more supplies than their weights, while the last constraint forces to move as much supplies as possible. We call this amount the total flow. Once the transportation problem is solved, and we have found the optimal flow f_{hk}, the EMD is defined as the work normalized by the total flow:

$$\mathrm{EMD}[S(P), S(Q)] = \frac{\sum_{h=1}^{N}\sum_{k=1}^{M} d_{hk} \cdot f_{hk}}{\sum_{h=1}^{N}\sum_{k=1}^{M} f_{hk}}. \tag{5}$$

The normalization is needed when the two signatures have different total weight, to avoid giving more importance to smaller signatures. In general, the ground distance d_{hk} can be any distance and will be chosen according to the problem at hand. We need to define a ground distance that match our purposes. For the extraction of range and domain blocks signatures we deploy a clustering process based on a metric distance as defined in (1). Such a distance was an Euclidean-based metric able to compare pixels and centroids both in the spatial color domains. The comparison is restricted in the spatial domain by r, that is, the medium spatial distance between pixels and the initial distribution of centroids. In the color space the search is limited to the centroids that differ less than the resolution of the human visual system (i.e., the $\mathrm{HVS}_{\mathrm{res}}$). To define the ground distance d_{hk}, we use a similar, but slight different approach. Although we still keep the boundary for the color component, and we use the $\mathrm{HVS}_{\mathrm{res}}$ value to normalize the Euclidean metric, we do not have elements to limit the search area in the spatial domain. Therefore, in the spatial domain, we do not set any constraint, we just normalize the distance component to the maximum measured Euclidean distance between centroids. Moreover, in a matching process based on signatures as above defined, to the success of the search, the importance of the spatial component is not the same as the relevance the color component. In fact, in two image blocks having a similar color distribution, the color position can be very different and this can lead to a weak best match algorithm. As to the above considerations, we propose the following measure for the ground distance:

$$d_{hk} = \sqrt{\lambda \mathbf{dis}_s^2(p_h, q_k) + (1-\lambda)\mathbf{dis}_c^2(p_h, q_k)}, \quad \lambda \in \mathbb{R}, \ 0 < \lambda < 1, \tag{6}$$

where $\mathbf{dis}_s(\cdot,\cdot)$ and $\mathbf{dis}_c(\cdot,\cdot)$ are the same as in (1), but for the former, here, $r = \max_{h,k} \| p_h - q_k \|_{\text{spatial}}$. In fact, the parameter λ in (6) weights the importance given to the color distance respect to the spatial distance and it is chosen as to maximize the quality in the reconstructed image. It is worth remarking that also d_{hk}, as well as $\mathbf{d}(\cdot,\cdot)$ of (5), is non-negative, symmetric and satisfies the triangle inequality, hence it is a true metric. To extract the fractal code, IFS look for similarities between range and domain blocks by comparing their signatures. IFS works with contractive transformations reducing the size of the domain blocks to the one of the range blocks. Therefore, the matching process compares signatures of same total weight. In this case, since the ground distance d_{hk} is a true metric, also the EMD as to (5) defines a metric space. Moreover, it can be shown that in this particular case

$$\begin{aligned} & \mathrm{EMD}[S(P), S(Q)] < d_{pq}, \\ & p = \frac{1}{w}\sum_{h=1}^{N} w_h p_h, \\ & q = \frac{1}{w}\sum_{k=1}^{N} w_k q_k, \\ & w = w_p = w_k, \end{aligned} \tag{7}$$

where w is the total weight of the two signatures and p, q their average centroids. In other words, the ground distance between the average centroids of two signatures of same total weight is a lower bound for the EMD between the two signatures [16]. This property is used by the IFS process to reduce the complexity of the similarities search algorithm. Using the EMD for the IFS best matching search has several advantages. In fact, comparing summary information of image blocks extracted by a clustering process leads to an increased robustness of the search process to the offset errors. This is not true for the pixel-based RMS approach. Moreover, it is less sensitive to quantization errors due to the intrinsic “averaging” nature of the clustering process.

The extension of the theory to video signals is straightforward. In fractal video coding [22] range and domain blocks become three-dimensional objects and thus, the number of isometries and massive transforms to be computed is higher. This fact dramatically raises the computational cost of the matching algorithm. Therefore, the application of fractal coding to video signals turns out to be possible only by following an accurate policy of data reduction and problem simplification. Three-dimensional zooming is achieved using the fractal code extracted from the sequence treated as a three-dimensional object.

3. Proposed Architecture

Within a framework of interactive HDTV applications, the user should select a scene of interest (i.e., a subsequence corresponding to the desired time interval) to be spatially zoomed and replayed in slow motion. The scene of interest is then passed to the proposed architecture shown in Figure 3 and explained in this section.

The scene of interest chosen by the user is at first decomposed in homogeneous shots in order to avoid that scene changes are going to be part of the fractal zoom process.

A set of frames containing the scene of interest is processed by video scene decomposition. In the proposed architecture a scene change algorithm is exploited for determining the optimal temporal window for maximizing the video quality of the zoom and slow motion processing. In fact, the scene of interest chosen by the user for slow motion needs to be preprocessed in order to be partitioned into homogeneous video shots in order to avoid that scene changes participate to the fractal zoom process.

A clustering-based segmentation approach is used for scene change detection. Basically, scene changes are identified on the basis of histogram variations and color variations where this variation is significant in one or both of them. Features function of YUV histogram variation and features function of subsampled YUV frame difference are discriminated choosing appropriate threshold values. The choice of such threshold is automatically performed by means of the Otsu method [23]. The scene change detection algorithm is presented in [23].

Let the homogeneous subsequence identified by the scene change algorithm be composed by M frames.

At first, being the computational complexity of the fractal encoder strictly proportional to the amount of data to be processed, frames are grouped into packets (GOPs) with length N. N is chosen according to the temporal activity of the sequence, so that higher values can be selected for slowly changing scenes without a significant time processing increase. Each GOP is treated as a single unit to be coded. The GOP size is chosen according to the temporal activity within the sequence, so that bigger sizes can be selected for slowly changing scenes without a significant time processing increase.

Packets are selected considering the temporal variance of the sequence, estimated by means of the Minimum Square Error (MSE) metric between frames:

$$\mathrm{MSE}(h,k) = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left(F_{i,j}^{h} - F_{i,j}^{h+k}\right)^2}{M \cdot N}, \quad h,k \in [1,2,\ldots,n], \tag{8}$$

where $F_{i,j}^{p}$ is the pixel (i,j) of the frame and, p is the frame position within the sequence, $M \cdot N$ the frame size, and n the number of frames of the sequence. Among the totality of frames composing the sequence, a certain number of key-frames are selected. A packet is defined as composed by a set of adjacent frames temporally located between two consecutives key-frames, as shown in Figure 4.

At the beginning of the division process the first frame of the sequence to be expanded is chosen as initial key-frame. More in general, once a frame h has been identified as the first key-frame for a packet, a successive frame k is marked as the ending key-frame for the packet if

$$\mathrm{MSE}(h,k) > Th, \tag{9}$$

where Th is a threshold selected so that

$$\mathrm{Th} = \frac{\mathrm{MSE}(1,n)}{2}. \tag{10}$$

In other words, for each packet the temporal variance must be lower than the 50% of the temporal variance of the whole sequence. Equation (10) assures at least a two-packet subdivision of the sequence to be expanded. According to (9) and (10), each packet can be composed by a variable number of frames. At the end of the packetization process, each packet is considered for coding as a single unit: in this manner the computational load and, thus, the time consumed for the coding are significantly reduced.

The drawback of this packetization process is that it introduces a discontinuity along the temporal axis. To limit this effect, time overlapping is used: each GOP is coded having as a boundary condition the motion information of the previous one. Owing to this, the presence of a buffer is necessary to assure the process being causal. A more general constraint is that the GOP size must be a multiple of R, size of the range block, and not smaller than D, size of the domain block. This guarantees the packet being partitioned into range and domain blocks, and not into portions of them.

Within each GOP an active scene detector is used to find the "active object" so that a group of three-dimensional blocks is extracted. Each frame is divided into tiles of M × M size. The EMD among corresponding tiles belonging to different frames is computed. If the EMD is higher than a prefixed threshold, tiles are grouped to form a three-dimensional block. The threshold is adaptively evaluated by averaging the EMD over all tiles composing the GOP. The set of the so extracted blocks defines the active object, the remaining blocks constituting the "background."

The active object is suited to be coded with a full three-dimensional fractal coder whereas the static background is processed with a two-dimensional one. Fractal coding is performed according to the IFS theory [9]: at first, data are partitioned into range and domain blocks; then, a domain pool is created by means of domain blocks and their contractive affine transforms. Each range block is then compared to the elements composing the domain pool by means of the EMD and a set of correspondences among range blocks, domain blocks, and affine transforms (i.e., the fractal code) is created.

Using fractal zoom during the decoding step leads to blockness distortion (Figure 5) along both the time and spatial dimensions. This problem derives from partitioning the video into nonoverlapping range blocks during the encoding process, and overall visual quality decreases when high zoom (i.e., above 4× factor) is performed. To contrast this effect, the coding Overlapped Range Blocks (ORBs) technique [12] is used. ORB coding is extended to the three-dimensional case for the active object coding. Background is encoded with a two-dimensional fractal code, since data does not change, on the temporal axis, for background blocks. Extending [12], eight different partitions of the active object and four partitions for the static background are computed. Eight different fractal codes for the active object are extracted and coded independently (Figure 6).

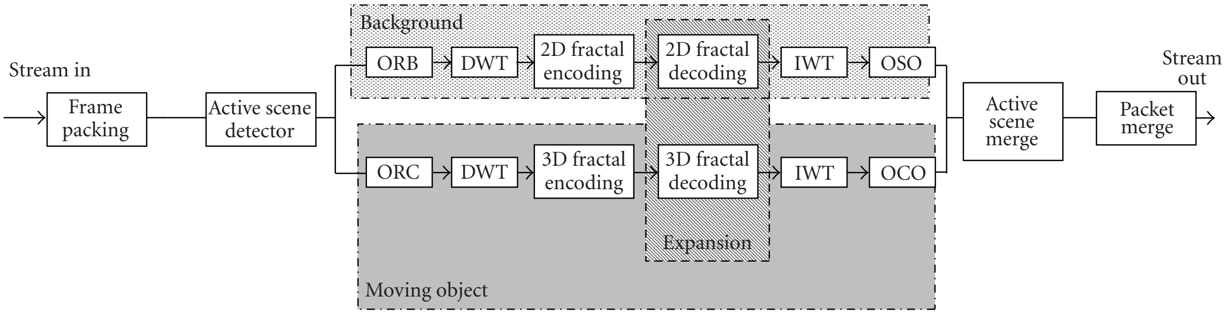


Figure 3: Flowchart of the proposed method.

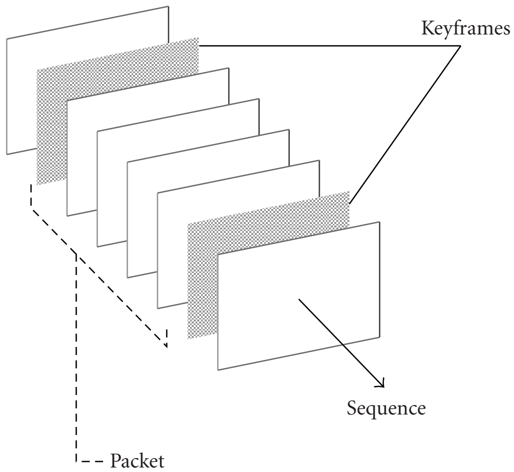


Figure 4: GOP extraction.

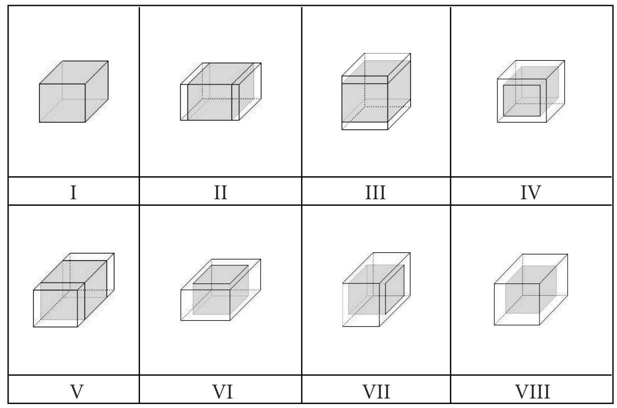


Figure 6: The set of partitions considered for the ORB/OCO process.

Figure 5: Example of block distortion on fractal interpolated frame.

At decoding time, the inverse process is applied, and the fractal zoom is performed. An Ordered Cube Overlapping (OCO) postprocess, defined as an extension of Ordered Square Overlapping (OSO) [12], merges the parts created by the overlapped partition of three-dimensional fractal code. The OSO presented in [12] is a windowed median filter that computes the median value from each partition generated by ORB. The technique is applied here in the three-dimensional case and the OCO computes the median among the eight ORB partitions. A drawback of using ORB and OCO is the growth of the computational cost of the fractal coding process.

To cope with high computational burden, a wavelet-based approach is used [7]. For the active object a three-dimensional wavelet subband analysis is computed. For the entire low pass component a fractal code is then extracted using ORB partitioning. For the high-pass components, the following coefficients classification procedure is performed [24]. Let S_m be the mth subband; we denote by $\{x_i^m\}$ the wavelet coefficients of S_m and by $p^m(x)$ the histogram of $\{x_i^m\}$. In $p^m(x)$, starting from the maximum $x_{\max}$ and moving to the tails of the distribution (see Figure 7), two thresholds are identified, that is, $t_1^m, t_2^m : \int_{t_1}^{t_2} p^m(x)dx = K$, $K \in (0, 1]$.

These thresholds identify the wavelet coefficients constituting the active zone for S_m, that is, $S_m^{az} = \{\forall x \in \{x_i^m\}, x \notin [t_1^m, t_2^m]\}$. In other words, an active zone is composed by those coefficients located on the distribution's tails identified by the above thresholds. After the classification process, a binary-value mask, indicating the position of active zone coefficients within the subband, is extracted. Those coefficients that do not belong to an active zone are discarded, while the S_m^{az} coefficients are ORB partitioned and then fractal encoded. The K parameter is unique for

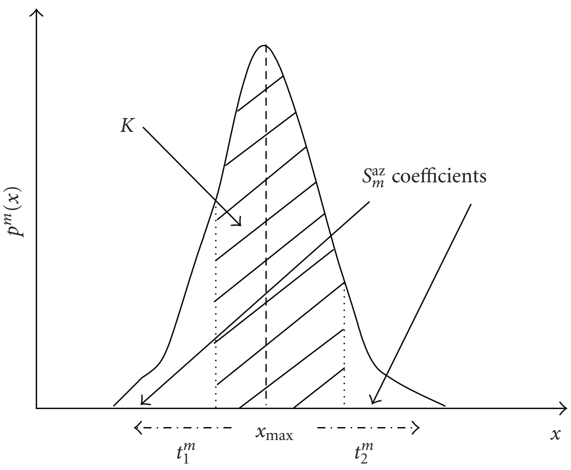


Figure 7: Wavelet coefficient classification process.

all the subbands and controls the speed up, and, on the other hand, the accuracy of the fractal coding process; higher values of K correspond to higher speed up factors, but also to lower final visual quality achieved. At decoding stage OSO/OCO filtering is applied independently to each subband. An additional advantage in terms of time saving of wavelet analysis is the "parallelization" of the entire process that increases the speed in a multithreaded environment.

At decoding time, the inverse process is applied, and the fractal zoom is performed. Since the extracted fractal code is resolution independent, during the decoding process an expansion can be performed independently along each dimension [19].

A three-dimensional (i.e., spatial and temporal) expansion of the active object and two-dimensional spatial zoom (i.e., frames of bigger size) of the background are performed. After the inverse wavelet transformation, an OSO/OCO filtering is performed on the background/active object, respectively. Combined ORB code and OSO/OCO filtering enhance visual quality performance of fractal, by coding reducing blocking artifacts generated by the block based nature of the IFS approach. Finally, an active scene merging and a packets merging processes are applied to release the desired output video sequence.

4. Experimental Results

We tested the effectiveness of the proposed method by comparing the result achieved to those obtained, under the same constraint (i.e., applying the same slow motion factors) by frame replica and classical interpolation techniques. Five HDTV test sequences in 10 seconds shots with 1280 horizontal pixels and 720 vertical pixels (lines), progressively scanned at 50 frames per seconds (namely, 720p/50), were used for experimental tests. Such sequences are freely available at [25]. The sequences are named *CrowdRun, ParkJoy, DucksTakeOff, IntoTree*, and *OldTownCross*. A snapshot from some of these sequences is shown in Figure 8. The first three sequences are classified as "difficult" for what concerns coding complexity while IntoTree and OldTownCross are classified as "easy."

In a framework of broadcasting HDTV, to measure the quality achieved we refer to the video quality assessment described on [26] and formalized in [27]. As to this, the perception of continuous motion by human vision faculties is a manifestation of complex functions, representative of the characteristics of the eye and brain. When presented with a sequence of images at a suitably frequent update rate, the brain interpolates intermediate images, and the observer subjectively appears to see continuous motion that in reality does not exist. In a video display, *jerkiness* is defined as the perception, by human vision faculties, of originally continuous motion as a sequence of distinct "snapshots" [27]. Usually, jerkiness occurs when the position of a moving object within the video scene is not updated rapidly enough. This can be a primary index of a poor performance for a slow motion algorithm. More in general, the total error generated by an incorrect coding of a moving object on a video sequence is representative of spatial distortion and incorrect positioning of the object. In [27] a class of full reference quality metrics to measure end-to-end video performance features and parameters was presented. In particular, [27] defines a framework for measuring such parameters that are sensitive to distortions introduced by the coder, the digital channel, or the decoder. Reference [27] is based on a special model, called the *Gradient Model*. Main concept of the model is the quantification of distortions using spatial and temporal gradients, or slopes, of the input and output video sequences. These gradients represent instantaneous changes in the pixel value over time and space. We can classify gradients into three different types that have proven to be useful for video quality measurement.

(i) The spatial information in the horizontal direction SI_h.

(ii) The spatial information in the vertical direction SI_v.

(iii) The temporal information TI.

Features, or specific characteristics associated with individual video frames, are extracted in quantity from the spatial and temporal information. The extracted features quantify fundamental perceptual attributes of the video signals such as spatial and temporal details. A scalar feature is a single quantity of information, evaluated per video frame. The ITU recommendation [27] divides the scalar features into two main groups: based on statistics of spatial gradients in the vicinity of image pixels and based on the statistics of temporal changes to the image pixels. The former features are indicators of the amount and type of spatial information, or edges, in the video scene, whereas the latter are indicators of the amount and type of temporal information, or motion, in the video scene from one frame to the next.

Spatial and temporal gradients are useful because they produce measures of the amount of perceptual information, or change in the video scene. Surprisingly parameters based on scalar features (i.e., a single quantity of information per video frame) have produced significant good correlation to subjective quality measurement (producing coefficients of

CrowdRun

(a)

IntoTree

(b)

Figure 8: Snapshot of test sequences.

correlation to subjective mean opinion score from 0.85 to 0.95) [26]. This demonstrates that the amount of reference information that is required from the video input to perform meaningful quality measurements is much less than the entire video frame. A complete description of all the features and parameters in [27] is beyond the scope of this paper. In the following a brief summary of the above feature will be given, a mathematical determination of the above features is provided in [27].

(i) *Blurring*. A global distortion over the entire image, characterized by reduced sharpness of edges and spatial detail. Reference [20] defines a Lost Edge Energy Parameter for measuring the blurring-effect, which causes a loss of edge sharpness and a loss of fine details in the output image. This loss is easily perceptible by comparing the Spatial Information (SI) of the output image with the SI of the input image. The lost edge energy parameter compares the edge energy of the input image with the edge energy of the output image to quantify how much edge energy has been lost.

(ii) *Tiling*. Distortion of the image characterised by the appearance of an underlying block encoding structure. Reference [27] defines an HV to non-HV edge energy difference parameter for quantifying the tiling impairment. In contrast to blurring which results in lost edge energy, tiling creates false horizontal and vertical edges. By examining the spatial information (SI) as a function of angle, the tiling effects can be separated from the blurring effects.

(iii) *Error Block*. A form of block distortion where one or more blocks in the image bear no resemblance to the current or previous scene and often contrast greatly with adjacent blocks. Reference [27] defines an Added Motion Energy Parameter for detecting and quantifying the perceptual effects of error blocks. The sudden occurrence of error blocks produces a relatively large amount of added temporal information. So the Added Motion Energy Parameter compares the temporal information (TI) of successive input frames to the TI of the corresponding output frames.

(iv) *Jerkiness*. Motion that was originally smooth and continuous is perceived as a series of distinct snapshots. Reference [27] defines a Lost Motion Energy and Percent Repeated Frames Parameter for measuring the jerkiness impairment. The percent repeated frames parameter counts the percentage of TI samples that are repeated; whereas the average lost motion energy parameter integrates the fraction of lost motion (i.e., sums the vertical distances from the input samples to the corresponding repeated output samples, where these distances are normalised by the input before summing).

To extract the performance metrics we deployed the Video Quality Metric (VQM) software developed by the ITS-Video Quality Research project [28] and compliant with [27]. All tests performed on the different test sequences produced similar outcomes that have been proven to be dependent to the natural temporal activity of the sequences. For the sake of concision, in the following are reported only the results obtained for *CrowdRun* and *IntoTree* video sequences. *CrowdRun* is a sequence presenting a lot of activity since half of the frame is composed by a fast running crowd of people. *IntoTree* is a sequence presenting less activity than *CrowdRun*. Results are reported for a window of 4 seconds corresponding to a cut of 200 frames of the complete video sequence.

Figures 9, 10, 11, 12 show experimental results for the *CrowdRun* sequence. The features mentioned above were computed for the proposed method (Fractal expansion) and for the frame replica and spline cubic interpolation method. The slow motion factors taken into exam are 2×, 4×, and 8×. A general overview of the outcomes shows the advantage in using the proposed technique for higher slow motion factors. Figure 9 shows that blurring distortion for CrowdRun sequence at 2× slow motion is almost the same for all the methods. As slow motion ratio increases the difference between the three techniques becomes more evident. The blurring distortion for the proposed method is lower than the other, and this result is more evident at 4× slow motion. Figure 10 compares the tiling feature. As expected frame replica presents the highest tiling while frame interpolation is superior to the proposed method for 2×

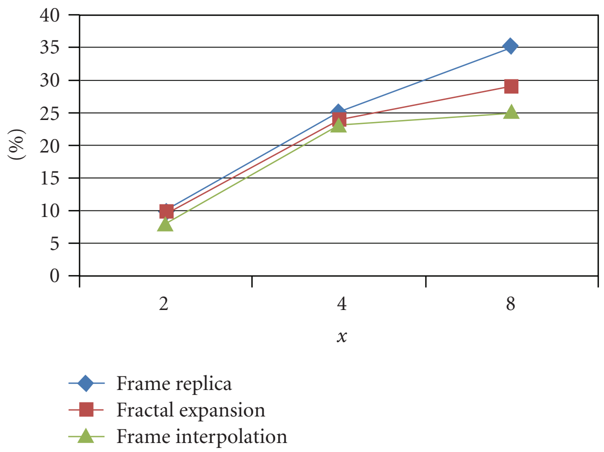


Figure 9: Measured blurring for *CrowdRun* sequence.

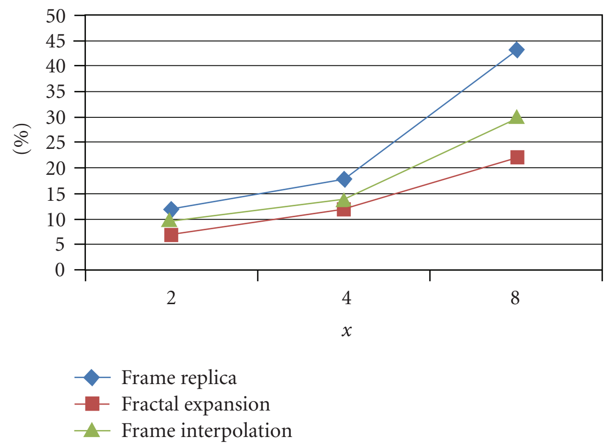


Figure 11: Measured error blocks for *CrowdRun* sequence.

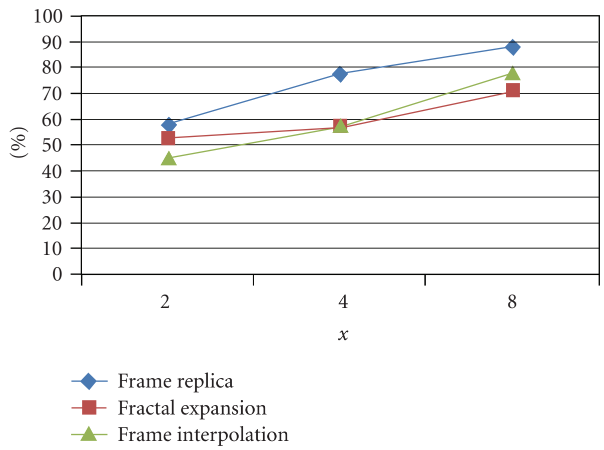


Figure 10: Measured tiling for *CrowdRun* sequence.

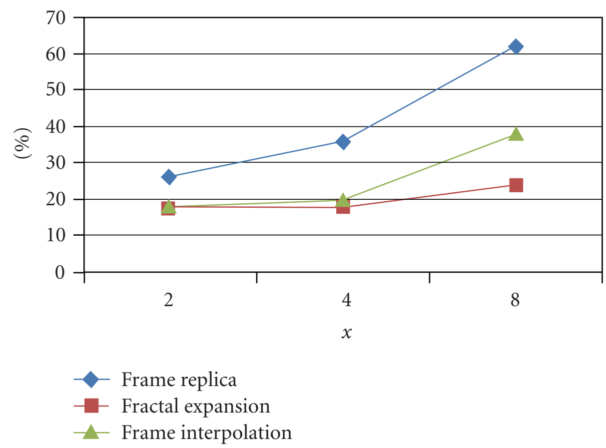


Figure 12: Measured jerkiness for *CrowdRun* sequence.

slow motion factor and vice versa for 8× slow motion factor. Figure 11 compares the error block features. Frame replica results in more error blocks than fractal interpolation and frame interpolation results in more error blocks than fractal expansion. Figure 12 shows the results for jerkiness feature. At 2× and 4× frame interpolation and fractal expansion are comparable while at 8× fractal expansion has a lower jerkiness than frame interpolation.

The use of a combined motion and subband analysis during the fractal coding and again the smoothing properties of the OSO filtering allow the method achieving high performance in terms of fluent motion flow during the presentation in slow motion of video sequences. However, for all the compared methods it is noticeable a degradation of the absolute performance in the presence of high and fast motion in the scene (e.g., for *CrowdRun*). While this is well known for classical frame replica and cubic spline interpolation, the weakness of the proposed algorithm in this case can be justified by the fairly simple method used to estimate the moving object within the sequences. In fact, more sophisticated schemes which assure superior accuracy on motion estimation can be applied and this will be a future task to be pursued during future research.

Figure 13 shows a region of a slow-motion frame from the ParkJoy video sequence. The image of the left is coded by fractal interpolation while the image on the rate is coded by cubic spline frame interpolation. It is evident that both images have a quality noncomparable with the original reference frames. Nevertheless the distortion generated by frame interpolation is subjectively worse than the fractal interpolated frame.

A further experiment session has been set up to compare the performance of the benchmark methods by Peak Signal to Noise Ratio (PSNR). Although it has been proven not to have strong correlation to subjective quality perception in case of video quality assessment [19, 20], PSNR is widely accepted as full reference objective metric for image and video coding.

PSNR experimental tests have been conducted as follows. The test sequences have been subsampled in time by discarding frames. Missing frames were then generated by means of frame replica, cubic spline, and fractal interpolation. PSNRs

(a) (b)

FIGURE 13: A region of *ParkJoy*, frame no. 117, 4× slow motion, coded by Fractal (a) and Cubic Spline (b) interpolation.

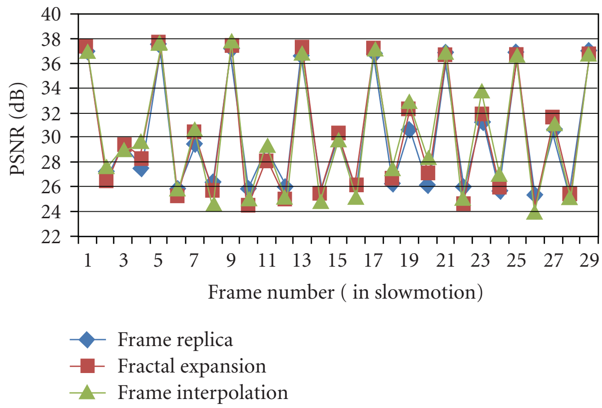


FIGURE 14: Measured PSNR (dB) for *IntoTree* (4× slow motion).

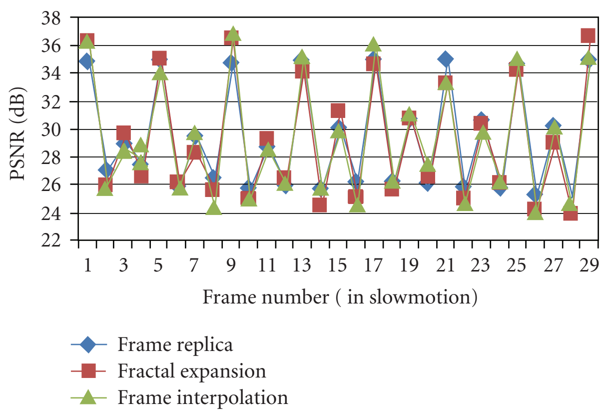


FIGURE 15: Measured PSNR (dB) for *CrowdRun* (4× slow motion).

on generated frames have been calculated. Figures 14 and 15 show the results for *IntoTree* and *CrowdRun* expanded with 4× slow motion factor. For all the compared methods, the average PSNR achieved with *IntoTree* is higher than the average PSNR achieved with *CrowdRun* due to the relative lower motion presence in *IntoTree*. A more in deep analysis shows, in most cases, also a prevalence of the proposed method over the others so as to confirm the results of both the subjective and objective ITU-R Recommendation BT.1683-based previous analysis.

5. Conclusions

In this work we have presented an alternative technique combining fractals theory and wavelet decomposition to achieve spatial zoom and slow motion replay of HD digital color video sequences, which can be integrated on the decoder set-top boxes irrespective of the specific broadcasting technology. Slow motion replay and spatial zooming are special effects used in digital video rendering. At present, most techniques to perform digital spatial zoom and slow motion are based on interpolation for both enlarging the size of the original pictures and generating additional intermediate frames. In our method fast scene change detection, active scene detection, wavelet subband analysis, and color fractal coding based on Earth Mover's Distance (EMD) measure are used to reduce computational load and to improve visual quality. Experiments show that the proposed scheme achieves better results in terms of overall visual quality compared to the state-of-the-art techniques. The proposed approach to video rendering is compliant with the new trend of convergence of digital TV systems and services.

References

[1] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," *IEEE Transactions on Image Processing*, vol. 12, no. 7, pp. 796–807, 2003.

[2] H. Pan, P. van Beek, and M. I. Sezan, "Detection of slow-motion replay segments in sports video for highlights generation," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '01)*, vol. 3, pp. 1649–1652, Salt Lake, Utah, USA, May 2001.

[3] T. Chen, "Adaptive temporal interpolation using bidirectional motion estimation and compensation," in *Proceedings of IEEE International Conference on Image Processing (ICIP '02)*, vol. 2, pp. 313–316, Rochester, NY, USA, September 2002.

[4] K. Hilman, H. W. Park, and Y. Kim, "Using motion-compensated frame-rate conversion for the correction of 3:2 pulldown artifacts in video sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 6, pp. 869–877, 2000.

[5] M. E. Al-Mualla, "Motion field interpolation for frame rate conversion," in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 652–655, Bangkok, Thailand, May 2003.

[6] C. W. Tang and O. C. Au, "Comparison between block-based and pixel-based temporal interpolation for video coding," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '98)*, vol. 4, pp. 122–125, Monterey, Calif, USA, May 1998.

[7] D. D. Giusto, M. Murroni, and G. Soro, "Slow motion replay of video sequences using fractal zooming," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 103–111, 2005.

[8] M. F. Barnsley and S. Demko, "Iterated function systems and the global construction of fractals," *Proceedings of the Royal Society of London, Series A*, vol. 399, no. 1817, pp. 243–275, 1985.

[9] A. E. Jacquin, "Image coding based on a fractal theory of iterated contractive image transformations," *IEEE Transactions on Image Processing*, vol. 1, no. 1, pp. 18–30, 1992.

[10] M. Polvere and M. Nappi, "Speed-up in fractal image coding: comparison of methods," *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 1002–1009, 2000.

[11] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[12] E. Reusens, "Overlapped adaptive partitioning for image coding based on theory of iterated function systems," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '94)*, vol. 5, pp. 569–572, Adelaide, Australia, April 1994.

[13] B. Hurtgen, P. Mols, and S. F. Simon, "Fractal transform coding of color images," in *Visual Communications and Image Processing*, vol. 2308 of *Proceedings of SPIE*, pp. 1683–1691, Chicago, Ill, USA, September 1994.

[14] Y. Zhang and L.-M. Po, "Fractal color image compression using vector distortion measure," in *Proceedings of IEEE International Conference on Image Processing (ICIP '95)*, vol. 3, pp. 276–279, Washington, DC, USA, October 1995.

[15] I. M. Danciu and J. C. Hart, "Fractal color compression in the $L^*a^*b^*$ uniform color space," in *Proceedings of the Data Compression Conference (DCC '98)*, p. 540, Snowbird, Utah, USA, March-April 1998.

[16] S. Cohen and L. Guibas, "The Earth Mover's Distance under transformation sets," in *Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 1076–1083, Kerkyra, Greece, September 1999.

[17] L. Atzori, D. D. Giusto, and C. Perra, "Automatic scene change detection in uncompressed video sequences," in *Proceedings of the Tyrrhenian Workshop on Digital Communications (IWDC '02)*, Capri, Italy, September 2002.

[18] Y. Rubner, L. J. Guibas, and C. Tomasi, "The Earth Mover's Distance as a metric for image retrivial," Tech. Rep. STAN-CS-TN-98-86, Stanford Computer Science Department, Stanford, Calif, USA, 1998.

[19] E. Polidori and J.-L. Dugelay, "Zooming using Iterated Function systems," in *Proceedings of the NATO ASI Conference on Fractal Image Encoding and Analysis*, Trondheim, Norway, July 1995.

[20] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, University of California Press, Berkeley, Calif, USA, 1967.

[21] S. T. Rachev, "The Monge-Kantorovich mass transference problem and its stochastic applications," *Theory of Probability and Its Applications*, vol. 29, no. 4, pp. 647–676, 1987.

[22] K. U. Barthel and T. Voye, "Three-dimensional fractal video coding," in *Proceedings of IEEE International Conference on Image Processing (ICIP '95)*, vol. 3, pp. 260–263, Washington, DC, USA, October 1995.

[23] O. Nobuyuki, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[24] M. Ancis and D. D. Giusto, "Image data compression by adaptive vector quantization of classified wavelet coefficients," in *Proceedings of the 6th IEEE Pacific RIM Conference on Communications, Computers, and Signal Processing*, vol. 1, pp. 330–333, Victoria, Canada, August 1997.

[25] ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat.

[26] S. Wolf, "Measuring the end-to-end performance of digital video systems," *IEEE Transactions on Broadcasting*, vol. 43, no. 3, pp. 320–328, 1997.

[27] ITU-R Recommendation, "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference," Tech. Rep. BT.1683, International Telecommunication Union, Geneva, Switzerland, 2004.

[28] Video Quality Research project, http://www.its.bldrdoc.gov/n3/video/.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 614178, 10 pages
doi:10.1155/2009/614178

Research Article

AL-FEC for Improved Mobile Reception of MPEG-2 DVB-T Transport Streams

David Gozálvez,[1] David Gómez-Barquero,[1] Thomas Stockhammer,[2] and Michael Luby[2]

[1] *Universidad Politécnica de Valencia (UPV), Mobile Communications Group, 46022 Valencia, Spain*
[2] *QUALCOMM Incorporated, Santa Clara, CA 95051, USA*

Correspondence should be addressed to David Gozálvez, dagoser@iteam.upv.es

Received 28 February 2009; Accepted 16 July 2009

Recommended by Sandro Scalise

We investigate the use of application layer FEC protection in DVB-T (Digital Video Broadcasting-Terrestrial) networks for the provision of mobile services. Mobile reception is characterized by variations of the received signal caused by fast fading and shadowing. DVB-T was originally designed for fixed and portable reception, and generally does not provide enough quality in mobile environments. The link layer protection mechanism MPE-FEC (Multi Protocol Encapsulation-Forward Error Correction) was standardized in DVB-H (Digital Video Broadcasting-Handheld) for the protection of mobile TV services. Although DVB-T itself does not incorporate any link or application layer protection mechanism, AL-FEC (Application layer Forward Error Correction) protection can be introduced in DVB-T in a backwards compatible way. By means of AL-FEC, it is possible to improve the robustness of DVB-T services for the provision of mobile TV. In this paper, we explain the concept of AL-FEC protection in DVB-T and evaluate its performance by means of laboratory measurements and dynamic simulations with shadowing. We study different configurations of AL-FEC and compare its performance with MPE-FEC. In this paper, we discuss some implementation aspects of AL-FEC in real scenarios and propose an implementation based on Raptor codes and hash sequences. We also present results obtained by a first AL-FEC prototype for DVB-T that demonstrates the feasibility of the approach.

1. Introduction

Digital Terrestrial TV (DTT) networks are being deployed worldwide, and it is planned that DTT services completely replace analogue TV in many European countries by latest 2012. DVB-T (Digital Video Broadcasting-Terrestrial) [1] is the European standard of DTT, and has been adopted by many countries all over the world to provide DTT services. DVB-T was designed for fixed and portable reception and generally does not provide enough robustness in mobile environments. Mobile reception is characterized by fluctuations of the received signal due to fast fading and shadowing. These fluctuations cause the loss of portions of information over time, and challenge the reception of DVB-T services in mobile environments. A main reason for this is the short interleaving performed in the physical layer (limited to one OFDM symbol which corresponds approximately to 1 millisecond).

The European digital mobile TV standard called DVB-H (Digital Video Broadcasting-Transmission System for Handheld Terminals) is a technological evolution of DVB-T, and was developed specifically for the provision of mobile TV services. DVB-H reutilizes the physical layer of DVB-T and introduces a set of enhancements in the link layer in order to adapt the transmission to mobile reception. These enhancements are aimed to reduce the terminal power consumption and counteract fast fading. A link layer protection mechanism called MPE-FEC (Multi Protocol Encapsulation-Forward Error Correction) [2] increases the robustness in mobile environments whereas a bursty transmission technique referred to as *time slicing* reduces the power consumption in receivers up to 90%. Simulations for DVB-H have shown that MPE-FEC provides gains up to 9 dB for mobile users when compared to DVB-T. Furthermore, the maximum Doppler tolerance increases by about 50% in mobile channels while reutilizing the physical layer of DVB-T [3].

AL-FEC (Application Layer Forward Error Correction) has been standardized in DVB-H for file delivery services. The advantage of AL-FEC is that it can spread the protection over large portions of information. AL-FEC takes advantage of the temporal diversity derived from user mobility by the use of extensive time interleaving (up to minutes or even hours), which increases the robustness of the transmitted information against fast fading and especially, against shadowing and signal outages. AL-FEC has been also proposed for DVB-H streaming services in the form of multi-burst protection [4]. Despite excellent performance, the main drawback of this approach is an increase in the channel switching times, which is considered as a crucial parameter in mobile TV usability.

Despite the fact that the physical layer of DVB-H is compatible with DVB-T, DVB-H encapsulates all the audiovisual information in IP (Internet Protocol) datagrams and generally simulcasts the services with lower quality than the MPEG-2 Transport Stream signal of DVB-T. Consequently, DVB-H requires the allocation of specific bandwidth for the transmission of the mobile TV content.

In some studies as well as even deployments, mobile reception of DVB-T has been verified. In order to enable mobile reception of current DVB-T services, antenna diversity techniques have been proposed. In [5], it is shown that the reception by means of two antennas results in a similar performance to MPE-FEC. However, for handset-based reception at UHF frequencies, multiple receiving antennas are generally impractical as the correlation distance of the antennas is far beyond the dimensions of typical handsets.

In this paper, we study the use of AL-FEC protection in DVB-T in order to increase the robustness of the transmitted information and to achieve reception in mobile channels. Specifically, we introduce mechanisms that allow the transmission of the additional FEC data needed for error correction (similar to the MPE-FEC in DVB-H) in a backward compatible way, that is, in such a way that the legacy DVB-T receivers are not impacted by this additional FEC. We also evaluate the mobile performance of AL-FEC in DVB-T and provide direct comparisons with MPE-FEC. Although different codes may be used in order to provide AL-FEC protection for DVB-T services, in this paper we consider the use of Raptor codes [6], as they are efficient and lightweight in terms of decoding complexity. This permits the FEC decoding to be done in generic software processors even on low-complexity devices and therefore does not require any upgrades in hardware. The requirements in terms of complexity and memory on mobile terminals are almost negligible as has been shown in the context of 3 GPP MBMS or other mobile TV standards that integrate Raptor codes.

The rest of the paper is organized as follows. In Section 2, we discuss the typical characteristics of mobile reception and the impairments of the mobile channel. In Section 3 we review the transmission of content in DVB-T networks and detail the implementation of AL-FEC in DVB-T. Section 4 is dedicated to the description of the simulation scenario and then, in Section 5, we present selected illustrative simulation results. In Section 6, we also address the issues of the practical implementation of AL-FEC in DVB-T. Section 7 provides considerations on DVB-T network planning for the provision of mobile TV services. Some concluding remarks are provided in Section 8.

2. Mobile Reception of DVB-T

Mobile channels are characterized by rapid variations of the received signal over time referred to as fast fading. Fast fading is caused by the Doppler shift of multiple propagation paths, which originate from the movement of the receiver with respect to the transmitter [7]. Fast fading results in the corruption of small portions of the data stream in a bursty manner due to ICI (Inter Carrier Interference). Higher velocities involve higher values of Doppler shift and thus, a greater degradation of the received information. If the user velocity is too high, the Doppler shift may increase above the values supported by the physical layer, corrupting great portions of information. The C/N (Carrier to Noise) sensitivity required for the proper reception of DVB-T tends to increase proportionally with the Doppler shift until a maximum Doppler value, from which reception is no longer possible. The protection applied by the physical layer of DVB-T extends only to the duration of one OFDM symbol (approximately 1 millisecond), and it is not capable to cope with the error bursts resulting from fast fading.

DVB-H reutilizes the physical layer of DVB-T, but integrates MPE-FEC at the link layer to repair the errors caused by mobile reception. MPE-FEC is an intra burst mechanism for which the protection is performed on a per burst basis [2]. MPE-FEC protection spreads over the duration of one burst (0.1–0.4 seconds) and is capable to counteract the effects of fast fading. By means of MPE-FEC it is possible to achieve a performance that is almost independent of the Doppler and in addition, the maximum Doppler supported by the system can be increased. In order to cope with mobile reception impairments in DVB-T systems, antenna diversity techniques have been proposed. In [5] it is shown that by means of 2 receiving antennas, it is possible to achieve a similar performance to MPE-FEC against fast fading.

The received signal in mobile channels is also characterized by slow variations known as shadowing. Shadowing results from the presence of large obstacles, such as buildings or hills that may block the line-of-sight between the receiver and the transmitter. Shadowing is generally modelled as a log-normal distributed variation of the received signal over the area of coverage [7]. When a user is moving in the presence of shadowing, the received signal may experience outages that corrupt longer periods of the data stream. Signal outages can be corrected if the protection of the physical or upper layers is spread over time. Such link layer protection has, for example, been standardized in [8, 9] for its use in DVB systems.

3. AL-FEC Protection of DVB-T Services

3.1. System Architecture. For the integration of AL-FEC into DVB-T, the architecture according to Figure 1 is

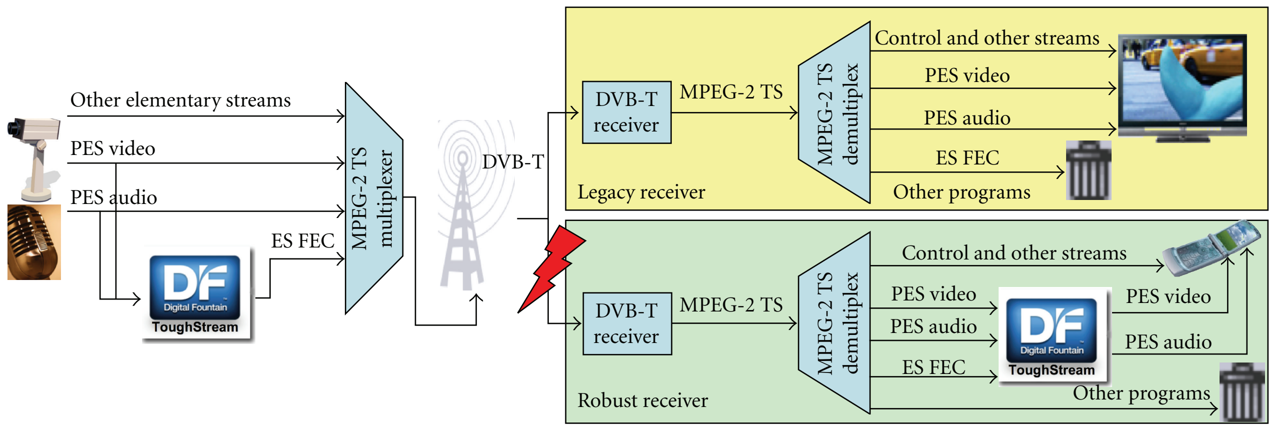


Figure 1: Proposed system architecture.

proposed. The elementary streams of the program to be protected are also processed by an FEC encoder at the transmitter. This entity generates an FEC elementary stream that is multiplexed with the MPEG-2 Transport Stream (TS) and distributed over DVB-T. Legacy receivers receive the MPEG-2 TS and should drop the FEC elementary stream. A robust receiver will make use of the FEC elementary stream to retrieve lost audio and video data, such that the original signal can be reconstructed. The FEC data needs to be multiplexed such that legacy receivers can drop it without impacting their operations. Furthermore, the robust receivers should be able to reconstruct as much data as possible by the use of the FEC elementary stream.

The FEC stream generated by the FEC encoder consumes part of the bit rate capacity at the physical layer and thus, the number of services carried per MPEG-2 TS may have to be reduced in order to accommodate the FEC data. The main difference of the proposed approach with respect to DVB-H lies in the fact that the same multimedia content transmitted to fixed receivers is protected for its use by mobile users. Therefore, no additional content is needed for the transmission of mobile services and only the capacity required for carrying the FEC data must be taken into account to support mobile reception.

Based on this short system description and the presented requirements we will now further detail the different components of the proposed system, in particular the MPEG-2 Transport Stream protocol as well as the Forward Error Correction scheme and methods.

3.2. MPEG-2 Transport Stream. In DVB-T all the content is multiplexed in an MPEG-2 TS and transmitted as a sequence of TS packets [10]. Each TS packet carries a header of 4 bytes and a payload of 184 bytes. The MPEG-2 TS contains all the data from the services multiplexed in the MPEG-2 TS along with signalling information, which is carried in the form of PSI/SI (Program Specific Information/Service Information) tables. Generally, several services (TV programs, radio programs, data channels . . .) are multiplexed in one MPEG-2 TS. The video, audio and data of each service in the MPEG-2 TS is referred to as Elementary Stream (ES). The header of each TS packet contains a 13-bit Packet IDentifier (PID) that uniquely identifies the ES that is carried inside the TS packet. The header of the TS packets also contains a Transport Error Indicator (TEI) bit and a Continuity Counter (CC) field that can be used for the detection of erroneous and missing packets.

Each ES is assigned a unique PID value inside the MPEG-2 TS. The associations between ESs and PID values are transmitted in the PSI/SI tables. DVB-T receivers parse the PSI/SI tables in order to identify the PID values of the ESs corresponding to the desired service. Then, the TS packets carrying the video, audio or data information of the service are demultiplexed by looking at the PID value of every TS packet.

3.3. Forward Error Correction. FEC mechanisms are designed to cope with the loss of information by transmitting additional FEC data. Erasure codes are often used in FEC mechanisms as they can regenerate lost portions of information transmitted over an erasure channel. In order to do so, the information to be protected is partitioned into source blocks, each of them constituted by k different source symbols. An erasure encoder is used in the transmitter to encode the source blocks and generate a total amount of n symbols per source block where $n > k$. If a systematic code is used, the original k source symbols are among the total n symbols generated by the encoding algorithm. The k original source symbols are transmitted along with the n-k repair symbols. Assuming an erasure channel, some of the n transmitted symbols are erased and are not available in the receiver. An erasure decoder is capable to recover the erased symbols if a sufficient number of source and repair symbols is received. An ideal erasure code is capable of recovering the k original source symbols if at least any k symbols among the n transmitted symbols are received. A practical low-complexity erasure code generally requires a small additional number of symbols in order to recover all original transmitted source symbols.

The protection provided by FEC mechanisms depends on the code rate and the protection period. The code rate is the proportion of source data with respect to the

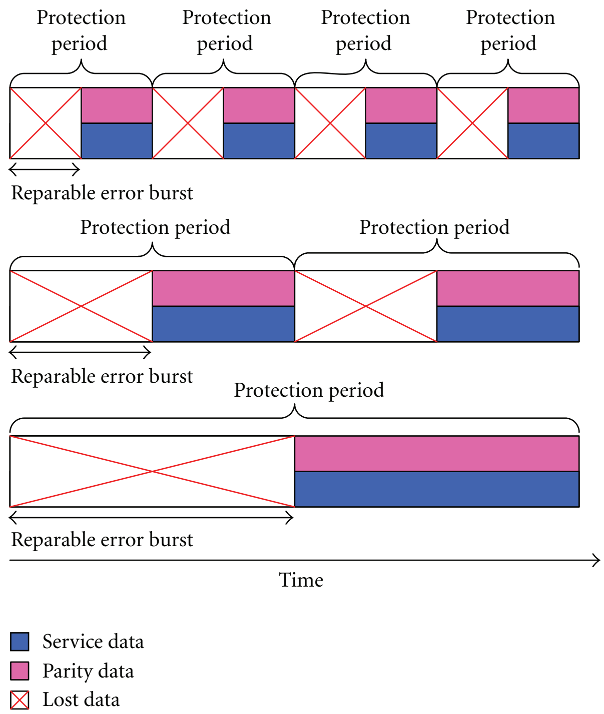


Figure 2: Example of the impact that the protection period holds on the error correction capabilities of FEC mechanisms.

total amount of information transmitted, accounting both source and repair data. The protection period refers to the duration of the information encoded in each source block. Figure 2 illustrates the effect of the protection period in FEC protection mechanisms. The code rate is assumed to be 1/2 and thus, it is possible to recover the service data by receiving at least half of the total transmitted data. The figure shows the burst correction capabilities for three different protection periods. The lost portions of data illustrated in the figure represent the maximum error burst duration that each protection period is capable of repairing. As can be seen in the figure, longer protection periods operate with greater portions of information and can recover from longer error bursts.

The protection period has an impact not only in the protection provided by the FEC mechanism but also in the network latency and, most importantly, in the receiver latency and channel switching time. The network latency can be defined as the amount of time that passes from the moment the information enters the transmitter till the moment is delivered to the media decoders in the receiver. The channel switching time refers to the amount of time between the instant when the user switches to a new channel and the instant when the new content is displayed to the user. Although network latency is not essential for the majority of services, the channel switching time is considered as a crucial criterion in mobile TV user experience, and must not be increased beyond certain values. There is a trade off between the level of protection that can be offered in mobile reception and the channel switching time that the user experiences as disturbing. Increasing the protection period has also an impact on the memory requirements in the receiver, as at least the size of the source data contained in one protection period must be stored in order to perform the decoding.

3.4. Application Layer FEC. The main idea of AL-FEC in DVB-T is to incorporate FEC protection by making use of erasure codes in a backward compatible way. In order to achieve this, the video and audio ESs must not be altered. Apart from the repair symbols, some additional information such as Source FEC Payload Identifier (ID) is necessary for the decoding process and must be passed to the receiver. For the purpose of connecting source and repair data, hash sequences of the source symbols are sent along with the repair symbols to provide this Source FEC Payload ID. By doing so, the source data is unmodified but the advanced receiver can still retrieve the necessary information by producing the same hash sequences. Moreover, the FEC data, including both the repair symbols and the hash sequences, must be encapsulated in a manner that ensures that legacy DVB-T receivers drop the TS packets carrying the FEC data without altering its proper operation. The MPEG-2 TS specification allows AL-FEC to be incorporated into the protocol stack of DVB-T in a transparent manner above the TS layer. The repair packets can be multiplexed into the MPEG-2 TS as another ES associated to the service, and will be discarded by the DVB-T receivers that do not incorporate AL-FEC. This is accomplished by assigning a specific new PID to the FEC elementary stream that is not recognized by the legacy receiver. By means of the TEI and the continuity counter fields in the TS packet header it is possible to detect erroneous or missing TS packets. As the source and repair packets are encapsulated in TS packets, the erasure of TS packets results in a symbol erasure channel. A source or repair symbol is considered erased if at least one of the TS packets carrying information of that symbol is lost. Longer source and repair packets are usually fragmented and encapsulated into multiple TS packets. As one erroneous or lost TS packet is sufficient to erase the entire source or repair packet, they tend to achieve a lower performance, especially in the presence of uncorrelated MPEG-2 TS packet errors.

In this paper we consider the use of Raptor codes [6] for AL-FEC in DVB-T services. They have been previously standardized in DVB systems for the provision of link layer FEC protection [8, 9]. Raptor codes are a computationally efficient implementation of fountain codes that achieve very close to ideal performance. Fountain codes are a class of erasure codes that can generate a very large amount of FEC data from a given source block and thus are considered as rate less (i.e., any amount of FEC data can be delivered for any source block size or protection period). Raptor decoding can be implemented in complexity and memory constrained receivers such as handset receivers without the need of dedicated hardware due to its low computational complexity and its efficient memory management.

In Figure 3 we compare the operation of MPE-FEC and AL-FEC. Because of the time slicing performed in DVB-H, the protection period achieved by MPE-FEC is limited to the

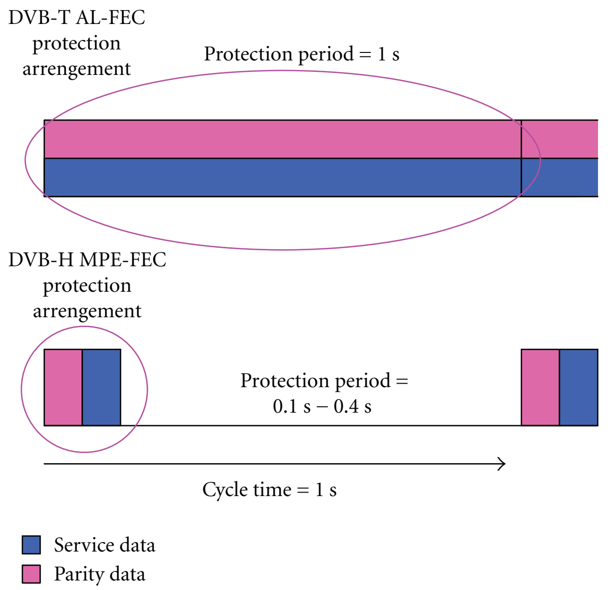


Figure 3: Comparison between the protection arrangement of MPE-FEC in DVB-H and AL-FEC in DVB-T.

burst duration of around 0.1–0.4 seconds. In DVB-H, when a new channel is selected by the user, the receiver must wait until the reception of the first burst of the service. Assuming no other delays, the channel switching time is equal to half the cycle time (i.e., period of time between bursts) on average and as high as one cycle time in the worst-case scenario. On the other hand, DVB-T does not perform time slicing, and the services are transmitted continuously over time. The protection period in AL-FEC can be configured up to 10 seconds and even more, and is only restricted by memory and channel switching time constraints. Although DVB-T does not perform time slicing and thus, it is not possible to achieve power saving, the development of more durable batteries has reduced the power consumption issues in handheld terminals. Furthermore, for a significant portion of mobile TV receivers such as netbooks or in-car receivers, battery lifetime is of less relevance.

Another limitation of MPE-FEC that can be overcome by AL-FEC is the dependency between code rate and protection period. Due to the nature of Reed Solomon encoding, in order to achieve different code rates other than the mother code rate (3/4), it is necessary to perform a padding/puncturing mechanism that shortens the burst duration. This is especially critical for code rates 2/3 and 1/2, for which the burst duration is, respectively, a 25% and a 50% shorter than in the case of code rate 3/4. On the contrary, the flexibility of Raptor allows the proposed Al-FEC implementation to deliver virtually any code rate for a given protection period.

4. Performance Evaluation

The simulations have been performed assuming a DVB-T physical layer configuration of FFT 8K, guard interval 1/4, nonhierarchical modulation 16QAM, and code rate 1/2. The transmission mode employed gives a total bit rate of approximately 9.95 Mbps. For the evaluation of AL-FEC in DVB-T services, a service of 2.5 Mbps has been protected by an ideal AL-FEC implementation. The size of the source and repair symbol is configured to 184 bytes, which corresponds to the payload of one TS packet and therefore each symbol is directly be mapped to one MPEG-2 TS packet. Initially, we have considered that no additional overhead (e.g., hash sequences) is required to perform the decoding, so all the FEC data is dedicated to the carriage of repair symbols. These assumptions correspond to an upper limit on the performance, but the expected deviation from real implementations is marginal as we will discuss later. We have also assumed an ideal erasure code for the encoding/decoding algorithm, so the source blocks can be decoded successfully without any additional FEC symbol overhead.

For the MPE-FEC evaluation, we have assumed a DVB-H service with a cycle time of 2 seconds where all the transmitted IP datagrams have a constant size of 512 bytes. The number of rows of the MPE-FEC frame has been configured to 512, and thus, each IP datagram fits exactly in one column of the MPE-FEC frame. The amount of columns of IP information and FEC data were configured according to each particular code rate.

The mobile performance has been evaluated by means of TU6 laboratory measurements. The laboratory measurement setup consists of a DVB-T modulator, a signal generator in charge of emulating the TU6 channel model and a DVB-T measurement system capable of recording the error data at the TS layer. By recording the error data at the TS layer it is possible to emulate the performance of FEC mechanisms in upper layers. The measurements were obtained in a range of CNR values from 0 to 30 dB and in a range of Doppler values from 5 to 80 Hz.

In order to study the influence of the shadowing, we have assumed a user moving at constant velocity across a log-normal CNR map defined by its standard deviation and correlation distance. The user velocity is given by the values of Doppler and frequency carrier configured in the simulations. The shadowing model outputs instantaneous CNR values that correspond to 100 millisecond time intervals. The instantaneous CNR values along with the Doppler determine the measured error data in each one of the 100 ms time intervals that will be passed to the link layer.

We have defined two different scenarios: a high diversity scenario with a correlation distance of 20 meters and a Doppler frequency of 80 Hz (which correspond to 144 km/h at 600 MHz), and a low diversity scenario with a correlation distance of 100 meters and a Doppler frequency of 10 Hz (which correspond to 18 km/h at 600 MHz). In both cases the standard deviation has been set to 5.5 dB which is the usual value employed in mobile TV planning for outdoor reception. The results have been averaged over 100 seeds of shadowing.

We have employed the PER (Packet Error Ratio) in order to measure the performance of both MPE-FEC and AL-FEC. The PER is defined as the percentage of packets in which there is at least one error. In the case of MPE-FEC, a packet

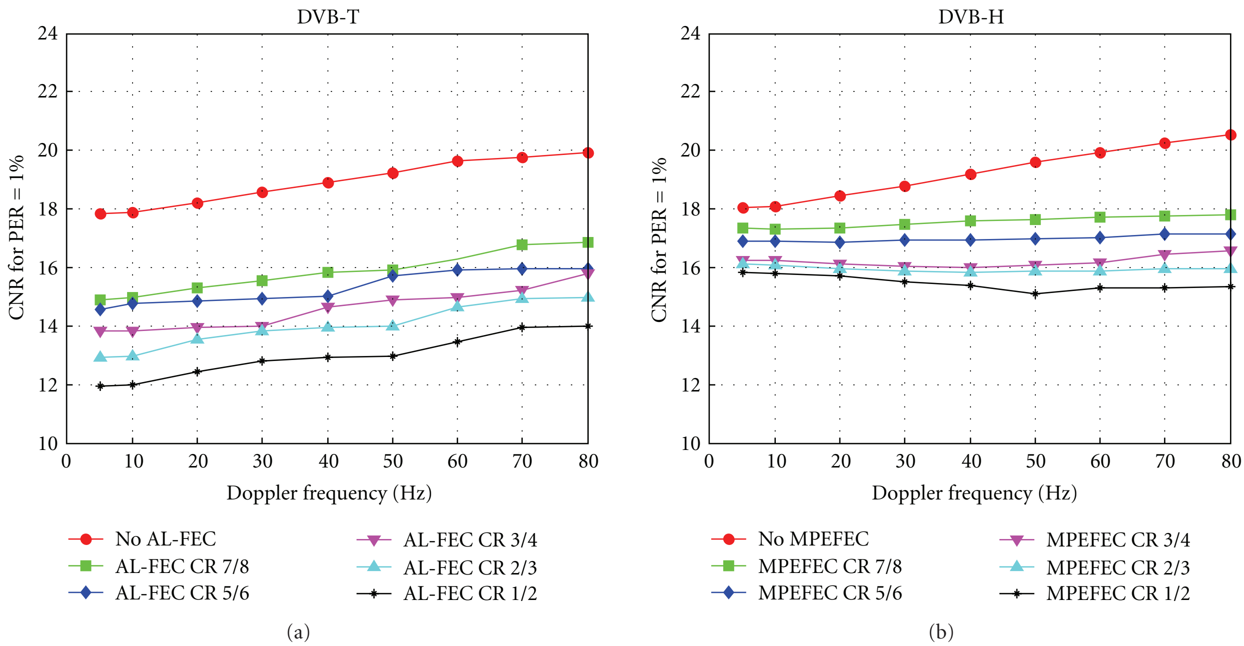


Figure 4: Mobile Performance of DVB-T (a) and DVB-H (b) services with AL-FEC in a TU6 channel.

corresponds to a MPE section (in which an IP datagram is encapsulated), whereas in the case of AL-FEC a packet corresponds to a source packet. Although this is not a one-to-one comparison it is a reasonable way to compare the performance of both mechanisms. We have considered a PER value of 1% as QoS criterion, which is more demanding than the MFER (MPE-FEC Frame Error Ratio) value of 5% generally employed in DVB-H evaluations.

5. Results and Discussions

5.1. Mobile Performance in the Presence of Fast Fading. First, we compare the mobile performance of DVB-T and DVB-H in Figure 4. AL-FEC has been configured to a protection period of 1 second in order to be comparable to the DVB-H service (cycle time equal to 2 seconds) in terms of channel switching time.

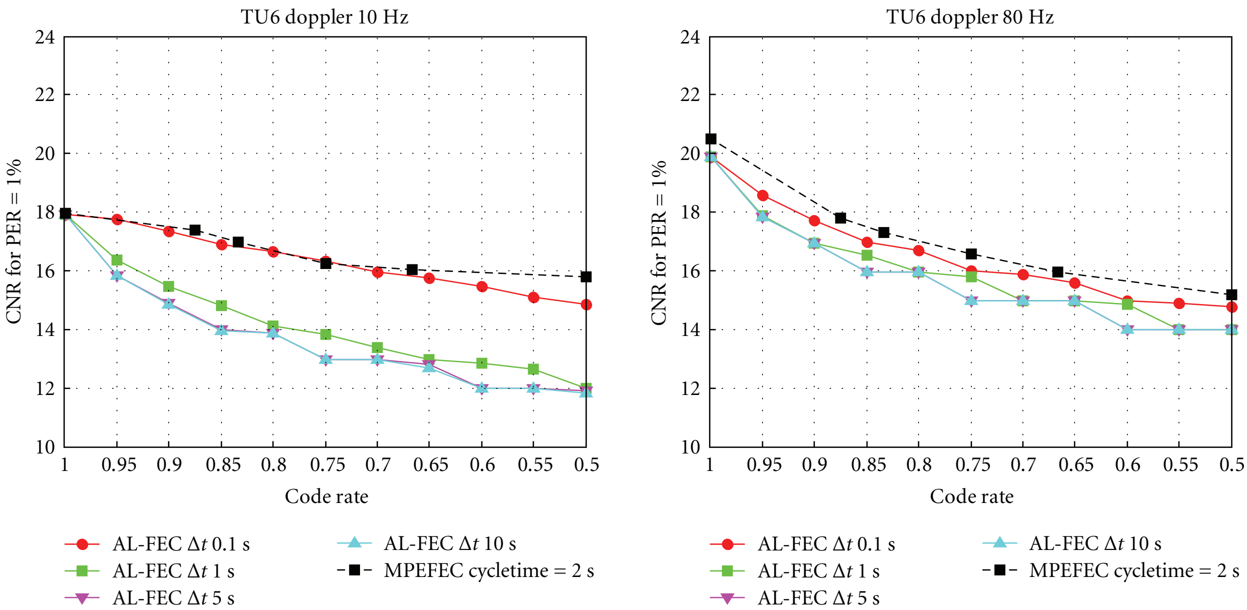
As can be seen in the figures, if no link or application layer protection is provided in DVB-H and DVB-T, both systems achieve a very similar performance. DVB-T and DVB-H employ the same physical layer, and the performance difference is due to the source packet size (528 bytes in the case of MPE-FEC and 184 bytes in the case of AL-FEC). As it is shown in the figures, the performance degradation due to the packet size is more severe for high Doppler values.

When MPE-FEC and AL-FEC are enabled, the performance of DVB-H and DVB-T increases considerably, achieving important gains in terms of CNR threshold. The gain obtained by MPE-FEC for the lowest code rate is up to 2 dB in the lower range of Doppler and up to 6 dB in the higher range. The performance difference between the lowest and highest code rates is around 2 dB. These values correspond quite well with the results presented in [3], which validates the methodology employed in this paper.

The gain obtained by AL-FEC is up to 6 dB in the entire range of Doppler. This represents an additional gain of 4 dB with respect to MPE-FEC in the case of low Doppler. This gain is originated by the longer effective protection period of AL-FEC. Despite the fact that MPE-FEC is configured with a cycle time of 2 seconds, the effective protection period is equal to the burst duration. With the configuration employed in the DVB-H service, the burst duration is approximately of 100 ms for the code rates 7/8, 5/6 and 3/4, of 80 ms for the code rate 2/3, and of 50 ms for the code rate 1/2. On the contrary, AL-FEC is configured with a protection period of 1 second that is independent of the code rate. The errors caused by low Doppler shifts tend to be more correlated over time, and longer protection periods can cope more effectively with this kind of error distributions.

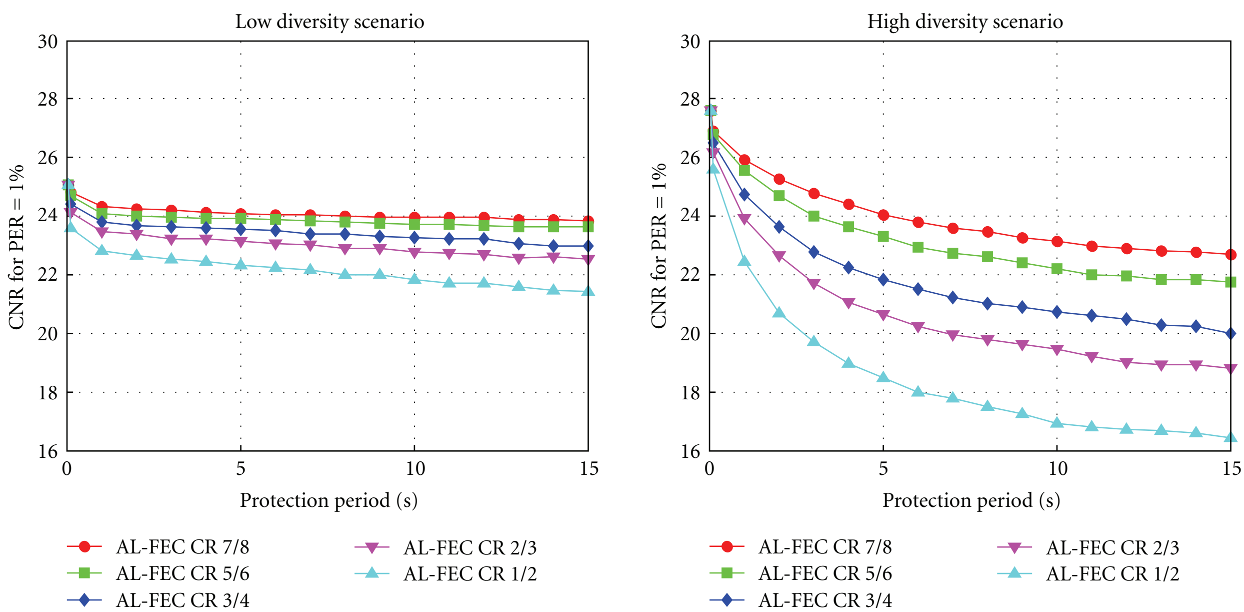
Figures 5 makes a direct comparison between MPE-FEC and AL-FEC with different values of protection period. As we have seen, the effective protection period achieved by MPE-FEC in the simulations is between 50 and 100 ms. In MPE-FEC, lower code rates operate with higher percentages of FEC data but involve shorter protection periods. As expected, in the case of code rates 7/8, 5/6, and 3/4, MPE-FEC obtains results that match those of AL-FEC with a protection period of 0.1 seconds. In the case of code rates 2/3 and 1/2, the performance of MPE-FEC degrades slightly with respect to AL-FEC due to the shorter burst duration. It is also remarkable the fact that with 10 Hz of Doppler, 1 second of protection period losses less than 1 dB with respect to 5 and 10 seconds.

Figure 5 shows again how higher values of Doppler degrade more severely the performance of MPE-FEC due



(a) (b)

FIGURE 5: Performance of AL-FEC in a TU6 channel configured with 10 Hz (a) and 80 Hz (b) of Doppler.



(a) (b)

FIGURE 6: Performance obtained by AL-FEC (a) in the low diversity scenario (correlation distance of 100 m and Doppler of 10 Hz) and (b) in the high diversity scenario (correlation distance of 20 m and Doppler of 80 Hz).

to the larger source packet size. In the figure we can also observe the little impact that the protection period has in a TU6 channel with high Doppler. Almost all the values achieve the same performance, and 0.1 seconds of protection period looses less than 1 dB with respect to 10 seconds. As it has been demonstrated, longer protection periods than 1 second do not provide a remarkable advantage in a TU6 channel.

5.2. Mobile Performance in the Presence of Fast Fading and Shadowing. In Figure 6 we can see the evolution of the performance of AL-FEC as the protection period increases for the two scenarios with shadowing defined for the simulations. We can observe an increase in the requirements of CNR caused by the effect of shadowing in mobile reception.

MPE-FEC performance is not presented as the results are approximately the same as with an AL-FEC configuration of 0.1 seconds. The results show the little impact of the protection period in low diversity conditions. In the low diversity scenario, there is almost no gain when increasing the protection period, especially in the case of the less robust code rates. Increasing the protection period from 0.1 to 15 seconds only yields a gain between 1 and 2 dB depending on the code rate. On the contrary, increasing the protection period from 0.1 to 15 seconds provides a gain from 4 to 10 dB in the high diversity scenario.

6. Implementation Issues of AL-FEC

6.1. FEC Implementation Issues. In this paper we have evaluated so far the performance of an ideal AL-FEC implementation in DVB-T. In a practical implementation some additional issues need to be addressed. A main issue is the transmission of the additional information required in reception for the decoding process. As already mentioned, we apply a scheme using hash sequences along with the repair symbols. The hash sequences represent at most a 10% of the total FEC data, so some very minor additional overhead must be taken into account in the configuration of the AL-FEC mechanism when compared to the results from above.

Another issue is the practical size of the source packets. Depending on the limitations of the hash mechanism employed, mainly the maximum number of source and repair packets per source block, the size of the source packets may have to be larger than 184 bytes. In this case, the loss of a single MPEG-2 TS packet results in a drop of more information in the source block. The simulation results have shown how the presence of isolated erroneous TS packets is more common at high Doppler values whereas it is less problematic at the low and medium range of Doppler.

An AL-FEC prototype for the protection of DVB-T services has been developed as a result of the collaboration between the Universidad Politécnica de Valencia and Digital Fountain. The prototype employs Raptor codes integrated into an IPTV streaming framework and operates with a source packet size of 1288 bytes and a repair packet size of 1472 bytes that corresponds exactly to a payload of 7 and 8 MPEG-2 TS packets, respectively. Figure 7 shows the performance obtained by the prototype. There is a degradation of 2 dB for the higher values of Doppler and less than 1 dB for the low and medium range of Doppler with respect to the results obtained in Figure 4. Note that this implementation is not yet fully optimized, but it shows the feasibility of the approach in a practical environment by reusing an IPTV streaming framework in the context. It is expected that future optimizations in the hash mechanism will allow AL-FEC to operate with smaller source and repair packets. Improved hash mechanisms specifically adapted to the integration in MPEG-2 TS packets are expected to move the results of real implementations even closer to the ideal performance shown in this paper.

The implementation of long protection periods is problematic in terms of channel switching time and memory requirements. As it has been explained, longer protection periods involve higher requirements of memory in the user

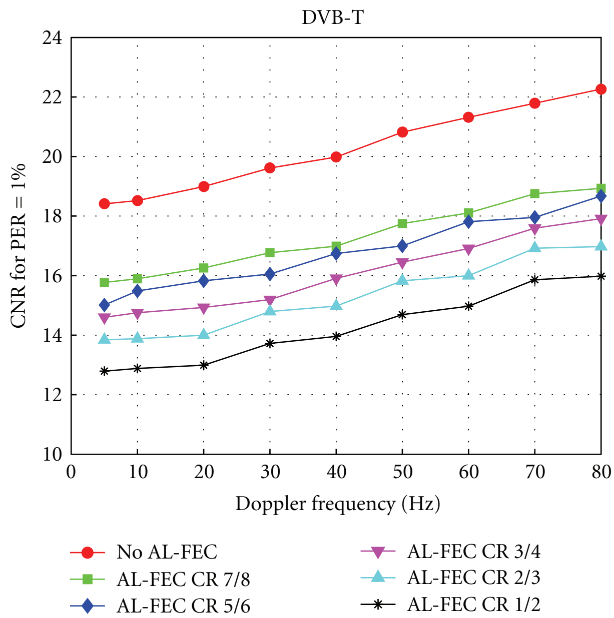


Figure 7: Mobile Performance of DVB-T services with a practical implementation of AL-FEC in a TU6 channel

terminals along with an increase in the channel switching time experimented by the user. Assuming a DVB-T typical service of 2.5 Mbps and a protection period of 10 seconds, a minimum memory capacity of 3 MB is required in order to store all the information of one source block. On the other hand, the use of computationally efficient erasure codes like Raptor avoids the implementation of dedicated hardware and allows the incorporation of AL-FEC as a software update.

Longer protection periods also have an important impact in channel switching time. Long channel switching times can degrade the user experience, and must be taken into consideration. Values of less than 0.5 seconds are not perceived by the user whereas values up to 2 seconds are considered to be tolerable [11]. As we have seen in the simulations, values of 1 second are capable to cope with fast fading and can even bring important gains in high diversity scenarios. Fast channel switching techniques are a hot topic today among mobile TV research as they can reduce the channel switching time perceived by the user with a small impact in the protection period. Although they are generally applied to bursty transmissions like DVB-H, they may also be incorporated in DVB-T services in order to improve the user experience when using long protection periods.

6.2. Deployment in Existing Systems. The available bandwidth for FEC data in DVB-T transmissions is also an important issue for the utilization of AL-FEC in existing DVB-T systems. The maximum amount of FEC data that can be multiplexed in an MPEG-2 TS depends on the available bandwidth and is limited. Existing DVB-T networks were not planned with mobile reception in mind and no specific bandwidth is assigned for the carriage of FEC data. Despite of this, it is possible to multiplex a limited amount of FEC data by means of the null TS packets generally present in an MPEG-2 TS. DVB-T transmitters insert null TS packets filled with stuffing data in the MPEG-2 TS in order to maintain a constant bit rate at the physical layer. We have performed preliminary studies in the MPEG-2 TS used in actual German and Spanish DVB-T transmissions to estimate the amount of null TS packets. The results reveal that depending on the multiplex, a percentage between 2% and 11% of the MPEG-2 TS corresponds to null TS packets that can be replaced by TS packets carrying FEC data.

DTT networks generally transmit several TV programs per MPEG-2 TS. Because of the limitation in the amount of available bandwidth to accommodate the FEC data, it may be desired to protect only few of the programs transmitted in an MPEG-2 TS. It is also possible to encode only a set of the video frames in order to increase the efficiency of the FEC data. Mobile reception in handheld terminals is generally performed with small displays that do not require the high frame rates of DVB-T services. The protection of Intra (I) and Predicted (P) frames only can increase the efficiency of the FEC data in a 33% without major penalties in the user experience (B frames usually represent 1/3 of the total amount of video frames). The arrangement of the video frames in each protection period can also affect the performance of the AL-FEC mechanism. Video frames in MPEG-2 video are grouped in Group of Pictures (GoP). By encoding an entire number of GoPs in each protection period, the errors do not propagate between protection periods, increasing the user experience for the same amount of uncorrected errors. The use of new video encoding standards such as H.264/AVC instead of the legacy MPEG-2 video coding, may also provide additional spare bandwidth for the accommodation of FEC data. By means of emerging video encoders and coding standards, it is possible to reduce the bit rate of current services without degrading the quality of the video experienced by the user.

7. Network Planning Discussion

Our results have shown that the AL-FEC protection of DVB-T services is capable to obtain gains of 6 dB in mobile channels. However, other considerations apart from the fast fading must be taken into account in the planning of DVB-T networks for the provision of mobile TV.

While fixed reception is generally performed with high gain antennas located in the roof of the buildings, mobile reception is characterized by the reception at ground level and the use of low gain antennas. The added penalization of the height loss and the use of mobile antennas represent more than 20 dB if external antennas are used and more than 27 dB if integrated antennas are used instead [12]. This degradation has an important impact on the link budget of DVB-T systems, especially in urban scenarios. Therefore, the provision of mobile TV in DVB-T networks planned for fixed reception cannot be performed with coverage levels comparable to those of fixed DTT services. In this kind of networks, the AL-FEC protection of DVB-T services or the use of antenna diversity techniques can only be aimed to the provision of a best effort kind of service available in areas with the best coverage conditions.

On the other hand, DVB-T networks planned for portable reception like in Germany, take into account the penalization due to the height loss and lower gain antennas. However, in order to provide the necessary bit rate for DVB-T services, DVB-T networks normally operate with higher modulation orders and less robust physical layer code rates than mobile TV networks such as DVB-H. Additionally, portable reception in DVB-T networks is planned for a Rayleigh channel model, which is approximately between 5 and 10 dB less demanding than mobile channels like the TU6 channel model. The combined gain of AL-FEC and antenna diversity techniques can be used in DVB-T networks deployed for portable reception in order to provide mobile DVB-T services with a similar coverage area to that of fixed DVB-T services.

8. Conclusions

In this paper we have investigated the use of AL-FEC for the provision of mobile DVB-T services. AL-FEC can be implemented in a backward compatible way and can be used in existing networks and services to extend the mobile coverage of DVB-T services. AL-FEC protection can be used in conjunction with antenna diversity techniques and hierarchical modulations in order to further enhance the vehicular reception of DVB-T services in existing networks. At the same time, newly deployed networks can be planned for the simultaneously provision of fixed and mobile DVB-T services.

The protection provided by AL-FEC depends not only on the proportion of FEC data transmitted along with the service information but also on the protection period. Long protection periods take advantage of the temporal diversity derived from user mobility, and achieve a better protection in the presence of shadowing.

We have provided direct comparisons between the AL-FEC and MPE-FEC protection of DVB-T and DVB-H services in a TU6 channel. The results show that AL-FEC achieves an additional 4 dB gain with respect to MPE-FEC in the low range of Doppler. The additional gain at low Doppler values is motivated by the fact that AL-FEC can be configured with higher protection periods than MPE-FEC while having a little impact in the channel switching time. Long protection periods have been also investigated in TU6 channels affected by shadowing. The results show that the gain derived from long protection periods is heavily conditioned to the temporal diversity. A protection period of 10 seconds can provide a gain of 10 dB in high diversity scenarios. The use of longer protection periods involves higher memory requirements along with an increment in the channel switching time. The use of memory efficient decoding algorithms can solve the memory problems whereas fast channel switching techniques may decrease the channel switching time perceived by the user.

The implementation of AL-FEC in DVB-T presents a series of practical issues. The size of the source packets and the amount of additional information that must be transmitted have an important impact in the performance of the system. Current implementations of AL-FEC in DVB-T can loss about 2 dB with respect to the ideal implementation, especially at high velocities. The amount of bandwidth available in current DVB-T transmissions in order to accommodate FEC data is also an important issue. Several possibilities such as the protection of only a few services per MPEG-2 TS have been discussed in order to increase the efficiency of the transmitted FEC data.

References

[1] ETSI, EN 300 744 v1.5.1, "Framing structure, channel coding and modulation for digital terrestrial television," November 2004.

[2] G. Faria, J. A. Henriksson, E. Stare, and P. Talmola, "DVB-H: digital broadcast services to handheld devices," *Proceedings of the IEEE*, vol. 94, no. 1, pp. 194–209, 2006.

[3] "Celtic WingTV Project: Services to Wireless, Integrated, Nomadic, GPRS-UMTS & TV Handheld Terminals, WingTV Simulation Re-port, Deliverable 15b," December 2006.

[4] D. Gómez-Barquero, D. Gozálvez, and N. Cardona, "Application layer FEC for mobile TV delivery in IP datacast over DVB-H systems," *IEEE Transactions on Broadcasting*, vol. 55, no. 2, pp. 396–406, 2009.

[5] "Advantages of Diversity Implementations on Mobile and Portable TV Receivers," Dibcom Whitepaper, May 2008.

[6] A. Shokrollahi, "Raptor codes," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2251–2567, 2006.

[7] S. Saunders, *Antennas and Propagation for Wireless Communications*, John Wiley & Sons, New York, NY, USA, 2000.

[8] ETSI, DVB Bluebook A131, "MPE-iFEC," November 2008.

[9] ETSI, DVB Bluebook A054r4.1, "Interaction channel for Satellite Distribution Systems," January 2009.

[10] ITU-T Recommendation H.222.0, "Generic coding of moving pictures and associated audio information: systems," April 2006.

[11] H. Fuchs and N. Farber, "Optimizing channel change time in IPTV applications," in *Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB '08)*, pp. 1–8, Las Vegas, Nev, USA, 2008.

[12] ETSI, "Implementation guidelines for DVB ter-restrial services; transmission aspects," Tech. Rep. TR 101 190 v1.3.1, ETSI, Sophia Antipolis, France, November 2008.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 303627, 9 pages
doi:10.1155/2009/303627

Research Article

The Implementation of a 2/4/8 Antennas Configurable Diversity OFDM Receiver for Mobile HDTV Application

Jing Gao,[1] Takashi Iida,[2] Hiroyuki Mizutani,[3] Sadanori Sakaguchi,[3] Shuji Murakami,[3] Atsushi Shimizu,[2] and Tomohisa Wada[1,3]

[1] *Graduate School of Engineering and Science, University of the Ryukyus, Okinawa 903-0213, Japan*
[2] *Automotive Information Systems Department, SANYO Electric Co., Ltd., Osaka 570-8677, Japan*
[3] *Development-I, Magna Design Net, Inc, Okinawa 901-0155, Japan*

Correspondence should be addressed to Jing Gao, gaojing722@yahoo.com

Received 1 February 2009; Accepted 7 October 2009

Recommended by Robert Briskman

Two pre-FFT adaptive array (AA) antenna combiners and a post-FFT carrier diversity (CD) combiner are integrated with a Japan Terrestrial digital TV (ISDB-T) OFDM receiver using 90 nm 7M1P CMOS process. A 2/4/8-antenna diversity receiver can be configured and a low-cost 4 antenna diversity reception system can be realized in one LSI by making use of the AA-CD two-stage diversity combining method. Mobile reception performance is increased by 1.63 times using a denoise filter circuit and SPLINE interpolator under urban 6-path Rayleigh fading (TU6) model with 2-antenna post-FFT carrier diversity (2CD) combing mode. The die area is 49 mm^2 and the power consumption is 310 mW.

1. Introduction

TV broadcasting technology is rapidly shifting to the digital domain and the services are expanding not only to home TV but also to ubiquitous devices such as cellar phones, PCs, and automotive TVs. Orthogonal Frequency Division Multiplexing (OFDM) is adopted as a modulation method for the terrestrial Integrated Services Digital Broadcasting(ISDB-T) standard in Japan. OFDM is wellknown as a high-spectral efficiency transmission method in a multipath environment [1].

When an OFDM receiver is used in an automobile, the Radio Frequency (RF) signal experiences a Doppler frequency shift as the automobile is moving as shown in Figure 1. The Doppler shift destroys the orthogonality between OFDM subcarrier signals and increases intercarrier interference (ICI). Therefore, it is a severe challenge to maintain the reception quality for a mobile ISDB-T receiver that is acceptable for human vision.

One well-known way to improve the performance of an OFDM receiver is to exploit spatial diversity by utilizing multiple antenna elements. In this paper, a mobile OFDM receiver LSI is described which includes two pre-FFT adaptive array (AA) antenna combiners and a post-FFT carrier diversity (CD) combiner. Using the LSI, several combinations of AA and CD diversity systems can be used to tradeoff performance of the mobile ISDB-T receiver with system cost. Additionally, directional antenna elements can be utilized to mitigate the Doppler shift. The performance of the proposed AA-CD scheme is evaluated by simulation and is validated in laboratory and field experiments.

The rest of the paper is organized as follows. In Section 2, diversity receiver fundamentals are explained. The LSI chip architecture and circuits will be described in Sections 3 and 4, respectively. In Sections 5 and 6, simulation results and measurement results including field experiments will be shown. Finally, conclusions are given in Section 7.

2. Diversity Receiver Fundamentals

2.1. Use of Directional Antenna. Doppler shift (f_d) is given as

$$\begin{aligned} f_d &= \frac{v}{c} f_c \cos(\theta) \\ &= f_{d\max} \cos(\theta), \end{aligned} \tag{1}$$

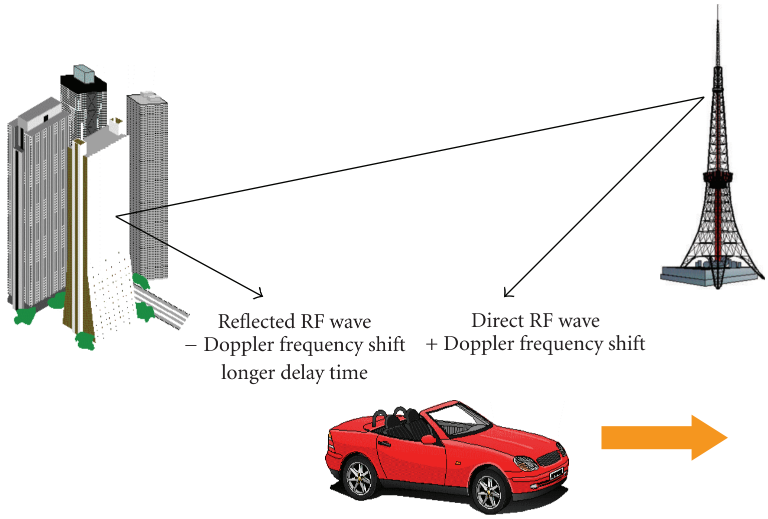


Figure 1: Problems in mobile reception.

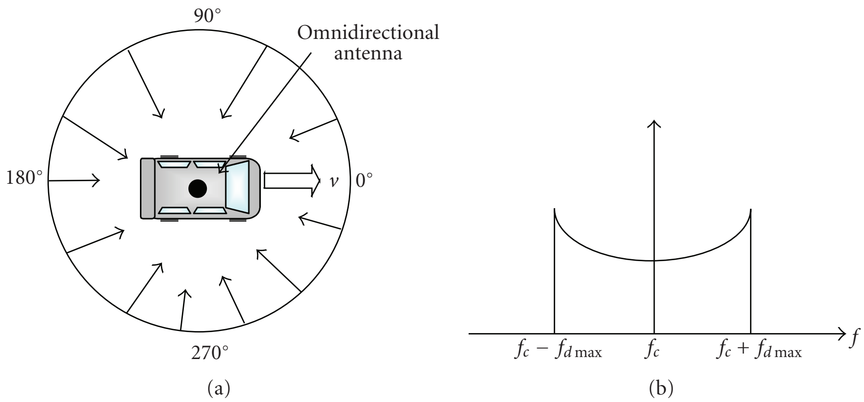


Figure 2: Doppler frequency shift for omnidirectional antenna. (a): In coming RF waves, (b): Doppler frequency distribution.

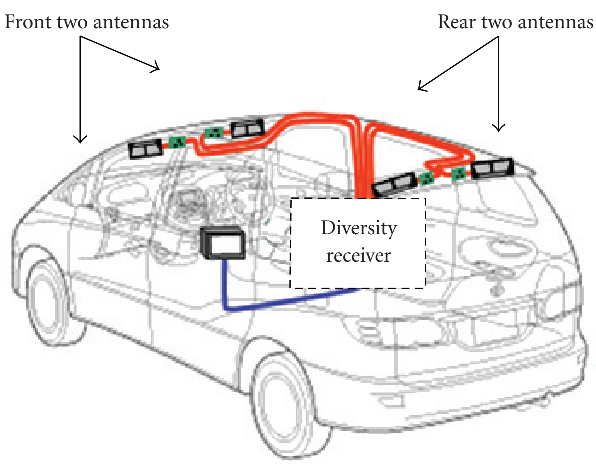


Figure 3: Antennas on vehicle.

where v and c are speed of the automobile and light, respectively, f_c is the carrier frequency, and θ is the Angle of Arrival (AOA) of the incoming signal. The maximum Doppler shift is determined as $f_{d\max} = (v/c)f_c$. When an automobile has one omnidirectional antenna on the roof as shown in Figure 2(a), the antenna receives all directional incoming RF waves, which is uniformly distributed within $[0, 2\pi]$. The spectrum of the received signal is distributed over $[f_c - f_{d\max}, f_c + f_{d\max}]$ as shown in Figure 2(b).

Figure 3 shows a typical 4 space diversity antenna configuration for an automobile. Two antenna elements are set in the front and the two are set in the rear. However as the body of the automobile is made from metal, the two front antenna elements seemingly experience a similar Doppler shift, as do the two in the rear. It has been experimentally confirmed that the directional characteristic of the front antenna is obviously distorted and concentrates in the forward direction, while the rear antenna focuses to the rear direction of automobile [2]. Consequently, as illustrated in Figure 4, the received signal from the front antenna element predominantly experiences a positive Doppler shift randomly distributed within $[f_c, f_c + f_{d\max}]$ while that of rear antenna element predominantly experiences a negative Doppler shift randomly distributed

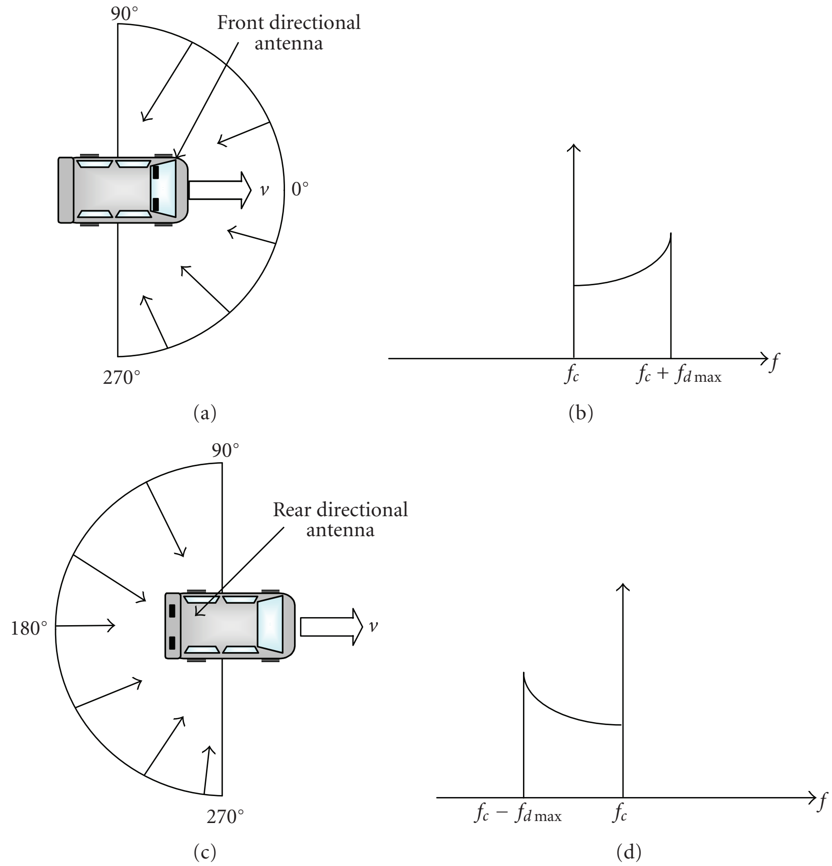


Figure 4: Doppler frequency shift for front and rear directional antenna: (a) and (b) front directional antenna; (c) and (d) rear directional antenna.

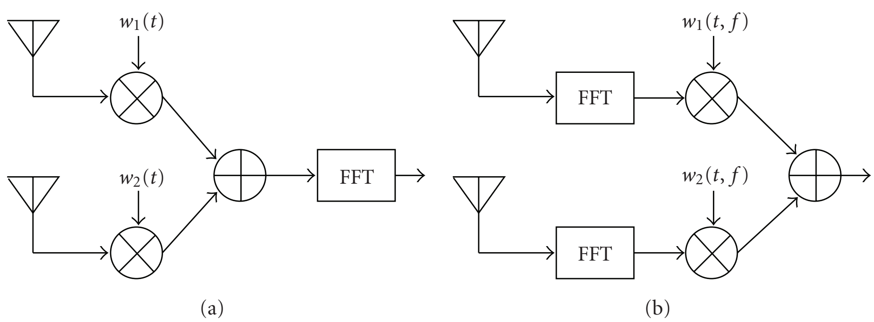


Figure 5: Two diversity combining methods. (a): pre-FFT adaptive array (AA) antenna, (b): post-FFT carrier diversity (CD).

within $[f_c - f_{d\max}, f_c]$. In other words, the performance of a mobile ISDB-T receiver is seemingly enhanced as the bandwidth of the Doppler spectrum is reduced by exploiting directional antenna elements.

2.2. Two Types of Diversity Combining Methods. There are two approaches to utilize the array antenna for multicarrier transmission: pre-FFT adaptive array (AA) antenna and post-FFT carrier diversity (CD) combining.

As shown in Figure 5(a), the AA scheme is a conventional method to employ the array antenna in which inputs from the array antenna are combined before OFDM demodulation [3–5]. Since this approach uses one set of coefficients such as $w_i(t)$ $(i = 1, 2)$ for each OFDM symbol, it is an attractive solution due to low computation complexity. However, for good performance, received antenna signals should have high correlation.

On the other hand, the post-FFT scheme is an advanced method to utilize the array antenna for a multi-carrier system as shown in Figure 5(b). Instead of combining before OFDM demodulation, inputs are demodulated using multiple OFDM demodulations. Subcarriers are then combined accordingly in the frequency domain using the diversity

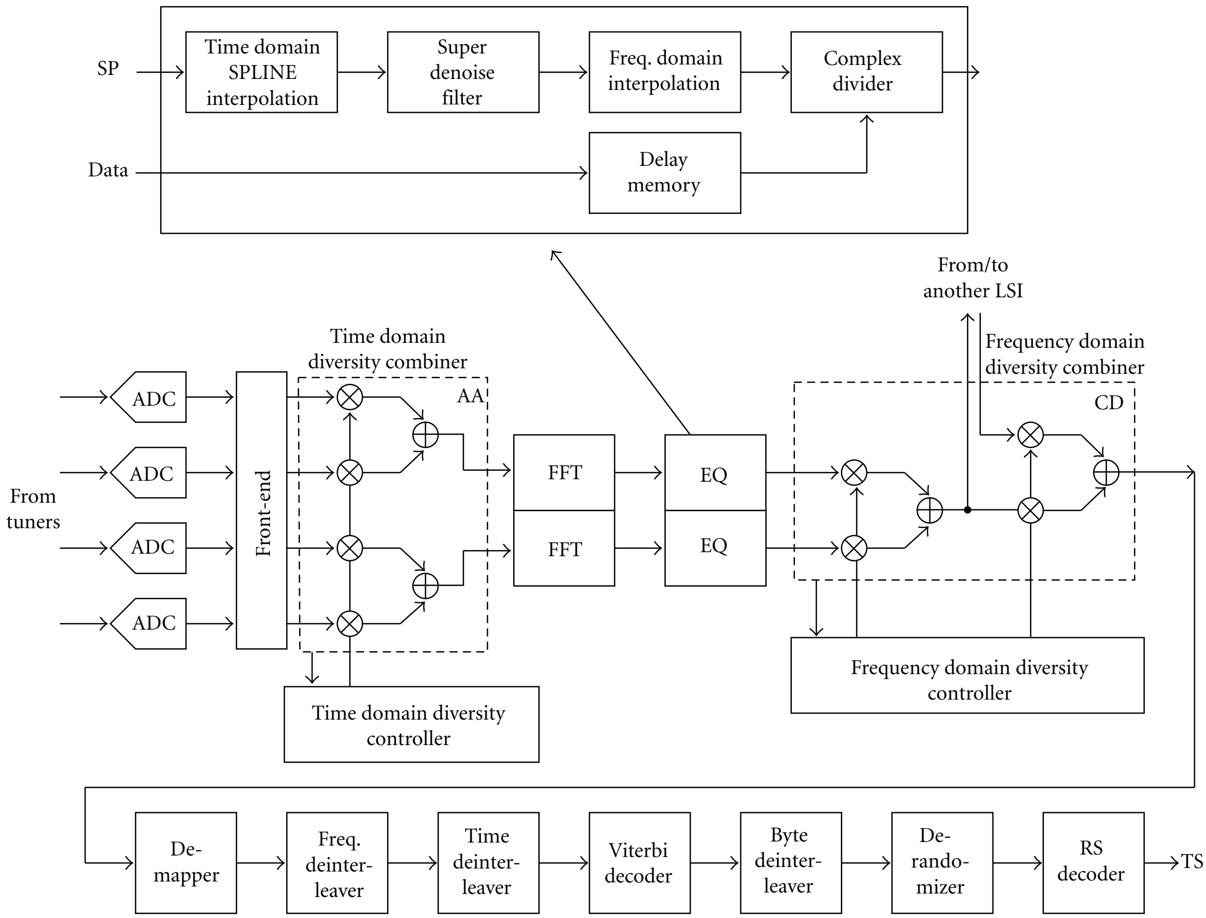


Figure 6: Detailed block diagram of receiver.

Table 1: Features of two diversity methods.

Type	Circuit	Performance	Restriction
Adaptive Array (AA)	Small	Moderate	High antenna signal correlation
Carrier Diversity	Large	High	None

combining method [6]. The Post-FFT scheme therefore is an optimum approach in term of maximizing signal-to-noise ratio (SNR) for each OFDM subcarrier. However, its computation complexity grows drastically as the number of antenna elements increases. The characteristics of the two diversity methods are summarized in Table 1.

3. Chip Architecture

Figure 6 shows the detailed block diagram of the receiver LSI. The UHF wireless signal is down-converted by tuners. A maximum of 4 tuner output signals are fed into one LSI and are converted to digital signals by ADCs. The first AA stage performs the array antenna signal processing and the second CD stage combines each OFDM subcarrier signal at the output of the Fast Fourier Transforms (FFT)/equalizers (EQ) processing. The two-stage diversity combiners halve the number of FFT/EQ circuits from 4 to 2 but maintain equivalent mobile reception performance compared to the 4 FFT/EQs case. The number of logic gates is reduced by 39% through the using of the two-stage architecture. To improve higher mobile performance, the equalizer introduces two new circuit blocks, the SPLINE interpolator and the super denoise filter, which will be explained in the following section.

4 types of diversity configurations are possible as shown in Figure 7. With each doubling of the number of antennas, SNR is improved by approximately 3 dB. For the 4-antenna case, there are two possible configurations. The 4AA mode is a one-chip solution that achieves the lowest cost and modest performance. The 4CD mode is a two-chip solution with higher performance. Since the antenna placement on a car is usually on the front or rear windshield glass, each antenna has directional gain fluctuation due to the car body. The 4-antenna system is preferable for compact to medium size cars to cover all signal angles of arrival.

4. Circuits

4.1. SPLINE Time Domain Interpolator. Figure 8 shows the newly introduced SPLINE interpolator. The subcarrier signals (see left) of FFT output are divided into two parts,

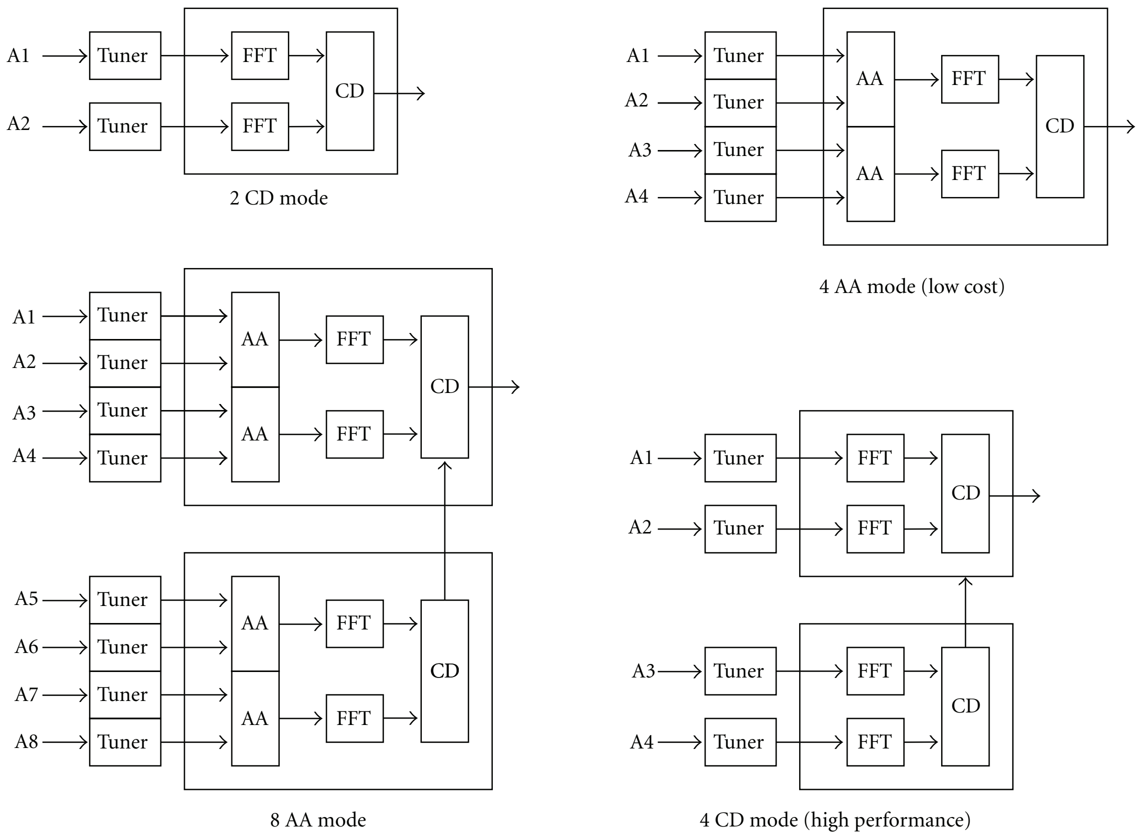


Figure 7: 2/4/8 antennas system configurations.

Table 2: Coefficients of 4-tap SPLINE FIR filter.

	C1	C2	C3	C4
I1	1/64	−6/64	25/64	44/64
I2	1/40	−6/40	29/40	16/40
I3	7/320	−42/320	303/320	52/320

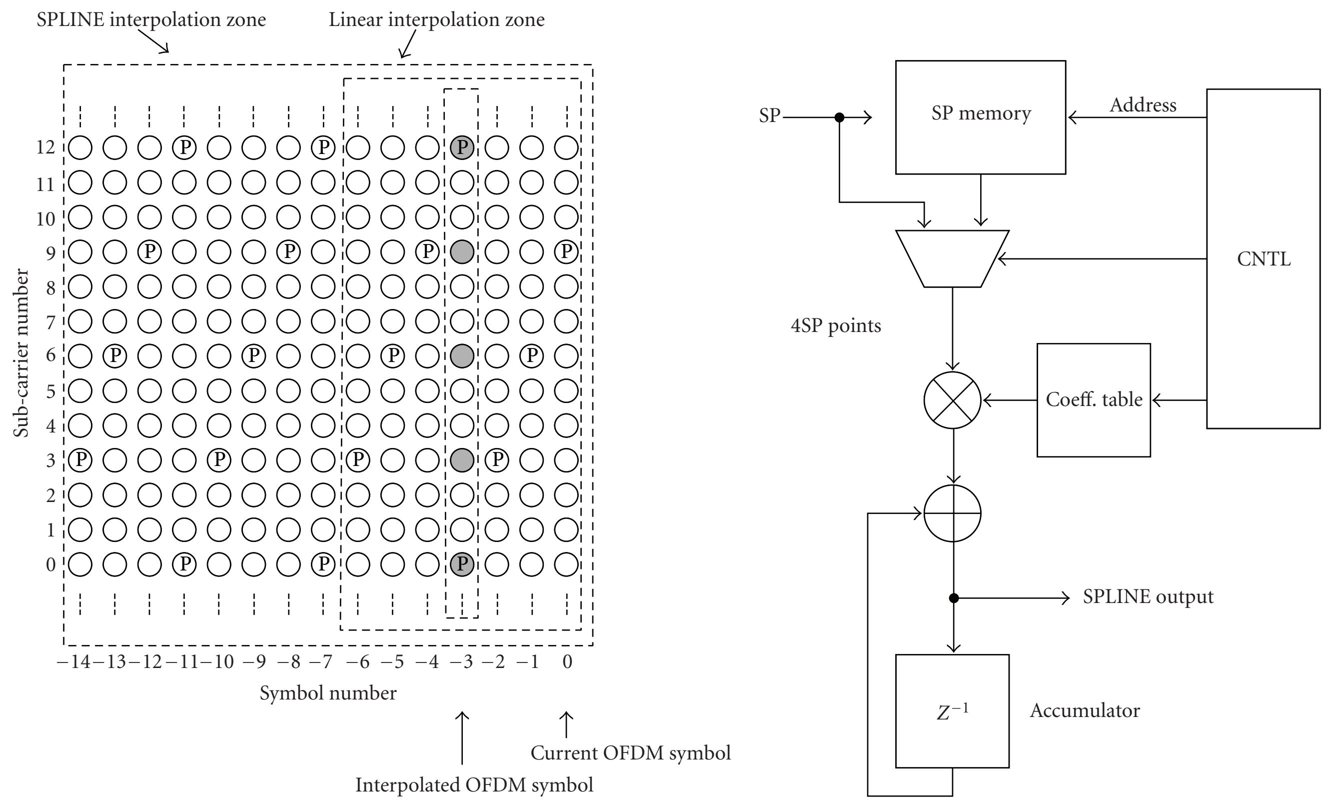
data (white circle) and scattered pilots (SPs, circle with P). Those SPs are located every 12 subcarriers in one OFDM symbol and the position of the SP shifts by 3 subcarriers every OFDM symbol. In previous work [7], 2-tap linear interpolation is performed by the two SPs in the linear interpolation zone. In this paper, the new SPLINE algorithm 4-tap interpolation is applied. Although the SPLINE uses two more past SPs in the SPLINE interpolation zone (total 4 taps), the 3 symbols of filter latency (see symbol number $= -3$) is the same as the previous work. Then the higher interpolation accuracy is obtained without increasing the delay RAM size in the equalizer.

The 4-tap FIR filter is implemented using a poly-phase accumulator. Figure 9 shows a logical view of the 4-tap poly-phase FIR filter. The position of the I1, I2, and I3 points can be computed by changing the filter coefficients of C1, C2, C3, and C4. Those coefficients are computed using the mathematical SPLINE function. Actual coefficients of the 4-tap SPLINE FIR filter are shown in Table 2.

4.2. Variable Bandwidth Super Denoise Filter. To reduce noise on the interpolated SP signals further, a variable bandwidth complex bandpass filter is introduced. The bandwidth and the center frequency are controlled by a delay-profile of the reception signal. To support fine control of the center frequency and a broad range of filter bandwidth, a two-stage serially connected circuit is used as shown in Figure 10. Both stages are identical 41-tap raised cosine Lowpass Filters (LPF). The bandpass position of each stage can be shifted using the complex rotation circuit, that is, complex multiplier at each input/output. Through the combination of the two stages, the complex filter bandwidth and center frequency can be controlled. The upper section of Figure 10 shows the example of a 1/4 bandwidth LPF. The shaded area is the complex bandpass filter bandwidth.

5. Simulation Results

In this section, the 4 operation modes (shown in Figure 7 above) 2CD, 4AA, 4CD, and 8AA are verified by computer simulation. To evaluate performance, directional characteristics of antenna elements are modeled based on [8]. Table 3 shows the system parameters. Mode 3 of the ISDB-T standard using 64QAM modulation and Guard Interval (GI) duration $T_g = T_e/8$ is used for the simulation. Bit Error Rate (BER) is measured without Error correction. Table 4 shows the specification of the simulation parameters.



(a) (b)

Figure 8: SPLINE time domain interpolator.

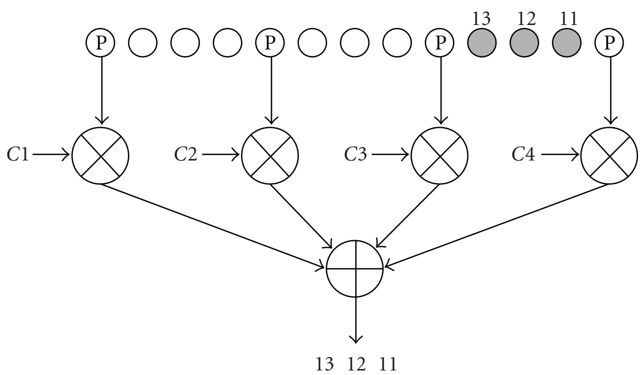


Figure 9: 4-tap FIR filter for SPLINE interpolation.

Table 3: OFDM system parameters.

System Parameters	
Carrier Frequency	563.143 MHZ
FFT size	8192
Number of used subcarrier	5167
Effective Symbol Duration	Te = 1008 us
Guard Interval Duration	Tg = Te/8
Digital Modulation	64 QAM

Table 4: 6-path simulation parameters.

Path	Angle (deg)	D/U (dB)	Delay
1	0	5	0.5∗(Tg/8)
2	30	0	0
3	90	8	3.2∗(Tg/8)
4	150	6	2.5∗(Tg/8)
5	210	1	0.3∗(Tg/8)
6	300	7	4.0∗(Tg/8)

Table 5: Summary of die.

Process Technology	90 nm 7M1P CMOS
Logic	1.8 M gates
Memory and ADC	18.4 M bit/4 ADC
Supply Voltage	1.2 V core, 3.3 V I/O
Active Power	310 mW typical
Die Size	7.0 mm × 7.0 mm
Package	12 mm × 12 mm 144FBGA

6 signals are received with different power and delay time. Since signal 2 at 30 degrees is the strongest, it is considered the desired signal while the others are considered undesired signals. The DU ratio (D/U) is the Desired to Undesired signal power ratio. Signal 1 at 0 degree is 5 dB lower power than 2.

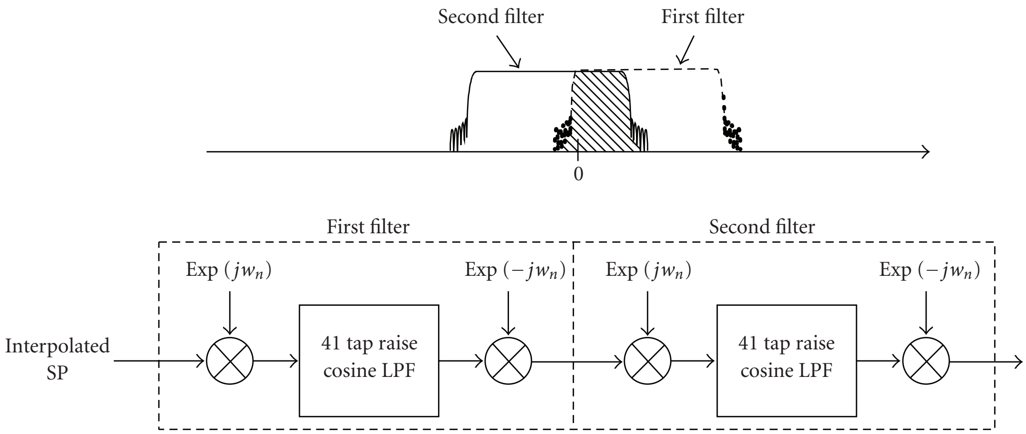


Figure 10: Variable bandwidth super denoise filter.

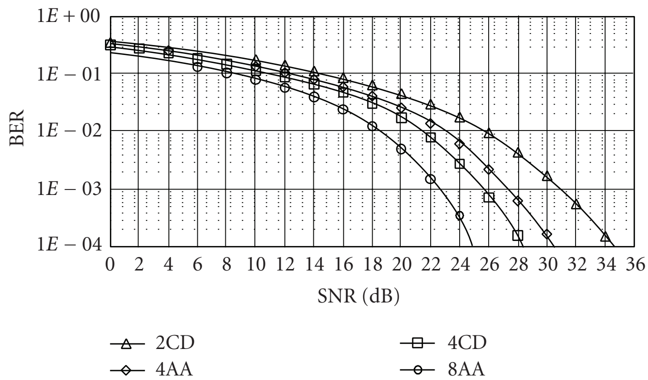


Figure 11: BER versus SNR with Doppler shift of 0 Hz.

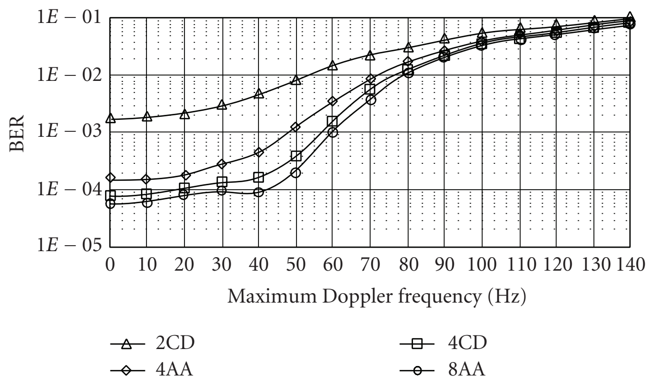


Figure 12: BER versus Maximum Doppler shift with SNR = 30 dB.

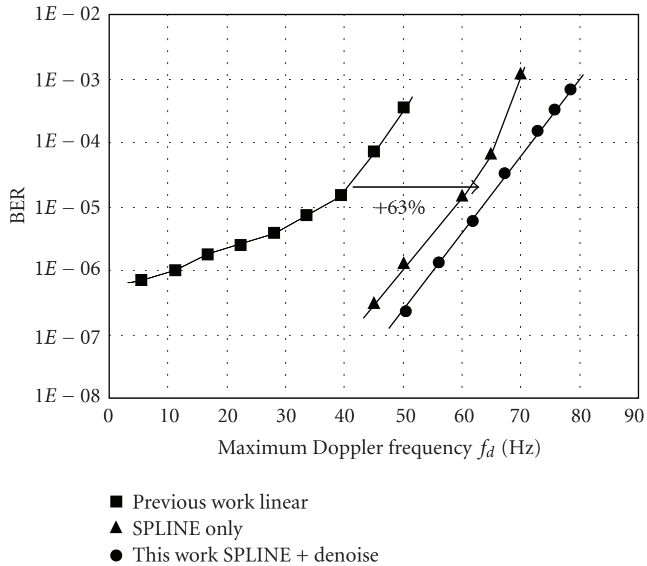


Figure 13: Mobile reception performance comparison of 2CD mode.

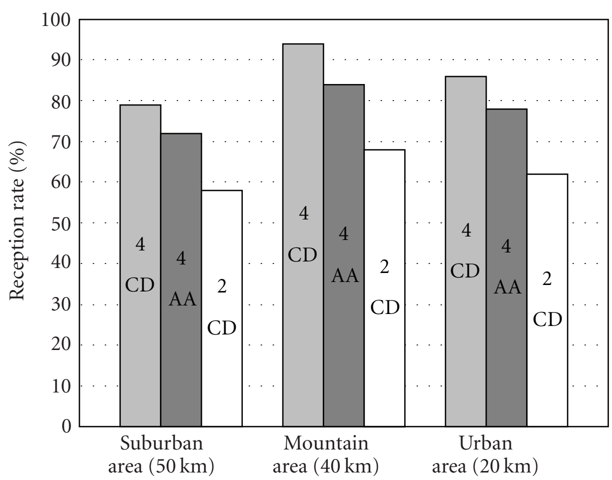


Figure 14: Field experimental result.

Figure 11 shows BER performances versus SNR with Doppler shift = 0 Hz. BER performance improves with increasing number of antenna elements. 4CD shows slightly better performance than 4AA. In addition, BER performances versus Maximum Doppler shift with SNR of 30 dB are shown in Figure 12. Moving from 2CD to the other modes, there are significant Doppler performance improvements. 2CD only uses 2 front antennas and the other modes

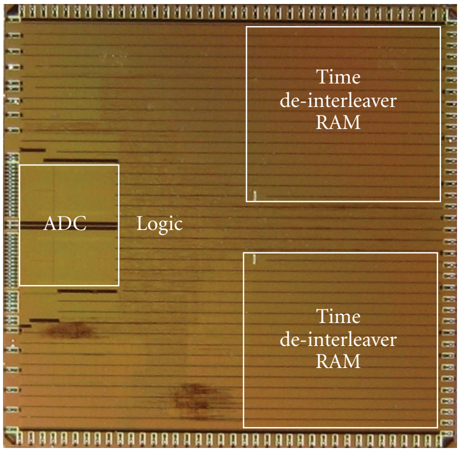


Figure 15: Die micrograph.

Figure 16: 2CD mode mobile TV reception demonstration.

use both front and rear antennas. 8AA has 4 front antennas and 4 rear antennas.

6. Measurement Results

Figure 13 shows the measurement results with typical urban 6 path Rayleigh fading (TU6) in 2CD mode. Thanks to the newly introduced circuits, the maximum Doppler performance is improved from 41 Hz to 67 Hz (1.63 times). The 1.63 times improvement corresponds to the moving speed improvement from 73 km/h to 120 km/h at UHF35ch (605.143 MHz).

Mobile reception performance was measured in field experiments at three severe test courses in Osaka Japan. The experimental UHF channel is 13 which is broadcast from the top of Ikoma Mountain between Osaka and Nara prefecture. 64QAM, code rate of 3/4, 12-segment HDTV broadcasting service is used. The three test courses are a suburban area 50 km from Ikoma broadcasting station (BS), a mountain area 40 km from the BS, and an urban area in the other side of a hill 20 km from the BS. Figure 14 shows the reception rate (error free reception duration to the total experiment duration) of those test courses. The one-chip solution 4AA mode showed close performance to the two-chip solution 4CD mode.

Figure 15 shows the die micrograph and the die features are summarized in Table 5. Two pre-FFT adaptive array (AA) antennas and a post-FFT carrier diversity (CD) combiner are integrated with an ISDB-T OFDM receiver in 90 nm 7M1P CMOS process occupying 49 mm^2 and dissipating 310 mW. The two stage diversity combiners result in a logic reduction of 39% over a conventional 4 FFT/EQs carrier diversity system. The total logic size and the memory size are 18 M gates and 18.4 M bits, respectively. The package is a 12 mm square 144 pin fine pitch ball grid array.

7. Conclusion

The LSI integrated two pre-FFT Adaptive Array (AA) and one post-FFT carrier diversity (CD) combiners with an ISDB-T compatible OFDM receiver in 90 nm 7M1P CMOS. A 2/4/8-antenna diversity receiver can be configured. Two stage diversity methods reduce logic circuits by 39% over the 4FFT CD configuration. SPLINE interpolator and Super denoise filter circuits improve the Doppler performance by 1.63 times. This corresponds to a 73 km/h to 120 km/h increase at UHF 35ch in 2CD mode. The die area is 49 mm^2 and the power consumption is 310 mW. The LSI was successfully utilized in a real mobile TV system product in a Japan ISDB-T application. Figure 16 shows a 2CD mode mobile TV reception demonstration using the product.

Acknowledgments

The authors thank Takuya Koujiya for the field experiment support, and Kenichi Satou, Atsushi Suyama for the tuner device support, and Kunio Morimoto for module design.

References

[1] J. A. C. Bingham, “Multicarrier modulation for data transmission: an idea whose time has come,” *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, 1990.

[2] S.-I. Matsuzawa, K. Sato, and K. Nishikawa, “Study of on-glass mobile antennas for digital terrestrial television,” *IEICE Transactions on Communications*, vol. E88-B, no. 7, pp. 3094–3096, 2005.

[3] S. Sakaguchi, M. Hori, H. Asato, et al., “An adaptive array direction control LSI for mobile digital HDTV receivers,” in *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE ’05)*, pp. 299–300, Las Vegas, Nev, USA, January 2005, Session: 7.4 Acquisition Systems 7.4-3.

[4] H. Asato, T. Tabata, D. H. Pham, et al., “A software-configurable adaptive array antenna system for ISDB-T reception,” in *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE ’06)*, pp. 463–464, Las Vegas, Nev, USA, January 2006, Session: 11.2 Signal Processing for Multiple Ant., DVB-T and DAPSK/DMT 11.2-1.

[5] D. H. Pham, T. Tabata, H. Asato, S. Hori, and T. Wada, “Joint hardware-software implementation of adaptive array antenna for ISDB-T reception,” *IEICE Transactions on Communications*, vol. E89-B, no. 12, pp. 3215–3224, 2006.

[6] S. Nakahara, H. Hamazumi, K. Shibuya, and M. Sasaki, "An application of diversity combination techniques to broadcasting wave relay station for ISDB-T," *ITE Technical Report*, vol. 25, no. 31, pp. 7–12, 2001.

[7] K. Kamiyama, H. Mizutani, H. Toma, et al., "A high-speed mobile OFDM receiver LSI with an iterative noise-reduction filter to enhance channel estimation," in *Proceedings of the IEEE International Conference on Consumer Electronics (ICCE '05)*, pp. 193–194, Las Vegas, Nev, USA, January 2005, Session: 5.1 Digital TV Broadcast: DVB & ISDB-T 5.1-3.

[8] M. Fujimoto and T. Hori, "Effect of antenna element characteristics on antenna pattern control of array antenna," Tech. Rep. RCS2004-177, IEICE, October 2004.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 263936, 8 pages
doi:10.1155/2009/263936

Research Article

Theoretical Models for Video on Demand Services on Peer-to-Peer Networks

Giovanni Branca,[1] Thomas Schierl,[2] and Luigi Atzori[1]

[1] *Department of Electrical and Electronic Engineering, University of Cagliari, 09123 Cagliari, Italy*
[2] *Fraunhofer Institute for Telecommunications, Heinrich Hertz Institut, 10587 Berlin, Germany*

Correspondence should be addressed to Giovanni Branca, giovanni.branca@diee.unica.it

Received 3 March 2009; Accepted 20 July 2009

Recommended by Maurizio Murroni

Peer-to-peer networks (P2Ps) are becoming more and more popular in video content delivery services, such as TV broadcast and Video on Demand (VoD), thanks to their scalability feature. Such characteristic allows for higher numbers of simultaneous users at a given server load and bandwidth with respect to alternative solutions. However, great efforts are still required to study and design reliable and QoS-guaranteed solutions. In this paper, within the scenario of P2P-based VoD services, we study the phenomenon of peer churns and propose four models of the peer behaviour to evaluate its impact on the system performance, which are based on the Gilbert-Elliot chain, the fluidic representation of the user behavior, and a queuing analysis of the system. The models are compared by computing the resources the system has to add on top of the P2P network to satisfy all the download requests. Simulations show important relationships between playback buffer length, peer request rate, peer average lifetime, and server upload rate.

1. Introduction

Last years have been characterized by an exponential growth of video traffic on the Internet, which has brought to significant investments in networks and systems aimed at the delivery of real-time high-rate streams. Several traffic analyses tell us that this growth will continue over the next decade, making video streaming applications the ones driving the Internet evolution during the near future [1]. Video on Demand (VoD) is one of these applications, which requires resources able to deliver a video whenever the customer requests it. Realizing a VoD system using the Internet requires architectures tailored to video characteristics. Even if advanced video coding technologies such as Scalable Video Coding (SVC) [2] allow for an efficient representation of the video content towards the transmission over packet networks, VoD service requires guaranteed bandwidths and constrained transmission delays that make it quite difficult to be provided in the traditional Internet architecture.

Typical VoD solutions can be grouped into four categories [3]: centralized, proxy-based, Content Delivery Network (CDN), and Hybrid architectures. In a centralized architecture, the source server manages all clients: it is the simplest and easiest way to implement a VoD system. This solution has the big disadvantages of having a single point of failure, requiring servers with high computational and transmission capabilities that generate unbalanced network loads. Proxy-based architectures are aimed at decreasing the central server load, introducing proxy-servers in strategic points of the network, typically close to the clients. CDNs can be seen as an extension of the proxy-based approach. Accordingly, the video requests are completely handled by edge servers, streaming the content directly to the clients. No requests are forwarded to the central server, as it instead happens in the proxy-based approach whenever the proxy does not have a copy of the requested content. Even if more robust than the centralized solution, major disadvantages limit the diffusion of the proxy-based and CDN approaches. The former *translates* a single point of failure into many points of failure, fractioning central server load to more servers. The latter may ensure high-quality services but it requires big investments for both network and servers deployment and management. Additionally, all these systems have scalability problems; that is, when the number of clients increases, the only way to satisfy all the incoming requests is to add new servers proportionally.

Hybrid architectures combine the employment of a centralized server with that of a peer-to-peer (P2P) network. Indeed, P2P technologies have been adopted for the deployment of important applications over the Internet, such as file sharing [4] and voice-over-IP (VoIP) [5]. Differently from file sharing, a P2P-VoD network must guarantee the video delivery to the end-user before rigid deadlines. In P2P-VoD, peers support the delivery of the video to other peers using a cache-and-relay strategy making use of their upload bandwidth so as to decrease server load and to avoid network congestions close to the server site. Advantages are a better use of resources and an increased system capacity that allow for the management of higher number of users. P2P networks are also used to realize video broadcast/multicast over the Internet [6]. This technology is attractive because the P2P paradigm has the intrinsic potential to scale with the number of active participants without requiring additional infrastructure deployments: a greater demand generates more resources.

In a peer-to-peer network each peer is free to join and leave the network without notice, bringing to the phenomena of peer churns. These peer dynamics are dangerous for VoD architectures, affecting the integrity and retainability of the service. In the past, many studies have addressed peer churns in file-sharing networks [7, 8], and some others focus on P2P-VoD systems proposing different techniques to avoid service disruption due to peer churns [9–11]. Differently from these works, this paper does not propose any new solution but analyses the user behaviour so as to develop models aimed at evaluating the impact of the peer churns on the system performance. Four models are then proposed. The first two rely on the Gilbert-Elliot model to represent the user connected and disconnected states; the third one is based on a fluidic analysis of the system; the last one makes use of the queuing theory to represent how the video chunk download requests are processed by the system. The models are compared by computing the resource that system has to add on top of the P2P network to satisfy all the download requests. The importance of an accurate modelling of the churns lies on the possibility to analyse important relationships between system parameters, such as playback buffer length, peer request rate, peer average lifetime, and server upload rate, which can then be used to drive the dimensioning and optimization of system resources while assuring user satisfaction.

The paper is organized as follows. Section 2 illustrates a common peer-to-peer Video on Demand scenario, which represents the basis of our analysis. In Section 3, the proposed theoretical models are described and in Section 4 numerical analyses are presented. Section 5 draws final conclusions.

2. The P2P-VoD Scenario

In a typical P2P-VoD scenario a centralized server receives video requests whereas a number of peers download and upload the same content. This is referred to as a single-video approach, and it differs from the multiple video approach because one peer can share only a video, which is the one it is playing back [12]. In case that all requested content cannot be provided by the peers, the server also streams the content accordingly.

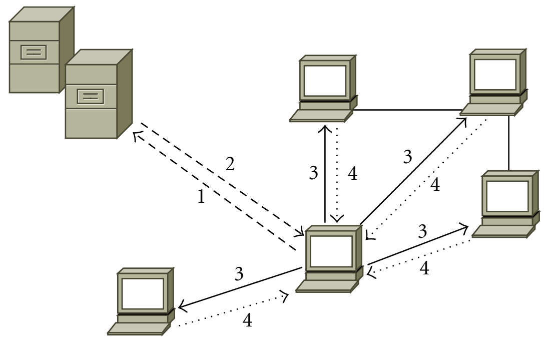


Figure 1: P2P-VOD system: 1: contact server; 2: receive peer list; 3: create application level connections; 4: send data.

Data and control information exchanges can be summarized in few steps. When a new peer joins the system, it contacts the server to know the available video contents. It chooses the video it is interested in and the server sends a list of possible peers that are viewing the chosen content; the peer then tries to create the necessary number of unicast connections with other peers to receive the content and start playing back. When a contacted peer had accepted a connection request, it starts to send useful data. This procedure is illustrated in Figure 1.

Each peer has a playback buffer used to decouple network dynamics from video playing. If a contacted peer does not have the requested data at that moment or it does not reply to the contacting peer, the latter starts creating another connection with the next peer according to the list provided by the server. The server takes charge of distributing a refreshed peer list to all peers whenever necessary, assuming a central role in the coordination of the VoD service.

The most critical problem in a P2P-VoD network is related to the dynamics of peer's participation. In a pure file-sharing network, it is not a serious problem: there are no deadlines to be respected, and it may not be a vital matter if the file download takes more than the expected or desired time. Instead, in the scenario of streaming applications we are considering, peer churns become an important issue which needs to be taken into account to make the system reliable enough to provide an acceptable QoS to the end-user.

The video content is divided in a sequence of video units, named chunks. To avoid playback interruptions, a peer must receive the correct sequence of these chunks before its playback deadline. Not to waste bandwidth, each peer can request only one chunk at time to one peer. We assume that each chunk is of the same transmission length T_{UT} (time to complete the transmission) and of the same playback length T_{UP} (time to finish the playback), both expressed in seconds. Typically T_{UT} is greater than T_{UP}, requiring more than one upstream peer (roughly T_{UT}/T_{UP} peers) per downloading peer on average to have a continuous playback of the video

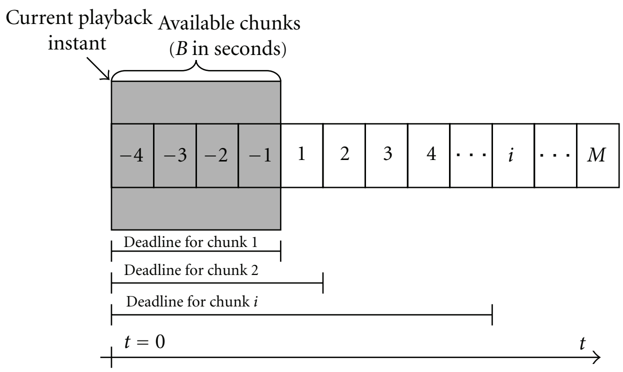


Figure 2: Playback buffer, video chunks, and associated deadlines.

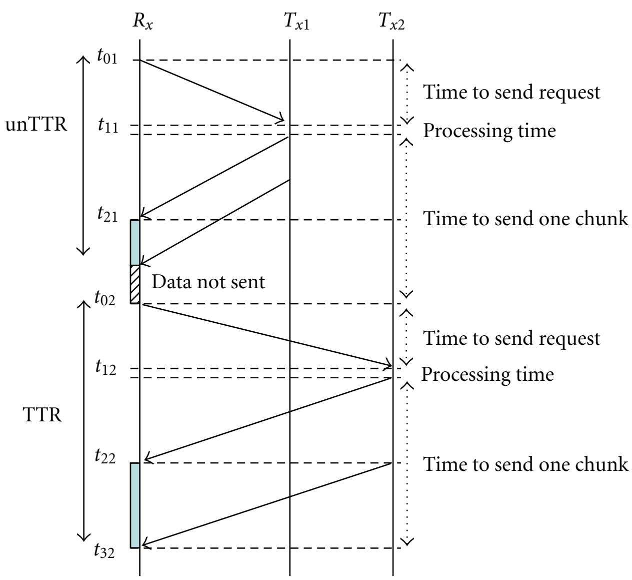


Figure 3: A peer churn event of length unTTR and successive successful request.

without server support. We assume that each peer access line has the same average upload bandwidth U, lower than the video streaming rate R. This is a frequent condition for Internet access in Small-Office Home-Office (SOHO) and domestic users, often characterized by asymmetric access lines.

Figure 2 shows the streaming time-line. The peer has filled its playback buffer and is then starting the playing back. Chunks are enumerated in an increasing progressive way, and the playback buffer B, measured in seconds, is of $4 * T_{UP}$ in length in this example. Chunks from 1 to M are not available yet and are in download phase from other peers at rate U. The deadline for every chunk i is

$$D_i = B + (i-1)T_{UP}. \quad (1)$$

Every time a disconnection occurs, the peer must contact a new available peer. We name the time necessary to complete a correct transmission Time-to-Redirect (TTR), as described in Figure 3.

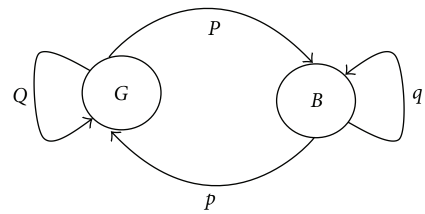


Figure 4: Gilbert-Elliot model.

On the contrary, the unTTR is the time wasted because of one disconnection and it is less or equal than TTR. For simplicity, we consider the worst case taking always unTTR equal to TTR. The TTR depends from many factors, such as nearness of other peers, popularity of video content, and network load.

Due to limited buffer capacity, peers can tolerate up to a maximum number of churns. When the total number of churns is becoming too high for a chunk transmission to a peer, the server takes part in the process by directly sending the chunk. In this scenario, it is interesting to evaluate which is the impact of churns to the whole system. In the following, we describe the proposed four theoretical models to represent peer churns in a P2P-VoD system.

3. Models

In this section we present the proposed models. The first two are based on the Gilbert Elliot (GE) model, the third one relies on Fluidic analysis, and the last is based on the Queueing theory.

3.1. GE Model. In this work, we initially model the peer behaviour using a two-state discrete-time process in which the time axis is measured in terms of TTR intervals. Such a process is then represented with a GE model [13, 14] drawn in Figure 4.

The transition probability P refers to the progress of the peer from the connected-state (good state G) to the disconnected-state (bad state B) during an interval TTR, whereas probability p refers to the inverse process.

Differently, Q and q refer to the probability to remain in the good and bad state, respectively, for an entire interval TTR. In our model, transition probabilities are changed time by time to represent changes in the user behaviour. This probability is taken randomly according to a uniform distribution because peer behaviour is considered stateless and peer participation is supposed very unpredictable. The uniform distribution is left constant for the entire session.

Based on the deadlines described in Section 2, the maximum tolerable number of disconnections is defined as

$$N_{\mathrm{DISC},i} = \frac{D_i}{\mathrm{TTR}}. \quad (2)$$

Each chunk has its own deadline, which has to be met not to interrupt video playback. The probability to satisfy the deadline condition for a generic chunk i is:

$$\psi_i = \sum_{k=1}^{N_{\mathrm{DISC},i}} P^{k-1} * Q = P_{S,i}. \quad (3)$$

This condition has to be fulfiled for every chunk that a peer is downloading. The probability to fulfil this condition is

$$\psi = \prod_{i=1}^{M} \psi_i. \quad (4)$$

Considering the streaming rate R and the number of peers N into the system, the total bandwidth W_{TOTAL} requested by the whole system is

$$W_{\mathrm{TOTAL}} = R * N. \quad (5)$$

Instead, the peers can provide an upload bandwidth W_{PEER} equal to

$$W_{\mathrm{PEER}} = \psi * N * U. \quad (6)$$

Finally, the bandwidth that the peers are not able to guarantee is the difference between (5) and (6): this is the bandwidth W_{SERVER} requested to the server:

$$W_{\mathrm{SERVER}} = W_{\mathrm{TOTAL}} - W_{\mathrm{PEER}}. \quad (7)$$

3.2. GE Extended Model. The GE model in Section 3.1 is characterized by transition probabilities selected randomly according to a uniform distribution, which is kept constant during the entire video. However, recent studies [15] on user accesses over time, arrival rates, and session lengths have shown that the user behaviour changes during the video playback session. It often happens that the user starts streaming the video and, after a while, he is not satisfied with the content then moves to another video. Accordingly, the probability that a user selects another video is a function of the time, and it decreases as the total amount of played back video increases. Indeed, the probability of streaming interruption is very low after half of the video has been already seen. In particular, it has been proved that the cumulative distribution function of video session lengths is well fitted by an exponential distribution.

Starting from these studies, we propose a GE model extension in which the probability of disconnection P is set according to an exponential distribution: in this way the stay-connect time of each peer is a monotonically increasing function of time, reflecting user trend to stay connected once a significant part of the video has already been watched. Probability of connection p is instead kept constant: its temporal variation's scale is very big if compared with disconnection probability variation, and for this reason it can be considered constant.

3.3. Fluidic Model. Recently, researchers have explored stochastic fluidic analytical models [16, 17] to model traffic in P2P networks. In these models, data transmission is seen like a fluid transferred through nodes, in a similar way to hydraulic models.

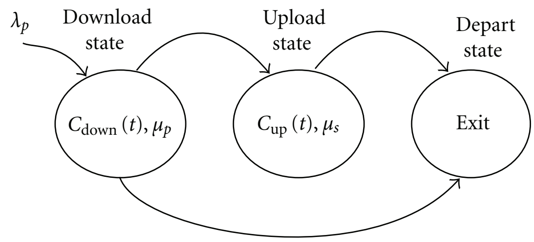


Figure 5: Peer state diagram.

Another study [11] develops a model for P2P-VoD in a broadcast environment. This model can be adapted to P2P-VoD with the hypothesis that peers in upload state can share all video in their memory, not only the first part. Peers can request aid from the server if the P2P network is not able to provide video data, which is the scenario we are considering in this paper. The state diagram of a peer has 3 states: download, upload, and depart, as shown in Figure 5.

When a peer joins the system, it goes in download state and can receive the first part of the video by the P2P network. Therefore, if its playback buffer is full, it goes to the upload state where it can share video parts already downloaded. Finally, a peer can leave the system and moving into the depart state.

The final target is reducing server load in the download state using upload capabilities of peers. From queueing theory point-of-view, the whole system can be approximated as a tandem queueing network with arrival and departure Poisson processes. Given the following:

$$\begin{aligned} &\lambda_p \quad \text{Arrival rate;} \\ &\mu_p = \frac{1}{\text{First\_part\_length}} \quad \text{Mean time in Download State;} \\ &\mu_s = \frac{1}{\text{Second\_part\_length}} \quad \text{Mean time in Upload state;} \\ &\gamma_p \quad \text{Mean Life Time;} \\ &C_{\mathrm{down}}(t) \quad \text{Number of peers in Download state;} \\ &C_{\mathrm{up}}(t) \quad \text{Number of peers in Upload state;} \end{aligned} \quad (8)$$

it can be developed a simple fluid model to study the system evolution. Peers number in the first state can be calculated considering their exponential distribution, which is proportional to ratio between peer's arrival rate and both mean life time and mean service time (9):

$$C_{\mathrm{down}}(t) = \frac{\lambda_p}{\gamma_p + \mu_p}\left[1 - e^{-(\gamma_p+\mu_p)t}\right]. \quad (9)$$

Instead, peers number variation in upload state is equal to the difference between peers coming from the download

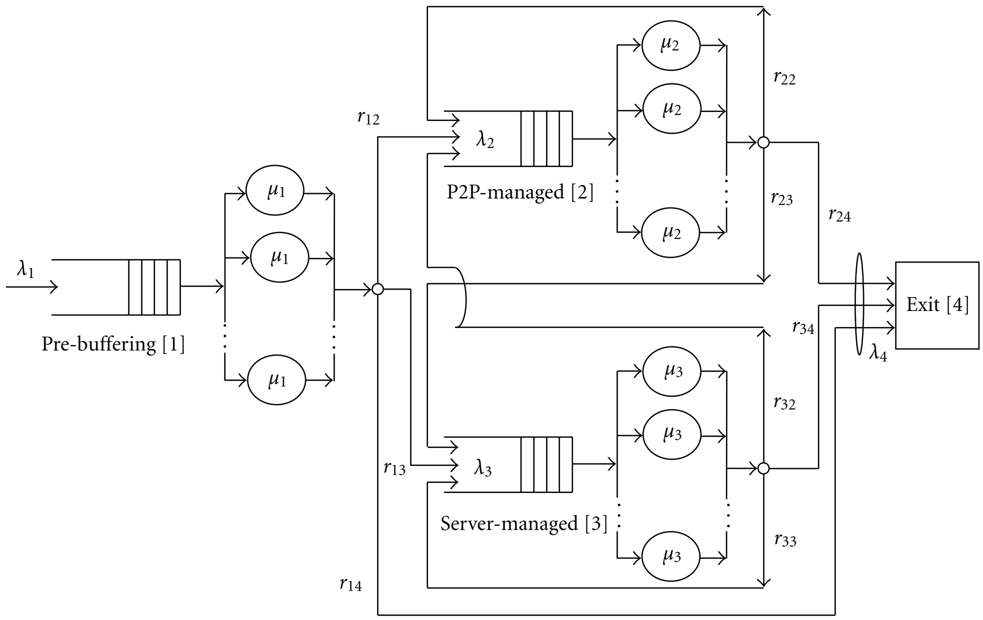


Figure 6: Proposed Queueing model of the P2P-VoD system.

state and the peers going to the exit state:

$$\frac{d}{d(t)}\left[C_{\mathrm{up}}(t)\right] = \mu_p C_{\mathrm{down}}(t) - \min\left(\mu_s, \gamma_p\right) C_{\mathrm{up}}(t). \quad (10)$$

The solution of differential equation (10) is the value of C_{up} as function of the time. The aggregate bandwidth of the P2P network W_{PEER} at time t is equal to $U * C_{\mathrm{up}}(t)$ and bandwidth W_{SERVER} requested by central server is

$$W_{\mathrm{SERVER}} = \left[C_{\mathrm{down}}(t) + C_{\mathrm{up}}(t)\right] * R - C_{\mathrm{up}}(t) * U. \quad (11)$$

3.4. Queueing Model. Queueing theory can be applied to a multiplicity of real problems, especially to transports and telecommunications fields, where each complex system is modeled by a set of queues connected each other. Each individual queue is called *node*, and the state of a queueing network is defined by the simultaneous distribution of customers in each node. In open networks the input rate to a queue i is given by

$$\lambda_i = \lambda_{0i} + \sum_{j=1}^{N} \lambda_j r_{ji}. \quad (12)$$

The term λ_{0i} is the arrival rate of tasks to ith node from outside, and r_{ji} are the routing probabilities that a served task is passed from node i to node j. The term λ_j is the arrival rate of tasks from internal nodes.

A simple queueing network model can be constructed splitting the life cycle of peer in four different phases or states. The first state is a "prebuffering state": peer joins the P2P network and buffers a certain quantity of data before to start video playing. When its buffer is full, it can be routed to the "P2P-managed state," to "Server-managed state" or can leave the system going in "Exit state." Each state is represented by an M/M/∞ queue except for the exit state. The proposed queuing model is shown in Figure 6.

The M/M/∞ queue model is chosen for its analytical tractability. The first queue exactly models the startup delay necessary to fill up the playback buffer. The aim is to collect enough data before starting the video playback to decouple the playback time from the transmission time. The buffer length is fixed, so that the service rate is constant:

$$\mu_1 = \frac{1}{B}. \quad (13)$$

When a peer has filled its buffer, it leaves the first queue and can be routed toward others queues or leave the system. Routing probability depends on the probability to leave the system α and probability to receive data from others peers P_{hit}. For each state, (12) has to be fulfilled as well as the constraint about outgoing routing probabilities for every i:

$$\sum_j r_{ij} = 1. \quad (14)$$

Additionally, the following routing probabilities apply:

$$\begin{aligned} r_{12} &= P_{\mathrm{hit}}, \\ r_{13} &= 1 - \alpha - P_{\mathrm{hit}}, \\ r_{14} &= \alpha. \end{aligned} \quad (15)$$

Exit state could be considered as another queue with service rate unitary: in truth, it is important to calculate only the overall arrival rate to evaluate model dynamics:

$$\begin{aligned} r_{12} &= r_{22} = r_{32}, \\ r_{13} &= r_{23} = r_{33}, \\ r_{14} &= r_{24} = r_{34}. \end{aligned} \quad (16)$$

The mean total number of peers in the system is

$$\overline{X}_{\mathrm{TOT}} = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} + \frac{\lambda_3}{\mu_3} = \rho_1 + \rho_2 + \rho_3. \tag{17}$$

Hit probability is calculated dynamically and is proportional to the arrival rate in queue 3 and in exit state:

$$P_{\mathrm{hit}} = 1 - \frac{\lambda_3}{\overline{X}_{\mathrm{TOT}}} - \frac{\lambda_4}{\overline{X}_{\mathrm{TOT}}}. \tag{18}$$

Considering the number of peers ρ_i in each queue i, the bandwidth requested by central server is

$$W_{\mathrm{SERVER}} = (\rho_1 + \rho_2 + \rho_3) * R - \rho_2 * U. \tag{19}$$

Finally, we need to specify the sense of mean service time in queues 2 and 3: every time-step long as mean service time, the next peers' status is set in relationship to the number of peers in queues 2 and 3. If the P2P system contains a sufficient number of peers so that the hit probability is high, this situation influences probability of routing toward P2P-managed state. Otherwise, P_{hit} decreases and it is more probable that a peer will forward to Server-managed state. Notice also that it is not possible to have peers in waiting line because there is always a servant free in an M/M/∞ queue.

4. Simulations

We have performed extensive simulations with different scenarios. The objective of the simulation analysis is to investigate the models behaviour varying the system parameters in order to assess the usefulness of such models in supporting the design and configuration of P2P-VoD architectures. Herein, we present the results when applying the following streaming parameters: transmission of video sequences of 100-minute length at 800 Kbps and an upload rate U of 600 Kbps. We choose these values according to the condition $U < R$, that reflects the most common situation of Internet access lines as explained in Section 2.

The simulations with the GE model has been conducted with a total of 50 peers in the system and changing the stay-connect probability Q every 10 seconds. This probability has been chosen according to a uniform distribution with different ranges, as shown in Table 1. The connection probability p has been kept constant and equal to 0.5 during all simulations.

In the GE extended model, the disconnection probability P follows the exponential distribution (20)

$$P(t) = T * e^{-(t/T)} \tag{20}$$

with the parameter T set so that the complementary stay-connect probability Q has the mean values of Table 1.

To evaluate the effectiveness of these two models, we have computed the requested server download rate at varying disconnection probabilities P. Figure 7 shows the results for the two models when the Time-To-Repair TTR has been set to 300 milliseconds and changing the buffer length from 0.6 to 4.2 seconds.

Table 1: Ranges of the uniform distribution for the probability Q of the GE model.

Range	Values	Mean
#1	0/0.1	0.05
#2	0/0.2	0.1
#3	0/0.3	0.15
#4	0/0.4	0.2
#5	0/0.5	0.25
#6	0/0.6	0.3
#7	0/0.7	0.35
#8	0/0.8	0.4
#9	0/0.9	0.45

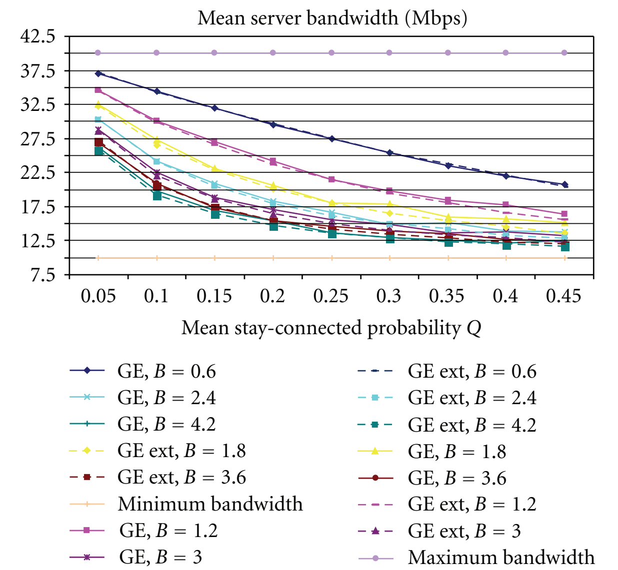


Figure 7: GE models comparison for different values of buffer length.

It can be noted that the two models show similar behaviours, as it was expected since the models are basically the same except the distribution of the connection probability. The shape of the plots shows that increasing the buffer length brings to lower requested bandwidth values. This is due to the fact that the deadline for each chunk is less stringent, allowing for finding an active peer from which successfully download the chunk. The curves are convex so that a higher benefit is obtained by increasing the buffer length at low values of the Q probability. The figure also shows that the total amount of average server bandwidth converges towards 10 Mbps ("minimum server bandwidth" in the figure), which indeed is the difference between the W_{TOTAL} bandwidth of 40 Mbps and the maximum theoretical bandwidth provided by the peers W_{PEER}, which is of 30 Mbps. Overall, this figure is a handy tool that helps the designer in finding the server resources that are required to satisfy the user requests on the basis of the playback buffer length and as far as the operator is able to estimate the peer stay-connected probability. Note that the curve

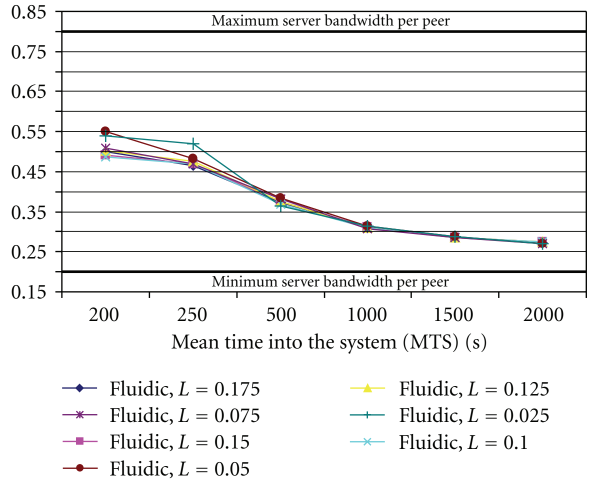


FIGURE 8: Mean server bandwidth requested by each peer in the Fluidic model with different peer input rates.

"maximum server bandwidth" in the figure represents the amount of server bandwidth that would be necessary without the support of the peer-to-peer network.

In the Fluidic and the Queuing Network models, one of the key parameters is the mean time a peer spends in the system, which is the peer Mean Time in the System (MTS). Whereas for the first model it is directly set by selecting the value of Mean Life Time γ_p, for the Queueing Network model the MTS is indirectly set through the probability to leave the system α, the sampling step Δ, and number of simulation samples N_s according to the following formula:

$$\mathrm{MTS} = \sum_{i=0}^{N_S} \left[(1-\alpha_i)^i \alpha_i \cdot i \cdot \Delta \right]. \quad (21)$$

Equation (21) has been used to find parameters values to achieve the desired MTS. As to the parameter P_{hit}, it has been initialized to 0.9, whereas successive values are dynamically calculated according to the model evolutions.

For the analysis of these models we have computed the mean server download bandwidth requested by each peer, while varying the following parameters: MTS and input arrival rate λ (L in Figure 8). These two parameters affect the number of peers into the system, which then cannot be directly set by us as in the GE models.

Figure 8 shows the requested bandwidth for the Fluidic model. Note that this time the resulting value has been divided by the number of peers in the system, which is different for any combinations of system parameters (see Table 2). In this figure we are also showing the upper and lower bandwidth limits: 800 Kbps is the rate requested to the server when no one peer is able to share video data, whereas 200 Kbps is the difference between the video rate R and the maximum upload rate U, which corresponds to the amount of bandwidth that should be provided by the server when all the active peers are successfully sending video content to another peer.

TABLE 2: Mean number of peers measured in the Fluidic Model.

Input rate	Mean number of peers					
0.175	34	41	80	146	198	240
0.150	29	36	69	125	170	205
0.125	24	29	58	104	141	171
0.100	19	23	46	83	113	137
0.075	14	18	34	62	85	102
0.050	9	11	23	42	56	68
0.025	4	6	11	21	28	34
MTS	200	250	500	1000	1500	2000

The shape of the plots shows a decreasing bandwidth requested as a function of MTS and implicitly with the increasing number of peers into the system: this behaviour confirms the implicit feature of system scalability of P2P systems. In fact, a bigger number of peers into the system generates more resources (upload bandwidth), reducing the bandwidth requested to server per peer into the system.

5. Conclusions

In this paper we have presented three mathematical models for the evaluation of the peer churn impact on the server resources in P2P-VoD systems. In the first model, the behaviour of each peer is represented by means of the Gilbert-Elliot model, where the two states are associated to the connected and disconnected states. The second and third models use a very different approach with respect to the GE one: a constant number of peers joins the system and the resources requests are related to the effective number of peers inside the system.

The simulations have shown that these models are an effective tool that help the designer in finding the server resources that are required to satisfy the user requests on the basis of the playback buffer length, as far as the operator is able to estimate the peer stay-connected probability. The longer the time each peer spends in the system, the lower the resource required to the server. In fact, an increase in the average stay-connected interval decreases the probability to waste time sending only useless partial chunks from peer to peer, which need to be resent from the beginning by another peer (if available) or by the server.

References

[1] Cisco White Paper, "Approaching the Zettabyte Era," June, 2008, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481374.html.

[2] H. Schwark, D. Marve, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[3] F. Thouin and M. Coates, "Video-on-demand networks: design approaches and future challenges," *IEEE Network*, vol. 21, no. 2, pp. 42–48, 2007.

[4] BitTorrent, http://www.bittorrent.com/.

[5] Skype, http://www.skype.com/.

[6] J. Liu, S. G. Rao, B. Li, and H. Zhang, "Opportunities and challenges of peer to peer internet video broadcast," *Proceedings of IEEE*, vol. 96, pp. 11–24, 2008.

[7] J. Tian and Y. Dai, "Understanding the dynamic of peer-to-peer systems," in *Proceedings of the International Workshop on Peer-to-Peer Systems (IPTPS '07)*, 2007.

[8] D. Stutzbach and R. Rejaje, "Understanding churn in peer-to-peer networks," in *Proceedings of the 6th ACM SIGCOMM conference on Internet Measurement*, pp. 189–202, September 2006.

[9] Y. Guo, S. Mathur, K. Ramaswamy, Y. Shengchao, and B. Patel, "PONDER: performance aware P2P video-on-demand service," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 225–230, Washington, DC, USA, November 2007.

[10] Z. Liu, Y. Shen, S. Panwar, K. W. Ross, and Y. Wang, "Efficient substream encoding and transmission for P2P video on demand," in *Proceedings of the Packet Video*, pp. 143–152, Lausanne, Switzerland, 2007.

[11] K. M. Ho and K. T. Lo, "A simple model for peer-to-peer video-on-demand system in broadcast environment," in *Proceedings of the International Conference on Information Networking (ICOIN '08)*, pp. 1–5, January 2008.

[12] C. Huang, J. Li, and K. W. Ross, "Peer-assisted VoD: making internet video distribution cheap," in *Proceedings of the International Workshop on Peer-to-Peer systems (IPTPS '07)*, Redmond, Wash, USA, 2007.

[13] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Systems Technical Journal*, vol. 39, pp. 1253–1265, 1960.

[14] E. O. Elliot, "Estimates of error rates for codes on burst-noise channels," *Bell Systems Technical Journal*, vol. 42, pp. 1977–1997, 1963.

[15] H. Yu, D. Zheng, and B. Y. Zhao, "Understanding user behavior in large-scale video-on-demand systems," in *Proceedings of EuroSys*, pp. 333–344, Leuven, Belgium, April 2006.

[16] R. Kumar, Y. Liu, and K. Ross, "Stochastic fluid theory for P2P streaming systems," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 919–927, May 2007.

[17] Y. Lu, A. Zhang, H. He, and Z. Deng, "Stochastic fluid model for P2P content distribution networks," in *Proceedings of Autonomous Decentralized Systems (ISADS '05)*, pp. 707–712, April 2005.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 617203, 9 pages
doi:10.1155/2009/617203

Research Article

A Software and Hardware IPTV Architecture for Scalable DVB Distribution

Georg Acher,[1] Detlef Fliegl,[1] and Thomas Fuhrmann[2]

[1] *BayCom GmbH, Guardinistr 47 81375 München, Germany*
[2] *Lehrstuhl für Netzarchitekturen und Netzdienste, Fakultät für Informatik, Technische Universität München, 85747 Garching bei München, Germany*

Correspondence should be addressed to Georg Acher, acher@baycom.de

Received 1 February 2009; Revised 3 September 2009; Accepted 30 September 2009

Recommended by Maurizio Murroni

Many standards and even more proprietary technologies deal with IP-based television (IPTV). But none of them can transparently map popular public broadcast services such as DVB or ATSC to IPTV with acceptable effort. In this paper we explain why we believe that such a mapping using a light weight framework is an important step towards all-IP multimedia. We then present the NetCeiver architecture: it is based on well-known standards such as IPv6, and it allows zero configuration. The use of multicast streaming makes NetCeiver highly scalable. We also describe a low cost FPGA implementation of the proposed NetCeiver architecture, which can concurrently stream services from up to six full transponders.

1. Introduction

In this paper we present NetCeiver, an IPTV architecture that is fully compatible to existing digital television broadcast systems. This allows an easy convergence of digital television and IPTV. During the transition, consumers can thus benefit from the advantages of both technologies.

Throughout this paper we use the term *convergence* to describe a way of delivering content independently of a specific transport medium. In traditional broadcast systems, this would be called simulcast. IPTV allows convergent deliverance of television content over IP. Convergence does not require restoring all physical capabilities of different transport media even though it is desirable. Therefore, the resulting feature set and quality of the encoded media streams may change.

In the following, we first briefly introduce some typical characteristics of the IPTV systems' basic operation schemes. We then examine traditional television broadcast. The comparison of both approaches allows us to investigate the possible means for convergence. Based on our analysis, we derive a novel architecture for transparently transporting broadcast TV services over IP.

1.1. IPTV. IPTV is a well-known term that describes IP-based television in general. Unlike other techniques there is no common specification nor standard that can be used to implement IPTV-based services for every purpose. In the past years several approaches have been made to achieve different solutions. Each solution has its own set of features and optimizations. Some IPTV systems are proprietary; others rely partly or fully on standards such as RTSP [1], SDP [2] or UPnP-AV [3]. This makes it difficult to easily compare IPTV implementations; more abstract characteristics need to be examined.

Roles. TV in the traditional meaning has one provider (server) and multiple receivers (clients). Typical IPTV applications show a similar setup. They only replace the transport medium by an IP-based technology.

However, some IPTV systems that are based on the Web 2.0 concept have given up the distinction between sender and receiver. YouTube, Vimeo, Make.tv, and the like transmit user generated content. Nevertheless they provide their services in a TV-channel-like fashion, either off-line or via live streaming.

Distribution. Professional IPTV systems, are typically based on a centralized infrastructure. They use dedicated hardware and network resources for their content distribution. This allows a reliable service, although at high costs.

A peer-to-peer (P2P) approach can eliminate most of the centralized IPTV infrastructure. P2P systems rely on a large number of participants (aka peers) that provide their service within a P2P overlay network [4]. Each peer is a sender and a receiver at the same time. It relays traffic within the P2P network. This results in better load balancing and fault tolerance. Compared to centralized solutions, P2P systems for IPTV are quite new. They do not yet fully exploit their potential [5, 6].

Interactivity. Often, advertisements list interactivity as important IPTV feature. The back channel over IP allows a wide variety of interaction schemes. In reality, these additional services are seldom really interactive and novel (e.g., rating, feedback, and tagging). Usually, they are only more visually attractive versions of traditional broadcast services such as electronic program guides or teletext. Often, they just combine TV with web browsing.

As an exception, typical video on demand (VOD) services provide interactivity in a different way: the customers can select the particular video they want to watch at a given or chosen time. As zapping in IPTV results in bidirectional communication with a server, this could be understood as interactivity as well.

1.2. Digital Television Broadcast Systems. Despite the fact that IPTV systems have been available on the market for quite some time, legacy distribution systems for digital television are still predominant. Most TV stations employ these systems, and as a result, they are more popular than any other IPTV solution—at least in western countries.

These legacy distribution systems broadcast according to well-established standards over terrestrial antenna, satellite transponders, or broadband cable networks. Inexpensive receivers provide consumers easy access to these services. Typically, their quality exceeds that of most current IPTV implementations.

Digital broadcast distribution systems differ by region. The most popular standards are DVB [7] in Europe, ATSC [8] in North America, ISDB [9] in Japan and Brazil, and DMB [10] in South Korea.

For simplicity we refer to all these systems as DVB, because they all use the MPEG2 transport stream format (MPEG2-TS) for their transmissions.

Digital television is widespread. Consumers accept it thoroughly. DVB provides high-quality audio and video transmissions, fast channel switching, and a vast number of additional services such as electronic program guides or conditional access (Pay-TV). Many of those services are fully standardized. As a result, consumers can trust that any commodity receiver is compatible with their preferred service provider. As stated above, this situation is still totally different in current IPTV systems.

If we had the same features for IPTV, we could rely on well-standardized client applications or set-top boxes that we could use with more than one IPTV service only. Unfortunately, this is unlikely to happen in foreseeable future, because IPTV systems have to meet too many different, mostly commercial interests.

In this paper we present a possible solution to this dilemma. Our NetCeiver approach is based on DVB. It makes all features of DVB broadcast systems available from within IP-based local area networks. Before presenting the NetCeiver architecture, in particular its communication protocols and hardware details, we take a look at other IPTV approaches.

1.3. State-of-Art IPTV Implementations. In the early 90's, a very first implementation of IPTV was based on the MBONE network [11]. This network enabled delivery of IPv4 multicast across nonmulticast capable networks. The original purpose of transmitting audio and video was providing a conferencing facility. Therefore a session announcement based on SAP [12] and SDP [2] is being used to show all multicast senders in a browser application. Based on this information it is possible to join multicast groups containing audio and video streams in RTP [13] format. Soon organizations like NASA started to broadcast shuttle launches and other popular events over the MBONE network.

Today, depending on the application, a vast number of IPTV standards and services exist. Still IPv4 multicast is being used in large-scale deployments [14]. This generally requires an infrastructure capable of multicast routing. Advanced IPTV systems seem to use a mixture of proprietary multicast and unicast protocols. As most of them not yet disclosed internals only speculations on their operation scheme can take place. Only few systems such as Universal Plug and Play [3] or DVB-IPTV [15] rely on open standards. These systems will be explained shortly.

Universal Plug and Play (UPnP) refers to a set of standards published by the UPnP Forum. It is primarily designed for usage in small or home networks. The UPnP system is based on IPv4 or IPv6 addressing and it uses the Simple Service Discovery Protocol [16]. Devices and services are described in XML formatted data structures and device functions are controlled by the Simple Object Access Protocol (SOAP) [17]. A subset of standards called UPnP-AV defines different components that can be used to implement IPTV. A server component called MediaServer and a client component called MediaRenderer have to be implemented. The standard defines different methods for the media access based on an URL scheme. The media streams can be encoded in several ways with different codecs depending on the capabilities of the single components. Today many UPnP-AV capable devices are commercially available. Standards like DLNA [18] integrate the UPnP specification which often results in DLNA/UPnP compliant devices.

The DVB-IPTV standard was originally introduced under the name Digital Video Broadcast Internet Protocol Infrastructure (DVB-IPI) in the year 2001 by the DVB project, an alliance of about 300 worldwide companies. DVB-IPTV covers a set of open and interoperable technical specifications. Its primary goal is the distribution of DVB services within a large-scale IP network. It is based on MPEG2 transport streams carried over IPv4 networks as unicast or multicast in RTSP format. For audio and video the same codecs as in broadcast systems have been selected. The standard also contains mechanisms for service discovery and service selection. Therefore the normal DVB Service Information (DVB-SI [19]) is being used in a XML coded data structure. Interactive content can be transported according to the DVB Multimedia Home Platform (DVB-MHP [20]) specification. Further DVB-IPTV will release a home network specification based on the DLNA/UPnP guidelines. Today operational DVB-IPTV systems can be found rarely as the standard is not yet fully developed. DVB-IPTV has a great potential as it could fully replace traditional television broadcast.

A general drawback of many IPTV architectures has been a high channel switching time. This is caused by different factors such as multicast request forwarding, buffering and decoding delay. Various approaches have been developed to reduce or hide these delays,

(i) Zapping mostly takes place between adjacent channels of a known list. When switching to a channel the client application joins the multicast group of the next channel in the list [21]. If zapping continues the stream of the next channel is already present. As soon as a channel gets watched for a while the multicast group of the next channel is being dropped again.

(ii) Advanced mechanisms use a so-called "GOP server" to enable an instant channel change (ICC) [21]: After selecting a new channel a multicast group gets joined, but also a "GOP server" is contacted via unicast. The server delivers a data burst containing a full group of pictures that enable the client to instantly display the new programe. After the multicast stream gets switched through to the client, the unicast connection is being closed.

(iii) The requirement of a specialized GOP server might reduce availability and cause load peaks within the network. Therefore another solution was found: so-called "secondary channel change stream" is provided for each regular channel [22]. As soon as the client switches to a new channel two corresponding multicast groups are being joined. The first one contains the regular high bandwidth stream, the second is a lower bitrate version carrying out only I-frame and associated audio. The latter enables the client to instantly display the new channel.

2. DVB to IP Network Bridging

The idea of delivering digital television broadcast to IP networks is not new at all. There already exist several implementations that receive, transcode, and distribute TV programes via IP. All these solutions provide random access to audio and video streams, which enables the user to watch TV on LAN-attached PCs or set-top boxes. But a closer look at these implementations reveals that they do not transparently bridge all digital broadcast features. Therefore only a fixed set of additional services such as the electronic program guide or teletext is converted for IP usage. Furthermore, these systems do not keep the broadcast character of the received streams which means they use unicast transport connections. DVB-IPTV is an exception to this. It reflects the broadcast character of DVB, but it is not yet widely accepted for production use and requires an expensive infrastructure.

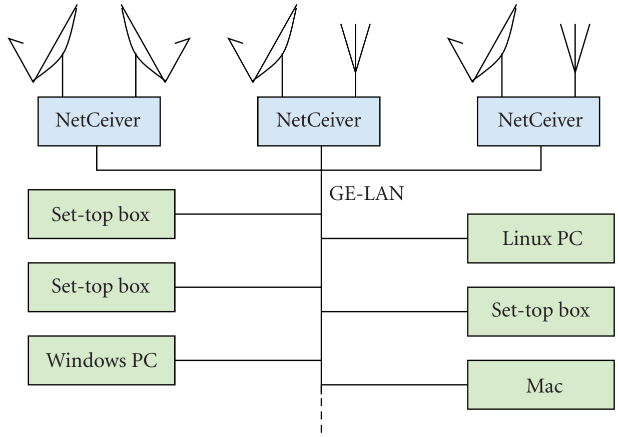


Figure 1: Usage scenario.

2.1. The NetCeiver Architecture. The primary goal of the NetCeiver architecture is to map full DVB services into IPTV without changing their quality, feature set, or broadcast character. It should be possible to access digital television services without prior (network) configuration, proprietary client software, or proprietary set-top boxes. Multiple DVB services, potentially received on different frequencies, should be available for multiple clients in the local IP network. The latter has been achieved in a scalable manner. As depicted in Figure 1, the NetCeiver can be seen as a head end for DVB distribution within IP-based networks.

Regarding our characterization in Section 1, NetCeiver is an infrastructure-based client/server system with a clear role assignment. Interactivity on one hand depends on the features of DVB distribution. NetCeiver does not support them as thoroughly as typical IPTV applications. On the other hand, NetCeiver benefits from all the capabilities of Internet access. The NetCeiver system provides a fully convergent solution in terms of our definition above. There are no restrictions when using DVB services from an IP network.

The NetCeiver architecture is based on highly specialized, cost efficient hardware with low power requirements. It provides slots for DVB tuners, and CAMs and has a Gigabit Ethernet interface (see below). It runs a Linux-based operating system, which allows resource efficient, simple and flexible software development.

A portable library provides IPTV client support for the NetCeiver architecture. It can be used on (embedded) Linux systems, Microsoft Windows, and MacOS. The library is designed to support full DVB access without any prior network configuration or other restrictions. Kernel drivers and daemons for Linux-and-Windows based systems emulate DVB capable hardware. This enables any DVB client software to transparently use the NetCeiver system.

2.2. Software Architecture. As stated above, we designed the NetCeiver system as an IPTV system for DVB services. For this purpose we had to first select a suitable transport protocol. Although many multimedia streaming frameworks already exist, we decided to find a new, more efficient way that meets the following requirements:

(1) transparent DVB access: support for all quality modes, the entire DVB feature set, and low zapping times (below 1-2 seconds),

(2) zero configuration: a self-organizing system that runs without any prior setup,

(3) scalability: multiple servers can serve multiple clients in an efficient way,

(4) low cost solution: server and client are designed to meet low cost requirements.

The rationale for these requirements is based on the way how digital television broadcast systems operate: DVB services are organized as a multiplex called *transport stream* (TS). Each TS contains a number of logical channels, which are identified by a unique 13-bit value, the *Packet Identifier* (PID). The so-called *Program Map Table* (PMT) lists and describes all PIDs that belong to a DVB service. Therefore it is necessary to selectively access single PIDs of a TS. A DVB receiver can bootstrap all available services (PMTs) by accessing the *Program Allocation Table* (PAT). The PAT is always located on PID 0. Studying various IPTV and streaming frameworks we learned that they do not well support this DVB operation scheme because of the following:

(1) none of the established protocols or streaming framework efficiently satisfies the requirement of transparently addressing single PIDs of DVB transport streams,

(2) some existing protocols, like UPnP-AV [3] or Multicast DNS [23], provide zero configuration support, but they are focused on already preassembled services. Implementing them for the NetCeiver system might be possible to some extent, but it would not be fully compatible with the predefined DVB schemes.

(3) the system has to be bandwidth efficient to be scalable: when different clients request the same PID, the corresponding data still should be sent over the network just once. This requirement can only be met by multicast capable streaming protocols,

(4) solutions like DVB-IPTV require a costly server infrastructure, where many different services and protocols have to be supported. On the client side, traditional DVB only set-top boxes need significant software changes; it is not easily possible to just replace the integrated tuner by a network data source.

2.2.1. IPv6 Multicast Based DVB Access. In order to meet these requirements, we designed the NetCeiver to selectively stream single PIDs of a DVB transport stream into a LAN using IPv6 multicast. Each stream is controled by the standard IPv6 multicast listener discovery protocol (MLDv2 [24]). This standard describes how to report multicast groups within a network segment: an IPv6 multicast capable router periodically sends out a multicast listener query to discover the presence of multicast listeners. Every multicast capable network node that has joined or recently left at least one multicast group reports its list of groups to the querier. Furthermore, a multicast node reports joining or leaving groups immediately without a prior listener discovery.

The NetCeiver exploits the MLDv2 protocol by sending out multicast listener queries periodically and receiving all multicast listener reports. This results in an up-to-date list of all joined multicast groups.

Up to this point, MLDv2 only handles the multicast management like IGMP for IPv4. But additionally the NetCeiver itself (ab)uses the very same join message from a client to determine all tuning parameters and sets up the tuner and streaming hardware. So by just blindly joining a not yet existing multicast group, and without other side-band communication, the full stream mechanism comes to life on demand.

To achieve easy scalability and transparency, we chose to broadcast each PID on a different multicast group instead of combining multiple PIDs for one service in one stream. Thus there is no need for managed "channel lists" on the NetCeiver. Also a client can arbitrarily request individual PID streams as with typical PID filtering in set-top box hardware. This eliminates any software changes above the low-level DVB drivers.

For a tune request, the NetCeiver needs to know which PID from which transponder or frequency has to be streamed on a given multicast group. It achieves this by applying an algorithm that transforms specific receiver settings into an IPv6 multicast group and vice versa.

An IPv6 address consists of 128 bits, which are written with hexadecimal characters in eight groups of 16 bits each [25]. Depending on the application there exist different address schemes. This means that not the entire 128 bits address space can be used freely. For example, multicast applications have to set the first 8 bits to 0xff, followed by a 8 bit scope identifier. Only the remaining 112 bits differentiate between the multicast groups.

When a specific PID of a DVB transponder is selected, our algorithm calculates a corresponding 112 bit multicast group address. The so-calculated multicast group uniquely identifies the tuning parameters. Hence, we can also calculate the inverse mapping. Moreover, 128 bits contain enough space to identify all relevant parameters: a delivery system (satellite, terrestrial, etc.), orbital position (for satellites),

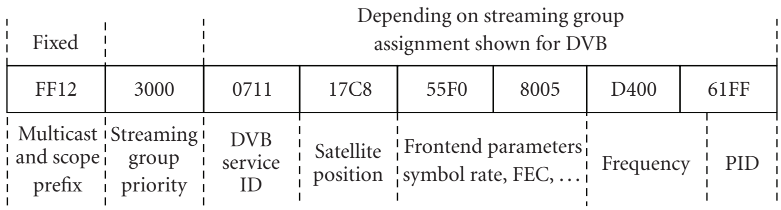


Figure 2: IPv6 multicast group decoding (example for DVB).

frequency, polarization, symbol rate, FEC modes, and PID. Figure 2 gives an example of this scheme. Thus a client tune request is issued just by joining a matching multicast group which contains all relevant information.

2.2.2. Zero Configuration Protocol. So far, we have just presented a transparent multicast-based mapping of DVB parameters. The next requirement deals with the zero configuration feature of the NetCeiver system. To this end, we defined a fixed set of multicast groups. All clients and servers have to join these groups, independent of any DVB streams (cf. Figure 3). All NetCeivers distribute their hardware capabilities and their current tuner allocations on these groups periodically. We use XML coded data structures for this purpose. They comply with the W3C RDF standard CC/PP (Composite Capability/Preference Profiles [26]).

Based on this information, a client can find available NetCeivers and configure itself automatically accordingly. When a client joins a multicast group for receiving a specific PID, it also joins a corresponding group that distributes the signal information from the tuner in charge.

When there are multiple NetCeivers on a network segment, it is also necessary to synchronize all NetCeivers so that they have consistent information about their state. This ensures a proper handling of requests if more than one NetCeiver could serve a multicast group. It also enables resilience in case one NetCeiver fails.

All of these features work in IPv6 without any prior network setup because every IPv6 capable network interface automatically receives a link local address. With such an address it is possible to communicate within a network segment. Unlike DVB-IPTV/UPnP, no DHCP or Auto IP mechanisms for IPv4 are required.

2.2.3. Scalability. The last requirement remaining is scalability. As said above, we wanted to have a resource and bandwidth efficient system that can handle multiple servers and clients. The unique multicast group calculation algorithm prevents the system from duplicating streams for different clients. The self-organization features allow a potentially large number of clients to concurrently feed from one or multiple NetCeivers.

Given the bandwidth limitation of a gigabit network, it is possible to stream a large number of DVB services: Assuming a bandwidth of 8 Mb in H.264 for HDTV and 5 Mb MPEG2 video in SDTV, About 100 HDTV or 160 SDTV programs could be streamed simultaneously. A small number of NetCeivers suffice to satisfy such a high demand: each NetCeiver can stream six full DVB-S2/DVB-C transponders, leading to a maximum bandwidth of about 300 Mb. Given a typical mixture of requests, the number of NetCeivers should not exceed five, unless the switching hardware can snoop MLDv2 and thus limit the traffic to the respective clients.

Currently, there exists a restriction when accessing a NetCeiver beyond router boundaries. Due to the operation scheme of multicast routing [27], it is impossible to propagate join operations for groups under certain conditions: traffic on multicast groups is being generated by the NetCeiver upon an incoming join request of a client. A router cannot determine where to forward this request for a yet inactive group of a specific NetCeiver.

3. Hardware

3.1. Overview. Figure 4 shows the NetCeiver hardware components. The central part is a Xilinx XC3S1600E FPGA [28] with nominally about 1.6 million gates. It acts as a dedicated System-on-Chip (SoC). We decided to use an FPGA because no standard SoC provides the required performance and data handling capabilities. Although, in general, an FPGA is more expensive than a standard SoC, it is nevertheless the key for the NetCeiver's low component cost (less than €50 excluding tuners). Moreover, the dedicated streaming logic also has a very good power efficiency, approximately 4–6 W for the base system and additionally 2–4 W for each S2-tuner. The FPGA does not require a heat sink so that the system does not need a forced cooling. This further increases its overall reliability.

The FPGA interfaces to three exchangeable tuner cards. Each card can deliver the transport stream (TS) of two tuners. Thus a NetCeiver board can use up to 6 DVB tuners in parallel. The overall input bandwidth peaks at about 35 to 40 MByte/s, depending on the tuned transponders.

Two slots for Conditional Access Modules (CAMs) can optionally decrypt the received streams. A flexible TS handling matrix in the FPGA can insert the CAMs into any tuner stream.

The NetCeiver requires only a few active components besides the FPGA: an AVR microcontroller with a MMC/SD-Card interface initializes the SoC with the FPGA bitstream and the uCLinux image. After bootup, the AVR allows simple block device IO to the flash card. It also acts as an IO expander for the GPIOs and the RS232 interfaces, and it helps in monitoring the analog voltages on the board.

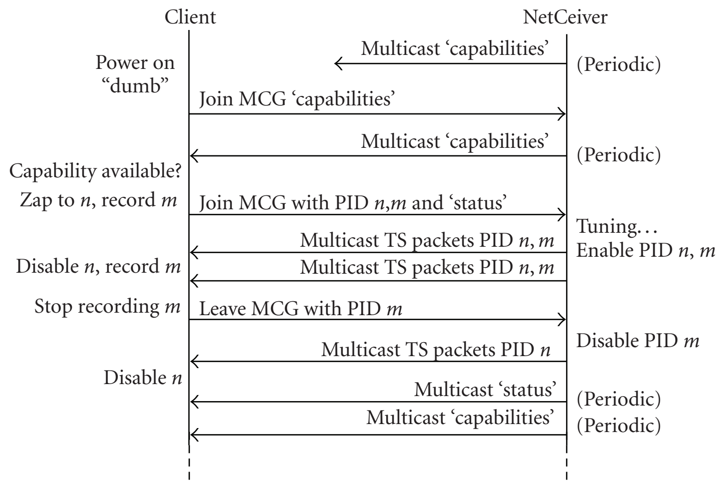


Figure 3: Zero configuration and implicit controlling via join and leave.

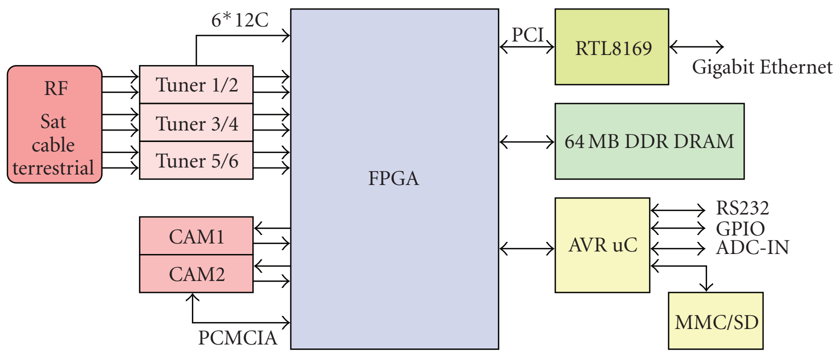


Figure 4: Hardware overview.

The FPGA has access to a 64-MByte DDR-DRAM. One half of the memory is used by the streaming hardware to buffer the TS packets; the other half is dedicated to the uCLinux-OS.

A PCI-connected Realtek 8169 Gigabit Ethernet interface handles the network connection. Even though network chips from Realtek used to have a bad reputation, this no longer true for the newer variants. The RTL8169 provides a high performance (above 100 MBytes/s) for very low cost. This makes it an ideal interface that did not require much development work. Additionally, the chip has two separate transmit queues. This made it easy for us to implement an independent hardware controlled streaming in parallel to the operating system. We explain this in detail later.

A summary of the technical specifications is given in Table 1; the PCB is shown in Figure 5.

3.2. FPGA Data Flow

3.3. Overview. Figure 6 depicts the SoC that is implemented inside the FPGA. It consists of the CPU core, the memory controller, a PCI subsystem, and the TS and network stream processor. Apart from the CPU, which uses IP cores provided by Xilinx in the Embedded Development Kit (EDK), we developed all other blocks specifically for the NetCeiver.

The main CPU block is realized by a Microblaze IP core [29] running at 66 MHz. The CPU core has instruction and data caches with 8 KB each. Although the Microblaze is a full-featured 32-bit RISC CPU and can run the MMU-less uCLinux, it requires only about 10% of the FPGA's resources.

The memory controller interfaces to a single-chip, 16 bit data bus DDR-DRAM clocked at 132 MHz. Internal provisions allow a high performance with a sustained memory bandwidth of about 400 MByte/s for longer burst accesses.

The PCI subsystem implements a 32-bit/33 MHz host bridge (initiator/target) to connect to the RTL8169. This allows us to use the 8169 Linux driver without modifications. The relatively slow CPU allows only about 2-3 MByte/s IP network bandwidth. This is sufficient for control connections, but obviously not for streaming multiple DVB channels. They require a much higher bandwidth.

For this purpose, the PCI core contains a bypass data path. It is enabled when accessing a specific memory area. This area does not map to system memory but to special registers in the network streaming controller. The second transmit queue of the RTL8169 reads on-demand hardware-generated DMA descriptors and packet data over this bypass.

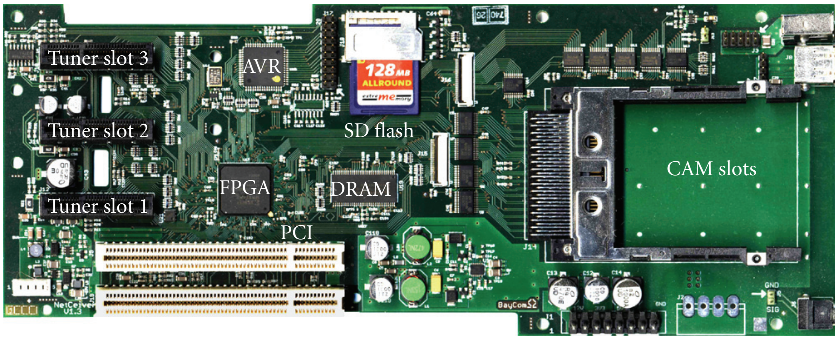


Figure 5: PCB of the NetCeiver.

Table 1: Technical specifications for the first NetCeiver HW implementation.

Tuners	6 TS inputs (each up to 9 MByte/s)
CAMs	2 (2 TS in, 2 TS out)
Network	RTL8169 Gigabit Ethernet
FPGA	XC3S1600
Internal clocks	66 MHz, except 33 MHz PCI, 132 MHz DDR & TS matrix
Memory	64 MB DDR-DRAM, 128 MB Flash
Power consumption	4–6 W (excl. tuners and CAMs)
OS	uCLinux
Boot time	Approx. 25s (depending on tuners)
Dimensions	Approx. 28 cm∗12 cm (16 cm∗12 cm w/o CAMs)
Height	8 cm (incl. network and tuner cards)
PCB	4 layer
Component cost	< €50 (at mid volume quantities, excl. tuners)

This fully offloads the CPU from the packet generation process for streaming. A benchmark using synthetic data packets showed that the RTL8169 and the PCI-bypass achieved a transmit rate of about 100 MB/s, which is about 80% of a Gigabit link capacity. This is fully sufficient for our purpose, even when the clients request six full DVB transponders.

On the input side, the TS data coming from the six tuners is routed through a matrix, which also allows to flexibly insert the decrypting CAMs into the stream. Afterwards, a table lookup that is based on the packet ID (PID) determines whether at least one client requires this PID at all. This *PID-info table* then also determines which way the packet should take. For the main data path network streaming FIFO areas (1 MB per tuner) in system memory buffer the packets. In total, there are 31 FIFO destinations possible for the PID-info. They are also used internally for monitoring and service information parsing.

One key element of the NetCeiver multicast protocol is that each PID is sent over a separate multicast address. Unfortunately, DVB delivers the PIDs in a randomly multiplexed order. They are not sorted to form bursts of one PID that could fill up the maximum transfer unit of a network packet. Thus we need to store and aggregate the packets. For this purpose, we scan 32 successive packets in the FIFO and sort them according to the PID. This allows us—at least in the case of the high bandwidth video PIDs—to fully utilize the MTU. This ensures the desired high throughput. Typically, a packet contains 1378 bytes, 62 bytes in the headers (Ethernet, IPv6 and UDP) and 1316 bytes in seven TS packets.

The network data generation unit then assembles the full network packet from such aggregated TS packets. The basic IPv6 multicast streaming header is based on templates that the software in the CPU core fills in advance. When the network chip accesses the header area over the bypass path, the data is finalized on the fly while being transferred. That means that the transfer unit inserts header fields such as the payload length or the destination IPv6 address into the template placeholders. After the header data it reads the packet data as is from the respective packet buffer.

Besides supporting IPv6 multicast, we made provisions in the PID-info table to enable IPv4 multicast and unicast, too. A unique ID, which can be propagated to the data generation engine, selects a linked list of individual connection data such as a destination MAC, an IPv4 destination address, RTP sequence counters, etc. This allows us to implicitly multiplex different PIDs to one destination address. As a result, NetCeiver also supports traditional streaming protocols and clients in hardware.

All the SoC components, the Microblaze CPU, the memory controller, the PCI module, the transport stream handler, and the network interface occupy only about 50%

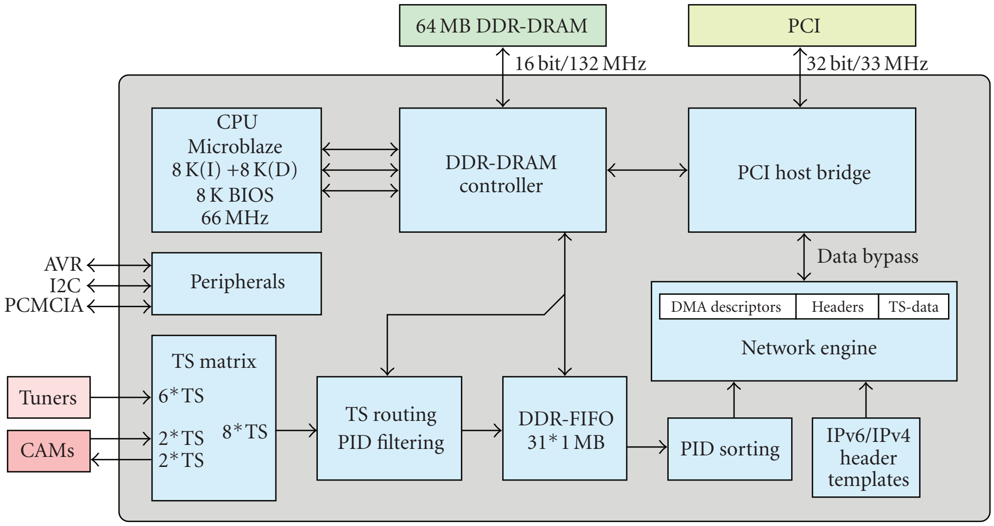


Figure 6: Internal components of the FPGA.

of the FPGA so far. This leaves plenty resources for future extensions.

4. Performance

The described TS handling and network packet generation runs in parallel to the CPU without any CPU intervention. The CPU only configures the IP header templates and updates the PID-info tables for each started or stopped streaming request. Our tests confirmed that the hardware is actually capable of streaming the full content of six tuner transport streams (the aforementioned 35–40 MB/s) without lost packets, although this scenario with all PIDs enabled is unlikely to occur in real life conditions. It merely demonstrates that at least the hardware data processing poses no performance bottlenecks.

For an interactive service, the zapping latency is important. Since we wanted no additional infrastructure like the mentioned GOP-server, our goal was not to differ too much from a locally connected tuner. This aspect is in our opinion very well handled by the dual use of MLDv2 for multicast management *and* tuning requests, as the binary protocol requires only little CPU resources for decoding and generation.

We have measured the time difference between issuing a tune request on a client and the arrival of the first transport stream packets. With 3 clients continuously tuning to random services with multiple PIDs on random DVB-S/S2 transponders, the latency varies between 70 ms and about 2 s with an average of about 0.5 s. The lower zapping times occur when a tuner already has a lock and only the PID filtering needs to be modified. The average latency is mainly determined by the tuning itself and includes the time to acquire a lock (50–200 ms) and the DiseqC-communication to the LNB (about 100 ms). Zapping times over 1s happen only with a probability of less than 10% and are mainly caused by missing a multicast listener report at the server which is sent by the client every second.

Since the NetCeiver uses XML to broadcast its capabilities and internal status, an exemplary experience for XML-handling on slow CPUs was also obtained. The internal status including all tuner parameters and satellite lists can be up to 20 KB in uncompressed XML. This size is caused by the CC/PP syntax with long descriptive tags. This report is generated every 10 s by using the popular libxml2-library facilities. The generation of this document needs about 7 s on the Microblaze. Further analysis showed that about 90% of the time is consumed by the malloc/free memory management of uCLib for the various XML result nodes. Based on this surprising result, the generation frequency was drastically reduced. Although solvable, this example shows that uncritical XML handling with standard frameworks can have a large performance impact on slow CPUs.

5. Conclusion and Outlook

In this paper we have presented the NetCeiver architecture for distributing DVB-like broadcast over IP networks. NetCeiver uses standard protocols such as IPv6 multicast with MLDv2. Our main design goals were zero configuration, scalable streaming to an arbitrary number of clients, and full service transparency for typical DVB receivers. Especially the zero configuration property was considered to be important for home and other non-administered environments.

We have developed a cost-efficient FPGA based hardware that implements our NetCeiver architecture. It allows easily deployment in the consumer market. Moreover, our design is power efficient and allows a very high network performance. We achieve this with a specialized stream handling inside the FPGA.

The NetCeiver architecture is available through a commercial vendor since late 2007. The customer feedback shows that it meets its design goals. The NetCeiver product works as a detachable DVB tuner component of a high-end set-top box. It is also available as a stand-alone head end for inexpensive and lightweight STB clients.

Yet, the development is not finished. In the future, we will extend the NetCeiver core functionality by VOD-services and add multicast security features. Although not as lightweight as the NetCeiver multicast protocol, industry standards like UPnP-AV and the related DVB-IPTV exist and should not be neglected. We plan to integrate the NetCeiver hardware into these frameworks.

Although DVB-IPTV is currently getting more attention, we think that our combined HW/SW design with a nonconventional interpretation of a low-level network protocol can contribute to the still heterogenic IPTV market, as it allows a low- and mid-scale deployment at a very low cost and almost no software changes in a typical DVB set-top box client.

References

[1] IETF, "Real Time Streaming Protocol (RTSP), RFC2326," April 1998, http://tools.ietf.org/html/rfc2326.

[2] IETF, "SDP: Session Description Protocol, RFC2327," April 1998, http://tools.ietf.org/html/rfc2327.

[3] UPnP Forum, "UPnP AV Architecture," December 2008, http://www.upnp.org/standardizeddcps/default.asp.

[4] Wikimedia Foundation, "Wikipedia, P2PTV," January 2009, http://en.wikipedia.org/wiki/P2PTV.

[5] T. Silverston and O. Fourmaux, "P2P IPTV measurement: a comparison study," 2006, http://arxiv.org/pdf/cs/0610133.pdf.

[6] Y.-F. Chen, Y. Huang, R. Jana, et al., "When is P2P technology beneficial for IPTV services," 2007, http://www.nossdav.org/2007/files/file-22-session4-paper1-chen.pdf.

[7] ETSI, "DVB standards," January 2009, http://www.etsi.org/WebSite/Technologies/DVB.aspx.

[8] The advanced television systems committee, "ATSC Standards," January 2009, http://www.atsc.org/standards.

[9] Digital broadcasting experts group, "ISDB-T System," 2008, http://www.dibeg.org.

[10] ETSI, "DAB standards," January 2009, http://www.etsi.org/WebSite/Technologies/DAB.aspx.

[11] Wikipedia, "Mbone," August 2009, http://en.wikipedia.org/wiki/Mbone.

[12] RFC Draft, "Session Announcement Protocol, RFC2974," October 2000, http://tools.ietf.org/html/rfc2974.

[13] RFC Draft, "RTP: a transport protocol for Real-Time Applications, RFC3550," November 1995, http://tools.ietf.org/html/rfc3550.

[14] D. Minoli, *IP Multicast with Applications to IPTV and Mobile DVB-H*, John Wiley & Sons, New York, NY, USA, 2008.

[15] DVB project, "DVB-IPTV 1.4: Transport of MPEG 2 TS Based DVB Services over IP Based Networks (and associated XML) (dTS 102 034 V1.4.1)," January 2009, http://dvb.org/technology/standards/index.xml#internet.

[16] Internet Draft, "Simple Service Discovery Protocol," October 1999, http://tools.ietf.org/html/draft-cai-ssdp-v1-03.

[17] W3C, "SOAP Version 1.2," April 2007, http://www.w3.org/TR/soap12-part1.

[18] Digital Living Network Alliance, "DLNA," September 2009, http://www.dlna.org.

[19] DVB Project, "Specification for service information (SI) in DVB systems (ETSI EN 300 468)," July 2009, http://dvb.org/technology/standards/index.xml#multiplexing.

[20] DVB Project, "MHP 1.2," July 2009, http://www.mhp.org/mhpgem12.htm.

[21] T. Janevski and Z. Vanevski, "Statistical analysis of multicast versus instant channel changing unicast IPTV provisioning," in *Proceedings of the 16th Telecommunications Forum (TELFOR '08)*, Belgrade, Serbia, November 2008.

[22] D. Banodkar, K. K. Ramakrishnan, S. Kalyanaraman, A. Gerber, and O. Spatscheck, "Multicast instant channel change in IPTV systems," in *Proceedings of the 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE '08)*, pp. 6–10, Bangalore, India, January 2008.

[23] RFC draft, "Multicast DNS," September 2008, http://files.multicastdns.org/draft-cheshire-dnsext-multicastdns.txt.

[24] IETF, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6, RFC3810," June 2004, http://tools.ietf.org/html/rfc3810.

[25] IETF, "IP Version 6 Addressing Architecture, RFC4291," February 2006, http://tools.ietf.org/html/rfc4291.

[26] W3C, "Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0," January 2004, http://www.w3.org/TR/CCPP-struct-vocab.

[27] IETF, "Protocol Independent Multicast—Sparse Mode (PIM-SM), RFC2362," June 1998, http://tools.ietf.org/html/rfc2362.

[28] Xilinx Inc., "Spartan-3E FPGA family data sheet," http://www.xilinx.com/support/documentation/index.htm.

[29] Xilinx Inc., "MicroBlaze processor," http://www.xilinx.com/tools/microblaze.htm.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 451897, 10 pages
doi:10.1155/2009/451897

Research Article

Concept for Mobility and Interconnections Aspects in Converged NGN-Based IPTV Architecture

Eugen Mikoczy, Radovan Kadlic, and Pavol Podhradsky

NGNlab, Department of Telecommunications, FEI, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovakia

Correspondence should be addressed to Eugen Mikoczy, mikoczy@ktl.elf.stuba.sk

Received 9 February 2009; Accepted 2 November 2009

Recommended by Alessandro Vanelli-Coralli

The progress and evolution trends in the area of ICT and ICT infrastructure based on convergence processes create new opportunities. Service providers and network operators can provide a wide spectrum of multimedia services and applications to end users. The IPTV services represent a specific group of multimedia services which are in the sphere of interest of the telecommunication technical community but also subscribers. Standardization bodies like ETSI TISPAN or ITU-T have defined standards for NGN-based IPTV architecture and services (IMS and non-IMS). The paper evaluates possibilities and potential architecture for concept of converged NGN IPTV. New vision of the converged NGN IPTV architecture is presented together with proposed enhancements compared to IMS-based IPTV where single converged platform can serve fixed, mobile, or wireless terminals. The concept for IPTV service roaming with mobility support and different interconnection scenarios are discussed with intention to show potential user benefits.

1. Introduction

The TV service started in last century as analogue terrestrial broadcasting service. Digital television as the terrestrial or satellite service started several years ago. The IPTV could be considered as second generation of the digital television.

The standardization institutions like the ITU-T (International Telecommunication Union) and ETSI (European Telecommunications Standards Institute) TISPAN (The Telecoms & Internet converged Services & Protocols for Advanced Networks) began with standardization process focused on convergence processes of the circuits switched and IP networks several years ago. First version of NGN concept was proposed as more voice oriented. The NGN architectures have been designed to provide a wide spectrum of NGN multimedia services and now these technologies also support Internet Protocol Television (IPTV) services. The IPTV is defined [1] as the set of multimedia services such as television/video/audio/text/graphics/data delivered over IP-based networks. NGN IPTV network supports the required level of QoS/QoE, security, interactivity, and reliability.

In early 2008, the ETSI TISPAN NGN in Release 2 defined two main architectural concepts: NGN-dedicated IPTV subsystem (NGN but not IMS solution) and IMS-based IPTV. ITU-T has identified also additional possible NGN-based IPTV architecture called the Converged NGN-based IPTV. Our vision of Converged NGN-based IPTV architecture is presented in the paper. The concept could be considered as evolution of IMS-based IPTV where single platform can be used to deliver IPTV over multiple type of access network (fixed-mobile converged IPTV). Selected scenarios supporting the mobility, interconnection, and service roaming within NGN-based IPTV networks are also presented.

2. NGN-Based IPTV Architecture

In the ETSI TISPAN NGN Release 2 and 3, several specifications of stage 1 (requirements) and stage 2 (architecture) the IPTV integration within TISPAN NGN standards is addressing are

(i) IPTV service requirements [2],

(ii) NGN integrated IPTV subsystem architecture [3] (formally NGN-dedicated IPTV in R2),

(iii) IMS support for IPTV architecture [4].

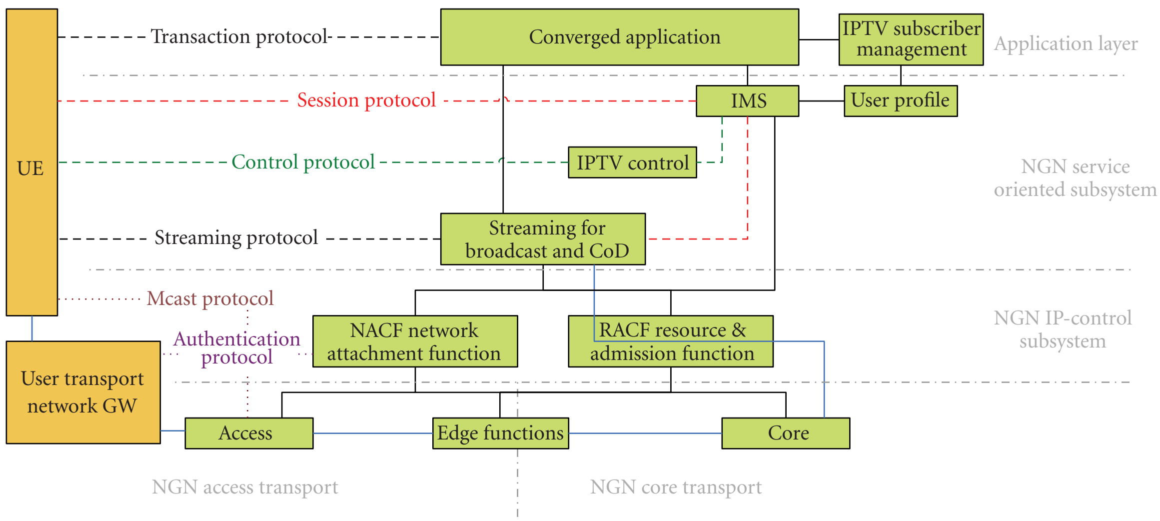


Figure 1: ITU-T concept of Converged NGN-based IPTV architecture [1].

The specifications about the implementation of the IPTV functions, interfaces, procedures, protocol recommendations (stage 3) have been finalized and more information for NGN-dedicated IPTV can be found in ETSI TS 183 064 [5] or for IMS-based IPTV in ETSI TS 183 063 [6]. The ITU-T NGN-based IPTV architecture is specified in recommendation of ITU-T Y.1910 named IPTV functional architecture [1]. Also other institutions like ATIS or Open IPTV forum are working on NGN-based IPTV.

IMS-based IPTV is new concept where first implementations based on ETSI TISPAN R2 have been already tested by MSF in 2008. ETSI is hosted in 2009 also IPTV plugtest where vendors and operators can test real interoperability of IMS-based IPTV servers and platforms (based on TISPAN R2 [6]). The most comprehensive set of new IPTV services will be implemented based on TISPAN Release 3 NGN-based IPTV [4].

Several scenarios are possible and really depend only on the operator's choice in which solution and migration scenarios are selected (if any) [7, 8]. The above mentioned ITU-T IPTV has been proposing the converged application framework where the non-IMS IPTV merges with the IMS-based IPTV architecture. But in this case just as purely unity of all elements supporting all interfaces from both architectures that make no sense from the complexity perspective, but also for both architectures if used in parallel, they are able to provide similar services (Figure 1).

ETSI TISPAN in release 3 (in [3] in informative annex) proposed possible migration scenarios between non-NGN, NGN non-IMS and NGN IMS-based architecture. There is also mentioned the last evolution step called converged NGN-based IPTV (this concept is not described in Release 3 because for now it is out of the scope). The following sections will describe potential concept of such architecture with additional goal to describe also the way of adaptation with multiple types of content sources, but also with multiple access and distribution networks converged to a single functional architecture.

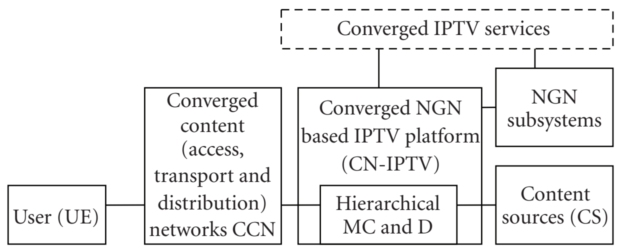


Figure 2: High-level view to IPTV domains.

3. Concept of Converged NGN-Based IPTV

The concept of converged NGN-based IPTV can be split into several subsystems and domains where each one plays an important role to provide the converged approach of multiservice NGN architecture (Figure 2):

(i) content sources,

(ii) converged NGN-based IPTV architecture—overall concept, description of functionalities,

(iii) concept of hierarchical content control and delivery subsystem,

(iv) content transport and distribution networks,

(v) converged IPTV services.

3.1. Functional Architecture for Converged NGN-Based IPTV. The proposed Converged NGN-based IPTV architecture (CN-IPTV) is an evolution of the IMS-based IPTV from ETSI TISPAN Release 3. We combined IPTV service control function (ISCF) for service control which can use the IMS for specific cases but not all signaling traffics need to pass core IMS (which improve performance and shorter delays). The concept is including media control and delivery organized in hierarchical MC&D architecture that is extended with specialized types of MDFs [1]. Additionally we elaborate possible integration with some DVB [9], 3GPP UMTS, and OMA BCAST component that can extend the mobility capabilities for IPTV services. DVB specified how DVB services could be provided over IP networks in specification called DVB-IPTV (formally DVB-IPI) and discussed also possible compatibility with TISPAN NGN-based IPTV [10]. Open Mobile Alliance defined service enabler primarily for mobile-related broadcasted/multicasted services (OMA BCAST [11]) but it could be used also as a base for adaptation of several types of access technologies to our proposed concept for converged NGN-based IPTV. OMA BCAST 1.1 explains the adaptation of the DVB-H, 3GPP MBMS, and WiMAX access technologies.

The proposed converged NGN-based IPTV architecture consists from following elements (Figure 3):

(i) ICAF: IPTV-Converged Application Functions,

(ii) SDSF: Service Discovery and Selection Functions,

(iii) ISCF: IPTV Service Control Function,

(iv) UPSF: User Profile Server Function (existing one),

(v) IMCF: IPTV Media Control Function,

(vi) IMDF: IPTV Media Delivery Function (P-Proxy, S-Serving, I-Interconnection),

(vii) IMS: IP Multimedia subsystem (P-/S-/ICSCF),

(viii) NASS: Network Attachment Subsystem,

(vx) RACS: Resource and Admission Control Subsystem.

Any IPTV architecture is not fully completed without other functionalities which we have also included in the proposed concept:

(i) IPTV supporting function (e.g., content preparation and manipulation),

(ii) IPTV management functions (e.g., content management),

(iii) IPTV security functions (e.g., content protection, IPTV service protection),

(iv) IPTV charging (based on NGN charging for online/offline charging but enhanced for IPTV specific scenarios),

(v) Interworking or interfacing with other NGN subsystems.

Three types of MDFs are introduced by hierarchical MC&D. MDFs are split into multiple specific functionalities with three architecture components described [8] as follows.

Interconnection-IPTV Media Delivery Function (I-IMDF): this element handles the media import and ingress of content from multiple content sources (ingress of media, metadata, content provider information, and interconnection to external domains):

(i) IPTV Headend or from content providers/originators or broadcasters,

(ii) from other IPTV service providers in case of interconnection or roaming or as offer of the content from service provider playing a role of content aggregator or Media Content Distribution Network (MCDN),

(iii) from the Internet sources like the Web-based TV or from the end users like user generated content,

(iv) the IMDF need to hide the IPTV service provider infrastructure for external domains but also provide necessary functionality to interconnect to heterogeneous content sources (which can hold a variety of coding, transport, signaling schemas) and convert to content/metadata/signaling to formats supported by Converged NGN-based IPTV.

Serving-IPTV Media Delivery Function (S-IMDF): this element handles the processing of contents (e.g., encoding, content protection and transcoding) and it is also responsible for the storage of contents and metadata as well as the propagation of content information. S-IMDF provides centralized oriented services such as Content on demand for long tail content (less popular content) or recording/storing of user independent content (n-PVR or Time shifted TV or Near CoD).

Primary-IPTV Media Delivery Function (P-IMDF): this element is the primary contact point of the users which provides also the streaming and downloading functionalities for all IPTV services according to the required quality, format, and the type of casting (multi/uni/broad-casting) for particular user's end device and network accessing to P-IMDF (personalization of user specific content delivery). This element could also store the most frequently accessed CoD assets or user's specific contents (specific user n-PVRs, user generated content). The P-IMDF could be responsible for the adaptation of IPTV service delivery to other access technologies or distribution networks.

(i) Preferred way is that the P-IMDFs are located as near as possible to UE, for example, near to edge of the network. These elements can be combined with specific elements of access network (e.g., in case of using OMA BCAST and 3GPP MBMS also with integrated control elements, in case of MBMS, e.g.). P-IMDFs can also integrate Broadcast-Multicast Service Centre (BM-SC). The BM-SC provides functions needed for user service provisioning and content delivery in MBMS capable UMTS. P-IMDF integration with 3GPP MBMS and PSS helps to provide IPTV services delivery over exiting UMTS mobile network.

(ii) P-IMDF can also support mobility and seamless handover between different technologies.

(iii) P-IMDF may require to transcode or adapt the content to required bitrate, codec, or content encapsulation to specific transporting technologies.

(iv) It can be used as security elements for the content protection (e.g., digital rights management and content encoding).

The element responsible for the service discovery and selection functions (SDSFs) could support multiple formats

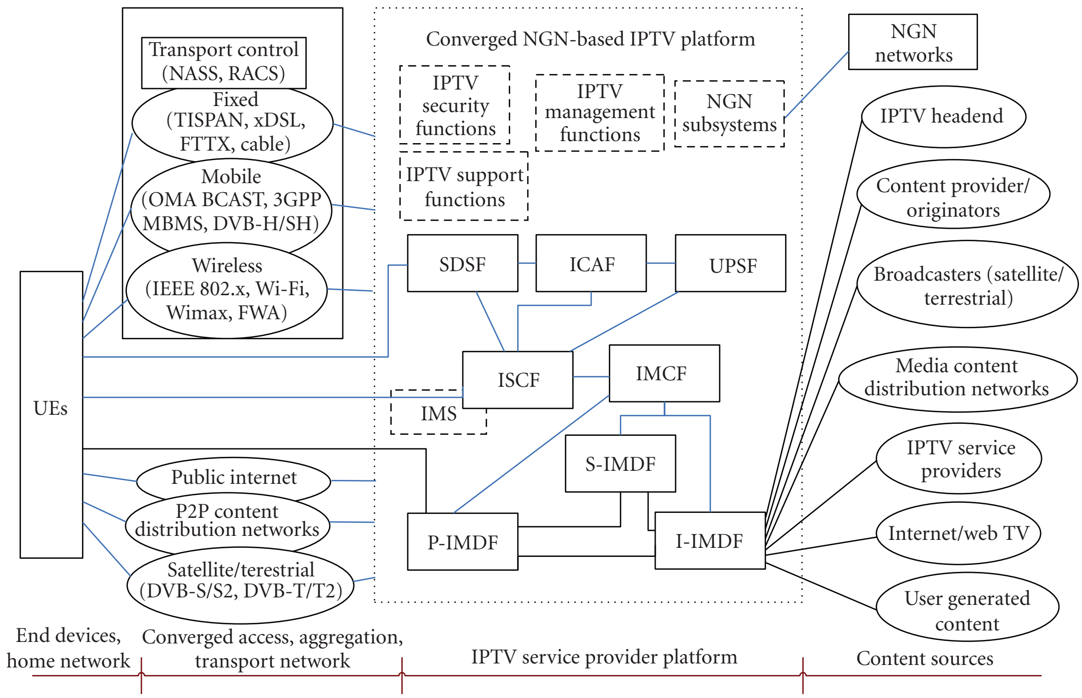


Figure 3: Proposed conceptual architecture for Converged NGN-based IPTV.

and mechanisms. But the main enhancement in proposed architecture is the SDSF's potential to aggregate metadata information from multiple sources (e.g., content provider, electronic program guide provider, internet, broadcasted service information) and provide them everywhere to the UE in personalized manner and allows them to integrate with other relevant information (presence, statistics, recommendations, etc.), too.

The last but not least element is (for sure) the IPTV converged application function (ICAF) which could provide combinational services and converged services with enhanced service logic and service orchestration. The ICAF can interact with other NGN application servers and subsystems and can be used for personalized service behavior based on user's preferences and settings.

The enhanced service control entity called IPTV Service Control Function (ISCF) is responsible (in conjunction with core IMS) for service control and for providing basic IPTV services and service interaction and also supports mobility on service control and application layers. Similar to other NGN-based IPTVs also the converged one needs to ensure relevant resource allocation and QoS handling (using existing interfaces to transport control). Main goal of Converged NGN-based IPTV is to provide the flexible service provider platform for IPTV services (and any NGN services as well) that may deliver personalized services over multiple access network with nomadic or seamless mobility. But we have to differentiate the IPTV service provider access infrastructure with guarantied quality (fixed technology as xDSL or FTTx, mobile technology like UMTS with IMS-based MBMS/PSS or wireless access technologies like WiFi or Wimax) from other additional distribution possibilities, like public internet (without QoS and unpredictable conditions), terrestrial/satellite distribution (mainly unidirectional for broadcasting with other technologies used for interaction, back channel signaling or unicast services), or the P2P content distribution network.

3.2. Protocol Stack for Converged NGN-Based IPTV. The protocol stack for such a complex system as Converged NGN-based IPTV platform will include almost all existing NGN protocols (Figure 4). Common layer for all of them is Internet Protocol (IP) which should be carried over different physical and data link technologies. Additionally, to IP-based NGN protocol family we can transmit the media over other type of broadcasting networks like DVB-T or DVB-S (where adaptation and other ways for interactive services and back channels for signaling may be needed).

4. Interconnection Models and Roaming Scenarios

In general, several scenarios with different issues that have to be solved for IMS-based IPTV or its evolution in form of Converged NGN-based IPTV for service roaming are possible. The network of the provider where the subscriber has made a contract for some service will be called the Home Network. The network of the provider where subscriber is just connected during roaming will be called the Visited Network. Main goal to support interconnection and roaming scenarios for IPTV is that given operator can interconnect other operators network over standardize interfaces to provide roaming or aggregate content and also IPTV services

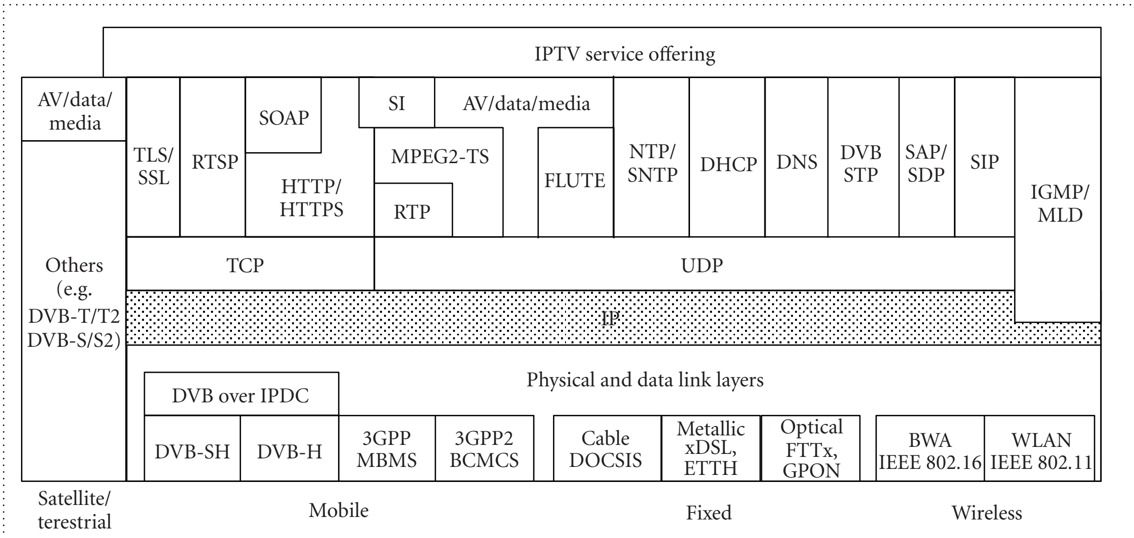


Figure 4: Protocol stack for Converged NGN-based IPTV platform.

which he could not provide by own platform (acquire them from partner operator). From user perspective main advantage is accessing his services outside of the Home Network or on move. The IPTV roaming also allows more variety of accessible local content (present only in Visited Network).

We can recognize at least the following basic scenarios for roaming and interconnection to home network (defined in [4] Annex H: Interconnection Models to support of Mobility Capabilities):

(i) remote data access to IPTV/content provider,

(ii) IMS interconnect to home IPTV provider,

(iii) visited—home network roaming between IPTV providers (served only from home network),

(iv) visited—home network roaming between IPTV providers (served from home or visited network).

There are several aspects which are crucial to solve the issues related to interconnection (from technical but also from service level agreement point of view):

(i) services which could be provided in roaming, if possible by partial coverage within Visited Network,

(ii) support for unicast, multicast, bandwidth negotiation, QoS issues,

(iii) model for storage, streaming, proxy of content between both domains,

(iv) service discovery and selection issues appearing from cross domain transport,

(v) content management (and generally IPTV management aspects),

(vi) IPTV security aspects,

(vii) roaming behavior,

(viii) content adaptation to support visited network conditions,

(ix) multi-multicast versus transcoding for multiple access technology,

(x) codec/transport/interconnection agreed between providers,

(xi) multi-multicast versus transcoding for multiple access technology,

(xii) codec/transport/interconnection agreed between providers.

4.1. Scenarios for Interconnect in NGN-Based IPTV

Scenario A: IMS-IPTV Subscriber Visiting Packet/Switched Provider without IMS. Because the Visited Network has no functions of the IMS, the subscriber must use some remote data IP connection (e.g., VPN or secure remote access) to connect to his Home Network. Over this connection there can be transferred all media and signaling to the subscriber directly from his Home Network. Because such a connection should go over any IP network (also best effort) also without resource reservation mechanisms, no QoS can be really ensured. This scenario may be used for IPTV services without guaranty like remote control or download of n-PVR recordings.

Scenario B: IMS-IPTV Subscriber Visiting IMS Provider without IMS-IPTV Solution. This scenario is the simplest one with IMS involved. Subscriber will use all IPTV elements from his Home Network. The elements of visited network must provide just few functions.

(i) Check if certain subscriber is allowed to roam (via IMS roaming) to his visited network (Network-to-Network check).

(ii) Collect and watch all data necessary for charging.

(iii) Control limitation of its own network and apply limits to subscriber based on roaming policy agreed by Hosted and Visited Network operators.

The quality of the IPTV service for the end customer is the same one as in home network, but no reuse of local media resources for the provider is possible (because in visited network there is no NGN-based IPTV platform).

Scenario C: IMS-IPTV Subscriber Visiting IMS Provider with IMS-IPTV Solution (Only Home Served). Additionally, to the previous scenario the following one expects the IMS-based IPTV platform in Visited Network; however all content and services are delivered only from Home Network. From the Visited Network, there could be used only some supporting elements like SDSF or IMDF. Useful functions are that ones like transcoding and content adaptation to adapt to media with parameters required in Visited Network.

Scenario D: IMS-IPTV Subscriber Visiting IMS Provider with IMS-IPTV Solution (Home or Partially Visited IPTV Platform Served). If there is an IPTV solution in Visited Network, it is possible that it has similar content of CoD, Bcast, PVR service. It is useful to use this content from local resource for the visiting subscriber. The content which is available in Visited Network is not needed to be transferred over an interconnection network. Therefore the interconnect bandwidth can be saved. For this purpose both providers must agree on the same identification of content, sharing service discovery and selection information, content security, billing clearing, and for roaming agreements (e.g., QoS, Service Level Agreement). It is possible for the provider to reuse the local media resources from Visited Network.

In the following section we will focus only on the most important scenarios B, C, and D which involve the IMS and these scenarios are therefore also applicable for Converged NGN-based IPTV.

4.1.1. Scenario B: IMS-IPTV Subscriber Visiting IMS Provider without IMS-IPTV Solution. The model of interconnection with the Visited Network within the Core IMS involved is used only for IMS type of roaming to access home IPTV service provider platform [4] (Figure 5).

The UE can request the IPTV services from the Home Network when connected the Visited Network. The Core IMS in the Visited Network or Home Network can request resources from the RACS in the Visited Network through the interface Gq'. The UE can be attached to the Visited Network through the interface e1 so that the NASS in the Visited Network can assign the IP address for the UE and discover the address of P-CSCF in the Visited Network.

The Core IMS in the Visited Network can transfer the IPTV service request from the UE through the interface Gm to the home Core IMS through the Ic connection of IBCFs (Intermediate Breakout Control Functions). The UE can connect to home SDSF through the interface Xa. To get configure parameters in ISCF through the interface Ut. Service initiation, control as well as media delivery, the UE can connect over existing interfaces the Home Network.

4.1.2. Scenarios C/D: IMS-IPTV Subscriber Visiting IMS Provider with IMS-IPTV Solution. The model of interconnection with the Visited Network based on the Core IMS and Converged NGN-based IPTV infrastructure (or IMS-based IPTV) is used for most advanced type of roaming to access the home IPTV service provider platform (also with CN-IPTV) where all services are provided only from Home Network (Scenario C) or from elements from both Home/Visited networks as shown on Figure 6 (with using IMDFs in Visited Network and I-IMDFs for interconnect and content adaptation on edge of both domain).

The UE can attach to the Visited Network through the interface e1/e3 towards the NASS in the Visited Network that can assign the IP address for the UE and discover the address of P-CSCF in the Visited Network. The Core IMS in the Visited Network or Home Network can request resources from the RACS in the Visited Network through the interface Gq'. The Core IMS in the Visited Network transfers the entire IPTV service request from the UE through the interface Ic to the Core IMS in the Home Network through the connection of their own IBCF (Intermediate Breakout Control Function).

The UE is connected to the Home SDSF (which can provide information about available content in this particular roaming case, because some of content could be restricted for roaming) to acquire service selection information through the interface Xa. The UE can be connected to the SCF in the Home IPTV network to configure parameters through the interface Ut (not shown on figure). The Core IMSs in both networks connect their own UPSF to get the user. But Home UPSF is responsible for service initiation authorization and user profile information required for personalization of IPTV services. The service initiation and control is provided over existing interfaces from Home Network (Gm to IMS and then other to interfaces towards ISCF, ICAF, or IMCF). The setup of media control and delivery channels is responsibile for the home ISCF with IMCF but media delivery itself could be provided from IMDFs from Home or Visited Network (we elaborate specific cases in Section 4.2) via Xc (for control) or Xd (for delivery). The End User has to subscribe the roaming services at the Home Network before he/she moves to the Visited Network. The exact details and mechanism agreed between IPTV service providers have to be negotiated in advance together with interconnect agreements, policy rules, and most probably also service level agreements (SLAs) to assure QoS on interconnect.

Following steps are proposed as potential procedures for NGN-IPTV interconnection with the Visited Network with CN-IPTV is presented (Figure 7).

(1) Network Attachment: in this step the UE attaches to the network (with NASS, receiving IP configuration, P-CSCF address discovery of Visited Network, etc.).

(2) UE performs IMS Registration in Visited Network. P-CSCF within Core IMS in Visited Network submits the registration request to the I-CSCF within the Core IMS in the Home Network.

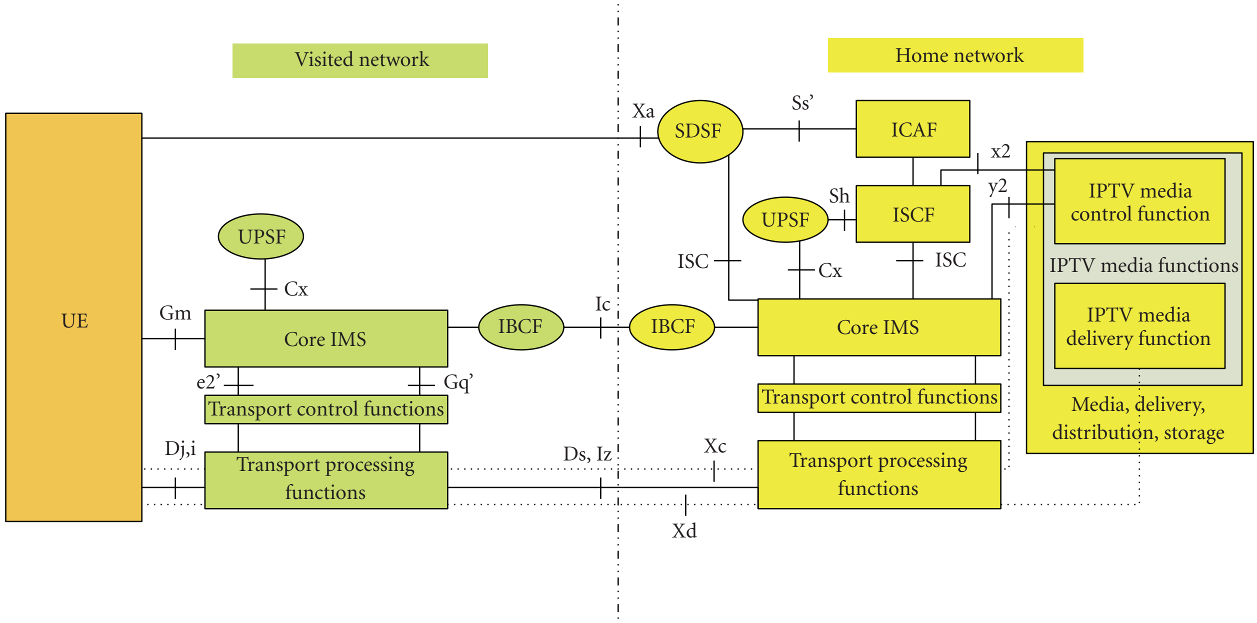


Figure 5: The model of interconnection with the Visited Network within Core IMS involved.

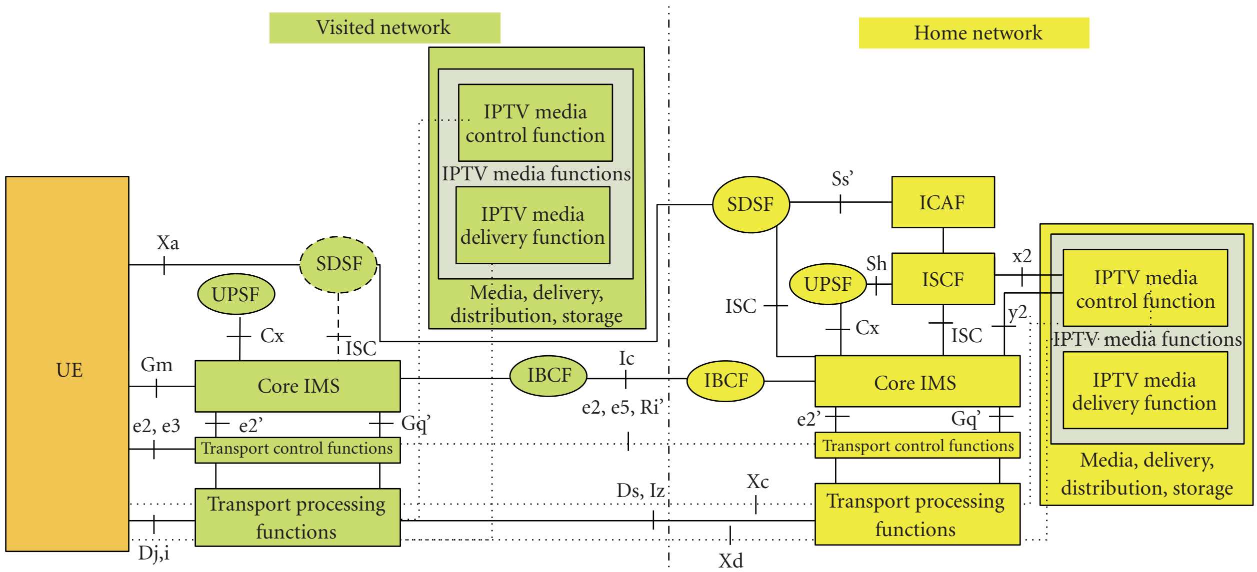


Figure 6: Model of interconnection with the Visited Network with CN-IPTV.

(3) Core IMS in the Visited Network returns parameters (e.g., P-CSCF address of Home Network, SDSF) to the UE,

(4) Home and visited network elements can require some exchanges of relevant information before allowing user and his/her terminal attached to IPTV services and select or initiate any service,

(5-6) UE performs service discovery and service selection contacting home SDSF directly or via visited SDSF (to compile a set of services and metadata which can be provided to users).

(7) User initiated via UE requests for selected service and sends it via visited IMS to home IMS and ISCF where it is processed (could be authorized by home UPSF (8), and also apply service logic with interaction with ICAF (9).

(10) After successful authorization ISCF initiates resource reservation in Home and/or Visited network using IMS mechanisms available there towards RACS (11) and after then setup also control and delivery channels to UE (12). Finally Home ISCF sends response to UE (13) and initiates content delivery through IMCF via appropriate IMDFs (14) to UE (15-16).

Accessing Services. The SDSF in Visited Network should collect information from Visited Network about services

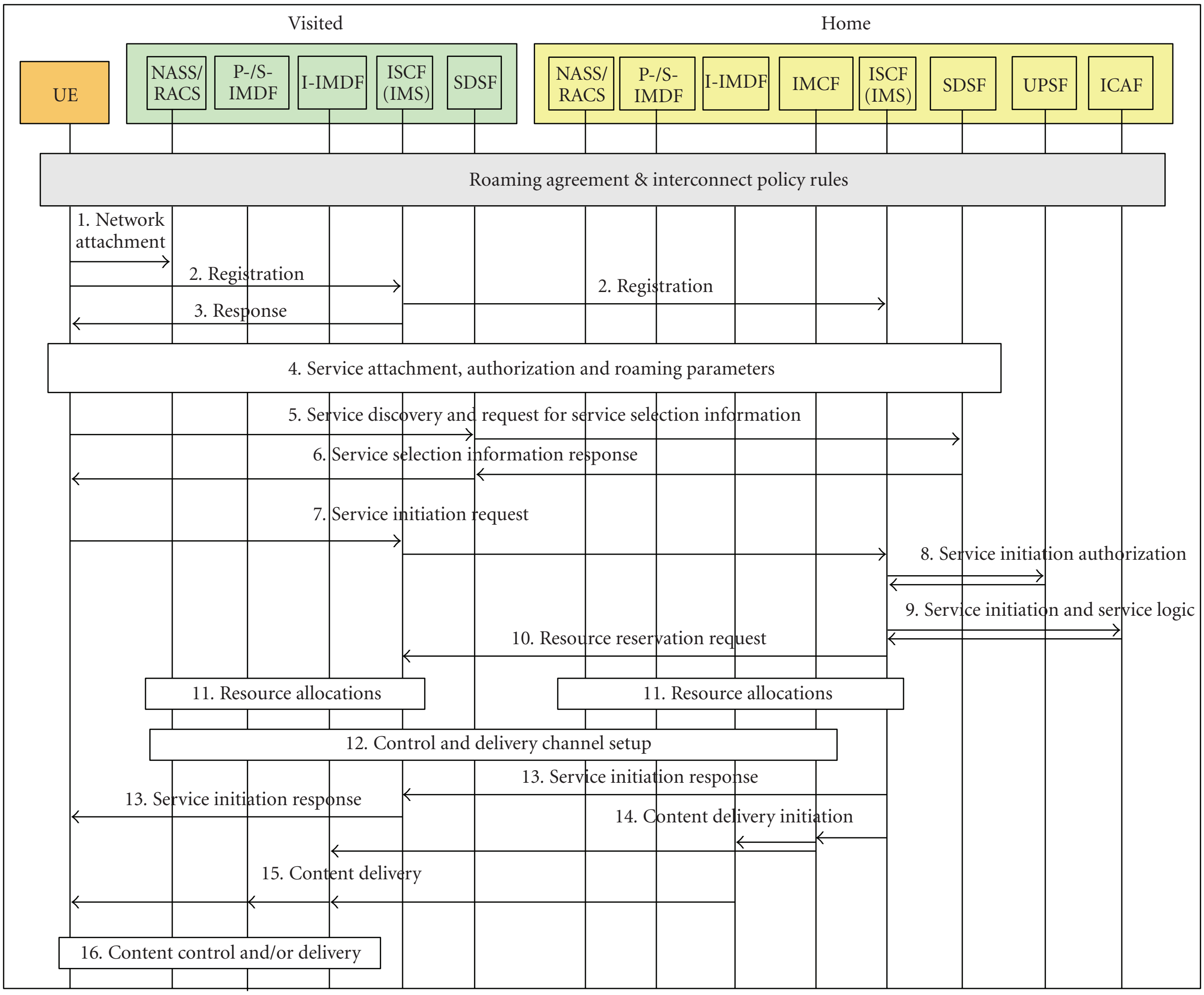


Figure 7: Procedures for interconnection with the Visited Network with CN-IPTV.

which subscriber can use in roaming but it also has access to relevant service information from Home Network. If this service is offered in Visited Network, the subscriber will get it directly from the Media distribution and Storage Subsystem of the Visited Network. If such service is not present in Visited Network, the UE will use it from the Media distribution and Storage Subsystem of Home network.

Since the UE or Visited Network can have some specific restrictions to format of content, the IMDFs in Visited Network must be able to transcode any stream to requested format and serve it to the UE. If there is no conjunction in functions of the Visited Network, I-IMDF, and UE possibilities, this media cannot be delivered to the customer. Therefore it is very useful for the IPTV providers in roaming relation to offer all the content in several bandwidth profiles (Mobile/WEB, SD, HD) with standardized common used codecs (3GPP, FLV, H.264, VC-1, MPEG2) preferably the same (to avoid transcoding). All UEs should support agreed and standardized codecs used in the Home Networks or during roaming in Visited Network.

Table 1: Examples of codec profiles for IPTV.

Profile	Video codec	Audio codec	Container	Bandwidth
Mobile	H.263	AAC	3GPP	256 kbit/s
WEB	H.264, baseline	mp4a	mp4	768 kbit/s
SD	H.264, main	MP2/AAC	MPEG2 TS	3.5 Mbit/s
HD	H.264, high	MP2/DD+	MPEG2 TS	12 Mbit/s

Table 1 is just example, but worldwide discussion and agreement should be found and some global table should be proposed for interconnection agreements. This will help all service providers to save costs for technology and will increase interconnection possibilities.

To save the resources in the provider's networks, it is useful to define a closed group of bandwidths and codes used for the streaming, for example, as follows [12].

Content Elements from Many Sources. If there is a service in the Visited Network which is not the same as that one, which was requested from the UE, it is possible to combine content/service that partially get this content/service from local source and from Home Network source. If the video stream with some source (e.g., CNN TV) is present in the Visited Network, but language is different as requested one, the MDF can request only this elementary language. This process requires global synchronization information, which has to be put in stream (e.g., RTP timestamp).

4.2. Interconnect from Service Perspective. In the following section are discussed technical scenarios for selected IPTV services like Linear TV, Live TV with Trickplay, Timeshift TV, and COD. We analyze scenarios and advantages from operator perspective and also user perspective especially when Visited network resources could be reused during roaming. Additionally two specific use cases for Advance PVR and UGC are explained as potentially attractive user services in roaming case. The main purpose for support of service roaming and terminal mobility is to enable to users the access to most of his personalized IPTV services and various contents from any terminal on the move with fixed terminal (e.g., nomadic mobility for moving STB from one location to other), portable devices (e.g., game console, laptops), or mobile (e.g., PDA, UMTS phone with DVB-H or MBMS/PSS).

4.2.1. Linear/Broadcast TV. Every channel of both operators must have a common channel identificator or at least synchronized metadata in both SDSFs. If the users can access some live channel which is present in visited network in the same quality, the SDSF must provide service selection information from the Visited Network. The provider of the Visited Network can agree with the Home Network provider that channels from the Visited Network can be offered to the subscriber (providing more channel); however, they are not a part of subscriber offer in the Home Network (e.g., some local channels provided for free or paid additionally to monthly fee). There should be other policies and addressing for multicast-based services in Home and Visited network or interconnect should not support multicast. In these cases I-IMDF from home network can adapt service to unicast and I-IMDF in Visited transform address back to multicast or deliver media to UE by unicast.

4.2.2. Trickplay of Live TV Stored in Network. In the case of Live TV channel with trickplay it is reasonable to make use of the local resource from the Visited Network. Otherwise the operators must agree which network should be used to store the data needed for trickplay (Visited/Home). If the Visited Network has not enough space for trickplaying another channel or the operators of the Visited Network do not want to spend this space, it is necessary to offer the service from the Home Network. This solution spends the bandwidth on the interconnect interface. The P-IMDF in Visited Network asks the Home Network to deliver the stream for the trickplay recording and then the trickplay session will start to serve the unicast stream from Visited P-IMDF to the UE.

4.2.3. Timeshift TV Service Served by Network. In the case of a channel which should be timeshifted (the channel with timeshift service also provided in the Visited Network), it is possible to use this resource from the Visited Network (caching/timeshifted in Visited PMDFs). Otherwise the operator of the Visited Network must store a huge amount of content from all the timeshifted channels from the Home Network operator offer, what can be unfeasible and wasting of resources and a subject for legislative problems. Therefore it is probable that this service will be offered from the Home Network. This will spend the bandwidth on the interconnect interface and can denied user request for timeshifting in roaming scenario.

4.2.4. CoD Service. Content on Demand (CoD) is very simple from the technical point of view, because it is defacto a catalogue and database of content (e.g., movie files/assets) and pure unicast delivery. If the UE searches CoD catalogue (e.g., in SDSF) and chooses the content which is accessible from the Visited Network, the UE will get it from this resource and do not need to consume the interconnect bandwidth. Otherwise content will be provided from Home Network IMDFs. If there are many visiting UEs from the same network and there is some popular content in it, it is useful to cache it anyway in the Visited Network. If there is any need for some adaptation of the content, this must be done by the transcoding in IMDF in prior to sending media to the UE over Visited Network.

4.2.5. Advanced PVR. If the UE requests recording (PVR—personal video recording) of the content which is already stored in the Visited Network storage, it should be served from this resource. If there is not such a content available, the UE will get the content from the Home Network.

The UE requests to record a channel, it should be done in the Home Network, because there is high probability of subscriber's presence in the Home Network, and therefore it is playback in this network.

In special case there are channels offered only in Visited Network, here the nPVR should be done in Visited Network and playback should be possible only in this network. In the case of client-based PVR (c-PVR) it is independent from roaming because every time it is stored in UE.

We also specified specific use case where the combination of both PVRs (c-PVR with combination of n-PVR) could be provided called Advanced or Hybrid PVR.

Multiple scenarios or situations when client PVR is disabled to record show that (e.g., end device is disconnected in time of scheduled recording or network parameters are not sufficient to stream or missing capacity of storage in local device) IPTV solution should record in network and later deliver to end device. In the case of different end device capabilities (resolution, encoding, etc.), the record by nPVR can be done in several formats but appropriate record will be distributed to end device storage to later preview.

4.2.6. User Generated Content. All user contents must be stored in the Home Network. If the user wants to upload or upstream some content, the UE will open the upload/upstream channel to his home storage in P-IMDF in the Home Network and stores it there. The playback of this content should be done from the Home Network, too. Handling of popular content should be done the same way as described in CoD. Advantage of the supporting UGC in roaming scenarios is specially when user would like share, for example, holiday videos or content from Visited Network as UGC for friends in Home Network (if interconnect allowed).

5. Conclusion

Available services, their quality, reliability, usability, and conformability are the main drivers from the user to select service of the content provider, of course apart from the price. Standardized IPTV solution could enable the IPTV providers operate IPTV services with better utilization and in cost effective manner.

Only standardized solutions could provide the interoperability among solutions, operators, and end devices. Next step could be the interconnection of operators, and IPTV service providers in order to provide user's mobility and roaming across multiple domains and access technologies.

The paper has presented the concept of Converged NGN-based IPTV which aggregates the content from multiple sources and provides it to user's end devices over various access technologies and terminals. Some scenarios explaining the mobility, interconnection, and subscriber roaming scenarios within Converged NGN-based IPTV networks are also presented in the paper. Stress has been given to IPTV services behavior in interconnect scenarios and implication in order to use them in an effective way in roaming.

The IPTV interconnection is one of the actual topics in ETSI TISPAN where the work on technical report which has to analyze possible mapping with other IPTV systems, interconnection issues and roaming scenarios, and hybrid concept has been just started.

Acknowledgments

The first author is a senior designer in Slovak Telekom where he has been actively participating on R&D project ScaleNet and ETSI TISPAN NGN Release 2/3 standardization with more than 80 agreed contributions included in several TISPAN IPTV-related specifications (e.g., IPTV-related work items WI0005, WI2048, WI2049, WI3127, WI3137, WI7029, WI1059, WI2070, 2074, WI2079, and WI3208). Some parts of this paper are based on contributions of an author within ETSI TISPAN. This paper also presents some of the results from participation in various research projects at Slovak University of Technology in Bratislava such as NGNlab project [13], European CELTIC EUREKA project Netlab [14], Slovak National research projects: AV project 4/0019/07: Converged technologies for next generation networks (NGNs), and Slovak National basic research projects VEGA 1/0720/09 and VEGA 1/4084/07.

References

[1] Recommendation ITU-T Y.1910 (09/2008), "IPTV functional architecture," ITU-T, 2008.

[2] ETSI, "TISPAN; service layer requirements to integrate NGN services and IPTV," TS 181 016 V2.0.0 (2007-11), ETSI, Sophia Antipolis Cedex, France, 2007.

[3] Draft ETSI, "TISPAN; IPTV architecture; NGN integrated subsystem for IPTV functions," TS 182 028 V3.2.6 (2009-06), ETSI, Sophia Antipolis Cedex, France, 2009.

[4] Draft ETSI, "TISPAN; IPTV architecture; IPTV functions supported by the IMS subsystem," RTS 182 027 V3.2.6 (2009-06), ETSI, Sophia-Antipolis Cedex, France, 2009.

[5] ETSI, "TISPAN; dedicated IPTV subsystem stage 3 specification," Technical Specification 183 064 V2.1.1 (2008-10), ETSI, Sophia Antipolis Cedex, France, 2008.

[6] ETSI, "TISPAN; IMS based IPTV stage 3 specification," TS 183 063 V2.4.2 (2009-06), ETSI, Sophia Antipolis Cedex, France, 2009.

[7] 18bTD221r2_WI02074_NGN_Integrated_IPTV_migration, TISPAN 18 bis contribution, NGN integrated IPTV—architecture & migration scenarios, ETSI, 2008.

[8] E. Mikoczy, D. Sivchenko, B. Xu, and J. I. Moreno, "IPTV systems, standards and architectures—part II: IPTV services over IMS: architecture and standardization," *IEEE Communications Magazine*, vol. 46, no. 5, pp. 128–135, 2008.

[9] ETSI, "DVB; transport of MPEG 2 based DVB services over IP based networks," Technical Specification 102 034 V1.2.1, ETSI, Sophia Antipolis Cedex, France, 2006.

[10] DVB Document A128, "DVB-IP Phase 1.3 in the context of ETSI TISPAN NGN, DVB," September 2008.

[11] OMA draft OMA-AD-BCAST-V1_0-20081209-C, "BCAST Mobile Broadcast Services Architecture," Candidate Version 1.0—2008, Open Mobile Alliance, 2008.

[12] ETSI, "DVB; specification for the use of video and audio coding in DVB services delivered directly over IP protocols," Technical Specification 102 005 V1.3.1, ETSI, Sophia Antipolis Cedex, France, 2007.

[13] NGNlab—NGN laboratory, Slovak University of Technology in Bratislava, http://www.ngnlab.eu/.

[14] "NetLab: Use Cases for Interconnected Testbeds and Living Labs," http://www.celtic-initiative.org//Projects/NETLAB/default.asp.

Hindawi Publishing Corporation
International Journal of Digital Multimedia Broadcasting
Volume 2009, Article ID 453471, 21 pages
doi:10.1155/2009/453471

Research Article

P2P and MPEG FGS Encoding: A Good Recipe for Multipoint Video Transmission on the Internet

Alfio Lombardo, Diego Reforgiato, and Giovanni Schembra

Dipartimento di Ingegneria Informatica e delle Telecomunicazioni, University of Catania, V.le A. Doria 6, 95125 Catania, Italy

Correspondence should be addressed to Diego Reforgiato, diegoref@diit.unict.it

Received 6 March 2009; Accepted 4 June 2009

Recommended by Robert Briskman

In the last years Peer-to-Peer (P2P) systems have gained ground for content sharing between communities, determining a real revolution on the Internet. The characteristics of P2P systems make them a very good choice for multimedia content distribution over IP networks. However, although P2P technology gives new opportunities to define an efficient multimedia streaming application, at the same time it involves a set of technical challenges and issues due to the best-effort service offered by the Internet and its dynamic and heterogeneous nature. The most of existent protocols for video communications over P2P mainly focus on tree topology maintenance, without paying any attention to the encoding problem. The idea of this paper is to propose a multipoint video broadcast framework over a heterogeneous content distribution P2P network. In the proposed system the source generates the video flow by using an MPEG-4/FGS encoder, in such a way that no losses occur at the Baselayer stream even in the presence of short-term bandwidth fluctuations. Although in the past the FGS was not employed due to its encoding complexity, today, thanks to advances in hardware technology, we were able to develop an MPEG-4/FGS encoder on low-cost PCs which turned out to be more feasible and appealing for its flexibility. The FGS layer is sent together with the Base layer, but with a lower priority. The source uses a rate controller to regulate the encoding rate of the Base layer. To this aim, a protocol is defined in order to provide the source with information related to the most stringent bottleneck link on the overlay network. A technique to reorganize the content distribution tree is proposed and discussed. To evaluate the performance of the proposed framework a case study is introduced; improvements obtained with respect to several reference cases where FGS is not applied are also shown.

1. Introduction

In recent years, with the widespread deployment of broadband access, broadcast video transmission over the Internet is becoming increasingly popular [1]. The main problem of such a kind of systems is to realize multipoint communication on the current Internet that does not natively support it. In the first generation of such systems the most widely used approach has been the employment of multiple parallel unicast streaming. According to this approach, multimedia data are transmitted by the source to each receiver in a point-to-point fashion: whenever a new client accesses the service, a dedicated stream is allocated until the end of it connection. However, since this approach requires a separate streaming bandwidth for each client, the multimedia source and the consumed network bandwidth resources inevitably grow linearly with the user population. Therefore this approach is not scalable with the number of users, and thus unfeasible in video broadcast scenarios, where the number of users is very high.

The first solution to the scalability problem was the application of IP network multicasting. In this way, unlike multiple parallel unicast transmission, a multicast multimedia stream can be shared by more than one receiver. The network switches/routers automatically replicate the multicast video for multiple receivers without adding any extra streaming workload on the multimedia server. However, IP multicast is not widely deployed, mainly due to practical and political issues. For example, multicast is not available in many low-cost Internet access points, like domestic ADSL.

In the meanwhile, Peer-to-Peer (P2P) systems have gained ground for content sharing between communities, determining a real revolution on the Internet. Unlike traditional distributed systems, P2P networks are self-organizing networks that aggregate large numbers of heterogeneous computers called nodes or peers (nodes and peers will be used interchangeably in this paper). In P2P systems, peers can communicate directly with each other for data sharing and data exchanging. Peers also share their communication and storage resources.

The characteristics of P2P systems make them a very good choice for video broadcast over IP networks [2]. According to the P2P approach, peers interested in the same video transmission organize themselves into an application-layer multicast tree. A peer in the application layer multicast tree receives video packets from it parent, then duplicates and forwards them to it children [3, 4]. Peers in the overlay network are cooperative in the sense that they share data and exchange group control information. A multitude of protocols have been proposed in previous literature to manage tree-structured overlay networks for video broadcast [5–8].

However, although P2P technology gives new opportunities to define an efficient multimedia streaming application, at the same time, it involves a set of technical challenges and issues due to the best-effort and dynamic nature of the service offered by the Internet, and the heterogeneity of terminals used by clients to access the service. First of all, although peers access the network through different access links with totally different bandwidth characteristics, often dominated by the uplink connections, P2P communication infrastructure is often built upon an overlay network whose topology does not depend on the underlying physical network. seconds, network bandwidth, delay and loss behaviors rapidly change in time. Last, but not least, a large number of different terminals are becoming very popular, ranging from powerful high-performance home receivers to mobile handheld video devices; they present different capabilities and requirements in receiving and decoding video content.

The work in [9] describes state-of-the-art strategies that allow the deployment of efficient streaming solutions in P2P systems, focusing on both delivery architectures and adaptive streaming mechanisms. These solutions enable resource-demanding and delay-constrained applications over unstructured networks, and demonstrate the powerful of the P2P paradigm, along with adaptive streaming mechanisms, providing an interesting alternative for low-cost and effective multimedia communication applications.

In this context many other protocols based over P2P have been defined in the last few years. Unfortunately, the most of them are mainly related to the maintenance of the tree topology (see [10] for further details); they are not focused to the encoding aspects which, especially in heterogeneous and dynamic scenarios like the one considered in this paper, are fundamental issues that have to be accounted very carefully. In fact, a common hypothesis for all of them is that the source should encode video flows with the knowledge of the instantaneous network bandwidth of all peer-to-peer connections over the tree, and modify it output rate according to bandwidth variations. Otherwise, if no instantaneous knowledge of the link bandwidths is available at the source, encoding should be performed with an average quality, providing

(i) "good" peers (i.e., peers with a high-speed Internet access and a high-performance receiver device) with a bandwidth waste and an inadequately low-quality level,

(ii) "bad" peers (e.g., peers accessing the service with handheld devices, or accessing the Internet with low-rate links) with unpredictable and large number of losses.

However, even when the source knows the instantaneous bandwidth of each link in the distribution tree, the decision of the encoding bit rate is very hard. In fact, the solution that avoids losses would be to follow the worst peers, but in such a case the overall quality of the broadcasted video should be dominated by them: one peer with a very poor bandwidth is able to strongly degrade the quality of the video received by all the other peers.

Some rare cases of works accounting video encoding in P2P multipoint video distribution are present in literature, but they are often limited to nonrealtime video on demand (VOD) applications. For example, [11] presents a P2P multicast protocol and analyzes the gains that video coding and prioritized packet scheduling at the application layer can bring to the overall streaming performance. A rate-distortion model which predicts end-to-end video quality was presented in throughput-limited environments, using a source that applies H.264 encoding with SP and SI frames in order to adaptively stop the error propagation due to packet loss. However, in that paper, retransmission requests issued by receiving hosts are used to recover the most important missing packets while limiting the induced congestion, and therefore cannot be applied in video broadcast scenarios, where retransmissions may be the cause of unacceptable delay jitter.

Keeping all this in mind, the idea of this paper is to apply fine granularity scalability (FGS) encoding [12–17] for video transmission. FGS is an evolution of the scalable hierarchical video encoding; it was defined some years ago to deliver multimedia applications in heterogeneous network environments with different bandwidth and loss behaviors.

An FGS stream has only two layers: a Base layer that must be received to make possible video decoding, and an enhancement layer, henceforward indicated as the FGS layer, which can be delivered optionally where bandwidth is available. FGS allows the source to adjust the relative sizes of both Base and FGS layers, therefore allowing the FGS layer to be broken up and allowing the decoder to decode any portion of the received FGS layer. The source or any intermediate node is responsible to do that.

The authors of this paper think that, although not widely applied in the past due to it encoding complexity, today, thanks to the evolution in hardware and software technologies, the FGS appears a good and feasible solution for multipoint multimedia broadcast systems for the immediate future. Besides, [18] is an invention of 2006 that discloses methods, devices and systems for effective fine granularity scalability coding and decoding of video data.

In fact, the MPEG-4 FGS encoder developed by the authors is able to encode and decode video in realtime. The encoder has been implemented in Visual C++ using the

Intel Integrated Performance Primitives (Intel IPP) [19], an extensive library of multicore-ready, highly optimized software functions for multimedia data processing, and communications applications. Running on a start-level dual-core personal computer equipped with 2 GB of RAM, it is able to encode about 100 fps and decode about 400 fps for CIF video streams.

The target of this paper is to propose a multipoint broadcast video transmission framework over a heterogeneous content distribution P2P network. More specifically, the purpose of this work is to make the following contributions:

(i) launching and encouraging the application of a rate-controlled FGS encoding in broadcast video transmission in P2P networks; to the best of our knowledge, this is the first work with such an idea;

(ii) defining the architecture and a network managing protocol for the proposed broadcasting P2P platform;

(iii) analyzing and discussing an algorithm to manage the P2P network tree topology in a case study, giving the readers the possibility to change and tune the algorithm parameters according to the domain of interest: people who want to use the proposed architecture can freely decide to define any other tree topology management protocol and apply that to the Topology Manager entity described along the paper.

In the proposed platform the source generates the Base layer of the MPEG-4 stream in such a way that no losses occur at the Base-layer stream even in the presence of short-term bandwidth fluctuations. The FGS layer is sent together with the Base layer, but with a lower priority. In this way, peers connected to other peers through bottleneck links guarantee an efficient transmission of the Base layer, discarding only portions of the FGS layer.

The source uses a rate controller to regulate the encoding rate of the Base layer. A protocol is defined in order to provide the source with the necessary information related to the bandwidth of the most stringent bottleneck link. The framework works on a tree-structured P2P network, and any tree construction and management protocols [5–8] can be adopted.

A case study is introduced to evaluate both the performance of the proposed framework and the improvements obtained with respect to a reference case in which FGS is not applied.

The paper is structured as follows. Section 2 provides a brief description of FGS. Section 3 describes the proposed platform and the definition of the protocol. Section 4 analyzes a case study; first, the simulator tool we have implemented to generate the bandwidth processes at the overlay network level is described. Then, a statistical analysis of the performance at both the overlay network and the application levels will be carried out. Specifically, performance is calculated analyzing video at the receiving side, accounting frame loss and encoding PSRN simultaneously. Performance comparison with other platforms is made in order to evaluate the improvements introduced by that proposed in this paper. Finally, Section 5 concludes the paper.

2. FGS: Overview and Some Statistics

In this section we will provide a brief overview of the main characteristics of the FGS encoding technique in order to facilitate the understanding of the remainder sections of the paper. For a more detailed description of FGS, the reader is referred to [12–14].

FGS is a scalable encoding technique. Generally, there are three types of scalability, that is, temporal scalability, spatial scalability and SNR scalability. In all the three cases, the Base-layer pictures are encoded based on subsampling with either less frame rate (for temporal scalability), smaller picture size (for spatial scalability), or coarser picture quality (for SNR scalability). Full-quality video is obtained by the combination of both Base and FGS layers.

FGS is an evolution of the scalable hierarchical video encoding. It emerged to deliver multimedia applications in heterogeneous network environments with different bandwidth and loss behavior. It was defined in [13] with the main target of achieving a good balance between coding efficiency and scalability.

Its encoding is designed to cover any desired bandwidth range while maintaining a very simple scalability structure. The basic idea of FGS is to encode a video sequence into two layers only, a Base layer and an enhancement layer, in the following called the FGS layer. A MPEG-4 encoding is used: the Base layer is obtained with a classical MPEG-4 encoder using nonscalable coding, whereas the FGS layer is coded using a fine-granular scheme. The latter encodes the difference between the original picture and the reconstructed one with the use of bit-plane coding of the DCT coefficients.

An important application of FGS regards multihop connections with heterogeneous bandwidths; in such a case, each node may truncate where desired. The encoder needs to know just the minimum bandwidth over which it has to code the content. The bit stream of the FGS layer may be truncated by intermediate nodes into any number of bits per picture. The decoder will be able to reconstruct an enhancement video by combining the Base layer and the truncated FGS layer received bit streams. The FGS-layer video quality is proportional to the number of bits decoded by the decoder for each picture.

As far as the source is concerned, the encoding is feasible both offline and online, thanks to the availability of new hardware and software routines.

The idea of this paper is to apply MPEG-4 FGS encoding for video transmission, which allows intermediate nodes to truncate the bit stream of each frame at any point, thus only degrading the quality proportionally to the current available bandwidth. The Base layer of the MPEG-4 stream is generated in such a way that it encoding bandwidth is a little lower than the current minimum bandwidth in the tree. Therefore losses are negligible at the Base-layer stream even in the presence of short-term bandwidth fluctuations (specifically, in our system the Base layer is generated frame-by-frame at 90% of the minimum link bandwidth in the whole network). The FGS layer is sent together with the Base layer, but with a lower priority. By so doing, peers connected to other peers through bottleneck links guarantee the transmission of the Base layer, discarding only portions of the FGS layer.

This scheme is sketched in Figure 1, where an example of frame transmission on the tree is depicted. The amount of frame bits transmitted along the tree is represented by rectangles: the light gray portion of the frame represents the Base layer part, which is the same over all the links from the source to all the peers in the network. The FGS-layer part of the frame is represented with dark gray portions, with a length proportional to the number of bits forwarded to the next peer in the tree within a frame interval. Each link shows the transmission bandwidth between two connected nodes of the tree; specifically, the indicated bandwidth is the uplink bandwidth of each node towards it children. In this figure we can observe that higher-level peers receive a larger amount of bits, and therefore a better quality; quality degrades along the tree at each bottleneck uplink. In the example, the Base layer is encoded at 0.5 Mbit/s in such a way that even $P_{2,7}$, that is, the node with the worst connection with it parent in the tree, is able to receive it with no losses; the quality perceived by $P_{1,2}$ is greater than that perceived by $P_{2,6}$ because of the portion of FGS-layer discarded by $P_{1,2}$ due to the bandwidth value of the link $\{P_{1,2} \rightarrow P_{2,6}\}$, which is lower than that of the link $\{S \rightarrow P_{1,2}\}$. Of course, the performance perceived by each peer and, as a consequence, the overall performance of the proposed system are strongly influenced by the position of each peer in the tree, and by the tree topology itself. For this reason, the tree organization strategy plays a fundamental role in the performance of the whole system, and the related issues will be deeply discussed in Section 3.3.

3. System Description

The target of this section is to describe the system we propose in this paper. It is a live video broadcast platform where a video source distributes a video stream to a number of clients in a multipoint fashion. Multipoint communication is achieved by applying a P2P approach, configuring a tree-structured overlay network where the root is the video source, while the other clients, henceforward referred to as peers, are internal nodes or leaves. In particular, Sections 3.1 and 3.2 present, respectively, the architecture of the video source and the generic peer; Section 3.3 describes the algorithm we have used to organize the overlay tree.

3.1. Architecture of the Video Source. The architecture of the video source is shown in Figure 2. As it can be seen, it core is the *MPEG-4/FGS Video Encoder*, which receives a raw video stream at it input and produces the MPEG-4 video flow, made up of two separate streams, one related to the Base layer and one to the FGS layer. A *Replicator* is needed in order to create as many streams (for both Base and FGS layers) as the number of the source's children; this number is upper bounded by the so-called fan-out parameter F, defined as the maximum upload connections each peer can support at the overlay network layer. The bits produced at each frame interval are queued in the *Base-layer Buffer* and *FGS-layer Buffer*, respectively, in order to avoid possible losses. Then they are first grouped in packets by a *Packetizer* and, next, sent to the *Intra-Flow Scheduler*, which applies a round-robin strict-priority scheduling algorithm that considers the data in the second buffer (FGS layer stream) only in case the first one (Base layer Buffer) is empty. More specifically, the buffer gives priority to I-frames since they are the most important to be received as discussed in Section 4.4.2. At the end of each GoP, if the buffers contain data (Base and FGS layer) they are emptied in order to avoid excessive buffering delays. At most $2F$ streams coming from these buffers end up into the *InterFlow Scheduler*, which applies a weighted round robin algorithm to send the streams with an amount of bandwidth proportional to the uplink bandwidth estimated towards each children of the source.

The amount of bits to be used in the Base layer, and the relative encoding quality, are determined by the *Rate Controller* through the quantizer scale parameter qsp, that is chosen in the range between 1 and 31: the greater the qsp value, the poorer the encoding quality. As usual, it uses the rate distortion curves of the movie being coded [20–22].

The *Rate Controller* obtains the needed information about the bandwidth from the *Bandwidth Statistic Manager*. The latter periodically receives the source's children uplink bandwidth estimation from the *Bandwidth Estimator* and, at the same time, the uplink bandwidth estimation by all the other peers which are internal nodes in the tree (see $B_{p_1} \cdots B_{p_N}$ of Figure 2 or B_{p_i} of Figure 3).

The qsp is chosen by the Rate Controller using the minimum value between the uplink bandwidth estimations made by all the peers in the tree at each interval.

EWMA Filter. In order to avoid excessively strong oscillations of both the encoding quality and the system behavior, the bandwidth values are first smoothed with an exponentially-weighted moving average (EWMA) filter with parameter $\beta^{(\mathrm{RC})}$, defined as follows:

$$\widehat{B}_n = \beta^{(\mathrm{RC})} \cdot \widehat{B}_{n-1} + \left(1 - \beta^{(\mathrm{RC})}\right) \cdot B_n, \qquad (1)$$

where $\widehat{B}_{n-1}$ and $\widehat{B}_n$ are the filtered bandwidths at the $(n-1)$th and nth update events, B_n is the instantaneous bandwidth at the nth update event, and $\beta^{(\mathrm{RC})}$ is the Rate Controller filter parameter whose range is between 0 and 1; values of $\beta^{(\mathrm{RC})}$ close to 1 give more importance to the history of the bandwidth process, achieving a process that is less sensitive to high "frequencies" (therefore able to smooth the short-term variations), but having slower responses to bandwidth changes; conversely, very low values of $\beta^{(\mathrm{RC})}$ give more importance to recent measures, achieving greater responsiveness to the process variations. In our implementation we have used the Rate Controller EWMA ($\mathrm{EWMA_{RC}}$) with $\beta^{(\mathrm{RC})} = 0.8$.

Another important block in the video source architecture is the *Topology Manager*, which decides and maintains the tree network topology by deciding the position of each peer within the tree. It receives the uplink bandwidths of all the

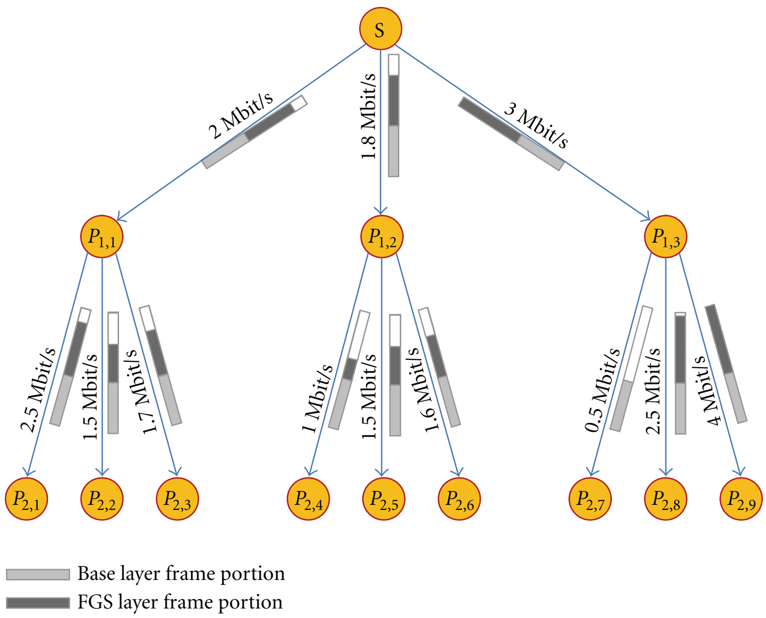


Figure 1: FGS video quality degradation along the tree due to the presence of bottlenecks.

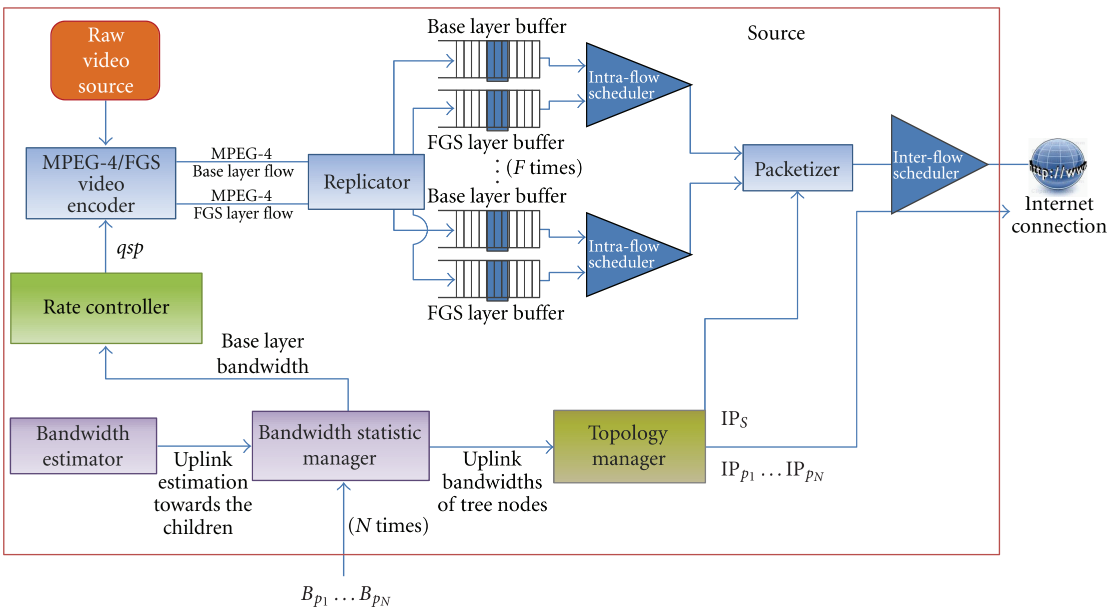


Figure 2: Architecture block diagram of the video source.

peers from the *Bandwidth Statistic Manager*, and implements an algorithm to manage both peer arrival and departure events. In detail, the Topology Manager is responsible of the following tasks

When a new peer arrives, this peer issues an admission request to the Topology Manager in order to receive both a position in the current tree and the IP address of the peer that will be it parent. The Topology Manager sends also the IP address of the new peer to it parent node. Specifically, the parent node inserts the IP of the new child in it Replicator in order to create another copy of the packets for the new child. The parent node creates also a new buffer pair (for the Base and the FGS layers) for it new child, and communicates to the Packetizer and the InterFlow Scheduler to handle the new node. The latter will be inserted in the weighted round robin buffer. Finally, the parent node inserts the IP of it new child in the Bandwidth Estimator.

When an existing peer departs, the process just described for peer arrival is reversed. The parent node deletes from the Replicator the IP of the departing peer. It also cancels the buffer pair (for the Base and the FGS layers) and communicates to the Packetizer and the InterFlow Scheduler

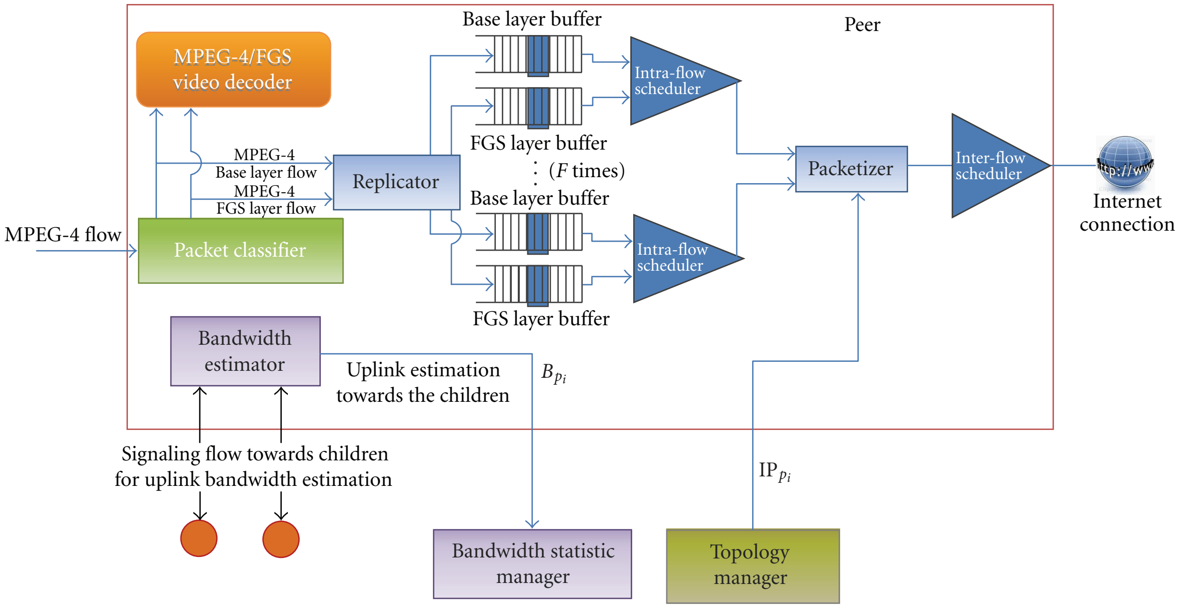


Figure 3: Architecture block diagram of the generic peer p.

to stop serving it old child node. The IP of the departing peer is also removed from the Bandwidth Estimator.

When the Topology Manager receives the information about the uplink bandwidth of each peer, periodically (in our implementation we have considered a period of $T_T = 30$ Seconds) it updates and optimizes the tree. To this end, using the same information received by the *Bandwidth Statistic Manager*, it implements an algorithm to manage topology modifications run-time after bandwidth variations. To avoid strong oscillations of peers when reorganizing them in the updated topology, the received bandwidth values are first passed under a EWMA filter, the so-called Topology Manager EWMA ($\mathrm{EWMA_{TM}}$) filter, which works like the one defined in (1) with parameter $\beta^{(\mathrm{TM})}$ (in our implementation we used $\beta^{(\mathrm{TM})} = 0.9$).

3.2. Architecture of a Generic Peer. The architecture of a generic peer is shown in Figure 3. It mainly performs three functions:

(1) video play-out,

(2) forwarding,

(3) bandwidth estimation.

Video bit streams, organized in IP packets, are received by the *Packet Classifier*, subdivided and queued in two different buffers, according to the type of packets: Base layer and FGS layer. At the same time, packets are given to the local *MPEG-4/FGS Video Decoder* for playback. As for the case of the source, a Replicator block is needed in order to create at most F streams (for both Base and FGS layer), one for each child of the generic peer. The bits produced at each frame interval are queued in the *Base-layer Buffer* and the *FGS-layer Buffer*, respectively. The two buffers are served by an *Intra-Flow Scheduler* as those discussed previously for the source. Of course, if estimation has been made correctly, the Base layer will find enough bandwidth on the transmission channels, and will not incur in any packet loss. On the other hand, the FGS-layer bits in each frame will be enqueued during the GoP, and transmitted only at the end of it, after the transmission of the Base layer of all the frames in the same GoP. If the time available for the GoP transmission elapses and bits of the new frame of the next GoP arrive, the Buffers are emptied deleting the remaining bits of the previous GoP. A smoothing operation is therefore performed in order to provide the Base layer with a higher priority.

For this reason, as already discussed previously, the decoding quality at lower levels of the tree will result worse. Each of the $2F$ streams coming from these buffers, together with the information about the peer topology sent by the *Topology Manager* of the source, ends up into a *Packetizer* responsible for packets creation and transmission on the Internet. Packets are sent to the *InterFlow Scheduler*, which, as the one discussed for the source, applies a weighted round robin algorithm and serves each flow with an amount of bandwidth proportional to the uplink bandwidth towards each source of children.

In addition to video play-out and forwarding, another important function performed by each peer is the uplink bandwidth estimation towards their children. This task is operated by the *Bandwidth Estimator*, which periodically sends the estimated bandwidth values to the *Bandwidth Statistic Manager* of the source discussed for the video source diagram of Figure 2. The purpose of this function was to achieve the best performance in small- and medium-size networks. Conversely, if we consider large networks, the algorithm can be slightly modified to be more scalable, but with worse performance. This part of the system is completely general and any bandwidth estimation algorithm can be plugged in. The choice of it goes beyond the purpose

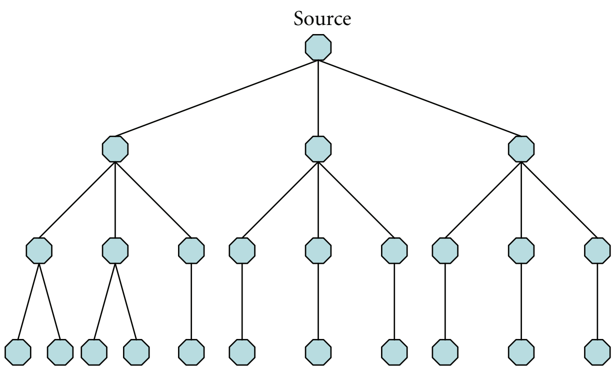
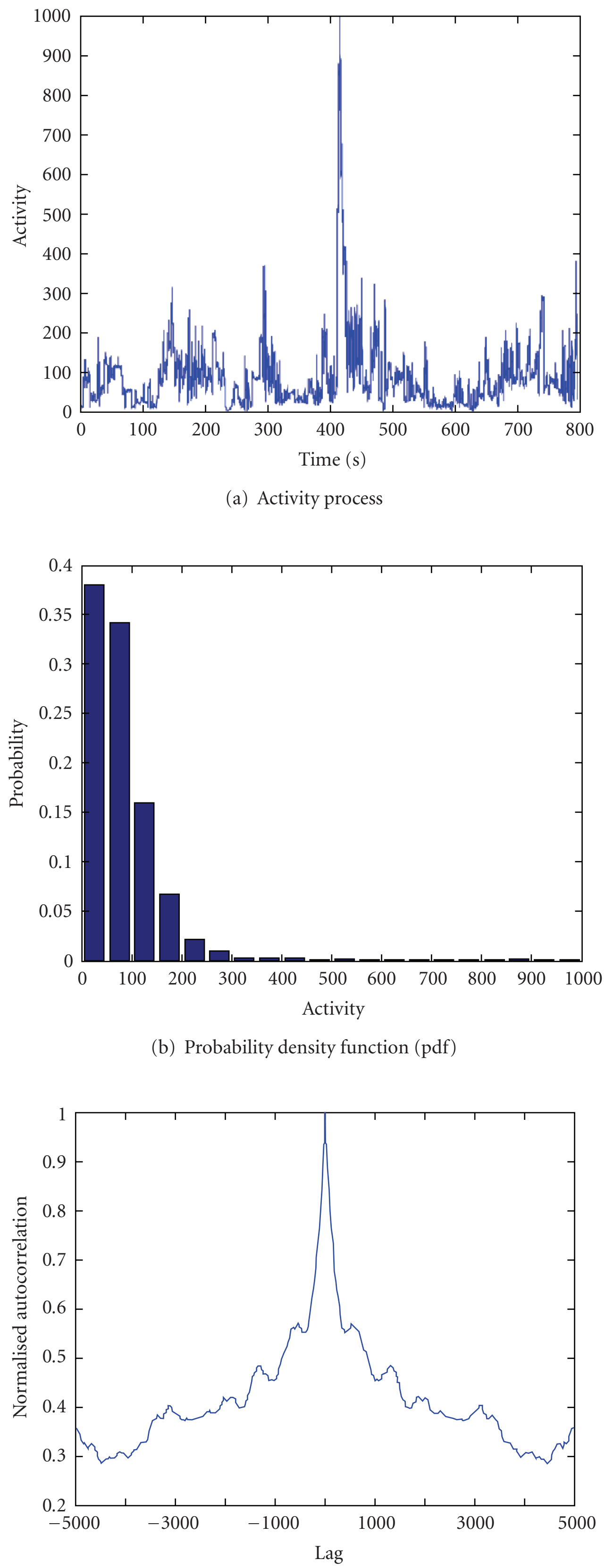


Figure 4: Construction of a tree with 24 peers and $F = 3$.



(a) Activity process

(b) Probability density function (pdf)

(c) Normalized autocorrelation function

Figure 5: Time behavior and first-and-second order statistics of the activity process of the sequence "BBC Planet Earth documentary".

of this paper. In addition, a sophisticated algorithm to predict the bandwidth in the short or middle term [23, 24] can be applied.

3.3. Organization of the Tree. In order to have our protocol working efficiently by applying FGS, it is obvious that a node with a low uplink bandwidth behaves as a bottleneck; if it was located in the upper part of the tree it would penalize all it descendants. Thus, the best peers should be located at the top of the tree.

For this reason, taking into account that link bandwidth varies in time, the Topology Manager has to optimize the tree topology runtime, in order to avoid configurations with bottlenecks at the highest levels of the tree. We will assume that the downlink bandwidth of each peer is higher than the corresponding uplink bandwidth. In fact:

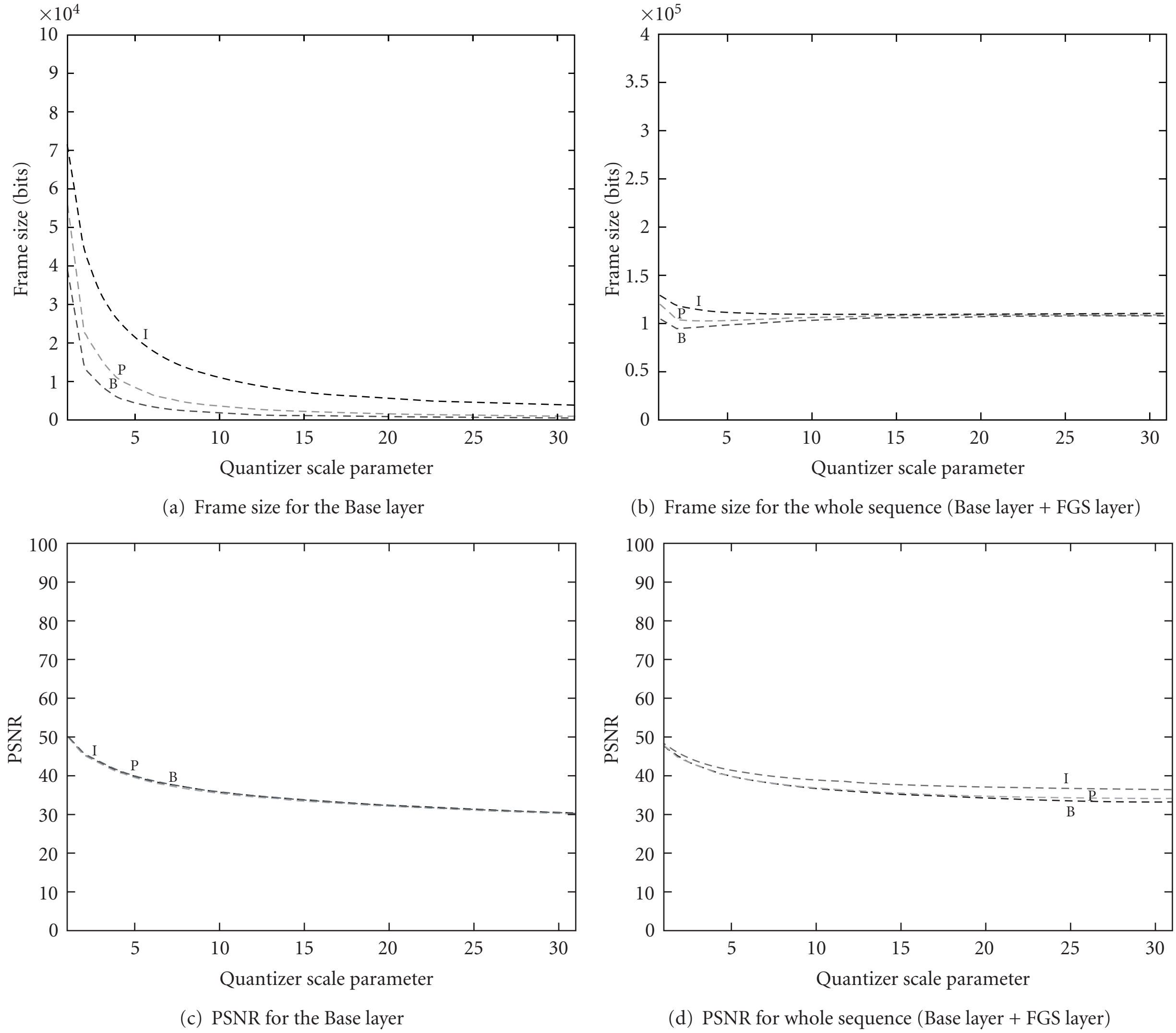
(1) there are many scenarios (e.g., ADSL) where the uplink bandwidth is lower than downlink bandwidth;

(2) the uplink bandwidth is shared with tree nodes whereas the downlink bandwidth of each peer is used as a whole to connect them with their parent only.

The Topology Manager obtains the bandwidth estimations every T_P seconds ($T_P = 10$ seconds in our implemented system). Then, every T_T seconds ($T_T = 30$ in our implementation) it updates the tree topology as follows: it chooses the F peers with the highest uplink bandwidth, and connects them as children of the source. Then, for each of these peers, the same operation is repeated, choosing the next F peers with the highest uplink bandwidths. This algorithm is run recursively until all the peers get a position in the tree. To optimize the bandwidth performance and to further balance the tree, the nodes in the last level (the leaves) are evenly distributed among the peers in the upper level of the tree (the parents nodes) in a round robin fashion; this avoids the scenario where some peer is overloaded with it uplink bandwidth whereas others are not. Figure 4 shows an example of a tree built as explained, with 24 peers and $F = 3$.

Basically, this tree is balanced and complete (all the levels are full except for the bottom level).

Let us note that changes in the topology structure can be deleterious for video decoding of peers; changing their



(a) Frame size for the Base layer

(b) Frame size for the whole sequence (Base layer + FGS layer)

(c) PSNR for the Base layer

(d) PSNR for whole sequence (Base layer + FGS layer)

Figure 6: Rate-Distortion curves of the video sequence "BBC Planet Earth documentary".

position in the tree can cause sudden changes in delays, thus increasing the delay jitter as well. In addition, let us observe that this does not happen only when a peer changes level in the tree, but also if either it changes it parent or one of it predecessors changes level or predecessor. However, two important observations can be made in favor of such a strategy:

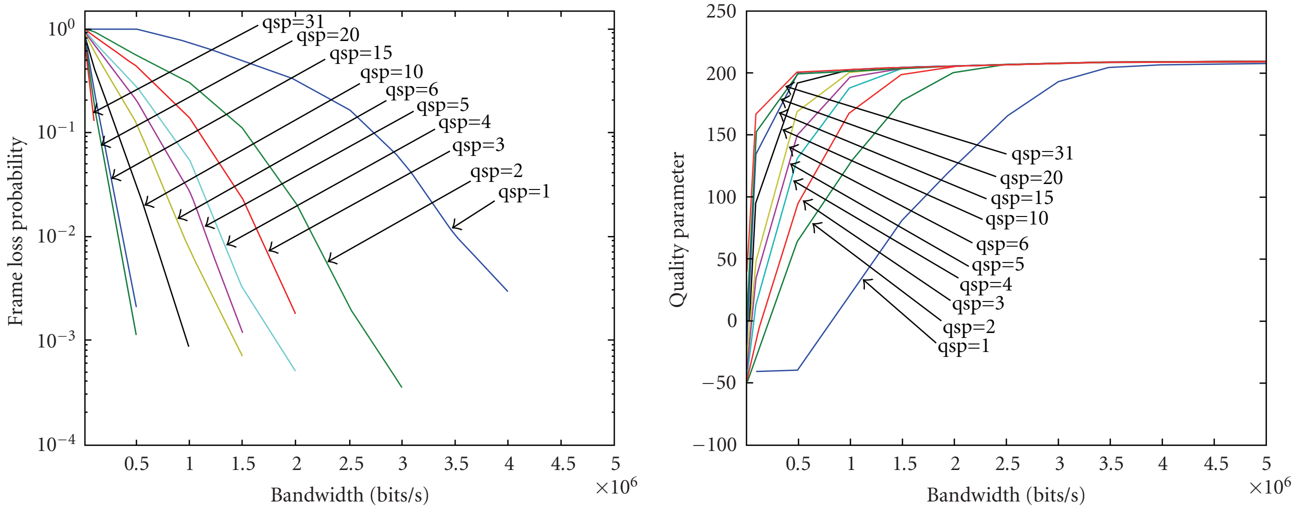
(1) since the bandwidth values are passed under the Topology Manager EWMA ($EWMA_{TM}$) filter (see Section 4), this strategy is not affected by occasional bandwidth changes. Changes in the tree topology occur in the event of serious and lasting bandwidth changes; in this case changes are likely to optimize performance;

(2) the delay jitter caused by topology changes can be recovered at destination through the application of intelligent compensation buffers and adaptive media play-out techniques (see e.g., [25, 26]).

Another observation is that both the interval durations T_P and T_T, and the parameters for all the EWMA filters used in our implementation were chosen empirically after a large number of experiments in order to optimize the tradeoff between system responsiveness to network bandwidth variations and the amount of signaling traffic.

4. Case Study

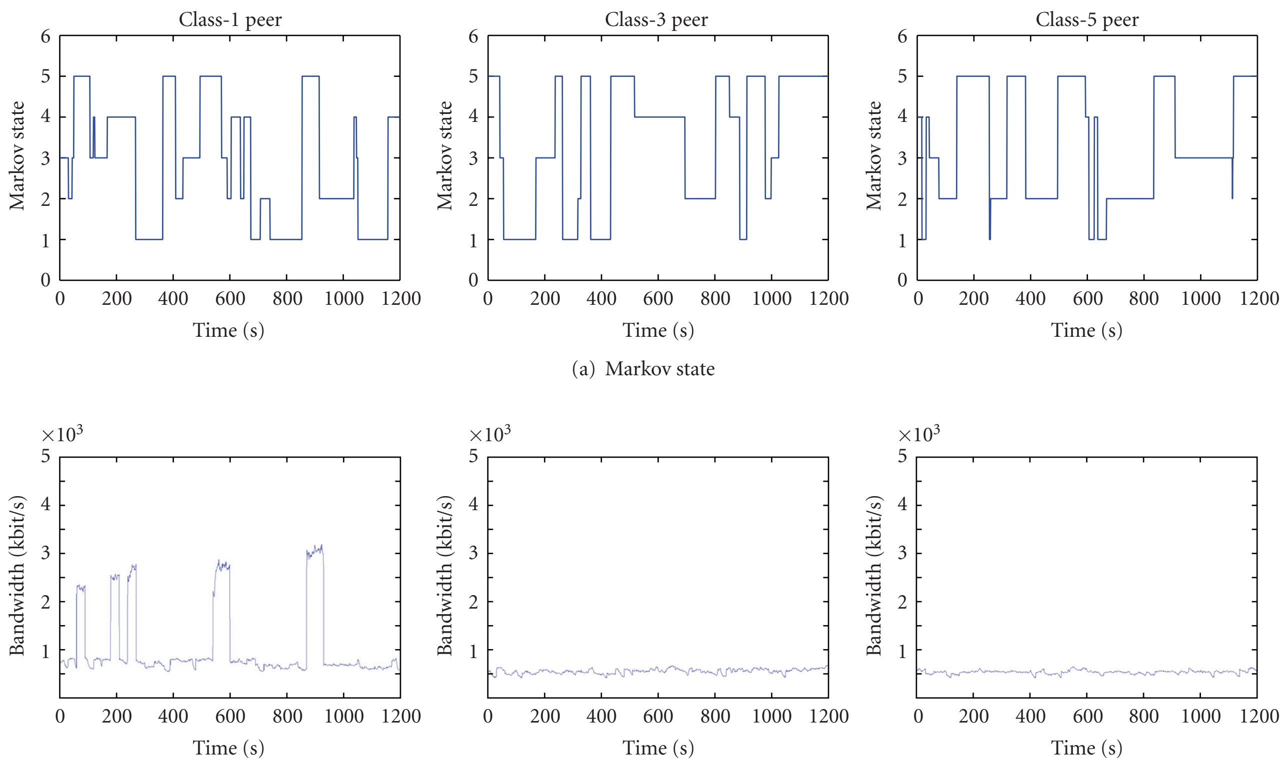
In this section we define a case study to analyze the performance of the proposed system. The target is to demonstrate the gain achieved by applying FGS encoding against classical video streaming over P2P, in terms of the peak signal-to-noise ratio (PSNR) evaluated on the video flow received at destination from each peer. More specifically, Section 4.1 introduces our case study; Section 4.2 describes the bandwidth generation simulator we developed to generate the processes of the uplink bandwidth at the overlay network level; Section 4.3 analyzes some statistics of the



(a) Frame loss

(b) Quality parameter

Figure 7: Frame loss percentage and quality parameter for the video sequence “BBC Planet Earth documentary.



(a) Markov state

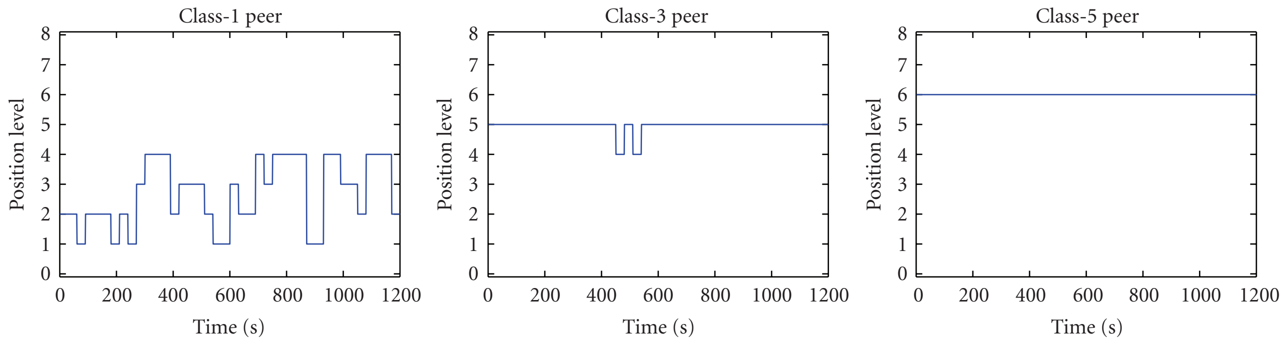
(b) Uplink bandwidth

Figure 8: Markov chain state and the generated uplink bandwidth for three representative peers of the tree.

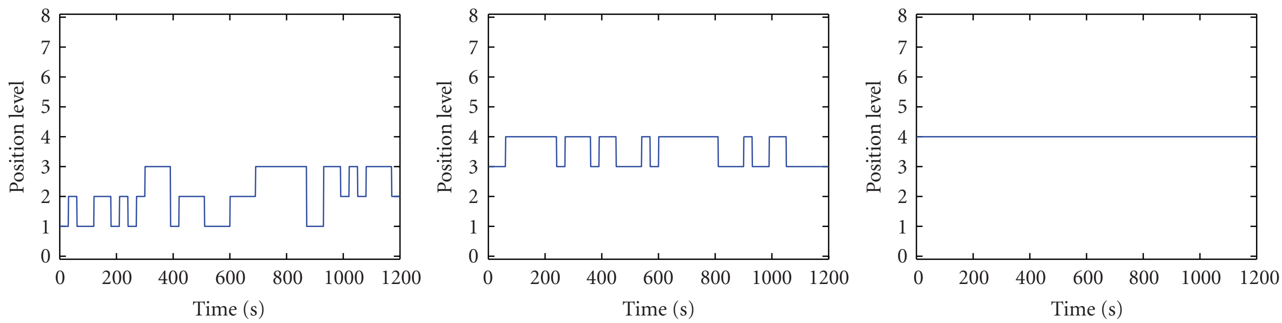
considered sequence; Section 4.4 contains both the results of the overlay network behavior analysis and the performance analysis at the application level. Finally, Section 4.5 shows the comparison of the platform we propose in this paper against other approaches currently used.

4.1. Case Study Description. According to what we said in the previous section, we consider a video distribution platform with centralized control by the Topology Manager.

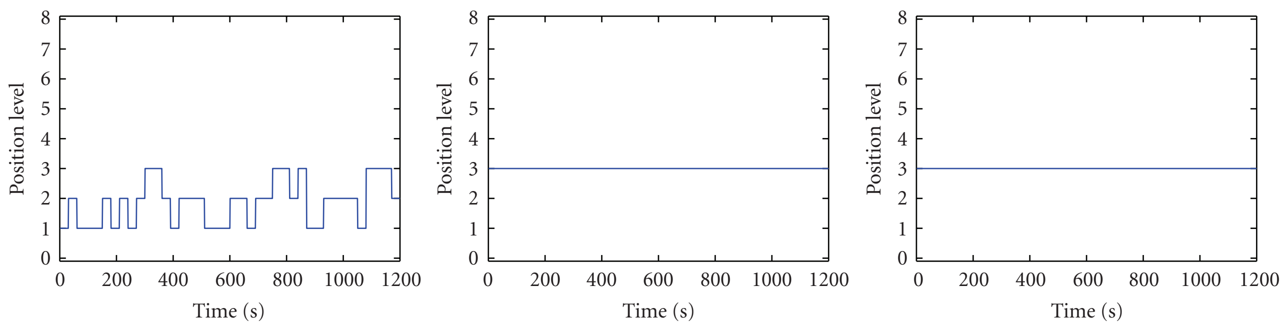
We will assume that all the peers, included the source, have set the same fan-out parameter, F, representing the



(a) $F = 2$



(b) $F = 3$



(c) $F = 4$

Figure 9: Time behavior of peer position in the tree of three representative peers.

maximum number of peers that can be attached as children in the distribution tree.

We will carry out a steady-state analysis, assuming that the number of peers in the network, hereafter referred to as N, remains constant for the whole duration of the simulation. In this way our study does not depend on the particular algorithm used to manage the topology structure when peer arrivals or departures occur. Therefore, the only job of the Topology Manager is to rearrange the tree according to bandwidth variations. The management of transitory peers will be discussed in the Future Work section.

Peers are grouped in $C = 5$ different classes, each characterized by different Internet access link performances, and different average values of the uplink bandwidth. As previously explained, the downlink bandwidth of each peer is assumed to be much higher than the corresponding uplink bandwidth. Therefore, along this section we will refer to the uplink bandwidth as the *bandwidth*, unless explicitly mentioned. Classes are organized with decreasing average bandwidth values: for example, peers in the third class have higher bandwidth values than peers in the fourth class. We have assumed that the source is a high-bandwidth server with an uplink of 5 Mbit/s.

4.2. Bandwidth Generation Simulator. In real scenarios bandwidth apparently available to a peer (e.g., the ADSL access of it domestic connection) might be not true, and the peer actually can make use only of a small fraction of it, for example due to the intensive use of file sharing programs or bandwidth sharing with other users in the same LAN. For this reason, in order to model the bandwidth variations in a realistic manner, we realized a bandwidth generation simulator at the overlay network level. It creates the desired bandwidth sequences first generating intermediate sequences, by using a modified version of the Switched Batch Bernoulli Process (SBBP) (see [20, 27]), the most general Markov

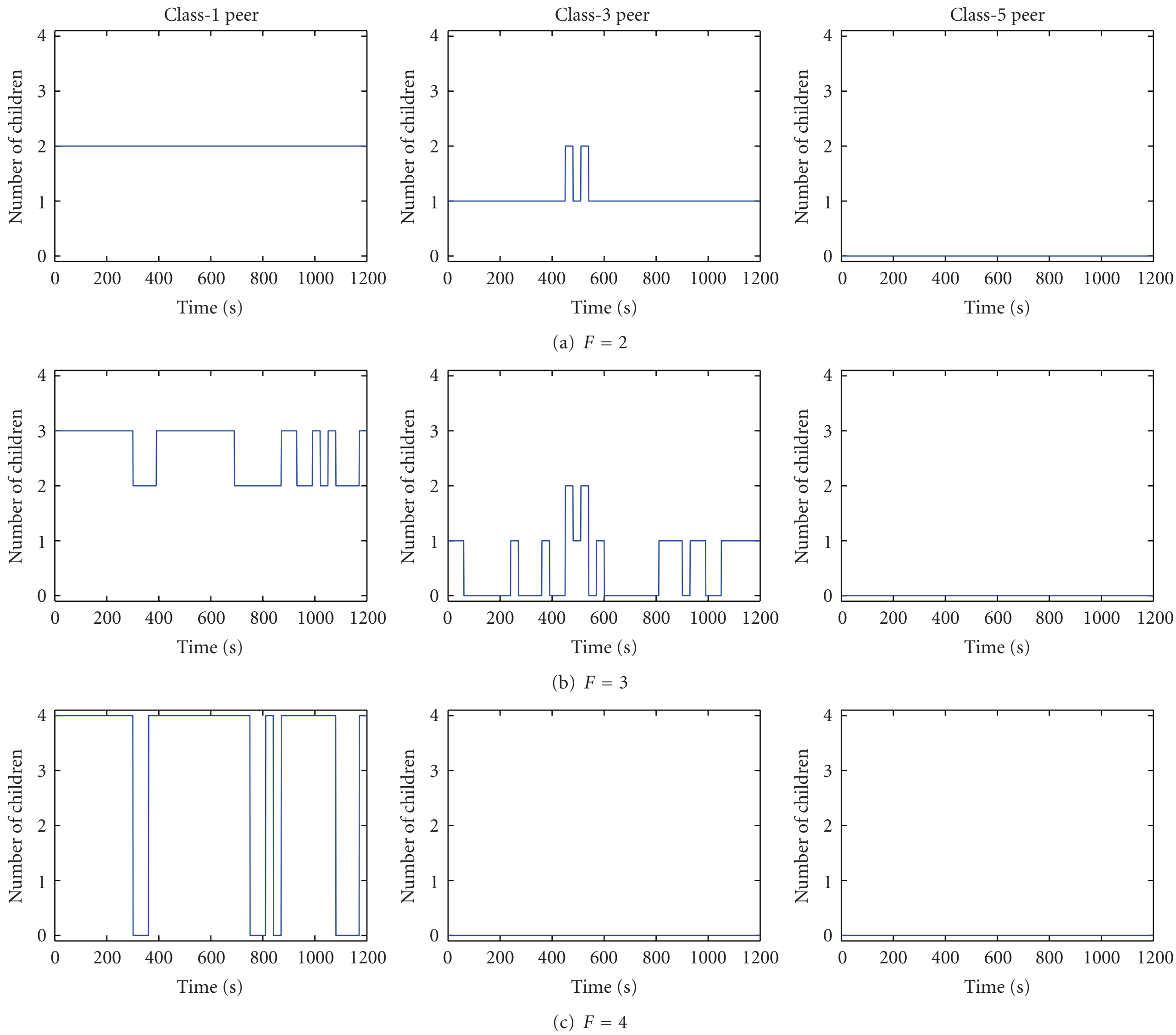


FIGURE 10: Time behavior of the number of children in the tree of three representative peers.

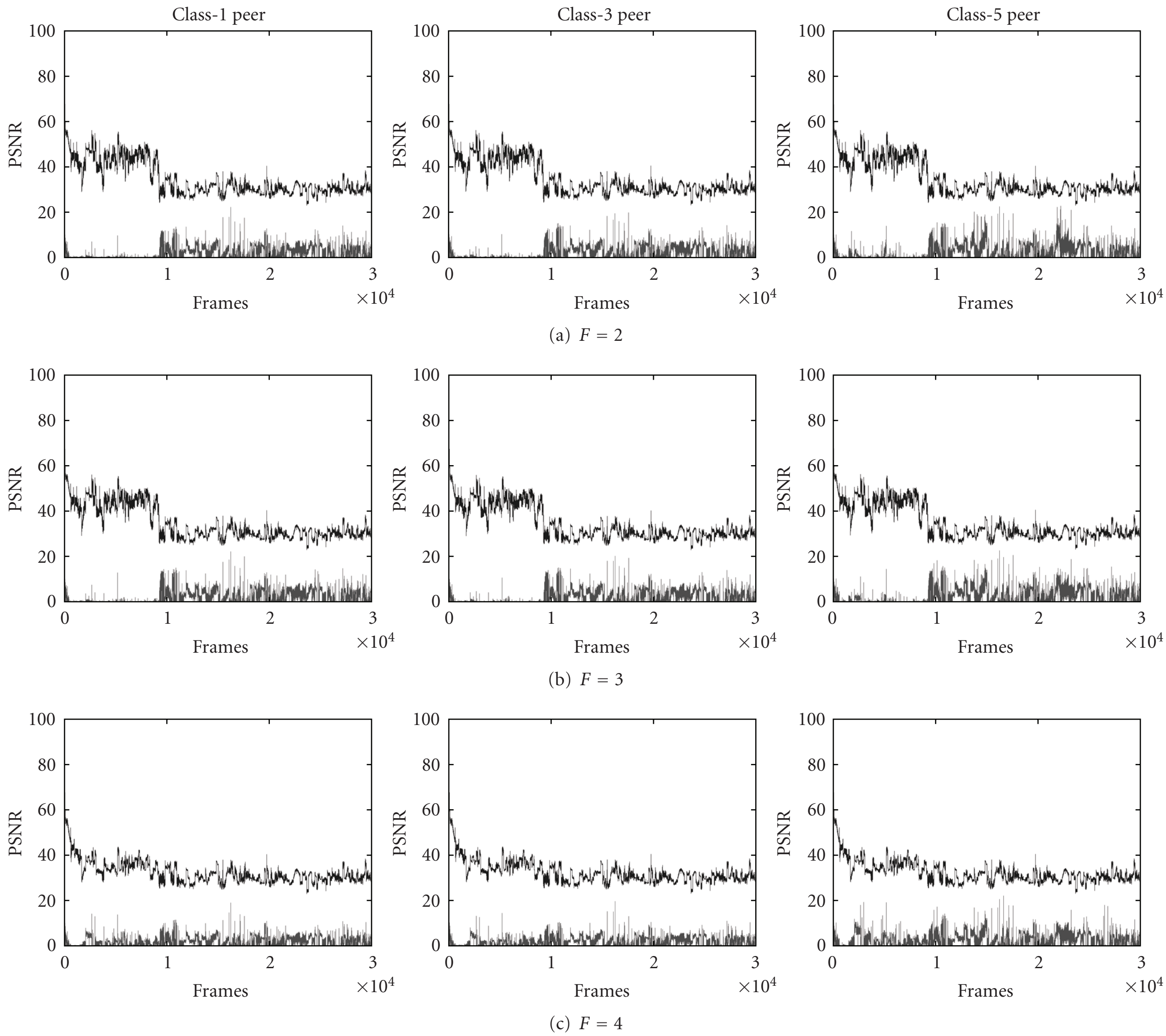
modulated process in the discrete-time domain. Then it creates the final sequences by applying a EWMA filter, in the following indicated as the *Bandwidth Variations EWMA (EWMA*$_{\mathrm{BV}}$*)*, with parameter $\beta^{(\mathrm{BV})}$ ($\beta^{(\mathrm{BV})} = 0.8$ in our simulations (see Section 3.1 for details on the EWMA filter). This EWMA filter is applied to the temporary sequences generated at the first step, to achieve the final smoothed traces. Therefore, the uplink of each peer is simulated with two blocks in cascade: a bandwidth generator block and a EWMA$_{\mathrm{BV}}$ filter block for the smoothing needed to obtain the desired dynamics. The bandwidth generation tool and the uplink bandwidth processes used for all the simulations are available at [28].

The first step of the generation algorithm works as follows. Let $i \in \{1, \ldots, C\}$ be the generic class, grouping peers with similar Internet accesses, and let $P^{(i)}$ be the generic peer belonging to the class i.

Let $BW_{P^{(i)}}(n)$ be the uplink bandwidth process of the peer $P^{(i)}$.

The process $BW_{P^{(i)}}(n)$ will be considered as modulated by a L-state underlying Markov chain, where the mean permanence is a geometricall-distributed random variable with mean value M_μ, assumed equally for all the classes. Once the underlying Markov chain leaves a state, it moves to one of the other states with the same probability. To this end we define the generic element of the transition probability matrix of the underlying Markov chain as follows:

$$Q^{(i)}_{[h,k]} = \begin{cases} \dfrac{1}{(L-1)}, & \text{if } h \neq k \text{ with } h, k \in \{1, \ldots, L\}, \\ 1 - \displaystyle\sum_{j \neq k} Q^{(i)}_{[h,j]}, & \text{if } h = k \text{ with } h, k \in \{1, \ldots, L\}. \end{cases} \quad (2)$$



(a) $F = 2$

(b) $F = 3$

(c) $F = 4$

Figure 11: PSNR for the Base layer (black) and gain achieved using FGS (gray).

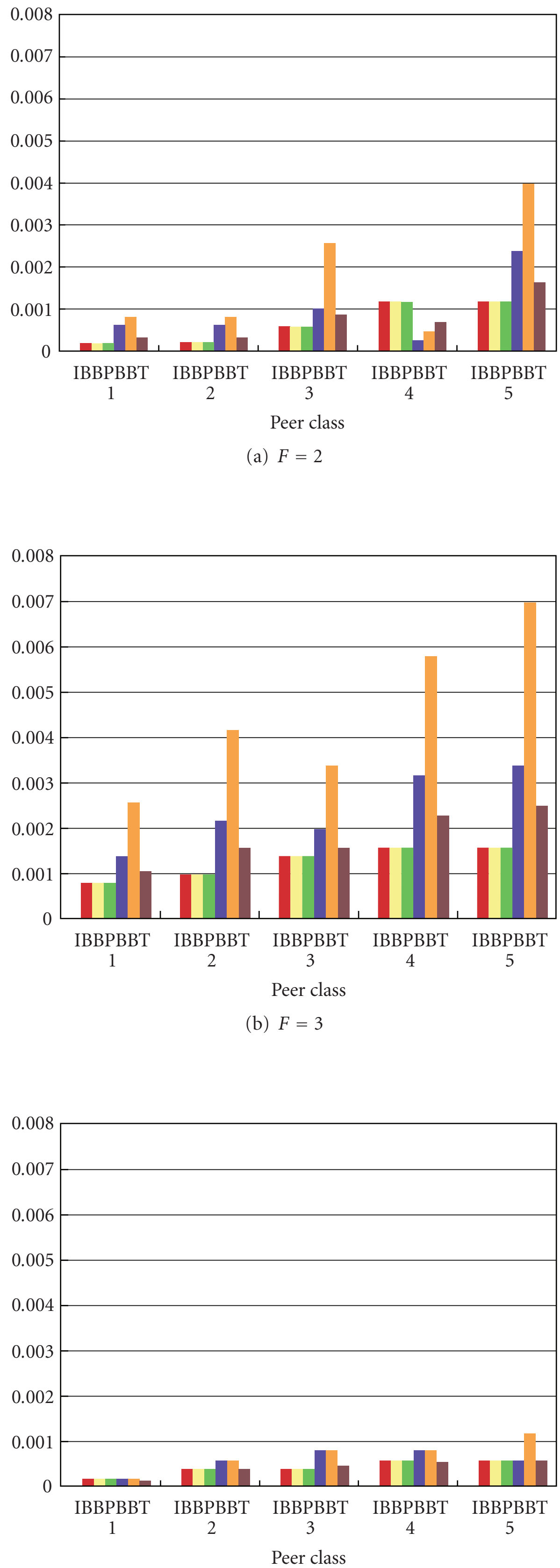
During the permanence of the Markov chain in the generic state h, with $h \in \{1, \ldots, L\}$, the value of the uplink bandwidth at the slot n, $BW_{P^{(i)}}(n)$, is randomly chosen according to a Gaussian distribution with a state-dependent mean value $\mu^{(i)}(h) = [(h - (i - 1)) \cdot 10\% + 1] \cdot \widehat{W}$, with $\widehat{W}$ equal to the average uplink value of the intermediate state $h = \lceil L/2 \rceil$, and a standard deviation $\sigma^{(i)}(h) = 0.08 \cdot \mu^{(i)}(h)$.

In our case study we have used a slot duration of 1 seconds, the number of states of the Markov chain $L = 5$, a mean duration of the permanence in each state of the underlying Markov chain of $M_\mu = 60$ slots, and $\widehat{W} = 370$ kbit/s.

Let us stress that the bandwidth generation technique and values described above are only provided for the reproducibility of the results achieved: any other bandwidth generation process can be used. In fact, other bandwidth processes have been applied by the authors, giving equally significant results and leading to the same conclusions discussed below.

4.3. Statistical Analysis of the Considered Sequence. For all the experiments along the paper we have considered a sequence of the video "BBC Planet Earth documentary". The sequence has a duration of 20 minutes and is encoded with a 176 × 144 QCIF format, at a frame rate of 25 frame/seconds, and using as Group of Pictures (GoP) structure with the pattern IBBPBB. For the sake of completeness, in this section we show a statistical analysis of this sequence.

The first parameter we analyze is the activity process of the sequence, which is a spatial property: the greater the spatial frequency range, the greater the activity. Activity has a strong impact on the behavior of the encoding results, in both frame size and PSNR: frames with a higher activity have a higher size; on the other hand, if a rate controller



(a) $F = 2$

(b) $F = 3$

(c) $F = 4$

Figure 12: Frame loss percentage for the five representative peers.

Table 1: Total number of levels in the tree and number of peers in the last level.

F	Number of levels	Peers in the last level
2	6	22
3	4	45
4	3	64

is applied to control the frame size, frames with a higher activity present a lower PSNR. Activity is determined by the scene of the movie only, and does not depend on the encoding technique and parameters used in the encoding process. Its time behavior for the considered sequence is shown in Figure 5, together with it first- and second-order statistics in terms of probability density function (pdf) and normalized autocorrelation function. Figure 6 presents the Rate-Distortion curves, calculated as in [20–22] for both the Base layer and the whole sequence (the latter is considered as the aggregation of both the Base and the FGS layers). The curves represent the frame size (Figures 6(a) and 6(b)) and the PSNR (Figures 6(c) and 6(d)) for the three encoding modes (I, P and B) as a function of the quantizer scale parameter, *qsp*. The strongly decreasing behavior of the curves for low values of *qsp* in Figure 6(b) is due to the corresponding behavior of the Base layer in the same range, since the Base layer is predominant over the FGS layer within this range. In Figure 6(d) we can observe that, as expected, the presence of the FGS layer allows the quality of the overall sequence to reach a high value even when the Base layer is encoded with high values of *qsp*, and therefore the overall PSNR is almost independent on the used quantizer scale. It follows that a low value for *qsp* would increase the frame size of the base layer too much, causing potential frame losses in case of bandwidth oscillations; on the other hand, a high value for *qsp* would increase the size of the aggregation of Base and FGS layers to be transmitted and decrease the PSNR. However, in this latter case, the encoded stream is more robust to network losses, thanks to the higher percentage of FGS bits, that can be truncated in any point of the stream. Therefore, in real scenarios, that is in presence of bandwidth oscillations, it is obvious the importance of the Rate Controller: according to bandwidth, it has to determine the size of the Base, and consequently the size of the FGS layer, in order to maximize the encoding quality, but getting frame losses as fewer as possible. For this reason, in our analysis we will consider the following quality of service (QoS) aspects simultanously:

(1) the encoding quality at the source side;

(2) the loss percentage in the network.

In fact, it is misleading considering the peak signal-to-noise ratio (PSNR) parameter at the source only, that is, for all the frames encoded by the source, since the quality perceived at destination is strongly degraded if network losses occur. On the other hand, it is misleading as well averaging PSNR over only the noncorrupted received frames because in this case the estimation would result not fair (e.g., because low-quality frames are small, and therefore more likely to be received).

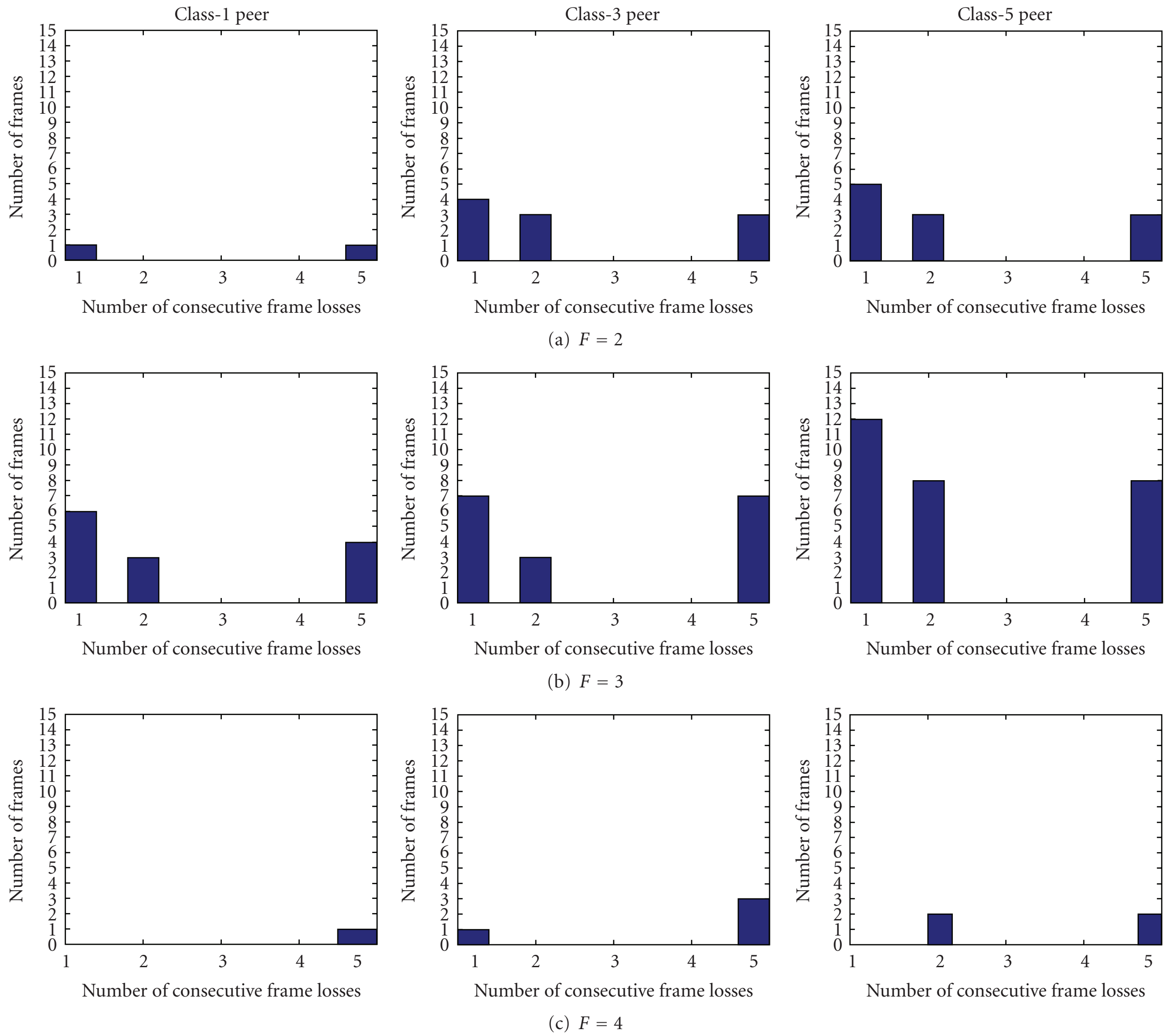


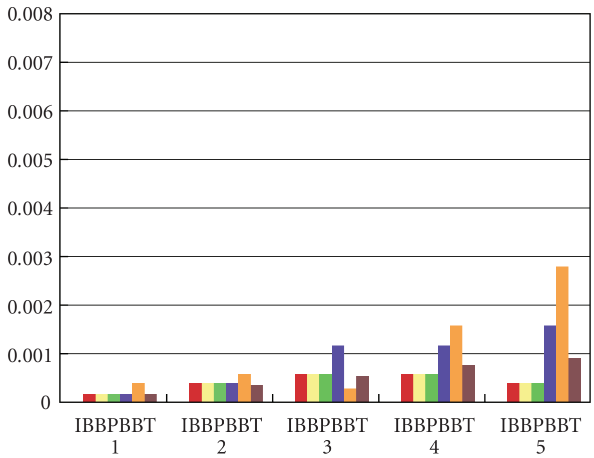
Figure 13: Histograms of consecutive frame losses.

Table 2: Average PSNR and Q for representative peers in each class.

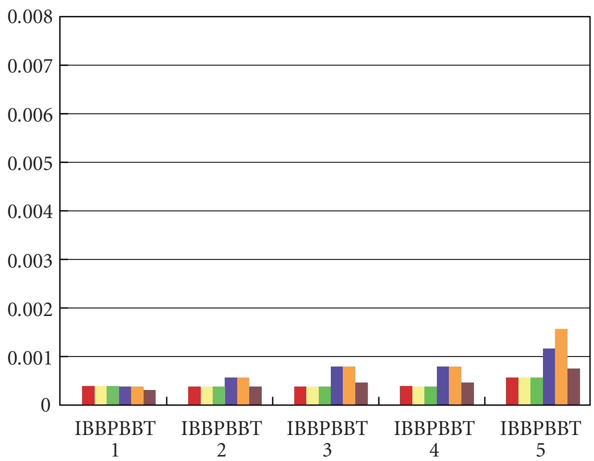
	PSNR			Q		
Class	$F = 2$	$F = 3$	$F = 4$	$F = 2$	$F = 3$	$F = 4$
1	36.2427	36.3263	34.5218	0.409232	0.446816	−0.365208
2	35.9045	36.0584	34.1065	0.257003	0.326285	−0.552064
3	35.8657	36.0403	33.9856	0.239575	0.318134	−0.60648
4	35.8563	36.0343	33.9228	0.235342	0.315454	−0.634722
5	34.4134	34.4915	32.3582	−0.413992	−0.378809	−1.33883

Table 3: Average PSNR and Q for representative peers in each class (*Dynamic* + *Base* method).

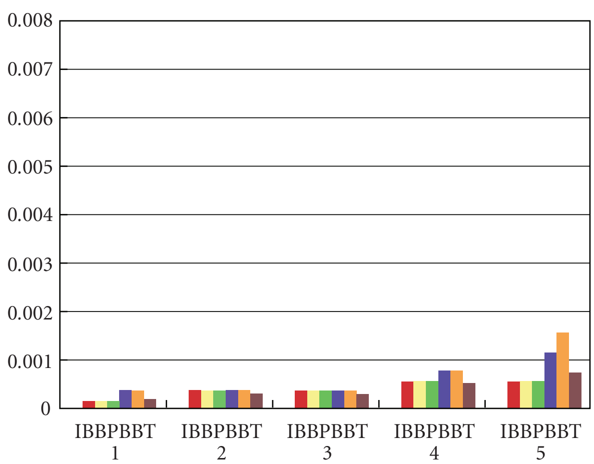
	PSNR			Q		
Class	$F = 2$	$F = 3$	$F = 4$	$F = 2$	$F = 3$	$F = 4$
1–5	34.186	34.108	32.3634	−0.411642	−0.371378	−1.33648



(a) $F = 2$



(b) $F = 3$



(c) $F = 4$

Figure 14: Frame loss percentage for the five representative peers (*Static + Base/FGS* and *Static + Base* methods).

For this reason, in order to take into account the encoding quality and the frame rate reduction simultaneously, we have used a *Quality parameter*, defined in [29] through a heuristic formula, which models the overall video quality at destination as a function of both the mean PSNR in dB of the received frames and the frame rate. More specifically, the *Quality parameter* is defined as follows:

$$Q = 0.45 \cdot psnr + \frac{(fr - 5)}{10} - 16.9, \tag{3}$$

where *psnr* is the PSNR value measured at the destination averaged only on the frames received at destination, *fr* is the frame rate of the video sequence received at destination, excluding frames corrupted due to damages of the Base layer for network losses. The constant coefficients in (3) were calculated in [29] by evaluating the data set obtained in a survey, and assuming a minimum acceptable frame rate of 5 frame/seconds. According to the above definition, the greater the PSNR and the frame rate at destination, the greater the Q parameter. Figure 7 shows the frame loss percentage and the Q value against the mean available network bandwidth for different values of *qsp*. Negative values of Q represent a decoded frame rate at destination of less than 5 fps, which is the minimum threshold assumed in [29] for an acceptable quality.

4.4. Numerical Results. This section contains numerical results concerning the overlay network behavior analysis (Section 4.4.1), and the performance analysis at the application level (Section 4.4.2).

4.4.1. Overlay Network Behavior Analysis. Here we will show results on some peers randomly chosen as representative of their corresponding class. Their uplink bandwidth has been generated using the tool described in Section 4.2.

We have fixed the number of peers to 85, including the source, and analyzed the performance for three different values of the fan-out parameter, $F \in \{2, 3, 4\}$. The 84 peers (the source has been considered as a high-bandwidth server with 5 Mbit/s uplink) have been randomly chosen from five different classes of peers (peers in different classes have different ranges of bandwidth values, as discussed in Section 4) according to a uniform distribution. The resulting peers were distributed as follows:

(i) 12 peers of class 1;
(ii) 23 peers of class 2;
(iii) 17 peers of class 3;
(iv) 19 peers of class 4;
(v) 13 peers of class 5.

We have generated traces for a time interval of 20 minutes, equal to the length of the considered movie trace. The last level of the obtained trees is partially filled for $F \in \{2, 3\}$, whereas it is completely filled for $F = 4$. Table 1 shows the number of tree levels and the number of peers lying in the last level for the three considered values of F.

Table 4: Average PSNR and Q for peers in each class for (*Static* + *Base/FGS* method) with $F \in \{2, 3, 4\}$.

	PSNR			Q		
Class	$F = 2$	$F = 3$	$F = 4$	$F = 2$	$F = 3$	$F = 4$
1	35.3316	35.4134	34.1984	0.3764	0.4189	−0.3744
2	35.2096	35.3491	34.1605	0.21363	0.30123	−0.4546
3	34.8912	34.773	33.6131	0.2111	0.2918	−0.5101
4	34.1513	34.0378	33.0019	0.2075	0.2744	−0.5321
5	33.6781	33.7644	32.5137	0.1012	0.1791	−0.559

Table 5: Average PSNR and Q for representative peers in each class (*Static* + *Base* method).

	PSNR			Q		
Class	$F = 2$	$F = 3$	$F = 4$	$F = 2$	$F = 3$	$F = 4$
1–5	30.6046	30.675	30.2024	−2.12792	−2.09624	−2.23894

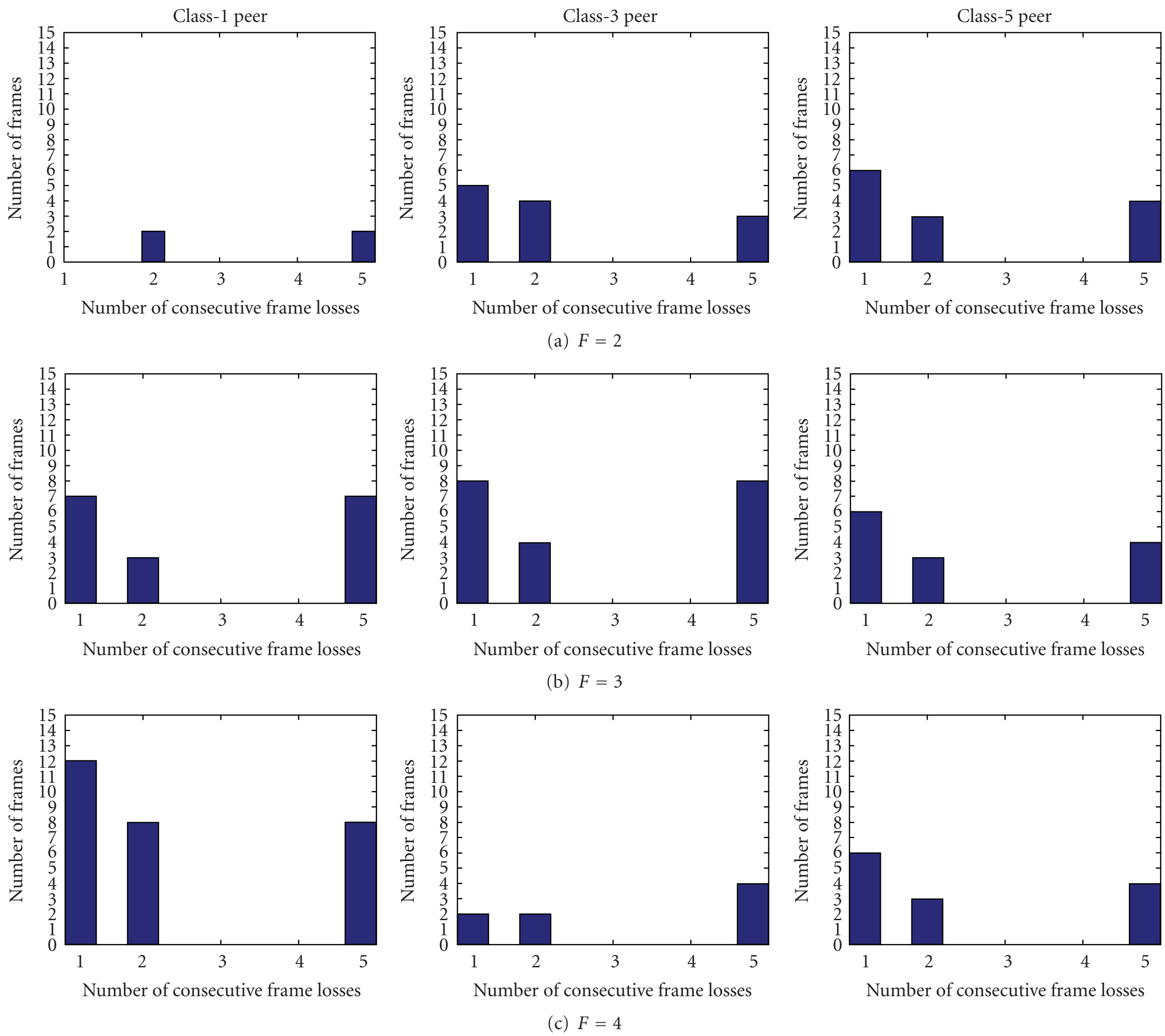
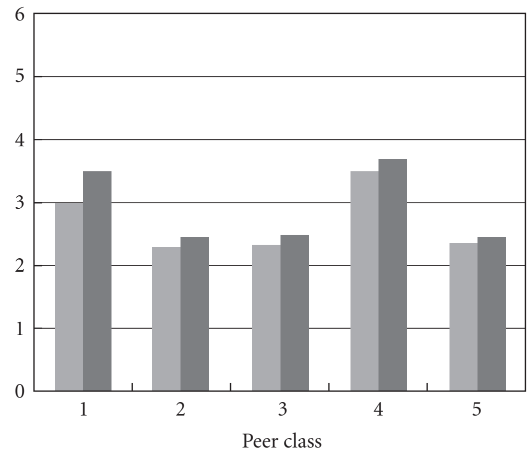
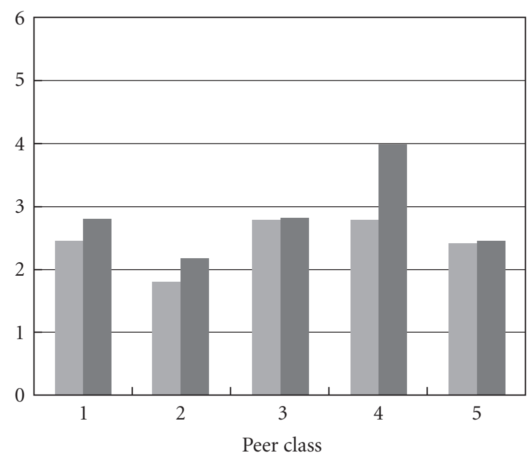


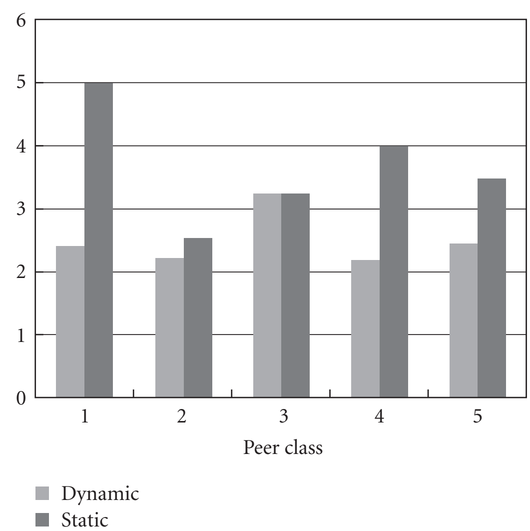
Figure 15: Histograms of consecutive frame losses (*Static* + *Base/FGS* and *Static* + *Base* methods).



(a) $F = 2$



(b) $F = 3$



(c) $F = 4$

Figure 16: Period of average length of consecutive losses for the five representative peers - *Static* and *Dynamic* methods.

In order to work with a more efficient scheme, as already introduced in Section 3.3, the leaf nodes have been assigned to the peers in the last but one level in a round-robin fashion.

We first show in Figures 8, 9, and 10 how the system works, analyzing the behavior of three representative peers (peer of class 1 on the left column, peer of class 3 in the middle one, and peer of class 5 on the right one) taken as representatives of their corresponding classes.

More specifically, for each considered peer, Figures 8(a) and 8(b) present the state of the underlying Markov chain of it uplink bandwidth, and the generated uplink bandwidth, respectively, and show how the Markov model is able to impress the desired temporal correlation on the bandwidth process. Then in Figures 9(a), 9(b) and 9(c) we observe how peers change level during their lifetime according to the behavior of the instantaneous uplink bandwidth from the source, and therefore are strongly influenced by the class they belong to.

Figures 10(a), 10(b), and 10(c) show the time behavior of the number of children for the considered peers. A peer of class 5 (third column) is usually a leaf of the tree, that is, it has no children, and it does not frequently become an internal node. Different observations hold for the considered peer of class 1 (first column), which has always a positive number of children for all the values of $F \in \{2, 3, 4\}$. For $F = 3$, the peer of class 1 changes level quite often, and, consequently, it gets either 2 or 3 children according to the level where it is situated. The peer of class 3 (second column) changes it level quite often for $F = 3$ and, for this reason, it number of children varies from 0 to 2.

The reader notes that Figures 8, 9, and 10 are strictly correlated. Peers with low bandwidth values get lower positions in the tree, whereas peers with higher bandwidth values lie in the upper levels.

For example, around the time instant 200 Seconds, the peer of class 1 (first column) has a bandwidth which rapidly increases (see Figure 8(b)) and, consequently, this peer is immediately moved from level 2 to level 1 for $F \in \{2, 3, 4\}$. Consequently, the number of children increases and it is exactly equal to F, for each $F \in \{2, 3, 4\}$.

4.4.2. Performance Analysis at the Application Level. In this section we will show some statistics concerning the quality of the video received from the source by the peers we are considering as representative of each class.

Table 2 shows the average PSNR and the average Quality parameter Q defined in Section 4.3 (3), measured at destination on the overall sequence by peers of each class for different values of F.

Figure 11 shows the PSNR per frame for the videos received by the three representative peers (one of class 1, one of class 3 and one of class 5) when both Base and FGS layers are transmitted (in black), and the PSNR difference between the video transmitted with both Base and FGS layers and the video transmitted with just the Base layer (in gray). For each value of F, when both Base and FGS layers are used, the PSNR is always higher, or, at the most, equal (difference equal to 0) to each other.

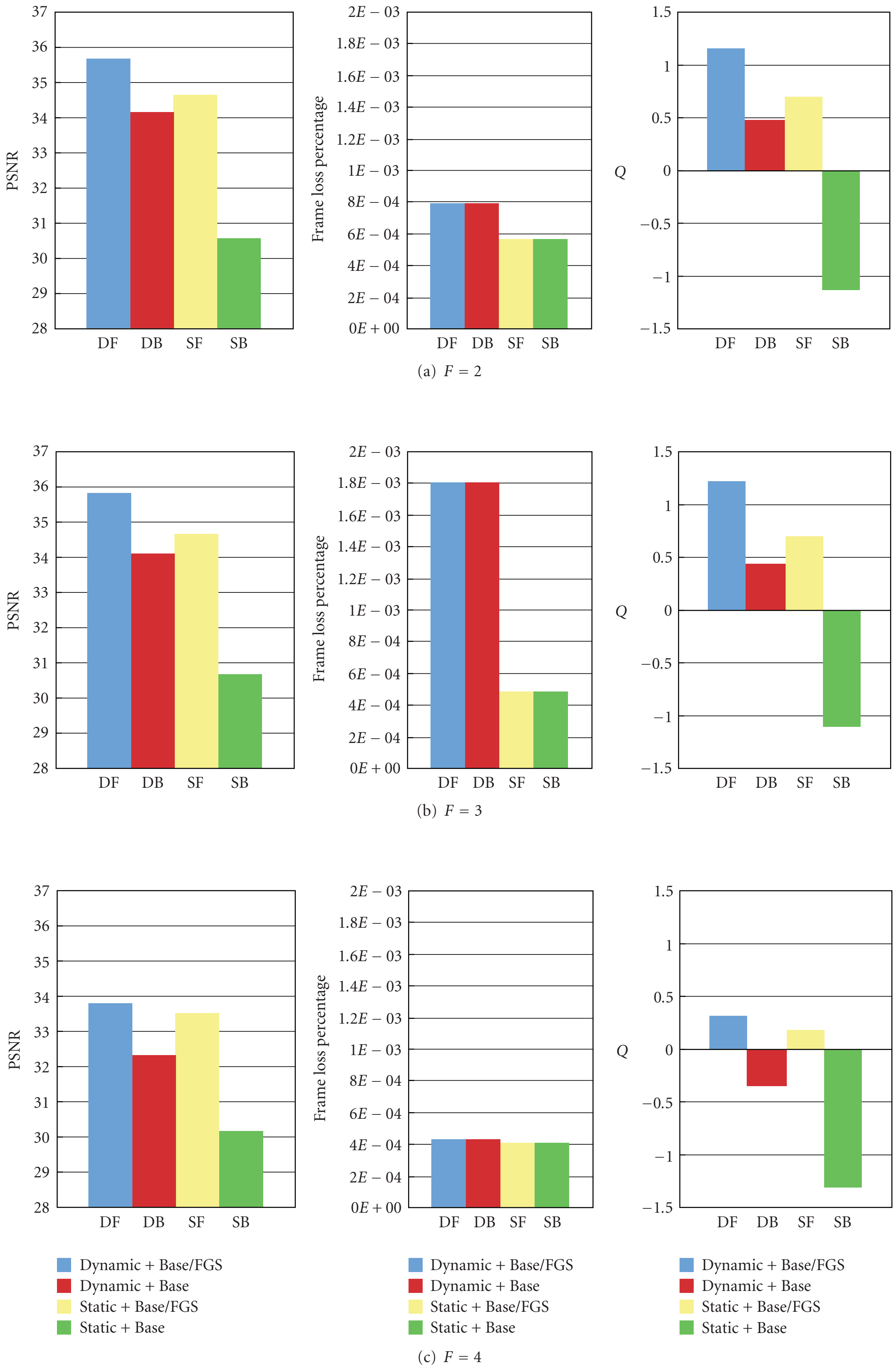


Figure 17: Global results for each method.

As discussed in Section 2, we have considered the video in QCIF format using the pattern IBBPBB as Group of Picture (GoP) structure. As a consequence of the application of this pattern, we remark that when a frame is lost, an error propagation action occurs as follow.

(i) If an I-frame gets lost then all the other frames of the same GoP and the two ending B frames of the previous GoP are lost as well.

(ii) If a P-frame gets lost then the two starting and the two ending B frames of the same GoP are also lost.

(iii) If a B-frame gets lost then nothing happens since B frames are not needed for the reconstruction of the other frames.

For each frame of the GoP, we have also computed it loss percentage (for $F \in \{2, 3, 4\}$) shown in Figure 12. The reader notes that, as introduced in Section 3.1, I-frames have the priority to be transmitted with respect to the other GOP frames because an I-frame loss causes 7 more frame losses. It is for this reason that, in our system, it is more difficult to incur in I-frame losses (i.e., charts in Figure 12 do not show any loss of I-frames). Each chart in the table has the peer class as x-axis, and the frame loss percentage as y-axis. For each value in the x-axis there are five columns related to the six GOP frames (since the one related to I-frame is always 0).

Figure 13 shows a histogram of the number of consecutive frames which are lost, calculated for the three representative peers and for $F \in \{2, 3, 4\}$. Let us note that there are no more than 5 consecutive frame loss: the reason is no I-frame has been lost (an I frame should have caused a number of eight consecutive lost frames: BBIBBPBB). Instead, in some cases, there are 5 consecutive lost frames, due to the corruption of the Base layer of a P-frame (in this case both the two B-frames before and the two B-frames after the lost P-frame cannot be decoded, and therefore the following pattern is lost: BBPBB). Let us note that a combination of losses that causes three or four consecutive lost frames does not exist.

4.5. FGS Performance Assessment. In this section we show how our video transmission system greatly improves on scenarios where either the tree is created but is not updated according to peer uplink bandwidth variations (and therefore it is static), or the enhancement layer for video transmission is not sent at all (i.e., FGS encoding is not applied). More specifically, in the following we will consider for comparison the four combinations of the above situations that can be classified and called as follows:

(1) *Dynamic + Base/FGS* (DF): a peer-to-peer network (with $F \in \{2, 3, 4\}$) where peers are organized with a tree topology which changes dynamically using both Base and FGS layers for video transmission; this is the novel method we propose in this paper;

(2) *Dynamic + Base* (DB):a peer-to-peer network (with $F \in \{2, 3, 4\}$) where peers are organized with a tree topology which changes dynamically; in such a case each peer receives the Base layer only of the whole movie;

(3) *Static + Base/FGS* (SF): a peer-to-peer network where peers are organized in a static tree (peer position is fixed in the tree for the whole transmission), and the source uses both Base and FGS layers for video transmission;

(4) *Static + Base* (SB): a peer-to-peer network where peers are organized in a static tree (peer position is fixed in the tree for the whole transmission); in such a case each peer receives the Base layer only of the whole movie.

For such scenarios, for each class of peers and values of $F \in \{2, 3, 4\}$ we show the PSNR, the Q value, the frame loss percentage and the distribution of consecutive frames losses.

We have already shown the above parameters for the novel method we propose in this paper (i.e., *Dynamic + Base/FGS*) in Table 2 and Figures 12 and 13. As far as the frame loss percentage and consecutive frame losses are concerned, let us notice that only the Base-layer frames contribute to their computation; therefore the methods *Dynamic + Base/FGS* and *Dynamic + Base* produce the same results (shown in Figures 12 and 13) for them. However, the PSNR and Q values for *Dynamic + Base* method are different since the FGS layer gives more details in terms of video quality. Table 3 shows such values. As expected, the PSNR and Q values for each combination of F and peer class are worse than the corresponding PSNR and Q values achieved by applying the *Dynamic + Base/FGS* method, and shown in Table 2.

The PSNR and Q values produced by *Static + Base/FGS* and *Static + Base* methods are shown in Tables 4 and 5, respectively. Similar considerations taken for the Dynamics methods hold: as shown in Figures 14 and 15 they produce the same results for the frame losses and the consecutive frame losses. In addition, from Figure 15 we can note that *Static* methods incur in a number of consecutive frame losses much higher than the ones of *Dynamic* methods.

As seen above for the Dynamics methods, the PSNR and the Q values for the *Static + Base/FGS* method are better than those produced by the *Static + Base* method.

In order to better analyze the impact of the applied method over the number of consecutive frames, we have calculated the average length of the periods of consecutive losses for both *Static* and *Dynamic* methods, for $F \in \{2, 3, 4\}$. The results are shown in Figure 16.

Finally, we have produced Figure 17 showing global results of each of the four methods above discussed for each of the variables of interest (PSNR, frame losses and Q value), and for each value of $F \in \{2, 3, 4\}$. Each value is calculated as the average of each performance parameter over all the peers in the network. It is evident how the PSNR and the Q values are much better for the proposed method (i.e., *Dynamic + Base/FGS*). *Dynamic* methods, for $F \in \{2, 3\}$, present a slightly higher value of frame losses: this is due to the size of the Base layer which is bigger, and therefore more vulnerable, than the one obtained using *Static* methods.

5. Conclusions and Future Work

This paper proposes a multipoint video transmission framework over a heterogeneous content distribution P2P network. The source generates the video flow by using an MPEG-4/FGS encoder, in such a way that the number of losses occurring at the Base-layer stream are minimized, even in the presence of short-term bandwidth fluctuations.

The FGS layer is sent together with the Base layer, but with a lower priority. The source uses a rate controller to regulate the encoding rate of the Base layer, according to an estimation of the bandwidth of the overlay network bottlenck link. A protocol is defined in order to provide the source with this information.

A case study is introduced to evaluate the performance of the proposed framework and the improvements obtained with respect to some reference cases in which FGS is not applied and/or the overlay network topology is not dynamically reorganized.The problem of organizing the tree in realtime is discussed and some techniques are proposed and compared.

As future work we plan to investigate the possibility to apply the proposed scheme for 3D-video streaming on several application scenarios: *three-dimensional television (3D-TV)*, *free viewpoint television (FTV)*, and *multi-view video coding (MVC)*.

Another task to take into account is to perform a statistical analysis of the delay, and in particular the delay jitter that, as discussed in Section 3.3, may result critical for frequent topology changes, particularly for realtime applications.

A feasible solution would be to design an adaptive playout buffer to reduce packet discarding at destination, and a topology management strategy which is able to wisely decide the position of each peer in the tree in order to avoid delay jitter that cannot be compensated even with an adaptive playout buffer.

Last but not least, in presence of high churn rate, the management of transitory peers will be investigated in order to maintain the architecture as more robust as possible.

Acknowledgment

This work is partially supported by the Italian MIUR PRIN 2007 project “Sorpasso”. Moreover, the work leading to this invention has benefited from a fellowship of the Seventh Framework Programme of the European Community [7° PQ/2007-2013] regarding the Grant Agreement n. PIRG03-GA-2008-231021.

References

[1] S. D. Sevetto and K. Nahrstedt, “Broadcast quality video over IP,” *IEEE Transactions on Multimedia*, vol. 3, no. 1, pp. 162–173, 2001.

[2] D. Xu, M. Hefeeda, S. Hambrusch, and B. Bhargava, “On peer-to-peer media streaming,” in *Proceedings of the 22nd International Conference on Distributed Computing Systems*, pp. 363–371, Vienna, Austria, July 2002.

[3] S. Banerjee, B. Bhattacharjee, and C. Kommareddy, “Scalable application layer multicast,” in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pp. 205–217, Pittsburgh, Pa, USA, 2002.

[4] Y. H. Chu, S. G. Rao, and H. Zhang, “A case for end system multicast,” in *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pp. 1–12, Santa Clara, Calif, USA, June 2000.

[5] V. N. Padmanaban, H. J. Wang, and P. A. Chou, “Resilient peer-to-peer streaming,” in *Proceedings of the 11th IEEE International Conference on Network Protocols*, pp. 16–27, Atlanta, Ga, USA, November 2003.

[6] M. Castro, P. Druschel, A. Kermarrec, A. Nandi, A. Rowstron, and A. Singh, “SplitStream: high-bandwidth multicast in cooperative environments,” in *Proceedings of the 19th ACM Symposium on Operating Systems Principles*, pp. 298–313, Bolton Landing, NY, USA, 2003.

[7] T. T. Do, K. A. Hua, and M. A. Tantaoui, “P2VoD: providing fault tolerant video-on-demand streaming in peer-to-peer environment,” in *IEEE International Conference on Communications*, vol. 3, pp. 1467–1472, June 2004.

[8] D. A. Tran, K. A. Hua, and T. T. Do, “A peer-to-peer architecture for media streaming,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 1, pp. 121–133, 2004.

[9] D. Jurca, J. Chakareski, J.-P. Wagner, and P. Frossard, “Enabling adaptive video streaming in P2P systems,” *IEEE Communications Magazine*, vol. 6, pp. 108–114, 45.

[10] T. Nguyen, K. Kolazhi, R. Kamath, S. Cheung, and D. A. Tran, “Efficient multimedia distribution in source constraint networks,” *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 523–537, 2008.

[11] E. Setton, J. Noh, and B. Girod, “Rate-distortion optimized video peer-to-peer multicast streaming,” in *Proceedings of the ACM Workshop on Advances in Peer-to-Peer Multimedia Streaming*, pp. 39–48, Singapore, November 2005.

[12] H. Radha and Y. Chen, “Fine-granular-scalable video for packet networks,” in *Proceedings of Packet Video Workshop (PV ’99)*, New York, NY, USA, April 1999.

[13] W. Li, “Overview of fine granularity scalability in MPEG-4 video standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 301–317, 2001.

[14] H. Radha, M. van der Schaar, and Y. Chen, “The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP,” *IEEE Transactions on Multimedia*, vol. 3, no. 1, pp. 53–68, 2001.

[15] K. Choi, K. Kim, and M. Pedram, “Energy-aware MPEG-4 FGS streaming,” in *Proceedings of the 40th Annual Design Automation Conference (DAC ’03)*, pp. 912–915, ACM, Anaheim, Calif, USA, 2003.

[16] J. Zhou, H.-R. Shao, C. Shen, and M.-T. Sun, “Multi-path transport of fgs video,” in *Proceedings of Packet Video Workshop (PV ’03)*, Nantes, France, April 2003.

[17] F. Wu, S. Li, and Y.-Q. Zhang, “A framework for efficient progressive fine granularity scalable video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 3, pp. 332–344, 2001.

[18] Y.-K. Wang, “Method, device and system for effective fine granularity scalability (FGS) coding and decoding of video data,” International Application N. PCT/IB2006/000631, International Filing Data 22.03.2006, 2006.

[19] "Intel(r) integrated performance primitives (intel(r) ipp)," http://www.intel.com/cd/software/products/asmo-na/eng/302910.htm.

[20] A. Lombardo and G. Schembra, "Performance evaluation of an adaptive-rate MPEG encoder matching intserv traffic constraints," *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 47–65, 2003.

[21] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression: Optimal Video Frame Compression and Object Boundary Encoding*, Springer, New York, NY, USA, 1997.

[22] C. F. Chang and J.-S. Wang, "A stable buffer control strategy for MPEG coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 6, pp. 920–924, 1997.

[23] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, New York, NY, USA, 1996.

[24] B. Krithikaivasan, Y. Zeng, K. Deka, and D. Medhi, "Arch-based traffic forecasting and dynamic bandwidth provisioning for periodically measured nonstationary traffic," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 683–696, 2007.

[25] G. Platt and J. Y. Khan, "Adaptive jitter compensation on the downlink of future mobile multimedia communications systems," in *IEEE International Conference on Communications*, vol. 3, pp. 1887–1891, May 2003.

[26] M. Kalman, E. Steinbach, and B. Girod, "Adaptive media playout for low-delay video streaming over error-prone channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 6, pp. 841–851, 2004.

[27] O. Hashida, Y. Takahashi, and S. Shimogawa, "Switched batch bernoulli process (SBBP) and the discrete-timeSBBP/G/1 queue with application to statistical multiplexer performance," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 3, pp. 394–401, 1991.

[28] "Bandwidth generator tool," http://www.diit.unict.it/arti/Tools/BandwidthSim.zip.

[29] G. Hauske, R. Hofmeier, and T. Stockhammer, "Subjective image quality of low-rate and low-resolution video sequences," in *Proceedings of the 8th International Workshop on Mobile Multimedia Communications*, Munich, Germany, October 2003.