Project acronym: CONCERTO

Project full title: Content and cOntext aware delivery for iNteraCtive multimEdia healthcaRe applicaTiOns

Grant Agreement no.: 288502

Deliverable 5.1

Comparative study of cooperative communications and networking solutions for MSHTM applications

Contractual date of delivery to EC:	T0+11
Actual date of delivery to EC:	31/10/2012
Version:	1.1

Lead Beneficiary:	NTUK
Participants:	VTT, NTUK, CNIT, BME, UOS, KU

Estimated person months:	14
Dissemination Level:	*PU*
Nature:	*R*
Total number of pages:	70

Keyword list:communication, transmission, media caching, HTTP streaming, mobility, LTE, wireless, QOS, radio resource, cross layer optimization, relay, coding, modulation, cooperative communications

Executive Summary

The MSHTM (Multi-Source and Holographic Tele-Medicine) applications require improvement of the communications system. This document describes the-state-of-the-art transmission solutions relevant for the CONCERTO scenario defined in [77], and identifies the limitation expected to be improved during the WP5 activities of the project. The document is organized by splitting the transmission in five layers, from the application layer to the physical layer. At the application layer, the use of media caching with free viewpoint video is presented. At the transport layer, a description of HTTP streaming in the context of multi-access network is presented. At the network layer, a description of mobility solutions at the IP layer based on a distributed mobility approach is presented. At the Link layer, three techniques based on LTE and one with WLAN are proposed: A description of the LTE quality of service mechanisms and its limitation for internet based traffic is presented. The next one relates to the allocation of radio resources uplink scheduling for LTE. The use of Relay, including mobile device used as a relay completes the techniques presented for LTE. For WLAN, a use of cross layer optimization for video streaming with SVC is presented. At the physical layer, a presentation of coding techniques and on cooperative communication solutions is proposed. Finally, a chapter dedicated to conclusions completes this deliverable.

Contents

List of Figures

1 Introduction

The MSHTM (Multi-Source and Holographic Tele-Medicine) applications require improvement of the communications system. This document describes, in the context of CONCERTO, the-state-of-the-art transmission solutions and identifies the limitation expected to be improved during the WP5 activities of project. An introduction to the specific problem of mHealth (Mobile Health) is presented in the next paragraph. Transmission can be viewed at different layers, such as the one defined by the 7 layer OSI model, and also at different location on the data path between the user terminal and a remote end server. The different paragraphs of the deliverable analyze different techniques in various contexts of interest for the CONCERTO project based in particular on the scenario defined in [77]. The deliverable has been structured with a reduced number of layers: Application, Transport, Network, Link, and Physical. Inside each layer presentation, different techniques are described, concerning the transmission from an end to end perspective or concerning a reduced transmission path such as the wireless transmission segment. This deliverable is the starting point of the different activities of WP5.

1.1 Introduction to scalabilty problems of future mHealth services

The emerging concept of mHealth (Mobile Health) represents a natural evolution of eHealth (Electronic Health) from the conventional telemedicine to the progressing applications of mobile and wireless telecommunication. Being part of modern telemedicine – which generally offers higher diagnosis and treatment quality standards, reduces medical costs, and provides possibilities to handle problems of the aging human society – mHealth rises in parallel with the rapid adoption of mobile communications and computing, sensor networking, and advanced wireless technologies into our daily life. Although mHealth is explained as a subgroup of eHealth, it is commonly considered as a unique healthcare paradigm thanks to its special capabilities to provide ubiquitous connectivity, personalized and accurate provision of services, automated and expedited healthcare delivery and reduced costs, such offering patients with more practicality and convenience [33]. Recent years have witnessed an increasing and remarkable advancement of mHealth systems and applications, mainly driven by the fact that for a significant part of the world's population access to various healthcare services is commonly limited by geographical distance, cost, and availability of qualified medical personnel, and by other challenges of societies like declining work force due to the ageing population, fast development of global medical knowledge coming with expensive new diagnostic procedures and infrastructure, and ever-growing expectations for the highest quality healthcare. New, immersive, multimedia-driven, pervasive and interactive mHealth services are rising to provide timely and prompt medical attention while also saving monetary resources. The key areas explored are how mHealth can facilitate wearable sensor system based personal healthcare monitoring, self-diagnosis, early detection and preventive care [49]; mobile-assisted telerehabilitation and therapy [195]; monitoring of soldiers in tactical environments [159]; medical care in emergency and mass casualty situations [184]; intelligent home care [148]; pandemic management [162]; mobile robotic systems for controlling medical devices in isolated areas [76]; personal and mobile wellness/dietary management [187]; and other specialized services like drug intake reminder, raising awareness of health issues, performing health surveys, maintenance of personalized medical records, and medical advice provision/teleconsultation.

However, like other services running in mobile and wireless environments, the efficiency and usability of mHealth applications are substantially impacted by the continuously varying environmental characteristics, scarce network resources, sparse radio bands and bandwidth, battery and computational power of mobile devices, fluctuating delay, jitter and other QoS parameter values. This clearly implies the need of scalable, context- and content-aware mechanisms making applications, service provision and delivery procedures adaptable to the extremely diverse mobile environments. The impacts of fluctuations in context and scarcity of resources are further aggravated by the current trends that prognosticate a massive traffic volume growth in mobile telecommunication systems during 2011-2020 [65]. To date, this traffic explosion is mostly driven by Internet applications providing almost an unlimited scale of interaction, information, and entertainment services for human users. But with the widespread deployment of autonomous, networked and inter-operating sensor technologies, another form of communications called M2M (Machine-to-Machine) or MTC (Machine Type Communication) is emerging nowadays which supposedly will be the leading traffic contributor for mobile Internet evolution in the near future and also has the potential to become a major enabler of successful live mHealth deployments [94]. Hundreds of billions of intelligent, resource constrained sensors and other devices are envisioned to be interacting without human intervention in the Internet of Things (IoT), and generating an enormous amount of data for monitoring, measuring, biomedical signal processing, remote controlling/intervention, safety, or surveillance functions in mHealth or other advanced application areas. An other prominent driving force of mobile traffic evolution is the advancement of high bi-trate multimedia applications. It is foreseen that due to the expansion of data-hungry services like television/radio broadcasting and high-definition Video on Demand, mobile video traffic volume will increase 25 fold between 2011-2016, accounting for over 70 percent of total mobile data traffic by the middle of the decade [65]. This trend is also visible in mHealth: the spreading of multimedia technologies and developments in mobile connectivity gives doctors, hospitals and medical institutions a new set of tools for managing patient care, disease history and billing of medical services, using 2D/3D imaging (e.g., X-ray and ultrasound scans, computer tomographs and magnetic resonance images) for diagnostic purposes and providing radically new types of multimedia services, such as multi-view, stereoscopic and holographic video communications for tele-diagnosis or even helping remote participation of prominent experts in life-saving operations.

The scale of the traffic volume and IoT expansion together with the extremely high requirements of diagnostically useful multimedia transmission techniques poses serious challenges for mobililty management of mHealth. The principal problems, which are is necessary to solve for mobile health to be a common medical and public health practice, are as follows: 1) provide mobile users with ubiquitous communication capabilities over any heterogeneous access infrastructures; 2) support seamless mobility across all available wireless technologies with efficient handover/mobility management; 3) offer medical QoS/QoE for multimedia data transmission during continuously varying network conditions; 4) deliver real-time medical data reliably, even in vertical (i.e., inter-system) handover situations; 5) support multihoming and simultaneous multi-access in order to possess potential to choose between available radio communication technologies; 6) ensure scalability to meet rapid growth of mobile Internet traffic volume and device penetration; 7) granting end-to-end security with efficient encryption of medical information, privacy, authentication, non-repudiation, and access control; and 8) optimize overall system performance within a multilayer/cross-layer management infrastructure.

2 Application layer

2.1 Media caching

Multimedia service is already very popular among users and has contributed to a significant amount of today's Internet traffic. In recent years the server and network bandwidths are the major limiting factors in the widespread usage of video streaming over the Internet, due to its high bandwidth requirements. The features of the digital video streams can be very different (e.g., resolution, bandwidth, duration) depending on the clients population and their requirements. Media objects can be accessed similar to conventional text and images using a video on demand mode, but in some cases the media content is generated real-time and only few second long buffer delay can be accepted. Video caching may decrease the overall load of the network and improve the estimated quality of the users in both cases. To reduce video access latencies as well as server/network loads, an effective means is to cache media data at proxies. Existing proxies are optimized for delivering conventional web objects (e.g., HTML pages or images), which do not meet the requirements of streaming applications. The main differences that have influence on the caching strategy are the stored data size, bandwidth usage, and level of interactivity. In case of 3D video or Free Viewpoint Video (FVV) the strategy must be reconsidered, because multiple high bitrate camera streams must be delivered and cached, or optionally re-coded requiring huge computational resources. Due to limited processing and storing capabilities of a caching proxy, scalability problems are arising by increasing the network complexity.

2.1.1 Related Work on Streaming Media Caching

The goal of media caching solutions is to reduce the network bandwidth cost and delivery latency in a scalable way. An old solution for bandwidth effective delivery is multicasting (broadcasting), but unfortunately IP multicast deployment in the Internet has been slow and even today remains severely limited in scope and reach. An alternative technique for reducing server loads, network traffic and access latencies is the use of proxy caches. Investigation of multimedia caching strategies has begun at the end of 1990's and numerous solutions were published since [278, 253, 35, 291]. The generic architecture of media streaming services with proxy servers can be illustrated as Figure 1 shows.

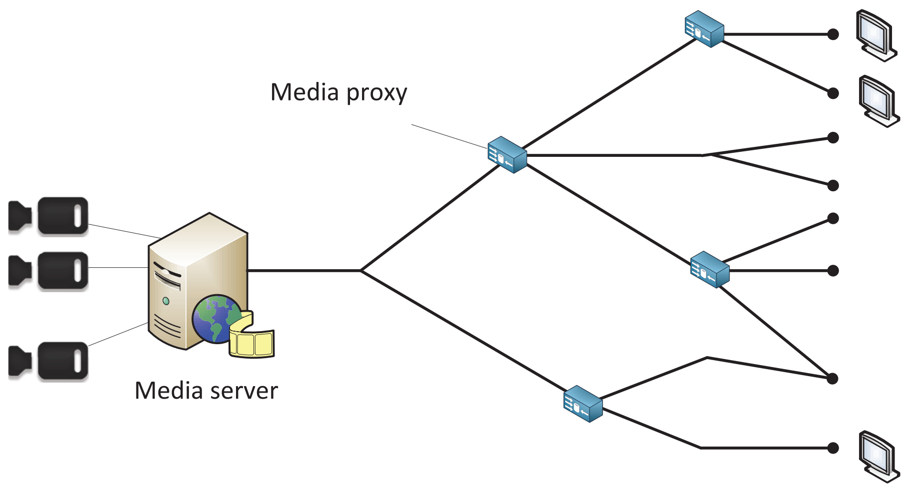


Figure 1: Video streaming with proxy caching.

In the first concepts of caching strategies a group of clients are receiving videos streamed across the Internet from a server via a single proxy. The clients always request playback from the server, but the proxy intercepts the client request and, if the video is present locally, streams directly to the client. If the video is not stored at the proxy, the latter contacts the server. Most existing caching algorithms assume homogeneous clients, which have similar configurations and capabilities. In this case a single version of the cached video stream would match the bandwidth and format demands of all client requests. Of course, what to cache and how to manage cache (e.g., cache placement) remains challenging. According to the selection of the portions to cache, existing algorithms serving homogeneous clients can be categorized [170] as sliding-interval caching, prefix caching, segment caching, and rate-split caching. Sliding-interval caching algorithms [234, 63] cache a sliding interval of a video stream to exploit sequential access of streaming media. If multiple requests for an object arrive in a short period, a set of adjacent intervals can be grouped and stored at the proxy. The cached stream portions will be released only after the last request has been satisfied. Sliding-interval algorithms can significantly reduce the consumed network bandwidth and startup delay for subsequent accesses. However, as the cached portion is dynamically updated with playback, the sliding-interval caching involves high disk bandwidth demands. The prefix caching algorithms [253] cache the initial portion (prefix) of the media stream at a proxy and upon receiving a client request, the proxy immediately delivers the prefix to the client. Meanwhile, the proxy server fetches the remaining portion (suffix) from the server and relays to the client. The advantage of the algorithms is that the startup delay for a playback can be remarkably reduced. Segment caching solutions [234, 299, 191, 91] splits the digital media stream into a series of segments, differentiating their respective utilities (e.g., access statistics, segment length) and making caching decision accordingly. Differently from the aforementioned caching algorithms partitioning a media object along the time axis, in case of rate-split caching [319] the upper part will be cached at the proxy and the lower part will remain stored at the origin server. This solution is particularly attractive for VBR streaming, as only the lower part of a nearly constant rate has to be delivered through the backbone network. Assuming homogeneity of clients is not always possible. Clients behind the same proxy often have quite different requirements on the same media stream, in terms of streaming rates or encoding formats. In order to serve different requests of the customers, a solution is to produce replicated streams of different rates or formats to fulfill the expected media features required by a subset of clients. Keeping more replicas [169] of the media object at the proxy leads to prohibitively high storage and bandwidth demands. An alternative solution is to transcode a media from one form to another with lower rate or different encoding format in an on-demand fashion [303]. Unfortunately, the resource-hungry computation overhead of transcoding prevents a proxy from serving a large, diverse client population. Transcoding process can be avoided if layered SVC video streams are delivered and only the most significant layer (base layer) is cached at the proxy. Another important issue is the cooperation of the proxies, because standalone proxies have lower performance then cooperating ones [29, 61]. Multiple proxies can aggregate cache space, balance loads, provide fault tolerance and improve system scalability. Most of the schemes include segmentation of streaming objects, dynamic caching, and self-organizing cooperative caching. A caching agent, the key component inside the network, communicates via a state distribution protocol and forms dynamic meshes for forwarding streaming media. 3D video is an emerging technology showing a great potential to become the next generation media. Streaming and rendering of dynamic 3D data in real-time requires tremendous network bandwidth and computing resources. Remote rendering [254, 255] is a promising solution to overcome these limitations. The analogy between remote 3D video rendering and proxy caching is based on the fact that also in the first case multiple camera streams must be cached and significant computing resources must be assigned to the 3D proxy. For mobile devices the 3D caching and rendering solutions are more important. There are two major challenges for rendering 3D video on mobile devices. First, 3D video requires very large network bandwidth for transmission depending on the total number of capturing cameras. A real-time 3D video stream may need a Gbps bandwidth, which is far beyond the capacity of wireless networks supported by most mobile devices. Second, although the most advanced smartphones are currently equipped with hardware capable to support OpenGL ES, rending 3D video frames is a very computation intensive task.

2.1.2 Optimization of 3D video caching

Compared to 2D media caching solutions, 3D video caching requires significantly higher storage capacity and computational resources if remote rendering is assumed. In case of 3D video multiple (even more than hundred) capture camera must be used that can be rendered according to the viewpoints requested by the clients. Moreover, the view synthesis algorithms make it possible to forward a video stream from a virtual viewpoint. Free viewpoint video (FVV) [260] and free viewpoint television (FTV) [273] are already hot topics in the 3D multimedia research area. Concerning 3D video caching, most of the arising issues are similar to the 2D case, but the possible solutions may be different:

- Where to deploy the proxies?
- How many proxies are needed?

- Which camera streams and segments should be cashed?
- Which proxies should be extended with rendering functionalities?
- What type of caching algorithm should be used?
- etc.

Of course the answers depend on the scenario in which the 3D video streaming service is provided. Numerous facts should be considered:

- What kind of video streaming service is offered: real-time video or video on demand (VoD)?
- Is the group of served clients homogenous or heterogeneous?
- What is the number of served clients and how are they located (dense or sparse)?
- What is the network topology and the features of the network components?
- How many capture cameras are used and what is the required bandwidth?
- Where is the view synthesis done: at the server, at the proxy or at the client?
- etc.

In an emergency scenario, where the ambulance is equipped with multiple cameras deployed around an injured person, the doctor and his assistant from a remote location should be able to perform the medical examination from any desired viewpoint. In this case real-time 3D video must be transmitted to a few clients only. It is also possible that part of the medical staff is sitting in the office and using a wide screen monitor, while other specialists are watching the scene from other viewpoints on their smartphone. The stream attributes can be adapted to the user equipment if the request message, originated by the user, contains the wanted resolution, bitrate, etc. Nevertheless, continuous feedback messages are needed for the individual viewpoint visualization. Virtual viewpoint generation is based on two or three camera streams that must be available at the place of viewpoint generation (Figure 2).

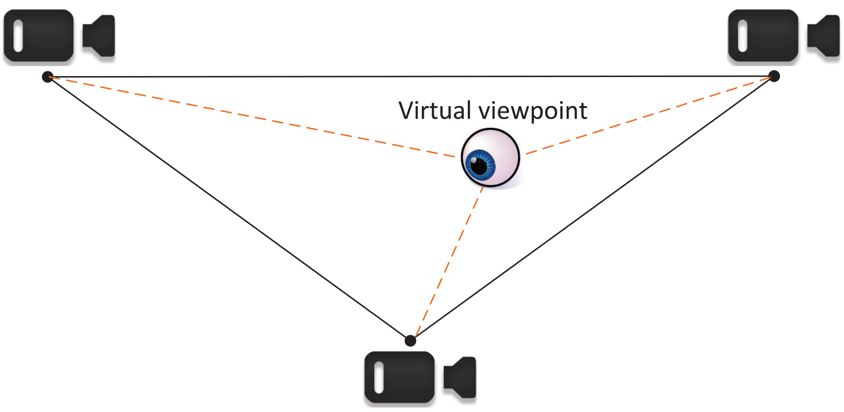


Figure 2: Viewpoint generation.

If the view synthesis is done at client side, two or three camera streams must be forwarded to the user equipment. While the virtual viewpoint is changing, different camera streams may be requested in feedback messages. Feedback messages are also needed if the rendering is at a proxy or at the media server, because the new viewpoint stream must be generated accordingly. For real-time streaming the benefit of caching is the shorter startup and joining time, as well the reduced bandwidth. Bandwidth efficiency can be improved further if rendering is also enabled at the proxy, because lower bitrate free viewpoint video stream is sufficient for the mobile client (Figure 3).

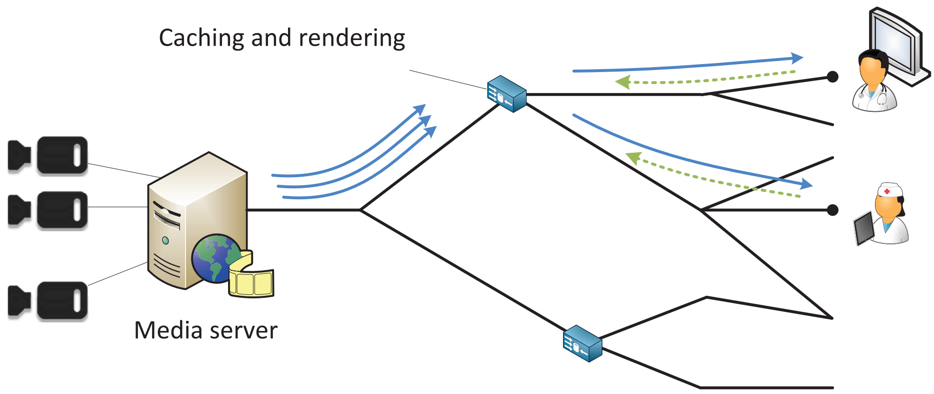


Figure 3: Real-time 3D streaming with caching and rendering.

Another scenario could be an online medical congress, where the camera streams are stored on a server and the users can independently start and stop the playout, freely changing their virtual viewpoint. In this case a 3D VoD service is offered to a large number of clients. Due to scalability issues of computational capacity, the view synthesis can not be done at the media server solely. The viewpoint generation must be a distributed process that requires two or three real camera streams as input for each independent virtual view. The synthesis and rendering functions can be distributed among the media server, the proxies and the client equipment. If the view synthesis is done at the clients, only those two or three real camera streams must be delivered that are needed for the given virtual viewpoint generation. If the view synthesis is distributed among the proxies, the number of served clients behind a proxy is limited, due to the tremendous computational resource requirement of the rendering process (Figure 4).

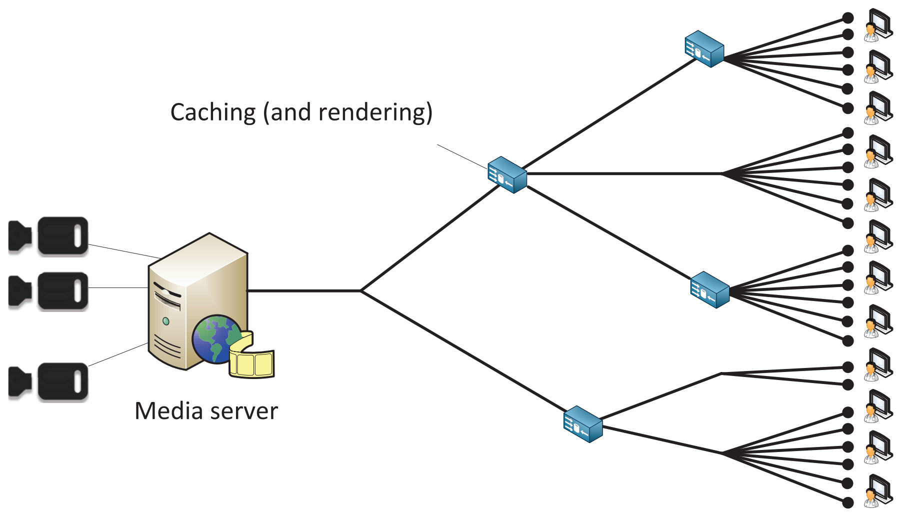


Figure 4: 3D VoD streaming with caching and rendering.

Basically a proxy may cache the segments of a video stream, but in case of 3D services and MSHTM applications it is preferred to support codec functionalities and viewpoint synthesis, so we propose a special proxy element design depicted in Figure 5.

Numerous possibilities can be investigated in order to improve the performance of 3D streaming. Among them we will focus on the proxy localization problem in order to answer where to deploy the proxy functionalities in the

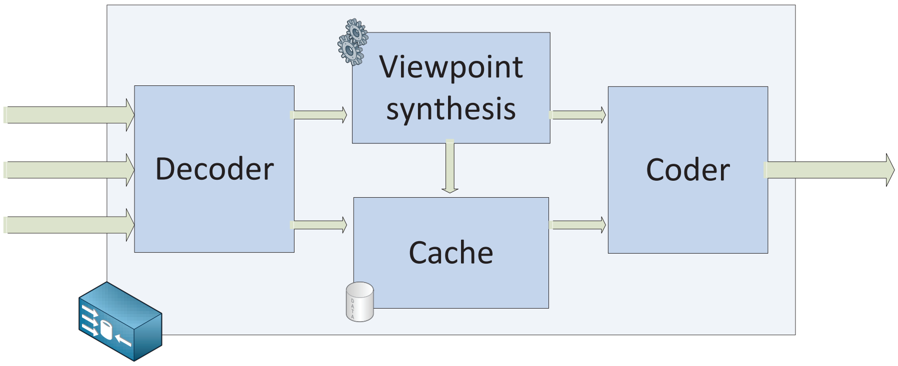


Figure 5: 3D caching and rendering proxy model.

network topology. In order to find the best proxy topology setup, the minimum requirements and the optimization goal must be declared first. The minimum requirements can be found by observing that the computational capacity of the proxy servers and the network links cannot be overloaded. Regarding the optimization goal, different objectives can be indicated:

- using minimum processing resources for rendering (the "green" approach: minimizing the consumption of the proxies)
- minimizing the overall bandwidth usage
- minimizing the video stream access delay
- minimizing the overall delay
- optimizing the overall video quality measured by the customers

In order to formalize the abovementioned goals, the system element parameters must be defined. These elements are the source, where the multiple camera streams are available, the proxy servers, the client, the network links and the camera streams. Each of these elements can be described by different parameters. Source parameters are the computational capacity and storage capacity. The same parameters are needed for the proxies as well. The client, as the element of the system, can be described based on the requested stream. A stream can be defined according to its resolution, bitrate and I-frame frequency.

2.1.3 Conclusions and Next Steps

3D video caching, especially with extended rendering functionalities, is a promising solution for real-time and on demand multimedia services. Numerous questions arise with reference to scalability issues of computational and network bandwidth resources, so we have carried on some interesting research topics in the field of 3D caching and remote rendering. In our research plan the optimal proxy localization issue will be primarily targeted. We will develop optimization algorithms to find the adequate placement of proxies by avoiding computational and bandwidth overloads in the free viewpoint multimedia system for MSHTM applications.

3 Transport layer

3.1 HTTP Streaming

HTTP streaming has gained popularity in the recent years as an alternative for traditional RTP/RTSP/UDP based streaming. The large overhead of TCP, caches, NATs and firewall issues are less problematic in HTTP streaming and also the evolved Internet infrastructure support efficient usage of HTTP [261]. The setup for HTTP streaming is easy and no distinct streaming server is required. From the client-side point of view, clients can access and play the content immediately and seek to point that are not yet downloaded. However, a small start-up delay can occur before the first video segment is generated. One of the strongest assets in HTTP streaming is the capability to adapt to the prevailing channel conditions, efficiently exploited in Smooth Streaming [6], HTTP Live Streaming (HLS) [4] and MPEG Dynamic Adaptive Streaming over HTTP (DASH) [5]. It's worth noticing that there is no single standardized method for HTTP streaming, only several so called *streaming solutions* primary developed in different companies (e.g., Apple, Microsoft). Two other solutions from this area of context are 3GP-DASH [2] (the predecessor of MPEG-DASH) and Adobe's streaming solution to flash, HTTP Dynamic Streaming (HDS) [3].

The bitstream structure in all HTTP streaming solutions is similar. The video file content is splitted into segments and manifest file tells which segments are available for downloading. The segments are usually only few seconds long sections of the original video, which can have several encoded versions to different target bit rate (and therefore quality). The order of the segments is defined in the playlist (manifest file), which can point any of the individual encoded segments. Usually XML format is used for the manifest file, but HLS uses .m3u8 format. The media formats for the content is MPEG-2(.ts) or ISOMBFF(.mp4). Playback and video adaptation is relatively simple; the client asks for the segments to be played. If client receives only small amount of data, it means that channel state is low and therefore the client can request for smaller size segments in order to minimize the number of packet losses. Figure 6 illustrates the principle of HTTP streaming.

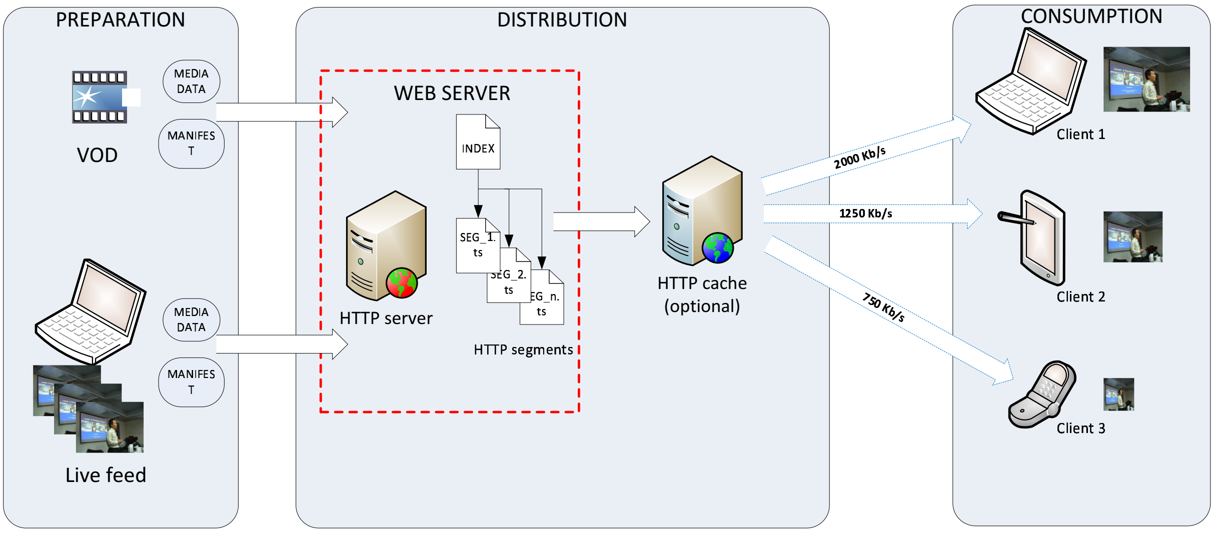


Figure 6: Overall principle of HTTP streaming.

3.2 HTTP streaming and Multi-paths

Sufficient transmission capacity is critical for ensuring the required robustness and reliability for medical applications in wireless networks and mobile contexts. The efficient utilization of today's heterogeneous multi-access network environments can provide one solution to the problem, thanks to the diverse networking capabilities of current mobile devices and their ability to utilize multiple access networks also simultaneously. In this section, we provide a brief introduction to the concept of multipath communication and its utilization for video stream delivery in CONCERTO.

3.2.1 Overview of Multipath Communication

Nowadays IP-based wireless networks comprise heterogeneous multiple access technologies. In this context, very frequently multiple and heterogeneous access networks are available in a given area. A typical example is the case of a wide area network (e.g., 3GPP EDGE/UMTS/LTE) covering an area where a WLAN hotspot is also present. This creates the opportunity for a user whose terminal supports both technologies to combine the transmission capacity available in the distinct access networks by splitting his/her traffic among the networks, and thus to maximize his/her communication bandwidth.

This capability, referred to as multi-homing and illustrated in Figure 7, is especially useful in scenarios where no single access network available to the user is capable of supporting his/her communication sessions due to technology-originated limitations or congestion. However, for such multipath communication to be useful also for mobile users, the implementation should support host mobility, meaning that the user's communication sessions are maintained (i.e., rerouted) dynamically as the user moves among the coverage areas of individual networks in a heterogeneous network environment.

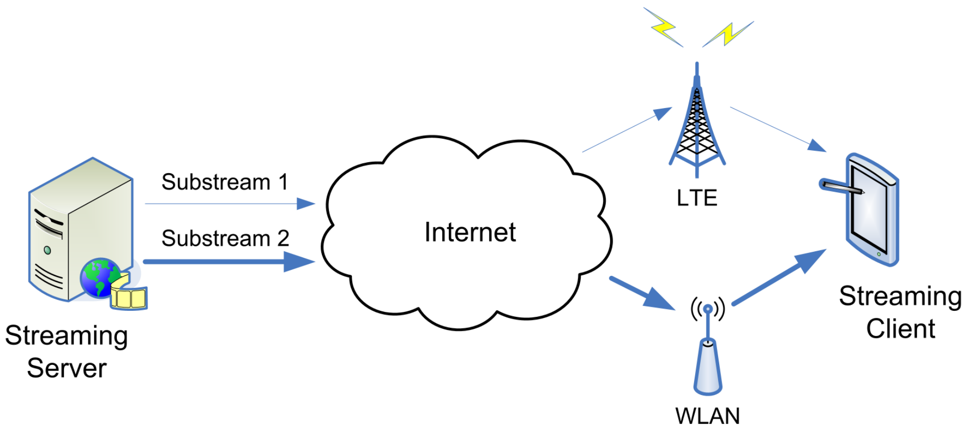


Figure 7: Multipath transmission concept in the context of video streaming: the server splits the video data for transmission over multiple paths and the client combines it before playout.

There are different solutions developed by the IETF for implementing multi-homing and mobility in IP-based heterogeneous networks. These can be grouped according to the protocol layer they are operating on: The network layer solutions include, for instance, the Mobile IPv6 protocol with a multi-homing extension [208, 288, 280] and SHIM6 [109, 200]. The Host Identity Protocol (HIP) based solution [194, 209] introduces an intermediate host identity layer between the network and transport layers. On the transport layer, we have the Stream Control Transmission Protocol (SCTP) [264] as well as the recent Multipath TCP (MPTCP) protocol [99, 100] that may be used for realizing multipath capability in TCP-based applications. At the application layer, there is a proposal similar to MPTCP also for RTP, namely the Multipath RTP (MPRTP) protocol [256]. Also the transport interface of the Scalable Video Coding (SVC) standard [252, 131] supports inherently the multipath transmission of video streams by allowing the transport of individual SVC layers over distinct RTP sessions (i.e., the RTP multisession transmission (MST) mode) [293, 294]. This is primarily intended for supporting multicast sessions, although some proprietary solutions for point-to-point sessions have also been proposed [266]. This latter approach, nevertheless, requires additional support from the video streaming client for dynamically requesting the SVC layers according to the available network interfaces/paths.

For the HTTP streaming discussed in Section 3.1, we propose to use the MPTCP protocol for realizing multi-path capability in heterogeneous IP networks in CONCERTO.

3.2.2 Multipath TCP

MPTCP allows communication over multiple network paths, concurrently. This enables more efficient resource usage as well as improved user experience through improved resilience to path failure and higher throughput compared to traditional TCP exploiting communications through a single path. MPTCP is seen as a natural choice for HTTP streaming, as HTTP streaming operates on top of TCP, and MPTCP essentially is a modified version of TCP that implements a multipath transport transparently to the application. In fact, MPTCP was designed to work with legacy applications through the standard TCP API. Nevertheless, any advanced features related to multi-homing and its control will require additional support from the applications. MPTCP also appears to the network as normal TCP connections alleviating any possible middlebox issues (e.g., firewalls) on the service availability.

The protocol stacks of standard TCP and MPTCP are illustrated in Figure 8 for comparison. In MPTCP, there is a functional decomposition of the multipath session management in order to make the functionality transparent and compatible to the applications as well as the underlying network. In other words, MPTCP appears to the network as a set of standard TCP sessions, indicated as subflows in the figure, to provide transport for the individual paths. On top, the MPTCP function manages the individual TCP subflows through implementing path management, packet scheduling, subflow interface, and congestion control.

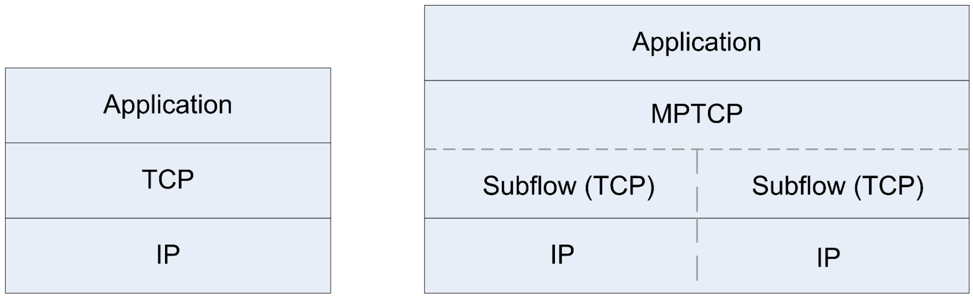


Figure 8: Standard TCP and MPTCP protocol stacks.

At the beginning of a new session, the hosts negotiate MPTCP support in the initial SYN exchange (using the MP_CAPABLE option), which allows them to determine their support for MPTCP as well as to exchange information (i.e., 64-bit host-specific keys) used for authenticating the establishment of additional subflows. After the initialization, additional subflows can be opened dynamically during the multipath connection's lifetime. This allows implementing sustainable connections also during host mobility as discussed in [224]. An additional subflow is initiated through normal TCP connection initialization but including the MP_JOIN option. For an additional subflow, the hosts may use either the same pair of IP addresses as for the first subflow, but different ports, or they can use any additional IP addresses the hosts may have. MPTCP also supports the addition or removal of IP addresses on a host and defines mechanisms for the hosts to signal a change in the available addresses either implicitly or explicitly.

After the establishment of multiple subflows, the sending host's MPTCP connection splits application data across the subflows. Additional TCP options are used to allow the receiver to reconstruct the received data in the original order. During normal MPTCP operation, the Data Sequence Signal (DSS) TCP option is used for signaling the data required for multipath transport. This data comprises the Data Sequence Mapping, which defines how the sequence space on the subflow maps to the connection level, and the Data ACK, for acknowledging the receipt of data at the connection level. MPTCP uses a 64-bit Data Sequence Number (DSN) to number all data sent over the MPTCP connection. Each subflow has its own 32-bit sequence number space and these sequence numbers have relevance only in the context of a subflow. The Data Sequence Mapping components of the DSS option allow mapping the subflow sequence space to the data sequence space. This also allows the retransmission of data on different subflows (mapped to the same DSN) in the event of failure. The Data ACK is a cumulative acknowledgement for the packets received successfully through the connection (i.e., MPTCP level). Its operation is similar to TCP SACK.

Besides having its own sequence space, each MPTCP subflow maintains its own congestion window so that it can adapt to conditions along the path. However, the subflows cannot implement the standard TCP congestion control algorithm as they are required to support fairness also in the case when multiple subflows share a common bottleneck link. Moreover, the multipath congestion control should efficiently move traffic away from congested areas to improve robustness and overall throughput (i.e., resource pooling).

One congestion control algorithm is proposed in [222] for MPTCP. This algorithm couples the additive increase function of the individual subflows and uses an unmodified TCP behaviour in case of a packet drop. As a summary from [223], the operation of the subflow congestion control algorithm is as follows:

- For each ACK on subflow r, increase the window w_r by $min(alpha/w_{total}, 1/w_r)$.
- For each loss on subflow r, decrease the window w_r by $w_r/2$.

where w_{total} is the sum of the congestion windows on all subflows in a connection whereas $alpha$ determines the aggressiveness of all the subflows and is calculated as described in [222]. The proposed algorithm meets its targets in terms of fairness, but has problems in fulfilling the resource pooling requirement [222]. Nevertheless, it is expected that there will be different congestion controllers for MPTCP, each aiming to achieve different properties in the resource pooling/fairness/stability design space, as well as those for achieving different properties in Quality of Service (QoS), reliability, and resilience. For instance, the paper [224] shows how MPTCP may be used for handling mobility in a heterogeneous network environment and to achieve different optimization goals such as higher throughput or lower energy consumption.

MPTCP disconnection is performed by sending a Data FIN packet to the other communicating host. This signals the successful transmission of all the packets and the end of the multipath connection.

MPTCP is primarily concerned with end-to-end multiple paths, where one or both of the end hosts are multi-homed. It may also have applications, where multiple paths exist within the network and can be manipulated by an end host, such as using different port numbers with Equal Cost MultiPath (ECMP). Compared to other transport protocols supporting multi-homing, e.g., SCTP, MPTCP aims at wide-scale deployment recognizing the importance of application and network compatibility goals.

From a standardization perspective, the specification of MPTCP is currently on-going in IETF in the corresponding working group. There is also an open source MPTCP project for Linux: the Multipath TCP Linux kernel implementation by the IP Networking Lab in UCL, Belgium (http://mptcp.info.ucl.ac.be/). The implementation was published in [41] with results from an experimental evaluation.

3.2.3 Considerations for Adaptive HTTP Video Streaming over Multiple Paths

In CONCERTO, we propose to use MPTCP to transport adaptive HTTP streams. For this purpose, the performance of the rate adaptation algorithms developed for single-path HTTP streaming are to be evaluated across multiple paths. All the requirements posed to the MPTCP API and cross-layer optimization are to be analysed in order to enable efficient utilization of the available access networks for mobile medical applications. The results of this work will be reported in the future deliverables of WP5.

4 Network layer

4.1 Distributed and dynamic mobility management

As presented in the paragraph 1.1, the emergence of mHeath services requires that significant progress are made for mobile users. In CONCERTO we propose a novel, jointly optimized distributed and dynamic mobility management system for the challenges imposed by MSHTM multimedia delivery of advanced mHealth services in evolving mobile environments.

Table 1: Drawbacks of centralized mobility management approaches.

Drawback	Description	Categories
Single Point of Failure	Any deliberate or accidental fault of the central node cuts of the whole mobile network.	Security, Stability
Inefficient Routing	Non-optimal routes via the central node often result in a longer route. It causes QoS deterioration and may lead to heavy network congestion in the core.	Resources, Speed
Signaling Overhead	If the mobile network serves high amount of mobile nodes, their signaling (binding) messages extremely load the central entity, which increases the latency, and decreases the stability.	Speed, Stability
Backhauling	A centralized approach leads to backhauling all traffic to the node, which has unfavorable operational consequences.	Stability
Latency	The latency and interruptions could be excessively high when updating binding, switching interfaces or networks.	Speed
Scalability	The load balancing or passing service over to another node are not solved.	Stability
Wasting Resources	Without scalability options the central node (and network) should to be calibrated for high stress or load.	Cost
Cost and complexity	The central node (and network) needs to be calibrated for high load, which requires very expensive hardware.	Cost

4.1.1 Problem statement

Mobility solutions at the network layer (for IPv4 and IPv6) have been specified by IETF years ago. The solutions could be categorized by the point of mobility management. If binding between the moving entity and the stationary handled by the moving object, it will be called host mobility. If the binding is initiated and updated by the network (Mobility Anchor Point, i.e. an Access Point), it will be called network mobility. All of them provide an IP address (or prefix) to the mobile nodes (MN), which ensures the IP session continuity for the node. This IP address remains constant even as the mobile node attaches to different access networks (AN). This document focuses only for, and primarily for host mobility. The common methodology (for host and network mobility too) requires a central entity, which could handle the moving of the objects. The binding process maintains the mapping between the constant, continuously available IP address, and the instantaneous address. This central service provides the global reachability for each mobile node; independently form its current point of attachment. This central entity is known as Home Agent (HA) in host mobility terminology (MIPv6) [208], and known as Local Mobility Anchor (LMA) in network mobility terminology (PMIPv6) [114]. Table 1 summarizes the drawbacks of the centralized approach, in which all responsibilities and tasks of mobility management is centralized by one node. The remaining part of this chapter categorizes the existing research directions and approaches, which would like to solve or handle the problem of centralized mobility architecture. Afterwards, it introduces the most useful (from implementation point of view), groundbreaking or innovative solutions.

4.1.2 Distributed and Dynamic Approaches

The drawbacks of centralized mobility architectures could be handled by two totally different directions which can be integrated in advanced approaches. On one hand, the central entity or anchor node could be distributed in the network topology. It means the multiplication of the Home Agent or Mobility Anchor Gateway, and some kind of distribution of the tasks. On the other hand, scalability problems could be managed from the signalling activity part of a system by balancing the necessity of mobility services and initiating a new mobile network only if it is really required. These kind of solutions focuse on the sessions which are alive during (or after) the switching between Internet points of attachment: running communication sessions will be kept alive, the previous anchor point will act as a mobility anchor point, and previous IP address will be mapped over the new one until the session will not be not terminated. The first approach is known as Distributed Mobility Management, the second as Dynamic Mobility Management (Table 2).

Table 2: Comparison of Distributed and Dynamic Mobility Management.

	Distributed	**Dynamic**
Pros	• Solves Single Point of Failure • Load balancing at the central node • Scalability • It could solve various issues (see later)	• Solves Single Point of Failure • Effective resource allocation • Very effective routing
Cons	Some kind of synchronization is necessary between the nodes of the central network, which may increase the signaling load.	The network node does not have a constant IP address. It won't be reachable globally. It is not effective for (very) long-term sessions.
References	see in next section	[46, 113, 171, 44, 166]

4.1.3 Partially and Fully Distributed Approaches

In conventional mobility management schemes the central entity manages the binding and tunnels the data packets to and from the mobile nodes. The first is called as the control pane, the second is called as the data pane. The distribution of the control pane requires some kind of synchronization between the anchor nodes, and the load of the data packets is significantly higher than the signaling packets. It follows that the distribution of the data pane only could be quite effective, and it could be simpler than the separation of both panes. These schemes are known as Partially Distributed Mobility Management schemes. If the architecture distributes the signaling tasks as well, it will be called as Fully Distributed Mobility Management. Table 3 compares the two approaches and Figure 9 summarizes the previous sections by depicting the available solutions.

4.1.4 Dynamic Mobility Management Solutions

Philippe Bertin et al. describes in their work [46] a flat IP architecture, especially from the point of view of UMTS networks. In this architecture, when the mobile node attaches to an access network, the AN issues an IPv6 prefix

Table 3: Comparison of Partially and Fully Distributed Schemes.

	Partially Distributed	Fully Distributed
Pros	• Effective routing • No backhauling • Simpler (no state synchronization between central nodes)	• Effective routing • No backhauling
Cons	Single Point of Failure	More signaling message: state synchronization between central nodes
References	[263, 181, 50, 212]	[97, 289, 30, 92]

to it, and the mobile node generates its IP address from it. The MN starts its sessions, and communicates with remote nodes. Once, it moves to another access network, it acquires a new IP address from it. All of the newly initiated session will use this newly issued IP address, but the existing ones will be forwarded between the old and the new access networks.

Fabio Giust et al. proposes [113] a dynamic architecture for network mobility. They distinguish the original and the current point of attachment. Until the session does not terminating, the packets are forwarded from the Anchor MAAR (Mobility Anchor and Access Router) to the Serving MAAR.

The basics of the work [171] of Pierrick Louin et. al are the same as the previously introduced concepts. Their DMA (Dynamic Mobility Anchoring) scheme forwards the existing sessions between the original point of attachment and the current one.

Additionally, there are several IETF Internet Drafts about this topic. A few of them focus for the problem, i.e. [86]; others try to specify network mobility schemes, i.e. [44]; finally some of them describes host mobility approaches, i.e. [166].

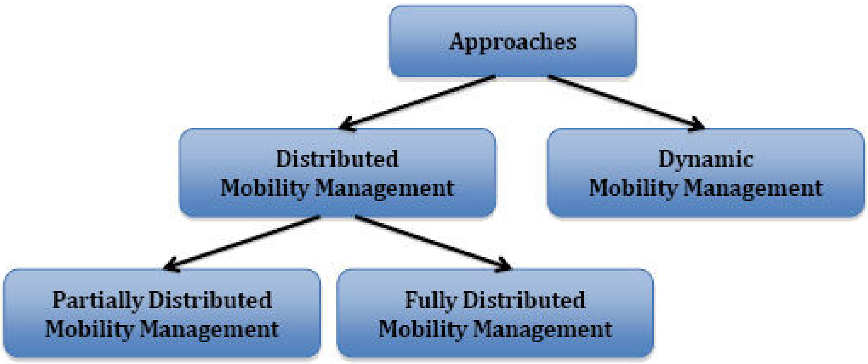


Figure 9: Available DMM approaches.

4.1.5 Partially Distributed Mobility Solutions

Mei Song et al. defined [263] a scheme, which organize the access networks into regions. The movement of the mobile nodes inside these regions is transparent for the central node. This scheme still requires on central entity, which handles the global binding, but a part of movement events (the size of these parts depend on the farseeing selection of them) does not load the central entity. This network architecture is very similar to the Hierarchical Mobile IPv6 (HMIPv6) [262] scheme.

Xuebin Ma et al. studied [181] the optimal settle of the HMIPv6 regions. Their scheme defines different network architecture for each mobile node. Different Access Routers act as MAP for each Mobile Node, and the size of the regions differs too. In this architecture the traffic load of the mobile nodes is distributed between the Access Routers. The Access Routers could to be arranged in ring structures, which provide the best resource allocation.

Chang Woo Pyo et al. proposed [212] a very similar scheme as the previous ones. They call the regions as domains. Any of the Access Routers within the domain could act as Domain Router. The movement inside the domains is transparent for the outside nodes.

Michael Boc et al. have watched [50] the distribution problem of network mobility architectures. The defined PMIPv6-based architecture separates the data and signaling paths. The LMA remains the responsible for signaling tasks, and newly defined Intermediate Anchors (IA) will be responsible for data paths, but the LMA has to select the corresponding IA.

4.1.6 Fully Distributed Mobility Solutions

Mathias Fisher et al. defined [97] a mobility management architecture, which distributes the management (signaling) and communication (data) tasks as well, by the distribution of the Binding Cache. It stores the Binding Cache in a Distributed Hash Table (DHT). Each Mobility Anchor manages its own part from the overall Binding Cache.

Most of the papers tries to distribute the Home Agent itself topologically, and defines a kind of synchronization mechanism for the parts. Ryuji Wakikawa et al. using [289] IPv6 Anycast Routing to advertise the same IPv6 prefix bay all of the Home Agents. The Home Agents are sending Binding Update Copy messages to each other, to ensure, that all of them shares the same status of the Mobile Network.

The Global HA to HA protocol specification is the closest to the standardization by IETF. Currently it is available [30] in draft format for the public. It specifies the same architecture as the previous one. It defines state synchronization methods between the Home Agents in the network.

There are available more papers in the literature, which summarize the possible solution trends and current research directions. Most of them follow the same categorization as we did (e.g., [62, 45]). From implementation point of view or overviewing the available solutions the Internet Drafts of IETF could be the best choices (e.g., [149, 39]).

4.1.7 Conclusions and Next steps

The proliferation mobile Internet applications and services tend to radically overload the current mobile and wireless architectures deployed nowadays. To fulfil the special reliability, QoS/QoE, and bandwidth requirements of the mHealth sector and MSHTM applications, we should distribute the mobility management, and we should optimize the flow routing between the peers. The recent mobility solutions have to apply the features defined by DMM and have to integrate various optimizations, such as cross-layer optimization, context- and content-awareness. We are currently working on a standard-based, implementable and integrated answer for all the issues of current Mobile IPv6 systems also within the unique context of mHealth use-cases.

5 Link layer

Among the communication layers, the Link layer play important roles more specifically on the radio access. With LTE, the link layer is in charge of providing the Quality of Services (QOS) mecanism, of managing the mobility of User Equipment, and of managing the radio ressources. Optimising the Link layer, or performing cross layer optimization across the link layer are expected to improve the support of MSHTM applications.

5.1 LTE Quality Of Service

5.1.1 3GPP Network architecture

The 3GPP Release 8 has defined in 2008 a major evolution of its network architecture, that introduced the LTE radio access specification. LTE specification largely improves the wireless link with a higher throughput and a lower cost per bit in order to follow the growth of data mobile transmission. Different types of user equipment (UE) will be supported with different bit rate capability. A typical UE (Categorie 3, as specified in [16]) will support a maximum bit rate of 100Mbits in downlink and 75 Mbits in uplink.

The network architecture that support LTE is all-IP based. There is no more circuit switched feature available. The services provided by the network can be real-time and non real time. An end-to-end class-based Quality-of-Service (QOS) architecture is defined for supporting the end to end services. The following paragraph will describe the major component of the architecture and focusing on the network access capabilities and the end to end QOS capabilities.

Notion of teleservices and bearer services The network provide an End-to-End service between the UE and a remote peer entities. The 3GPP TS 22.105 [7], presents the notion of Teleservices, or End to end services, and the notion of bearer services. This is illustrated in the figure 10. A bearer services is defined as the telecommunication services providing the capability of transmission of signals between access points. The bearer services is a notion of low layer of the OSI model (Layer 1, 2 and 3). The user data transported by the bearer service can be view as IP PDUs.

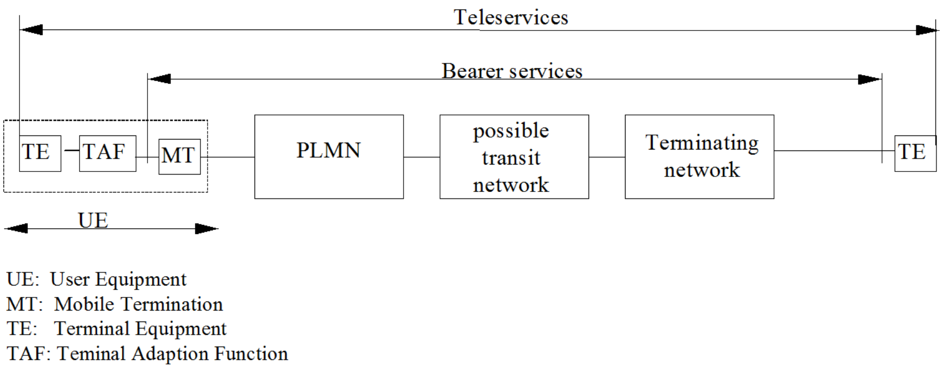


Figure 10: Basic Telecommunication Services

5.1.2 3GPP EPS Architecture

Figure 11 illustrates the overall 3GPP network architecture as presented in TS 23.401 [12] expected for the LTE era. It is marked by the elimination of the circuit-switched domain and a simplified access network. The functional entities depicted in this figure are highly flexible, and can be physically co-located, or reside in dedicated hardware according to the network operators needs. The EPS as Evolved Packet System, is composed of 2 two main parts: the E-UTRAN and the Evolved Packet Core (EPC). The result is a system characterized by its simplicity, a non-hierarchical structure for increased scalability and efficiency, and a design optimized to support real-time IP-based services. The services can be realized by the IMS connected to the PDN Gateway and the PCRF.

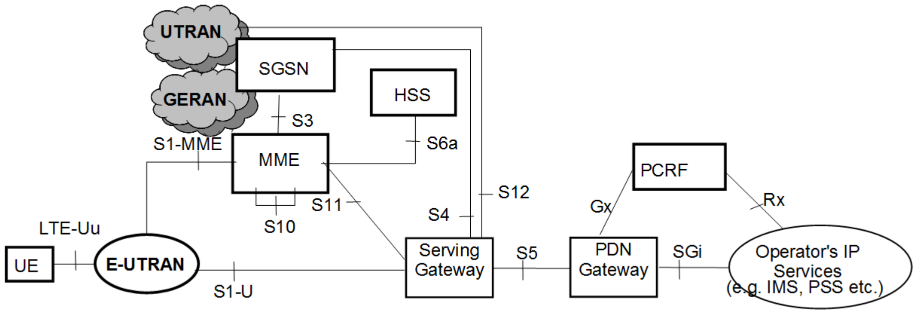


Figure 11: EPS Reference Architecture

Evolved Packet Core (EPC) The EPC, presented in the figure 12 is the IP-based core network defined by 3GPP in release 8 for use by LTE and other access technologies. The goal of EPC is to provide a simplified all-IP core network architecture to efficiently give access to various services such as the ones provided in IMS (IP Multimedia Subsystem). EPC consists essentially of a Mobility Management Entity (MME) and access Gateways for routing of user IP packets.

The EPC is composed of Mobility Management Entity (MME), Serving Gateway (SGW), PDN Gateway (PGW), Policy and Charging Rules Function (PCRF) and Home Subscriber Server (HSS). The following description is limited to the main EPC function regarding the management of UE and UE user plane.

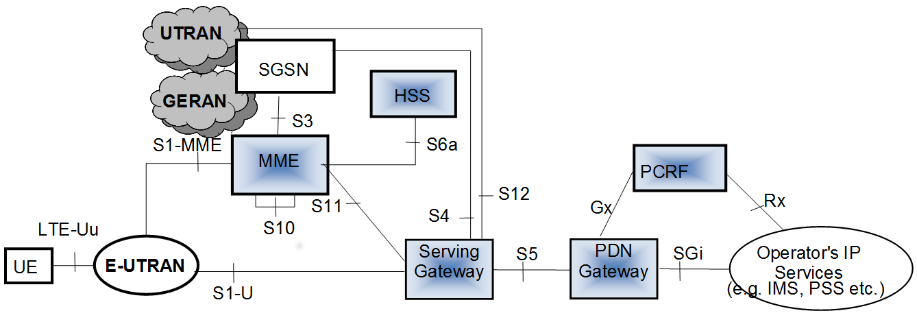


Figure 12: The EPC Network elements

The Mobility Management Entity (MME) realizes the following functions:

- Handling of the NAS signalization exchanged with the UE, that allow managing securely the network attachment with the support the HSS, and the UE mobility based on the notion of tracking area.
- Management of the EPS bearer that support the transport of UE user plane information to/from the relevant PDN gateway. Each EPS bearer will be associated with a specific QOS.

The Serving Gateway (SGW) realizes the following functions:

- Act as a mobility anchor for inter-eNodeB handover
- Perform the packet marking in the uplink and the downlink (DiffServ Code Point in IP packets), based on the QCI of the associated EPS bearer;

The PDN Gateway (PGW) realizes the following functions:

- Interconnect with the external PDNs;
- Perform packet inspection and per-user based packet filtering;

- IP address allocation for UEs;
- Perform the packet marking in the uplink and downlink (DiffServ Code Point in IP packets), based on the QCI of the associated EPS bearer;
- uplink and downlink service level gating control: based on filters (quintuples of source/destination IP addresses and ports, as well as protocol type) the flow of IP packets is allowed or denied;
- Uplink and downlink service level rate enforcement: the amount of IP packets flowing through the node is kept within the applicable limits. Methods to achieve this are rate policing (may lead to dropping of packets)and shaping (e.g., short time buffering), per individual Service Data Flow(SDF);
- Uplink and downlink rate enforcement based on the aggregate maximum bit rate for an APN; it applies to all SDFs of the same APN that are associated with Non-Guaranteed bit rate QCIs. The methods are the same as for the above item;
- downlink rate enforcement based on the accumulated maximum bit rates of the aggregate of service data flows with the same guaranteed bit rate QCI;

The Home Subscriber Server (HSS) is a database that contains user-related and subscriber-related information.

The Policy and Charging Rules Function (PCRF): this entity encompasses policy control decision and flow based charging control functionalities. It provides network control regarding the service data flow detection; gating, QOS and flow based charging towards the PCEF (Policy and Charging Enforcement Function) that is part of the PGW. More details on this entity is provided in chapter 5.1.5.

E-UTRAN The E-Utran (Evolved Universal Terrestrial Access Network), presented in the figure 13 is the radio access infrastructure of an LTE network [17]. It is composed of connected eNodeB (eNB). On one side the eNB interfaces with the user equipment (UE), for transferring over the wireless interface the user plane and control plane information. On the second side, the eNB interfaces with the EPC with the S1 interface. The S1 interface is used to connect to the MME and to the SGW. The eNB performs all the management of the radio resources and the transport of user plane data between the UEs and the EPC.

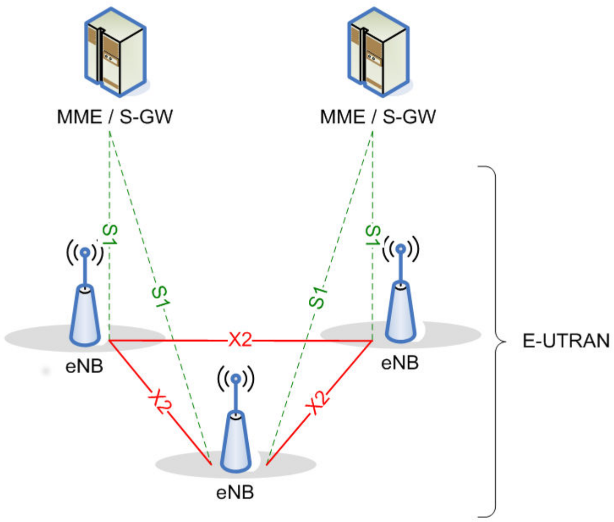


Figure 13: The E-UTRAN Architecture

As the eNB performs the resource management, it performs the allocation of the uplink and downlink radio resources to/from each UE of a cell. This allocation is performed by a scheduler function realized by the MAC layer, that takes into account each radio channels (radio bearers) QOS requirements for all the cell UEs.

5.1.3 Access Network Mechanisms

The e-UTRAN provides mechanisms to control the access of UE to the networks, and thus is the first mechanisms that can allow the access to the network by UEs, depending on the network availability and in particular of the load conditions.

Access network procedure A 3GPP networks broadcasted constantly over the radio interface cell, messages named System Information Block or SIB. In order to access to the network services, after its power-up, a UE first try to get access to the network, by executing the cell selection procedure. During this procedure, the UE scans the supported radio frequency in order to find a radio cell meeting criteria (cell belong to a suitable PLMN, cell is not barred etc ...). In order to limit the access to a particular type of UEs, the UE is associated with an Access Class value stored in the SIM. More details are provided below:

- Classes 0 to 9: are assigned to commercial users. During network congestion phase, the access to a cell can be controlled by a limiting algorithm based on a Barring Factor parameters and a Barring Time parameter.
- Access Class 10: is reserved for the "‘Emergency call"’. If this class is not barred, then any user can use the cell for this particular specific service.
- Classes 11 and 15: are reserved for network operator administrative purpose.
- Classes 12, 13 and 14: Remaining classes are reserved for Public Safety users: Class 12 is for Security Services, Class 13 is for Public Utilities and Class 14 is for Emergency Services.

Any number of these classes may be barred at any one time.

5.1.4 EPS QOS concept

The 3GPP network main feature is to provide a Layer 2 data link between the UE and the PGW that is used to transport the user data, with a defined Quality of Service (QOS). This QOS assure that the user data are transported with a guaranty on the priority, latency, error rates and bit rates. It enables operators to provide services and subscriber differentiation. The simplest solution to ensure a required QOS, could be to over-dimension all the critical resources on the network. Nevertheless, this is not a feasible approach in real life. So to make a QOS mechanism successful, a delicate balance between the complexity and usefulness of the QOS mechanism need to be realized.

An end-to-end class-based QOS architecture has been defined for LTE. This mechanism is based on the concept of data flows and bearers. The class-based QOS approach adopted for LTE has been simplified in comparison to the QOS mechanisms defined for 3G/HSPA. It allows bearers to be mapped to a limited number of discrete classes named QCI.

Bearer To realise an End-to-End service between a UE and a remote peer entity, the 3GPP has defined the bearer service notion, aiming to transport the User IP PDU with a required QOS. In order to provide the end to end bearer services, different bearers are created between the different entities considered. The EPS provides to the UE the IP connectivity to a PDN, like the Internet, or others IP network supporting services such as Voice over IP (VoIP). As presented in the figure 14 the EPS bearer provides a logical transport channel between the UE and the P-GW with a defined quality of service (QOS). The logical link between the P-GW and the peer entity is represented by an external Bearer. EPS bearers represent the layer 2 transport channel between a UE and the P-GW. It allows the UE to access to operator services or to remote peer entities reachable on the IP network connected to the PGW. EPS bearer service includes all aspects to enable the provision of a contracted QOS. It is the basic traffic separation element that enables differential treatment for traffic with differing QOS requirements. Multiple bearers can be established between an UE and the PGW, in order to provide different QOS streams or connectivity to different PDNs [7]. The EPS bearer is mapped on an E-RAB (Evolved Radio bearer) between the UE and the S-GW and a S5/S8 bearer between S-GW and P-GW.

Radio Bearer represents the transport channel between a UE and a eNB. It is used to transmit both control plane (RRC, NAS messages) and user plane (IP traffic). A default EPS bearer is established during attachment and maintained throughout the lifetime of the connection and generally with no guarantee QOS. Dedicated EPS bearers

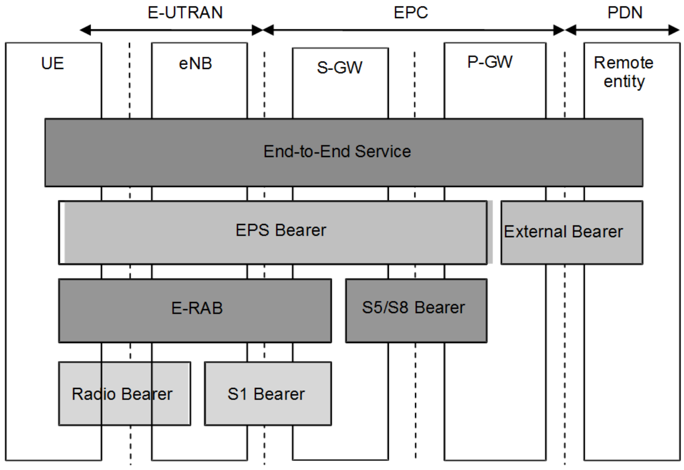


Figure 14: EPS Bearer Architecture

can be established, dynamically, as a result of service requests. Each bearer can be established with a different PDN. The user data are transported between the E-UTRAN, SGW and PGW by using IP tunnels for supporting the UE mobility.

QOS parameters describes the properties of each bearer, including bit rates, packet delay, packet loss, bit error rate, and scheduling policy in the radio base station.

Each bearer is associated with the following parameters:

- QOS Class Indicator (QCI)
- Allocation and Retention Priority (ARP)
- Guaranteed Bit Rate (GBR)
- Non-GBR Bearers (Non-GBR)
- Maximum Bit Rate (MBR
- Aggregate Maximum Bit Rate (AMBR)

QOS Class Indicator (QCI): The QCI defines the treatment of IP packets received on a specific bearer. Each functional node packet treats forwarding of traffic traversing a bearer. QCI values have impacts on several node-specific parameters, such as link layer configuration, scheduling weights, and queue management. It represents an index referring to a number of different sets of minimum QOS characteristics, such as priority, delay, etc. required by a service (cf [10] The figure 15 summarizes series of standardized QCI types which were defined by the 3GPP. At first, a majority of operators will likely start with three basic service classes: voice, control signaling, and best-effort data. In the future, dedicated bearers offering premium services such as high-quality conversational video can be introduced into the network.

Allocation and Retention Priority (ARP): The 3GPP standards provide mechanisms to perform admission control checks during the establishment of bearers and bearer path changes during handovers. Each bearer is associated with a priority indicator (1 to 15), a pre-emption capability flag indicating that the bearer can preempt lower priority bearer in case of lack of resources, and a pre-emption vulnerability flag indicated that the bearer could be or not preempted once established. The ARP is defined based on the subscription right of the user.

Bit rate : A GBR indicates the bandwidth reserved and thus guarantied for the bearer. A Maximum Bit Rate (MBR) is specified for real-time services only as the upper bound of the bit rate that the bearer may temporarily get. For example, rate adaptive codecs for voice or video services may benefit for having bandwidth at least what GBR defines, and occasionally consume more bandwidth up to MBR [9].

QCI	Resource Type	Priority	Packet Delay Budget (NOTE 1)	Packet Error Loss Rate (NOTE 2)	Example Services
1	GBR	2	100 ms	10^{-2}	Conversational Voice
2		4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3		3	50 ms	10^{-3}	Real Time Gaming
4		5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100 ms	10^{-6}	IMS Signalling
6		6	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100 ms	10^{-3}	Voice, Video (Live Streaming) Interactive Gaming
8		8	300 ms	10^{-6}	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		9			

Figure 15: Standardized QCI characteristics

Aggregate Maximum Bit Rate (AMBR): There are two types of AMBR, APN-AMBR and UE-AMBR. APN-AMBR is the maximum available bit rate available for Non-GBR bearers and for all PDN connections of the same APN. UE-AMBR is the maximum available bit rate available for Non-GBR bearers of a UE. , APN-AMBR is enforced in the UE for uplink traffic and in the PDN gateway for downlink traffic. UE-AMBR is Enforced in the eNodeB for both uplink and downlink traffic.

Service data flows The notion of Service Data Flows (SDFs) is defined as a particular type of user traffic to which a specific QOS is applied. The filtering of the SDF is performed at both entry point of the EPS bearer: In the UE and in the PGW. The filtering of the user plane packet is performed by using traffic flow templates (TFT). TFT's contain packet filtering information used to identify and map packets to specific bearers. The filters are configurable by the network operator ands is commonly referred to as a 5-tuples include : The source and destination IP address, the source and destination port number or ranges and the protocol identifier [13]. The figure 16, extracted from TS 23.401 [12], presents the filtering point in the UE for the uplink traffic and in the PDN GW for the downlink traffic. The filtering procedure will identify for each SDF the suitable bearer that will be used to transport the packet, in respect to the defined QOS.

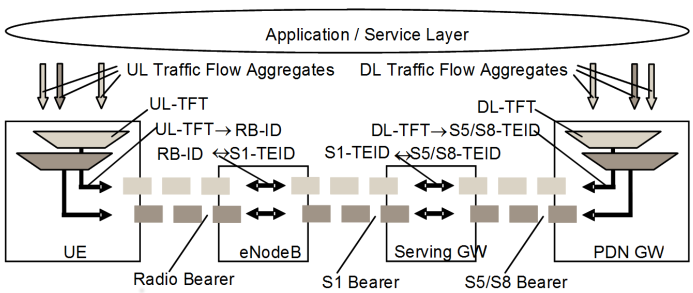


Figure 16: Traffic Flow Template filtering

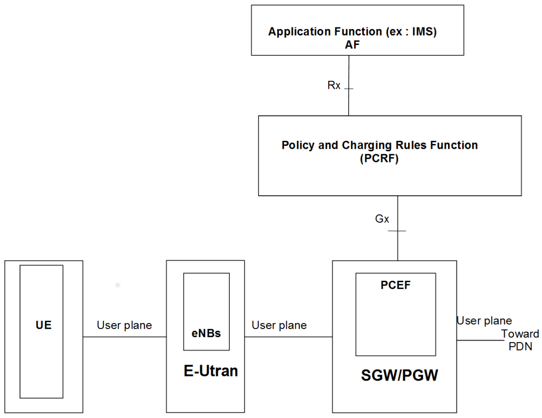


Figure 17: Concept for Policy and Charge

5.1.5 LTE Policy and Charging Control

Policy control and charging are critical in enabling the network operator to control and monetize these new capabilities. The term of policy concerns operator's policy which provides a wide range of options of how an operator may differentiates users and their data traffic. It concerns also service based-policy which determines how user data transmission occurs in the network. PCC architecture is presented at the figure 17 There are two types of relevant policy:

- QOS policy which describes how QOS parameters shall apply to packet flow.
- Filtering policy which defines whether a packet flow is allowed or not.

The PCC functions include:

- PCRF (policy and charging rules function) provides policy control and flow based charging control decisions.
- PCEF (policy and charging enforcement function) implemented in the serving PDN gateway. It enforces gating and QOS for individual IP flows as described in [Renvoi Service data flows]. It also provides usage measurement to support charging.

On the architecture diagram of the figure [add renvoie], it is important to note the role of AF, or application function. The AF is the network entities that will request the creation of dedicated bearers associated with a specific QOS.

5.1.6 IMS architecture description

The IP Multimedia Subsystem (IMS) is expected to offer the access to services requiring a defined Quality Of Service (QOS). IMS was specified by a joined work between ETSI and 3GPP. The user application interacts with the IMS subsystem by using SIP signalization.

The IMS architecture is presented in figure 18. When the IMS is used in combination with a 3GPP EPC core network, the SIP proxy server (X-CSCF), interfaces by the Rx interfaces with the PCRF (Note : The interfaces is not represented on this diagram). In order to create a bearer with a suitable QOS, an application running on a UE will communicate, using the SIP protocol, with the IMS CSCF. Following the end to end negotiation of QOS with the remote and to end parties, the P-CSCF interfaces with the PCRF to create specific QOS bearers suitable for its needs. At the time of LTE network, the IMS is expected firstly to be used for providing voice services (VOLTE : Voice Over LTE) and secondly for other services such as defined in the RCS [Reference]. In case of multiple network connection, the GMSA has defined a profile to ensure the end to end QOS of Voice and SMS services in [PRD IR.92].

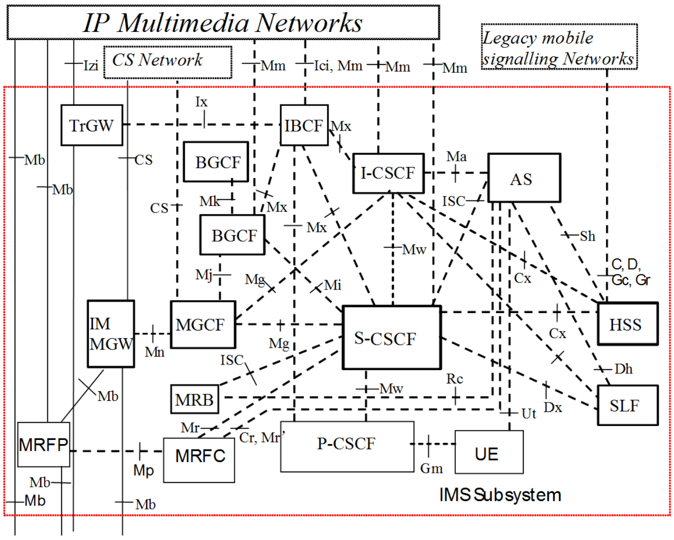


Figure 18: IMS Basic Architecture

5.1.7 Conclusion

3GPP has defined network architecture suitable to ensure an end to end QOS based on the creation of bearers. The defined architecture is expected to meet the end user expectation. As the architecture is defined from an end to end perspective, all the interconnected network are expected to meet the QOS requirements of the end users such as presented in the figure 19

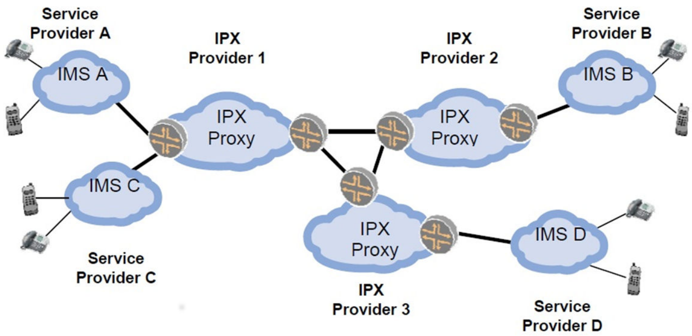


Figure 19: End to end QOS : Interconnection of network supporting QOS

In case the end user application is located on the internet where the traffic is only best effort, then only best effort traffic QOS is possible. In this case the network have limited standardized way to manage the user traffic, and perform traffic differentiation. The network is also more prone to reach congestion. Improving the management of user plane traffic, and in particular, the best effort traffic, remains an important goal for the network. To improve this situation, The 3GPP has perform in 2012 a study named FS-Upcon (Feasibility Study on User plane CONgestion management) [9] has been in 2012 with the goal of identifying use cases causing high level traffic in the RAN, and to propose requirements for handling user plane traffic. In most of the use cases, only best effort QOS is expected. From a research perspective, innovative solutions, possibly based on cross layer signaling, are expected to to improve the limitations of the current networks for best effort traffic.

5.2 Radio Resource Allocation and Scheduling in LTE Uplink

The preliminary definition of the scenario use cases [77] provides different sets of quality of service (QoS) requirements, in terms of guaranteed bit-rate, maximum end-to-end delay and packet loss. Among them, *Ambulance and emergency area*, *Emergency area with multiple casualties* and *Surgical Assistance* scenarios should exploit the uplink of existing and next-generation wireless systems.

While the 3G cellular technologies only offer QoS support for voice services, beyond-3G and 4G systems, such as the Long Term Evolution (LTE) [16] and its upgraded versions, are expected to provide QoS support for multimedia application according to different class quality indicators (CQI). In particuar, the Technical Specification (TS) 23.203 [11] provides different class-of-service priorities in terms of minimum guaranteed bit-rate (GBR), packet delay budget (PDB) and packet error loss rate (PELR). Figure 15 lists the requirements for each particular quality class identifier (QCI) and the corresponding priority. According to the proposed requirements for abovementioned use cases, mobile terminals should receive a priority value at least equal to 4, corresponding to packet delay and packet error rate lower than 150 ms and 10^{-3}, respectively, with a minimum guaranteed bit-rate. However, none of such requirements can be accomplished without efficient radio resource allocation (RRA) and multiuser scheduling at the MAC layer [1] for both the downlink and the uplink, and these matters are not defined by the LTE standards.

Assuming a cross-layer exchange of information, such as as channel state information (CSI) from the PHY layer and/or source content information from the upper layers, RRA strategies are often categorized in channel-aware and/or content/context-aware schedulers. A channel-aware scheduler takes advantage of the temporal, frequency and multi-user diversity inherently provided by the wireless channel. Each user in a cellular scenario experiences different channel gains on the available radio resources, according to path loss, shadowing and fast fading. Therefore the scheduler can exploit such information to allocate the radio resources to those users whose channel conditions are favorable. The content-aware schedulers take into account the specific content (or context) of the source data, such as, for example, rate-distortion relationship and play-out deadline of video sources, in order to provide differentiated services for heterogeneous applications. In this section we will mainly focus on Channel-aware schedulers.

While the research community has provided a large amount of proposals to solve the RRA problem for LTE/LTE-like downlink, the uplink RRA problem has gathered attention only starting from 2004 and a sensible lower number of contributions can be found.

In the following we will provide an overview of the last improvements in the RRA methodologies for the LTE uplink. We first describe the LTE physical layer techniques, to better understand benefits and constraints characterizing the LTE uplink. Then, we provide a review of the most interesting scheduler proposals with relation to the CONCERTO use cases.

5.2.1 LTE Uplink Overview

The LTE standard [16] was developed by the 3rd Generation Partnership Project (3GPP), starting from 2004, to address the increasing demand for better QoS and the growth of bandwidth consuming multimedia applications. Compared to its predecessors LTE provides several enhancements, as higher peak data-rate and spectral efficiency, flexible resource allocation, lower user and control plane latency and reduced devices complexity. In order to meet such requirements, orthogonal frequency division multiplexing (OFDM) has been chosen for the first time as the access technology for both downlink and uplink in cellular networks.

The OFDM modulation scheme divides the input data stream into several parallel sub-streams with reduced data rate, hence with increased symbol duration, and each sub-stream is transmitted on a separate orthogonal subcarrier. The higher symbol duration improves the robustness againts the channel delay spread due to multipath propagation. At the beginning of each OFDM data symbol a cyclic prefix (CP) is inserted as guard interval. Typically, the CP is constituted by a replica of the final part of the symbol itself. The CP insertion can completely eliminate the inter-symbol interference (ISI) as long as the CP duration is longer than the channel delay spread. OFDM modulation can be realized through efficient inverse fast Fourier transform (IFFT), which allows to manage a large number of subcarriers (up to 2048) with low complexity. In an OFDM system, transmission resources are available in both time and frequency domain, in terms of OFDM symbols subcarriers respectively. The time and frequency resources can be organized into subchannels for allocation to individual users leading to orthogonal

[1] In the remainder of this section we will use the terms RRA and multiuser scheduling interchangeably.

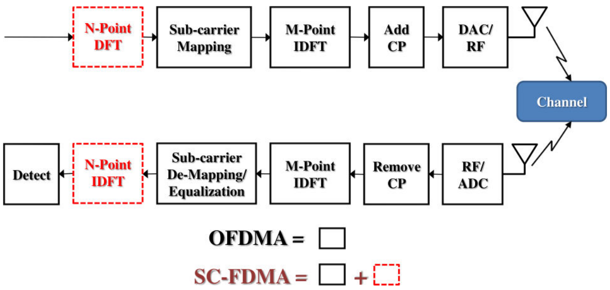


Figure 20: Transmitter and receiver structure of SC-FDMA and OFDMA modulation. SC-FDMA introduces a N-point DFT ($N > M$) precoding to spread the data power over the entire allocated bandwidth.

frequency division multiple access (OFDMA), also known as multi-user OFDM. Within LTE, downlink and uplink transmissions are organized into radio frames with 10 ms duration with two different radio frame structures according to the supported duplexing schemes, namely time division duplexing (TDD) and frequency division duplexing (FDD). Each 10 ms radio frame is divided into ten equally sized sub-frames. In its turn, each sub-frame consists of two equally sized slots. The entire bandwidth is divided into several subcarriers with subcarrier spacing of 15 kHz, allowing six or seven OFDM symbols for each subcarrier according to the two supported CP durations. In the uplink, three constellations are allowed on each subcarrier, *i.e.*, quadrature phase shift keying (QPSK), 16-quadrature amplitude modulation (QAM) and 64-QAM. In order to reduce the feedback information and the RRA complexity, adjacent subcarriers are grouped into physical resource blocks (PRB) of 12 subcarriers each, which represents the minimun amount of allocable resources.

An important drawback of the OFDMA schemes is that the transmitted time-domain waveform exhibits very pronounced envelope fluctuations, resulting in a high peak-to-average power ratio (PAPR). Signals with high PAPR require highly linear power amplifiers to avoid excessive inter-modulation distortion. To achieve this linearity, the amplifiers have to operate with a large back-off from their peak power, resulting in increasing costs and power consumption. Such problem is clearly more critical in the uplink transmission where the cost and power consumption of mobile must be kept as lower as possible. For this reasons, single carrier frequency division multiple access (SC-FDMA) [197], also known as discrete Fourier transform (DFT)-spread frequency division multiple access (FDMA), has been introduced for the LTE uplink. SC-FDMA offers advantages similar to OFDMA but provides a lower PAPR by introducing a DFT precoding process at the transmitter (see Fig. 20), aimed at spreading the power over the entire allocated bandwidth. Such an advantage is paid with an increased ISI at the receiver, so that some kind of adaptive frequency domain equalization is required. This trade-off is well balanced in 4G cellular systems, since higher costs due to complex signal processing at the base station (BS) are acceptable if they entail simple and low-consumption amplifiers in the mobile terminals.

There are two types of SC-FDMA: localized-FDMA (L-FDMA), in which the subchannels assigned to a user are adjacent to each other, and interleaved-FDMA (I-FDMA) in which users are assigned with subchannels distributed over the entire frequency band. Only L-FDMA is taken into account by LTE and this imposes an additional constraint (*i.e.*, contiguous subcarriers) with respect to traditional OFDMA resource allocation. A detailed overview of SC-FDMA schemes can be found in [232] and [197]. In the next subsection we will review the RRA problem formulation for SC-FDMA-based uplink systems.

5.2.2 Resource Allocation and Scheduling

RRA and multiuser scheduling in OFDM-based systems aim at assigning the available resources (i.e. power and subchannels), [2] at each time transmission interval (TTI) to the different users, according to a given trade-off among spectral efficiency, fairness, and general QoS constraints. The basic idea is to map the available resource and/or performance criteria into the corresponding utility or weight values and optimize the aggregate utility under specified PHY layer constraints.

In order to provide a comprehensive overview and to better understand the challenges inside the SC-FDMA

[2] In this review we mainly focus on scenarios where user equipments (UEs) and eNB are equipped with one antenna each. Further investigation can also consider the problem of allocating the spatial resources in case of multi-antenna configurations.

RRA problem, we will first briefly review the OFDMA RRA problem by introducing some mathematical symbolism.

We consider a scenario where K users indexed by the set $\mathcal{K} = \{1, \ldots, k, \ldots, K\}$ are served by one enhanced-NodeB (eNB). Assuming an OFDM modulation scheme, the total available bandwidth B is divided into S orthogonally subchannels, indexed by the set $\mathcal{S} = \{1, \ldots, s, \ldots, S\}$, with a subcarrier spacing ΔB shorter than the coherence bandwidth.

Let $U(\mathbf{R})$ be the system utility dependent on the achievable rate vector $\mathbf{R} = [R_1, \ldots, R_K]$. The rate R_k achievable by user k depends on the set $\mathcal{S}_k$ of allocated subchannels and on the transmission power $p_{k,s}$ assigned to each subchannel $s \in \mathcal{S}_k$. The instantaneous rate R_k is bounded by the well-known Shannon formula, *i.e.*,

$$R_k = \sum_{s \in \mathcal{S}_k} \Delta B log_2(1 + p_{k,s}\gamma_{k,s}) \tag{1}$$

where $\gamma_{k,s}$ is the channel gain experienced by user k on subchannel s on a particular TTI. More generally, rates can be also evaluated according to the ergodic approximation, considering a suitable large time window on which instantaneous rates are averaged. Ergodic rate approximation allows to exploit the temporal diversity of the wireless channels but lacks of stringent QoS delay support.

The utility function is usually selected as concave and monotonically increasing. Example are MAX-SNR scheduler ($U(\mathbf{R}) = \sum_{k \in \mathcal{K}} R_k$), proportional fairness scheduler ($U(\mathbf{R}) = \sum_{k \in \mathcal{K}} \log(R_k)$), or objective and subjective video-based utilities, such as the peak-to-signal-noise ratio (PSNR) or the mean opinion score (MOS) metrics [87]. Most of the researchers tend to opt for a weighted sum-rate (WSR) approach ($U(\mathbf{R}) = \mathbf{w}^T\mathbf{R}$), where the weights $\mathbf{w} = [w_1, \ldots, w_K]$ can be derived from several utility functions or constraints, according to the target trade-off between fairness and efficiency.

The first scenario to gather a major attention by the research community was the OFDMA downlink. In the last twelve years, a large amount of scientific papers has provided heuristic, sub-optimal and optimal solutions to the RRA problem according to a multitude of goals. In this case the problem can be stated as:

$$\max_{\mathcal{S}_1, \ldots, \mathcal{S}_K} U(\mathbf{R}) \tag{2a}$$

$$s.t. \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} p_{k,s} \leq \bar{P} \tag{2b}$$

$$R_k \geq \bar{R}_{0k}, \quad \forall k \in \mathcal{K} \tag{2c}$$

$$\mathcal{S}_k \cap \mathcal{S}_{k'} = \emptyset, \quad \forall k, k' \in \mathcal{K}: \quad k \neq k' \tag{2d}$$

where $\bar{P}$ is the power budget at the BS and $\bar{R}_{0k} > 0$ is the minimum bitrate that should be provided to GBR users ($\bar{R}_{0k} = 0$ for non-GBR users).

The optimal solution of such a problem, can be derived through dual decomposition by decoupling subchannel and power allocations. The starting point is to enable subcarrier sharing or, in other words, to relax the subchannel exclusivity constraints (2d), leading to a convex achievable rate region. Similar frameworks were proposed by Wong and Evans [296], and Wang and Giannakis [290], who proved that quasi-optimal solutions for both ergodic and instantaneous rate cases has linear complexity with respect to number of subchannels and user. In both cases, the optimal power is obtained through water-filling solutions and the users are scheduled according to a winner-takes-all policy for each subchannel.

A first step towards the RRA uplink is the introduction of the distributed power constraints, *i.e.*, the transmission power of each k-user should be less than a predefined power budget $\bar{P}_k$. The general OFDMA uplink RRA problem can be re-formulated as :

$$\max_{\mathcal{S}_1, \ldots, \mathcal{S}_K} U(\mathbf{R}) \tag{3a}$$

$$s.t. \sum_{s \in \mathcal{S}_k} p_{k,s} \leq \bar{P}_k \; \forall k \in \mathcal{K} \tag{3b}$$

$$R_k \geq \bar{R}_{0k}, \forall k \in \mathcal{K} \tag{3c}$$

$$\mathcal{S}_k \cap \mathcal{S}_{k'} = \emptyset, \quad \forall k, k' \in \mathcal{K}: \quad k \neq k' \tag{3d}$$

Yaacoub, Dawy and others have provided several solutions for the OFDMA uplink RRA problem, considering continuous (capacity-based) instantaneous rates [311], as well as continuous and discrete ergodic rates due to realistic adaptive modulations [312]. Their objective was to maximize the weighted sum of the achievable rates. The authors showed that also the optimal solution of the uplink case can be derived in the dual domain assuming subcarrier sharing, also resulting in a winner-takes-all policy and excusive subchannel allocation.

After the introduction of SC-FDMA for the LTE uplink in 2004, the RRA problem has gathered increasing interest. As mentioned above, SC-FDMA and, more specifically, L-FDMA only consider contiguous subchannel allocation, making the problem much harder to solve. As an example, assuming the LTE test case with $S = 24$ subchannels and $K = 10$ users, an exhaustive search would require to evaluate all the possible $5.26 \cdot 10^{12}$ subchannel allocations. In the following we summarize the constraints that each LTE scheduler should satisfy in the uplink:

- *Exclusivity*: each subchannel can be allocated to at most one user. This allows to eliminate intra-cell interference (non-exclusive methods, such as super-position coding, allowing more than one user in each subchannel are considered too complex).
- *Adjacency* : if more than one PRB is allocated to one user, they must be contiguous in the frequency domain (L-FDMA).
- *Total Distributed Power Constraints* : the total transmission power of each user should be less then a predefined power budget $\bar{P}_k$. Differently to the downlink scenario, the power is limited on a per-user basis.
- *Peak Power Constraints*: the peak power transmitted by each user on each subchannel should be less than some peak power level $\tilde{P}_{k,s}$. Such constraint becomes trivial when constant power-allocation is considered in the problem solution.

We can now turn to formalize a typical SC-FDMA uplink problem. Let us define $s_{k,min} = \min_s(\mathcal{S}_k)$ and $s_{k,max} = \max_s(\mathcal{S}_k)$ as the lowest and highest allocated subchannel, respectively. We can then define

$$\mathcal{A}_k = \{s_{k,min}, s_{k,min} + 1, \dots, s_{k,max} - 1, s_{k,max}\} \tag{4}$$

as the set of all allocable subcarriers which satisfy the contiguous subchannel constraint. The SC-FDMA RRA problem can be generally formulated as

$$\max_{\mathcal{S}_1,\dots,\mathcal{S}_K} U(\mathbf{R}) \tag{5a}$$

$$s.t. \sum_{s\in\mathcal{S}_k} p_{k,s} \leq \bar{P}_k \ \forall k \in \mathcal{K} \tag{5b}$$

$$p_{k,s} \leq \tilde{P}_{k,s}, \quad \forall k \in \mathcal{K}, s \in \mathcal{S} \tag{5c}$$

$$R_k \geq \bar{R}_{0k}, \forall k \in \mathcal{K} \tag{5d}$$

$$\mathcal{S}_k \cap \mathcal{S}_{k'} = \emptyset, \quad \forall k, k' \in \mathcal{K}: \quad k \neq k' \tag{5e}$$

$$\mathcal{A}_k \cap \Big(\bigcup_{j\in\mathcal{K}\setminus\{k\}} \mathcal{S}_j \Big) = \emptyset, \quad \forall k \in \mathcal{K} \tag{5f}$$

According to such problem formulation, the techniques traditionally used to solve OFDMA RRA problem (*e.g.*, dual decomposition) cannot be directly applied, primarily due to the subchannel adjacency restriction. For this reason, the first scheduler proposals were based on heuristic approaches, *e.g.*[168] [167] [196],[231][59] [34].

The first proposal published by Lim *et al.*[168] considers an equal-bit equal-power (EBEP) allocation for each subchannel in order to maximize a proportional fairness utility function. Instead of solving the optimization problem, they provide a subcarrier allocation scheme which improves the marginal utility using both L-FDMA and I-FDMA schemes. In both cases the heuristic scheduler aims at finding the subchannel which has the highest marginal utility for a given user, defined as the difference between the utility obtained with and without the allocation of the s-th subchannel to the user. The heuristic algorithm ensures contiguous or interlevead subchannel allocation, according to L-FDMA and I-FDMA schemes, respectively. Specifically, when a set of subchannels to allocated to a user, the algorithm removes such set from the available subchannels, without checking if other users can exhibit a larger marginal utility gain. The complexity is linear in both the number of users and subchannels.

The authors show that a significant gain, in terms of aggregate throughput, can be achievable adopting a localized scheme compared to interleaved one. They also extended their work to take into account the outdated CSI as in pratical high-mobility scenario [196].

In [231], the authors provide three different solutions, namely the first maximum expansion (FME), the recursive maximum expansion (RME) and the minimum area difference (MAD), based on a proportional fairness metric. FME and RME algorithms are extensions to the solution proposed in [168]. Both algorithms assign the subchannels starting from that with the highest metric of the $S \times K$ matrix representing the channel gain. Then, they expand the allocation for a selected user on both the right and the left sides of the subchannel with the largest metric. However, the subchannels are not dropped at each allocation as in [168]. In fact, the algorithms verify if another user can obtain a larger marginal utility improvement. Finally, the MAD algorithm aims at providing the minimum difference between the cumulative utility of different users and the maximum utility value for any given subchannel.

A comparative study of the performance of these scheduler proposals was provided in [88] with respect to the search-tree based packet scheduling (STBPS) proposed in [59]. They also proposed a modified version of the FME algorithm. The performance comparison is carried out according to throughput, spectral efficiency and fairness for different number of users in the systems. All the schemes are able to exploit multi-user diversity by increasing throughput and spectral efficiency as the number of users increases. However, as expected, all algorithms lacks for QoS support due to the absence of GBR constraints in the problem statements. The STBPS algorithm exhibits better performance in terms of data-rate fairness with respect to all the considered schemes but is penalized by a lower cumulative throughput. On the other hand, MAD provides the highest efficiency, in terms of both throughput and spectral efficiency, and exhibits a fairer behavior with respect to FME and RME.

An extended comparative study was provided by the same authors in [248], where mixed traffic is used as prescribed by 3GPP for pratical performance evaluation. In their analysis an heuristic localized gradient algorithm (HLGA) as proposed in [34] was also included. HLGA is designed to include hybrid-ARQ (H-ARQ) schemes as specified by LTE, where a subset of PRBs are reserved for H-ARQ processing of previous unsuccessful transmissions. Performance evaluations were carried out in terms of per-user throughput, packet loss and fairness. However, none of the abovementioned framework is able to provide optimal solutions to understand the overall benefits of each proposal in terms of performance and complexity.

In 2009 Wong *et al.*[297] provided a novel reformulation of RRA problem in (5) without GBR contraints, considering the weighted sum rate as utility. The problem is translated into a pure binary integer program (BIP), namely a set partitioning problem. It allows to compute the optimal allocation without resorting to exhaustive enumeration. Moreover, the authors propose a greedy algorithm to reduce the computational complexity in solving the set partitioning problem. The proposed solution is based on the marginal utilities maximization of the abovementioned scheduler and exhibits the worst-case linear complexity when $S \gg K$. However, an undesirable gap is still present with respect to the optimal solution in terms of the cumulative distribution function (CDF) of the spectral efficiency.

Ahmad and Assaad [31] take advantage of such reformulation to transform the BIP problem into a canonical dual problem in the continuous space. It is proved analytically that under certain conditions, the solution of the canonical dual problem is identical to the solution of the primal problem. The performance evaluation in terms of the WSR CDF shows a significant reduction of the gap with respect to the optimal solution compared to the greedy algorithm proposed in [297]. They also investigated the problem of minimizing the sum-power while satisfying the target data rate constraint for all the users [32], providing a similar solution.

5.2.3 Conclusion and Discussion

The SC-FDMA scheme selected for the uplink transmission in LTE has set new challenges for the scheduling functionality of the MAC layer. The localized (contiguous) allocation of the subchannels, required to ensure low PAPR on the trasmitter side, introduces novel contraints leading to a problem that can not be solved through the classical methods applied to OFDMA uplink and downlink scenarios.

The first heuristic proposals to solve such NP-hard problems generally maximize a marginal utility according to Max-SNR or proportional fair rules. Such approaches lacks of QoS support in terms of GBR and maximum tolerable delay, which are the first requirements for the CONCERTO use cases. However, such rules only identifies the metric to maximize and, as a consequence, can be easily translated into more sophisticated utilities. As an example, in [151] and [75], the utility is selected according to a weighted proportional fair paradigm, where the weights are evaluated according to the delay accumulated by the head-of-line packet and the overall delay constraints according to a modified largest weighted delay first (MLWDF) algorithm. All these heuristic approaches can be seen as interesting trade-offs between simplicity of implementation, computational complexity and limited performance.

So far, the best proposal, in terms of weighted sum-rate performance, seems to be the canonical dual approach proposed in [31] and [32], which has been proved to be optimal under certain conditions. The combination of the RRA approach of [31] to schedule the physical layer resources with the packet scheduling and weights selection of [151] or [75] could be a good starting point to guarantee efficient resource utilization and QoS support for the CONCERTO use cases. Finally, we remark that, since the CONCERTO project specifically addresses multi-source and holographic tele-medicine (MSHTM) transmissions over wireless networks, the integration of source content-awareness in the RRA algorithms is a key issue that is still open in the context of 4G uplink.

5.3 LTE Relay

5.3.1 Relaying techniques for LTE systems

This section provides a description and some comments of the state-of-the-art of relay-aided transmission in LTE advanced (LTE-A) wireless networks.

According to the preliminary definitions of the use cases and requirements in [77], 4G cellular technology (supporting both uplink and downlink video transmission) is an important wireless component of the CONCERTO scenarios and relay-aided transmission is one of the main transmission techniques currently under investigation for LTE-A and WiMAX mobile systems [313] to extend coverage and to increase throughput for cell-edge terminals. The latter issue is particularly relevant in CONCERTO scenarios where high quality video transmission has to be supported, with data-rates higher than 1 Mbit/s per video even in situation with scarce coverage.

The principle behind relaying is very simple. By taking advantage of the broadcast nature of wireless communications, some nodes in the wireless network are able to retransmit the received signals to help the transmission among other nodes. In other words, these wireless nodes, which can be either user terminals or access network components, act as relays. This concept is often indicated as cooperative communication and can be used in improving network connectivity, link capacity and reliability, as well as enhancing power and spectrum efficiency.

The idea of cooperative communication can be traced back to the ’70s when Van der Meulen in [81] analyzed a three-terminal relay channel, and derived upper and lower bounds on its channel capacity. Later, in [67] capacity of the cooperative relay channel was investigated with more detail by considering two cooperation protocols, namely decode-and-forward (DF) and compress-and-forward (CF), and by defining the theoretical basis for the subsequent research work. In the early 2000s, the interest of academy and industry in cooperative communication techniques started to grow intensively. Various cooperation schemes have been designed for enhancing the performance of wireless communication networks [21, 136]. Recently, relay-aided communication has been adopted in LTE (Release 10) as a key technology for future generation commercial wireless communication systems [18, 15].

In general, in 4G wireless networks relay nodes are deployed as network components that help other transmissions but do not have their own data to send. Inter-user cooperation has not been included in current wireless standards due to the critical issues related to the consumption of terminal resources for benefit of other users. Compared to the traditional eNB, the relay nodes transmit at lower power and cover smaller areas. User terminals can either communicate directly with the eNBs or via the relay stations. Moreover, since the user data is available at both the transmitter and the relay, joint transmission and processing is also allowed to further improve the performance. This last option is under investigation for LTE Release 11 and beyond [163]. It is important to remark that the introduction of relaying techniques in the new releases of LTE has to preserve the compatibility with the previous releases. Therefore, from a user terminal perspective, the relay nodes must be completely transparent or appear as regular eNBs [123].

5.3.2 eNB Relaying techniques

Two main categories of relays are considered: amplify-and-forward (AF) relays and DF relays. As their name suggest, AF relays simply amplify and forward the analog received signal. They are transparent to both user terminal and eNBs, but, as analog repeaters, they also amplify noise and interference. On the contrary, DF relays detect and decode the received signal and regenerate it before forwarding. They introduce a small delay at the physical layer but they are more robust in low-SNR environment. Depending on the network topology and the quality of the backhaul link between the source and the relay, one relay category may outperform the other in terms of system capacity or diversity. Generally speaking, for systems with good backhaul links, DF based schemes are more favorable, while for systems with relatively poor backhaul links, AF based schemes are more advantageous.

Two families of DF relays were taken into account to become part of LTE-A [18]. Type-I relays are non-transparent to the user terminals, have their own physical "identity", and manage all the necessary physical channels to appear as a regular eNB to the user terminals. They perform IP packet forwarding at layer 3. Type-II relays do not have their own physical identity and are transparent to the user terminals, i.e., the terminal is unaware of their existence and identity. The user terminal receive control signals from the eNB, whereas the relay does not transmit common reference signal and other control information. Type-II relays simply exploit early decoding of transmitted signal for participating in the retransmissions. An example is when the relay is activated only for H-ARQ retransmission when a NACK message is received. Since the relay transmitter causes interference to its own receiver, simultaneous operation of backhaul and access links is not feasible. Two practical methods for decoupling outgoing and incoming signals are considered in the 3GPP standard:

- Frequency multiplexing (outband relaying): no air interface changes are required, but an additional LTE carrier is needed.
- Time multiplexing (inband relaying): duplexing of transmission/reception leading to half-duplex operations on the backhaul and access links.

Currently the 3GPP specifications support Type-I relays with both inband and outband relaying. From a communication point of view, coverage extension and throughput enhancement are obtained through a two-hop communication. On the other hand, the technical discussion is still open for possible specifications of Type-II relays and cooperation techniques for relay-aided communications involving joint transmission and/or joint processing of received signals. In the latter case, the deployment of cooperation techniques leads to a two-phase relay-aided communication, where the destination performs joint processing of source and relay signals and cooperative diversity gain is also achieved. More details on architectures and protocols for relay operations in LTE release 10 are discussed in [123], whereas possible schemes to support cooperation with both Type-I and Type-II relays are discussed in [163].

Although in the last decade a lot of papers have investigated theoretical issues, performance analysis, cooperation strategies, coding schemes, and parameter optimization for relay-aided communication schemes, only from 2008 the application to LTE or International Mobile Telecommunications - Advance (IMT-A) system has been specifically addressed. The earliest works for LTE [129][161][43] mainly investigated coverage improvement and business impact considering realistic environment. Performance studies are also found in [66, 198] with reference to deployment issues and comparison between Type-I DF relays and AF relays. An interesting work is [251], in which it is shown that, with a proper design, the delay due to relaying is almost negligible when addressing the end-to-end delivery latency.

More recently, the scientific community is working around several open issues related to the use of relay technology in LTE wireless systems and the design of system parameters and optimization algorithms. They can be outlined as follows: relay selection strategies and multi-relay operations [42, 199, 47], resource allocation and scheduling algorithms [164, 182, 314, 78, 230, 189], interference coordination/mitigation methods [51, 315, 316, 227], energy efficiency issues [249, 95, 96] and network coding-based cooperation techniques [60, 207]. Part of these issues are not addressed in the LTE specifications and are left to manufacturers. Others may have a significant impact on the development of the next releases.

Concerning RRA and scheduling, [164, 182, 314] consider the downlink of Type-I relay-aided systems focusing the analysis on how to partition the resources betwen relay aided and non aided user terminals. Proportional fair scheduler is mainly considered as basic scheme at both the eNB and relay. [78] proposes an optimal asymmetric resource allocation algorithm for DF Type-I relay systems, where both the time slot duration and the number of subcarriers for the two hops may be different. In partcular, the authors show that it outperforms symmetric resource allocation. The uplink of relay-aided LTE systems is investigated in [230] and in [189]. The former work formulates a constrained optimization problem and describes a solution which divides the utility maximization of all nodes, including relay nodes, into sub-problems of utility optimization at each individual user terminal. The latter work proposes a buffer-based channel dependent scheduler whose objectives are to minimize packet loss due to buffer overflow and to increase resource utilization efficiency. In spite of some solutions focusing on specific aspect, this research field is still open.

The design of interference coordination/mitigation methods is in general a very important issue. In fact, to fully exploit the benefits of relaying, the inter-cell interference which is increased due to the presence of other relay nodes should be limited and controlled. A classical way for handling this issue is to allocate or control the power of the different network components. In [51] different power control optimization strategies have been proposed for 3GPP urban and suburban scenarios and it is shown that the power control is a crucial means to increase the cell-edge and system capacities, and to mitigate inter-cell interference. The works in [315, 316, 227] investigate some interference coordination techniques. The first one evaluates three inter-cell interference mitigation methods for LTE-A relay systems based on a proper coordination of the backhaul subframe allocation among macro and relay cells. The second one proposes an effective combination of relay networks and soft frequency reuse (SFR), with a dedicated relay topology (here, the relay is shared among adjacient cells) and a SFR-based resource allocation scheme. The third work investigates an interference coordination scheme based on prioritized scheduling applied on top of the relay-based cell expansion.

5.3.3 User Equipement Relaying techniques

Although LTE-A does not support user terminals (UE) acting as relays, there is an ongoing industry interest for such techniques. Research on device-to-device communication over LTE and LTE-A has been performed in the recent years by EU founded projects [295] and [202]. QUALCOMM has developed the FlashLinqtm technologies aiming to use the direct device-to-device communication for realising proximity-based services [175]. In June 2011, QUALCOMM proposed, to the System Architecture group of 3GPP (SA), to start a new study on proximity-based services using direct communication between devices. The study item was not straightly accepted, but at the following meeting in September, the study item [247] was amended and supported by several major telecom operators (Telecom Italia, NTT Docomo, ATT, etc ...) and manufacturers (NEC, Motorola). The main use cases targeted have been classified in 3 categories: Commercial/social use, Network offloading and Public Safety. Since November 2011, the interest of D2D has increased in particular on Public Safety uses case, due to the choice of LTE technologies by the USA for building a nation-wide public safety network. In addition, it is expected that a full definition of the system architecture and protocols for this feature will be defined in the 2013-2014 timeframe. The idea of UE-Relay appeared at the end of 1990's in the 3GPP TR 25.924 [14]. The direct communication of UE's was studied in the European project Winner+. The project evaluated the control of the D2D link under network coverage. Evaluation of the potential gain on the cell capacity was also performed. UE-Relay can provide major benefits for public safety. It can first limit the power consumption of a UE. The UE can communicate with a close UE-Relay to reach the infrastructure, thus reducing the required energy. It can also extend the infrastructure coverage. A specific use case for public safety has been defined in TR 22.803 [8] for the 3GPP Proximity services.

5.3.4 Conclusion

In the perspective of CONCERTO scenarios, which take into account LTE-A wireless infrastructure, the most significant open issues to address are the design of resource allocation strategies and the analysis of interference coordination/mitigation methods and algorithms. The resource allocation problem for relay-aided communication can be connected to the other challenges of CONCERTO, that is cross-layer optimization and support of high quality video transmission. Also the study of network coding based cooperation techniques appears at early stage and may be included in the CONCERTO framework to address uplink relay-aided communication from multiple video sources. Moreover, by keeping in mind that 3GPP has recently identified public safety use cases including UE-relay for extending the network coverage, it is also interesting to consider for the CONCERTO framework the investigation of relay components which can be temporarily deployed in emergency scenarios where a local coverage and throughput enhancement is needed to support multiple on-site video terminals.

5.4 Cross-layer Optimization for Video Streams

Some of the CONCERTO scenarios can utilize the rate adaptation capability of video streaming to adjust the video bitrate to network capacity for optimal Quality of Experience (QoE) and network resource usage. This can greatly expand the video service availability especially in wireless networks and mobile context where the used network connection cannot always support the QoS requirements of the whole video stream. In some situations, the ability to communicate with even the lowest quality video may be critical for some of the medical services considered in CONCERTO. Moreover, the rate and congestion adaptation capability of the medical video streaming services is an important tool to avoid causing a congestion collapse in resource limited networks with potentially large amount of concurrent users. Thus, the CONCERTO video streaming system should ensure seamless and ubiquitous access to media through heterogeneous networks and terminals by implementing dynamic scalability across the whole delivery chain. In this section, we are mainly considering real-time video streaming over UDP, that is, lacking inherent support for congestion control and reliability.

5.4.1 Overview of Video Adaptation Approaches

For video streaming, the inability of wireless networks to guarantee the required bandwidth and QoS for the services has boosted the development of novel video coding and adaptation solutions to improve the robustness and QoS for video transmission. For instance, the Scalable Video Coding (SVC) technology [252] provides both bitrate and device capability adaptation, which are especially useful in heterogeneous network environments. In addition, several algorithms and protocols for controlling the video stream bitrate of UDP-based streams to match the available network capacity have been proposed in the literature. The typical solution of adapting video streams in the application layer has been studied, for instance, in [274, 147]. In this case, video bitstream adaptation takes place in the server or an intermediate network node and the decision-making relies on client feedback information of the streaming performance (e.g., delay and loss metrics). However, due to this very feedback signaling requirement, application layer adaptation is not responsive enough to quick wireless link capacity fluctuations.

To overcome the deficiencies of application layer adaptation solutions in wireless networks, several proposals for prioritizing and adapting video streams in the data link layer have appeared during the recent years. Video traffic can be prioritized over other less critical traffic with different MAC layer QoS solutions supported by most wireless standards (e.g., IEEE 802.11 WLAN, IEEE 802.16 WiMAX, 3GPP). The standard QoS architectures, however, provide QoS support in per-flow basis, that is, all video packets receive equal QoS treatment from the MAC. This is insufficient for video streams that would benefit from per-packet QoS differentiation in order to ensure that the most important video packets (e.g., SVC base layer ones) get transmitted also under limited transmission capacity. This problem has been addressed in the literature by several proposals for adapting video streams in the MAC layer, including [211, 165, 144, 48, 93]. In the proposals, video adaptation in the MAC layer employs selective packet discarding and prioritized transmission to ensure that the most important video packets are transmitted over the wireless link with the highest probability. Such capability typically assumes cross-layer signaling between the application and MAC layers to communicate the video packet priorities. However, the proposed solutions do not take into account fairness towards other traffic types. One solution for implementing QoS differentiation within the video traffic category without hampering the QoS for other traffic types in WLAN is provided in the recently finalized IEEE 802.11aa standard [127].

Nevertheless, all the MAC-level solutions regulate video streaming only in the scope of the wireless link, thus potentially wasting transmission resources in the wired core network. Thus, it can be acknowledged that local adaptation within a single system layer is not the most efficient way to achieve dynamic scalability. Therefore, we propose to employ cross-layer video adaptation in CONCERTO.

In a cross-layer scheme, both application layer and MAC layer adaptations are implemented, as proposed in [267]. The application layer adaptation is performed by exploiting information on bandwidth availability from lower layers (*e.g.,* using [274]), whereas MAC layer adaptation is done by using side information on video packets, such as scalable video coding (SVC) layer information, priority and achievable rate-distortion improvement.

Regarding the adaptation of multiple video streams, cross-layer optimization has been investigated, within the framework of SVC, in several contributions, *e.g.,* [185]. Most of the proposed optimization schemes aim at minimizing the aggregate distortion of the multiple video programs.

However, different video sequences have different complexities, hence the relationship between rate and quality differs from one video sequence to another. Assuming taht the physical resources are shared among different video sequences, strategies aimed at minimizing the aggregate distortion of each video could lead to a high level of distortion for the most complex videos, typically requiring higher rates. Nevertheless, the end-user expectation is to receive the best feasible quality independently of the particular video complexity. In order to optimize the transmission strategy based on the end user video quality, the rate should be allocated among the videos based on a fairness criterion [64]. Moreover, the trade-off between the goal of reducing the bit rate and the goal of keeping the distortion at acceptable levels can be afforded dynamically, in order to perform adaptation to different conditions.

Some contributions exist in the literature that consider fairness-oriented rate adaptation, but they exploit the fine granularity scalability (FGS) tool, e.g., [292] [318]. However, FGS mode has been removed from SVC, due to its complexity. A multi-stream rate-adaptation scheme aimed at minimizing the distortion at the end-user while preserving fairness is proposed in [64]. It is based on the knowledge of the total bandwidth to be shared by the multiple videos and on an accurate rate-distortion (R-D) model of medium grain scalable medium grain scalability (MGS) encoded video.

R-D models enable to predict the minimum bit rate required to achieve a target quality. The rate of a video sequence is expressed in bytes/s, while the distortion is defined in terms of mean square error (MSE). The PSNR is more often used to express the quality of a video sequence. The time required to model the R-D curve for a given sequence may drive the decision on the methodology/algorithm to be adopted for the R-D modeling. On the other hand, the performance of the streaming system is directly affected by the accuracy of the R-D model [74]. R-D models are often categorized in analytical, semi analytical and empirical models. Analytical R-D models are used to predict rate and distortion of video sequences prior to the encoding process but they often incur in a loss of accuracy. Empirical models require the computation of all R-D points set resulting in a high complexity. Semi-analytical models aim at reducing such complexity by deriving parametrized functions that follow the shape of analytically derived functions, but are evaluated through curve fitting from a subset of the rate- distortion empirical data points.

Many R-D models have been proposed in the literature for real time and non-real time video streaming, see for example, [74][186] and references therein. In [150] the authors present a detailed analysis of the R-D relationship in FGS coders and provide an accurate semi-analytical square root R-D model, which requires at least two empirical points. Enhanced analytical R-D models for H.264/AVC were proposed for coded video sequences in [186]. However, the parameter extraction is performed after transformation and quantization in the encoding process and the numerical results show that they exhibit a low accuracy.

In [120] a technique to further reduce the complexity of semi-analytical models by keeping similar performance on the accuracy has been investigated. The proposed framework introduces new functions dependent only on the uncoded video streams. The coefficients of these new functions can be estimated off-line through a prior knowledge of the parameters of a set of sample video sequences, and then used for any future video sequence. The R-D model proposed in [120] only uses two parameters which are calculated taking into account the characteristics of the video sequences through a spatial and a temporal index extracted from the original raw video streams. This model has been finally applied as a component of a fairness-oriented multi-stream rate adaptation scheme.

5.4.2 Cross-layer Video Adaptation Solution for IEEE 802.11 WLANs

In this section, we briefly introduce one approach for implementing cross-layer video adaptation in CONCERTO. The solution has been proposed in [267], and an interested reader is referred to the article for more details.

The proposal defines an architecture and implementation approach for cross-layer adaptive video streaming required for ubiquitous video stream delivery in the future Internet, and may thus be used in CONCERTO for implementing robust healthcare services for scenarios permitting adaptive video services. The solution relies on the SVC technology for implementing wireless bandwidth-adaptive video streaming without adding any extra redundancy to the streaming. It is also based on an end-to-end architecture for scalable video transmission enhanced with cross-layer signaling and adaptation capabilities. The SVC-encoded video bitstream adaptation is handled jointly in the application (i.e. source) and MAC layers.

In the following, we describe in more detail the implementation of the video adaptation on the different protocol layers. We also give an overview of preliminary simulation results to attest the usefulness of the approach.

Application Layer Adaptation The application layer adaptation is implemented using a TCP Friendly Rate Control (TFRC) based adaptation algorithm [274], and the corresponding feedback information delivery may be realized using a cross-layer signaling framework, such as the one defined in the OPTIMIX architecture [160]. The TFRC based adaptation performs well in terms of TCP friendliness and smoothness and it is well suited for multimedia applications. The available network bandwidth is calculated in TFRC using the following Equation 6:

$$T = \frac{s}{R\sqrt{\frac{2p}{3}} + t_{RT0}(3\sqrt{\frac{3p}{8}})p(1+32p^2)} \tag{6}$$

where T is the proposed data rate for the multimedia stream, s is the packet size, R is round-trip time, p is the packet loss rate and t_{RT0} is TCP retransmission time. t_{RT0} can be estimated by setting it to $4R$ and the packet size *s* can be considered fixed (e.g., 1450 bytes) since fixed size segments are expected to be used [147]. Round-trip time *R* is also considered as an estimate calculated from the relation to packet loss rate value.

The packet loss rate *p* is calculated in the client by monitoring the number of the received packets. The client calculates the packet loss rate periodically (e.g., every 0.5 s) neglecting zero values and sends it to the video streaming server via the selected end-to-end signaling protocol (e.g., trigger delivery [160]). In addition, the client applies a weight to the calculated current and earlier packet loss rate values using Equation 7:

$$\sum_{k=0}^{4} E_k \times M_k, M_k \in \{0.45, 0.3, 0.1, 0.1, 0.05\} \tag{7}$$

where *E* denotes the erroneously received packets in the client. *M* is a coefficient to accentuate the most recent values. As indicated, for example, in [190], the use of weighted coefficients for previous values increases the accuracy and reliability of predictive events. Furthermore, in order to remove large fluctuations from the estimated data rate provided by TFRC, an average of four last data rate estimates is calculated and used as the target bit rate. The available bandwidth estimates calculated using the TFRC algorithm can vary quite a lot since TFRC reacts easily to even small levels of packet loss. The target bitrate calculated by TFRC is utilized in the bitstream adaptation process on a regular time interval (e.g., once in a second).

The application layer bitstream adaptation process for SVC-encoded streams utilizes the scalability features of SVC streams in order to adapt the video stream according to the calculated target bitrate from TFRC. Firstly, the SVC priority information contained in the NALU header is a key tool for SVC bitstream adaptation. Any part of an SVC bitstream can be simply extracted by removing NALUs based on these parameters. The priority information is contained in specific fields of the SVC NALU header, namely the *dependency_id*, *temporal_id*, *quality_id*, and *priority_id*, indicating the spatial, temporal, quality, and priority layers the SVC NALU belongs to, respectively. Secondly, the Scalable Supplemental Enhancement Information (SSEI) NALUs can be used for defining the bit rates and other characteristics of the SVC stream and its sub-streams. The SSEI information can be send in-band as well as out-of-band.

MAC Layer Adaptation The paper [267] primarily considers IEEE 802.11 WLAN networks. Therefore, the proposed MAC layer adaptation solution employs a MAC layer QoS architecture based on the IEEE 802.11e Enhanced Distributed Channel Access (EDCA) [128] and IEEE 802.11aa MAC Enhancements for Robust Audio Video Streaming [127] standards for MAC-level video packet differentiation. EDCA can be used implementing distributed QoS management for WLANs and the .11aa amendment enables efficient video traffic differentiation *within* the video access category (AC_VI) of EDCA without hampering the QoS allocated to other traffic types. In the solution, the EDCA queuing and medium access control mechanisms are extended by adding extra video queues and a video scheduler to achieve differentiated treatment for SVC video packets. Figure 21 illustrates the MAC layer QoS architecture built on top of EDCA. The architecture supports three queues in AC_VI, each having a different priority: high, medium and low, and a video scheduler, which selects video packets from the queues accordingly for transmission.

The prioritized video packet handling is accomplished by placing the incoming video packets into the video queues based on their type (e.g., SVC base layer or enhancement layer) with a packet classifier, and triggering the video scheduler every time the video category is granted access to the medium by EDCA. The number of additional video queues used depends on the level of video packet differentiation required in the system. When invoked, the video scheduler selects the next video packet for transmission considering factors like the type of each head of line (HOL) video packet, the time the packet has spent in the queue, and whether there are higher priority video packets currently stored in the queues. The packet classifier needs to access additional information to detect the types of the incoming higher layer packets. For this, the DiffServ Assured Forwarding (AF) per-hop behaviours [121] can be utilized as discussed in [265]. Besides supporting prioritized transmission of packets within AC_VI, the MAC QoS solution implements active queue management. Video traffic typically has strict QoS requirements in terms of end to end delay as frames arriving past their playout deadline are useless to the receiver. Thus, the queue management discards video packets from the queues, if a specific maximum queueing time (e.g., 500 ms) is passed. Finally, during congestion, the system implements priority-based discarding of video packets. Finite buffers are assumed, meaning that under congestion incoming packets that do not fit into their corresponding queue will be dropped. With priority-based discarding of video packets, the MAC always drops the packets belonging to the least important SVC layer first to fit the more important layers into the queues; thus, ensuring that the most

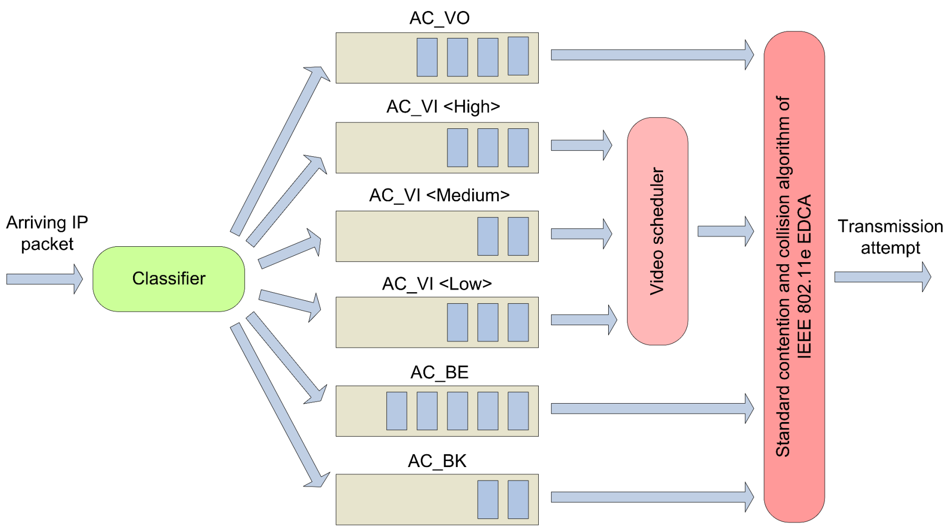


Figure 21: MAC-layer QoS architecture for adaptive SVC transmission.

important base layer packets have the highest probability to get transmitted. In case two or more enhancement layers are placed into the same queue, the lower importance packets need to be marked as drop eligible, to allow priority-based discarding.

In addition, to save processing in the WLAN AP, the MAC layer video adaptation solution may be dynamically enabled/disabled whenever the link capacity starts deteriorating and recovers. For this, the AP collaborates with the MS to constantly monitor the condition of the wireless link between them through a MIH-based approach originally presented in [210]. In the paper [267], the parameters monitored in the AP side, namely the rate of video packets dropped due to buffer overflow or exceeding of retransmission limit and the video queue size, are AC_VI specific. This is to ensure that the MAC-level prioritized transmission and adaptation of SVC are triggered on based on the link conditions experienced by the video traffic.

Experimental Evaluation To attest the advantages of the described cross-layer adaptation approach, we include here an overview of the results obtained in a simulation study and published in the paper [267]. More details of the experimental evaluation and results can be found from [267].

The simulations were conducted in the OMNeT++ environment [1]. The simulation model used included a video server, an IEEE 802.11g AP as well as a wireless MS, receiving a video stream from the server through the AP, and an IPv6 wired network connecting the BS and the server. During the simulations, the wired network did not introduce any packet loss or congestion. The video server and the MS used an RTP/UDP/IPv6 protocol stack for the SVC encoded video stream. The wireless MAC in the AP implemented the QoS architecture introduced in this section. The channel access for the four EDCA ACs was controlled using the default EDCA function (EDCAF) parameters [128]. The wireless physical layer simulated a log-normal shadowed uncorrelated Rayleigh fading channel with additive white Gaussian noise (AWGN) and without path-loss. The physical layer did not use adaptive coding and modulation, but the modulation is kept fixed throughout the simulation run. More details on the used simulation setup and parameters can be found from [267].

The evaluation was done for four different cases: no adaptation, MAC layer adaptation, application layer adaptation, and combined MAC and application layer adaptation for comparison. Figures 22-24 present example simulation results obtained for the four test cases. The results originate from a scenario, including the sending of a three-layer quality scaled SVC bitstream (Foreman) with average layer bitrates of 530 Kbps (base layer), 1 Mbps (base layer and first enhancement), and 2.3 Mbps (base layer and two enhancements) from the server to the MS while progressively increasing and decreasing interfering UDP traffic with a peak rate of 2.3 Mbps across the same resource limited WLAN channel between the AP and MS.

The results show how the application- and MAC-level adaptation complement one another in optimizing the

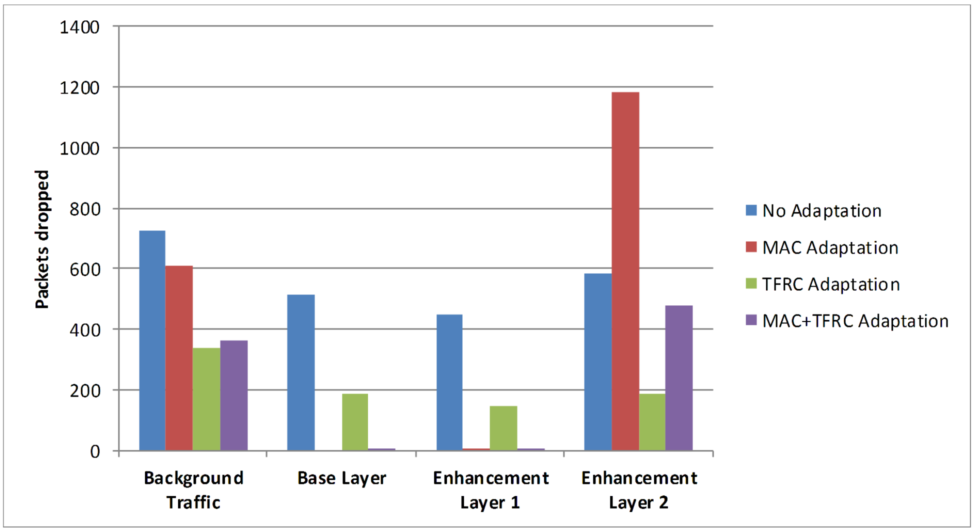


Figure 22: Total number of packets dropped in the WLAN AP.

QoE for the video user and saving network as well as terminal resources by reducing the number of useless transmissions under congestion. The scheme provides the best performance in selective discarding of SVC packets in AP MAC while minimizing the unnecessary transmissions in the wired network as shown in Figure 22. On the other hand, Figures 23 and 24 illustrate how the performance QoE-wise is almost as good as the best performing sole MAC layer adaptation.

5.4.3 Conclusions and Next Steps

Cross-layer video adaptation can be used for obtaining performance enhancements for video streaming in resource limited wireless networks in terms of both QoE and network resource usage. This section provided a state-of-the-art analysis on available cross-layer video adaptation solutions. Additionally, we presented an example of realizing cross-layer adaptive video streaming in IEEE 802.11 WLAN using SVC and cross-layer signaling. The cross-layer optimization results obtained for video streaming with SVC and WLAN technologies are very promising in terms of implementing adaptive video streaming services for the purposes of the CONCERTO project. Thus, the proposed approach will be extended to cover also other access network and video coding technologies during the project. The results of this continuation work will be presented in the future deliverables of WP5.

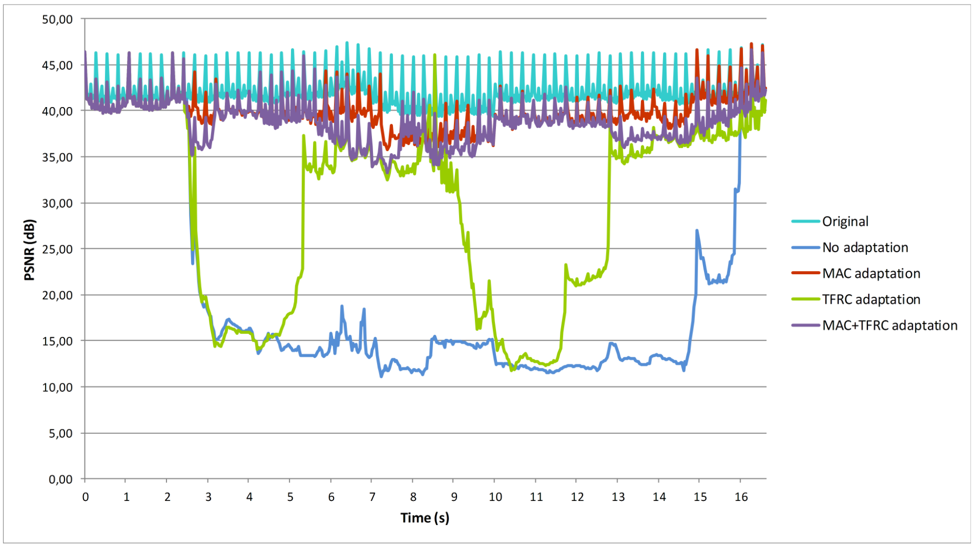


Figure 23: Received video quality measured in PSNR.

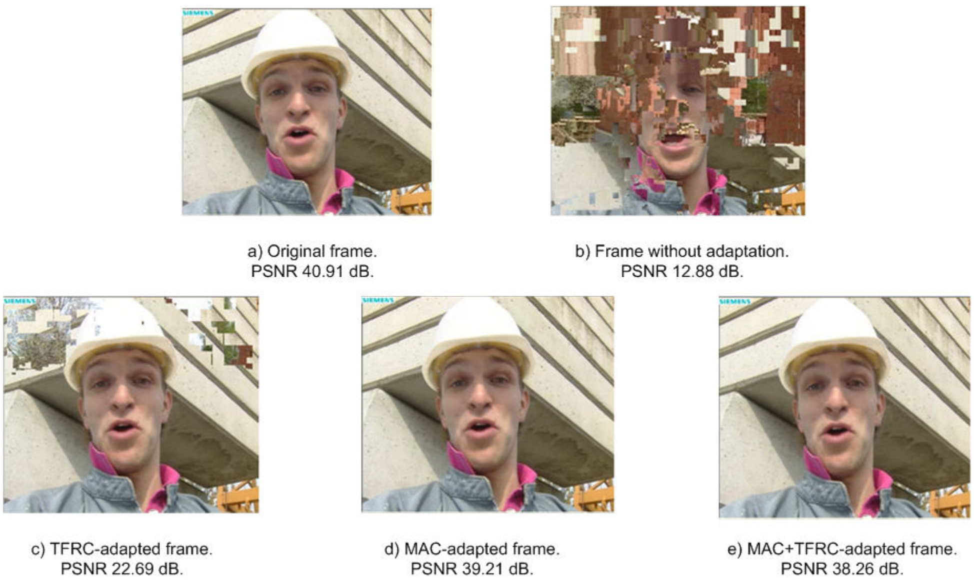


Figure 24: Visual frame quality comparison corresponding to the measured average PSNR values.

6 Physical layer

Cooperative communication is a new paradigm where each mobile unit collaborates with one or possibly several partners for the sake of reliably transmitting both its own information as well as the information of its partners jointly. More specifically, one or in fact several source nodes can transmit their signals to their respective destination nodes via one or several relay nodes. Cooperative communications is capable of increasing the capacity, transmission reliability, energy efficiency and coverage area of the overall system. As a benefit of these advantages, the IEEE 802.16j standard was approved by the IEEE-SA Standards Board on 13th May 2009. It is currently being developed for increasing both the transmission rate and coverage area of the IEEE 802.16e standard by employing fixed or nomadic relay terminals. More explicitly, the requirements for the MSHTM applications targeted in the Concerto project are as follows:

1. extended the coverage area in case of emergency situations, which occur outside the non-cooperative coverage area;
2. reliable communication for transmitting important medical information between the hospital and ambulances;
3. creating low-latency connections for enabling real-time video conferencing between medical staff;
4. low-complexity solutions for mobile terminals.

In a nut-shell, cooperative communication constitutes a promising solution for satisfying the above requirements. However, powerful channel coding and modulation schems are required in order to fulfil the potential offered by cooperative communications.

In this Chapter, we will present the states-of-the-art of powerful coding and modulation schemes as well as that of cooperative communication schemes, before outlining our research directions in these areas.

6.1 Coding and Modulation

Forward Error Correction (FEC) or **Channel Coding** in the context of digital communication has a history dating back to the middle of the twentieth century. In recent years, the field has been revolutionized by iterative detection aided codes, which are capable of approaching the theoretical limits of performance, namely the *channel capacity*. Important milestones in the area of channel coding are described in Table 4.

When the concatenated coding philosophy [104] was conceived back in 1966, it was deemed to have an excessive complexity and hence the resultant codes failed to stimulate immediate research interests. It was not until the discovery of Turbo Codes (TC) by Berrou *et al.* in 1993 [52], that efficient iterative decoding of concatenated codes became a reality at a low complexity by employing low-complexity constituent codes. There are two major types of iteratively decoded concatenated coding schemes, as discussed below:

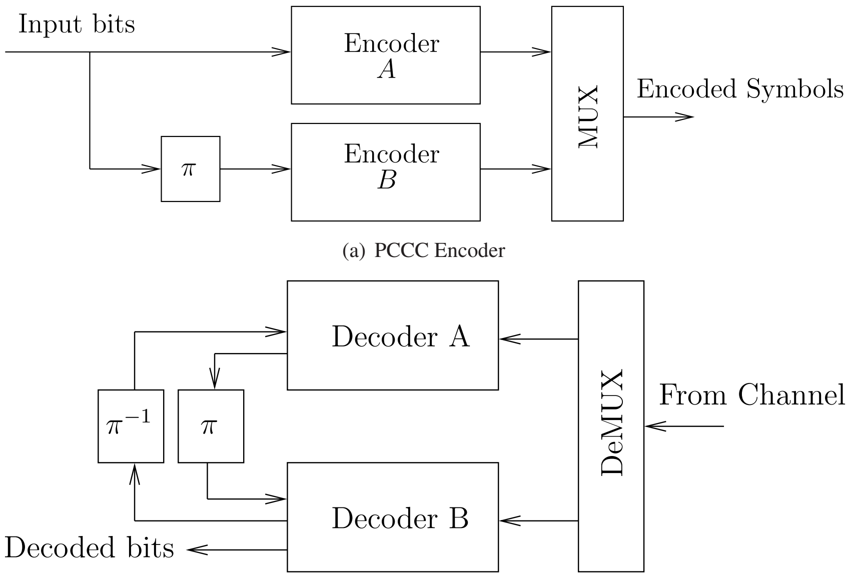
6.1.1 Parallel Concatenated Convolutional Codes

Classic TCs [52] consist of two or more parallel constituent codes [54]. The component codes are usually systematic codes, because their systematic nature simplifies the iterative exchange of information between the constituent decoders. In general, each component encoder independently encodes its input information and an interleaver (π) - also often termed as a scrambler - is used between the two constituent encoders to make both their input data and their encoded data statistically independent of each other, as shown in Figure 25(a).

Again, the encoders used in classic TCs are almost always RSC encoders, which output both the original information bits that are also referred to as systematic bits and the corresponding parity bits. Hence two codewords are generated by the two RSC codes, both of which contain the same original information bits, but typically these bits are only transmitted from one of the output streams. If both RSC encoders are half-rate encoders, the resultant TC becomes a third-rate code. However, the number of parity bits transmitted from the two streams can be appropriately adjusted by simply discarding the required fraction of parity bits. This so-called puncturing operation tacitly assumes that these bits were set to zero and hence the corresponding zeros have to be inserted in the right bit-positions at the decoder's input. In a nutshell, the redundant parity bits of both encoders may be transmitted, plus a single copy of the systematic information bit. At the decoder shown in Figure 25(b), two RSC

Year	*Milestone*
1948	Shannon's Capacity Theorem [57].
1950	Hamming codes were discovered by Hamming [233].
1954	Reed [126] and Muller [73] present Reed-Muller (RM) codes.
1955	Elias introduces convolutional codes [203].
1957	Prange introduces cyclic codes [83].
1959	Hocquenghem [19] and ...
1960	Bose and Chaudhuri [225] proposed BCH codes.
	Reed and Solomon defined (RS) codes over certain finite Galois fields [130].
	Peterson designed a BCH decoder [300].
1961	Peterson's book on Error Correction Codes (ECC) [301].
1962	Gallager invents LDPC codes [215].
	2400 BPS modem commercially available (4-PSK)(see [112]).
1963	Fano algorithm introduced for decoding convolutional codes [214].
	Massey describes threshold decoding [140].
1966	Forney's introduction of concatenated codes [104]
	and generalized minimum distance decoding [105].
1967	Berlekamp designs an efficient algorithm for BCH/RS decoding [79].
	Rudolph initiates the study of finite geometries for coding [154].
	4800 BPS modem commercially available (8-PSK)(see [112]).
1968	Berlekamp, documents Algebraic Coding Theory [80].
	Gallager publishes, Information theory and reliable communications [216].
1969	Jelinek defines the stack algorithm for decoding convolutional codes [90].
	Massey introduces his BCH decoding algorithm [141].
	Reed-Muller code used on Mariner deep space probes.
1971	Viterbi algorithm for Maximum Likelihood (ML) decoding of convolutional
	codes [28].
	9600 BPS modem commercially available (16-QAM) see [112].
1972	Bahl *et al.* invents the Maximum A-Posteriori (MAP) algorithm [152].
	Chase introduces his soft-decision-based block decoding algorithm [68].
	Peterson and Weldon revise their book [302].
1973	Forney further interprets the Viterbi algorithm [110].
1974	Bahl *et al.* describe the symbol based MAP algorithm [153].
1975	Sugiyama *et al.* invokes the Euclidean algorithm for decoding [310].
1977	MacWilliams and Sloane write The Theory of Error Correcting Codes [98].
	Voyager deep space mission uses a concatenated RS/convolutional code (see [220]).
1978	Wolf introduces trellis-decoding of block codes [137].
1980	14,400 BPS modem commercially available (64-QAM) (see [112]).
	Sony and Phillips standardize the compact disc, including
	a shortened RS code.
1981	Goppa introduces Algebraic-Geometry (AG) codes [281, 285].
1982	Ungerböck invents trellis-coded modulation (TCM) [107].
1983	Textbook on Error control coding by Lin and Costello [240].
	Blahut publishes his channel coding book [226].
1984	14,400 BPS TCM modem commercially available(128-TCM) (see [112]).
1985	19,200 BPS TCM modem commercially available(160-TCM) (see [112]).
1988	Divsalar and Simon discover multiple trellis-coded modulation [70].
1989	Hagenauer and Hoeher present the Soft-Output Viterbi Algorithm (SOVA) [133].
1990	Koch and Baier describe a reduced complexity MAP algorithm [286].
1992	Zehavi introduces Bit-Interleaved Coded Modulation (BICM) [84].
1993	Berrou, Glavieux, and Thitimajshima discover turbo codes [52].
	Honary, Markarian and Farrell $et\ al.$ presented low complexity trellis decoding
	of array [38] and Hamming codes [37].
1994	The Z_4 linearity of certain families of nonlinear codes is announced [36].
	Erfanian, Pasupathy and Gulak describe the Max-log-MAP algorithm [138].
1995	MacKay revives LDPC codes [183].
	Wicker publishes his textbook [250].
	Robertson, Villebrun and Hoeher desribe Log-MAP algorithm [206].
1996	Hagenauer, Offer and Papke propose turbo-BCH codes [134].
	33,600 BPS modem (V.34) modem is commercially available (see [111]).
	Sidorenko, Markarian and Honary presented a novel trellis design technique [283]
	for block and convolutional codes resulting in low complexity Viterbi decoding.
1997	Tarokh, Seshadri and Calderbank introduce space-time trellis coding (STTC) [284].
	Nickl, Hagenauer and Burkett report approaching the Shannon limit over
	Gaussian channels [118] within 0.27 dB.
	Schlegel writes his book on trellis coding [58].
	Ritcey and Li introduce Bit-Interleaved Coded Modulation with Iterative
	Decoding (BICM-ID) [305].
1998	Turbo trellis-coded modulation (TTCM) introduced by Robertson and Wörz [205].
	Alamouti introduces space-time block coding [259].
	Guruswami and Sudan present a list decoder for RS and AG codes [282].
1999	Ritcey and Li combine TCM with BICM-ID [306].
2000	Aji and McEliece [258] (and others [101]) synthesize
	several decoding algorithms using message passing ideas.
	Proakis publishes fourth edition of his textbook [139].
2002	Hanzo, Liew, and Yeap characterize turbo algorithms in [156].
	Siwamogsatham and Fitz introduce MTCM assisted STBC [242].
2003	Jafarkhani and Seshadri propose super-orthogonal STTC (SOSTTC) [117].
	Koetter and Vardy extend the GS algorithm for soft-decision
	decoding of RS codes [219].
2004	Lin and Costello publish second edition of their textbook [241].
2005	Moon publishes his textbook [279].
	Simon and Alouini write Digital Communications over Fading Channels [192].
	Song *et al.* introduce SOSTTC combined with QAM [26].

Table 4: Milestones in channel coding (1948-2006) [279, 156]



(a) PCCC Encoder

(b) PCCC decoder

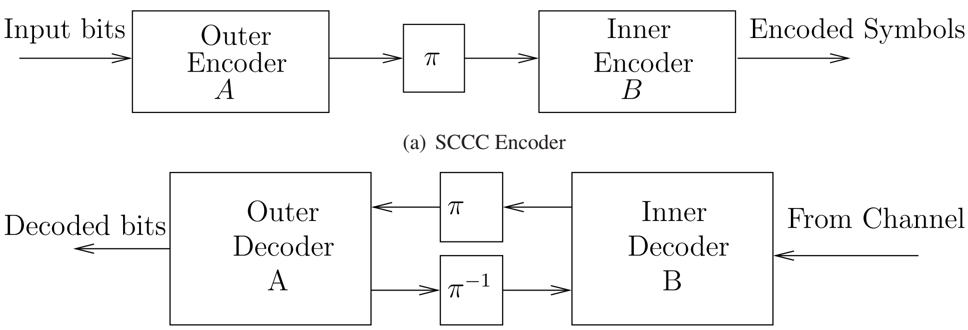
Figure 25: The schematic of a PCCC encoder and decoder.

decoders are used, which iteratively exchange their so-called soft-information, before making a hard-decision after a sufficiently high number of iterations.

The RSC constituent codes of classic TCs may also be replaced by other constituent codes. Inspired by this turbo coding concept various other coding arrangements, such as Turbo Trellis Coded Modulation (TTCM) schemes were proposed in [257], [236] and [205], which have a similar architecture to classic TCs, but employ Trellis Coded Modulation (TCM) constituent codes [269]. The appealing philosophy of TCM schemes is that they combine channel coding and modulation in an ingenious way, where the modulated signal constellation is extended to an increased number of constellation points, so that more bits per symbol can be transmitted for the sake of absorbing the parity bits. This way the constellation points have a reduced Euclidian distance amongst them, which potentially results in an increased Bit Error Ratio (BER), but this is more than compensated by the error correction capability of the Forward Error Correction (FEC) codec. It was also shown by Robertson *et al.* in [205] that TTCM is capable of outperforming classic TC.

6.1.2 Serial Concatenated Convolutional Codes

The serial concatenation of an outer and an inner encoder is shown in Figure 26(a). These codes were discovered by Benedetto *et al.* [239]. Typically the inner code is a weaker code and the outer code is a stronger code, which are separated by an interleaver as shown in Figure 26(a). The SCCCC decoder is shown in Figure 26(b).



(a) SCCC Encoder

(b) SCCC decoder

Figure 26: The schematic of an SCCC encoder and decoder.

To obtain higher code rates we may employ puncturing. SCCC codes have been shown to yield a performance comparable, and in some cases superior, to TC. The major scientific contributions on iterative detection and its convergence analysis are summarised in Tables 5 and 6.

Year	*Milestone*
1962	Gallager invented LDPC codes [215].
1966	Forney [104] proposed a novel concatenated coding scheme.
1974	Bahl *et al.* [153] invented the Maximum A-Posteriori Probability (MAP) algorithm.
1990	Koch and Baier describe a reduced complexity MAP algorithm [286].
1993	Berrou, Glavieux, and Thitimajshima invented the TCs [52].
1994	Erfanian, Pasupathy and Gulak describe the Max-log-MAP algorithm [138].
1995	Robertson *et al.* [206] proposed the log-MAP algorithm that results in similar performance to the MAP algorithm but at a significantly lower complexity.
	Divsalar *et al.* [69] applied turbo principle to multiple PCCCs.
	Douillard *et al.* [56] presented turbo equalisation, where iterative decoding was invoked for exchanging extrinsic information between a soft-output symbol detector and an outer channel decoder in order to overcome the multipath propagation effects in Gaussian and Rayleigh channels.
1996	Benedetto *et al.* [235] extended the turbo principle to serially concatenated block and convolutional codes.
1997	Loeliger proposed turbo-like codes using a single trellis for their decoding [115].
	Benedetto *et al.* [237] proposed an iterative detection scheme where iterations were carried out between the outer convolutional code and an inner TCM decoder.
	Caire *et al.* [102, 103] presented the BICM concept along with its design rules.
	Ritcey and Li [305] introduced Bit-Interleaved Coded Modulation using Iterative Decoding (BICM-ID).
1998	Robertson and Wörz introduced turbo trellis-coded modulation (TTCM) [205].
	Benedetto *et al.* [238, 239] studied multiple SCCCs combined with interleavers.
	Benedetto *et al.* proposed self-concatenated trellis coded modulation (SECTCM) schemes.
	ten Brink *et al.* [246] introduced a soft demapper between the multilevel demodulator and the channel decoder in an iteratively detected coded system.
1999	Wang *et al.* [307] proposed iterative multiuser detection and channel decoding for coded CDMA systems.
	Acikel and Ryan [201] designed high-rate punctured TCs.
2000	Divsalar *et al.* [71, 72] employed unity-rate inner codes for designing low-complexity iterative schemes for bandwidth/power limited systems having stringent BER requirements.
	ten Brink [243] proposed the employment of EXIT charts for analysing the convergence behaviour of iteratively detected systems.
2001	Lee [125] studied the effect of precoding on SCCC systems for transmission over ISI channels.
	ten Brink [245, 244] extended the employment of EXIT charts to three-stage PCCCs.
	El Gamal *et al.* [116] used SNR measures for studying the convergence behaviour of iterative decoding.
	Ramamurthy and Ryan [221] proposed the serial concatenation of convolutional differential encoders (accumulate codes), whose performance is better than those of PCCCs.
2002	Tüchler *et al.* [178] simplified the compuutation of EXIT charts.
	Tüchler *et al.* [179] compared several algorithms predicting the decoding convergence of iterative decoding schemes.
	Tüchler *et al.* [176] extended the EXIT chart analysis to three-stage SCCCs.

Table 5: Major concatenated schemes and iterative detection (1962-2002).

6.2 Cooperative Communication

Traditional direct transmission has its shortfalls, because when the MS roams at the fringe of the cell's coverage region while a conversation is in progress, initiating a handoff might not be possible due to the unavailability of unused channels or the lack of sufficient signal level at the adjacent cell. The call may be dropped in that scenario. Cooperative communication comes to our help in this case. It has the potential of extending the coverage area of a cell by creating an alternative transmission path from the MS to the base station (BS) via the introduction of a

Year	*Milestone*
2003	Sezgin *et al.* [25] proposed an iterative detection scheme, where a block code was used as an outer code and STBC as an inner code.
2004	Tüchler *et al.* [177] proposed a design procedure for creating systems exhibiting beneficial decoding convergence depending on the block length.
2005	Lifang *et al.* [157] showed that non-square QAM constellations can be decomposed into a parity-check block encoder having a recursive nature and a memoryless modulator. Iterative decoding was implemented in combination with an outer code for improving the system performance.
	Brännström *et al.* [89] considered EXIT chart analysis for multiple concatenated codes using 3-dimensional charts and proposed a way for finding the optimal activation order.
	Luo and Sweeney proposed the employment of cross-entropy as a novel method of predicting the convergence threshold of a TC, which achieved without imposing the usual conditions of either having a Gaussian distribution for the *a priori*/extrinsic information or perfect knowledge of the source information [213].
	Douillard and Berrou [55] showed that double-binary TCs are capable of achieving a better performance in comparison to classic TCs [52].
2006	Chatzigeorgiou *et al.* proposed a novel technique of finding the transfer function of a punctured TC designed for optimal performance [124].
2008	Carson *et al.* proposed a novel optimal bit-to-symbol mapping scheme for an 8PSK modulated BICM-ID system for transmission over quasi-static fading channels [298].
	Ng *et al.* [268] used EXIT charts and union bound analysis to compare the performance of near-capacity TTCM schemes.
	Maunder *et al.* [228] designed irregular variable length codes for the near-capacity operation of joint source and channel coding aided systems.
2009	Berrou *et al.* [53] proposed a low-complexity decoding algorithm for improving the performance of TCs in the 'turbo-cliff' region with the introduction of a rate-1 post-encoder applied in a classic TC scheme at the cost of imposing 10% increase in complexity.

Table 6: Major concatenated schemes and iterative detection (2003-2009).

relay, as shown in Figure 27. Another advantage of this is the creation of independent paths between the MS and the BS, namely the direct path between the two and the one via the relay.

There are various protocols that may be implemented at the relay channel. These can generally be organised into fixed and adaptive relaying schemes [146]. In fixed relaying schemes the channel resources are shared between the source and the relay in a time-invariant manner. They can be further divided into Amplify-And-Forward (AAF), Decode-And-Forward (DAF), Compress-And-Forward (CAF) and Coded Cooperation [146, 132]. The AAF scheme relies on a relay, which amplifies the received signal and then transmits it to the destination. Although the noise is also amplified along with the signal, we still gain spatial diversity by transmitting the signal over two spatially independent channels [143]. The DAF scheme has a relay which decodes the received signal transmitted by the source, re-encodes it and then forwards it to the destination, which combines all the independently faded signal replicas [143]. In CAF relaying [271, 106] the relay transmits a quantised and compressed version of the received signal in the form of source encoded symbols. At the destination, the source encoded i.e., compressed version of the relay's transmitted signal is decoded by mapping the received bits into a set of values that estimate the source's transmitted message, which are then combined with the message directly received from the source. Finally, in coded cooperation [275] incremental redundancy is introduced by the relay, which is then combined at the destination with the codeword sent by the source, resulting in a codeword benefitting from an increased amount of redundancy. While in some codes the information and redundancy are encoded in such a way that they are inseparable and only perfectly error-free decoding can separate them, some redundancy can be removed from the codeword in the case of punctured concatenated codes.

Major cooperative communications techniques have been outlined in Table 7. The basic idea behind cooperative communications can be traced back to the philosophy of the relay channel, which was introduced in 1971 by van der Meulen [81]. Although full-duplex relaying and the associated capacity theorem derived for the discrete memoryless relay channel model have been proposed by Cover and El Gamal [271], practical cooperative diversity schemes were only proposed much later in [23, 143, 85, 119]. In [22] Sendonaris *et al.* generalised the conventional relay model, where there is one source, one relay and one destination, to multiple nodes that transmit their own data as well as serve as relays for each other. The scheme of [22] was referred to as "user cooperation diversity". Sendonaris *et al.* presented in [23, 24] a simple user-cooperation methodology based on a DAF signalling

Year	*Milestone*
1971	van de Meulen [81] introduced a simple relay channel modeled by three terminals: a source, a destination and a relay. He studied the problem of transmission of information as effectively as possible from the source to the destination assuming that the relay cooperate in the transmission process [82].
1979	Cover and El Gamal [271] provided a thorough capacity analysis of the full-duplex relay channel.
1998	Sendonaris $et\ al.$ [22] generalised the relay model to multiple nodes that transmit their own data as well as serve as relays for each other.
2002	Hunter $et\ al.$ [275] introduced coded cooperation to achieve diversity in which the idea of cooperation was combined with the classic error-control-coding.
	Dohler $et\ al.$ [172] introduced the concept of virtual antenna arrays that emulates Alamouti's STBC for single-antenna-aided cooperating users.
2003	Sendonaris $et\ al.$ [23, 24] presented a simple user-cooperation diversity based algorithm, where a cooperative CDMA system is implemented.
	Laneman $et\ al.$ [142] developed different cooperative diversity protocols for exploiting spatial diversity in a cooperation scenario.
	Valenti and Zhao [188, 40] proposed a turbo coding scheme in a relay network.
2004	Laneman $et\ al.$ [143] developed cooperative diversity protocols and compared the performance of DAF, AAF, selection relaying and incremental relaying in terms of their outage behaviour.
	Nabar $et\ al.$ [229] analysed the spatial diversity performance of various signalling protocols.
	Janani $et\ al.$ [174] presented two extensions to the coded cooperation framework [275]: increased the diversity of coded cooperation via ideas borrowed from space-time codes and applied turbo codes in the proposed relay framework.
	Stefanov $et\ al.$ [27] analysed the performance of channel codes that are capable of achieving the full diversity provided by user cooperation in the presence of noisy interuser channels.
2005	Azarian $et\ al.$ [145] proposed cooperative signalling protocols that are capable of striking an attractive diversity-multiplexing tradeoff.
	Sneessens $et\ al.$ [122] proposed a soft decode-and-forward signalling strategy that can outperform the conventional DAF and AAF.
	Hu $et\ al.$ [217] advocated Slepian-Wolf cooperation that exploits distributed source coding in wireless cooperative communication.
	Yu [180] compared the AAF and DAF signalling schemes in practical scenarios.
	Kramer $et\ al.$ [106] addressed the information-theoretic aspects and considered DAF and CAF schemes for the wireless relay channels with many relays.
2006	Hunter $et\ al.$ [276, 277] further developed the idea of coded cooperation [275] by computing BER and FER bounds as well as the outage probability of coded cooperation.
	Li $et\ al.$ [309] employed soft information relaying in a BPSK modulated relay aided system employing turbo coding.
	Hu $et\ al.$ [218] proposed Wyner-Ziv cooperation as a generalisation of the Slepian-Wolf cooperation [217] combined with a compress-and-forward signalling strategy.
	Hϕst-Madsen [20] derived upper and lower bounds for the capacity of four-node ad hoc networks having two transmitters and two receivers using cooperative diversity.
2007	Bui $et\ al.$ [270] proposed soft information relaying where the relay's LLR values are quantised, encoded and superimposed, before being forwarded to the destination.
	Khormuji $et\ al.$ [193] improved the performance of the conventional DAF strategy by employing constellation rearrangement in the source and the relay.
	Bao $et\ al.$ [304] combined the benefits of AAF as well as DAF and proposed a new signalling strategy referred to as decode-amplify-forward.
	Xiao $et\ al.$ [158] introduced the concept of network coding in cooperative communications.
2008	Yue $et\ al.$ [108] compared the multiplexed coding and superposition coding in the coded cooperation system.
	Zhang $et\ al.$ [287] proposed a distributed space-frequency coded cooperation scheme for communication over frequency-selective channels.
	Wang $et\ al.$ [272] introduced the complex field network coding approach that can mitigate the throughput loss in conventional cooperative signalling schemes and attain full diversity gain.
2009	Hanzo $et\ al.$ [155] presented low-complexity cooperative MIMO codes and distributed turbo codes designed for two users cooperating for the sake of improving their attainable BER performance.
	Liu $et\ al.$ [146] authored a book on cooperative communications and networking.

Table 7: Major cooperative communications techniques (1971-2009).

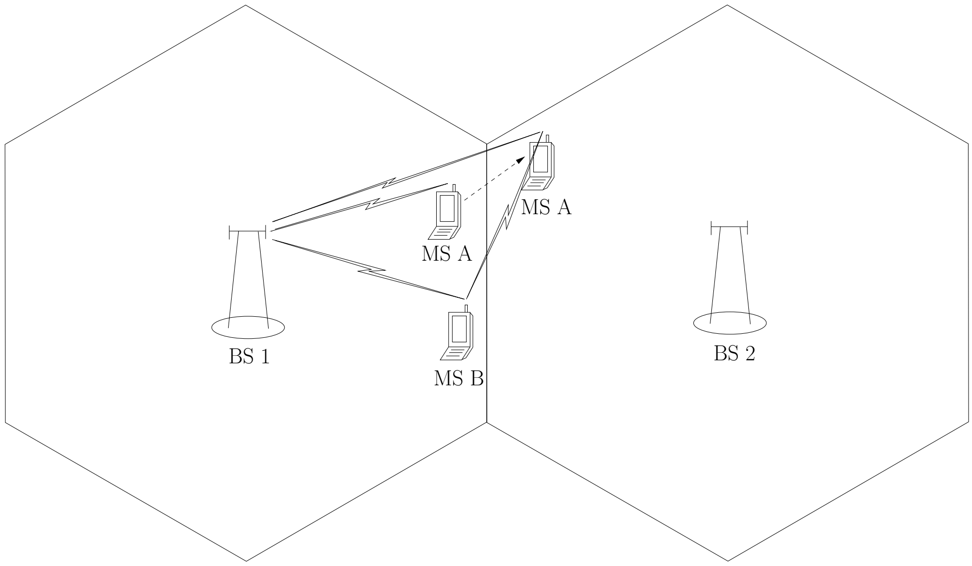


Figure 27: An example of a cooperative communication scenario.

scheme using CDMA. Dohler $et\ al.$ [172] introduced the concept of VAAs that emulates Alamouti's STBC for single-antenna-aided cooperating users. Space-time coded cooperative diversity protocols designed for exploiting spatial diversity in a cooperative scenario were proposed in [142]. In practice, each mobile collaborates with either a single or with a few partners for the sake of reliably transmitting both its own information and that of its partners in a concerted action, which emulates a virtual MIMO scheme.

Cooperative communications have been shown to offer significant performance gains in terms of various performance metrics, including improved diversity gains [142, 143, 204] as well as multiplexing gains [145]. Hunter $et\ al.$ [275] proposed the novel philosophy of coded cooperation schemes, which combine the idea of cooperation with the family of classic channel coding methods. Its extension to the framework of coded cooperation was presented in [174], where the diversity gain of coded cooperation was increased with the aid of ideas borrowed from the area of space-time codes. Additionally, a turbo coded scheme was proposed in [174] in the framework of cooperative communications. The performance benefits of channel codes in a coded cooperation aided scenario were quantified in [27]. Laneman $et\ al.$ proposed fixed (DAF and AAF), selection and incremental relaying protocols and compared them in [143].

Distributed coding [308] constitutes another attractive cooperative diversity technique, where joint signal design and coding are invoked at the source and relay nodes. Distributed turbo codes [40, 174] have also been proposed for cooperative communications, although typically under the simplifying assumption of having a perfect communication link between the source and the relay nodes. These are half-duplex relay-aided systems, where the source transmits to both the relay and destination during the first transmission period and after decoding the information from the source the relay re-encodes it and sends it to the destination in the second transmission period. Hence half-duplex systems do not suffer from multiple-access interference, which results in a simplified receiver structure at the cost of halving the spectral efficiency. As a more realistic design alternative, a turbo coded cooperation aided system having an imperfect source-relay (SR) communication link has been proposed in [317, 135]. In [317] the source node continues its transmission of the rest of the codeword in the second transmission period with the aim of achieving an improved bandwidth efficiency. Still referring to [317], the signals arriving from the source and relay are superimposed at the destination, where a *Maximum A Posteriori Probability* (MAP) detector and a turbo decoder exchange extrinsic information, which were shown to be capable of operating near the capacity of ergodic flat fading channels. The scheme proposed in [135] considers a more complex irregular Low-Density Parity-Check (LDPC) coded near-capacity system designed using EXIT charts and a design-procedure similar to that of [317]. It is demonstrated in [155, 173] that in the presence of Rayleigh fading, DAF cooperation-assisted systems are expected to outperform their non-cooperative counterparts. However, an error floor is observed in [155], which may be mitigated by using soft-relaying [122, 309].

6.3 Conclusion

We have presented the state-of-the-art in coding and modulation techniques in Section 6.1 and that of cooperative communications in Section 6.2. As seen in Section 6.1, the state of the art coding and modulation designs rely heavily on iterative detection. However, when the channel SNR is lower than the threshold, all transmitted bits in a modulated symbol will be corrupted completely partly due to error propagation during the iterative detection. A more advanced coded modulation scheme referred to as Turbo Trellis Coded Hierarchical Modulation (TTCHM) will be designed in the CONCERTO project, where the receiver can still retrieve at least some of the modulated bits, when the channel SNR is lower than the detection threshold. Furthermore, a low-complexity non-coherent modulation scheme referred to as StarQAM will be developed in the CONCERTO project, in order to enable the receiver to detect the modulated signal without the need to perform complex channel estimation.

We will then apply these advanced coded modulation techniques in the context of cooperative communication arrangements as outlined in Section 6.2. More specifically, a distributed TTCHM scheme can be designed with the aid of a relay node, in order to enable reliable transmission at low channel SNR values. The effects of channel estimation errors will also be studied. Finally, as a design alternative, non-coherent schemes dispensing with channel estimation will be investigated. Our related results will be disseminated in future WP5 deliverables.

7 Conclusion

This deliverable presented different techniques for the transmission of multimedia data required for realizing the CONCERTO use cases. The techniques presented will be improved during the WP5 activities, and the results will be described in subsequent deliverables (D5.2 and D5.3). The WP5 activities are aiming to improve the limitation identified and will allow to realize significant progress in the transmission of multimedia data, in particular on wireless network, for different healthcare and safety use case scenarios. These results will be demonstrated in WP6 together with techniques developed in other work packages in the CONCERTO simulation chain based on OMNeT++ or with network equipments tested in the CONCERTO demonstrator.

8 Acronyms

3GPP 3rd Generation Partnership Project

AF amplify-and-forward

ARQ adaptive repeat request

BIP binary integer program

BS base station

CDF cumulative distribution function

CF compress-and-forward

CP cyclic prefix

CQI class quality indicators

CSI channel state information

DF decode-and-forward

DFT discrete Fourier transform

EBEP equal-bit equal-power

eNB enhanced-NodeB

FDD frequency division duplexing

FDMA frequency division multiple access

FGS fine granularity scalability

FME first maximum expansion

GBR guaranteed bit-rate

H-ARQ hybrid-ARQ

HLGA heuristic localized gradient algorithm

I-FDMA interleaved-FDMA

IFFT inverse fast Fourier transform

IMT-A International Mobile Telecommunications - Advance

ISI inter-symbol interference

L-FDMA localized-FDMA

LTE Long Term Evolution

LTE-A LTE advanced

MAD minimum area difference

MGS medium grain scalability

MLWDF modified largest weighted delay first

MOS mean opinion score

MSE mean square error

MSHTM multi-source and holographic tele-medicine

OFDM orthogonal frequency division multiplexing

OFDMA orthogonal frequency division multiple access

PAPR peak-to-average power ratio

PELR packet error loss rate

PDB packet delay budget

PRB physical resource blocks

PSNR peak-to-signal-noise ratio

QAM quadrature amplitude modulation

QCI quality class identifier

QoS quality of service

QPSK quadrature phase shift keying

R-D rate-distortion

RME recursive maximum expansion

RRA radio resource allocation

SC-FDMA single carrier frequency division multiple access

SFR soft frequency reuse

STBPS search-tree based packet scheduling

SVC scalable video coding

TDD time division duplexing

TTI time transmission interval

WSR weighted sum-rate

References

[1] Omnet++ network simulation framework.

[2] 3gp-dash: http://www.3gpp.org/ftp/specs/html-info/26247.htm, Access date: June 2012.

[3] Http dynamic streaming: http://www.adobe.com/products/hds-dynamic-streaming/features.html, Access date: June 2012.

[4] Http live streaming: https://developer.apple.com/resources/http-streaming/, Access date: June 2012.

[5] Mpeg-dash: http://dashpg.com/, Access date: June 2012.

[6] Smooth streaming: http://www.iis.net/download/smoothstreaming, Access date: June 2012.

[7] 3GPP TS 22.105. Service aspects; services and service capabilities. Technical report, version 8.4.0 Release 8, 2007.

[8] 3GPP TR 22.803. Feasibility study on proximity-based services. Technical report, version 1.0.0 Release 12, 2012.

[9] 3GPP TS 22.805. Feasibility study on user plane congestion management. Technical report, version 12.0.1 Release 12, 2012.

[10] 3GPP TS 23.203. Policy and charging control architecture. Technical report, version 8.14.0 Release 8, 2012.

[11] 3GPP TS 23.203. Policy and charging control architecture. Technical report, version 10.7.0 Release 10, 2012.

[12] 3GPP TS 23.401. General packet radio service (gprs) enhancements for evolved universal terrestrial radio access network (e-utran) access. Technical report, version 8.16.0 Release 8, 2012.

[13] 3GPP TS 24.008. Mobile radio interface layer 3 specification; core network protocols. Technical report, version 8.18.0 Release 8, 2012.

[14] 3GPP TR 25.924. Opportunity driven multiple access (odma). Technical report, version 1.0.0 Release 4, 1999.

[15] 3GPP TS 36.216. Evolved universal terrestrial radio access (E-UTRA); physical layer for relaying operation (release 10). Technical report, version 10.3.1, 2011.

[16] 3GPP TS 36.300. Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; stage 2. Technical report, version 10.8.0 Release 10, 2012.

[17] 3GPP TS 36.300. Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; stage 2. Technical report, version 8.12.0 Release 8, 2012.

[18] 3GPP TR 36.814. Evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects (release 9). Technical report, version 9.0.0, 2010.

[19] A. Hocquenghem. Codes correcteurs d'erreurs. *Chiffres*, 2:147–156, 1959.

[20] A. Host-Madsen. Capacity bounds for cooperative diversity. *IEEE Transactions on Information Theory*, 52(4):1522–1544, April 2006.

[21] E. Erkip A. Sendonaris and B. Aazhang. User cooperation diversity - part I: System description. *IEEE Trans. Comm.*, pages 1–28, Vol. 51, No. 11, 2003.

[22] A. Sendonaris, E. Erkip and B. Aazhang. Increasing uplink capacity via user cooperation diversity. In *IEEE International Symposium on Information Theory*, Cambridge, MA, August 1998.

[23] A. Sendonaris, E. Erkip and B. Aazhang. User cooperation diversity part I: System description. *IEEE Transactions on Communications*, 51(11):1927–1938, 2003.

[24] A. Sendonaris, E. Erkip and B. Aazhang. User cooperation diversity part II: Implementation aspects and performance analysis. *IEEE Transactions on Communications*, 51(11):1939–1948, 2003.

[25] A. Sezgin, D. Wuebben and V. Kuehn. Analysis of mapping strategies for turbo-coded space-time block codes. In *Proceedings of IEEE Information Theory Workshop*, pages 103–106, Paris, France, March/April 2003.

[26] A. Song, G. Wang and X.G. Xia. Some super-orthogonal space-time trellis codes based on non-PSK MTCM. *IEEE Transactions on Wireless Communications*, 4(3):1214–1221, May 2005.

[27] A. Stefanov and E. Erkip. Cooperative coding for wireless networks. *IEEE Transactions on Communications*, 52(9):1470–1476, September 2004.

[28] A. Viterbi. Convolutional codes and their performance in communication systems. *IEEE Transactions on Communications [legacy, pre - 1988]*, 19:751–772, October 1971.

[29] S. Acharya and B. C. Smith. Middleman: A video caching proxy server. In *Proceedings of 10th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'00)*, 2000.

[30] R. Wakikawa adn R. Kuntz, Z. Thu, and L. Zhang. Global HA to HA Protocol Specification, 2012.

[31] A. Ahmad and M. Assaad. Polynomial-complexity optimal resource allocation framework for uplink SC-FDMA systems. In *Proc. IEEE Global Telecommunications Conf. (GLOBECOM 2011)*, pages 1–5, 2011.

[32] A. Ahmad and M. Assaad. Power efficient resource allocation in uplink SC-FDMA systems. In *Proc. IEEE 22nd Int Personal Indoor and Mobile Radio Communications (PIMRC) Symp*, pages 1351–1355, 2011.

[33] Shahriar Akter, Pradeep Ray, and John D'Ambra. Continuance of mhealth services at the bottom of the pyramid: the roles of service quality and trust. *Electronic Markets*, pages 1–19. 10.1007/s12525-012-0091-5.

[34] M. Al-Rawi, R. Jantti, J. Torsner, and M. Sagfors. Opportunistic uplilnk scheduling for 3G LTE systems. In *Proc. 4th Int. Conf. Innovations in Information Technology IIT '07*, pages 705–709, 2007.

[35] J. Almeida, D. Eager, and M. Vernon. A hybrid caching strategy for streaming media files. In *Proc. SPIE/ACM Conference on Multimedia Computing and Networking*, January 2001.

[36] A.R. Hammons, P.V. Kumar, A.R. Calderbank, N.J.A. Sloane and P. Sole. The Z_4-linearity of Kerdock, Preparata, Goethals and related codes. *IEEE Transactions on Information Theory*, 40:301, 1994.

[37] B. Honary and G. Markarian. Low-complexity trellis decoding of Hamming codes. *Electronics Letters*, 29(12):1114 –1116, June 1993.

[38] B. Honary, G. S. Markarian and P. G. Farrell. Generalised array codes and their trellis structure. *Electronics Letters*, 29(6):541–542, March 1993.

[39] J. Korhonen B. Patil, C. Williams. Approaches to Distributed mobility management using Mobile IPv6 and its extensions, 2011.

[40] B. Zhao and M.C. Valenti. Distributed turbo coded diversity for relay channel. *IEE Electronics Letters*, 39:786–787, May 2003.

[41] S. Barré, C. Paasch, and O. Bonaventure. Multipath TCP: From theory to practice. In *Proceedings of the 10th international IFIP TC 6 conference on Networking - Volume Part I*, pages 444–457, Valencia, Spain, 2011.

[42] I. Ben Chaabane, S. Hamouda, and S. Tabbane. A novel relay selection scheme for LTE-advanced system under delay and load constraints. In *Proc. IEEE Wireless Communications and Networking Conf. Workshops (WCNCW)*, pages 263–267, 2012.

[43] T. Beniero, S. Redana, J. Hamalainen, and B. Raaf. Effect of relaying on coverage in 3GPP LTE-advanced. In *Proc. IEEE 69th Vehicular Technology Conf. VTC Spring 2009*, pages 1–5, 2009.

[44] CJ. Bernardos and JC. Zuniga. PMIPv6-based distributed anchoring, 2012.

[45] P. Bertin, S. Bonjour, and J.-M. Bonnin. Distributed or centralized mobility? In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1 –6, 30 2009-dec. 4 2009.

[46] P. Bertin, Servane Bonjour, and J.-M. Bonnin. A distributed dynamic mobility management scheme designed for flat ip architectures. In *New Technologies, Mobility and Security, 2008. NTMS '08.*, pages 1 –5, nov. 2008.

[47] A. Bhamri, F. Kaltenberger, R. Knopp, and J. Hamalainen. Smart hybrid-ARQ (SHARQ) for cooperative communication via distributed relays in LTE-advanced. In *Proc. IEEE 12th Int Signal Processing Advances in Wireless Communications (SPAWC) Workshop*, pages 41–45, 2011.

[48] G. Bianchi, A. Detti, P. Loreti, C. Pisa, F.S. Proto, W. Kellerer, S. Thakolsri, and J. Widmer. Application-aware H.264 scalable video coding delivery over wireless LAN: experimental assessment. In *Second International Workshop on Cross Layer Design (IWCLD'09)*, pages 1–6, Palma de Mallorca, Spain, 2009.

[49] G. Blumrosen, N. Avisdris, R. Kupfer, and B. Rubinsky. C-smart: Efficient seamless cellular phone based patient monitoring system. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a*, pages 1 –6, june 2011.

[50] M. Boc, A. Petrescu, and C. Janneteau. Anchor-based routing optimization extension for proxy mobile ipv6 in flat architectures. In *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pages 1 –5, oct. 2011.

[51] O. Bulakci, S. Redana, B. Raaf, and J. Hamalainen. Impact of power control optimization on the system performance of relay based LTE-advanced heterogeneous networks. *Journal of Communications and Networks*, 13(4):345–359, 2011.

[52] C. Berrou, A. Glavieux and P. Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1. In *IEEE International Conference on Communications*, volume 2, pages 1064–1070, Geneva, May 1993.

[53] C. Berrou, A. Graell i Amat, Y. Ould-Cheikh-Mouhamedou and Y. Saouter. Improving the distance properties of turbo codes using a third component code: 3D turbo codes. *IEEE Transactions on Communications*, 57(9):2505 – 2509, September 2009.

[54] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: turbo-codes. *IEEE Transactions on Communications*, 44(10):1261–1271, October 1996.

[55] C. Douillard and C. Berrou. Turbo codes with rate-m/(m+1) constituent convolutional codes. *IEEE Transactions on Communications*, 53(10):1630 – 1638, October 2005.

[56] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier and A. Glavieux. Iterative correction of intersymbol interference: turbo equalization. *European Transaction on Telecommunications*, 6:507–511, September/October 1995.

[57] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379/623–423/656, July/October 1948.

[58] C. Schlegel. *Trellis coding*. John Wiley & Sons, New York, September 1997.

[59] F. D. Calabrese, P. H. Michaelsen, C. Rosa, M. Anas, C. U. Castellanos, D. L. Villa, K. I. Pedersen, and P. E. Mogensen. Search-tree based uplink channel aware packet scheduling for UTRAN LTE. In *Proc. IEEE Vehicular Technology Conf. VTC Spring 2008*, pages 1949–1953, 2008.

[60] Lei Cao, Jinyun Zhang, and N. Kanno. Multi-user cooperative communications with relay-coding for uplink IMT-advanced 4G systems. In *Proc. IEEE Global Telecommunications Conf. GLOBECOM 2009*, pages 1–6, 2009.

[61] Y. Chae, K. Guo, M. M. Buddhikot, S. Suri, and E. W. Zegura. Silo, rainbow, and caching token: Schemes for scalable, fault tolerant stream caching. In *IEEE Journal on Selected Areas in Communications*, volume 20, pages 1328–1344, 2002.

[62] H Anthony Chan, Hidetoshi Yokota, Jiang Xie, Pierrick Seite, and Dapeng Liu. Distributed and dynamic mobility management in mobile internet: Current approaches and issues. *Journal of Communications*, 6(1), 2011.

[63] S. Chen, B. Shen, Y. Yan, Basu S., and X. Zhang. Srb: Shared running buffers in proxy to exploit memory locality of multiple streaming media sessions. In *Proceedings of the 24th IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2004.

[64] Sergio Cicalò, Abdul Haseeb, and Velio Tralli. Fairness-oriented multi-stream rate adaptation using scalable video coding. *Elsevier Signal Processing: Image Communication*, 2012.

[65] Cisco. Global mobile data traffic forecast update, 2011-2016. Technical report, Cisco VNI White Paper, Feb 2012.

[66] C. Coletti, P. Mogensen, and R. Irmer. Performance analysis of relays in LTE for a realistic suburban deployment scenario. In *Proc. IEEE 73rd Vehicular Technology Conf. (VTC Spring)*, pages 1–5, 2011.

[67] T. M. Cover and A. A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Info. Theory*, pages 572–584, vol. 25, 1979.

[68] D. Chase. A class of algorithms for decoding block codes with channel measurement information. *IEEE Transactions on Information Theory*, IT-18(1):170–182, January 1972.

[69] D. Divsalar and F. Pollara. Multiple turbo codes for deep-space communications. TDA Progress Report 42-121, Jet Propulsion Laboratory, Pasadena, CA, May 1995.

[70] D. Divsalar and M.K. Simon. Multiple trellis coded modulation (MTCM). *IEEE Transactions on Communications*, 36(4):410–419, April 1988.

[71] D. Divsalar, S. Dolinar and F. Pollara. Low complexity turbo-like codes. In *2nd International Symposium on Turbo Codes and Related Topics*, pages 73–80, Brest, France, September 2000.

[72] D. Divsalar, S. Dolinar and F. Pollara. Serial concatenated trellis coded modulation with rate-1 inner code. In *IEEE Global Telecommunications Conference (GLOBECOM)*, volume 2, pages 777–782, San Francisco, CA, 2000.

[73] D. E. Muller. Application of boolean switching algebra to switching cicuit design. *IEEE Transactions on Computers*, 3:6–12, September 1954.

[74] M. Dai, D. Loguinov, and H. M. Radha. Rate-distortion analysis and quality control in scalable internet streaming. 8(6):1135–1146, 2006.

[75] O. Delgado and B. Jaumard. Scheduling and resource allocation for multiclass services in LTE uplink systems. In *Proc. IEEE 6th Int Wireless and Mobile Computing, Networking and Communications (WiMob) Conf*, pages 355–360, 2010.

[76] C. Delgorge, F. Courreges, L.A. Bassit, C. Novales, C. Rosenberger, N. Smith-Guerin, C. Bru, R. Gilabert, M. Vannoni, G. Poisson, and P. Vieyres. A tele-operated mobile ultrasound scanner using a light-weight robot. *Information Technology in Biomedicine, IEEE Transactions on*, 9(1):50 –58, march 2005.

[77] CONCERTO-MSR document. Preliminary definition of use cases and requirements. Technical report, version 1.0, 2012.

[78] Linhao Dong, Xu Zhu, and Yi Huang. Optimal asymmetric resource allocation for multi-relay based LTE-advanced systems. In *Proc. IEEE Global Telecommunications Conf. (GLOBECOM 2011)*, pages 1–5, 2011.

[79] E. Berlekamp. Nonbinaty BCH decoding. *Intl. Symp. on. Info. Th.*, 1967.

[80] E. Berlekamp. *Algebraic coding theory*. McGraw-Hill, New York, 1968.

[81] E. C. van der Meulen. Three-terminal communication channels. *Advanced Applied Probability*, 3(1):120–154, 1971.

[82] E. C. van der Meulen. A survey of multi-way channels in information theory: 1961-1976. *IEEE Transactions on Information Theory*, 23(1):1–37, January 1977.

[83] E. Prange. Cyclic error-correcting codes in two symbols. *AFCRC-TN-57, 103, Air Force Cambridge Research Center, Cambridge, Mass.*, 1972.

[84] E. Zehavi. 8-PSK trellis codes for a Rayleigh fading channel. *IEEE Transactions on Communications*, 40:873–883, May 1992.

[85] E. Zimmermann, P. Herhold and G. Fettweis. On the performance of cooperative relaying protocols in wireless networks. *European Transactions on Telecommunications*, 16(1):5–16, 2005.

[86] H. Chan (Ed.). Problem statement for distributed and dynamic mobility management, 2011.

[87] A. El Essaili, Liang Zhou, D. Schroeder, E. Steinbach, and W. Kellerer. QoE-driven live and on-demand LTE uplink video transmission. In *Proc. IEEE 13th Int Multimedia Signal Processing (MMSP) Workshop*, pages 1–6, 2011.

[88] Khalid Elgazzar, Mohamed Salah, Abd-Elhamid M. Taha, and Hossam Hassanein. Comparing uplink schedulers for LTE. In *Proceedings of the 6th International Wireless Communications and Mobile Computing Conference*, IWCMC '10, pages 189–193, 2010.

[89] F. Brännström, L.K. Rasmussen and A.J. Grant. Convergence analysis and optimal scheduling for multiple concatenated codes. *IEEE Transactions on Information Theory*, 51(9):3354–3364, September 2005.

[90] F. Jelinek. Fast sequential decoding algorithm using a stack. *IBM J. Res. Develop.*, pages 675–685, Nov 1969.

[91] H. Fahmi, M. Latif, S. Sedigh-Ali, A. Ghafoor, P. Liu, and L. Hsu. Proxy servers for scalable interactive video support. In *IEEE Computer*, volume 43, pages 54–60, 2001.

[92] Z. Faigl, L. Bokor, P. Neves, K. Daoud, and P. Herbelin. Evaluation of Two Integrated Signalling Schemes for the Ultra Flat Architecture using SIP, IEEE 802.21, and HIP/PMIP Protocols. *Comput. Netw.*, 55(7):1560–1575, May 2011.

[93] Y. Fallah, P. Nasiopoulos, and H. Alnuweiri. Scheduled and contention access transmission of partitioned H.264 video over WLANs. In *IEEE Global Communications Conference*, pages 2134–2139, Washington DC, USA, 2007.

[94] Zhong Fan and Siok Tan. M2m communications for e-health: Standards, enabling technologies, and research challenges. In *Medical Information and Communication Technology (ISMICT), 2012 6th International Symposium on*, pages 1 –4, march 2012.

[95] R. Fantini, D. Sabella, and M. Caretti. An E3F based assessment of energy efficiency of relay nodes in lte-advanced networks. In *Proc. IEEE 22nd Int Personal Indoor and Mobile Radio Communications (PIMRC) Symp*, pages 182–186, 2011.

[96] R. Fantini, D. Sabella, and M. Caretti. Energy efficiency in LTE-advanced networks with relay nodes. In *Proc. IEEE 73rd Vehicular Technology Conf. (VTC Spring)*, pages 1–5, 2011.

[97] M. Fischer, F.-U. Andersen, A. Kopsel, G. Schafer, and M. Schlager. A distributed ip mobility approach for 3g sae. In *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1 –6, sept. 2008.

[98] F.J. MacWilliams and J.A. Sloane. *The theory of error-correcting codes*. North-Holland, Amsterdam, 1977.

[99] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar. Architectural guidelines for multipath TCP development. IETF Request for Comments: 6182, March 2011.

[100] A. Ford, C. Raiciu, M. Handley, and O. Bonaventure. TCP extensions for multipath operation with multiple addresses. IETF Internet-Draft: draft-ietf-mptcp-multiaddressed-09, June 2012.

[101] F.R. Kschischang, B.J. Frey and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, February 2001.

[102] G. Caire, G. Taricco and E. Biglieri. Bit-interleaved coded modulation. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, page 96, Ulm, Germany, June/July 1997.

[103] G. Caire, G. Taricco and E. Biglieri. Bit-interleaved coded modulation. *IEEE Transactions on Information Theory*, 44(3):927–946, May 1998.

[104] G. Forney. *Concatenated codes*. MIT Press, Cambridge, 1966.

[105] G. Forney. Generalized minimum distance decoding. *IEEE Transactions of Information Theory*, 12(2):125–131, April 1966.

[106] G. Kramer, M. Gastpar and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory*, 51(9):3037–3063, September 2005.

[107] G. Ungerböck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, 28(1):55–67, 1982.

[108] G. Yue, X. Wang, Z. Yang and A. Host-Madsen. Coding schemes for user cooperation in low-power regimes. *IEEE Transactions on Signal Processing*, 56(5):2035–2049, May 2008.

[109] Alberto García-Martínez and Marcelo Bagnulo. The shim6 architecture for IPv6 multihoming. *IEEE Communications Magazine*, 48(9):152–157, 2010.

[110] G.D. Jr. Forney. The Viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–278, March 1973.

[111] G.D. Jr. Forney, L. Brown, M.V. Eyuboglu and J.L. III. Moran. The V.34 high speed modem standard. *IEEE Communications Magazine*, 34(12):28–33, December 1996.

[112] G.D. Jr. Forney, R.G. Gallager, G.R. Lang, F.M. Longstaff and S.U. Qureshi. Efficient modulation for band-limited channels. *IEEE Journal on Selected Areas in Communications*, 2(5):632–647, September 1984.

[113] F. Giust, A. De La Oliva, C.J. Bernardos, and R.P.F. Da Costa. A network-based localized mobility solution for distributed mobility management. In *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pages 1 –5, oct. 2011.

[114] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil. Proxy Mobile IPv6. RFC 5213 (Proposed Standard), August 2008. Updated by RFC 6543.

[115] H.-A. Loeliger. New turbo-like codes. In *IEEE International Symposium on Information Theory*, page 109, Ulm, Jun/Jul 1997.

[116] H. El Gamal and A.R. Hammons. Analyzing the turbo decoder using the Gaussian approximation. *IEEE Journal on Selected Areas in Communications*, 47:671–686, February 2001.

[117] H. Jafarkhani and N. Seshadri. Super-orthogonal space-time trellis codes. *IEEE Transactions on Information Theory*, 49(4):937–950, April 2003.

[118] H. Nickl, J. Hagenauer and F. Burkett. Approaching Shannon's capacity limit by 0.27 dB using simple Hamming codes. *IEEE Communications Letters*, 1(5):130–132, September 1997.

[119] H. Ochiai, P. Mitran and V. Tarokh. Design and analysis of collaborative diversity protocols for wireless sensor networks. In *IEEE Vehicular Technology Conference, VTC-'04 Fall*, pages 4645–4649, Los Angeles, USA, September 2004.

[120] A. Haseeb, M.G. Martini, S. Cicalo, and V. Tralli. Rate and distortion modeling for real-time MGS coding and adaptation. In *Wireless Advanced (WiAd), 2012*, pages 85 –89, june 2012.

[121] J. Heinanen and et al. Assured forwarding phb group. IETF Request for Comments: 2597, 1999.

[122] H.H. Sneessens and L. Vandendorpe. Soft decode and forward improves cooperative communications. In *6th IEE International Conference on 3G and Beyond*, pages 1–4, Washington, DC, November 2005.

[123] C. Hoymann, W. Chen, J. Montojo, A. Golitschek, and X. Shen C. Koutsimanis. Relaying operation in 3GPP LTE: Challenges and solutions. *IEEE Communications Magazine*, pages 156–162, Feb. 2012.

[124] I. Chatzigeorgiou, M.R.D. Rodrigues, I.J. Wassell and R. Carrasco. A novel technique to evaluate the transfer function of punctured turbo codes. In *IEEE International Conference on Communications, ICC '06.*, volume 3, pages 1166–1171, June 2006.

[125] I. Lee. The effect of a precoder on serially concatenated coding systems with an ISI channel. *IEEE Transactions on Communications*, 49(7):1168–1175, July 2001.

[126] I. Reed. A class of multiple-error-correcting codes and the decoding scheme. *IEEE Transactions on Information Theory*, 4(4):38–49, September 1954.

[127] IEEE. MAC enhancements for robust audio video streaming, 2012. IEEE Std 802.11aa-2012.

[128] IEEE. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, 2012. IEEE Std 802.11-2012.

[129] R. Irmer and F. Diehm. On coverage and capacity of relaying in LTE-advanced in example deployments. In *Proc. IEEE 19th Int. Symp. Personal, Indoor and Mobile Radio Communications PIMRC 2008*, pages 1–5, 2008.

[130] I.S. Reed and G. Solomon. Polynomial codes over certain finite fields. *J. Soc. Ind. Appl. Math.*, 8:300–304, June 1960.

[131] ITU-T. *H.264: Advanced video coding for generic audiovisual services*, 2007.

[132] J. Abouei, H. Bagheri and A. Khandani. An efficient adaptive distributed space-time coding scheme for cooperative relaying. *IEEE Transactions on Wireless Communications*, 8(10):4957–4962, October 2009.

[133] J. Hagenauer and P. Hoeher. A Viterbi algorithm with soft-decision outputs and its applications. In *IEEE Global Telecommunications Conference, GLOBECOM '89.*, pages 1680–1686, Dallas, TX, November 1989.

[134] J. Hagenauer, E. Offer and L. Papke. Iterative decoding of binary block and convolutional codes. *IEEE Transactions on Information Theory*, 42(2):429–445, March 1996.

[135] J. Hu and T. M. Duman. Low density parity check codes over wireless relay channels. *IEEE Transactions on Wireless Communications*, 6(9):3384–3394, September 2007.

[136] D. N. C. Tse J. N. Laneman and G. W. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Trans. Inform. Theory*, pages 3062–3080, vol. 50, no. 12, 2004.

[137] J. Wolf. Efficient maximum likelihood decoding of linear block codes using a trellis. *IEEE Transactions on Information Theory*, 24(1):76–80, January 1978.

[138] J.A. Erfanian, S. Pasupathy and G. Gulak. Reduced complexity symbol dectectors with parallel structures for ISI channels. *IEEE Transactions on Communications*, 42:1661–1671, 1994.

[139] J.G. Proakis. *Digital communications*. McGraw Hill Higher Education, 4th edition edition, December 1, 2000.

[140] J.L. Massey. *Threshold decoding*. MIT Press: Cambridge, Mass., 1963.

[141] J.L. Massey. Shift-register synthesis and BCH decoding. *IEEE Trans. Information. Theory*, IT-15:122–127, 1969.

[142] J.N. Laneman and G.W. Wornell. Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks. *IEEE Transactions on Information Theory*, 49(10):2415–2425, 2003.

[143] J.N. Laneman, D.N. Tse and G.W. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, 50(12):3062–3080, December 2004.

[144] H.-H. Juan, H.-C. Huang, C. Huang, and T. Chiang. Cross-layer mobile WiMAX MAC designs for the H.264/AVC scalable video coding. *Springer Wireless Networks*, 16:113–123, January 2010.

[145] K. Azarian, H. El Gamal and P. Schniter. On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels. *IEEE Transactions on Information Theory*, 51(12):4152–4172, December 2005.

[146] K.J.R. Liu, A.K. Sadek, W. Su and A. Kwasinski. *Cooperative communications and networking*. Cambridge University Press, 2009.

[147] I. Kofler, J. Seidl, C. Timmerer, H. Hellwagner, I. Djama, and T. Ahmed. Using MPEG-21 for cross-layer multimedia content adaptation. *Springer Journal on Signal, Image and Video Processing*, 2:355–370, 2008.

[148] I. Korhonen, J. Parkka, and M. Van Gils. Health monitoring in the home of the future. *Engineering in Medicine and Biology Magazine, IEEE*, 22(3):66 – 73, may-june 2003.

[149] R. Kuntz, D. Sudhakar, R. Wakikawa, and L. Zhang. A Summary of Distributed Mobility Management, 2011.

[150] Do-Kyoung Kwon, Mei-Yin Shen, and C.-C. J. Kuo. Rate control for h.264 video with enhanced rate and distortion models. 17(5):517–529, 2007.

[151] Sungoh Kwon and Neung-Hyung Lee. Uplink QoS scheduling for LTE system. In *Proc. IEEE 73rd Vehicular Technology Conf. (VTC Spring)*, pages 1–5, 2011.

[152] L. Bahl, C. Cullum, W. Frazer and F. Jelinek. An efficient algorithm for computing free distance. *IEEE Transactions on Information Theory*, 18(3):437–439, May 1972.

[153] L. Bahl, J. Cocke, F. Jelinek and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Transactions on Information Theory*, 20(2):284–287, March 1974.

[154] L. D. Rudolph. A class of majority logic decodable codes. *IEEE Trans. Inform. Theory*, 13:305–307, May 1967.

[155] L. Hanzo, O. Alamri, M. El-Hajjar and N. Wu. *Near-capacity multi-functional MIMO systems*. IEEE Press - John Wiley, April 2009.

[156] L. Hanzo, T.H. Liew and B.L. Yeap. *Turbo coding, turbo equalisation and space time coding for transmission over fading channels*. Wiley: IEEE Press, Chichester, UK, 2002.

[157] L. Lifang, D. Divsalar and S. Dolinar. Iterative demodulation, demapping and decoding of coded non-square QAM. In *IEEE Transactions on Communications*, volume 53, pages 16–19, January 2005.

[158] L. Xiao, T. Fuja, J. Kliewer and D. Costello. A network coding approach to cooperative diversity. *IEEE Transactions on Information Theory*, 53(10):3714–3722, October 2007.

[159] E. Lai and K.E. Friedl. Digital soldiers: Transforming personalized health in challenging and changing environments. In *Wearable Micro and Nano Technologies for Personalized Health (pHealth), 2009 6th International Workshop on*, pages 5 –8, june 2009.

[160] C. Lamy-Bergot, R. Fracchia, M. Mazzotti, S. Moretti, E. Piri, T. Sutinen, J. Zhuo, J. Vehkaperä, G. Feher, G. Jeney, G. Panza, and P. Amon. Optimisation of multimedia over wireless IP links via X-layer design: an end-to-end transmission chain simulator. *Springer Multimedia Tools and Applications*, 55(2):261–288, November 2011.

[161] E. Lang, S. Redana, and B. Raaf. Business impact of relay deployment for coverage extension in 3GPP LTE-advanced. In *Proc. IEEE Int. Conf. Communications Workshops ICC Workshops 2009*, pages 1–5, 2009.

[162] JunHua Li and P. Ray. Applications of e-health for pandemic management. In *e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on*, pages 391 –398, july 2010.

[163] Q. Li, R. Q. HU, Y. Qian, and G. Wu. Cooperative communications for wireless networks: techniques and applications in LTE-advanced systems. *IEEE Wireless Communications*, pages 22–29, April 2012.

[164] G. Liebl, T. M. de Moraes, A. Soysal, and E. Seidel. Fair resource allocation for inband relaying in LTE-advanced. In *Proc. 8th Int Multi-Carrier Systems & Solutions (MC-SS) Workshop*, pages 1–5, 2011.

[165] G. Liebl, T. Schierl, T. Wiegand, and T. Stockhammer. Advanced wireless multiuser video streaming using the scalable video coding extensions of H.264/MPEG-4-AVC. In *IEEE International Conference on Multimedia and Expo (ICME'06)*, pages 625–628, Toronto, Canada, 2006.

[166] M. Liebsch. Per-Host Locators for Distributed Mobility Management, 2011.

[167] Junsung Lim, H. G. Myung, Kyungjin Oh, and D. J. Goodman. Channel-dependent scheduling of uplink single carrier FDMA systems. In *Proc. VTC-2006 Fall Vehicular Technology Conf. 2006 IEEE 64th*, pages 1–5, 2006.

[168] Junsung Lim, H. G. Myung, Kyungjin Oh, and D. J. Goodman. Proportional fair scheduling of uplink single-carrier FDMA systems. In *Proc. IEEE 17th Int Personal, Indoor and Mobile Radio Communications Symp*, pages 1–6, 2006.

[169] J. Liu, X. Chu, and J. Xu. Proxy cache management for fine-grained scalable video streaming. In *Proceedings of IEEE INFOCOM*, 2004. Hong Kong.

[170] Jiangchuan Liu and Jianliang Xu. Proxy caching for media streaming over the internet. In *IEEE Communications, Feature Topic on Proxy Support for Streaming on the Internet*, August 2004.

[171] P. Louin and P. Bertin. Network and host based distributed mobility. In *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pages 1 –5, oct. 2011.

[172] M. Dohler, E. Lefranc and H. Aghvami. Space-time block codes for virtual antenna arrays. In *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 1, pages 414–417, September 2002.

[173] M. Elfituri, W. Hamouda and A. Ghrayeb. A convolutional-based distributed coded cooperation scheme for relay channels. *IEEE Transactions on Vehicular Technology*, 58(2):655–669, February 2009.

[174] M. Janani, A. Hedayat, T. Hunter and A. Nosratinia. Coded cooperation in wireless communications: space-time transmission and iterative decoding. *IEEE Transactions on Signal Processing*, 52(2):362–371, February 2004.

[175] Li Junyi V. Park T. Richardson G. Tsirtsis M. S. Corson, R. Laroia. Toward proximity-aware internetworking. *IEEE Wireless Comm.*, pages 26–33, Vol 17, issue 6, 2010.

[176] M. Tüchler. Convergence prediction for iterative decoding of threefold concatenated systems. In *IEEE Global Telecommunications Conference (GLOBECOM)*, volume 2, pages 1358–1362, Taipei, Taiwan, November 2002.

[177] M. Tüchler. Design of serially concatenated systems depending on the block length. *IEEE Transactions on Communications*, 52(2):209–218, 2004.

[178] M. Tüchler and J. Hagenauer. EXIT charts of irregular codes. In *Conference on Information Science and Systems*, pages 748–753, Princtone, NJ, March 2002.

[179] M. Tüchler, S. ten Brink and J. Hagenauer. Measures for tracing convergence of iterative decoding algorithms. In *Proceedings of the 4th International ITG Conference on Source and Channel Coding*, pages 53–60, Berlin, Germany, January 2002.

[180] M. Yu and J. Li. Is amplify-and-forward practically better than decode-and-forward or vice versa? In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, March 2005.

[181] Xuebin Ma, Tao Wen, Quan Guo, and Gang Wang. An enhanced dynamic and distributed domain-based handoff scheme for mobile ipv6. In *Computer and Information Technology, 2007. CIT 2007. 7th IEEE International Conference on*, pages 375 –380, oct. 2007.

[182] Zhangchao Ma, Wei Xiang, Hang Long, and Wenbo Wang. Proportional fair resource partition for LTE-advanced networks with type i relay nodes. In *Proc. IEEE Int Communications (ICC) Conf*, pages 1–5, 2011.

[183] MacKay and Neal. Good codes based on very sparse matrices. In *IMA: IMA Conference on Cryptography and Coding, LNCS lately (earlier: Cryptography and Coding II, Edited by Chris Mitchell, Clarendon Press, 1992)*, 1995.

[184] David Malan, Thaddeus Fulford-jones, Matt Welsh, and Steve Moulton. Codeblue: An ad hoc sensor network infrastructure for emergency medical care. In *In International Workshop on Wearable and Implantable Body Sensor Networks*, 2004.

[185] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos. Channel aware multiuser scalable video streaming over lossy under-provisioned channels: Modeling and analysis. *IEEE Trans. on Multimedia*, 10(7):1366–1381, 2008.

[186] H. Mansour, V. Krishnamurthy, and P. Nasiopoulos. Rate and distortion modeling of medium grain scalable video coding. In *Proc. 15th IEEE Int. Conf. Image Processing ICIP 2008*, pages 2564–2567, 2008.

[187] E. Mattila, J. Parkka, M. Hermersdorf, J. Kaasinen, J. Vainio, K. Samposalo, J. Merilahti, J. Kolari, M. Kulju, R. Lappalainen, and I. Korhonen. Mobile diary for wellness management; results on usage and usability in two user studies. *Information Technology in Biomedicine, IEEE Transactions on*, 12(4):501 –512, july 2008.

[188] M.C. Valenti and B. Zhao. Distributed turbo codes: towards the capacity of the relay channel. In *IEEE Vehicular Technology Conference, VTC-'03 Fall*, volume 1, pages 322–326, October 2003.

[189] M. Mehta, S. Khakurel, and A. Karandikar. Buffer-based channel dependent uplink scheduling in relay-assisted LTE networks. In *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, pages 1777–1781, 2012.

[190] T. Melia, A. de la Olivia, I. Soto, C.J. Bernardos, and A. Vidal. Analysis of the effect of mobile terminal speed on wlan/3g vertical handovers. In *Proceeding of IEEE Global Communications Conference*, pages 1–6, San Francisco, USA, 2006.

[191] Z. Miao and A Ortega. Scalable proxy caching of video under storage constraints. In *IEEE Journal on Selected Areas in Communications*, volume 20, pages 1315–1327, Sept. 2002.

[192] M.K. Simon and M.S. Alouini. *Digital communications over fading channels*. John Wiley Sons, Inc., 2nd edition edition, 2005.

[193] M.N. Khormuji and E.G. Larsson. Improving collaborative transmit diversity by using constellation rearrangement. In *IEEE Wireless Communications and Networking Conference*, pages 803–807, Kowloon, March 2007.

[194] R. Moskowitz and P. Nikander. Host identity protocol (HIP) architecture. IETF Request for Comments: 4423, May 2006.

[195] S.G. Mougiakakou, I. Kouris, D. Iliopoulou, A. Vazeou, and D. Koutsouris. Mobile technology to empower people with diabetes mellitus: Design and development of a mobile application. In *Information Technology and Applications in Biomedicine, 2009. ITAB 2009. 9th International Conference on*, pages 1 –4, nov. 2009.

[196] H. G. Myung, Kyungjin Oh, Junsung Lim, and D. J. Goodman. Channel-dependent scheduling of an uplink SC-FDMA system with imperfect channel information. In *Proc. IEEE Wireless Communications and Networking Conf. WCNC 2008*, pages 1860–1864, 2008.

[197] Hyung G. Myung, Junsung Lim, and David J. Goodman. Single carrier FDMA for uplink wireless transmission. *IEEE Vehicular Technology Magazine*, 1(3):30–38, 2006.

[198] S. Nagata, Y. Yan, X. Gao, A. Li, H. Kayama, T. Abe, and T. Nakamura. Investigation on system performance of l1/l3 relays in LTE-advanced downlink. In *Proc. IEEE 73rd Vehicular Technology Conf. (VTC Spring)*, pages 1–5, 2011.

[199] Wei Ni, I. B. Collings, and Ren Ping Liu. Relay handover and link adaptation design for fixed relays in IMT-advanced using a new markov chain model. *IEEE Transactions on Vehicular Technology*, 61(4):1839–1853, 2012.

[200] E. Nordmark and M. Bagnulo. Shim6: Level 3 multihoming shim protocol for IPv6. IETF Request for Comments: 5533, 2009.

[201] O. F. Acikel and W. E. Ryan. Punctured turbo-codes for BPSK/QPSK channels. *IEEE Transactions on Communications*, 47(9):1315–1323, September 1999.

[202] OneFIT+. functional and system architecture. Technical report, D2.2, 2011.

[203] P. Elias. Coding for noisy channels. *IRE Conv. Rept.*, pages 37–47, 1955.

[204] P. Mitran, H. Ochiai and V. Tarokh. Space-time diversity enhancements using collaborative communications. *IEEE Transactions on Information Theory*, 51(6):2041–2057, June 2005.

[205] P. Robertson and T. Wörz. Bandwidth-efficient turbo trellis-coded modulation using punctured component codes. *IEEE Journal on Selected Areas in Communications*, 16(2):206–218, February 1998.

[206] P. Robertson, E. Villebrun and P. Hoeher. A comparison of optimal and sub-optimal MAP decoding algorithms operating in the Log domain. In *Proceedings of International Conference on Communications*, pages 1009–1013, Seattle, USA, June 1995.

[207] Mugen Peng, Changqing Yang, Zhongyuan Zhao, Wenbo Wang, and Hsiao-Hwa Chen. Cooperative network coding in relay-based IMT-advanced systems. *IEEE Communications Magazine*, 50(4):76–84, 2012.

[208] C. Perkins, D. Johnson, and J. Arkko. Mobility support in IPv6. IETF Request for Comments: 6275, 2011.

[209] S. Pierrel, P. Jokela, J. Melén, and K. Slavov. A policy system for simultaneous multiaccess with host identity protocol. In *1st IEEE Workshop on Autonomic Communication and Network Management (ACNM'07)*, pages 71–77, Munich, Germany, May 2007.

[210] E. Piri, M. Uitto, J. Vehkaperä, and T. Sutinen. Dynamic cross-layer adaptation of scalable video in wireless networking. In *IEEE Global Communications Conference*, Miami, Florida, USA, 2010.

[211] T. Pliakas, G. Kormentzas, and S. Tsekeridou. Joint scalable video coding and packet prioritisation for video streaming over IP/802.11e heterogeneous networks. In *Proceedings of the 3rd International ICST Mobile Multimedia Communications Conference*, Nafpakos, Greece, August 2007.

[212] Chang Woo Pyo, Jie Li, and Hisao Kameda. A dynamic and distributed domain-based mobility management method for mobile ipv6. In *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, volume 3, pages 1964 – 1968 Vol.3, oct. 2003.

[213] Q. Luo and P. Sweeney. Method to analyse convergence of turbo codes. *Electronics Letters*, 41(13): 757 – 758, June 2005.

[214] R. Fano. A heuristic discussion of probabilistic decoding. *IEEE Transactions on Information Theory*, 9(2):64–74, April 1963.

[215] R. Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8(1):21–28, January 1962.

[216] R. Gallager. *Information theory and reliable communication*. Wiley, New York, 1968.

[217] R. Hu and J. Li. Exploiting Slepian-Wolf codes in wireless user cooperation. In *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, pages 275–279, June 2005.

[218] R. Hu and J. Li. Practical compress-forward in user cooperation: Wyner-Ziv Cooperation. In *IEEE International Symposium on Information Theory*, pages 489–493, Seattle, WA, July 2006.

[219] R. Koetter and A. Vardy. Algebraic soft-decision decoding of Reed-Solomon codes. *IEEE Transactions on Information Theory*, 49(11):2809–2825, November 2003.

[220] R. McEliece and L. Swanson. *Reed-Solomon codes and their applications*. IEEE Press, New York, 1994.

[221] R. Ramamurthy and W.E. Ryan. Convolutional double accumulate codes (or double turbo DPSK). *IEEE Communications Letters*, 5(4):157–159, April 2001.

[222] C. Raiciu, M. Handly, and D. Wischik. Coupled congestion control for multipath transport protocols. IETF Request for Comments: 6356, October 2011.

[223] Costin Raiciu, Sebastien Barre, Christopher Pluntke, Adam Greenhalgh, Damon Wischik, and Mark Handley. Improving datacenter performance and robustness with multipath TCP. In *ACM SIGCOMM 2011 (SIGCOMM '11)*, pages 266–277, 2011.

[224] Costin Raiciu, Dragos Niculescu, Marcelo Bagnulo, and Mark Handley. Opportunistic mobility with multipath TCP. In *MobiArch'11*, pages 7–12, Bethesda, Maryland, USA, June 2011.

[225] R.C. Bose and D.K. Ray-Chaudhuri. On a class of error correcting binary group codes. *Information and Control*, 3:68–79, March 1960.

[226] R.E. Blahut. *Theory and practice of error control codes*. Addison-Wesley, Reading, MA, USA, 1983.

[227] Zhe Ren, Saleh, A. B., Bulakci, O., Redana, S., B. Raaf, and J. Hamalainen. Joint interference coordination and relay cell expansion in LTE-advanced networks. In *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, pages 2874–2878, 2012.

[228] R.G. Maunder, J. Wang, S.X. Ng, L.-L. Yang and L. Hanzo. On the performance and complexity of irregular variable length codes for near-capacity joint source and channel coding. *IEEE Transactions on Wireless Communications*, 7(4):1338–1347, April 2008.

[229] R.U. Nabar, H. Bolcskei and F.W. Kneubuhler. Fading relay channels: performance limits and space-time signal design. *IEEE Journal on Selected Areas in Communications*, 22(6):1099–1109, August 2004.

[230] R. Ruby, A. Mohamed, and V. Leung. Utility-based uplink scheduling algorithm for enhancing throughput and fairness in relayed LTE networks. In *Proc. IEEE 35th Conf. Local Computer Networks (LCN)*, pages 520–527, 2010.

[231] L. Ruiz de Temino, G. Berardinelli, S. Frattasi, and P. Mogensen. Channel-aware scheduling algorithms for SC-FDMA in LTE uplink. In *Proc. IEEE 19th Int. Symp. Personal, Indoor and Mobile Radio Communications PIMRC 2008*, pages 1–6, 2008.

[232] M. Rumney. 3GPP LTE: Introducing single-carrier FDMA. *Agilent measurement journal*, 4:18–27, 2008.

[233] R.W. Hamming. Error detecting and error correcting codes. *Bell Syst. Tech. Journal*, 29:41–56, 1950.

[234] Chen S., Shen B., Wee S., and Zhang X. Designs of high quality streaming proxy systems. In *Proc. IEEE INFOCOM*, 2004. Hong Kong.

[235] S. Benedetto and G. Montorsi. Iterative decoding of serially concatenated convolutional codes. *Electronics Letters*, 32(13):1186–1188, June 1996.

[236] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara. Bandwidth efficient parallel concatenated coding schemes. *Electronics Letters*, 31(24):2067–2069, November 1995.

[237] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara. A soft-input soft-output APP module for iterative decoding of concatenated codes. *IEEE Communications Letters*, 1(1):22–24, January 1997.

[238] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara. Self-concatenated trellis coded modulation with self-iterative decoding. In *IEEE Global Telecommunications Conference*, volume 1, pages 585–591, Sydney, NSW, Australia, 1998.

[239] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara. Serial concatenation of interleaved codes: performance analysis, design and iterative decoding. *IEEE Transactions on Information Theory*, 44(3):909–926, May 1998.

[240] S. Lin and D.J. Jr. Costello. *Error control coding: Fundamentals and applications*. Prentice-Hall, Englewood Cliffs, N.J., 1983.

[241] S. Lin and D.J. Jr. Costello. *Error control coding: Second edition*. Prentice-Hall, Upper Saddle River, New Jersey, 2004.

[242] S. Siwamogsatham and M. Fitz. Robust space-time coding for correlated Rayleigh fading channels. *IEEE Transactions on Signal Processing*, 50(10):2408–2416, October 2002.

[243] S. ten Brink. Designing iterative decoding schemes with the extrinsic information transfer chart. *AEÜ International Journal of Electronics and Communications*, 54(6):389–398, November 2000.

[244] S. ten Brink. Convergence behavior of iteratively decoded parallel concatenated codes. *IEEE Transactions on Communications*, 49(10):1727–1737, October 2001.

[245] S. ten Brink. Convergence of multidimensional iterative decoding schemes. In *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 270–274, Pacific Grove, CA, USA, 2001.

[246] S. ten Brink, J. Speidel and R.-H. Yan. Iterative demapping and decoding for multilevel modulation. In *IEEE Global Telecommunications Conference (GLOBCOM)*, volume 1, pages 579–584, Sydney,NSW, 1998.

[247] 3GPP SA. Proposal for a study on proximity-based services. Technical report, SP-110638, 2011.

[248] M. Salah, N. A. Ali, A.-E. Taha, and H. Hassanein. Evaluating uplink schedulers in LTE in mixed traffic environments. In *Proc. IEEE Int Communications (ICC) Conf*, pages 1–5, 2011.

[249] A. B. Saleh, O. Bulakci, S. Redana, B. Raaf, and J. Hamalainen. Evaluating the energy efficiency of LTE-advanced relay and picocell deployments. In *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, pages 2335–2340, 2012.

[250] S.B. Wicker. *Error control systems for digital communication and storage*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1994.

[251] D. C. Schultz, R. Pabst, and B. Walke. Analytical estimation of packet delays in relay-based imt-advanced networks. In *Proc. IEEE Vehicular Technology Conf. VTC Spring 2008*, pages 2411–2415, 2008.

[252] H. Schwarz, D. Marpe, 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*, 17:1103–1120, September 2007.

[253] S. Sen, J. Rexford, and D. Towsley. Proxy prefix caching for multimedia streams. In *INFOCOM '99. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1310 –1319 vol.3, mar 1999.

[254] Shu Shi, Won J. Jeon, Klara Nahrstedt, and Roy H. Campbell. Real-time remote rendering of 3d video for mobile devices. In *Proceedings of the 17th ACM international conference on Multimedia*, MM '09, pages 391–400, New York, NY, USA, 2009. ACM.

[255] Shu Shi, Mahsa Kamali, Klara Nahrstedt, John C. Hart, and Roy H. Campbell. A high-quality low-delay remote rendering system for 3d video. In *Proceedings of the international conference on Multimedia*, MM '10, pages 601–610, New York, NY, USA, 2010. ACM.

[256] V. Singh, T. Karkkainen, J. Ott, S. Ahsan, and L. Eggert. Multipath RTP (MPRTP). IETF Internet-Draft: draft-singh-avtcore-mprtp-05, July 2012.

[257] S.L. Goff, A. Glavieux and C. Berrou. Turbo-codes and high spectral efficiency modulation. In *IEEE International Conference on Communications*, pages 645–649, New Orleans, LA, 1994.

[258] S.M. Aji and R.J. McEliece. The generalized distributive law. *IEEE Transactions on Information Theory*, 46(2):325–343, March 2000.

[259] S.M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, 1998.

[260] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand. 3d video and free viewpoint video - technologies, applications and mpeg standards. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 2161–2164, july 2006.

[261] Iraj Sodagar. Industry and standards: The mpeg-dash standard for multimedia streaming over the internet. *IEEE Computer Society*, pages 62–67, 2011.

[262] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier. Hierarchical Mobile IPv6 (HMIPv6) Mobility Management. RFC 5380 (Proposed Standard), October 2008.

[263] Mei Song, Jianwen Huang, Ruijun Feng, and Junde Song. A distributed dynamic mobility management strategy for mobile ip networks. In *ITS Telecommunications Proceedings, 2006 6th International Conference on*, pages 1045 –1048, june 2006.

[264] R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang, and V. Paxson. Stream control transmission protocol. IETF Request for Comments: 2960, October 2000.

[265] T. Sutinen and J. Huusko. Mac layer qos architecture for optimized and fair transmission of scalable video. In *IEEE Consumer Communications and Networking Conference*, Las Vegas, Nevada, USA, 2011.

[266] T. Sutinen and H. Rivas. Cross-layer assisted network interface selection for multi-interface video streaming. In *The 20th International Conference on Computer Communications and Networks (ICCCN)*, pages 1–6, Maui, Hawaii, USA, August 2011.

[267] Tiia Sutinen, Janne Vehkaperä, Esa Piri, and Mikko Uitto. Towards ubiquitous video services through scalable video coding and cross-layer optimization. *EURASIP Journal on Wireless Communications and Networking*, 2012, 2012.

[268] S.X. Ng, O.R. Alamri, Y. Li, J. Kliewer and L. Hanzo. Near-capacity turbo trellis coded modulation design based on EXIT charts and union bounds. *IEEE Transactions on Communications*, 56(12):2030–2039, December 2008.

[269] S.X. Ng, T.H. Liew, L-L. Yang and L. Hanzo. Comparative study of TCM, TTCM, BICM and BICM-ID schemes. In *IEEE Vehicular Technology Conference*, pages 2450–2454, Rhodes, Greece, May 2001.

[270] T. Bui and J. Yuan. A decode and forward cooperation scheme with soft relaying in wireless communication. In *IEEE 8th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, Helsinki, June 2007.

[271] T. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Transactions on Information Theory*, 25(5):572–584, 1979.

[272] T. Wang and G.B. Giannakis. Complex field network coding for multiuser cooperative communications. *IEEE Journal on Selected Areas in Communications*, 26(3):561–571, April 2008.

[273] M. Tanimoto, M.P. Tehrani, T. Fujii, and T. Yendo. Free-viewpoint tv. *Signal Processing Magazine, IEEE*, 28(1):67–76, jan. 2011.

[274] K. Tappayuthpijarn, G. Liebl, T. Stockhammer, and E. Steinbach. Adaptive video streaming over a mobile network with TCP-friendly rate control. In *The 2009 International Conference on Wireless Communications and Mobile Computing (IWCMC'09)*, pages 1325–1329, Leipzig, Germany, 2009.

[275] T.E. Hunter and A. Nosratinia. Cooperation diversity through coding. In *IEEE International Symposium on Information Theory*, 2002.

[276] T.E. Hunter and A. Nosratinia. Diversity through coded cooperation. *IEEE Transactions on Wireless Communications*, 5(2):283–289, February 2006.

[277] T.E. Hunter, S. Sanayei and A. Nosratinia. Outage analysis of coded cooperation. *IEEE Transactions on Information Theory*, 52(2):375–391, February 2006.

[278] R. Tewari, H. M. Vin, A. Dan, and D. Sitaram. Resource-based caching for web servers. In *Proc. SPIE/ACM Conference on Multimedia Computing and Networking*, January 1998.

[279] T.K. Moon. *Error correction coding: Mathematical methods and algorithms*. John Wiley Sons, Inc., 1st edition edition, 2005.

[280] G. Tsirtsis, H. Soliman, N. Montavont, G. Giaretta, and K. Kuladinithi. Flow bindings in mobile IPv6 and network mobility (NEMO) basic support. IETF Request for Comments: 6089, January 2011. (Updates: 5648).

[281] V. Goppa. Codes on algebraic curves. *Soviet Math. Dokl.*, 24:170–172, 1981.

[282] V. Guruswami and M. Sudan. Improved decoding of Reed-Solomon and algebraic-geometric codes. In *IEEE Symposium on Foundations of Computer Science*, pages 28–39, 1998.

[283] V. Sidorenko, G. Markarian and B. Honary. Minimal trellis design for linear codes based on the Shannon product. *IEEE Transactions on Information Theory*, 42(6):2048–2053, November 1996.

[284] V. Tarokh, N. Seshadri and A. Calderbank. Space-time codes for high data rate wireless communications: Performance criterion and code construction. In *Proc IEEE International Conference on Communications '97*, pages 299–303, Montreal, Canada, 1997.

[285] V.D. Goppa. *Geometry and codes*. Dordrecht [Netherlands] ; Boston : Kluwer Academic Publishers, 1988.

[286] W. Koch and A. Baier. Optimum and sub-optimum detection of coded data disturbed by time-varying intersymbol interference. In *IEEE Global Telecommunications Conference, GLOBECOM '90.*, pages 1679–1684, San Diego, CA, USA, December 1990.

[287] W. Zhang, Y. Li, X.-G. Xia, P.C. Ching and K.B. Letaief. Distributed space-frequency coding for cooperative diversity in broadband wireless Ad Hoc networks. *IEEE Transactions on Wireless Communications*, 7(3):995–1003, March 2008.

[288] R. (Ed.) Wakikawa. Multiple care-of addresses registration. IETF Request for Comments: 5648, October 2009.

[289] Ryuji Wakikawa, Guillaume Valadon, and Jun Murai. Migrating home agents towards internet-scale mobility deployments. In *Proceedings of the 2006 ACM CoNEXT conference*, CoNEXT '06, pages 10:1–10:10, New York, NY, USA, 2006. ACM.

[290] Xin Wang and G. B. Giannakis. Resource allocation for wireless multiuser OFDM networks. *IEEE Trans. on Information Theory*, 57(7):4359–4372, 2011.

[291] Y. Wang, Z.-L. Zhang, D. Du, and D. Su. A network conscious approach to end-to-end video delivery over wide area networks using proxy servers. In *Proc. IEEE INFOCOM*, Apri 1998.

[292] Yu Wang, Lap-Pui Chau, and Kim-Hui Yap. Joint rate allocation for multiprogram video coding using fgs. 20(6):829–837, 2010.

[293] S. Wenger, Y. Wang, and M. Hannuksela. RTP payload format for H.264/SVC scalable video coding. *Journal of Zhejiang University SCIENCE A*, 7(5):657–667, 2006.

[294] S. Wenger, Y.-K. Wang, T. Schierl, and A. Eleftheriadis. RTP payload format for scalable video coding. IETF Request for Comments: 6190, 2011.

[295] Winner+. Innovative concepts in peer-to-peer and network coding. Technical report, D1.3, 2009.

[296] I. C. Wong and B. L. Evans. Adaptive downlink OFDMA resource allocation. In *Proc. 42nd Asilomar Conf. Signals, Systems and Computers*, pages 2203–2207, 2008.

[297] I. C. Wong, O. Oteri, and W. Mccoy. Optimal resource allocation in uplink SC-FDMA systems. *IEEE Transactions on Wireless Communications*, 8(5):2161–2165, 2009.

[298] W.R. Carson, I.J. Wassell and M. Rodrigues. Optimal 8PSK mappings for BICM-ID over quasi-static fading channels. In *5th International Symposium on Turbo Codes and Related Topics*, pages 128–133, September 2008.

[299] K. L. Wu, P. S. Yu, and J. L. Wolf. Segment-based proxy caching of multimedia streams. In *Proceedings of World Wide Web Conference*, 2001. Hong Kong.

[300] W.W. Peterson. Encoding and error correction procedures for the Bose-Chaudhuri codes. *IRE Trans. Inform. Theory*, IT-6:459–470, September 1960.

[301] W.W. Peterson. *Error-correcting codes*. Cambridge, MA and New York MIT Press and Wiley, 1961.

[302] W.W. Peterson and E.J. Jr. Weldon. *Error-correcting codes*. MIT Press: Cambridge, Mass., 2nd edition edition, 1972.

[303] Tang X., Zhang F., and Chanson S. T. Streaming media caching algorithms for transcoding proxies. In *Proceedings of 31st International Conference on Parallel Processing (ICPP'02)*, 2002.

[304] X. Bao and J. Li. Efficient message relaying for wireless user cooperation: Decode-Amplify-Forward (DAF) and Hybrid DAF and coded-cooperation. *IEEE Transactions on Wireless Communications*, 6(11):3975–3984, November 2007.

[305] X. Li and J.A. Ritcey. Bit-interleaved coded modulation with iterative decoding. *IEEE Communications Letters*, 1(6):169–171, November 1997.

[306] X. Li and J.A. Ritcey. Trellis-coded modulation with bit interleaving and iterative decoding. *IEEE Journal on Selected Areas in Communications*, 17(4):715–724, April 1999.

[307] X. Wang and H.V. Poor. Iterative (turbo) soft interference cancellation and decoding for coded CDMA. *IEEE Transactions on Communications*, 47(7):1046–1061, July 1999.

[308] Y. Li. Distributed coding for cooperative wireless networks: An overview and recent advances. *IEEE Communications Magazine*, 47(8):71–77, August 2009.

[309] Y. Li, B. Vucetic, T.F. Wong and M. Dohler. Distributed turbo coding with soft information relaying in multihop relay networks. *IEEE Journal on Selected Areas in Communications*, 24(11):2040–2050, November 2006.

[310] Y. Sugiyama, M. Kasahara, S. Hirasawa and T. Namekawa. A method for solving key equation for Goppa codes. *Inf. and Control*, 27:87–99, 1975.

[311] E. Yaacoub and Z. Dawy. A game theoretical formulation for proportional fairness in LTE uplink scheduling. In *Proc. IEEE Wireless Communications and Networking Conf. WCNC 2009*, pages 1–5, 2009.

[312] Elias Yaacoub, Ahmad M. El-Hajj, and Zaher Dawy. Uplink OFDMA resource allocation with discrete rates: optimal solution and suboptimal implementation. *Transactions on Emerging Telecommunications Technologies*, 23(2):148–162, 2012.

[313] Y. Yang, J. Xu H. Hu, and G. Mao. Relay technologies for wimax and lte-advanced mobile systems. *IEEE Communications Magazine*, pages 100–105, Oct. 2009.

[314] Su Yi and Ming Lei. Backhaul resource allocation in LTE-advanced relaying systems. In *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, pages 1207–1211, 2012.

[315] Yiwei Yu, E. Dutkiewicz, Xiaojing Huang, and M. Mueck. Inter-cell interference coordination for type I relay networks in LTE systems. In *Proc. IEEE Global Telecommunications Conf. (GLOBECOM 2011)*, pages 1–6, 2011.

[316] Y. Yuda, A. Iwata, and D. Imamura. Interference mitigation using coordinated backhaul timing allocation for LTE-advanced relay systems. In *Proc. IEEE Int Communications (ICC) Conf*, pages 1–5, 2011.

[317] Z. Zhang and T.M. Duman. Capacity-approaching turbo coding for half-duplex relaying. *IEEE Transactions on Communications*, 55(10):1895–1906, October 2007.

[318] Xi Min Zhang, A. Vetro, Y. Q. Shi, and Huifang Sun. Constant quality constrained rate allocation for fgs-coded video. 13(2):121–130, 2003.

[319] Z.-L. Zhang, Y. Wang, D. Du, and D. Su. A proxyserver-based approach to end-to-end video delivery over wide-area networks. In *IEEE/ACM Transactions on Networking*, volume 8, pages 429–442, 2000.