**Remote Collaborative Real-Time Multimedia Experience over the Future Internet**

**ROMEO**

**Grant Agreement Number: 287896**

# D2.1

# Report on the mission scenarios, service definition and requirements, review of the state-of-the-art in 3D media services

| Document description | |
|---|---|
| Name of document | Report on the mission scenarios, service definition and requirements, review of the state-of-the-art in 3D media services |
| Abstract | This document defines the reference scenarios and the requirements in line with ROMEO objectives as well as includes an overview of the state-of-the-art technologies in 3D media delivery services. |
| Document identifier | D2.1 |
| Document class | Deliverable |
| Version | 1.0 |
| Author(s) | A. Akman, E. Ç. Öztürk, S. Ö. Pelvan, S. Çiftçi (TTA)<br>H. Ibl, J. Lauterjung (R&S)<br>C. Hartmann, M. Meier, M. Weitnauer, N. Just, D. Driesnack, P. tho Pesch, M. Laabs (IRT)<br>M. Albano, E. Logota (IT)<br>I. Politis, K. Birkos, A. Lykourgiotis, A. Kordelas (UP)<br>E. Ekmekcioglu, V. De Silva, I. Elfitri, C. Kim, S. Dogan, A. Kondoz (US)<br>X. Shi, B. Evans (MULSYS)<br>S. Rouco (TID)<br>B. Demirtaş (ARC)<br>O. Guye, J. Maillard (VITEC)<br>D. Doyen (TEC)<br>N. Daskalov (MMS) |
| QAT team | All partners (same as above) |
| Date of creation | 14-Nov-2011 |
| Date of last modification | 16-Dec-2011 |
| Status | Final |
| Destination | European Commission |
| WP number | WP2 |
| Dissemination Level | Public |
| Deliverable Nature | Report |

# TABLE OF CONTENTS

# LIST OF FIGURES

## LIST OF TABLES

# 1 INTRODUCTION

## 1.1 Purpose of the document

This document (D2.1 Report on mission scenarios, service definition and requirements, review of the state-of-the-art in 3D media services) aims to serve as a basis to ROMEO use cases and initial system architecture, which will have a big impact on the design and development of system components in Research and Technological Development (RTD) work packages (WP3-6). To achieve this purpose, the document defines the reference scenarios, specifies a minimum set of service requirements and gives a summary of the state of the art technologies on 3D media platforms.

## 1.2 Scope of the work

This document covers the reference scenarios to be accomplished, the service requirements to be met and the state-of-the-art technologies to be interested in by ROMEO system.

Three different scenarios with the proposed use cases highlighting the ROMEO objectives are defined in the document to be considered for the demonstration as well as performance evaluation purposes in WP7. The first scenario defines the interaction of mobile and fixed users with the proposed system as well as the behaviour of the system against changing network conditions. The focus of the second scenario is to provide a description of how collaborative users are handled and how user generated data is shared between them. Finally, the last scenario describes the way that QoS is maintained over a fixed area network for ROMEO users.

Based on the reference scenarios and the ROMEO objectives, ROMEO service requirements are defined. In addition to scenario specific requirements, terminal, media coding and networking requirements are specified in an end-to-end aspect, stretching from user terminals all the way up to content providers. These requirements will be used to define the initial architecture of the ROMEO system.

In the last part, the document gives an overview of the state-of-the-art technologies to be considered and enhanced within the scope of RTD work packages (WP3-6).

## 1.3 Objectives and achievements

This document achieves to form a consensus on the reference scenarios and requirements, which will be used to define the top to down system architecture of the ROMEO system. The reference scenarios will be used for the demonstration and performance evaluation of the system.

## 1.4 Structure of the document

This deliverable is structured as follows:

- Section 2 defines the reference scenarios for the project.
- Section 3 specifies the initial set of service requirements for project.
- Section 4 gives the state-of-the-art on 3D media systems.

# 2 REFERENCE SCENARIOS

In this part, three different reference scenarios are going to be defined. These will be then used as a base for the requirements and architecture as well as the demonstration of the platform.

## 2.1 Reference scenario 1: Consumption of 3D audio/video over hybrid delivery paths

In this scenario, ROMEO is delivering a 3D live stream using either only one of the available networks or a combination of them (i.e., Digital Video Broadcast (DVB) and Peer-to-Peer (P2P) over Internet Protocol (IP)) at a given time to maximize the Quality of Experience (QoE) offered to the user, including adaptation in response to changing network conditions, triggering different media delivery options (adjusting number of views, quality of views, etc). Within this scenario, user connection to P2P overlay, peer churn, mobility, roaming and handover use cases are provided.

*ROMEO objective mapping[1]:*

*Objective 1:* Support for live and collaborative real-time 3D media delivery to fixed and mobile users by simultaneous use of broadcast and IP networks

*Objective 2:* New architectures to jointly optimize multi-view 3D video and spatial audio over broadcast and QoE aware P2P streaming networks serving fixed and mobile users

**Use cases for reference scenario 1**:

*Starting condition:*

User wants to join a live stream.

*Use case 1 – User connection and live stream reception with changing network conditions:*

1. User switches on the platform and begins to receive the broadcast.
2. User does not want to receive any additional views. (Scenario ends here)
3. User wants to receive additional views via the network so s/he sends his/her credentials.
4. The main server checks his credentials and:
    a. Either rejects the user. Scenario ends. Or
    b. Confirms the user.
5. User sends his/her network and device capabilities to the Topology Management Unit.
6. The Topology Management Unit[2] tries to insert this user into the P2P overlay:

[1] Mapping between the ROMEO objectives stated in Section 1.1 of the ROMEO Description of Work (DoW).

[2] The Topology Management Unit is responsible for forming and maintaining the P2P overlay, and it will be investigated whether it will be centralised (in the main server), distributed (among peers) or a hybrid solution.

a. Without meeting the minimum requirements to receive the live stream over P2P, user receives the main 3D stereoscopic and 5.1 audio contents via DVB only, and waits for the network conditions to become sufficient to connect to the P2P overlay for receiving additional views. Scenario ends.

b. With sufficient network conditions, the Topology Management Unit locates the optimum place for the user to be inserted in the P2P overlay.

7. User starts to receive the main stereoscopic 3D view from DVB and additional views as well as object-based audio from the P2P network.

8. User terminal monitors network conditions periodically (period to be determined later) and takes the following steps against changing conditions:

    a. With deteriorating network conditions, the steps are going to be taken in the following order (The thresholds will be determined later)).

        i. Quality of the views received over the P2P network is decreased without having a noticeable effect on QoE (by discarding the enhancement layer and receiving the base layer only).

        ii. Some views will be dropped. (View dropping strategy will be determined later).

        iii. When it is not possible to receive the stream from the P2P network, user continues to receive the main 3D stereoscopic view and 5.1 audio from DVB only.

    b. With improving network conditions, the reverse steps defined in the deteriorating case will be executed.

*Use case 2 – Peer disconnection:*

1. User disconnects from P2P overlay while live stream continues:

    a. With acknowledgement:

        i. User sends a disconnect request to the Topology Management Unit.

        ii. The Topology Management Unit tries to reform the overlay so that QoE of the other users is not affected with this disconnection.

        iii. The Topology Management Unit maintains the topology (connections of the served peers of the peer requesting to disconnect) and sends disconnect acknowledgement so that user can disconnect in a controllable manner.

    b. Without acknowledgement:

        i. User suddenly disconnects from the P2P overlay.

        ii. Peers served by the disconnected peer try to receive the enhancement layer or additional views via another path where possible.

        iii. The P2P topology is reformed with remaining users.

2. The disconnected user wants to re-connect to the P2P overlay.

3. The Topology Management Unit treats the user as a new user.

*Mobile user specific use cases:*

Remaining use cases consider mobile users only.

*Use case 3 – Intra-system handover:*

*Starting condition:*

The mobile user wants to join a live stream and is connected to a 4$^{th}$ Generation (4G)/Long Term Evolution (LTE) access network. An enhanced Media Aware Network Element (MANE), which is an intelligent router with the ability to adapt scalable video content according to network conditions, monitors all access networks in the vicinity of the mobile user and stores network statistics to a central database.

1. The mobile user requests all the 3D video views from the P2P overlay to be able to select the view to watch.
2. Handover functionality within MANE calculates the threshold values of physical and network layer parameters and decides whether to perform:
    a. Video adaptation is performed (e.g. layer dropping, packet scheduling, quantization parameter adaptation, etc.) based on the prevailing 4G/LTE access network conditions and user preferences.
    b. Mobility control is performed to initiate a horizontal handover procedure.
3. The mobile user terminal periodically updates the MANE about QoE measurements, physical network layer statistics (i.e. SNR, packet losses, etc.) and requested view-points and refreshes its client profile.

*Use case 4 – Multi-homing transmission:*

*Starting condition:*

During an active live streaming session through 4G/LTE, the mobile user enters a Wireless Fidelity (Wi-Fi) hotspot that is already monitored by MANE as in the previous scenario.

1. The mobile user terminal connects to the Wi-Fi and updates its client profile to MANE.
2. MANE, taking into account the current QoE feedback of the mobile/portable user, determines the number of views (portable users can support up to 2 views simultaneously) that will be delivered to the mobile user and divides these views into two categories:
    a. Views that will be sent over the 4G/LTE interface.
    b. Views that will be sent over the Wi-Fi interface.
3. MANE establishes connections to LTE and Wi-Fi access networks.

*Use case 5 – Inter-system handover:*

*Starting condition:*

During an active live streaming session, the mobile user is connected to a Wi-Fi hotspot and moves into an LTE coverage area.

1. The mobile user sends to MANE a QoE feedback of degradation to the physical and/or network and/or application statistics.
2. MANE decides that a vertical handover (IEEE802.21 compatible) must be initiated and performs the following steps:
    a. Determination of the views (that the user can select from) that will be available to the mobile/portable user through the new access network.
    b. Triggering of the mobility protocol (Mobile IP) to execute the handover.
3. LTE connection is established and the Wi-Fi connection is released.

## 2.2 Reference scenario 2: Collaborating group and sharing user generated content

In this scenario, ROMEO is providing a real-time audio-visual communication channel to connect collaborating users while they are receiving the live stream via their DVB and/or P2P connections. In this scope, users can initiate a collaborating group and specify the members of the group. Users can generate or use existing content and share it with the members of the collaborating group for further social interaction. The three use cases below provide further details for this scenario.

*ROMEO objective mapping:*

*Objective 3:* Provide a real-time audio-visual communication overlay to bring the remote collaborators together for joint interaction and enjoyment.

**Use cases for Reference Scenario 2:**

*Starting condition:*

A group of users wants to form an Audio/Video (A/V) collaborating group during a live stream. Each user joins the P2P overlay as in Scenario 1.

*Use case 1 – Formation of the collaborating group:*

1. One of the users in the group sends a request to the main server as initiator (admin) of the group to initiate a collaborating group with defined access rights for other users that will join this group later.
2. After users join the collaborating group according to the defined access rights, the admin user starts collaboration in the group.
3. At this point, the main server calculates the minimum requirements for new peers to join the collaborating group based on the connected users' conditions.
4. An A/V link will be formed between the collaborating users to let them communicate in real time through video conferencing infrastructure while consuming the live stream in a synchronised way.
5. In the case of a peer disconnecting and re-joining /a new peer joining request:
    a. If the peer satisfies the minimum requirements calculated at the initiation of the group, it can join the collaborating group by consuming 3D content through both DVB and P2P networks.
    b. If the peer does not satisfy the minimum requirements determined by the main server, it can still join the collaborating group by getting the main 3D stereoscopic view through DVB only.

*Use case 2 – User generated content upload:*

1. User creates, encodes, digitally signs and uploads his/her content to the main server.
2. User who provides the content gives certain access rights to certain users in the collaborating group while uploading.

3. The main server authenticates the user generated content using digital signature methods.
4. The authenticated content is then pre-processed; metadata is embedded to it and packet is digitally signed with the same key as the user by the main server to give access rights to the users only provided with the user key.
5. The main server indexes the content and stores it in a repository with its access rights.

*Use case 3 – User generated content download:*

1. User searches and discovers the authenticated content in the repository of the main server.
2. User sends a request to the main server to download this content.
3. The main server checks the user's access rights and:
    a. Rejects the user request. Scenario ends. Or
    b. Lets the user download the content.
4. User checks the content integrity after downloading it.

## 2.3 Reference scenario 3: P2P quality of service (QoS) over fixed access networks

In this scenario, more than one user is connected to the Internet using the same connection simultaneously. One of the users joins a live stream using ROMEO, and thus other users' connection priorities are adjusted by the Virtual Terminal Equipment accordingly to allow for meeting the expected QoE for the ROMEO user. Related use case is given below.

*ROMEO objective mapping:*

*Objective 4:* Terminal equipment virtualisation for dynamic content adaptation

**Use cases for reference scenario 3:**

*Use case 1 – QoS adjustment:*

*Starting condition:*

User A and User B are in the same home network. User A is downloading a large file consuming most of the bandwidth contracted to the operator (a graph of the download statistics is displayed on a Personal Computer (PC) monitor).

1. User B joins a live stream as described in Scenario 1.
2. The QoE of the live stream visualization is not at the expected level, and hence 2 different actions can be performed:
    a. The access network has already identified the P2P flow streams and Broadband Remote Access Server (BRAS) of the access network reacts to the new conditions by providing the required bandwidth for the P2P access at the expense of the background traffic mainly allocated for file downloading and other services.
    b. The end user makes use of the ROMEO QoS service in order to be able to watch the 3D live streaming content at the expected level of QoE. This action implies that the user connects to the ROMEO web portal and prioritizes the live stream over the background traffic, so as to provide the required bandwidth to the P2P flows.
3. The result is that background traffic bandwidth is shrunk until live streaming meets the QoE requirements.
4. The live stream is displayed meeting the required QoE regardless the bandwidth usage of any other application running in the home.

# 3 SERVICE REQUIREMENTS

This part of the document sets the initial requirements which will then be used to design the architecture of the platform.

## 3.1 Overall system and scenario requirements

ROMEO will bring new 3D media (3D multi-view video and spatial audio) to European citizens at home as well as on the move. In order to deliver the high bandwidth, high quality 3D media to mobile and fixed users with guaranteed minimum QoE for all users, ROMEO will combine the DVB technology with the P2P Internet technology. In addition to this, ROMEO will focus on the delivery of live 3D media to a group of users all at the same time to enable them to collaborate, and jointly enjoy it. With these motivations, the following overall system and scenario requirements are determined within the scope of the project:

- Fixed and mobile users shall be able to use broadcast and IP networks to receive live 3D media.
- Fixed and mobile users shall receive live 3D media with an unnoticeable delay.
- ROMEO shall provide a real-time audio-visual communication overlay to bring the remote collaborators together for joint interaction and enjoyment.
- 3D live stream and collaboration between users shall be synchronised to have proper interaction between users according to stream.
- Collaborating users shall be able to upload and share their own generated content to the IP network main server. The content generator shall be able to determine the access rights of other collaborating users.

### 3.1.1 Requirements for multi-view media delivery services

In ROMEO, 3D multi-view contents are delivered to home fixed-users and mobile users via DVB-T2/T2-Lite/NGH, wireless networks including LTE, 3G (3$^{rd}$ Generation) and Wi-Fi, as well as the IP-based Internet. Generally speaking, the DVB stream, which carries the main stereoscopic 3D video and the 5.1 audio, encoded using standardised down-mixing and/or compression codec, is always available to fixed and mobile users. All other views and its enhancement data, in a 3D multi-view, are transmitted via mobile wireless networks and P2P networks. This is also applicable for audio objects, including the corresponding object description, which will be transmitted as an enhancement layer via P2P. The idea behind this structure is to use the DVB for providing the main stereoscopic view and P2P for providing additional contents, e.g., additional views. As such, the requirements for multi-view media delivery service are as following:

- The main view's Advanced Video Coding (AVC) compatible base layer with sufficient broadcast quality, which is stereoscopic 3D video and 5.1 channel audio, shall always be received through DVB if available.
- ROMEO system shall deliver at most 1 stereoscopic view to mobile terminals and all available camera views (if bandwidth suffices) to compatible portable and fixed terminals (there may be device specific limitations on the number of simultaneously decodable camera views for portable (e.g., at most 2 camera views) and fixed (e.g., at most 4 camera stereoscopic views) terminals).
- Furthermore the object-based scene description could possibly be delivered via DVB for further development approaches.
- P2P shall carry an object-based scene description audio stream and (if the DVB stream is not available) the 5.1 channel audio stream.

- All other views and their enhancement data for 3D multi-view rendering purposes shall be sent through the P2P network. The multi-view contents could be delivered to users via different networks depending on the user terminal capabilities and network conditions:
    - Fixed terminals shall be served via high-speed broadband.
    - Mobile terminals shall be served via Wi-Fi hotspot and/or mobile networks such as 3G and LTE.
    - If a user device has multiple interfaces and all of the interfaces are able to retrieve data from the P2P network, the device shall use interfaces in the following order for receiving data:
        - DVB-T2/T2-lite/NGH,
        - One of the following devices, in order of priority:
            - Fixed broadband connection,
            - Wi-Fi interface connection to the hotspot,
            - 3G or LTE mobile connection.
- The camera views comprising the main stereoscopic 3D pair shall also be encoded using SVC simulcast (both views using scalable high profile at Level 4 or over) with two spatial resolution layers (HD as full spatial resolution and preferably SD as lower resolution) and quality enhancement layers to be stored in the seed server and transmitted over P2P.
- The rest of the cameras, and the depth map videos (including the depth of the main stereoscopic view), shall be encoded in the same way as in the previous requirement.
- The SD frame compatible coded version (AVC compliant) can also be included in the P2P streaming server, considering peers with mobile terminals that are outside the DVB coverage and that cannot decode two 1920x1080p25, or two 720x576p25 streams simultaneously, which are normally streamed to them over P2P.
- Audio and video streams shall be synchronised for all views (play-out synchronisation).

### 3.1.2 Requirements for broadcast media delivery networks: DVB-T2/T2-Lite/NGH

For the overall system and the proof-of-concept implementations, two different broadcast links are required. The first one is a DVB-T2 link to fixed and portable terminals; and the second one is a DVB-T2 Lite/ NGH link to mobile terminals.

- All terminals, either fixed or mobile, shall be able to receive DVB streams.
- DVB streams shall not be encrypted.
- DVB-T2 Lite/NGH standard shall be implemented for mobile users.
- The broadcast link to fixed and portable terminals shall provide the main 3D stereoscopic view.
- The 3D-stereoscopic video signal to fixed and portable receivers shall be encoded as a DVB-compliant signal.
- The main 3D stereoscopic view for fixed and portable terminals shall be transmitted in one Physical Layer Pipe (PLP) of the DVB-T2 signal.
- The bit rate of the main 3D stereoscopic view for fixed and portable terminals shall be limited to 7 to 8 Mbps (The complete DVB-T2 multiplex typically carries 3 to 4 different 3D-stereoscopic video signals for fixed and portable terminals with an overall bit rate of up to 30 Mbps.)
- The DVB-T2 signal to fixed and portable terminals shall be compliant to standard ETSI EN 302 755 (DVB-T2 Base).

- The broadcast link to mobile terminals shall provide the main view either as a traditional 2D signal or a 3D-stereoscopic video signal.
- The 2D view for mobile terminals shall be in HD/SD depending on the hardware capabilities of the terminal. The main 3D stereoscopic view to mobile terminals is encoded as an SD-frame compatible (AVC compliant) with standard-compliant signalling. (To maintain the same bit rate as for a 2D SD signal, the horizontal or vertical resolution of the 3D SD signal is reduced by 50 %.)
- The bit rate of the 2D HD/SD or 3D-stereoscopic SD view for mobile terminals shall be limited to 4 Mbps (Annex I.10 “T2-Lite PLP data rate limitations” in [1]).
- The DVB-T2 signal to mobile terminals is compliant to [1] (DVB-T2 Lite).
- The output frequency of the test transmitter shall be adjustable to any channel in the TV Ultra High Frequency (UHF) band.
- The input signal to the test receiver platform shall be a DVB-T2 Base or DVB-T2 Lite signal, which can be transported in any channel in the TV UHF band.
- The output signal of the test receiver platform shall be fed to a display via a High-Definition Multimedia Interface (HDMI) carrying video and audio for the selected programme.
- The test receiver platform shall measure a multitude of Quality of Service (QoS) parameters of the received DVB-T2 signal. A sub-set of these parameters can be selected for monitoring certain QoS aspects permanently. The parameters cover the physical layer (RF level, RF frequency offset, etc.), the I/Q domain (Modulation Error Ratio (MER), amplitude imbalance, phase error, etc.) and the Transport Stream domain (packet loss rate, missing packets, CRC checks on certain packets, etc.).
- The test receiver platform shall be able to combine the different signals received via the DVB-T2 and P2P links in such a way that differences in latency or delay are compensated.
- The test receiver platform shall be able to select either the DVB-T2 or the P2P signal as a fall-back solution for displaying in the case that the reception quality of one of the signals is not sufficient.

#### 3.1.3 Requirements for network architecture through a virtual router with QoS

- Each user shall own Optical Network Terminal (ONT) equipment.
- Access node shall be an Optical Line Terminal (OLT).
- Virtual routers shall be implemented at BRAS Node.
- Authentication, Authorization and Accounting (AAA) server shall be used to authenticate subscribers.
- Dynamic Host Configuration Protocol (DHCP) server shall be used for Virtual Routers Public IP assignation.
- A dedicated web server shall offer the Virtual Router Configuration Portal to the User.

### 3.2 Terminal requirements

#### 3.2.1 Specifications on 3D displays

##### 3.2.1.1 Fixed users’ terminals

- Fixed terminals shall have display units to cover both multi-view 3D content and stereoscopic content to present 3D video to the end-user.

#### 3.2.1.2 Mobile users' terminals

- Mobile terminal display (MTD) shall provide minimum one stereoscopic view that corresponds to two camera views (left and right).
- Mobile terminal display shall have a minimum as Wide Video Graphics Array (WVGA).
- Mobile terminal display shall be able to display both 2D and 3D content
- Mobile terminal display shall have low power operation, suitable for battery operated systems.
- Mobile terminal display shall have Thin and lightweight package.
- Mobile terminal display shall have the following interfaces: Electrical interface: MIPI DSI, Parallel or Unipro.
- Mobile terminal display shall have Backlit technology.

### 3.2.2 Specification on rendering capabilities

#### 3.2.2.1 Fixed users' terminals

- Fixed terminals shall use JavaScript Object Notation-Remote Procedure Call (JSON-RPC) for inter-module communication.

#### 3.2.2.2 Mobile users' terminals

- Mobile terminals shall use JSON-RPC for inter-module communication.
- Mobile terminals shall provide power management capabilities in order to maximize the battery life, ensuring a minimum of 2 hours of operation during High Definition Television (HDTV) decoding/playback.
- Mobile terminals shall provide a means for integrating stereoscopic Liquid Crystal Display (LCD).
- Mobile terminals shall support content streaming using the abovementioned interfaces.
- Mobile terminals shall provide sufficient processing power to decode/encode at least two H.264 Main Profile simulcast coded streams (720x1280p25) simultaneously or a H.264 Main Profile multi-view coded stream (two views).
- Mobile terminals shall provide means to implement custom type H.264 decoder as a base of multi-view or view+depth type of stereo video encoder.
- Mobile terminals shall provide means for stereo camera system connection.
- Mobile terminals shall be capable to run High Level Operating Systems (HLOS) (Android, Linux, Symbian or Windows).
- Mobile terminals shall provide at least 1GB of low-power (LP-DDR2, 3) RAM memory, 64GB non-volatile memory (e.g. NAND, e-MMC).
- Mobile terminals shall provide means for memory extension (e.g. SD, microSD).
- Mobile terminals shall support capacitive touch screen interface for user interaction.

### 3.2.3 Networking related specifications

#### 3.2.3.1 Fixed users' terminals

- Fixed device shall have DVB-T2 reception unit and high bandwidth broadband interface.
- 1Gbit/s Ethernet adaptor shall be used for P2P 3D signal to be received.
- DHCP Client shall receive IP address from the Virtualised Broadband Access Router.

#### 3.2.3.2 Mobile users' terminals

- Mobile device shall have interfaces for 3G/LTE, Wi-Fi, and DVB-T2-Lite/NGH for receiving 3D content.
- Mobile terminals shall support a variety of interfaces – USB, Wi-Fi, 3G/LTE modems, Bluetooth for connection to other devices.
- Mobile terminals shall provide sufficient processing power to execute DVB-T2-Lite/NGH stack, GUI and encoding/decoding.

### 3.2.4 Specifications on users terminals video interface

- An interface capable of supporting 3D 1080-pixel video streams (e.g. HDMI 1.4a, DVI-D) is needed on the terminal equipment.

### 3.2.5 Requirements for virtualised broadband access router

- A DHCP Server must be used to assign IPv4 addresses to user equipment.
- A DHCP Client must be used to receive public IPv4 address.
- Network Address Translation (NAT) capacity must be used to redirect traffic between two different IP address spaces.
- Port mapping capacity must be used to forward traffic to specific ports/addresses.

## 3.3 Media coding requirements

### 3.3.1 Requirements for visual formats for different display types

In the field of ROMEO, there will be different kinds of displays addressed. We can classify them in terms of views they can render. Starting from the most demanding one in term of volume of transmitted data to the least demanding, we shall have:

- Multi-view auto-stereoscopic displays (displaying 8 views or more).
- Auto-stereoscopic displays with 2 views (portable display, single user).
- Stereoscopic displays with glasses.
- 2D displays.

Addressing correctly multi-view displays means that the renderer placed before the display will receive all data required to feed the different views. It is planned to transmit a maximum of 4 views and 4 additional depth maps. The renderer will handle the data and recreate missing views to feed the display. Figure 1 illustrates the view synthesis principle using 2 views (left and right in this case) and a disparity map associated to these left and right views. It is possible to synthesize M views from the input 2 views. Then, with 4 views and the associated disparity maps, it will be possible to address the dedicated number of views required by the display.

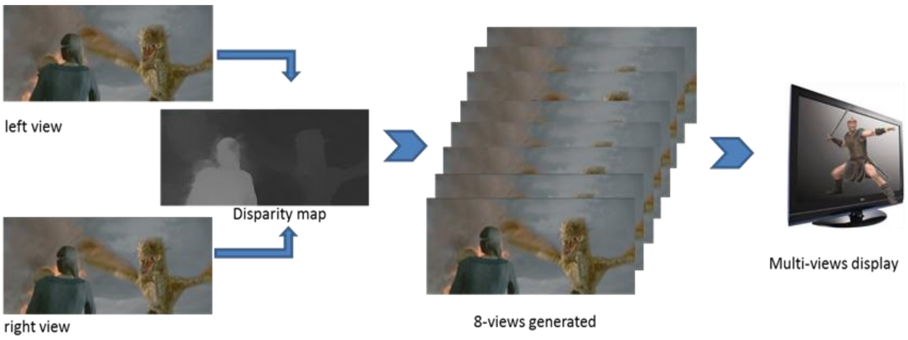


*Figure 1 - View synthesis principle using 2 views*

In terms of the requirements of a multi-view display, there are two different levels of information. If network conditions are sufficient, it is possible to receive at the end-user side 4 views and 4 associated disparity maps. Each of them has HD resolution 1080x1920. From 4 views + 4 disparity maps, it is possible to create the number of views required by the display (e.g., 8 views or 28 views) and to show a range of disparity large enough for compelling 3D pictures.

Figure 2 illustrates the 8-view positioning compared to the input 4 views. Synthesis is possible thanks to the disparity maps in between the input views.

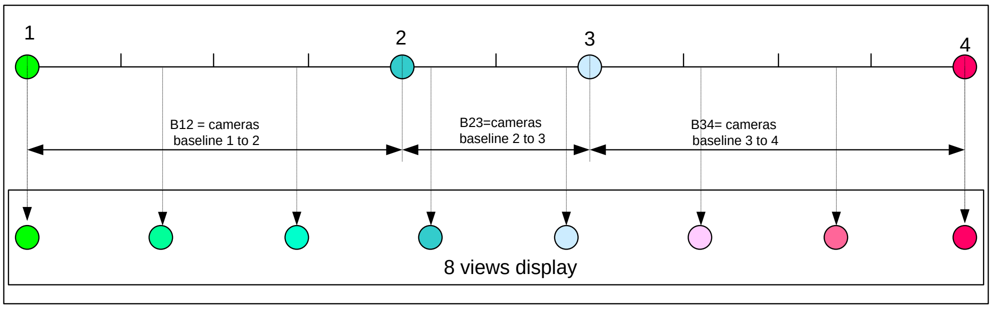


*Figure 2 - 8-view synthesis from 4 views + 4 disparity maps*

The second level that should be considered if network conditions are not optimum is to get at least 2 views and their associated disparity maps. In this case, the multi-view display will only render the corresponding range of disparity, from view 2 and view 3 (as illustrated in Figure 3).

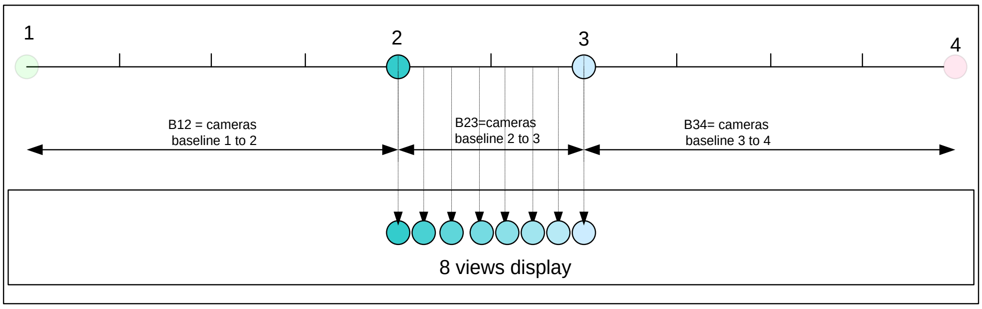


*Figure 3 - 8-view synthesis from 2 central views + 2 disparity maps*

With lower network conditions where only 2 views (without additional disparity information) are available, multi-view displays will not be considered. In this particular case, only auto-stereoscopic 2-view and stereoscopic displays can be addressed adequately.

### 3.3.2 Requirements for multi-view capturing setup

In the different ROMEO reference scenarios, it is planned to have a default 3D video format that will be used to supply a baseline 3D stereo video content over DVB.

The project is planning to demonstrate the usefulness of Scalable Multi-view Video plus Depth (SMVD) video coding for handling the delivery of 3D multi-video streams. It can be performed if, in addition to the N raw video streams, it can be also provided to the M depth maps synchronised with the raw streams as an input to a Multiple Description Scalable Multi-view Video plus Depth ((MD-) SMVD) coder.

Subsequently, the following list of requirements can be deduced that should satisfy a multi-view capturing setup:

- Small form factor cameras must be used in order to satisfy to the human mean intraocular distance for providing 3D stereo video for a parallel arrangement.
- A mounting system enabling to accept several cameras for providing 3D multi-view video with a minimum of 4 cameras.
- An adjustable mounting system enabling to tune the inter-camera distance and their orientation starting from a parallel arrangement up to convergent ones according to the kind of rendering that can be expected.
- Frame accurate synchronisation of all the cameras between themselves during video acquisition.
- Geometry and colour calibration for multi-view picture alignment and colour balancing before any digital parallel pre-processing.
- Multi-view disparity map computation before video encoding;
- Synchronisation of audio and video capture for enabling audio and video multiplexing.
- At least HD multi-view video capture for offline video processing (1080x1920p50).
- Lower video format for user generated data capturing and processing (720x1280p50).

### 3.3.3 Requirements for video codecs for multi-view content

According to the different use case scenarios considered in ROMEO, flexibility (e.g., mobile, portable and fixed terminals with different screen sizes) and extensive bandwidth scalability through resolution and quality scalability should be sought in the envisaged multi-view codec. In this way, it can be ensured that the best available representation of the encoded 3D media can be delivered to users that have different network access capabilities.

It is also required that the envisaged scalable multi-view encoder design generates side information (e.g., metadata) that can be exploited by the video decoders and the depth image based video renderers. Or, at least the envisaged codec should supply mechanisms to support incorporating extra side information in the source bit-stream. Packetisation and encapsulation stage should take care of the reliable delivery of such overhead information to most of the users. For instance, the metadata should aid the 3D video player to compensate for corruptions/losses in the coded colour and depth map streams.

Based on the envisaged P2P overlay requirements depicted in the use case scenarios, a multiple-description streaming mechanism should be coupled with such overlay, where layers within the coded 3D media bit-stream should be distributed or redundantly repeated over a set of paths, such that more critical media parts (e.g., base layers of visually salient/key views) will be ensured to reach a greater set of users.

Previous research showed that utilising optimal Forward Error Correction (FEC) schemes in the context of joint source and channel coding comprised an efficient option for error resilient video communications [2]. Optimisation in this regard stands for the exact knowledge on the channel behaviour (e.g., channel erasure rate, and hence the packet loss rate). However, it is known that P2P topologies comprise highly mismatched conditions for potential forward error correction schemes, where it is preferable to have prior knowledge on the network state to insert optimal redundancy. The evident peer instability within the P2P system may cause sudden drops in the upload and download capacities of other peers. A multiple description coding scheme does not have any ties to the network information, and hence its design is not affected by the prior consideration of the network state. Furthermore, the implementation of a scalable Multiple Description Coding (MDC) decoder, where redundant stream packets are simply discarded in cases of repetition, is simpler and requires lesser delay in the user client. Because, every received description stream can be decoded straightaway, without the necessity of decoding FEC related codewords in advance. This feature is very much in line with the real time media playback requirements of ROMEO.

The above mentioned requirements on the 3D video codec system can be summarized in the following list:

- Spatial resolution scalability (taking into account the range of targeted user terminal devices).
- Quality scalability (assuming highly dynamically changing network states of multiple users/peers).
- Viewpoint scalability (to support selective view streaming – free-viewpoint applications and reduce inter-view coding dependencies).
- Metadata extraction (offline), metadata exploitation (for increased coding performance or improved adaptation quality) and supplying means for carrying metadata information in the bit-stream.
- Suitable multiple-description generation out of the encoded 3D video bit-stream (by means of signalling) that would maximise the chances of delivering critical video components to a greater set of users.

Decoder side:

- Real time decoding of scalable coded 3D video stream (with/without enhancement layers) including metadata exploitation.

#### 3.3.4 Specifications and requirements for spatial audio codec

The basic delivery format should be 5.1 surround. It will be encoded and delivered using standardised multichannel audio down-mixing and/or compression codecs, such as MPEG Surround and AAC. As an enhancement layer, both fixed and mobile user terminals may render an object-based audio scene for spatial audio playback. Within the object based transmission there will be metadata providing general information for 3D audio reproduction systems beyond 5.1 surround as well as the actual audio signals of the individual audio objects. The rendering system for a specific 3D audio reproduction system can use this object-based audio scene description to play the 3D audio scene.

The DVB transmission will contain the basic channel-based 5.1 audio stream. For further development approaches, the audio object scene description could also be delivered via DVB, if investigations are successful regarding this point. The P2P transmission however, will be object based. Optionally there can be additional audio and metadata objects to enhance the overall spatial impression delivered via P2P.

#### 3.3.4.1 MPEG surround

The first choice to encode 5.1 audio signals will be the MPEG Surround standard. However, MPEG AAC standard will also be investigated and considered as an option provided that it can deliver higher quality audio reproduction than MPEG Surround. A block diagram of MPEG Surround (MPS) is shown in Figure 4. Multiple audio signals are typically down-mixed as mono or stereo signals which are encoded further by an existing audio coder. Channel level differences (CLDs), inter-channel coherences (ICCs), and channel prediction coefficients (CPCs) are basically extracted as the spatial parameters. The residual signal is computed and transmitted for high quality audio reconstruction while in low bit rate implementation it can be replaced by synthetic signal created by decorrelator. The spatial parameters and the residual signal are then transmitted as side information.

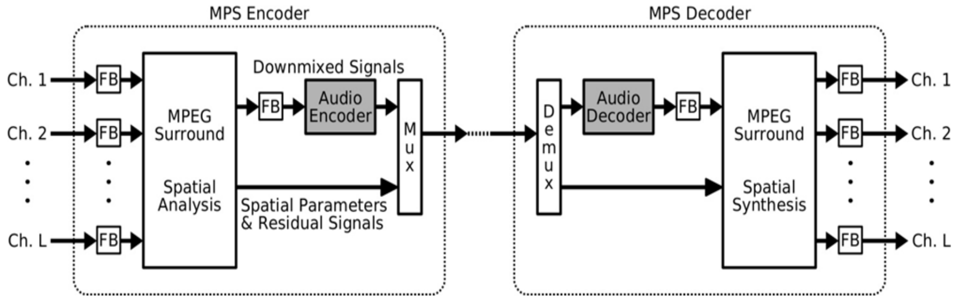


*Figure 4 - Block Diagram of MPEG Surround (MPS)*

The requirements and specifications for MPS implementation is detailed as follows:

Encoder side:

- A 5.1 audio system shall be used as input which means that the number of channels, L = 6.
- Audio signal in each channel shall be sampled and then decomposed by analysis filter bank.
- When MPEG Surround is used, sub band signals shall be grouped as 20 parameter bands sharing common spatial parameters (CLD and ICC).
- MPEG Advanced Audio Coding (AAC) shall be adopted as audio encoder for further encoding down mixed signals.
- The transmission bit rate shall be in the range of 300 – 600 kb/s.

Decoder side:

- Spatial audio shall be rendered using 5.1 audio streams for fixed users. For the mobile users, a binaural signal will be rendered if the mobile device fulfils the necessary requirements. WFS system will be investigated for higher quality audio rendering.

Improved performance is expected by performing MPS encoder in the form of analysis by synthesis (AbS). The analysis by synthesis spatial audio codec is illustrated in Figure 5. This approach includes the advantage of a closed-loop system, for improving the quality of multichannel audio reproduction.

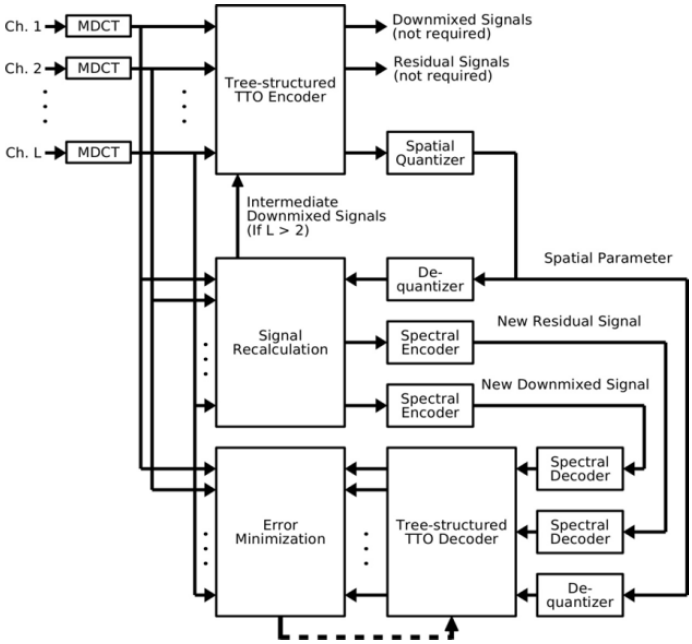


*Figure 5 - The analysis by synthesis spatial audio codec*

The requirements and specifications for MPS in the form of AbS are detailed as follows:

- To obtain the minimum approximation error, both the synthesized and approximated signals shall be synchronised and matched. Hence, each audio channel is proposed to be transformed by means of Modified Discrete Cosine Transform (MDCT) instead of a filter bank.
- A very high speed processor shall be used for the AbS encoder to perform error minimization procedure.
- The improvement achieved by AbS codec compared to MPEG Surround standard shall be measured by segmental Signal-to-Noise Ratio (segSNR).

#### 3.3.4.2 Audio scene description language

Within recent years, a number of 3D audio playback systems became more and more mature while still being active research topics. The most prominent of these systems include Wavefield Synthesis (WFS), (Higher-Order) Ambisonics (HOA) or Binaural Audio. The main problem with these approaches is that 3D audio content is typically produced for only one playback system and cannot be used easily with one of the others.

To overcome this problem, some approaches for a unified audio scene description language were developed and MPEG-4 Audio BIFS is the most prominent one. But up to date, none of them has been able to gain broad acceptance, either because of technical, practical or aesthetical issues.

Therefore, an audio scene description language will be proposed that may be interpreted by any audio rendering system provided by the end user terminal. Considering the comparatively high video data rate, reducing the audio bandwidth by a scalability concept will not be part of the investigations.

The requirements and specifications for the audio scene description language are detailed as follows:

- Description of the position, directivity and trajectories of individual audio objects shall be required.
- Both point sources as well as wide sources (e.g. ambient sounds) shall be supported.
- Description of the environment (physical and/or perceptual) shall be possible.
- The description language shall provide an extension mechanism to support, for example, future playback systems.
- The description format must be streamable.

Basic reference implementations for 5.1 and binaural rendering will be provided.

## 3.4 Networking requirements

### 3.4.1 Fixed user requirements

#### 3.4.1.1 Specification on P2P overlay formation and maintenance:

- User join/churn:
    - User join/churn shall not affect other users' QoE.
    - User shall be able to gracefully leave P2P network with a prior notice in order not to affect other users' QoE.
    - User's leaving abruptly without a prior notice may affect the QoE of its served peers temporarily.
    - User shall join the P2P overlay based on its terminal capabilities and its available network conditions.
    - User shall satisfy minimum P2P overlay requirements to join.
    - Admission control shall be enforced, being a session or a peer's request to join an overlay shall be admitted only when the available resource capabilities of the overlay meet the QoS requirements of the request.
- Network monitoring:
    - All users shall monitor their network conditions (bandwidth, jitter, delay, throughput, packet loss, signal strength, etc.) on all interfaces periodically.
    - Mobile users shall be able to provide feedback regarding their network conditions to MANE.
- Network security:
    - P2P encryption shall be optional for all users.
    - IP network main server shall check the access rights of the uploaded content to let users download it.
    - User shall send his/her credentials to the main IP network server and his/her terminal capabilities to P2P network management unit while joining the IP network.

#### 3.4.1.2 Synchronisation requirements:

- For P2P and DVB networks:
    - The main stereoscopic view sent through DVB shall be available on P2P server, but shall not be required to be sent via P2P to end users that have DVB connection.
    - Encapsulation format for P2P transmission over IP-based networks (fixed high speed broadband, Wi-Fi, 3G and LTE) shall be aligned with the encapsulation format of broadcasting network (e.g. DVB-T2) to achieve synchronisation of multimedia streams.
    - The multi-view data received via P2P networks at the user terminal shall be buffered for synchronisation purposes. The buffer size depends on the

capability of terminal device and the overall system requirement for decoding and rendering delay.
    - There shall be a mechanism to synchronise the P2P buffer and the DVB stream at the user terminal side. It shall operate in such a way that the multi-view data from P2P and the primary stereoscopic 3D view data from DVB shall be delivered to the multi-view and audio decoders within a time difference which shall not cause any noticeable visual and aural effects.
    - Timing information, e.g., timeline, shall be integrated into DVB and P2P transmission, so the different delivery paths can be synchronised in the receiver.
    - Metadata shall be available through both broadcast and IP network. Metadata through broadcast network shall be used as a backup.
- For collaborating users:
    - Collaborating users must receive live streaming content with an unnoticeable delay.
    - For demonstration purposes, there shall be 6 collaborating users (2 users for each terminal type – mobile, portable and fixed).
    - There shall be an A/V overlay communication link between collaborating users.
    - Any user satisfying the minimum requirements of collaborating group requirements (in terms of its available bandwidth, delay, etc.) shall be able to join the group.
    - A collaborating user shall be able to receive the content through its DVB interface in unfavourable P2P network conditions.

### 3.4.2 Mobile user requirements

#### 3.4.2.1 Mobility requirements in P2P networks

Having P2P session preservation is a challenge. Most of the P2P techniques in the state-of-the-art have got one important module that is used to relate the unique P2P ID of the peer, to the information to reach it in the network, which can be a sequence of hops in the wireless world, or an IP and port in the context of Internet.

ROMEO will initially target a scenario made up of an infrastructure and a set of users that communicate with the aforementioned infrastructure, potentially changing their point of attachment to it, but NOT creating an ad-hoc network. With this limitation on the scenario, it is possible to cope with the mobility of the terminals using (and expanding) techniques borrowed from the Mobile IP world. In particular, the first approach that will be studied will consider empowering a mobile terminal with more than one P2P unique identifier for each domain it visits. The IP resolution module of the P2P system maps the P2P identifiers to the IP, port information to reach the user, and the network part of the IP address will be used to reach the correct point of attachment of the mobile terminal.

This approach leads to the following requirements for the ROMEO system:

- A peer shall have one main identifier in the P2P world. This identifier is used by the ROMEO application as a source when sending data.
- A peer must be able to accept more than one network-based P2P identifier, and must be able to use the correct one while communicating, based on context information (the network it is using to send packet to the P2P system, etc).

- The P2P layer of the client application shall be abstracted on the client side by adding a layer that translates the main identifier of the client to the network-based identifier when sending packet, and the other way around when receiving packets from the P2P network.
- A (possibly distributed) resolution system must be able to align automatically in terms of network-based P2P identifiers while managing the P2P overlay. As long as one of the network-based P2P identifiers of a peer is active, and as long as the peer wants to receive a given stream, the P2P system will consider that the network-based P2P identifier (a virtual P2P peer) must receive the stream.
- The resolution system shall be informed by either the peer or the network when a network-based P2P identifier is not active anymore (the peer handed over to another network, or the peer disconnected from the P2P overlay, or the P2P is not receiving a stream anymore).

#### 3.4.2.2 Requirements for multi-homing and Media Independent Handover

Multi-homing, Media Independent Handover (MIH), and smooth video rate adaptation are functionalities required in order to ensure seamless 3D video service continuity across heterogeneous wireless networks. The requirements for ensuring such functionalities include:

- Mobile Terminal (MT) with multiple network interfaces:
  The mobile terminal shall be possible to receive either all views in one interface, or concurrently receive different views within the 3D media stream to different interfaces. In particular, MT is required to have at least one Wi-Fi 802.11 g/n network interface, one LTE interface and one DVB interface. Within the context of ROMEO, the aforementioned interfaces will support multi-homing, in order to enable MT to receive video data (views, layers) over different access technologies and exploit their available resources. Multi-homing capability will be implemented by MANE which will handle Layer-2 data flows to MT. Particularly, DVB interface will be used to connect the MT to the DVB network at all times in order to ensure video streaming with signalling minimum QoS and synchronisation among video views received over heterogeneous wireless access technologies.
- MANE to support seamless service continuity:
  An enhanced MANE with additional functionalities shall provide seamless service continuity across heterogeneous wireless access networks. Particularly, the Media Independent Handover framework (IEEE 802.21 MIH) will be integrated within MANE. In particular, MIH Information services, MIH Command services and MIH Control services shall be implemented with respect to monitor all available networks in the vicinity of MT, collect information from the network and the client side, make decisions upon handover and/or rate adaptation, creating such events and control their execution. Functionalities designed in accordance to ROMEO networking environment, including network selection, handover decision, video adaptation and mobility control shall ensure the selection of the best candidate network and seamless handoff as well as guaranteeing quality levels and seamless service continuity to MT.
- Smooth multi-view video rate adaptation:
  The video adaptation functionality within MANE will be responsible for smoothly adapting video rate according to current network conditions. The decision and extend of the rate adaptation will be based on information from the network and the MT side. The network related information will be available from the MIH Information service. Moreover, MT side information regarding the QoE of an ongoing video stream will be monitored in real time by the QoE monitoring functionality incorporated in the

enhanced MANE. In case of a handover, the video rate will be adapted smoothly (either increased or decreased) to meet the available bandwidth limitations thus ensuring high levels of QoE across heterogeneous networking environment.

- MANE shall determine the level of quality and the content delivery of its mobile users.

### 3.4.3 Requirements for sharing user generated content

#### 3.4.3.1 User generated content requirements

- Only collaborating users shall be able to share their own generated content.
- Collaborating user shall be able to:
  - Share captured 2D/3D as well as pre-stored content with other collaborating users.
  - Sign the captured content.
  - Upload the content to P2P main seed server.
  - Authorize other collaborating users to download the content.
  - Search other user generated content.
  - Download the user generated content if he/she has access rights.

#### 3.4.3.2 Mobile content capturing requirements

- Resolution and performance:
  - Video: 720x1280p25
  - Still: 8Mpix
- Audio:
  - 2 channels @ 48 KHz
- Minimal processing:
  - Video: Temporal noise filtering, Spatial noise filtering, Video stabilisation
  - Still: Spatial noise filtering, Global Brightness Contrast Enhancement (GBCE)
  - Audio: Automatic Gain Control (AGC), Adaptive noise control

### 3.4.4 Specifications for content and network optimization

#### 3.4.4.1 Quality of Experience based joint content and network requirements

- There shall be an adaptation mechanism within the P2P delivery network to advise the P2P network so that it only delivers essential view data and drops some view data if the network condition is deteriorated. The principles of 3D multi-view P2P delivery shall be:
  - Only necessary view data shall be delivered.
  - Other views and enhancement data shall also be delivered if bandwidth is sufficient and network conditions allow such activity.
  - If network conditions have deteriorated to a degree where not enough resource is available, or network is congested, or the delay is intolerable, some of non-essential view data shall be removed from the delivering stream so that the change in users' service quality of experience is unnoticeable or QoE only deteriorates gradually and gracefully.
  - The removal of view data in adverse network conditions shall be in the order of enhancement data, other non-primary views from the farthest to the nearest user view.
    - Key view frames shall be prioritized in transmitting through P2P network and a metadata description shall be included to generate the missing frames at the video decoder in an optimum way.

- Both horizontal and vertical handover decision shall be taken by considering media source QoS requirements such as bit rate, delay, jitter, etc. as well as the traffic conditions of the network to be handed over.
- Based on the network resources available, maximum number of views shall be selected and coded for transmission. Untransmitted views will be interpolated with the available views.
- Source coding configurations of multiple descriptions of MVD content shall be optimized for end-to-end rate-distortion performance given the condition of the network (e.g. Packet Loss Rate)

#### 3.4.4.2 Quality of Experience based network optimization requirements

- A Media Server must be used from which the user downloads some media content for the in order to implement QoS demonstration in (Reference Scenario 3).
- Correct values of QoS and Traffic Shaping at OLT and BRAS shall be defined to detect and prioritize different streams automatically and under user selection.
- Each peer on a P2P overlay (e.g., interface capacity- bandwidth) must be willing to forward packets of other peers.
- Each peer shall allow for configuring its schedulers, to let the system reserve the peer’s bandwidth to assure QoS to data flows.
- The QoS management information to be enforced on peers shall be conveyed to the peers by means of appropriate signalling protocol such as the Internet Engineering Task Force (IETF) proposed Next Steps in Signalling (NSIS). The bandwidth will be over-reserved and dynamically controlled on the overlay to reduce undue signalling and related overheads with increased resource utilization.
- A good view of the network topology and the related peers´ resource capabilities shall be monitored and maintained to efficiently improve the performance of the QoS delivered.

# 4 STATE OF THE ART

The overview of the existing technologies has been provided under this section.

## 4.1 Media delivery systems

### 4.1.1 Multi-view media delivery services

ROMEO system architecture proposes a full end-to-end solution for Multi-view 3D experience to end users. The system contains content capturing, broadcasting, P2P, communication overlay and fixed/mobile user terminal blocks.

Currently, 3D media is delivered to end-users via different channels separately. The two mostly used channels are broadcasting and the internet. Those fixed users, such as home users, can receive the 3D media via DVB (e.g. DVB-T2/S2/C2) [3], [4], [5] or an internet connection offered by their service and/or content providers offer such services. However, the DVB and internet delivery is always separated and not employed at the same time in this case. To mobile users, 3D contents are delivered through either wireless hot-spot access or their service providers' cellular networks if their handheld devices are capable of doing so. Of course, there is always the traditional delivery service provided by the retail chains where the 3D contents are carried on physical media.

Looking at the media delivery services over DVB, users can have choices of DVB over terrestrial, satellite, cable or handheld. DVB-T2, which offers higher bit rate and is more suitable for HDTV to home than its predecessor, is currently broadcasting in the UK (Free view HD), Italy (Europe 7 HD) and several other European countries. DVB-S2 is widely used to delivery 3D contents directly to home users in many countries throughout the world, such as SKY in UK, Germany and Italy, Freesat in UK and Ireland and DirectTV in the U.S. Meanwhile, DVB-C2 technology is about to take off in the delivery services market. DVB-NGH [6] is just emerging as the technique to provide high data rate services for handheld devices, along with other techniques such as 3G cellular system.

Multimedia content delivery over the internet is dominated by P2P technique today. According to [7], in Europe the traffic of the P2P file sharing accounts for 30.1% of fixed access traffic and streaming 33.2% in peak period in March 2011. It is anticipated that it will remain so for a foreseeable period. In terms of decentralized P2P networks, BitTorrent [8] is the most widely used technique which allows users to join a swarm of hosts to download from and upload to each other simultaneously. Using BitTorrent, content can be delivered to users with lower bandwidth efficiently. It is a very useful and efficient way to deliver large amount of media to small mobile handheld devices [9]. Other proprietary P2P systems such as Tribler [10] and CoolStreaming [11] are mainly designed for file-sharing. There are projects under FP7 (e.g. P2P-Next [12] , NAPA-WINE [13], etc.) mainly address the P2P delivery issues of 2D content distribution.

There are several EU projects on 3D media processing, delivery and presentation to the end-user. These projects are SEA, 20-20 3D MEDIA, 3D4YOU, P2P-Next, MUSCADE, MOBILE3DTV, DIOMEDES and SKYMEDIA. SEA has aspects related to context-aware networking delivery platform for 2D/3D media. 20-20 3D MEDIA has aspects of creating complete 3D media-capable chain while considering distribution and networking of spatial media extensively. 3D4YOU has features of defining 3D media delivery formats for broadcasting for home entertainment applications. P2P-Next has aspects on P2P content delivery for user centric media distribution. MOBILE3DTV is working on delivery of stereoscopic 3D video over DVB-H. MUSCADE is considering scalable 3D multi-view video via broadcast to home. DIOMEDES is concentrating on combining DVB with P2P to deliver high quality 3D media to home. Although MUSCADE [14] and DIOMEDES [15] address 3D multi-view content delivery, they do not consider user collaboration. SKYMEDIA has aspects of delivery of 3D multi-view video.

Commercial world has an increasing interest over 3D media. 3D media is viewed by the end user with glasses at home or cinemas. 3D media at commercial site can be considered in the following areas; delivery via broadcasting, via Video on Demand (VoD) and via storage devices such as DVDs.

Each of the methods described above uses stereoscopic video data to be processed at the display. Currently commercial solutions for display side are using active or passive glasses for end users to experience 3D content. The content is delivered either in a single frame with left and right resolution is reduced to half (side-by-side or top-bottom methods: Figure 6), or left and right frames are packed in a single frame with full resolution (frame-packing: Figure 7)). Frame Packing is enabled in HDMI standard only for the time being (HDMI1.4a).

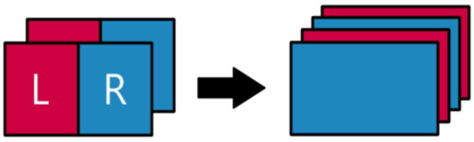


*Figure 6 - 3D Stereoscopic image in single frame displayed as 3D image at display*

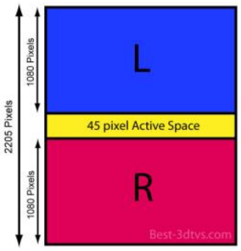


*Figure 7 - 3D stereoscopic image with 2 frames in one frame*

Since broadcasting has limited bandwidth over storage area, stereoscopic content is delivered within a single frame of video. On the other hand storage media or VoD has the advantage of unlimited bandwidth for delivery to end-user. In this scenario the content has full left and right frames in a single frame. This solution has more data compared to the broadcasting method. As a result it will have more high quality 3D data. Commercially blu-ray disks are very common in distributing 3D media to the end user. There are several blu-ray players that support 3D content. Players transmit data to display units via HDMI 1.4a. Display units differentiate the content using HDMI information and switches to 3D mode. There is also a broadcasting VoD application. Some broadcasters transfer 3D data to their set-top boxes at their subscriber’s request. All data is transferred to a storage area in the set-top box and user views the content whenever he wants. Content in this style is again side-by-side or top-bottom format.

### 4.1.2 Broadcasting methods

#### 4.1.2.1 DVB-T2

The DVB-T2 Transmission System is specified in the standard ETSI EN 302 755. This standard describes the building blocks and the signal structure at the modulator/ transmitter side in an unambiguous way. As in the standards for other DVB Baseline Systems, the receiver is not specified. The only requirement for the receiver is its capability to receive, demodulate and decode the DVB-T2 signals.

The main building blocks described in the standard are (from the input interface to the output interface of the modulator/ transmitter:

- Mode Adaptation (incl. Input Interface, in case of a multiple PLP signal: also Input Stream Synchroniser, Delay Compensation, Null Packet Deletion, CRC-8 Encoder, Baseband (BB) Header Insertion)
- Stream Adaptation (incl. in case of a multiple PLP signal: Scheduler, Frame Delay, In-band Signalling, Padding Insertion, BB Scrambler)
- FEC Encoding, Bit Interleaver, De-multiplexing bits to cells, Mapping cells to constellation, Constellation Rotation, Cell Interleaver, Time Interleaver
- L1 Signalling Generation, L1 FEC Encoding, L1 Demux and Mapping,
- Frame Builder and Frequency Interleaver
- Orthogonal Frequency Division Multiplexing (OFDM) Generation (incl. Multiple Input Single Output (MISO) Processing, Pilot Insertion, Inverse Fast Fourier Transform (IFFT), Peak to Average Power Ratio (PAPR) Reduction, Guard Interval Insertion, P1 Symbol Insertion, Digital to Analog (D/A) Conversion)

The protocol structure of the DVB-T2 signals is based on frames. An overview of the composition of a T2 frame is given in Figure 8.

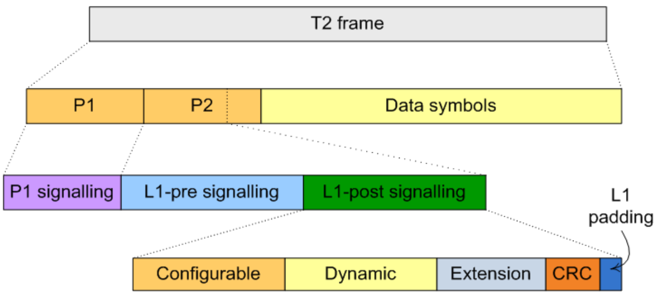


*Figure 8 - Composition of a T2 frame (from EN 302 755)*

When implementing a DVB-T2 modulator/ transmitter, the user has the choice between a multitude of parameters to adapt the system to the requirements of the services, the network and the terminal population. Table 1 illustrates as an example the OFDM related parameter set for an 8 MHz channel.

#### 4.1.2.2 DVB-T2 Lite

The variant for DVB-T2 Lite was created to specify a system that addresses mobile terminals. In principle, DVB-T2 Lite is a subset of DVB-T2 plus a few additions.

The main differences are:

- 2 new code rates added (1/3, 2/5)
- Optional scrambling added to L1-post signalling
- 32k and 1k Fast Fourier Transform (FFT) sizes removed
- Scattered pilot pattern PP8 removed
- 3 combinations of FFT size, guard interval and pilot pattern removed

- Long FEC blocks removed
- Time-interleaving memory halved
- Code rates 4/5 and 5/6  removed
- Code rates 2/3 and 3/4 not used with 256-QAM (Quadrature Amplitude Modulation)
- Rotated constellations not used with 256-QAM
- Maximum data rate significantly reduced (to 4 Mbit/s)

| Parameter | | 1K mode | 2K mode | 4K mode | 8K mode | 16K mode | 32K mode |
|---|---|---|---|---|---|---|---|
| Number of carriers $K_{total}$ | normal carrier mode | 853 | 1 705 | 3 409 | 6 817 | 13 633 | 27 265 |
| | extended carrier mode | NA | NA | NA | 6 913 | 13 921 | 27 841 |
| Value of carrier number $K_{min}$ | normal carrier mode | 0 | 0 | 0 | 0 | 0 | 0 |
| | extended carrier mode | NA | NA | NA | 0 | 0 | 0 |
| Value of carrier number $K_{max}$ | normal carrier mode | 852 | 1 704 | 3 408 | 6 816 | 13 632 | 27 264 |
| | extended carrier mode | NA | NA | NA | 6 912 | 13 920 | 27 840 |
| Number of carriers added on each side in extended carrier mode $K_{ext}$ (see note 2) | | 0 | 0 | 0 | 48 | 144 | 288 |
| Duration $T_U$ | | 1 024$T$ | 2 048$T$ | 4 096$T$ | 8 192$T$ | 16 384$T$ | 32 768$T$ |
| Duration $T_U$ µs (see note 3) | | 112 | 224 | 448 | 896 | 1 792 | 3 584 |
| Carrier spacing $1/T_U$ *(Hz) (see notes 1 and 2)* | | *8 929* | *4 464* | *2 232* | *1 116* | *558* | *279* |
| Spacing between carriers $K_{min}$ and $K_{max}$ $(K_{total}-1)/T_U$ (see note 3) | normal carrier mode | *7,61 MHz* | *7,61 MHz* | *7,61 MHz* | *7,61 MHz* | *7,61 MHz* | *7,61 MHz* |
| | extended carrier mode | NA | NA | NA | *7,71 MHz* | *7,77 MHz* | *7,77 MHz* |
| NOTE 1: Numerical values in italics are approximate values.<br>NOTE 2: This value is used in the definition of the pilot sequence in both normal and extended carrier mode.<br>NOTE 3: Values for 8 MHz channels. | | | | | | | |

| Bandwidth | 1,7 MHz | 5 MHz | 6 MHz | 7 MHz | 8 MHz | 10 MHz (see note) |
|---|---|---|---|---|---|---|
| Elementary period $T$ | 71/131 µs | 7/40 µs | 7/48 µs | 1/8 µs | 7/64 µs | 7/80 µs |
| NOTE: This configuration is only intended for professional applications and is not expected to be supported by domestic receivers. | | | | | | |

*Table 1 - OFDM parameters (from EN 302 755)*

**Table 17: Maximum bit-rate and recommended configurations for 8 MHz, 32 K 1/128, PP7**

| Modulation | Code rate | Absolute maximum bit-rate | | | Recommended configuration | | |
|---|---|---|---|---|---|---|---|
| | | Bitrate Mbit/s | Frame length $L_F$ | FEC blocks per frame | Bitrate Mbit/s | Frame length $L_F$ | FEC blocks per frame |
| QPSK | 1/2 | 7,49255 | 62 | 52 | 7,4442731 | 60 | 50 |
| | 3/5 | 9,003747 | | | 8,9457325 | | |
| | 2/3 | 10,01867 | | | 9,9541201 | | |
| | 3/4 | 11,27054 | | | 11,197922 | | |
| | 4/5 | 12,02614 | | | 11,948651 | | |
| | 5/6 | 12,53733 | | | 12,456553 | | |
| 16-QAM | 1/2 | 15,03743 | 60 | 101 | 15,037432 | 60 | 101 |
| | 3/5 | 18,07038 | | | 18,07038 | | |
| | 2/3 | 20,10732 | | | 20,107323 | | |
| | 3/4 | 22,6198 | | | 22,619802 | | |
| | 4/5 | 24,13628 | | | 24,136276 | | |
| | 5/6 | 25,16224 | | | 25,162236 | | |
| 64-QAM | 1/2 | 22,51994 | 46 | 116 | 22,481705 | 60 | 151 |
| | 3/5 | 27,06206 | | | 27,016112 | | |
| | 2/3 | 30,11257 | | | 30,061443 | | |
| | 3/4 | 33,87524 | | | 33,817724 | | |
| | 4/5 | 36,1463 | | | 36,084927 | | |
| | 5/6 | 37,68277 | | | 37,618789 | | |
| 256-QAM | 1/2 | 30,08728 | 68 | 229 | 30,074863 | 60 | 202 |
| | 3/5 | 36,15568 | | | 36,140759 | | |
| | 2/3 | 40,23124 | | | 40,214645 | | |
| | 3/4 | 45,25828 | | | 45,239604 | | |
| | 4/5 | 48,29248 | | | 48,272552 | | |
| | 5/6 | 50,34524 | | | 50,324472 | | |

*Table 2 - Extract from DVB-T2 standard on bit rates for different modes*

The result is a very robust DVB-T2 signal that has a limited capacity. Where the DVB-T2 Base version provides a capacity of up to 50 Mbps in an 8 MHz Radio Frequency (RF) channel that can in an extreme case be allocated to one service, the DVB-T2 Lite system assumes that 4 Mbps for a service to mobile terminals is sufficient, although several such services can be transmitted in one RF channel.

## 4.2 Display and terminals

### 4.2.1 Display technologies

3D video can be displayed over several types of displays. Some of these displays are HD displays with stereoscopic 3D display capability, multi-view lenticular displays, and multi-view light-field displays.

In commercial market of fixed user mobile display, 3D media is displayed via stereoscopic displays that need active or passive glasses for 3D experience. Multi-view or light field solutions are still in research topics to be able to use in end-user consumer goods.

Stereoscopic displays receive stereo video and display each of the videos one after another. User uses a passive or active shutter glasses to perceive 3D video over these stereoscopic displays. Content is delivered via Low Voltage Differential Signalling (LVDS) standard to the displays. Each frame contains information about the right and left frame sequence. Using frame information sequence, each frame is displayed respectively. Difference between passive glasses technology display and active glasses technology display lies on the method they use to deliver left and right frame information to the user. In active shutter glasses technology, there is an active Infrared signal from display unit to active shutter glasses. Each pulse shows which eye of the glasses to turn on. On the other hand, passive glasses technology displays use orthogonal polarization of the light. Since glasses are also polarized, left image can be observed by left eye and right image can be observed on right eye only.

Multi-view lenticular displays are used to display multi-view video without a need to any glasses. These displays receive all views and display each view in a very high refresh rate to a specific location in the room. For example for an 8 view display it is required to refresh at 480 Hz for a 60Hz multi-view 3D video. The main principle of this technology is to direct a specific view to a specific area. Every user is capable of viewing neighbouring views that are recorded by a serial camera array.

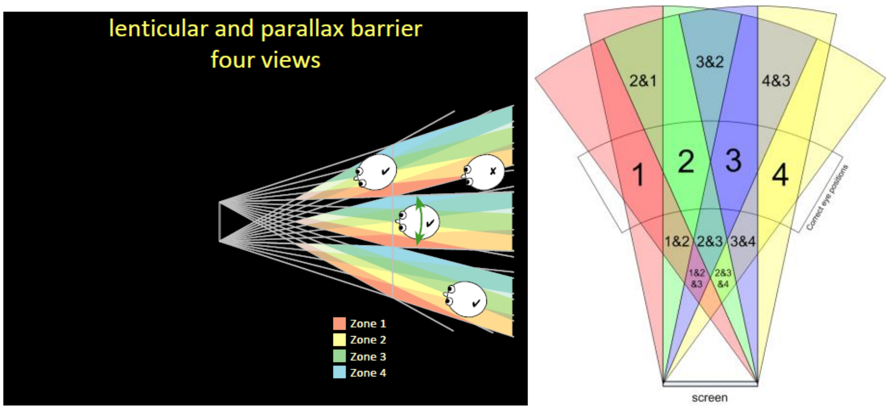


*Figure 9 - Different view areas for users [16]*

The basic operational principle of a multi-view auto stereoscopic display is to “cast” different images towards each eye of the observer. This is done by a special optical layer, additionally mounted on the screen surface which redirects the light passing through it. There are two common types of optical filters – lenticular sheet which works by refracting the light, and parallax barrier which works by blocking the light in certain directions. In both cases, the intensity of the light rays passing through the filter changes as a function of the angle, as if the light is directionally projected. These two technologies are shown in Figure 10 below.

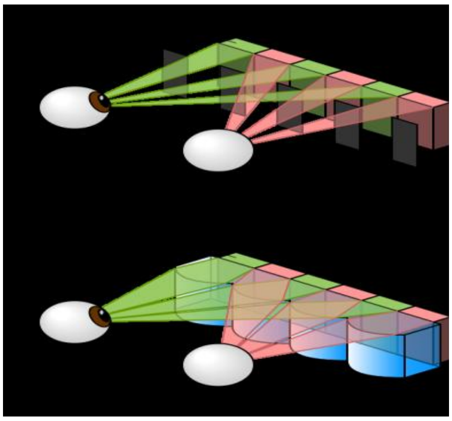

*Figure 10 - Lenticular sheet*

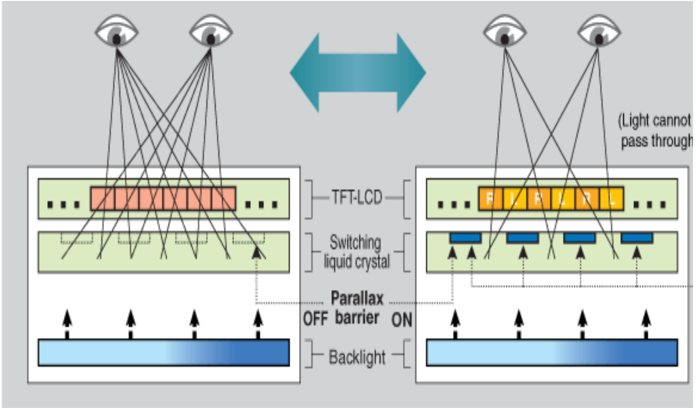


*Figure 11 - Switchable parallax barrier*

A parallax barrier, which comprises an array of slits spaced at a defined distance from a high resolution LCD behaves like a lenticular lens array. The parallax effect is created by this lattice of very thin vertical lines, causing each eye to view only light passing through alternate image columns.

Indeed, if the definition of the LCD is 1920x1080 (width * height), and if a vertical grid of slits is placed in front of this display to visualize 8 views, the definition of each view is 240x1080 which is quite unbalanced. Solutions have been proposed to better balance the loss of resolution.

System proposed by Newsight is based on an oblique pattern of vertical slits that allows to decrease the definition in both direction and so to have a balanced width / height ratio. Originality is also that the grid is not based on pixel but on components of pixels (colours). It exploits the fact that LCD panels have the Red Green Blue (RGB) sub pixels in the vertical stripe configuration. Slits are aligned with oblique views patterns and horizontally distant with each other from 8/3 pixels (see double arrow). The interlacing pattern repeats periodically every 2 columns and 3 lines (e.g. from pixel (3; 4) to pixel (5; 7)).

As a result, screen definition is reduced by a factor of 3/8 and 1/3 respectively on horizontal and vertical dimensions. So a screen with 1920x1080 LCD definition can presents 8 views of 1920*3/8 x 1080*1/3 = 720 x 360.

The main advantage of parallax barrier is the ability to switch it off, so the display works in 2D mode, thus providing backwards compatibility with 2D content. The main disadvantage of the parallax barrier is that it blocks part of the light, resulting in a lowered brightness of the display. In order to compensate for that, one needs an extra bright backlight, which decreases the battery life in mobile applications. Comparing the cost of developing and manufacturing parallax barrier and lenticular sheet, the former is much cheaper than the latter.

A second method to display a 3D multi-view video is Light-field Displays. This displays use light's propagation to different areas of the room. Display has units called voxels and each voxel emits different light beams according to the recorded view data. This method intends to build whole videos to be visible to end user to create full 3D perception.

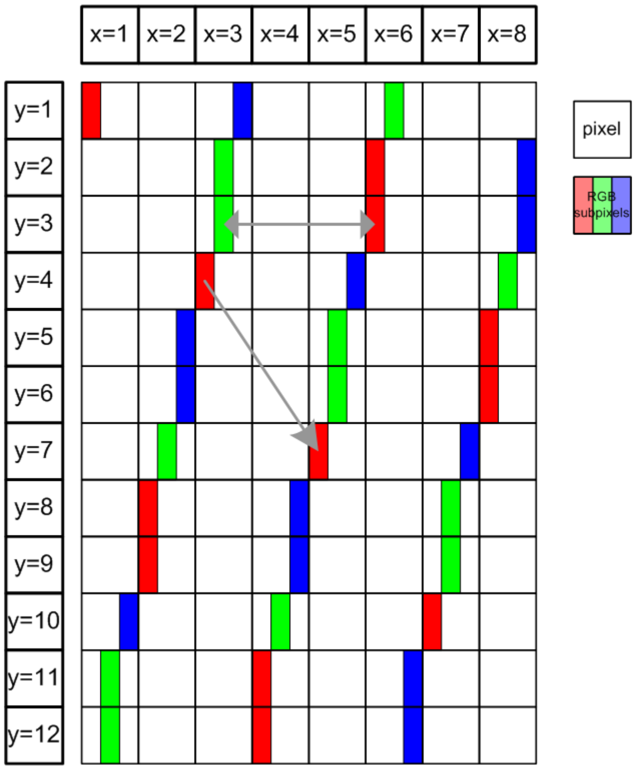


*Figure 12 - Pixel corresponding to the same view (1 among 8 views)*

Currently, there is a wide range of 3D display technologies, but not all of them are appropriate for mobile use. For example, wearing glasses to aid the 3D perception of a mobile device is highly inconvenient. The limitations of a mobile device, such as screen size, Central Processing Unit (CPU) power and battery life limit the choice of a suitable 3D display technology.

Another important factor is backward compatibility – a mobile 3D display should have the ability to be switched back to “2D mode” when 3D content is not available.

#### 4.2.2 Mobile platforms

For mobile display units, CPU is a critical component to achieve required processing capabilities. Moreover, power consumption is also a major challenge for mobile devices under huge processing needs. There are several CPUs on the market that are dedicated for mobile smart devices that can support 2D/3D HD data for mobile devices. A brief explanation of these devices is added to Appendix A.

### 4.3 Audio video format and capturing technologies

#### 4.3.1 Multi-view capturing technologies

The video streams captured by a rig of cameras should be pre-processed in order to solve the optical differences of the different cameras, the alignment of cameras on the rig and to adapt the video streams to the external capture conditions like colour distribution and exposure parameters.

Each camera has to be calibrated separately at first. A camera calibration is the process that enables to determine all the relevant parameters that describe the way how an image is captured by a camera block (optics + sensors). Thanks to them it is possible to deduce the real world coordinates of an object from its image coordinates.

It includes all the parameters that determine the relation between real world spatial coordinates and the correspondent coordinates inside the captured image.

Two kinds of parameter must be differentiated:

| **Intrinsic parameters**<br>**(depends on the camera itself)** | | **Extrinsic parameters**<br>**(depends on the camera position)** | |
|---|---|---|---|
| F | Focal point | **R** | Rotation matrix |
| $k_{u},$ $k_{v}$ | Scale factors in x and y axis | | |
| $c_{u},$ $c_{v}$ | Focal point projection in the image | | |
| $s_{uv}$ | Skew coefficient between the x and y axis (often 0) | **T** | Translation vector |

*Table 3 – Stereo system camera parameters*

The intrinsic parameters model the defaults of the camera itself (lens distortion, not centred sensor, etc.). The extrinsic parameters denote the coordinate system transformation from the 3D world coordinates to 3D camera coordinates.

The common procedure employed to calibrate cameras is as following:

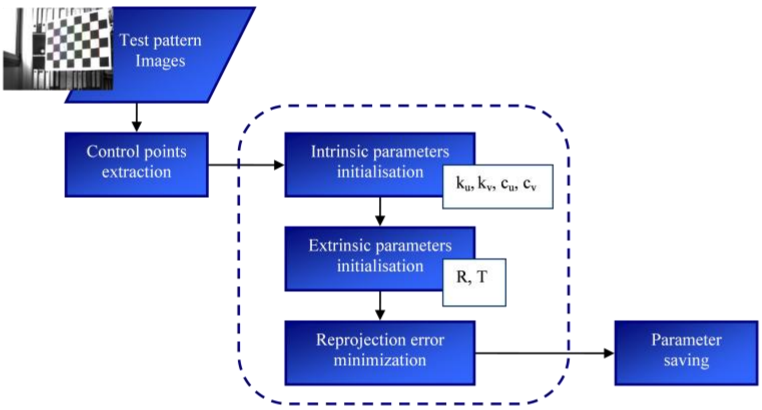


*Figure 13 - Calibration procedure*

The point positions on the test pattern are known and the goal of this manipulation is to find the intrinsic and extrinsic parameters that enable to obtain the best correspondence between the control point and their projection. Once all the parameters have been computed, it is possible to remove defaults (especially distortion) in the image captured.

Concerning stereo or multi-view video acquisition, the process must be completed by computing the geometrical relationships between two or more cameras in space and estimating their parameters. This process is very similar to which has been described above, except that the rotation and translation vectors will describe a geometrical relationships between the different cameras of a rig.

Thus, for a given 3D point P in object coordinates, the single-camera calibration algorithm is used separately for the two cameras to put P in the camera coordinates for the left and the right cameras respectively. Given many joint views of test pattern control points, the single-camera calibration algorithm is used to solve rotation and translation parameters of the test pattern views for each camera separately. The stereo calibration algorithm then plugs these left and right rotation and translation solutions into equations to solve the rotation and translation parameters between the two cameras. It must be generalized for the alignment of multi-view streams captured by a rig of several cameras.

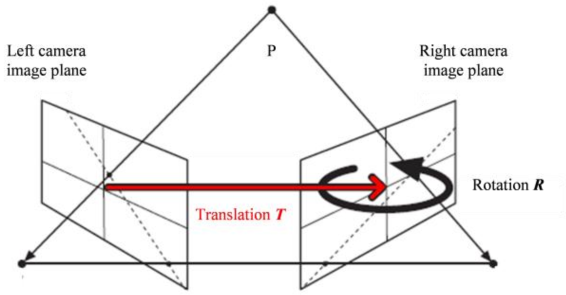


*Figure 14 - Stereo calibration rotation and translation vectors*

It is easier to run image processing algorithms when the two image planes are exactly aligned. Unfortunately, as discussed previously, a perfect aligned configuration is rare with a real multi-camera system. The goal of stereo rectification is to project the image planes of the two cameras so that they reside in the exact same plane, with image rows perfectly aligned into frontal parallel configuration.

Given the rotation matrix and the translation vector between the stereo images, an algorithm developed by Bouguet and based on Tsai and Zhang publications [17] [18] [19], can be used to achieve the stereo rectification. The result is rectified images (previously calibrated and undistorted) which are perfectly row aligned. In a multi-view environment, each camera must be calibrated and rectified in relation to one camera in reference position of the set.

Another major problem of multi camera systems is the colour calibration of cameras. Variations in camera parameters, different illumination conditions, or changes in viewpoint often cause changes in colour values of corresponding regions in the two images of a scene. Those variations can lead to major problems during 3D processing. Therefore, a colour calibration must be fulfilled to make the 3D algorithms more accurate. Its goal is to set the same colour distribution to all images.

Two different kinds of colour calibration algorithms can be found: the first one is based on the global statistics computed all over the images; the second one is relying on the correspondence of pixels retrieved from geometry calibration. The first one is generally requiring less computational effort while providing still good results.

The key of the colour calibration chosen is to work in a decor related colour space. It implies that the modification of a pixel colour requires the modification of the value of each colour channel in a coherent way. Recently, Ruderman et al. developed a colour space, called lαβ, which minimizes correlation between channels for many natural scenes [20]. This space is based on data-driven human perception research that assumes the human visual system is ideally suited for processing natural scenes (The authors discovered lαβ colour space in the context of understanding the human visual system).

Once the images captured are correctly rectified and colour-calibrated, disparity maps can be computed. Disparity map computation (that is also called "dense stereo matching" or "stereo correspondence") is a very wide-ranging research topic that mobilizes a lot of scientists around the world and a lot of algorithms have been proposed [21], [22]. Like it was explained before, its main principle is to find, for a given point in the first image of a stereo calibrated image couple, the corresponding point in the second image in order to deduce the distance between them.

The development of accurate dense stereo matching methods can be very difficult, especially in the image areas including occlusion, object boundaries or fine structure that can be appeared blurred inside the depth map. It is also challenging due to lower repetitive textures, illumination differences and recording conditions.

In the scope of real-time applications, Heiko Hirschmüller [23] has proposed an algorithm based on a Stereo Processing by Semi global Matching and Mutual Information method which is providing interesting results. Moreover, it can work even if the cameras are not perfectly calibrated, which can be the case in the rough conditions in which the image will be captured (although a calibration-rectification have been done).

In order to have the best results as possible during 3D processing, the disparity map has to be very accurate. Therefore a refinement step can be done to enhance it in order to decrease the number of border artefacts, occlusion problems and disparity value errors.

It is supposed that in a coloured image, connected pixels of a similar colour belong to the same object and thus have the same depth. Therefore a colour based segmentation of the original captured image will help to refine the disparity map.

The refinement algorithm has two steps. First, the original captured image is smoothed using a variant of anisotropic diffusion [24]. Secondly, colour based segmentation is applied on the resulting image. Each segment is reported on the disparity map and is filled with the mean value of all the disparity pixels that belong to it.

### 4.3.2 Video codecs for multi-view content

3D video constitutes more than one camera view. Stereoscopic 3D video stands for a dual camera system, where each eye view corresponds to one camera view. Based on the different representation formats for stereoscopic and multi-view videos, different coding schemes have evolved over time to compress such media efficiently. Pure colour based (e.g. multi-view, stereo L-R) or depth based formats (e.g. colour-plus-depth or MVD) are usually compressed using the legacy block based coding standards, such as MPEG-4 Part 10/ H.264 AVC [17], or its extensions, such as SVC [25] or Multi-view Video Coding (MVC) [26]. AVC has also introduced new profiles to support coding of stereoscopic videos, such as Stereoscopic High Profile [27]. Both AVC derived codecs, namely SVC and MVC, get their base layer fully compliant with AVC specifications in order to be compliant with legacy deployed material. The purpose of SVC is to provide a universal media bit-stream that can be decoded by multiple decoders of different capacities to produce the reconstructed media at different states. SVC also provides dynamic adaptation to a diversity of networks, terminals and formats. This is extremely useful for simple adaptation of transmission, efficient storage and beneficial for transmission services with uncertain resolution, channel conditions and device types. The MVC amendment uses both temporal and inter-view redundancy in order to compress multiple camera views with a typical 20% to 50% lesser overhead compared to equivalent 2D content. With MVC, most of the stream syntax remains identical with respect to the AVC and SVC amendment, and the base view is encoded with AVC. This means that an AVC compliant decoder fed with an MVC stream will be able to decompress the base view. Figure 15 shows a generic multi-view video service flow, depicting a number of use cases in the user terminal, where any of the aforementioned coding standards could be utilized.

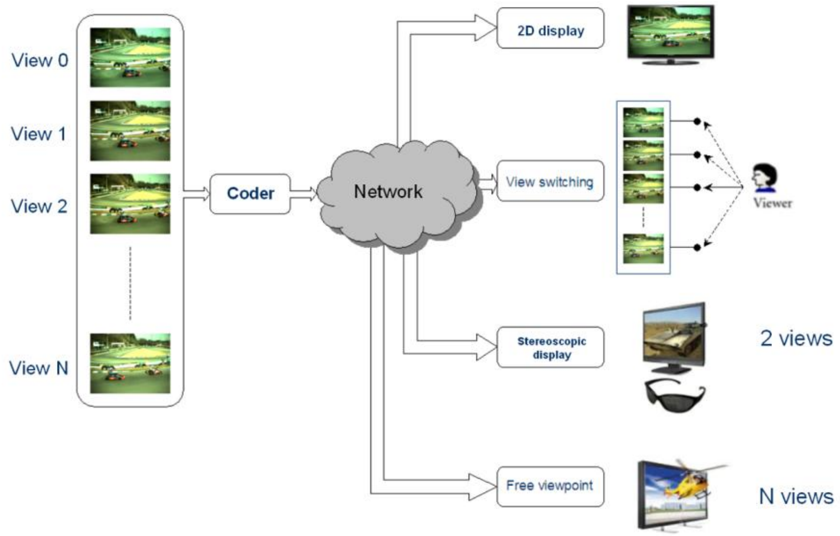


*Figure 15 - Generic Multi-view video service framework*

Simulcast coding technique (coding both left and right eye video components simultaneously with independent encoders) is the most natural and old way of compression stereo or 3D video with more than two viewpoints. Being usually the most computationally efficient and fast way of encoding 3D videos, it is not bandwidth efficient in most cases, since the statistical dependencies vastly existing among different views are not exploited. Nevertheless, utilization of asymmetric coding techniques in the context of simulcast 3D video coding (e.g. assigning different quantization step sizes for coding separate viewpoints) can yield considerable bandwidth savings without sacrificing the perceived 3D video quality. Cues of human visual system need to be taken into account carefully while assigning different quantization step sizes, in stereoscopic or multi-view video coding, in order not to introduce eye-fatigue problems after long time watching. Besides, it is also possible to signal the method(s) used to pack the frames of a stereoscopic video pair in the video bit-stream, utilizing supplemental user data (e.g. Supplemental Enhancement Information – SEI messages defined in AVC standard). The AVC standard has already defined such an SEI message that signals the frame packing method. The Frame Packing Arrangement SEI message tells the decoder that the left and right view frames are packed into a single high-resolution video frame (e.g. HD frame compatible) either in a top-to-bottom, side-by-side, checkerboard, or any other arrangement. It is also possible to signal the temporal interleaving scheme used for left and right view frames that are not packed into a single HD frame (e.g. Full Stereo). Packing both views into a single video frame makes it possible to use existing AVC conformant hardware equipment, decoders & set-top boxes to decode 3D video streams immediately without having to upgrade them to decode other 3D video specific standards such as MVC and AVC Stereoscopic High Profile. It is possible to exploit the frame compatible 3D video services as the base layer of a more extensive 3D video service, where the additional enhancement layer would exploit spatial prediction from the frame compatible base layer at the lower spatial resolution to produce Full HD stereoscopic output. SVC standard that offers scalability in spatial, Signal to Noise Ratio (SNR) (e.g. quality) and temporal dimensions is suitable in such scenarios, since the base layer of it (e.g. the frame compatible stereo video) can be decoded with AVC conformant decoders.

Inter-view prediction method, as an extended version of motion compensated prediction already existing in modern block based video coders, such as AVC, can compensate for spatial redundancies existing between two views of the same scene. Instead of motion (in temporal direction), disparity in the inter-view domain is estimated and compensated in this prediction scheme. In this case, frames are not only predicted from temporal references, but also from inter-view references. The prediction is done adaptively, which means that the best predictor among temporal and inter-view references can be selected on a block basis in terms of bit-rate - distortion cost. Stereo High Profile of AVC allows coding two views in a stereo video using inter-view prediction for one of the views. The other view is encoded independent of the other view and can be decoded straightaway with legacy AVC decoders conforming to a particular profile. MVC standard employs the same idea and offers the same profile in addition to another multi-view high profile. Research has shown that coding multi-view video with adaptive inter-view prediction can yield remarkably better rate-distortion performance compared to independent (e.g. simulcast) coding [28]. Especially if the stereo or multi-view data set consists of densely packed cameras (i.e. very small disparity), the efficiency of the inter-view prediction becomes more prominent and bit-rate savings up to 50% can be reached with respect to simulcast stereo video coding. Depth based representations for 3D video can also be encoded using the legacy video coding standards described, such as AVC, MVC and SVC. Major advantages of using depth based representations (e.g. MVD or colour-plus-depth) are backwards compatibility with 2D, independency regarding display and capture technology, direct compatibility with most “2D to 3D” algorithms, user-controlled global depth range, and good compression efficiency (low overhead).

Depth maps constitute different texture characteristics than colour videos, whereas coding them using the same tools usually generate considerably lower bit-rates than colour videos. Depth videos mainly consist of larger homogeneous texture areas and sharp transitions along object boundaries at various depth levels. Hence, only low and very high frequencies are dominant in the frequency spectrum of depth map videos. On the other hand, depicted video coding standards are typically designed to preserve low frequencies and in case high compression ratios are applied, depth transitions are blurred significantly. This in turn affects the final rendered 3D scene quality, because the depth values at the object borders have more visual significance than any other image areas in the depth map. Applying the same coding tools directly on the depth maps targets only the reconstruction fidelity of depth map pixels, whereas more dedicated tools based on the legacy standards should target the perceptual quality of the synthesized videos using the reconstructed depth pixels. Hence, rate-distortion rationale for depth coding needs modification in that sense and some of the current research works target adapting existing video coding standards to depth maps. Other than this, some works have applied the asymmetric coding principle on the depth maps, but in spatial resolution. Non-uniform post decoding filter is applied on the pre-encoder decimated depth videos to rescale them to full resolution, while preserving the depth transitions [29]. There have been other coding techniques developed for depth map coding, which are not based on block based video encoders described previously. For example, platelet based coding described in [30] tries to represent each block in a depth map frame using linear equations, exploiting the fact that depth map textures exhibit the characteristics of piecewise linear functions. Nevertheless, it is currently a common practice to employ the video coding tools that are applied to compress colour viewpoints in compressing depth maps.

Recent trends in 3D video coding adapt themselves to the application context (e.g. multi-view broadcasting) as well as to the envisaged 3D multimedia formats. It is certain that although computationally intensive multi-view coding techniques, such as MVC standard, can save considerable bit-rates with respect to simulcast coding, they are not capable of maintaining the total bit-rate in manageable limits with ever increasing camera numbers. MPEG has published a call for proposals on a new 3D video coding standard [31], which considers a framework of multi-view 3D scene generation from a limited number of cameras (e.g. 2, 3). The sophisticated depth image based rendering techniques coupled with the multi-view decoders will generate virtual views, the number of which is much higher than the number of originally coded camera views. At the same time, there are on-going European research projects that target broadcasting of multi-view media to end users through a number of networks (e.g. DIOMEDES, MUSCADE). DIOMEDES project adopts SVC standard to encode different camera views and their depth maps in a multi-view video set, where colour viewpoints have different quality layers for flexible bandwidth adaptation at client terminals. Furthermore, SVC is coupled with a visual attention model based quantization level assignment in individual quality layers, which in turn results in more flexible bit-rate distribution among layers, while reducing quality gaps in between them. As many number of camera as necessary can be incorporated in the system. Free-viewpoint navigation capability within user devices is enabled by selective streaming and downloading, where intelligent packet scheduler in the user terminals can adapt the bit-stream and select the viewpoints to be downloaded. Hence, every user would only download the layers and views necessary for them to render the desired viewpoint. Utilization of SVC simulcast multi-view coding provides an extensive flexibility, such that there is no interdependency among views preventing the users to download redundant views or frames for the purpose of decoding only. Quality scalability also offers improved bandwidth adaptation range, where the system can adapt the individual cameras' qualities before truncating them completely. Since a P2P communication overlay is exploited, the intelligent downloading client can even request packets of a particularly important camera's frame from multiple neighbours that results in an adaptive multiple description system. Parallel SVC decoder instances run concurrently and fast to achieve the desired playback frame rate.

In MUSCADE, a 4-view plus depth system is envisaged, where a narrow baseline stereoscopic view pair in the middle of the multi-view rig can straightaway be displayed on compatible 3D displays after decoding. Other two views, which are farther apart from the central stereoscopic video pair along with their depth maps, enable both multi-view applications and free-viewpoint applications. MVC standard is adopted, while a layered view scalability framework is envisaged. Base layer that consists of the stereoscopic video pair can be decoded with legacy stereo profile of MVC and displayed on compatible displays with just proper pixel interlacing. The second layer consists of a depth enhanced stereoscopic video, where the additional depth maps can be exploited in adjusting the stereo comfort by arranging the baseline depending on the user and display (e.g. size). Both depth and colour view pairs are encoded separately in two different MVC encoders. The third layer includes the side cameras' colour and depth components, which are also encoded as dependent views on one of the camera in the stereoscopic pair (i.e. MVC views), and can help more advanced multi-view applications, where a depth image based renderer can synthesize desired number of output views. The system can adapt at three different levels, depending on the application context, as well as the capacity of the transmission network utilized.

### 4.3.3 Audio codecs for multi-view content

#### 4.3.3.1 Human spatial hearing

Due to having two ears, humans can localise sound direction. Sound waves coming from a source reach the two ears at different times because they travel different distances. Shadowing effects by the human head also cause the sound pressure at each ear to have different levels. This means that human hearing system can simply detect the position of a sound source by processing mainly the inter-aural time differences (ITDs) and inter-aural level differences (ILDs) [32]. Shadowing by human head, reflection and diffraction by surrounding environments can also cause differences in the frequency spectrum of the signals received by the left and the right ear, which are used by the human auditory system to further improve the localisation of sound sources.

However, there are some positions such as in front or in the back of human head, which would produce identical ITDs and ILDs. If a sound comes from these locations, humans will get confused and cannot precisely determine the source position. This phenomenon is known as cone of confusion [33] and can be solved by head movements [34]. Using the changes of the ILDs and ITDs, which are referred to as dynamic binaural cues, the human auditory system is able to assign the correct position to the sound source. Another situation that makes human brain fail to detect proper positions of sound sources is when there are two or more sound sources, located in different locations, emitting the same sound with small time delays. These sounds will be perceived as a single sound source localised at the position of the sound source emitting the sound first. Such a situation is called the precedence effect [35].

The human head's size, the unusual shape of pinna, as well as the reflections off the shoulder and body will also affect the ability to localise a sound source. All of these cues can be represented as a head-related transfer function (HRTF) [36]. This function is a frequency response of the human auditory system. It is measured for a sound signal coming from every possible source direction in an anechoic chamber. If the measurements also include a real sound source (possibly a loudspeaker), the acoustic properties of a room and the HRTF of a human head, these are commonly called Binaural Room Transfer Functions (BRTF). Converted into in the time domain, impulse responses are generated. These are called Binaural Room Impulse Responses (BRIR) [36].

#### 4.3.3.2 Spatial audio processing: Recording and reproduction

In order to maintain its spatial effect, sound is usually recorded, transmitted and reproduced in several channels such as two (stereo) or more. Most frequently used number of audio channels is five which is known as 5.1 audio systems. Three channels are located in front of the listeners which are the centre, left and right channels and two channels are positioned in the back, which are the rear/side channels. The 5.1 is for low frequency enhancement (LFE) channel where the bandwidth required for this channel is much lower than the others. Other channel configurations are also available such as 7.1 and 10.2 audio systems [33]. Principally, all of these systems are intended to give more spatial effects to the listener.

Each channel of a multichannel audio content is usually recorded separately by one microphone. Each microphone is usually intended to capture the sound signals coming from a specific area. Otherwise, all channels can be recorded by a microphone array which provides a set of microphone signals such as B-format. Some commercial microphone products such as Holophone [37] and SoundField [38] are available to capture high quality spatial audio and record as B-format microphone signals.

Three dimensional audio can be reproduced by playing audio signals with a number of loudspeakers. Generally, audio signals recorded by a specific channel configuration should be reproduced by the same configuration.

The convolution of channel based audio signals with BRIRs generates a binaural signal. If this binaural signal is played back through headphones, it gives the listener a spatial impression, comparable to the one in this room at the recording position. This method is called binaural room synthesis [39] and exists as some rare implementations for professional production. For mobile applications, headphones can be an interesting option to reproduce multichannel audio signals by transcoding multichannel audio into binaural audio. Using a convolution with BRIRs for generating a binaural signal on mobile devices is not available at the starting date of ROMEO.

#### 4.3.3.3 Spatial audio coding

Efficient coding techniques play essential role in delivering high quality multichannel audio such as those used in home entertainment, digital audio broadcasting, computer games, music downloading and streaming services as well as other Internet based applications such as teleconferencing [40]. The traditional way for compressing multichannel audio is to encode each audio channel by a mono audio coder such as Dolby AC-3 [41] and MPEG AAC [42]. Even though high quality audio reproduction is achieved, however, in this method, the number of bits, to be transmitted, increases linearly with the number of channels.

A recent concept proposed for encoding multichannel audio signals is performed by extracting the spatial cues and down mixing multiple audio channels into a mono or stereo audio. The down mixed signals are subsequently compressed by an existing audio encoder and then transmitted accompanied by the spatial parameters as side information. Any receiver system which cannot handle multichannel audio can simply remove this side information and just render the down mixed signals. This provides the coder backward compatibility which is important for implementation in various legacy systems. In addition, by utilising the spatial parameters, the down mixed signals can be directly up mixed in the decoder side into channel configuration that can be different from the one used in the encoder side. This technique has been known as spatial audio coding (SAC) [43].

Various SAC techniques have been proposed such as binaural cue coding (BCC) [44], [45] and MP3 Surround [46]. Inter-channel level difference (ICLD), inter-channel time difference (ICTD), and inter-channel coherence (ICC) are extracted as spatial parameters that are basically determined based on exploitation of human spatial hearing [42]. Techniques such as parametric stereo (PS) [47], [48], [49] and MPEG Surround (MPS) [50], [51] may also utilize other parameters such as inter-channel predictability [52] and utilize signal processing techniques such as decorrelator [53].

#### 4.3.3.4 ROMEO implementations

For the ROMEO project, the audio signals will be recorded using various multiple, individual microphones to provide capable content for an object-based scene description. Through a renderer, a discrete 5.1 mix will be created and afterwards compressed by means of standardised spatial audio codec. A novel spatial audio coding technique will be investigated by performing the MPEG Surround in the form of AbS [54]. At the decoder side, 5.1 audio systems will be used for audio reproduction. For better spatial audio reproduction, WFS will be investigated as well as object-based binaural reproduction for the mobile terminals [55].

## 4.4 P2P network specifications and optimization

### 4.4.1 P2P live streaming

This section will be divided into two parts. First a brief introduction to P2P systems will be given. In the second part a summary of some of the existing P2P live streaming methods will be summarized.

#### 4.4.1.1 P2P networking

A brief formal definition of P2P Networking is “a set of technologies that enable the direct exchange of services or data between computers” [56].

As can be understood from the above definition a P2P system is a set of computers that communicate between each other to share some services. So far P2P systems have been used extensively in applications like:

- File Sharing (Kazaa, Napster, Gnutella, BitTorrent …)
- Voice over IP (Skype …)

One of the new areas of application has become live streaming media, since client-server based solutions (Content Delivery Networks (CDN)) have become costly for the service providers. To overcome this problem, many research groups and companies have been researching in this area and come up with different solutions. In the next section, at first a brief summary of the main methodologies will be given. This will be followed by the implementations done by other research groups.

#### 4.4.1.2 Streaming topologies

##### 4.4.1.2.1 Tree-based P2P streaming

In tree-based approaches, an overlay construction algorithm runs participating peers into multiple trees [57], [58]. Each peer determines a typical number of trees to participate based on the same parameters that the designer configured. For minimizing the effects of churning and effectively utilizing the network resources, peers are organized into multiple diverse trees as illustrated in Figure 16. At this stage each peer most generally is chosen to be an internal node in only a diverse tree, and a leaf node, in all the other diverse trees. Then, the MDC encoded content is sent through a specific tree. As a result, the main concern of these types of designs is tree construction methods.

##### 4.4.1.2.2 Mesh-based P2P streaming

In the mesh-based approach, participating peers form a randomly connected mesh overlay [59], [60]. According to this approach each peer tries to serve a list of members, which are randomly selected (Of course there are some limitations e.g. how many nodes a peer will maintain. This number is generally called the outgoing degree). Also each peer maintains a specific number of peers, from which it receives the stream, called parents (The number of these peers are called incoming degree). After being connected, any peer contacts a bootstrapping node to receive a set of peers that it can potentially connect and begin streaming. The bootstrapping node also maintains the outgoing degrees of nodes.

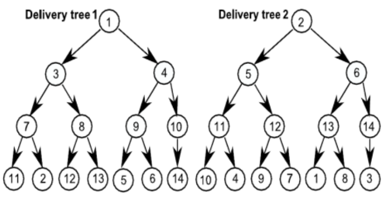


*Figure 16 - Multiple-Tree Overlays*

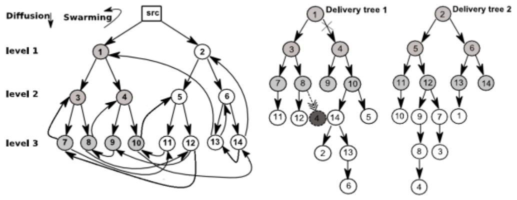


*Figure 17 – Mesh Overlay*

### 4.4.2 Existing protocols

In this section only the two main protocols are explained in detail. For some of the extensions a detailed explanation is given in Appendix B.

#### 4.4.2.1 BitTorrent protocol

BitTorrent is a widely used P2P protocol for sharing files. This protocol has been published in 2003 by Bram Cohen. This protocol has a mesh based overlay.

Before detailing the protocol below are some definitions:

- Seeder: A peer that provides the complete file.
- Leecher: A peer that still downloads the file.
- Initial Seeder: A peer that provides the initial copy is called the initial seeder.
- Tracker: Server that coordinates the file distribution.

The protocol functions as below ([8], [61]):

1. To share a file or group of files, a peer first creates a .torrent file, a small file that contains metadata about the files to be shared, and information about the tracker.
2. Peers first obtain a .torrent file, and then connect to the specified tracker, which tells them from which other peers to download the pieces of the file.
3. The .torrent file contains the following information:
    a) The URL of the tracker (announce)
    b) Pieces <hash1,hash 2,….hash n> (info)
    c) Piece length (info): This is a fixed size for all pieces.
    d) Name (info): The directory name where the file should be saved.
    e) Length of the file (info)
4. The tracker returns to the peer that wants to download the content with the following information:
    a) IP address, port, peer id
    b) State information (Completed , downloading or stopped) for a random list of peers
5. Initial seeder chops the file into many pieces (called chunks) and shares it with other users.
6. Leecher first locates the .torrent file that directs it to a tracker, which tells which other peers are downloading that file. As a leecher downloads pieces of the file, replicas of the pieces are created. More downloads mean more replicas are available for other peers to download the content.
7. When a peer finishes downloading a piece and checks that the hash matches. It sends to other peers that it has the piece.
8. As soon as a leecher has a complete piece, it can potentially share it with other downloaders. Eventually each leecher becomes a seeder by obtaining all the pieces, and assembles the file and verifies the checksum.

In BitTorrent one of the most important aspects is the piece selection methods and these are:

1. Strict Priority
   - First Priority
2. Rarest First
   - Determine the pieces that are most rare among your peers, and download those first.
   - This ensures that the most commonly available pieces are left till the end to download
3. Random First Piece
   - Special case, at the beginning
   - No pieces at all, important to get a full copy as soon as possible
   - Select a random piece of file and begin downloading it
4. Endgame Mode
   - Near the end, missing pieces are requested from every peer containing them. When the piece arrives, the pending requests for that piece are cancelled.
   - This ensures that a download is not prevented from completion due to a single peer with a slow transfer rate.
   - Some bandwidth is wasted, but in practice, this is not too much

Incentive Mechanisms:

1. Choking Algorithm:
   - Choking is a temporary refusal to upload. It is one of BitTorrent's most powerful ideas to deal with free riders (those who only download but never upload).
2. Tit-for-tat strategy is based on game-theoretic concepts.
   - Reasons for choking:
     - Avoid free riders
     - Network congestion
     - A good choking algorithm caps the number of simultaneous uploads for good Transmission Control Protocol (TCP) performance. Avoids choking and unchoking too quickly, (known as fibrillation).
   - Peer changes the peers that are choked every ten seconds.
3. Optimistic Un-choking
   - A BitTorrent peer has a single "optimistic unchoke" to which it uploads regardless of the current download rate from it. This peer rotates every 30 seconds.
   - Reasons:
     - To discover currently unused connections are better than the ones being used
     - To provide minimal service to new peer Upload-Only mode
     - Once download is complete, a peer has no download rates to use for comparison nor has any need to use them. The question is which nodes to upload to?
     - Policy: Upload to those with the best upload rate. This ensures that pieces get replicated faster, and new seeders are created fast

#### 4.4.2.2 VidTorrent protocol

VidTorrent is a protocol designed for streaming media. The protocol is based on a multi-tree overlay, thus relies on a MDC based algorithm.

VidTorrent Architecture as described in [62] and [63]:

1. Distributed protocol that dynamically creates and maintains an overlay forest.

2. The forest is composed of independent trees that carry different parts of the stream. Trees are constructed and maintained by clustering nearby nodes.
3. Aggressively probes for bandwidth and adapts to changing network conditions and node failures.

A single tree requires each node to have enough upstream capacity for the full stream to contribute. Below are the reasons to use a multiple tree:

1. Residential broadband connections are typically asymmetric.
2. Splitting the stream to multiple trees reduces the granularity and allows us to leverage residual upstream capacity of the system.
3. VidTorrent partitions the stream to multiple interleaved sub-streams, distributed by independent trees
4. Nodes reconstruct the stream by aggregating the sub-streams.

Node Join:
1. A new peer wants to join the stream.
2. Initial peer (tracker) sends a list of available parent peers.
3. Peer probes possible parents to choose the best among them, and inserts itself to the tree. Peer probes the possible parents by:
    - Bandwidth test (very basic implementation, just send a package, get the trip time and divide the size to the time.)
    - Round Trip Time test

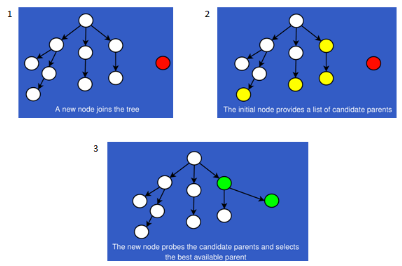


*Figure 18 – A node joins tree*

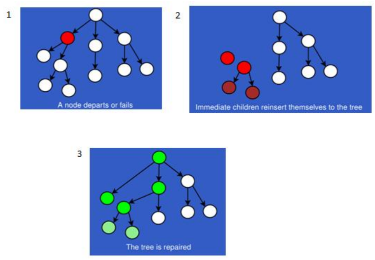


*Figure 19 – Peer churn*

Peer Churn:
1. A peer churns.
2. Children of this peer detect the problem.
3. They act as new peers and try to rejoin tree.

In case of churn, for VidTorrent it is likely to cause a pause in the stream for peers. To overcome this problem:

- Each node keeps a buffer of a few seconds for each tree it participates.
- Nodes synchronise their buffers with their parents after failures and provide the illusion of an uninterrupted stream.

There is not really a piece selection in VidTorrent. Each parent sends what they have in their buffer to their children. This is actually a result of the MDC encoding allowing independent parts to be decoded.

#### 4.4.3 P2P related projects

Below are some selected implementations for the P2P live streaming. These are the ongoing projects in the area. For some of the finished projects please refer to Appendix B.

##### 4.4.3.1 Saracen

Saracen is an EU funded project for live streaming 2D and single view 3D stereoscopic content. As described in [64]:

“The Saracen initiative will research on distribution of multimedia streams supported through innovative techniques, both as regards media encoding and media distribution including P2P schemes, where the peer nodes will participate in an end to end media distribution system making use of social networking related information.”

As explained in [65], Saracen P2P protocol is an alteration on the BitTorrent protocol for live streaming. It is called the P2P SVC protocol. The following changes are made to the BitTorrent protocol:

1. Live stream is divided into chunks of 2 seconds. These chunks are then treated as the whole file to be downloaded in BitTorrent which means that they are divided into smaller packages and then are downloaded by the peers. Here a seed is a node in the topology which has at least all the needed parts to play the 2 second video.
2. SVC P2P does not use equally sized blocks as in BitTorrent.
3. SVC P2P does not use rarest first or random piece allocation algorithms. Instead Priority Sliding Window (PSW) is used for picking pieces from neighbouring peers. An additional consideration is the encoding method used.
   As described in [65]:

   “PSW is an adaptation algorithm that chooses the adequate pieces already stored at each peer, in a download time period of 2 seconds, and sets a Chunk download window of size3 (i.e., 6 seconds of video) that prioritizes the base layer files. In the Chunk download window the number of Layers can increase or decrease, depending on several measured conditions (e.g., available bandwidth, latency, CPU load, etc.). As the chunk download window slides every 2 seconds, priority assignments can be given to higher enhancement layer for next Chunks, if conditions are favourable.”

4. For peer selection, SVC P2P has a slight modification to BitTorrent. Peers which are geographically nearer will be given a higher priority.

##### 4.4.3.2 DIOMEDES

DIOMEDES is an EU funded project. As described in [15]:

“Work will focus on new methods for the compression and delivery of multi-view video and multi-channel audio to users. The DIOMEDES approach is to develop a 3D P2P distribution system, which will be designed jointly with novel video compression techniques. The compression, distribution, and security technologies developed within DIOMEDES will be demonstrated in the form of an integrated test bed.”

The P2P system used in DIOMEDES is based on BitTorrent. The differences between the used methods and BitTorrent are [15]:

1. Since SVC is used for video encoding, base layer chunks have a higher priority than the enhancement layer chunks.

2. To achieve streaming a sliding window is introduced for picking chunks. This means the priority of the packages nearer to play out have higher priority.
3. To solve slow start problem, fairness concern for first couple of chunks are avoided. If a peer requests first chunks of a stream, other peers will try to upload these pieces if they have capacity.

The approach followed by DIOMEDES project [15] is to manage peers into clusters according to the requested views. When a peer connects for the first time, tracker forwards a random subgroup of peers. In the following connections, peers with similar buffer map pattern are preferred, and when downloading a chunk, a peer will contact peers that have downloaded similar number of enhancement chunks.

#### 4.4.3.3 P2P-Next

P2P-Next aims to build an infrastructure to build P2P streaming applications. This infrastructure will then be shared with public as an open source project for further use.

The architecture for P2P-Next is based on BitTorrent. The differences between P2P-Next and BitTorrent are, as defined in [66]:

1. SVC is used for video encoding, thus base layer and enhancement layer chunks have different priorities.
2. The chunk picking policy is also altered with a sliding window. The nearer the chunk to the playing times, the higher the priority it has.
3. Base layer chunks are downloaded in sequential order whereas for enhancement layer chunks a rarest first for a given window is used.
4. The list for chunks to be downloaded is updated at each I-Frame. The chunks still in the list before the update are discarded.
5. Peer selection is based on the performance of the peers. Each time a download from a peer fails, the bad performance counter is incremented. Peers with a larger bad performance counter are then pushed down the peer list.
6. Chunks with high priority are downloaded from good peers, whereas lower priority chunks are downloaded from bad peers (based on their bad performance count).

### 4.4.4 Ensuring QoS by class-based network resource over-provisioning

In the ROMEO project, a major requirement is supporting real-time multimedia streaming with Quality of Service (QoS) guarantee in terms of bit rate, delay, and jitter and packet loss to remote collaborating users to be reached synchronously via both DVB and P2P overlay networks. In such a scenario, application functions must specify the performance level that they require from the network for every QoS-sensitive traffic flow by means of appropriate QoS control metrics (e.g., bit rate, delay, jitter, etc.) on one hand. On the other hand, the network control system must provide the ability to treat admitted traffic flows differently to ensure that each flow is receiving the QoS contracted to it. Integrating these control capabilities in the P2P overlay network is fundamentally an end-to-end issue with two major requirements: (1) individual network elements (peers and IP routers) along a communication path/tree followed by an application's data packets must implement appropriate mechanisms to control the quality granted to those packets (e.g., proper configuration of schedulers on nodes (e.g. [[67][68])); (2) a mechanism to convey the application's QoS management information to the network nodes along the paths/trees (e.g., by means of appropriate signalling protocol such as Resource Reservation Protocol (RSVP) [69]). In particular, such control usually encompasses mechanisms for end-to-end admission control, resource reservation control, traffic flow control, packet scheduling, monitoring and maintenance of the QoS delivered. An example is the delivery of a sequence of audio and video to remote collaborating users in the ROMEO network.

From resource reservation control perspective, a network may agree on a traffic contract with the application software and reserve the required bandwidth, during a session establishment phase, in the nodes along the paths/trees that will be used for the traffic transport. Then, the achieved QoS performance in terms of data rate and delay may be monitored while the scheduling priorities are controlled in the network nodes to allow for dynamic QoS adaptation upon need. When an application terminates, the related resource reservation shall be released in order to be reused by future applications. However, maintaining the reservation states of each flow on every node along related paths/trees (e.g., IntServ using RSVP [69]) is not scalable. Therefore, QoS models based on Classes of Service (CoSs) [70][71] have been proposed to prevent performance degradations of the per-flow approaches. In class-based networks (e.g. DiffServ [72], Multiprotocol Label Switching - MPLS [73])), flows are classified into a set of CoSs at network borders (e.g. ingress routers) or central stations (e.g. Bandwidth Brokers), based on predefined policies regarding QoS, protocols, application types, etc. This way, while the reservation states must be maintained per-flow at the network border or the central stations for the purpose of fine-grained control, they are simply maintained per CoS on interior nodes inside the network. As a result, the control complexity is pushed to the edge nodes and leaves the interior nodes simpler. However, per-class reservations driven by per-flow signalling as in [74] introduce undue overheads since the control operations are triggered with the increasing number of session demands where excessive control messages not only put heavy processing load on interior nodes' CPU, but also consume more bandwidth, memory and energy, while they affect the session setup time.

Alternatively to per-flow QoS signalling control approaches, over-provisioning techniques allow reserving more bandwidth than currently required by a CoS. Hence, multiple flows can be merged on a path/tree without signalling the nodes on the path/tree as long as the over-reserved resources are still available. Also, there will be no need to signal the network to explicitly release resources when a session terminates. Thus, the control signalling and the related overheads can be significantly reduced to improve the system overall performance. However, in dynamic network scenarios, unpredictable cross-traffic load on links makes the approach very challenging, mainly in terms of waste of resources, session blocking and QoS violation. In literature, the technique proposed to over-reserve resources for aggregate flows in the Border Gateway Reservation Protocol (BGRP) [75] - a Sink-Tree-Based Aggregation Protocol is static and not suitable for dynamic network environment. The Simple Inter-Domain QoS Signalling Protocol (SIDSP) [73] system over-provisions the so-called virtual trunks of aggregate flows. However, bandwidth over-reservation based on predictive algorithm (e.g. based on past history) without appropriate mechanism to dynamically control the shared resource between the trunks is not efficient, since it can lead to waste of bandwidth. The Multi-user Aggregated Resource Allocation (MARA) [76] proposed functions to dynamically configure over-reservation parameters for CoSs. It deals with the waste of bandwidth by attempting to grant a congested CoS with a portion of residual over-reservation of remaining classes. As studied in Class-based resource Over-pRovisioning (COR) [77], the MARA scheme also wastes resources, especially when the network is near congestion. Besides, COR shows improved results when compared with MARA in terms of waste and session blocking. However, its algorithm does not include any architectural control protocol to efficiently deal with high cross-traffic challenges such as in the ROMEO P2P network.

Considering the aforementioned limitations of existing over-reservation proposals and the relevance of the approach in the scope of ROMEO, it is important to investigate: (1)

appropriate architectural control protocol to allow exploiting resource over-reservation to improve control performance with reduced overheads; (2) ways for dynamic configuration and readjustment of the over-reservations parameters for CoSs in a way to increase the resource utilization without QoS violation or unnecessary waste of resource; (3) mechanisms to assist P2P mobility since over-reservation allows for lower session setup time, which is important to achieve timely handover for mobile users.

#### 4.4.5 Mobility in P2P networks

The conjunction of mobility and P2P systems has been studied in the past. The term “mobility” has been given two different meanings in this context:

- An infrastructure less system composed of mobile units, also called a Mobile ad-hoc Network (MANET)
- An infrastructure facilitating the mobility of a mobile unit that is connected via the infrastructure to the internet.

In the first case, the P2P overlay is deployed over an infrastructure less MANET, while in the second the scenario incorporates P2P networks coping with wireless handovers into infrastructure wireless networks.

Most of the research work, especially in algorithms and protocols for P2P in mobile networks, relates to the MANET scenario, like for example the work described in [78], [79], and [80].

ROMEO’s research interests are into the second type of mobility. Considering mobility in infrastructure networks, most related work focuses on WLAN. The paper [81] investigates some problems that arise due to the design aspects of BitTorrent, and P2P networks in general, being incompatible with typical characteristics of wireless and mobile environments. In particular, it considers Power Saving Modes, mobility, use of simultaneous bi-directional TCP connections, identity management for keeping the incentives under mobility, interaction between the rarest-first fetching strategy and the frequent disconnections of mobiles. Then, the paper presents a backward-compatible extension to BitTorrent, which controls the slow-start characteristics of TCP, the use of session-based incentives, and increases the number of chunk uploads that are requested for peers having frequent disconnections.

The work in [82] compares PPLive [83], SopCast [84], and TVAnts [85], which are three of the most successful P2P-TV systems, to check the network awareness of the systems. The experiment involved deploying 40 peers over 7 sites in 4 different European countries (Italy, France, Hungary, and Poland). Even though the paper is not concerned with mobility, its comparison methodology can be applied to ROMEO project, and the criteria of network awareness described in the paper can be considered in ROMEO's overlay design.

The general structure of ROMEO network is similar to the one proposed and analyzed in [86], where an internet-wide P2P overlay deliver multimedia data to end routers using mainly wired connections, and then data consumers are attached to such end routers via wireless connections. The P2P network is divided into clusters, each cluster centred on a super-peer. The paper uses the concept of “super-peer”, which is in general a peer supposed to have better bandwidth and longer lifetime. We can consider that most super-peers are installed on the end routers, to move responsibilities from the wireless terminal to them. The super-peers will index the resources available in the peers, and they will enforce a control on queries of resource discovery, to limit the expectations imposed on normal peers. Normal peers are supposed to leave the network frequently and hence the query for resource discovery should give preference to either a resource on a super-peer, or a resource on a normal peer in the same cluster of the peer performing the query.

[87] presents a comparison of different P2P streaming systems, all of them loosely based on BitTorrent, which are used to stream multimedia content. The authors compare SopCast [84], PPLive [83], TAVnts [85], PPStream [88], and ESM [89], which differ for the transport protocol they employ to deliver multimedia content. The paper considers a scenario where mobile terminals perform handovers between different networks. The main result of the paper is that the systems that employ TCP are still able to function well, since they use short-lived TCP connections, which are compatible with usage in wireless networks. Thus, ROMEO project will be able to decide which transport protocol to use to deliver data, since it has been shown that even TCP will not cripple the P2P system. Regarding mobility, the paper compares a scenario where a peer maintains its IP address, and then it considers Mobile IP, where peers have got tunnels with their original location to receive data from the overlay when changing their IP address in a new network.

Finally, regarding the protocols that were considered for P2P streaming, IETF is working towards a standardized P2P Streaming Protocol (PPSP) [90]. Their work is still preliminary, but it considers also P2P streaming in cellular mobile networks, and it can be well into the targets of ROMEO to impact on the IETF methodology to consider also handovers in the mobile P2P scenario.

### 4.4.6 Multi-home mobile and Media Independent Handover

The integration of complementary wireless technologies in order to achieve “anywhere, anytime and on any device” ubiquitous services is the core idea of Next Generation Networks (NGN). Mobility Management (MM) is a major factor towards a seamless provision of multimedia applications in such a heterogeneous environment. The Mobile Internet Protocol (MIP) is the most well known Network-layer solution for MM. In Internet Protocol each endpoint is identified by a unique IP Address. When a mobile node (MN) changes its point of attachment its IP Address changes too, losing its TCP/IP connection. The MIP was introduced to alter this problem by allowing the MN to roam while keeping its connectivity to the Internet.

#### 4.4.6.1 Definitions

Handover is the process of network association of a Mobile Node while it moves across different networks. The mobility concept is defined in different ways by the relevant standardization for a:

1. Nomadism: It is the ability of the user to change his network point of attachment while, he is on the move. When the network point of attachment is changed, the user's service session is completely stopped and it can be resumed later on.
2. Session Continuity: It refers to the ability that the end-user’s terminal can switch to a new network point of attachment while maintaining the ongoing session from the old point of attachment to the new one. This may include a session break and resume, or a certain degree of service interruption or loss of data while changing to the new access point.
3. Seamless handoff: the handoff algorithm should minimize the packet loss. It is sometimes referred to as smooth handoff. Transparent migration of ongoing data flows between two access points belonging to independent heterogeneous technologies is achievable, and tools and mechanisms for supporting this type of mobility should be placed within the next generation networking architectures.

#### 4.4.6.2 Handover taxonomy

The handover taxonomy can be grouped into three general categories:

1. Intra-system Handover: The user associates itself with a new network of attachment (Access Point (AP)) that it is the same technology with the previous one (AP)
2. Inter-system Handover: The user associates itself with a new network of attachment (AP) that is different (heterogeneous) as compared with the previous one. It is also referred as vertical handover
3. Inter-operator Handover: It refers to the ability to extend the connectivity of a session in a location that is different from the home location where the service has been registered

#### 4.4.6.3 IETF mobility management protocols

Handling mobility in IP based networks regards the maintenance of Network Point of Attachment, while the end-user moves across different networks. As proposed, the mobility management protocols can be grouped at the following categories.

##### 4.4.6.3.1 Mobile IPv4 (MIPv4)

The MIPv4 [91] was introduced to allow the MN to gain access to the Internet by utilizing two IP Addresses, one for identification called Home Address (HoA), the other for routing called Care of Address (CoA). According to MIPv4, every domain must have a router called Home Agent (HA), so as to allow roaming of its users and a router called Foreign Agent (FA), so as to accept visitors. Every time a MN enters a Foreign Network, it registers with the FA and obtains a CoA. Then the FA informs the HA of the MN’s current CoA. Every time a Correspondent Node (CN) establishes a connection with the MN, it sends packets to MN through the HA. This deficient way of communication is called Triangular Routing and is one of the main downsides of the protocol. Some other drawbacks of the protocol are the limited number of IP addresses and the vulnerability against malicious attacks.

##### 4.4.6.3.2 Mobile IPv6 (MIPv6)

The MIPv6 [92] was introduced as a standard that surpasses the drawbacks of the previous version. A fundamental sub-protocol of Mobile IPv6 is the Return Routability Protocol (RRP). The RRP is a procedure by which a CN carries out a minimal verification that a MN owns an address (HoA) and is reachable at another (CoA). With the RRP the CN can be informed for the CoA and send packets directly to the MN, solving the problem of Triangular Routing. Moreover, MIPv6 also eliminates the use of FA as the MN is capable of performing its own mobility procedures, involves more sophisticated security procedures and provides a wider range of IP addresses.

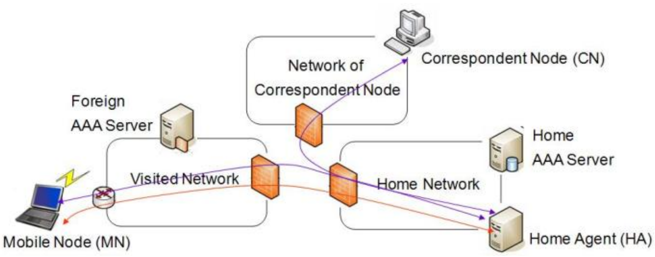


*Figure 20 - MIPv4*

#### 4.4.6.4 Route optimization in mobile IP (MIP-RO)

MIP-RO [93] solves the Triangular Routing problem and makes feasible direct communication between MN and CN. Using the base Mobile IP protocol, all datagrams destined to a MN are routed through the MN's HA, which then tunnels each datagram to the MN's current location. With MIP-RO, the MN can send a registration packet to the CN to inform it for its new CoA. CN may cache the binding of a MN, and then tunnel their datagrams for the MN directly to CoA, bypassing the MN's HA. Route Optimization is supported by MIPv6 while it is a proposed extension for MIPv4.

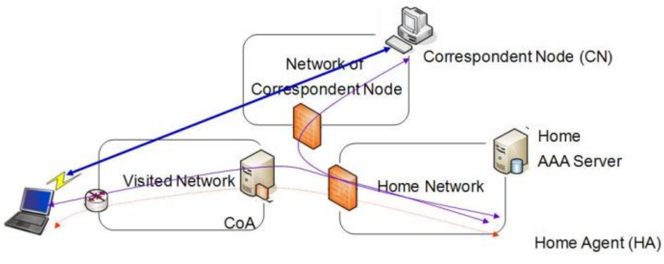


*Figure 21 – MIP-RO*

#### 4.4.6.4.1 Local mobility

Although Mobile IP enables the MN to maintain its connectivity to the Internet when it roams across IP subnets it has been slowly deployed in real networks as it suffers from high Handover Latency. During handover, there is a period during which the mobile node is unable to send or receive packets because of link layer handover latency and IP protocol operations. This handover latency resulting from standard Mobile IP is often unacceptable to real-time traffic. Separating local from global mobility reduces adequately the handover latency. The following extensions are proposed standard of the IETF designed to deal with local mobility and compatible with both versions of Mobile IP protocol.

#### 4.4.6.4.2 Hierarchical mobile IP (HMIP)

HMIP [94], [95] introduces hierarchical mobility management in MIP, and by that separates local from global mobility. In HMIP, global mobility is managed by the MIP protocols, while local handovers are managed locally. To do so, a new Mobile IP node called the Mobility Anchor Point (MAP) is used, which is basically a local HA. The MN registers to the local MAP rather than the HA and CN and, as a result, only one registration message needs to be transmitted by the MN before traffic from the HA and all CNs is re-routed to its new location. Consequently, HMIP reduces Mobile IP signalling load and improves the Mobile IP handover latency. Moreover, the MN hides its location (CoA) from both the HA and CNs while using Mobile IP route optimization. All in all, HMIP minimizes the impact on Mobile IP or other IP protocols offering additional reliability and scalability.

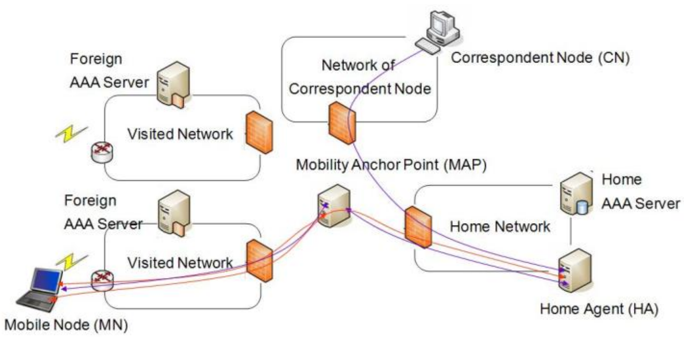


*Figure 22 - HMIP*

#### 4.4.6.4.3 Mobile IP Fast Handovers (FMIP)

FMIP [96],[97] enables the MN to quickly detect that it has moved to a new subnet by providing the new access point and the associated subnet prefix information when the MN is still connected to its current subnet. The protocol introduces signalling message exchange, so that handover can be initiated on the previous network and arriving packets can be tunnelled from the current to the new location of the MN while the MN performs a link layer handover. As soon as link connectivity is established, the new Access point forwards arriving and buffered packets to the MN. Therefore, the MN is capable of receiving packets before initiating a layer 3 handover and reducing the overall handover latency.

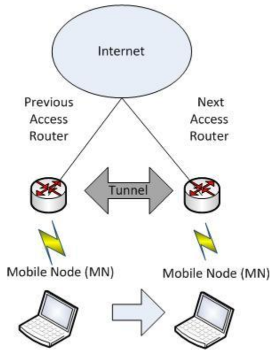


*Figure 23 – FMIP*

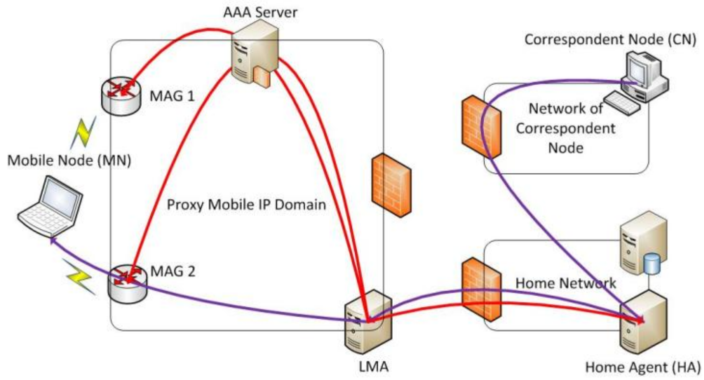


*Figure 24 – PMIP*

#### 4.4.6.4.4 Proxy Mobile IP (PMIP)

PMIP [98], [99] is a protocol implementing MM procedures in network part without involving the MN. For that purpose, two new functional entities are introduced in PMIP, the Local Mobility Anchor (LMA) and the Mobility Access Gateway (MAG).The LMA is the topological anchor point for the MN’s home network prefix, receiving any packets that are sent to the MN by any node in or outside the PMIP domain. The LMA is Home Agent with enhanced capabilities for supporting PMIP. The MAG is a new functional entity that emulates the MN’s home link on the access link. To do so, the MAG sends Router Advertisement messages, containing the MN’s home network prefix. Typically, it is a function running on an Access point. PMIP supports multi-homing by allowing the MN to connect to a domain through multiple interfaces for simultaneous access. By this means, PMIP does not require move detection and address configuration procedures to be performed during handover reducing handoff delay and signalling. Additionally, PMIP is compatible with current devices and reduces signalling exchange through the wireless link.

### 4.4.6.5 Media Independent Handover framework

In the context of NGNs, it is important to define the framework that will allow mobility protocols to handle ongoing session and demanding services uniformly among heterogeneous link-layer interfaces. Media Independent Handover (MIH) framework can handle these upcoming challenges [100]. IEEE Working Group has recently proposed IEEE 802.21 standard to enable handover and interoperability between heterogeneous networks with context-awareness in mobile terminals [101].

One of the main ideas behind IEEE 802.21 is to provide a common interface for managing events and control messages exchanged between networks devices that support multiple interfaces, both wired and wireless. The concept is also supported by IETF MIP shop WG (Mobility for IP: Performance, Signalling and Handoff Optimization), which aims to optimize Fast Handover IP Mobility across heterogeneous wireless networks. The contribution of Media Independent Handover framework is centred on the following objectives:

1. Inter-System (Remote Access Technology (RAT)) Handover: The MT will be capable of performing vertical handovers between heterogeneous wireless access technologies. The handover may be triggered/initiated either by the MT or the RAT.
2. Seamless handover-service continuity: One of the main objectives of MIH functionality is the continuation of the service during and after the handover procedure.
3. Support for Multi-homing: Tackles issues regarding multi-interface operation and application and flow mobility.
4. QoS aware handover: Provides QoS guarantees for both best effort and real-time traffic. This is accomplished by determining the available bandwidth at each RAT. Furthermore, co-channel and inter-channel interference may limit the transfer rates of a wireless network, degrading QoS.
5. QoE and Session Characteristics adaptation: Upon a vertical handoff, session characteristics may be updated. As an example, in case of delay-tolerant applications these functions may necessitate the media adaptation functionalities.
6. Security Support: Handover signalling should be secured against possible attacks and in the same time, security primitives should be transferred from the old Point of Attachment (PoA) to the new one.

The MIH framework collects information from different layers and selecting the Best Available conditions. The following section illustrates the proposed MIH framework.

##### 4.4.6.5.1 Triggering/Information services

The services provided by MIH function include:

1. Media Independent Event Service (MIES): Detects and delivers triggers from both local and remote interfaces corresponding to dynamic changes in link characteristic, status and quality.
2. Media Independent Command Service (MICS): Enables higher layers to control lower layers (physical, link layer). Higher layers may control the reconfiguration or selection of an appropriate link through a set of handover commands.
3. Media Independent Information Service (MIIS): Provides a framework and corresponding mechanisms for discovering and obtaining information of existing networks able to facilitate handovers. The MIIS provides link layer parameters such as channel information, MAC address and security information of PoA.

The MIH function and its relationship with upper and lower layer elements are shown in Figure 25.

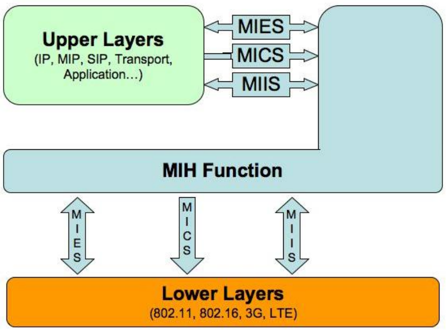


*Figure 25 - Media Independent Handover function key services*

#### 4.4.6.5.2 MIH information service

The MIIS offers information that is needed to perform the handover and is linked to the appropriate module of each MIH enabled entity. It collects the following statistics:

1. Access Network:
   a) List of available networks
   b) List of Authorized Users
   c) Target QoS parameters for on-going sessions
   d) Number of Connections per Point of Attachment
   e) Load Factor (uplink, downlink)
   f) Available bandwidth
2. MT
   a) Discovered Networks
   b) Context Transfer Parameters
   c) Physical (PHY) parameters
   d) Network QoS Parameters
   e) Measured QoE Parameters

#### 4.4.6.5.3 MIH event service

The current 802.21 specification defines events that may be relevant to handover decision, which may originate from PHY, Media Access Control (MAC) or MIH either at MT or network PoA.

The MIES collects the information provided from MIIS and process it in order to trace any change in physical, network or application parameters that possibly affect the ongoing session. Whenever such changes are detected, MIES creates the appropriate triggers to the media independent handover module. In the heterogeneous wireless network environment of our approach, a hierarchical structure of decision factors, each of which has a related trigger event, is considered.

The decision factors and trigger events that are currently being considered are the following:

#### 4.4.6.5.4 Link information

The trigger event for link information is Link_SNR_Change and includes:

1. Trigger Event Type: {Link_SNR_Change}
2. Source: {Lower layers}
3. Parameter: {Network identifier, SNR threshold, Bandwidth}

Link_SNR_Change indicates changes of the SNR and the corresponding signal strength during an ongoing connection. Network identifier refers to the current network ID of trigger source. SNR value indicates the threshold value of the SNR below which handover is imminent. A low SNR threshold value will cause degradation of the perceived QoS, while on the other hand high threshold value will cause frequent vertical handovers between heterogeneous access networks.

#### 4.4.6.5.5 Application QoS

Application_QoS_Change event indicates changes in the perceived quality and includes:

1. Trigger Event Type: {Application_QoS_Change}
2. Source: {Upper layers}
3. Parameter: {Network identifier, relative PSNR (rPSNR), R-factor, Traffic class}

Relative PSNR is an estimation of the received video stream quality measured as PSNR at the receiver. This parameter is determined by the MT and is based on the video source coding characteristics and the current network conditions in order to estimate the impact of the video impairments at the receiver, using the packet losses, delay and encoding rate, as described in [102], [103]. The coding parameters are provided to the MT by the Real Transport Control Protocol Extended Report (RTCP-XR) that is sent periodically by the video server. RTCP-XR defines block types for use in Quality of Service reporting for video over IP [95]. In addition, R-factor indicates the perceived quality of service of VoIP sessions and it depends on the voice codec, the network impairments, end-to-end delay and delay jitter.

##### 4.4.6.5.6 Network context

Network_Context_Change event indicates changes in the network and includes:

1. Trigger Event Type: {Network_Context_Change}
2. Source: {Upper layers}
3. Parameter: {Network identifier, Network context}

##### 4.4.6.5.7 MIH command service

The higher layers use the MICS primitives to control the functions of the lower layers. MICS commands are used to gather information about the status of connected links, as well as to execute higher layer mobility and connectivity decisions to the lower layers. The mobility management protocols combine dynamic information regarding link status and parameters, provided by the MICS with static information regarding network status or other higher-layer service information provided by the MIIS, in order to help in the decision making. MIH commands can be both local and remote. These include commands from upper layers to the MIH and from the MIH to lower layers. Some examples of MICS commands are MIH Poll, MIH SCAN, MIH CONFIGURE and MIH SWITCH. The commands instruct an MIH device to poll connected links to learn the most recent status, to scan for newly discovered links, to configure new links and to switch between available links.

All commands are designed to help in the handover procedure, but the routing of user packets is left to the mobility management protocols.

### 4.4.7 Media packetisation for P2P networks

MVC exhibits packetisation similarities with H.264 SVC [104]. There are three transmission modes for multiple-view video, referring to Single Session Transmission (SST), Multi-Session Transmission (MST) and MANE-based transmission as shown in Figure 26 and Figure 27. In SST, all MVC packets are carried in a single RTP session utilizing only a single transport address/port. In MST two or more RTP sessions are used to carry the MVC data. The MANE-based system is more complicated and it is described below.

In the case of SST, each RTP session may either carry the base view bit stream or the base view with a number of different non-base views using a point-to-point unicast session. For each client, the server “constructs” a sequence of RTP packets depending on the terminal needs by aggregating Network Abstraction Layer (NAL) units of the needed views.

In the case of MST, each RTP session may carry either the base view bit stream, or more than one non-base view bit streams, or eventually the base view with a number of non-base views. MST should be used in a multicast RTP session where different receivers may request different views of the multiple-view bit stream.

In the case of unicast applications, each MVC video view is transmitted in a different IP address (Single Session Transmission, it is an IP Multicast) while in multicast applications each MVC video view is transmitted in its own IP multicast group. The number of the RTP sessions must be at least equal to the number of different views, otherwise a few of the RTP sessions encapsulate more than one non-base views.

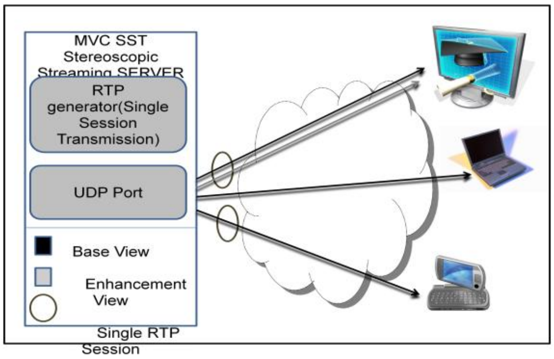


Figure 26 - MVC SST transmission

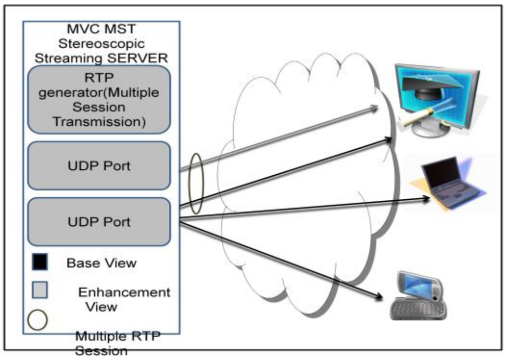


Figure 27 - MVC MST transmission

The third option includes MANE, as illustrated in Figure 28. MANE generally resides in the path between the server and the clients [105]. In a MANE-based system, the server keeps using MST transmission. However in this case, MANE collects all RTP sessions and depacketise them. It then customizes NAL Units according to the client's needs through Adaptation Decision Taking Engine (ADTE) and aggregates new packets for SST to clients through a single transport address using in a single RTP session [106]. The need of such an intermediate system comes from the existence of a lot of limitations in a real network environment such as the existence of firewalls or Network Address Translation (NAT) protocol. MANE is able to take all important information through RTP and NAL Unit Headers.

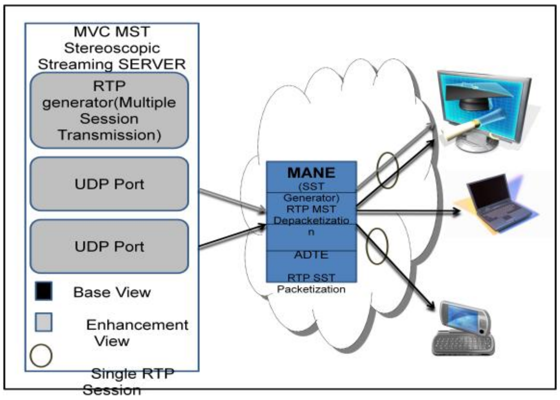


Figure 28 - Media-Aware Network Element (MANE) transmission

RTP packets may use any of the three payload structures as Figure 29 shows. When Single NAL Unit mode is used, each RTP packet contains only one NAL Unit. Aggregation packets contain more than one NAL Units by merging them and conversely Fragmentation Units contain part of a NAL Unit. Single NAL Unit is allowed to contain NAL Unit types from 1 to 23 and 30 as well as 31. For the encapsulation of each NAL Units in RTP packet, an RTP header must be inserted before each NAL Unit Header. The RTP header shall always include the first 12 octets composed of all the bit fields of the first 4 octets as well as the timestamp bit field and the Synchronisation Source (SSRC) as shown in Figure 29, while it may contain more octets for specifying Contributing Source (CSRC) depending on the needs of the application [104].

Aggregation packets were introduced in order to avoid problems among heterogeneous networks using different MTU size. By aggregating NAL Units, overhead is reduced at the RTP headers. Each NAL Unit being carried in an RTP packet is encapsulated in an aggregation unit first. There are five versions for aggregation units as shown in Table 4. The first four STAP-A, STAP-B, MTAP-16, and MTAP-24 are used also to packetize H.264/AVC data, while for H.264/MVC a new type of aggregation unit named NI-MTAP [105] has been introduced. Regardless of the RTP Header, every aggregation unit encapsulated in one aggregation packet contains one additional aggregation header which consists of 1 octet and has the form of a H.264/AVC NAL Unit Header. Depending on the version of the aggregation unit that has been used, some additional information is introduced regarding whether the NALs refer either to base or non-base view. In case of using STAP-A there is only one additional header (2 octet long), which indicates the size of the NAL Unit [104].

Fragmentation Units allows the NAL Units to be fragmented at the application layer which provides some advantages. For example when bit errors occur during transmission, fragmentation of a NAL Unit allows the utilization of FEC/retransmission mechanism only at the specific fragment of the NAL Unit and not in the entire Single NAL Unit [107]. Each fragmentation unit must be transmitted consecutively with upward sequence numbers. There are two versions of fragmentation unit as shown on Table 4, FU-A and FU-B. The basic difference between them is that the latter version uses the Decoding Order Number (DON) mechanism in order to retrieve the correct decoding order of the NAL Units instead of the former one who uses timestamps [105]. Both versions except from the RTP header employ two additional headers. The first one, named FU-indicator, consists of 1 octet and has the form of a H.264/AVC NAL Unit Header. The second header, namely FU Header, consists of 1 octet [104]. In the case of FU-B, one more additional header named DON must be used which consists of 2 octets. It should be noted that the H.264/AVC NAL header after the fragmentation of the NAL Unit must be deleted, while the information of the first three bits (2 bit fields) moves to FU indicator and the last 5 bits (Nal_Unit_Type) moves to the last 5 bits of FU Header. During the depacketisation procedure, the H.264/AVC NAL Unit Header is reconstructed using the FU headers. The next 3 octets containing the NAL Unit Header extension lies at the payload.

For more details for the RTP Header, and for basic RTP packet payload structures introduced and used also for H.264/AVC refer to [107]. For new payload structures (NI-MTAP) refer to [105].

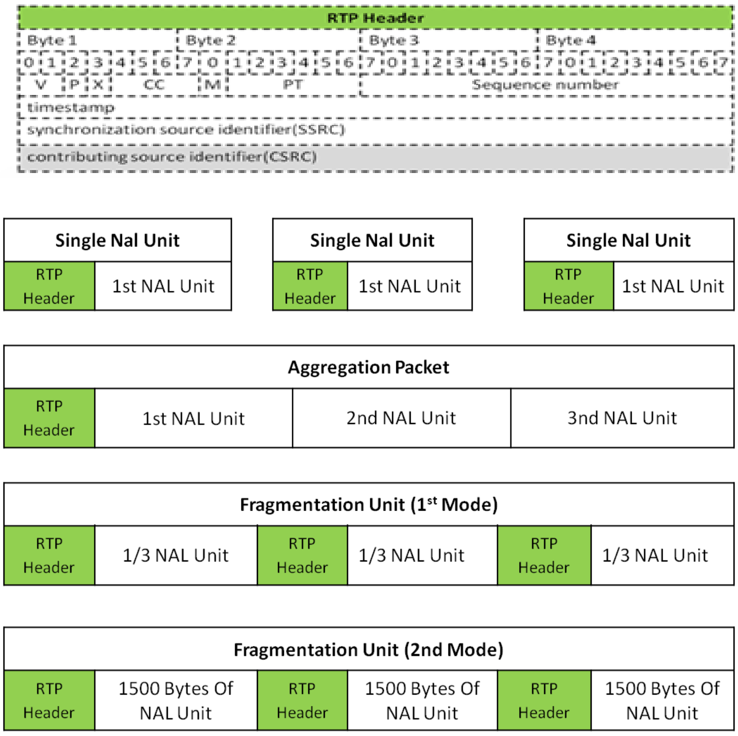


*Figure 29 - RTP packet payload structures for MVC*

| *Type* | *Packet Type* | *Packet Type Name* |
|---|---|---|
| 0 | Reserved | |
| 1 to 23(-14,-20) | NAL unit | Single NAL unit packet |
| 14 | NAL Unit | Prefix NAL Unit packet |
| 20 | NAL Unit | coded slice of MVC extension |
| 24 | STAP-A | Single-time aggregation packet |
| 25 | STAP-B | Single-time aggregation packet |
| 26 | MTAP-16 | Multi-time aggregation packet(2 byte offset) |
| 27 | MTAP-24 | Multi-time aggregation packet(3 byte offset) |
| 28 | FU-A | Fragmentation unit |
| 29 | FU-B | Fragmentation unit |
| 30 | NAL unit | PACSI NAL Unit packet |
| 31 | NI-MTAP | NI-MTAP NAL Unit packet |

*Table 4 - MVC RTP packet types specified from IETF*

There are three basic packetisation modes for SST namely Single NAL, non-interleaved mode and interleaved mode. Single NAL Unit mode is being used for systems defined in accordance with the recommendation of ITU-T for H.241 [108]. Non-interleaved-mode is used for systems that are not compliant with the above recommendation and finally interleaved mode is suitable for systems that do not have latency restrictions and do not need decoding order transmission. All NAL units from type 1 up to type 29 can directly be used as packet payload for all the above modes while for the newly introduced types, a PACSI NAL Unit (type 30) can be used directly as payload only in Single NAL Unit Mode and a NI-MTAP can only be used in non-interleaved mode.

Four new modes have been introduced for MST. The newly introduced modes namely are Non-interleaved timestamp based mode (NI-T), Non-interleaved cross-session decoding order number (CS-DON) based mode (NI-C), Non-interleaved combined timestamp and CS-DON mode (NI-TC) and finally Interleaved CS-DON (I-C) mode [94]. The new modes re-use the three basic modes for SST for each of their individual RTP sessions. The two basic differences of the above modes is whether or not decoding order transmission must be used and what mechanisms they use in order to recover the correct decoding order through multiple sessions [106], [109]. NI-T, NI-C, NI-TC modes contrary to I-C do allow random packet transmission and for that reason they are targeted for low latency systems. NI-T use timestamps to recover the correct decoding order against NI-C and I-C who uses CS-DON mechanisms. The NI-TC mode provides both of the mechanisms referred above. If a system needs to transport both base and non-base views at low latency mode then one of the three first modes must be used in parallel with non-interleaved mode while for higher latency, I-C in parallel with Interleaved mode must be used for transport both base and non-base views [109].

## 4.5 Terminal equipment virtualisation

### 4.5.1 Problem statement

Since the introduction of Internet Protocol Television (IPTV) services by access network providers in their triple play commercial offers (Voice, IPTV and Internet), there has been the need for adding more and more infrastructure within the user's home to deploy such services. This trend started for the digital subscriber loop and cable accesses and continues in the deployment of fibre to the home, and is also common in satellite television services.

The addition of more and more devices and with more and more advanced functionalities within the user's home increases considerably the deployment costs: service providers are usually responsible for supplying the end to end service – i.e. including not only the network connectivity devices but also the needed Set-Top-Boxes, recorders, etc. – and, as such, they have not only to deal with initial investment in the devices, but also with the operational costs that such devices might bring and the incidents in the service.

This increase in capital and operational expenditures is the main driver for service providers to look for alternative solutions to the deployment of advanced devices within the user's home.

In the aim of reducing the number of devices, one of the initiatives is to analyze the possibility of developing a new access network architecture based on the virtualisation within the network of the layer-3 home routing gateway functionality. These way providers could remove the Broadband Access Routers at customer premises.

Avoiding any CPE (Customer Premises Equipment) and the new Layer2 Access network architecture could have some more advantages for the company such as:

- CAPEX Savings. Service providers will not only save on these installation but also in future hardware upgrades for new services (IPv6 functionality for example).
- OPEX Savings. Operations will be simplified as the physical router will not break. Service providers will have a complete Layer 2 visibility of home networks. Software upgrades will be done at the service providers' premises.
- Improve the easiness of development and deploy new services over the network.

### 4.5.2 State of the art

Current state of the art in home and access network architectures [110], [111], [112], [113] relies on a layer-3 device (the routing gateway) in home premises that performs different functionalities depending on the access technology (xDSL modem, Gigabit Passive Optical

Network (GPON) modem, etc.) and/or the services provided (Internet access, IPTV, VoIP, etc.). Such functionalities include NAT, local DHCP, Internet Group Management Protocol (IGMP) proxy-routing, Point-to-Point Protocol (PPP) sessions, routing, etc. This routing gateway is the base equipment for Telco services, enabling Internet access. Advanced services rely on additional devices (e.g. IPTV needs a set-top-box, generic Value Added Service (VAS) need a Home Gateway). This model, based on the installation of different devices in home premises, implies a high cost for service providers in both initial installation and operational support as they are typically responsible for the end-to-end service.

This problem is even worse in GPON deployments as nowadays its home premises modems, the (Optical Network Terminal) ONTs, do not include layer-3 functionality themselves and delegate those in a separate routing gateway within the users' homes. As a result, an additional device (the ONT) must be installed in the customer's premises for fibre accesses.

The fact of having a new (layer-2) device in the home, the ONT, is a driver to propose a different paradigm for the home network. Since fibre deployments are in an early deployment stage, this could be the right moment for a radical shift in the way broadband fibre services are delivered.

The architecture proposal for home virtualisation is thus specially focused on GPON accesses, and is based on the following requisites:

- Home and access networks should be layer-2 based networks with layer-2 visibility among them, where the need for routing gateways in the home network is suppressed.
- Installation and maintenance procedures should be simplified and Plug & Play client architecture should be achieved.
- Devices and services should have the capacity of self-provision.
- The vast majority of layer-3 functionality should be moved from the home network to the service provider network, whether distributed across the network or virtualised in a single node.
- Devices in the home network should have visibility among them to minimize the bandwidth usage in the access network.

### 4.5.3 Technology advancement

The objective of this section is to present and elaborate several alternatives for the different network architecture in the area of Home Routing Gateway Virtualisation. Such a virtualisation should bring a reduction in service providers' capital and operational expenditures, as well as ease the deployment of new services, since they are not expected to be constrained by the hardware and software capacities of the equipment available in the customer's premises, but based on equipment located in the network that might get advantage of the layer-2 visibility of the home network from the access network.

### 4.5.4 Solution selection

In order to remove and virtualise the Broadband Access Router from customer premises, all the necessary functionality has to be carefully studied and implemented at any point of the Network so the perception of the user is, at least, as good as if the physical router would be at his home.

1. For instance, the DHCP server that will allocate IP addresses to user terminals is now located in the access network and different ways of addressing have been considered. Although differentiate the addressing range among different homes would ease the Optical

Line Termination (OLT) to identify user traffic origin, it is been decided for the sake of scalability to use an Overlapped address system. In turn, it is been decided not to apply special addresses to specific devices to define a determined behaviour preferring otherwise a multiservice based addressing system.

2. Another key concept is the switch the user has in the home router. Without the physical router, this capability has to be assumed by the ONT. Regarding this new role of the ONT it is been decided that ONT ports should not be dedicated to specific services or devices, so they should be multiservice ports to allow any service/device to be connected at any port.
3. As a consequence of previous requirements, it is not feasible to maintain the current delivery system in which 1 service is delivered through 1 VLAN because this would mean that each port would be associated to 1 service. Moreover, the delivery of the different services to the home throughout separated VLANs usually is a complication in the service provisioning and especially in this architecture in which it was the Broadband Access Router the entity in charge of routing user traffic on the corresponding VLAN. For all the above mentioned it is decided to deliver triple play service over a single VLAN (on Ethernet ports, Plain Old Telephone Service (POTS) port will send the traffic through another VLAN) .Due to 1 VLAN is used for all services, the OLT will forward the traffic through the MAN (OLT operating in QinQ mode) instead of rerouting the different VLANs to the different services.
4. Finally, another functionality shifted to the access network is the Network Address Translation (NAT). There were different possibilities with the NAT performance depending on the addressing system used. It is been decided to perform NAT to the internet service. The NAT server could be located somewhere in between of the BRAS and the OLT. Most likely within the BRAS where the public IP address is allocated to the Virtual Broadband Access Router.

### 4.5.5 Conclusion

This section has presented the main drive for thinking in device virtualisation as a good opportunity to reduce the device complexity within the home network in a fibre deployment, reduce the expenses associated to these deployments, and get into the home LAN offering L2 services without additional equipment. To achieve this end, several strategies of network architecture virtualisation have been under study selecting the alternative that suits better with the objectives defined.

# 5 CONCLUSION

In this document main reference scenarios are defined and an initial set of service requirements have been specified. An overview of the state-of-the-art technologies is also presented to identify the possible research directions within ROMEO project.

This document will form a basis for use cases and a top to down architecture of the whole platform to be designed by the ROMEO partners, which will be a second deliverable of the project at the end of May 2012.

# 6 REFERENCES


[1] DVB Document A122 (July 2011): Digital Video Broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2).

[2] M. Stoufs, A. Munteanu, J. Cornelis, and P. Schelkens, ""Scalable Joint Source-Channel Coding for the Scalable Extension of H.264/AVC", IEEE Transactions on Circuits and Systems for Video Technology, vol.18, no.12, pp.1657-1670, Dec. 2008.

[3] http://www.dvb.org/technology/dvbt2

[4] http://dvb.org/technology/fact_sheets/DVB-S2_Factsheet.pdf

[5] http://www.dvb.org/technology/dvbc2

[6] http://www.dvb.org/technology/dvb-ngh

[7] Sandvine Intelligent Broadband Networks: "Global Internet Phenomena Spotlight – Europe, Fixed Access, Sping 2011", 13 May 2011.

[8] B. Cohen, "Incentives Build Robustness in BitTorrent", 2003.

[9] Sandvine Intelligent Broadband Networks: "2010 Mobile Internet Phenomena Report", 19 March 2010

[10] http://www.tribler.org

[11] http://www.coolstreaming.us

[12] http://www.p2p-next.org/

[13] http://napa-wine.eu/cgi-bin/twiki/view/Public

[14] http://www.muscade.eu/

[15] http://www.diomedes-project.eu/ "DIOMEDES Deliverable 4.2 Streaming and Networking", 2010.

[16] N. A. Dodgson "multi-view autostereoscopic 3D display" presentation

[17] R. Y. Tsai, "A versatile camera calibration technique for high accuracy 3D machine vision metrology using off -the-shelf TV cameras and lenses", IEEE Journal of Robotics and Automation 3 (1987): 323–344.

[18] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations", Proceedings of the 7th International Conference on Computer Vision (pp. 666–673), Corfu, Sep. 1999.

[19] Z. Zhang, "A flexible new technique for camera calibration," IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000): 1330–1334.

[20] E. Reinhard, M. Ashikhmin, B Gooch, and Peter Shirley, "Colour transfer between images", IEEE Computer Graphics and Application (Sep/Oct 2001): 0272-1716

[21] Dense stereo matching algorithm benchmark: http://vision.middlebury.edu/stereo/eval/

[22] D. Sharstein, and R. Szelisky, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", International Journal of Computer Vision (2001).

[23] H. Hirschmüller, "Stereo Processing by Semiglobal Matching and Mutual Information", IEEE Transactions on pattern analysis and machine intelligence, vol. 30, no.2, Feb. 2008.

[24] S. T. Birchfield, B. Natarajan, and C. Tomasi, "Correspondence as Energy-Based Segmentation",Image and Vision Computing, 25(8):1329-1340, Aug 2007.

[25] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard", IEEE Trans. Circuits Syst. Video Technol., vol. 17, Special Issue on Scalable Video Coding, no. 9, pp. 1103–1120, Sep. 2007.

[26] A. Vetro, P. Pandit, H. Kimata, A. Smolic, and Y.-K. Wang, "Joint draft 8 of multi-view video coding", Hannover, Germany, Joint Video Team (JVT) Doc. JVT-AB204, July 2008.

[27] G. J. Sullivan, A. M. Tourapis, T. Yamakage, and C. S. Lim, “Draft AVC amendment text to specify constrained baseline profile, stereo high profile, and frame packing SEI message”, London, UK, Joint Video Team (JVT) Doc. JVT-AE204, July 2009.
[28] A. Vetro, T. Wiegand, and G. J. Sullivan, “Overview of the stereo and multi-view video coding extensions of the H.264/AVC standard”, Proc. IEEE, 2011.
[29] K.-J. Oh, S. Yea, A. Vetro, and Y.-S. Ho, “Depth reconstruction fi and down/up sampling for depth coding in 3-D video”, IEEE Signal Processing Letters, vol. 16, no. 9, pp. 747–750, Sep. 2009.
[30] Y. Morvana, P. H. N. de With, and D. Farin, “Platelet-based coding of depth maps for the transmission of multi-view images”, 2006.
[31] http://mpeg.chiariglione.org/meetings/daegu11/daegu_press.htm
[32] J. Blauert, “Spatial Hearing, the Psychophysics of Human Sound Localization”, Cambridge, MA: MIT Press, 2001.
[33] F. Rumsey, “Spatial Audio”, 2nd ed. Oxford, England: Focal Press, 2001.
[34] P. Mackensen, “Auditive localization head movements, an additional cue in localization”, Ph.D. dissertation, Technischen Universitat Berlin, Berlin, Germany, Jan. 2004.
[35] U. Zolzer, “DAFX - Digital Audio Effects”, 1st ed. Chichester, England: JohnWiley Ltd, 2002.
[36] C. I. Cheng, and G. H. Wakefield, “Introduction to head-related transfer functions HRTF’s: Representations of HRTF’s in time, frequency, and space”, J. Audio Eng. Soc., vol. 49 no. 4, pp. 231–249, 2001.
[37] Internet draft. [Online]. Available: http://www.holophone.com/
[38] Benefits of a soundfield system. Internet draft. [Online]. Available: http://www.soundfield.com/soundfield/soundfield.php
[39] D. Menzel, H. Wittek, G. Theile, and H. Fastl, “The Binaural Sky: A Virtual Headphone for Binaural Room Synthesis”, In: Tonmeistersymposium, 2005 57585960616263
[40] M. Bosi, “High-quality multichannel audio coding: trends and challenges”, J. Audio Eng. Soc., vol. 48, no. 6, pp. 588–595, 2000.
[41] C. C. Todd, G. A. Davidson, M. F. Davis, L. D. Fielder, B. D. Link, and S. Vernon, “AC-3: Flexible perceptual coding for audio transmission and storage”, Presented at the 96th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 1994.
[42] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, “ISO/IEC MPEG-2 advance audio coding”, J. Audio Eng. Soc., vol. 45, no. 10, pp. 789–814, 1997.
[43] J. Herre, C. Faller, S. Disch, C. Ertel, J. Hilpert, A. Hoelzer, K. Linzmeier, C. Spenger, and P. Kroon, “Spatial audio coding: Nextgeneration efficient and compatible coding of multi-channel audio”, Presented at the 117th Convention of the Audio Engineering Society, San Fransisco, CA, USA, Oct. 2004.
[44] F. Baumgarte and C. Faller, “Binaural cue coding-part I: Psychoacoustic fundamentals and design principles”, IEEE Trans. Speech Audio Process., vol. 11, no. 6, pp. 509–519, Nov. 2003.
[45] C. Faller and F. Baumgarte, “Binaural cue coding-Part II : Schemes and applications”, IEEE Trans. Speech Audio Process., vol. 11, no. 6, pp. 520–531, Nov. 2003.
[46] J. Herre, C. Faller, C. Ertel, J. Hilpert, A. Hoelzer, and C. Spenger, “MP3 Surround: efficient and compatible coding of multi-channel audio,” Presented at the 116th Convention of the Audio Engineering Society, Berlin, Germany, May 2004
[47] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, “Parametric coding of stereo audio”, EURASIP J. Appl. Signal Process., vol. 2005, pp. 1305–1322, 2005.

[48] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegard, “Low complexity parametric stereo coding”, Presented at the 116th Convention of the Audio Engineering Society, Berlin, Germany, May 2004.
[49] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, “Advances in parametric coding for high-quality audio”, Presented at the 114th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, March 2003.
[50] S. Quackenbush and J. Herre, “MPEG Surround”, IEEE Multimedia, vol. 12, no. 4, pp. 18–23, 2005.
[51] J. Hilpert and S. Disch, “The MPEG Surround audio coding standard [Standards in a nutshell]”, IEEE Signal Processing Mag., vol. 26, no. 1, pp. 148–152, Jan. 2009.
[52] G. Hotho, L. F. Villemoes, and J. Breebaart, “A backward-compatible multichannel audio codec”, IEEE Trans. Audio, Speech Language Processing, vol. 16, no. 1, pp. 83–93, Jan. 2008.
[53] J. Herre, H. Purnhagen, J. Breebaart, C. Faller, S. Disch, K. Kjorling, E. Schuijers, J. Hilpert, and F. Myburg, “The reference model architecture for MPEG spatial audio coding”, Presented at the 118th Convention of the Audio Engineering Society, Barcelona, Spain, May 2005.
[54] I. Elfitri, B. Gunel, and A. Kondoz, “Multichannel audio coding based on analysis by synthesis”, Proceedings of the IEEE, vol. 99, no. 4, pp. 657-670, April 2011.
[55] G. Theile, “Wave field synthesis – A promising spatial audio rendering concept”, Proc. Of the 7th Int. Conf. On Digital Audio Effects (DAFx'04), pp. 125-132, Naples, Italy, Oct. 2004.
[56] M. Hoffman, L. R. Beaumont, “Content Networking: Architecture, Protocols and Practice”, The Morgan Kaufmann Series in Networking, 2005.
[57] S. Banerjee, B. Bhattacharjee, and C. Kommareddy, “Scalable Application Layer Multicast”, 2002.
[58] Y. Chu, A. Ganjam, T. Ng, S. Rao, K. Sripanidkulchai, J. Zhan, and H. Zhang, “Early Experience with an Internet Broadcast System Based on Overlay Multicast”, Jun. 2004.
[59] X. Zhang, J. Liu, B. Li, and T.-S. P. Yum, “DONet/CoolStreaming: A Datadriven Overlay Network for Live Media Streaming”, 2005.
[60] M. Zhang, J.-G. Luo, L. Zhao, and S.-Q. Yang, “A Peer-to-Peer Network for Live Media Streaming using a Push-Pull Approach”, 2005.
[61] http://www. bittorrent.org/beps/bep_0003.html
[62] D. Vyzovitis and I. Mirkin, “Design and Implementation of a Scalable Real-time Streaming Protocol”, 2006.
[63] D. C. Vyzovitis, “An Active Protocol Architecture for Collaborative Media Distribution”, 2002.
[64] http://www.saracen-p2p.eu/
[65] R. P. N. R. S. Cruz and M. S. Nunes, “Scalable Video Distribution in Peer-to-Peer”, 2010.
[66] N. Ramzan, E. Quacchio, T. Zgaljic, L. Celetto, E. Izquierdo, F. Rovati, and S. Asioli, “Peer-to-Peer Streaming of Scalable Video in Future Internet Applications”, 2010.
[67] A. Demers, S. Keshav, and S. Shenker, “Analysis and simulation of a fair queueing algorithm”, ACM SIGCOMM’89, vol.19, Pages 1-12.
[68] S. J. Golestani, “A self-clocked fair queueing scheme for broadband applications”, INFOCOM’94, Networking for Global Communications, vol. 2, pp. 636-646, Canada.
[69] R. Braden, L. Zhang, S. Berson, and S. Herzog, “Resource ReSerVation Protocol (RSVP) --Version 1 Functional Specification”, RFC 2205, Sept. 1997.
[70] A. G. Ash et al., “Y.1541-QOSM -- Model for Networks Using Y.1541 QoS Classes draft-ietf-nsis-y1541-qosm-10”, Feb. 2010.
[71] ETSI TS 123 107 Universal Mobile Telecommunications System (UMTS); Quality of Service (QoS) concept and architecture.

[72] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Services”, IETF RFC 2475.
[73] E. Rosen, A. Viswanathan, and R. Callon, “Multiprotocol Label Switching architecture”, IETF RFC 3031.
[74] A. Neto, E. Cerqueira, A. Rissato, E. Monteiro, and P. Mendes, “A Resource Reservation Protocol Supporting QoS-aware Multicast Trees for Next Generation Networks”, In Proc. 12th IEEE Symp. on Computers and Communications, Aveiro, Portugal, Jul. 2007.
[75] P. Pan, E. Hahne, and H. Schulzrinne, "BGRP: A Tree-Based Aggregation Protocol for Inter-domain Reservations", Trans. of Communications and Networks Journal, vol. 2, pp. 157-167, Jun. 2000.
[76] A. Neto, E. Cerqueira, M. Curado, E. Monteiro, and P. Mendes, “Scalable Resource Provisioning for Multi-User Communications in Next Generation Networks”, Global Telecommunications Conference, 2008, IEEE GLOBECOM 2008.
[77] E. Logota, A. Neto, and S. Sargento, “COR: an Efficient Class-based Resource Over-pRovisioning Mechanism for Future Networks”, IEEE Symposium on Computers and Communications (ISCC), June 2010.
[78] A. Yu and S.T. Vuong, “A DHT-based Hierarchical Overlay for Peer-to-Peer MMOGs over MANETs”, Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International, pp.1475-1480, 4-8 July 2011.
[79] A. Mondal, S. Madria, and M. Kitsuregawa, “CLEAR: An Efficient Context and Location-Based Dynamic Replication Scheme for Mobile-P2P Networks”, Database and Expert Systems Applications, LNCS, pp. 399-408, ISBN 978-3-540-37871-6, 2006.
[80] M. Sbai, C. Barakat, J. Choi, A. Hamra, and T. Turletti, “Adapting BitTorrent to Wireless Ad Hoc Networks”, Ad-hoc, Mobile and Wireless Networks, LNCS, pp. 189-203, ISBN 978-3-540-85208-7, 2008.
[81] Z. Zhuang, S. Kakumanu, Y. Jeong, R. Sivakumar, and A. Velayutham, "Mobile Hosts Participating in Peer-to-Peer Data Networks: Challenges and Solutions", Springer Journal on Wireless Networks, pp 2313-2333, vol.16, no.8, 2010.
[82] D. Ciullo, M. A. Garcia, A. Horvath, E. Leonard, M. Mellia, D. Rossi, M. Telek, and Paolo Veglia, “Network Awareness of P2P Live Streaming Applications: A Measurement Study”, IEEE Transactions on Multimedia, vol. 12, no. 1, Jan. 2010.
[83] http://www.pptv.com/
[84] http://www.sopcast.org/
[85] http://tvants.en.softonic.com/
[86] C.-M. Huang, T.-H. Hsu, M.-F. Hsu, "Network-aware P2P file sharing over the wireless mobile networks", IEEE Journal on Selected Areas in Communications, vol.25, no.1, pp.204-210, Jan. 2007.
[87] I. M. Moraes, M. E. M. Campista, L. H. M. K. Costa, O. C. M. B. Duarte, J. L. Duarte, D. G. Passos, C. V. N. de Albuquerque, and M. G. Rubinstein, “On the impact of user mobility on peer-to-peer video streaming”, IEEE Wireless Communications, vol 5, n 6, pp 54-62, Dec. 2008.
[88] http://www.pps.tv/
[89] http://esm.cs.cmu.edu/
[90] http://tools.ietf.org/html/draft-ietf-ppsp-problem-statement-06
[91] C. Perkins, “IP Mobility Support for IPv4. IETF RFC 5944”, 2010.
[92] D. Johnson, “Mobility Support in IPv6. IETF RFC 6275”, 2011.
[93] C. Perkins, “Route Optimization in Mobile IP draft-ietf-mobileip-optim-11.txt”
[94] E. Fogelstroem, “Mobile IPv4 Regional Registration IETF RFC 4857”, 2007.
[95] H. Soliman, “Hierarchical Mobile IPv6 (HMIPv6) Mobility Management IETF RFC 5380”, 2008.

[96] R. Koodli, “Mobile IPv4 Fast Handovers, IETF RFC 4988”, 2007.
[97] R. Koodli, “Mobile IPv6 Fast Handovers, IETF RFC 5568”, 2009.
[98] S. Gundavelli, “Proxy Mobile IPv6 IETF RFC 5213”, 2008.
[99] K. Leung, “Mobility Management using Proxy Mobile IPv4. draft-leung-mip4-proxy-mode-02”, 2007.
[100] “IEEE 802.21/D10.0. Draft Standard for Local and Metropolitan Area Networks: Media Independent Handover Services”, IEEE Draft, April 2008.
[101] G. Lampropoulos, A.K. Salkintzis, and N. Passas, “Media-Independent Handover for Seamless Service Provision in Heterogeneous Networks”, IEEE Communications Magazine, vol. 46, no. 1, pp. 64-71, Jan. 2008.
[102] Y. Liu, “Delay Bounds of Chunk-Based Peer-to-Peer Video Streaming”, IEEE/ACM Transactions on Networking, vol.:18 (4), 2010.
[103] G. Bianchi, N. B. Melazzi, L. Bracciale, F. Lo Piccolo, S. Salsano, “Fundamental Delay Bounds in Peer-to-Peer Chunk-Based Real-Time Streaming Systems”, 21st Int’l Teletraffic Congress, pp.1-8, 2009.
[104] IETF Draft, ”RTP Payload Format for H.264 Video”, draft-ietf-avt-rtp-rfc3984, February 2005.
[105] IETF Draft, ”RTP Payload Format for SVC Video”, draft-ietf-avt-rtp-svc, October 2010.
[106] K.-D. Seo, J.-S. Kim, S.-H. Jung, and J.-J. Yoo, ”A Practical RTP Packetization Scheme for SVC Video Transport over IP Networks”, ETRI Journal, no. 2, April 2010, DOI: 10.4218/etrij.10.1409.0031
[107] G. B. Akar, A. M. Tekalp, C. Fehn and M. R, Civanlar, ‘Transport Methods in 3DTV—A Survey’, IEEE Trans. On Circuits and Systems for Video Technology, Vol. 17, 2007, DOI: 10.1109/ TCSVT.2007. 905365
[108] ITU-T H.241, “Extended video procedures and control signals for H.300-series terminals”, Dec. 2009
[109] S. Liu, Y. Chen, Y. Wang, M. Gabbouj, M. M. Hannuksela, and H. Li, “Frame Loss Error Concealment For Multi-view Video Coding”, June 2008, DOI: 10.1109/ISCAS.2008.4542206
[110] DSL Forum TR-101: “Migration to Ethernet-Based DSL Aggregation”, April 2006
[111] BroadBand Forum TR-156: “Using GPON Access in the context of TR-101”, Dec. 2008.
[112] Open IPTV Forum OIPF-T1-R2: “Functional Architecture”, September 2009
[113] Home Gateway Initiative: “Home Gateway Technical Requirements: Residential Profile”, April 2008.
[114] P. Maymounkov and D. Mazieres, “Kademlia: A Peer-to-peer Information System Based on the XOR Metric”, 2005.
[115] A. Bakker, J.J.D. Mol, J. Yang, L. d’Acunto, J.A. Pouwelse, J. Wang, P. Garbacki, A. Iosup, J. Doumen, J. Roozenburg, Y. Yuan, M. ten Brinke, L. Musat, F. Zindel, F. van der Werf, M. Meulpolder, J. Taal, and B. Schoon, “Tribler Protocol Specification”, 2009.
[116] J. Gmol, J. Poulwelse, M. Meulpolder, D. Epema, and H. Sips, “Give-to-Get: Free-riding Resilient Video-on-demand in P2P Systems”, 2008.
[117] Y. Ding, J. Liu, and H. Jiang, “Peer-to-Peer Video On Demand with Scalable Video Coding”, Computer Communications, Elsevier, vol.:33 (2), pp.1589-1597, 2010.
[118] O. Abboud, K. Pussep, A. Kovacevic, and R. Steinmetz, “Quality Adaptive Peer-to-Peer Streaming Using Scalable Video Coding”, Lecture Notes in Computer Science, vol.: 5842, pp.41-54, 2009.
[119] Y. Zhou, D. M. Chiu, and J. C. S. Lui, “A Simple Model for Chunk-Scheduling Strategies in P2P Streaming”, IEEE/ACM Transactions on Networking, vol.:19 (1), pp.42-54, 2011.

[120] R. Birke, E. Leonardi, M. Mellia, A. Bakay, T. Szemethy, C. Kiraly, R. Lo Cigno, F. Mathieu, L. Muscariello, S. Niccolini, J. Seedorf, and G. Tropea, “Architecture of a Network Aware P2P-TV Application: the NAPA-WINE Approach”, 2010.
[121] http://www.ist-sea.eu/
[122] S. Zezza, E. Magli, G. Olmo, and M. Grangetto, “SEACAST: A Protocol for Peer-to-Peer Video Stream Supporting Multiple Descrition Coding”, 2009.
[123] D. Childers, “Video and Audio Streaming With Bittorrent [internet]. Versão 10”, Knol. 2009 jan 18. Disponível em: http://knol.google.com/k/david-childers/video-and-audio-streaming-with/11g5z3s8cl2s7/1
[124] X. Hei, Y. Liu, and K. W. Ross, "Inferring Network-Wide Quality in P2P Live Streaming Systems," JSAC, vol.25, no. 10, Dec. 2007.
[125] http://www.scvi.net/stream/soft.htm
[126] http://www.scvi.net/stream/soft2.htm
[127] http://goalbit.sourceforge.net/
[128] http://tools.ietf.org/html/draft-ietf-ppsp-problem-statement-06

## APPENDIX A

### Apple A5

The Apple A5 is a package on package (PoP) system-on-a-chip (SoC) designed by Apple and manufactured by Samsung to replace the Apple A4. The chip commercially debuted with the release of Apple's iPad2 tablet, followed by iPhone 4S smart phone.

The A5 contains a dual-core Advanced RISC Machines (ARM) Cortex-A9 CPU with ARM's advanced SIMD extension, marketed as NEON, and a dual core PowerVR SGX543MP2 GPU. This GPU can push between 70 and 80 million polygons/second and has a pixel fill rate of (2 billion pixels)/second. Apple lists the A5 to be clocked at 1 GHz on the iPad 2's technical specifications page, though it can dynamically adjust its frequency to save battery life.

Apple claims that the CPU is twice as powerful and the Graphics Processing Unit (GPU) is up to nine times as powerful as its predecessor, the Apple A4. The A5 package contains 512 MB of low-power DDR2 RAM clocked at 533 MHz. The A5 is estimated to cost 75% more than its predecessor; the price difference is supposed to diminish as production increases.

### Nvidia Tegra

Nvidia Tegra is a system-on-a-chip series developed by Nvidia for mobile devices such as smart phones, personal digital assistants, and mobile Internet devices. The Tegra integrates the ARM architecture processor CPU, GPU, Northbridge, Southbridge, and memory controller onto one package. The series emphasizes low power consumption and high performance for playing audio and video.

### Tegra 3 (Kal-El) series

- Processor: quad-core ARM Cortex-A9 MPCore, up to 1.5 GHz
- 12-Core Nvidia GPU with support for 3D stereo
- Ultra low power GPU mode
- 40 nm process by Taiwan Semiconductor Manufacturing Company (TSMC )
- Video output up to 2560×1600
- NEON vector instruction set
- 1080p MPEG-4 AVC/h.264 High Profile video decode

The Tegra 3 is functionally a quad-core processor, but includes a fifth "companion" core. All cores are Cortex-A9s, but the companion core is manufactured with a special low power silicon process. This means it uses less power at low clock rates, but more at higher rates; hence it is limited to 500 MHz. There is also a special logic to allow running state to be quickly transferred between the companion core and one of the normal cores. The goal is for a mobile phone or tablet to be able to power down all the normal cores and run on only the companion core, using comparatively little power, during standby mode or when otherwise only a small CPU performance is needed. According to Nvidia, this includes playing music or even video content.

### Qualcomm Snapdragon

Snapdragon is a family of mobile system on chips by Qualcomm. Qualcomm considers Snapdragon a "platform" for use in smart phones, tablets, and smart book devices.

The Snapdragon application processor core, dubbed Scorpion, is Qualcomm's own design. It has many features similar to those of the ARM Cortex-A8 core and it is based on the ARM v7 instruction set, but theoretically has much higher performance for multimedia-related SIMD operations.

All Snapdragon processors contain the circuitry to decode high-definition video (HD) resolution at 720p or 1080p depending on the Snapdragon chipset. Adreno, the company's proprietary GPU technology, integrated into Snapdragon chipsets (and certain other Qualcomm chipsets) is Qualcomm's own design, using assets the company acquired from AMD.

**STE Novathor™ U8500**

The NovaThor™ U8500 is the first integrated smart phone platform to offer the latest SMP (Symmetric Multi-Processing) dual core technology in a high-performance, low-power and cost-optimized solution for multiple operating systems. The U8500 is the first mobile platform with full High-Definition 1080p progressive-scan camcorder capabilities. With its combination of a dual-core SMP processor and a high-end 3D graphics accelerator, the U8500 enables a full web-browsing experience for next-generation smart phones.

*Features*

- Full HD 1080p camcorder, multiple codecs supported (H264 HP, VC-1, MPEG-4)
- High-resolution, touch screen display support up to (Wide eXtended Graphics Array) WXGA
- Simultaneous dual display support up to dual eXtended Graphics Array XGA
- High performance 3D graphics
- Dual camera support with Integrated ISP 20 Mpixel and 5 Mpixel
- Wi-Fi, Bluetooth, Global Positioning System (GPS) and Frequency Modulation (FM) enabled platform
- Built-in USB 2.0, HDMI out
- Support for multiple operating systems
- Optional support for mobile TV standards

*Technology*

- Highly efficient, low-power ARM dual Cortex™- A9 processor
- Dual multimedia DSP for low-power, flexible media processing
- High-bandwidth LP-DDR2 interface
- ARM® Mali™ 400 GPU and NEON®CPU extensions
- State-of-the-art HSPA (High-Speed Packet Access) Release 7 modem
- Unique audio architecture with a wide range of audio codecs supported
- Advanced power saving architecture enabling class-leading audio and video playback times

**OMAP 5**

The 5th generation OMAP, OMAP 5 SoC uses a dual-core ARM Cortex-A15 CPU with two additional ARM Cortex-M4 cores to offload the A15s in less computationally intensive tasks to increase power efficiency, two PowerVR SGX544MP graphics cores and a dedicated TI 2D BitBlt graphics accelerator, a multi-pipe display sub-system and a signal processor. They respectively support 24 and 20 megapixel cameras for front and rear 3D HD video recording. The chip also supports up to 8 GB of dual channel DDR3 memory, output to four HD 3D displays and 3D HDMI 1.4 video output. OMAP 5 also includes 3 USB 2.0 ports and a SATA 2.0 controller.

*OMAP5430 Key Benefits*

- Designed to drive Smartphone's, Tablets and other multimedia-rich mobile devices
- Multi-core ARM® Cortex™ processors
    - Two ARM Cortex-A15 MPCore processors capable of speeds up to 2 GHz each
    - Two ARM Cortex-M4 processors for low-power offload and real-time responsiveness
- Multi-core POWERVR™ SGX544-MPx graphics accelerators drive 3D gaming and 3D user interfaces
- Dedicated TI 2D BitBlt graphics accelerator
- IVA-HD hardware accelerators enable full HD 1080p60, multi-standard video encode/decode as well as 1080p30 stereoscopic 3D (S3D)
- Faster, higher-quality image and video capture with up to 24 megapixels (or 12 megapixels S3D) imaging and 1080p60 (or 1080p30S3D) video
- Supports four cameras and four displays simultaneously
- Packaging and memory: 14mm x 14mm, 0.4mm pitch PoP dual-channel LPDDR2 memory

*Support Features*

- Up to three simultaneous, high-resolution, colour-rich LCD display support
- HDMI 1.4a output to drive HD displays, including S3D
- MIPI serial camera and serial display interfaces
- MIPI® SLIMbusSM
- MMC/SD
- USB 3.0 On-The-Go (OTG) Super Speed with integrated PHY interface
- Comprehensive software suite supporting all major mobile OSes that is fully integrated and tested for real-world use cases to reduce development time and costs
- OMAP Developer Network provides programs and media components for manufacturers to create distinctive products delivered to market quickly

Based on the current development activities and projects, current display technologies will remain on the market within coming several years. On the other hand quality of the current technologies is expected to develop making auto stereoscopic mobile displays appealing to mass customers.

# APPENDIX B: EXISTING WORK ON P2P NETWORKS

**Kademlia protocol**

This is a protocol for distributed hash table. This protocol has been used in BitTorrent Enhanced Protocol for distributing peer addresses.

As described in [114]:

Kademlia searches the available nodes through their 160 bits Node IDs. By building a binary tree based on these Ids, Kademlia tries to locate other nodes. So the look up table is actually a table based on the Node Id of each node.

Each node keeps a bucket of possible neighbouring nodes. In this bucket (called a k-bucket) for each node, a triplet (< IP address; User Datagram Protocol (UDP) port; Node ID>) is stored in the bucket. The buckets are sorted according to the last time they are seen.

When a new node receives any message from another node:

1. If the node is already in the corresponding bucket, the recipient moves it to the tail of the node.
2. If the corresponding node has less than k elements, the recipient adds the node to the tail.
3. If appropriate bucket is full. Recipient lets the last node in the bucket to decide
    a) If the last node does not respond, the new node is added and the tail is removed.
    b) If it responds, the sender is discarded.

There are four calls in Kademlia protocol:

1. PING: probes a node to see if it is online.
2. STORE: instructs a node to store a key value pair for later use.
3. FIND_NODE: takes a 160-bit ID as an argument. The recipient of a RPC returns IP address; UDP port; Node ID i triples for the k nodes it knows about closest to the target ID. These triples can come from a single k-bucket, or they may come from multiple k-buckets if the closest k-bucket is not full. In any case, the RPC recipient must return k items (unless there are fewer than k nodes in all its k-buckets combined, in which case it returns every node it knows about).
4. FIND VALUE behaves like FIND NODE—returning IP address; UDP port; Node ID i triple—with one exception. If the RPC recipient has received a STORE RPC for the key, it just returns the stored value

Kademlia locates the k closest nodes to some given node ID which is called a node lookup. Lookup is done as:

1. The initiator starts by picking a number of from its closest non empty buckets and sends them FIND_NODE calls.
2. In the recursive steps, initiator sends the FIND_NODE to nodes it has acquired from previous calls.
3. Each time the initiator picks the nearest nodes to the destination node from the received lists.
4. Each time a node does not respond it is removed from consideration till it responds.
5. If destination is not found, initiator sends the address to the remaining closest neighbour.

Joining the network:

1. A new node (U) must have a contact to an already participating contact (W).

2. U inserts W into its k-bucket.
3. U then makes a node lookup for its own node ID.
4. U refreshes its buckets and during these, U both populates its own k-buckets and inserts itself into other nodes' k-buckets as necessary.

**Tribler protocol**

This is a live streaming protocol based on the BitTorrent protocol. To support the needs of streaming media there are some alterations to the existing protocol. These are explained in [115] in detail as:

1. Tribler created the Give-To-Get algorithm for distributing the stream.
2. For its peer selection, a peer in the system will try to keep a list of ten neighbours. If there are less than 10 neighbours peer will try to discover new neighbours.[116]
3. Chunks are obtained from neighbours by requesting them. Each peer keeps its neighbour informed about its chunks. [116]
4. Tribler keeps a list of neighbours which it updates periodically to decide whom to choke. This is based on the performance of the neighbours. Here the performance is based on the amount of data uploaded from this neighbour thus the name" give to get". [116]
5. A Sliding priority window is implemented for chunk selection. The priority is based on the nearness of the piece to the current playback time. Pieces are discarded when their playtime is over. The priority levels for a piece with timestamp **i** and current playtime m can be summarized as:
   • High priority: $m \leq i < m + h$. If **p** has already started playback, it will pick the lowest such **i**, otherwise, it will pick i rarest first.
   • Mid priority: $m + h \leq i < m + (\mu + 1)h$. Peer p will choose such an **i** rarest first.
   • Low priority: $m + (\mu + 1) h \leq i$. Peer p will choose such an **i** rarest first.
6. For the integrity of the pieces a Merkle Hash is used by the Tribler protocol.

**Chunk selection and scheduling for P2P streaming**

In the following, existing chunk selection and scheduling mechanisms are presented. These mechanisms aim at offering playback continuity and quality. Some of them are suitable for scalable video streaming while some others have a more generic orientation. Theoretical studies on chunk dissemination strategies are useful in providing insight into the advantages of certain approaches and leading the way to the design of more robust streaming techniques.

In [117], the chunk scheduling problem in P2P video on demand with SVC is formulated as an Integer Linear Programming (ILP) problem. The study is based on a sliding window that determines the chunks currently of interest. The sliding window has a fixed width, representing the number of consequent chunks under consideration in the playback time line and a variable height, representing the number of layers. The height of the sliding window can be adjusted to meet the current capacity faced by the peer. Chunks of different layers are assigned different weights depending on their importance. The objective is to determine whether and which other peer a peer will retrieve each chunk from. An efficient solution is hard to be derived and global knowledge regarding peers' upload and download bandwidth is required. Therefore, a zigzag scheduling algorithm is proposed.

A zigzag ordering of the chunks included in the sliding window defines the ordering of transmissions in case not all the chunks can be transmitted. The zigzag ordering follows three basic principles: a) it captures the dependencies between frames and layers in the SVC hierarchy, b) chunks in nearer time instants are more important and c) higher-layer chunks at nearer time instants are more important than lower-layer chunks at farther time instants.

Chunks are transmitted according to the zigzag line. Each chunk is transmitted by the peer with the least number of requested chunks and among a set of possible providers, the one with the highest bandwidth is selected.

In [118], a multi-coefficient modelling approach for block selection is followed. A series of blocks constitutes a chunk. The priority of a block is given by:

$$\text{Priority}(\text{Block}(t,D,T,Q)) = -A\,t - B\,(a\,D + b\,T + c\,Q)$$

In this formula, t is the index for temporal layer, D, T and Q are the total layer numbers for spatial, temporal and quality layers respectively, A and B are the chunk priority factors, and a, b and c define different weights for blocks inside a chunk that belong to different layers.

The priority is separated into two parts. The first part describes urgency due to playback constraints and the second part describes importance based on the order of layer. The A parameter actually controls smoothness while the B parameter controls quality assuming long buffering times.

Authors in [102] analyze the problem of delay in chunk-based P2P live video streaming systems. In these systems, an inherent source of delay is the fact that the uploading bandwidth of a peer cannot be utilized to upload a chunk until the peer completes the download of that chunk. Three streaming strategies are compared: single-tree, multi-tree and snowball chunk dissemination. In snowball chunk dissemination, peers who receive a chunk in the $k^{th}$ time slot upload the chunk for K-k times, where K is the number of time slots for all peers to receive the chunk. Theoretical delay bounds are derived and it is shown how bandwidth heterogeneity among peers and network impairments like propagation delays and bandwidth variations can affect performance. For the homogeneous case, it is proved that there exists a feasible chunk schedule such that all chunks are disseminated to all peers within the minimum chunk delay time.

The Dynamic Snowball (DSB) Streaming Algorithm is proposed for dynamic environments. The algorithm tries to achieve two seemingly contradicting objectives: a) older chunks are pushed as quickly as possible and b) newer chunks get enough upload bandwidth to quickly grow the upload bandwidth for them. DSB is a heuristic that mimics the static snowball chunk dissemination and it works in rounds. In each round:

- For each newly completed chunk transmission the chunk is added to a buffered chunk set of the destination peer.
- For each active chunk that has not been uploaded to all peers, the demand factor is calculated as the fraction of the number of peers without that chunk to the number of peers with that chunk.
- For each peer, the expected workload is calculated as the sum of the demand factors of the chunks in the peer's buffered chunk set.
- For each active chunk, starting from the oldest, two peers are chosen: a source peer with the lowest expected workload that has that chunk and a destination peer with the lowest expected workload that does not have that chunk. The source peer uses all its upload bandwidth to upload that chunk.

A similar study appears in [103]. Stream diffusion metric is used to derive fundamental delay bounds in chunk-based P2P streaming systems. In order to attain the performance bounds, the authors apply a Round-Robin chunk scheduling policy. Different results characterize the performance of the system using a single unbalanced tree and unbalanced multiple trees as dissemination structures.

In [119], two chunk selection and scheduling strategies are compared against random peer selection. The Rarest First strategy aims at increasing the scalability of the system by delivering a chunk to as many peers as quickly as possible. The Greedy strategy aims at maximizing playback continuity from a peer's local perspective. For each strategy, a chunk selection function is defined. The chunk selection function is supposed to maximize the buffer occupancy probability, i.e. the probability that a slot in the play buffer is filled with the correct chunk at a certain time instance. The chunk selection function of the Rarest First algorithm is such that the latest chunk distributed by the initial source that is missing from all the local peers' buffer is expected first. The chunk selection function of the Greedy algorithm is such that the empty buffer location closest to the playback time is filled first.

A mixed-strategy, based on the intuition behind the strengths of the previous strategies is also proposed. According to this strategy, the buffer space is partitioned into two parts. If the Rarest First strategy fails to find chunks to download for the buffer locations belonging to the first part, the Greedy strategy is followed for the buffer locations of the second part. In this case, the buffer occupancy probability depends also on the partitioning scheme.

It is evident from above that the chunk selection and scheduling strategies highly affect the performance of live P2P streaming systems. Parameters as peers' uploading and downloading capacity, physical network impairments and peer dynamics need to be considered by the mechanisms that handle the diffusion of chunks through the network. The problem is further complicated by the specific characteristics of the streamed video content like scalability, multi-viewing and 3-dimensionality. Therefore, the importance of each video chunk is mapped into a multi-dimensional space. Given the dynamic nature of the service level offered by the wireless access technologies, the vector that describes chunk selection has to be properly accommodated by each peer's current capacity.

**Napa-Wine (Network-Aware P2P-TV Application over Wise Networks)**
This is another EU funded project. This project aims to stream TV content via a P2P network to users thus eliminating the need for a centric approach.

As described in [120]:

Napa-Wine is an innovative, network cooperative P2P-TV application targeted to favour cooperation between the application and the transport network layer.

Napa-Wine uses a mesh based topology. It has a distributed system for overlay management. There are different types of repositories (In Napa-Wine repositories act somehow like trackers for BitTorrent) storing the overall data provided by the peers in the system.  The first one is called a P-REP which store a list of active peers, their available network conditions (throughout their life time), and path traversal times made by these peers to other peers. In addition to these other repositories (with different names) store uncalculated path traversal times (based on predictions),  topology data like AS graph and peering points, the ranking of routing paths, routing tables, information on the routing optimization and the network topology provided by the ISP. All peers can connect to these repositories via a Repository Controller API which serves them with any data requested.

When a peer wants to join a live stream, it connects to a repository to get an initial list of possible neighbours then the set of neighbours of every node in the system is dynamically optimized over time. In addition to their repository data, peers use a gossip based method to find more suitable neighbours to share stream. The choice for neighbours depends on inter-peer throughput, the measured round trip time (RTT), the chunk loss rate, etc. Addition of neighbours is instead based on a mixed strategy of optimization (e.g., adding the most stable and resourceful peers) and randomization to avoid fragile topologies (e.g., group of peers with too few connections toward the rest of the overlay).

In Napa-Wine, more than one chunk selection method has been tried. Throughout the experiments it was found out that deadline based algorithm that diffuse the chunk with the smallest deadline to a peer that is in the best possible state to further diffuse it with prioritizing the high upload bandwidth peers.

**SEA (Seamless Content Delivery)**
Sea is an EU funded project. As described in [121]:

"SEA (SEAmless Content Delivery) project aims to offer a new experience of seamless video delivery, maintaining the integrity and wherever applicable, adapting and enriching the quality of the media across the whole distribution chain."

SEACAST protocol:

1. Uses a dynamic multiple-tree topology,
2. A distributed tree building protocol giving peers only a local knowledge of the topology
3. MDC to distribute the video into independent parts.

SEACAST is based on the VidTorrent protocol. VidTorrent is a multipath multi-tree protocol implemented by the Multimedia Lab in MIT. The differences between VidTorrent and SEACAST are [122]:

1. Instead of TCP, RTP/UDP is used for transport.
2. MDC is used for encoding the video thus enabling a multipath multi-tree overlay.
3. Real Time Streaming Protocol (PRTSP)/Session Description Protocol (SDP) signalling has been implemented to describe streaming sessions.
4. DIVA MDC has been integrated with SEACAST, and replaces the trivial odd/even frame data splitting present in VidTorrent.
5. VidTorrent has major problems to scale to a large number of nodes. Even though these problems are not completely solved, at present SEACAST is able to scale to a sufficient number of nodes (about 50) to enable validation on a local test-bed; full functionality is expected to be available shortly.

**Additional projects**
A thorough introduction to streaming multimedia via BitTorrent-based protocols appears on [123]. The paper describes the basics of the BitTorrent protocol, relating it to the requirements of streaming services, and then it provides information regarding current streaming software and libraries based on the P2P paradigm.

Regarding the tracker, which is the component deciding which serves a list of peers in response to a request for a list of peers to exchange chunks with, the original approaches, like for example the "vanilla" BitTorrent, select peers at random in the set known by the tracker. It is desirable to use a strategy to perform a smart choice of peers. For example, [124] proposes to use buffer maps (a data structure that indicates the chunks that are available in a peer) to monitor network behaviour of a peer, and to match peers based on the state of their buffer maps, and with the secondary goal to match peers with similar network resources.

The website of SCVI has got two web pages reporting P2P streaming software. In particular, [125] reports open source P2P streaming software, while [126] lists proprietary streaming software. The lists pretend to be exhaustive.

One of the most interesting software suites reported in the list of open source P2P streaming systems is called Goalbit [127]. It has got a number of the functionalities that will be included into ROMEO. These include in particular, peer casting (P2P-based streaming) and real-time collection of client's performance. The suit follows the standard PPSP [128], which is an IETF effort to create a standard protocol for P2P streaming.

## APPENDIX C: GLOSSARY OF ABBREVIATIONS:

| | |
|---|---|
| **A** | |
| AAA | Authentication, Authorization and Accounting |
| AAC | Advanced Audio Coding |
| ADTE | Adaptation Decision Taking Engine |
| AGC | Automatic Gain Control |
| AP | Access Point |
| ARC | Arcelik A.S. |
| ARM | Advanced RISC Machines |
| A/V | Audio / Video |
| AVC | Advanced Video Coding |
| **B** | |
| BCC | Binaural Cue Coding |
| BGRP | Border Gateway Reservation Protocol |
| BRAS | Broadband Remote Access Server |
| BRIR | Binaural Room Impulse Responses |
| BRTF | Binaural Room Transfer Functions |
| **C** | |
| CDN | Content Delivery Networks |
| CN | Correspondent Node |
| CoA | Care of Address |
| CoS | Class of Services |
| CPE | Customer Premises Equipment |
| CPU | Central Processing Unit |
| CSRC | Contributing Source |
| **D** | |
| D/A | Digital to Analog |
| DHCP | Dynamic Host Configuration Protocol |
| DON | Decoding Order Number |
| DoW | Description of Work |
| DVB | Digital Video Broadcast |
| **F** | |
| FA | Foreign Agent |
| FEC | Forward Error Correction |
| FFT | Fast Fourier Transform |
| FM | Frequency modulation |
| FMIP | Fast Mobile Internet Protocol |
| FTP | File Transfer Protocol |
| FU | Fragmentation Unit |
| **G** | |
| GBCE | Global Brightness Contrast Enhancement |
| GM | General Meeting |

| | |
|---|---|
| GPON | Gigabit Passive Optical Network |
| GPS | Global Positioning System |
| GPU | Graphics Processing Unit |
| GUI | Graphical User Interface |
| **H** | |
| HA | Home agent |
| HD | High Definition |
| HDMI | High-Definition Multimedia Interface |
| HLOS | High Level Operating System |
| HMIP | Hierarchical Mobile IP |
| HRTF | Head Related Transfer Function |
| HoA | Home Address |
| **I** | |
| ICC | Inter-Channel Coherence |
| ICLD | Inter-Channel Level Difference |
| ICTD | Inter-Channel Time Difference |
| IETF | Internal Engineering Task Force |
| IFFT | Inverse Fast Fourier Transform |
| IGMP | Internet Group Management Protocol |
| ILD | Inter-aural Level Differences |
| IP | Internet Protocol |
| IPR | Intellectual Property Rights |
| IPTV | Internet Protocol Television |
| IRT | Institut für Rundfunktechnik GmbH |
| IT | Instituto de Telecomunicações |
| ITD | Inter-aural Time Differences |
| **J** | |
| JSON-RPC | JavaScript Object Notation-Remote Procedure Call |
| **L** | |
| LCD | Liquid Crystal Display |
| LFE | Low Frequency Enhancement |
| LMA | Local Mobility Anchor |
| LTE | Long Term Evolution |
| LVDS | Low Voltage Differential Signalling |
| **M** | |
| MAC | Media Access Control |
| MAG | Mobility Access Gateway |
| MANE | Media Aware Network Element |
| MANET | Mobile Ad-Hoc Network |
| MAP | Mobility Anchor Point |
| MARA | Multi-user Aggregated Resource Allocation |
| MDC | Multiple Description Coding |
| MDCT | Modified Discrete Cosine Transform |
| MD-SMVD | Multiple Description Scalable Multi-view Video plus Depth |
| MICS | Media Independent Command Service |

| | |
|---|---|
| MIES | Media Independent Event Service |
| MIH | Media Independent Handover |
| MIIS | Media Independent Information Service |
| MIP | Mobile Internet Protocol |
| MISO | Multiple Input Single Output |
| MM | Mobility Management |
| MMS | MM Solutions AD |
| MN | Mobile Network |
| MPEG | Moving Pictures Experts Group |
| MPLS | Multiprotocol Label Switching |
| MST | Multi-Session Transmission |
| MT | Mobile Terminal |
| MTAP | Multi-Time Aggregation Packet |
| MTD | Mobile Terminal Display |
| MULSYS | MulSys Limited |
| MVC | Multi View Coding |
| **N** | |
| NAL | Network Abstraction Layer |
| NAT | Network Address Translation |
| NEM | Networked Electronic Media |
| NGN | Next Generation Networks |
| **O** | |
| OFDM | Orthogonal Frequency Division Multiplexing |
| OLT | Optical Line Terminal |
| ONT | Optical Network Terminal |
| **P** | |
| P2P | Peer-to-Peer |
| PAPR | Peak to Average Power Ratio |
| PC | Personal Computer |
| PHY | Physical |
| PLP | Physical Layer Pipe |
| PMIP | Proxy Mobile Internet Protocol |
| PoA | Point of Attachment |
| PoP | Package on Package |
| POTS | Plain Old Telephone Service |
| PPP | Point-to-Point Protocol |
| PPSP | Peer-to-Peer Streaming Protocol |
| PSW | Priority Sliding Window |
| **R** | |
| R&S | Rohde & Schwarz GmbH & Co. KG |
| RAT | Remote Access Technology |
| RF | Radio Frequency |
| RGB | Red Green Blue |
| RISC | Reduced Instruction Set Computer |
| RO | Route Optimization |

| | |
|---|---|
| RRP | Return Routability Protocol |
| RSVP | Resource ReServation Protocol |
| RTCP-XR | Real Time Control Protocol Extended Report |
| RTP | Real-time Transport Protocol |
| RTSP | Real Time Streaming Protocol |
| **Q** | |
| QAM | Quadrature Amplitude Modulation |
| QAT | Quality Assurance Team |
| QoE | Quality of Experience |
| QoS | Quality of Service |
| **S** | |
| SAC | Spatial Audio Coding |
| SD | Standard Definition |
| SDP | Session Description Protocol |
| segSNR | segmental Signal-to-Noise Ratio |
| SEI | Supplemental Enhancement Information |
| SIDSP | Simple Inter-Domain QoS Signalling Protocol |
| SMVD | Scalable Multi-view Video plus Depth |
| SoC | System on Chip |
| SSRC | Synchronisation Source |
| SST | Single Session Transmission |
| STAP | Single-Time Aggregation Packet |
| SVC | Scalable Video Coding |
| **T** | |
| TEC | Technicolor R&D France |
| TFT | Thin Film Transistor |
| TID | Telefónica I+D |
| TTA | Türk Telekomünikasyon A.Ş. |
| **U** | |
| UDP | User Datagram Protocol |
| UHF | Ultra High Frequency |
| UP | University of Patras |
| US | University of Surrey |
| **V** | |
| VAS | Value Added Service |
| VITEC | VITEC Multimedia |
| VoD | Video On Demand |
| **W** | |
| Wi-Fi | Wireless Fidelity |
| WFS | Wave Field Synthesis |
| WG | Working Group |
| WP | Work Package |
| WVGA | Wide Video Graphics Array |
| WXGA | Wide eXtended Graphics Array |
| **X** | |

| | |
|---|---|
| XGA | eXtended Graphics Array |
| **Numbers** | |
| 4G | $4^{th}$ Generation |