1. Multimedia Systems Architecture

1.1. Components of a Multimedia System

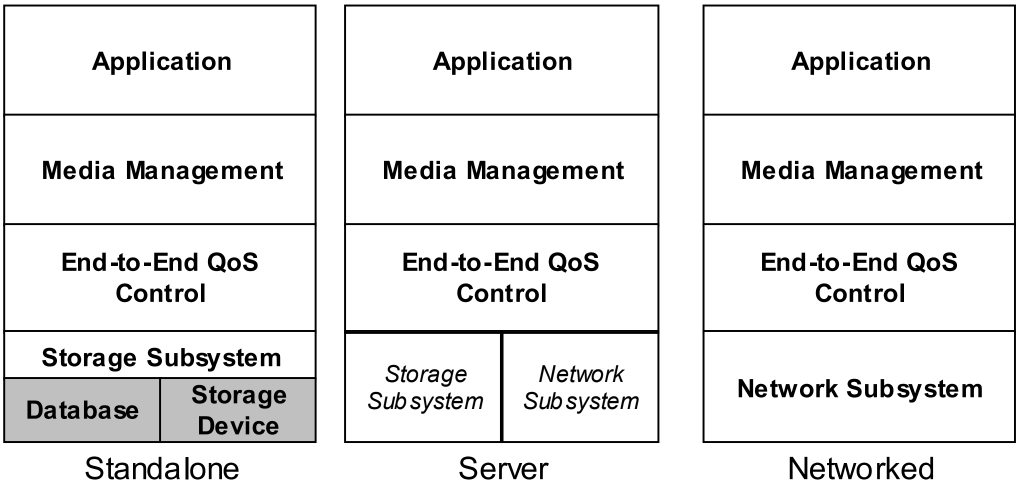
Multimedia applications can be categorized as ‘live’ or ‘orchestrated.’ Live applications are those which involve interaction among the users of the application, such as videoconferencing, and data is generated and reproduced on the fly. Orchestrated applications are those that reproduce multimedia data that was previously generated and stored in some medium, where the user can control and restart the play-back according to his requirements. Examples include CAI packages, pay-per-view movies, and business presentations.

In order to satisfy the requirements of these different categories of applications, a range of multimedia system architectures would be required, with different emphasis on the capabilities of individual components.

Multimedia systems can either be stand alone or networked. The architecture of a stand alone vs. a networked multimedia system are similar, the only difference being the presence of a storage subsystem vs. a network subsytem at the lowest level. A networked multimedia server would incorporate both subsystems.

The storage subsystem consists of databases to organize and keep track of the various multimedia objects that make up a given presentation, and storage devices that the databases and media contents reside in.

1.2. RT (Real Time) Multimedia Architecture



The RT architecture is a layered archicture, where the lower layer provides services that are utilized by higher layers to perform their functions. Using the RT model, we can analyze different aspects of multimedia systems by mapping their functions to one or more layers in the architecture.

- Storage Subsystem and Network Subsystem (Layer 1): These services are central to the functioning of the multimedia system and is normally provided as part of the operating system services.
- End-to-End QoS (Layer 2): This layer deals with maintaining connections between the source of multimedia content and the destination. For standalone systems, the source originates from the multimedia database, while networked systems sources from the multimedia server. The connections provide a logical link between the source and destination to provide a reliable channel of multimedia content with specified quality of service (QoS). QoS

comprises bandwidth, delay, delay jitter and packet loss probability.

- Media Management (Layer 3): Generic multimedia services such as media synchronization and media stream management are provided by Layer 3. This ensures that the temporal and spatial relationships among the various multimedia streams are maintained for the system.
- Application (Layer 4): Interface with the user, to provide access and controls to the presentation.

1.2.1. Impact of Application Requirements on Subsystems

The individual requirements of the application in terms of system resources can cause significant impact on required subsystem peak and average throughput (bandwidth or capacity). Allocating for ***peak throughput*** of the application results in wasted bandwidth, while allocating for ***average throughput*** results in significant delays in transmitting the real-time information. These delays might cause the destination system to lose synchronization among the various media streams. Media management and QoS are important for resolving some of these issues.

1.3. Storage Subsystem (Layer 1)

1.3.1. Storage Devices

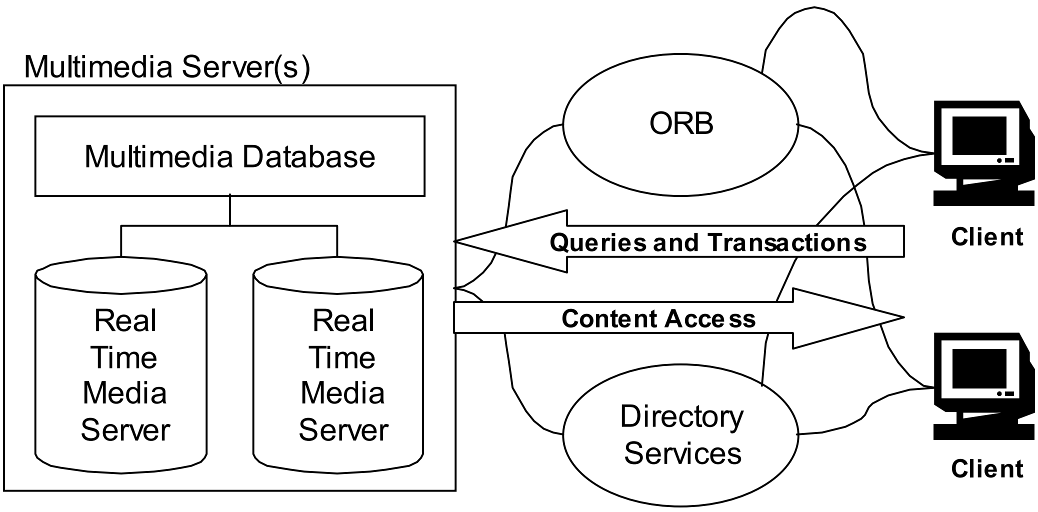
The storage subsystem consists of a data management component (databases), and storage devices that provide the physical storage media, such as Hard Disks, RAID, optical jukebox, etc. The parameters of interest are: throughput, access time, and storage capacity. Management of the storage devices is the task of the operating system. The file system supported by an OS is an important factor for multimedia storage, since the file sizes are usually huge and timing requirements stringent for real-time recording and playback.

1.3.2. Database

For standalone systems, the file system provided by the OS can be used to implement a simple database based on a hierarchical directory structure, where each subdirectory can represent a particular component in the database. In contrast, actual databases are used to manage the different media contents generated for multiple multimedia presentations, especially for multimedia servers. The choice of databases are important, since most traditional Relational DBMS are not able to handle multimedia content which are free-form and of much larger size than what can be stored in normal database tables. Normal relational databases are not able to handle the different media types such as text, graphics, audio and video, as well as the nonstationary nature of audio and video streams. Most RDBMS are designed for batch and non-real time operation. Consequently they are not suitable for use as multimedia databases.

However, extensions to RDBMS have been implemented, where a new data object type, BLOB (Binary Large Object) can be used to store multimedia information. This is often implemented as a separate storage server for managing those time sensitive content, and links to the objects are stored in normal RDBMS tables for access. However, searching the contents of individual BLOBs is not possible as the RDBMS does not understand how to interpret the content of the BLOBs.

Object oriented multimedia databases provide encapsulation for the data regardless of media type, and integrated access methods for searching and manipulating the data as part of the stored object.



The following characteristics are important for a multimedia database:

- Storage of various media types in a transparent manner, regardless of text, graphics, audio or video
- Handle access to multimedia data in a suitable and timely manner
- Search (query) for specific information in non-traditional media types, such as searching for a particular face in a database of images of different persons

- Maintain transaction integrity and version control for operations involving multimedia data, such as editing a collection of multimedia objects and storing the updated versions
- Locate a specific multimedia object in a network of distributed multimedia servers, via Directory Services and Object Request Broker (ORB) middleware such as Common Object Request Broker Architecture (CORBA) or Distributed Component Object Model (DCOM)

Directory Services and ORBs are designed to allow a system to locate a given object that exists on the network dynamically. This enables the system to be reconfigured and information to be stored at different locations without affecting the operation of a given application that depends on the services provided by the object.

ORBs are typically used to interface heterogeneous sytems in a distributed environment, and is typically implemented as ‘middleware’ that operates on top of the network subsystem to perform translation and conversion between data formats in different platforms. CORBA is an example of such middleware for heterogeneous systems, while DCOM is targeted for Windows-based systems.

1.4. Network Subsystem (Layer 1)

The network subsystem is critical for the real-time performance of networked multimedia presentations.

1.4.1. Network Components

Typical network components are:

- Information Source: Usually the server that is accessed by various clients. However, in a videoconferencing system, each client is also an information source
- Information Sink: The network component that ‘consumes’ the information generated by the source. This is typically the clients
- Hub: The computer systems are connected to a hub which acts as a repeater to transmit the data sent from one system to all the other systems attached to the same hub. The hub provides the similar functionality as a common network segment (bus) except that each machine is attached directly to the hub in a star configuration. Some hubs function as a bridge to interconnect different LAN segments or LAN technologies as well
- Switch: A switch isolates systems connected to it by only sending data that is intended for a given system to that recipient. This allows for simultaneous parallel transmission of data among different systems. E.g., if system A needs to communicate with system B (but not with C or D), and system C with system D (but not with A or B), then the two paris of systems can communicate simultaneously. In contrast, a hub would only allow one data transmission at any given time regardless of the recipient
- Router: The router connects a Local Area Network (LAN) to a larger network environment, such as a campus-wide backbone, or a Wide Area Network (WAN). It performs checking on the data transmitted to determine if it is intended for other LANs and

forwards it as necessary. It can also interface between different LAN technologies and protocols (e.g., Ethernet, FDDI, ATM, etc.)

- Backbone network: A high speed network that interconnects different LANs in a given organization together, to provide Intra-networking capabilities
- Widea Area Network (WAN): Provides access to the Internet where different organizations and users are able to exchange information. The links are typically low-speed (compared to LANs) as the cost of maintaining those links are high

1.4.2. Ethernet

Ethernet is the most common LAN standard today. It is a connectionless technology based on the concept of distributed access control, where each system determines whether it is possible to transmit data using the Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocol. This protocol is non-deterministic, where a station is not guaranteed access to the network within a given time interval. Instead, if the network traffic is overloaded, it could result in a system being unable to transmit data for an indefinite period.

As such, it is not optimal for multimedia applications where real-time delivery of information is important. However, the advantage of Ethernet is that it is a broadcast type network, where a system need only send one frame (addressed to a given group) and it would be received by all interested recipients.

Nonetheless, Ethernet based multimedia systems have been developed using a ‘best-effort’ approach, and by ensuring that adequate bandwidth is present to avoid the congestion problem. The evolution of Ethernet from 10 Mbps to 100Mbps and Gigabit speeds help to ensure that it will remain as a popular low-cost solution to LAN requirements.

1.4.3. ISDN

ISDN is a connection oriented technology, where systems that wish to exchange information must first establish a connection (circuit) among them (typically between two systems). A basic ISDN configurations provides 2 B channels (64 kbps each) for data or voice, and 1 D channel (16 kbps) for signalling and low speed data.

Since ISDN is connection oriented, performance is guaranteed to the user for real-time applications. However, the connection oriented nature and the limited bandwidth of ISDN channels (multiples of 64 kbps) means that it is not suitable as a backbone network connecting different LANs. ISDN is typically used for low-speed WAN access, and is becoming more common for remote/home access to the office, but is still not widely deployed in Malaysia.

1.4.4. Frame Relay

Frame Relay is a virtual circuit technology derived from ISDN and X.25 that allows multiple logical connections from one LAN environment to the WAN to be carried over the same interface. It has minimal error checking and error correction capabilities, thus ensuring that the Frame Relay operates at high speed.

Providing virtual circuits for different logical connections in Frame Relay is done via Statistical Time Division Multiplexing (STDM), where each logical connection ('virtual circuit') transmits data as variable length frames that are then grouped with frames from other logical connections to be transmitted over the Frame Relay 'cloud'. The 'cloud' is managed by the Frame Relay Service Provider which ensures that the connections are routed to the correct output port towards the destination and provide dynamic bandwidth allocation and QoS based on user requirements.

Dynamic bandwidth allocation refers to the ability to increase the allocated bandwidth on an application by application basis. This is determined during the virtual circuit setup. This enables applications that require additional bandwidth intermittently during its operation to be accomodated without preallocating the bandwidth and therefore underutilizing the extra bandwidth.

QoS parameters are used to negotiate and verify that the network service is operating as promised. Committed Information Rate (CIR) or throughput refers to guaranteed bandwidth for that virtual circuit that the network is expected to carry without any loss. The Committed Burst Size (B_c) over an averaging period (T_c) gives the CIR.

i.e., $CIR = B_c / T_c$.

Excess Burst Size (B_e) is the additional traffic that will be carried on a 'best-effort' basis, while any data beyond ($B_c + B_e$) will be discarded.

1.4.5. Asynchronous Transfer Mode (ATM)

ATM extends the virtual circuit and QoS concepts of Frame Relay to achieve an even higher performance networking infrastructure by specifying that all data is to be transmitted in fixed sized ‘cells’ that are 53 bytes in length (5 bytes header, 48 bytes data). The fixed sized cells simplify the processing of each cell, as well as improve the statistical multiplexing capabilities via fine grain control over each virtual circuit and bandwidth allocation. ATM cell processing is done in hardware to achieve very high performance and throughput.

ATM is designed to accommodate various different types of traffic, from real time voice and video traffic, to ‘elastic’ traffic such as file transfers and traditional Ethernet-based network applications. Real time traffic are often implemented using circuit-switched (connection oriented) technology such as ISDN and Frame Relay, while elastic traffic is carried over connectionless technology such as Ethernet. ATM is able to carry both types of traffic through the use of virtual circuits that have different QoS requirements. This can be regarded as tradeoffs between ‘delay variations’ and ‘cell loss probability.’

ATM can be viewed to have three main layers, namely the Physical, ATM and ATM Adaptation Layer (AAL). The Physical layer can operate over various types of media, to provide a scalable bandwidth from 25 Mbps to 155 Mbps and beyond. The function of the ATM layer is just to perform cell routing, flow control and virtual circuit multiplexing and demultiplexing. Since the operation of the two layers are simple in nature, they can be designed as hardware based switching circuitry.

The AAL layer is where the QoS is negotiated and managed. Several classes of traffic are defined, with Class A and B being real-time traffic, while Class C and D being elastic traffic. There is also an Available Bit Rate (ABR) and an Unspecified Bit Rate (UBR) class.

The various classes are identified using various AAL Type numbers. AAL1 refers to connection oriented Constant Bit Rate (CBR) real-time traffic, such as that generated by a videoconferencing application. AAL2 is Variable Bit Rate (VBR) real-time traffic, which can arise if the videoconferencing application utilizes compression (e.g., MPEG) to reduce the data rate. AAL3/4 and 5 are used for Variable Bit Rate (VBR) elastic traffic, such as data tranmission where timing is not critical. AAL 3/4 may be connection oriented or connectionless, while AAL5 is connection oriented.

AAL5 has been used to carry existing TCP/IP based Ethernet networks as it maps well to the ‘best-effort’ approach of Ethernet networks, with little extra overhead compared with AAL3/4. ATM can be used to replace both LAN and WAN connections as it is able to scale in bandwidth to accommodate different requirements.

The protocols for multimedia applications with stringent real-time requirements are based on AAL1 and AAL2 since they provide QoS guarantees that are negotiated at the beginning of the connection. The QoS parameters for ATM are: Cell Loss Rate (CLR), Cell Transfer Delay (CTD), and Cell Delay Variation (CDV).

CLR determines what happens when the instantaneous bandwidth capacity on the ATM switch is exceeded, if certain cells can be discarded. CTD affects the priority used for multiplexing of different virtual circuits, to ensure that cells arrive at the destination within a predetermined time interval. CDV is a measure of how much jitter is introduced into the data stream by the ATM network. This is important to ensure the smoothness of multimedia video and audio playback. If the network is able to meet the QoS request, then the connection is ***accepted***. Otherwise, it would either be renegotiated or else the connection is ***rejected***.