MULTIMEDIA COMMUNICATION SYSTEM

Unit 9

9.1. Definition of multimedia communication

- Multimedia applications have several requirements – as well for the data transmission as for controlling interactivity. For structuring control and transmission functionality and implement common protocols
- From the communication perspective, we can divide the higher layers of the Multimedia Communication System (MCS) into two architectural subsystems:
  - *Application Subsystem*
  - *Transport Subsystem*

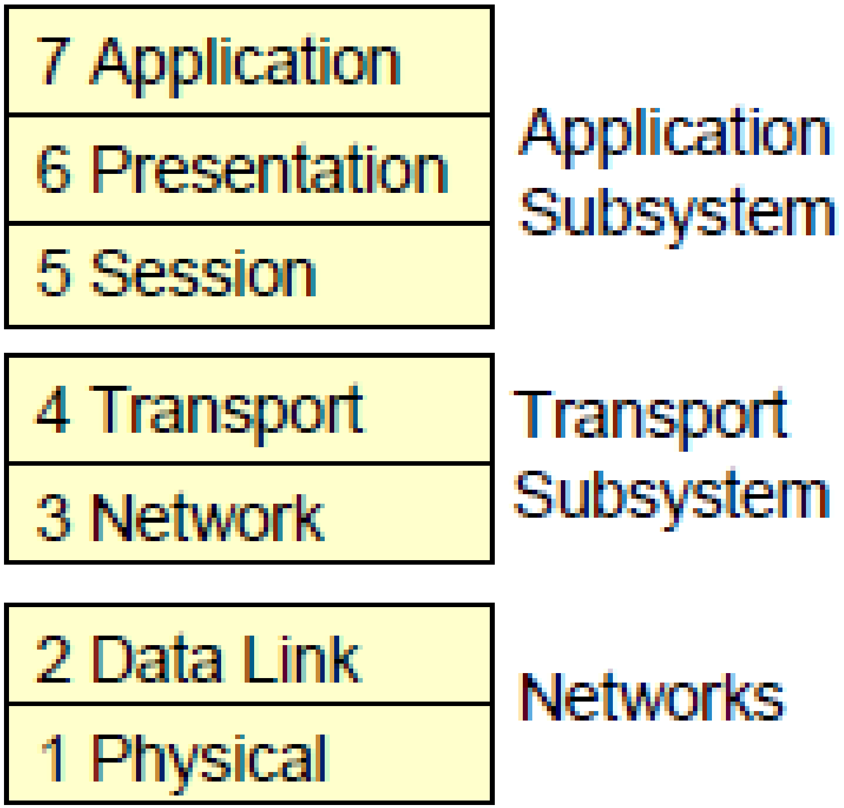
9.1. Definition of multimedia communication

Application Subsystem

This subsystem includes the software and the tools with which the end user directly interacts. E.g. of application can be email, video conferencing software etc.

- Responsible for the management and service issues for group cooperation and session orchestration
- Supporting a large scale of multimedia applications, e.g.
  - Multimedia Mail
  - Virtual Reality Applications
  - Video Conferencing
  - CSCW (Computer Supported Cooperative Work)

9.1. Definition of multimedia communication



9.1. Definition of multimedia communication

Transport Subsystem

- Transport and network layer protocols for multimedia applications (streaming)

9.2. Application Subsystem

- Streaming multimedia data often is used in *cooperative computing*
  - Beneath streaming services (transport subsystem), some control functionality for cooperation support is needed
  - Cooperative Computing is generally known as *Computer Supported Cooperative Work* (CSCW).
- Tools for cooperative computing:
  - Electronic mail
  - Shared whiteboards
  - Screen sharing tools
  - Application sharing
  - Text-based conferencing systems
  - Video conference systems (e.g. MBone Tools, ProShare from Intel, PictureTel,Teles Online, NetMeeting from Microsoft)

9.2. Application Subsystem

Cooperation Dimensions

Computer-Supported cooperation may be categorized according to the following parameters:

- *Time*
  - Asynchronous cooperative work (not at the same time)
  - Synchronous cooperative work (at the same time)
- *User Scale*
  - Single user, two users ("dialogue, point to point", direct cooperation) or groups with more than two users
  - Static or dynamic groups, depending on if the members are pre-determined or not
- *Control*
  - Centralized, i.e. controlled by a "chairman"
  - Distributed, i.e. control protocols provide consistent cooperation
- *Locality*
  - Cooperation at the same place
  - Tele-cooperation of users at different places

9.2. Application Subsystem

- **Application Subsystem**

  This subsystem includes the software and the tools with which the end user directly interacts. E.g. of application can be email, video conferencing software etc.

- **Collaborative Computing**

  Collaborative computing is the computer supported cooperative work supported by the networks, PCs and the software that facilitates the cooperation. The examples of collaborative computing tools are *electronic mail, bulletin boards, screen sharing tools, text-based conferencing systems, telephone conference systems, conference rooms* and *video conference systems.*

- **Collaborative Dimensions** The collaboration dimensions are
  - *Time*
  - *User Scale*
  - *Control*

9.2. Application Subsystem

- ***Time:***

According to time there can be two types of collaborative work and they are *asynchronous* where the cooperative works do not happen at the same time; while the other is the *synchronous* where the cooperative works happen at the same time.

- ***User Scale:***

The user of the application can be a single user interacting to the other user or to a group of users. Email between two individuals is a user to user interaction where as Email to a group is a user to group interaction. Video conferencing is also a user to group interaction application.

9.2. Application Subsystem

There can be following types of groups.

- Dynamic or Static Group: When the new users can join the group for cooperative work in the real time the group is said to be dynamic where as when the number of members and their membership is predefined it is static group.
- The members of the group may simply be a *participant* of the collaborative work or he/she may be the *co-coordinator, conference initiator, conference chairman, a token holder* or just an *observer.*
- The members of the group may have *homogenous* or *heterogeneous* characteristics. For e.g. they may belong to different ethnic group or they may differ in the level of intelligence.

9.2. Application Subsystem

- Control:

Control during the collaboration can be *centralized* or *distributed*. Centralized control means there is a chairman who controls the collaborative work and every group member reports to him. Distributed control means every group member has control over his/her own tasks in the collaborative works.

9.2. Application Subsystem

Group Communication Architecture

Group communication is a cooperative activity which may be synchronous or asynchronous which may be central control or a distributed control. A group communication architecture consists of:

- Support model
- System model
- Interface model

Group Communication Architecture

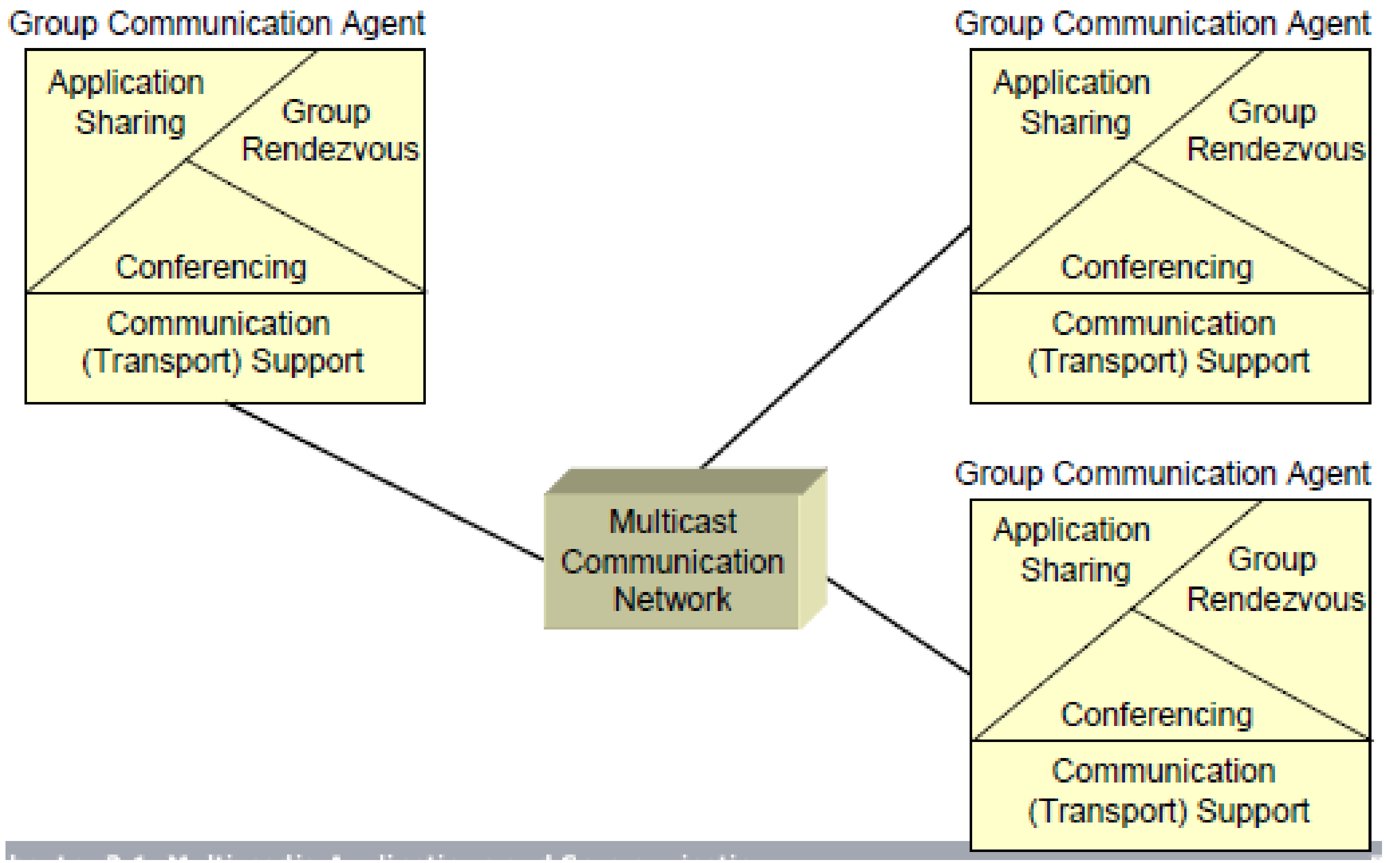


Figure: Communication Support Model

Group Communication Architecture

- Group communication:
  - Synchronous communication
  - Asynchronous communication
- Support Model: *Group communication agents* (cooperating via a multicast network)
  - *Group rendezvous* (organization of meetings and delivering information)
  - *Shared applications* (simultaneous replication and modification of information to multiple users, e.g. telepointing, joint editing)
  - *Conferencing (*audio/video*)*
- System Model:
  - Client/server model
- Interface Model:
  - Exchanging information within the support model (object oriented)

Group Communication Architecture

The Support Model:

It includes *group communication agents* that communicate via a multi-point multicast communication network as shown in the figure above. Group communication agents may use the following for their collaboration:

Group Rendezvous:

It represents methods which allows one to organize meetings, and to get information about the group, ongoing meetings and other static and dynamic information. The rendezvous can be Synchronous where the Directory services allows the access to information stored in a knowledge base about the conference, registered participants, authorized users and name and role of the participants.

Shared Applications:

It refers to the techniques which allow the replication of the information which can be delivered to all participants of the collaborative work. It may use the Centralized Architecture or Replicated Architecture. In the former a single copy of the application runs at one site say server.

The Support Model:

- **Conferencing**:

It represents the service which manages the multiple users to communicate and interact with each other by the use of multimedia data. Thus conferencing is basically a management service that controls the communication among multiple users via multiple media, such as video and audio, to achieve simultaneous face-to-face communication.

Group Communication Architecture

- **The System Model:** It is based on a client-server model where the clients are applications that provides interface to the users who interact with the system while servers refers to function which makes it possible for the clients to communicate with each other and manage the communication.
- **The Interface Model:** It includes the *user presentation protocols* and *group work management protocols.* User presentation protocol is the interface available to the end users from which they can initiate, join, manage, communicate and terminate the conference. Group work management protocols specify the communication between the client and the servers for services like registration and querying the status of the conference.

Session Management

Session Management

A session is the total logged in time of a user or it can be the entire conference from its commencement to its termination. The management of session is a very important task and it should consider several issues like allowing users to join and leave the conferencing, selection of the coordinator, distributing information between users.

Architecture of Session Management It consists of following components:

- Session manager
- Media agent
- Shared workspace agent

Session Management

- *Session Manager:* It includes local and remote functionalities. The local functionalities includes the
  - Membership Control management: Authenticating the users or allowing members to join the session.
  - Control Management: It may involve floor management which involves the distribution and sharing of the information and resources available to the conference.
  - Media Control Management: It is required for the synchronization between the different media.
  - Configuration Management: It refers to the exchange and optimization of the QoS parameters.
  - Conference Control Management: It consists of functionalities for starting, changing and closing the conference.

Session Management

- ***Media Agents**:*

  These are responsible for decisions specific to each type of media. Each agent performs its own control mechanism over the particular medium, such as mute, unmute, change video quality, start sending, stop sending etc.

- ***Shared Workspace Agent**:*

  The shared workspace agent transmits shared objects (e.g., telepointer coordinate, graphical or textual object) among the shared applications.

9.3. Transport Subsystem : Requirements, Transport Layer, Network Layer

Transport Subsystem

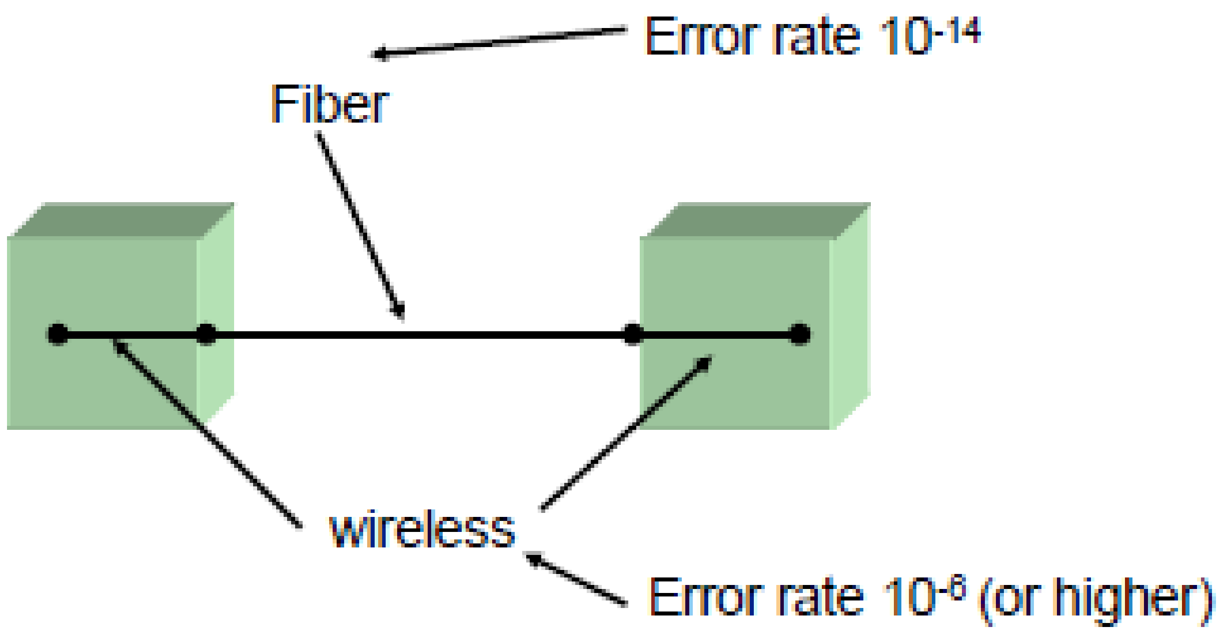
Multimedia applications have high requirements on network protocols:

1. High data throughput
   - Deliver as much data as possible in short time
2. Fast data forwarding
   - Deliver data as fast as possible
3. Service guarantees
   - Deliver data with regard to negotiated policies (throughput, delay)
4. Multicasting
   - 1:*n* and *m*:*n* point communication [*a*:*b* denotes: *a* senders and *b* receivers] *Needed: special transport resp. network protocols*

Most transport protocols are designed for unreliable and "relatively slow" networks Modern high speed networks need ne transport protocols that take into account the properties of new network technology

An end-to-end connection consisting of three parts: wireless link, fiber, wireless link:

The different error rates at different parts of a connection complicate the design of suitable end-to-end transport protocols



Data Throughput (fast & effective)

- Audio and video data typically have a stream-like behavior
- Even in compressed mode, they demand high data throughput (16 kbit/s for compressed audio, 64 kbit/s for original PCM-audio, 2 Mbit/s for MPEG-coded video)
- In a workstation or in a network several of those streams may coexist
- Telephone services or video conferencing demand real-time computing of the data streams

This requires not only suitable transport protocols, but also high performance workstations which are able to compute several multimedia streams simultaneously and can transmit the packets in appropriate speed to the network interface

Service Guarantees / Multicasting

Service guarantees

- Service guarantees are important for the acceptance of multimedia applications
- Multimedia applications need guarantees, such as: throughput ≥ minvalue, delay ≤ maxvalue1, jitter ≤ maxvalue2
- To give service guarantees, resource management must be used – without this, in end-systems and switches/routers multimedia systems cannot provide reliable QoS to their users because transmission over unreserved resources may lead to dropping or delaying of packets

Service Guarantees / Multicasting

Multicasting

- Multicasting is important for multimedia applications in terms of sharing resources like the network bandwidth
- Many multimedia applications, such as video conferencing, have multicast characteristics

Transport Subsystem

Multimedia applications generate and consume a huge amount of data. They produce high demands for the underlying infrastructure and communication architecture:

- Data should be copied directly from adapter to adapter to reduce copy overhead, e.g. from the video board to the network interface by using **direct memory access** (**DMA**). With DMA technique, the application itself never really touches the data, it only sets the correct switches for the data flow by connecting sources to sinks
- The data transfer involved by the layered structure of the protocols form a bottleneck, hence other mechanisms must be found
- For error-recovery some protocols use retransmission techniques which impose requirements on buffer space for queues at the expense of larger end-to-end delays
- The synchronous behavior of most multimedia data streams must be mapped onto the asynchronous transfer mode of the underlying networks

Are the “old“ transport protocols like TCP and UDP suitable for multimedia transmission and, if not, which new protocols exist or have to be developed?

Transport Subsystem

Requirements

- **User and Application Requirements**

  Networked multimedia applications by themselves impose new requirements onto data handling in computing and communication because they need: *Sustainable data throughput, fast data forwarding, service guarantees, and multicasting.*

- *Data Throughput*

  This requirement wants the processing of the system to be **fast** and **effective**.

- *Fast Data Forwarding*

  The users or application wants **very low end to end delay** and jitter when communicating multimedia data. The holding time should be very less due to the real time requirement.

Transport Subsystem

- ***Service Guarantees**:* The loss of the information is undesired and the system or protocol used must ensure that the information is delivered to the intended destination.
- ***Multicasting** :* It is important for sharing the bandwidth and the communication protocol processing at end systems.
- **Processing and Protocol Constraints** Processing system and protocols have constraints which need to be considered while processing and transmitting multimedia information.
  - Following the "**shortest possible path**" for quicker delivery
  - Buffer management
  - Segmentation and reassembly
  - Re-transmission on error
  - Error-recovery
  - Asynchronous transfer

Transport Subsystem

Transport Layer

Transport protocols, to support multimedia transmission, need to have new features and provide the following functions:

- Timing information
- Semi-reliability
- Multicasting
- NAK (none-acknowledgement)-based error recovery mechanism
- Rate control

Internet Transport Protocols

- *TCP (Transmission Control Protocol)* It was designed to provide a **reliable end-to-end** byte stream over an unreliable inter-network.
- Each machine supporting TCP has a TCP transport entity, either a library procedure, a user process, or part of the kernel. In all cases, it manages TCP streams and interfaces to the IP layer.
- A TCP entity accepts user data streams from local processes, breaks them up into pieces not exceeding 64KB, and sends each piece as a separate IP datagram.

Internet Transport Protocols

- UDP (User Datagram Protocol) The Internet protocol suite supports a connectionless transport protocol, *UDP (User Datagram Protocol).*
- UDP provides a way for applications to send encapsulated **(summarized) IP datagrams and send them without having to establish a connection**.
- UDP transmits *segments* consisting of an 8-byte header followed by the payload

Real-time Transport Protocol (RTP)

- RTP is a **UDP protocol** used in the **client server** environment and in the real-time multimedia applications.
- The multimedia application consists of multiple audio, video, text, and possibly other streams.
- These are fed into the RTP library, which is in the user space along with the application. This library then multiplexes the streams and encodes them in RTP packets, which it then stuffs into a socket. At the other end of the socket, UDP packets are generated and embedded in IP packets.

Xpress Transport Protocol (XTP)

- XTP integrates transport and network protocol functionalities to have **more control over the environment** in which it operates. XTP is intended to be useful in a wide variety of environments, from real-time control systems to remote procedure calls in distributed operating systems and distributed databases to bulk data transfer.
- It defines for this purpose six service types: *connection, transaction, unacknowledged data gram, acknowledged datagram, isochronous stream* and *bulk data.*

Network Layer

- **Internet Protocol** In the TCP/IP protocol stack the network layer protocol is the *Internet Protocol (IP)* and, in order to transfer packets of information from one host to another, it is the IP in the two hosts, together with the IP in each Internet gateway and router involved that perform the routing and other harmonization functions necessary. The IP in each host has a unique Internet-wide address assigned to it.
- This is known as the host‟s *Internet address* or, more usually, its *IP address.* Each IP address has two parts: a *network identifier* and a *host identifier.*

Routing

- Routing is used for guiding the packets from its source to the destination.
- Routers are dedicated for this purpose.
- Routing is based upon the congestion information, shortest path method.
- These routers may be administered by a common authority and are called *Autonomous Systems (AS)* for which protocol- *Interior Gateway Protocol (IGP).* ASs of gateways exchange reachability information by means of *Exterior Gateway Protocol (EGP).*

Internet Group Management Protocol (IGMP)

- Multicasting in Internet is done with the help of multicast routers.
- About once a minute, each multicast router sends a hardware multicast to the hosts on its LAN asking them to report back on the groups their processes currently belongs to.
- Each host sends back responses for all the class D addresses it is interested in.
- These query and response packets use a protocol called *IGMP (Internet Group Management Protocol)*.
- **It has only two kinds of packets: query and response**, each with a simple, fixed format containing some control information in the first word of the payload field and a class D address in the second word.

Resource Reservation Protocol (RSVP)

- In order to ensure that the real-time traffic flows does not exceed that which is allocated for it, the resources required for each flow are reserved in advance of each packet flow starting.
- **The resources can be bandwidth and buffer capacity.**
- The protocol used to do this is called *Resource Reservation Protocol.*
- Because many of the new real-time applications involve multiple participants, RSVP is used to reserve resources in each router along either a unicast or a multicast path.
- When making a reservation, a receiver can specify one or more sources that it wants to receive from. It can also specify whether these choices are fixed for the duration of the reservation or whether the receiver wants to keep open the option of changing sources later.
- The routers use this information to optimize bandwidth planning. Once a receiver has reserved bandwidth, it can switch to another source and keep that portion of the existing path that is valid for the new source.

9.4. Quality of Service(QoS) and Resource Management

- During a multimedia communication, the services in the multimedia systems need to be parameterized.
- Parameterization of the services is defined in ISO standards through the notion of Quality of Service (QoS).
- Each service can be characterized by a quality of service. As a simple example some services are reliable i.e. they do not loose data while some are unreliable as they may loose data.
- The parameters can be bandwidth, maximum and minimum end to end delay, jitter, buffer allocation etc. There are several issues that need to be addressed and they are:

9.4. Quality of Service and Resource Management

QoS Layering

The QoS requirement is associated with each layer of the OSI model, or the TCP/IP model. However the QoS for multimedia communication system (MCS) consists of three layers: application, system and devices.

Application means the software and the program parameters, where as system refers to the overall system of communication and then in the network.

QoS-layered model for the MCS

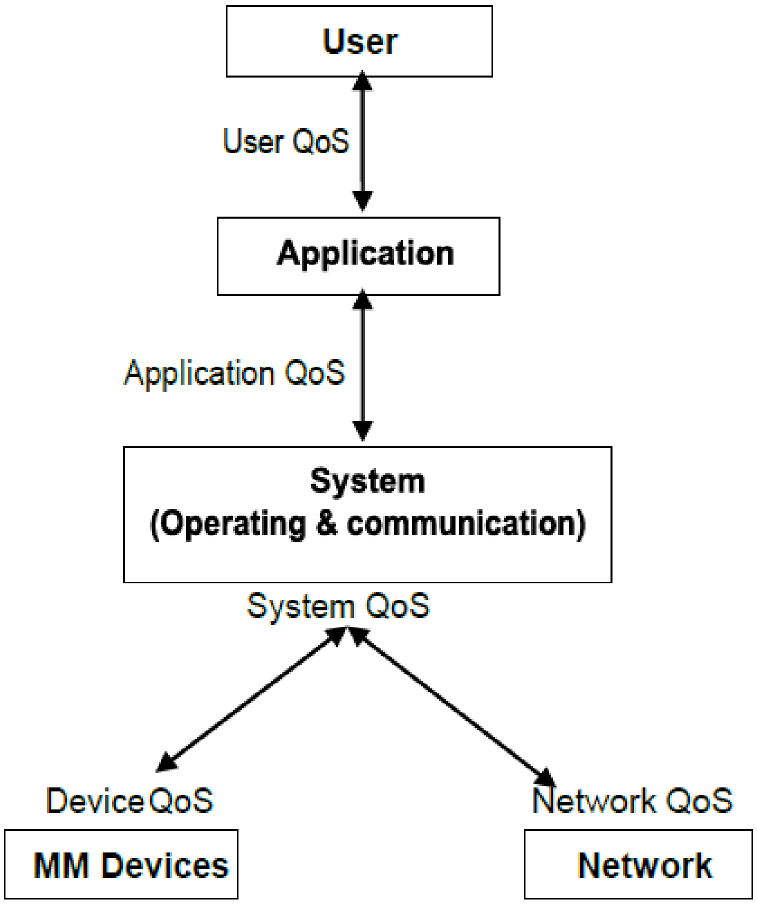


Figure: QoS-layered model for the MCS

QoS-layered model for the MCS

- **Service Objects**

  Services are performed on different objects, for example, media sources, media sinks, connections and Virtual Circuits (VCs), hence QoS parameterization species these *service objects.*

- **QoS Description**

  The QoS is described in terms of the required parameters by the end systems.

  The *application QoS* parameters may include media quality, transmission delay, jitter, synchronization.

  The system QoS parameters describe requirements on the communication services and OS services resulting from the application QoS. They may be specified in terms of both *quantitative* (bandwidth, PDU size, buffer size) and *qualitative criteria* (level of synchronization, order of data delivery, recovery).

QoS-layered model for the MCS

- *The network* QoS parameters describe requirements on network services. They may be specified in terms of network load (packet size, service time) and network performance (congestion, delay).
- *The device QoS parameters* typically specify timing and throughput demands for media data units.

QoS Parameter Values and Types of Service

There are three major type of service and they are:

- *Guaranteed Services*
- *Predictive Services*
- *Best-effort services*

In the **Guaranteed services** the QoS parameter values are deterministic or statistical in nature. It may ensure the lossless transmission of data i.e. reliable transmission.

In the **Predictive Services** the QoS parameter are predictable with the help of the past parameters. Though the exact value of the parameters may not be known a rough estimate can be made.

In the **Best-effort services**, the QoS parameters depend on the load of the network. It ensures that the best possible service is provided to the multimedia data.

9.4. Quality of Service and Resource Management

Resource Resource is a system used for processing, storing, manipulating data.

- The resource can be ***active* and *passive*.** The active resource can be a CPU which processes data or manipulates data where as the passive resource is a bandwidth which only serves a particular purpose.
- A resource can be *exclusive* i.e. used by a single process or it may be *shared* where it is shared between various processes.
- The resource may be *single* or it may be *multiple.*

Resource Management Architecture

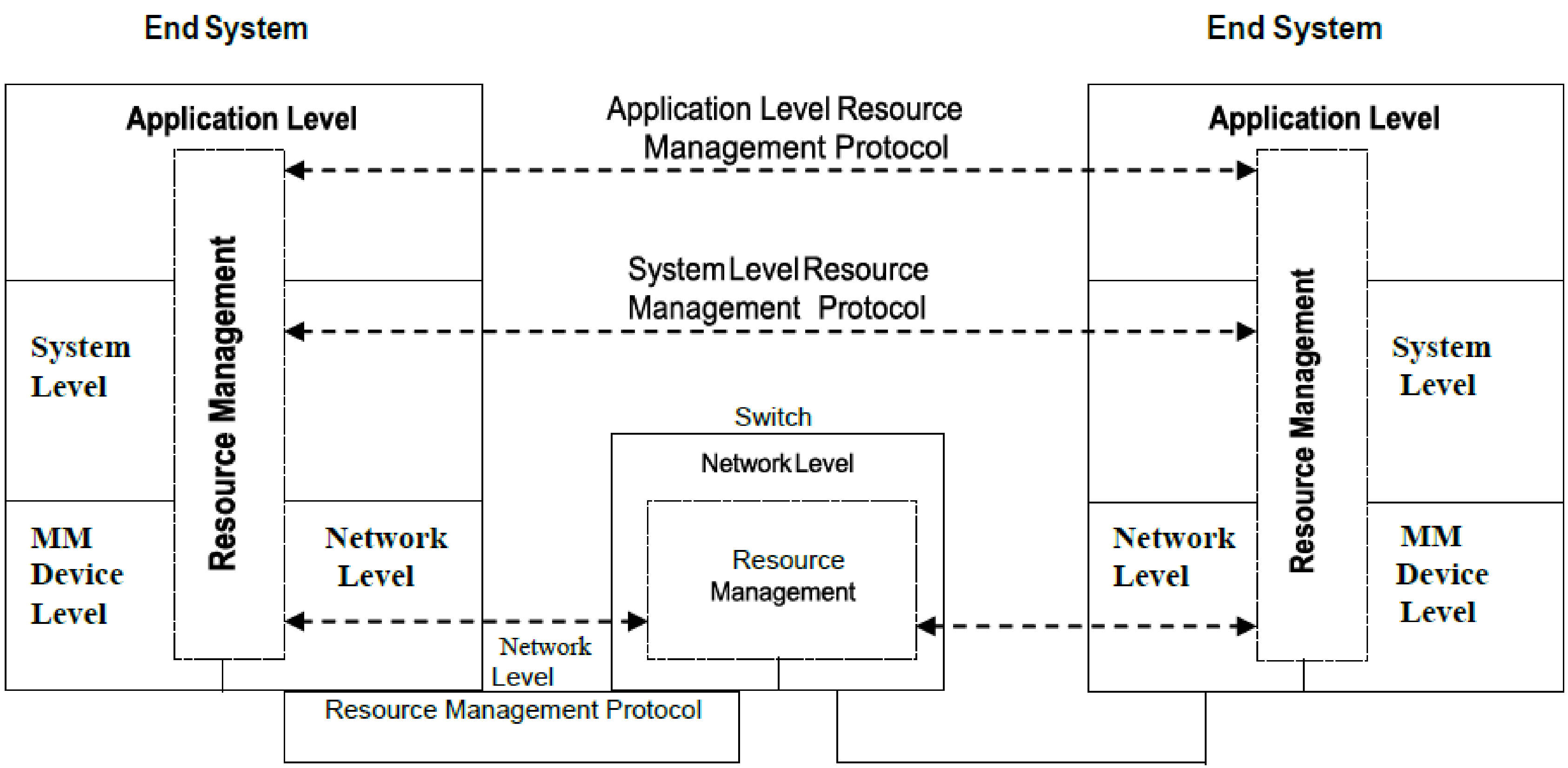


Figure: Resource Management in MCSs

Resource Management Architecture

The main goal of resource management is to offer guaranteed quality of service. It addresses three main actions

- Reserve and allocate resource
- Provide resources according to QoS specification
- Adapt to resource changes during on-going multimedia data processing.

Relation between QoS and Resources

The QoS parameters and their corresponding resources are mapped for the management of resources. For e.g. the end to end delay QoS parameter determines the behavior of transmission services along the path between source and sink with respect to packet scheduling, queuing and task scheduling. Description of a possible realization of resource allocation and management shows the QoS and resource relation.

QoS Negotiation (Cooperation)

- *Bilateral Peer-to-Peer Negotiation:* Negotiation occurs between the two service users and the service provider is not involved.
- *Bilateral Layer-to-Layer Negotiation:* It occurs between the service user and the service provider.
- *Unilateral Negotiation:* The user and the provider cannot modify the QoS parameter. It is based on “take it or leave it” model.
- *Hybrid Negotiation:* The negotiation between host and sender is bilateral layer-to-layer negotiation and negotiation between network and host-receiver is unilateral.

QoS Negotiation

- *Triangular Negotiation for Information: Exchange* The user specifies its required QoS parameters while the provider may change it according to the possibility and availability before confirming that the caller agrees upon it.
- *Triangular Negotiation for a Bounded Target:* It is similar to the above method but in this type of negotiation the caller specifies both the target value and the minimum required value. If the parameter that the provider does not meet the minimum required value the caller rejects the provider.
- *Triangular Negotiation for a Contractual Value:* In this case, the QoS parameters are specified through a minimal requested value and bound of strengthening. The goal of this negotiation is to agree on a contractual value, which in this case is the minimal request QoS parameter value.