Multimedia communications

Georgios Tziritas
Computer Science Department
http://www.csd.uoc.gr/~tziritas

Computer communication networks

- Computer networks are essential to modern computing environment.
- Multimedia communications and networking share all major issues and technologies of computer communication networks.
- The ever-growing demands from numerous conventional and new generation multimedia applications have made networking one of the most active areas for research and development.
- Various network services and protocols are becoming a central part of most contemporary multimedia systems.

Quality-of-Service for multimedia communications

Challenges in multimedia network communications arise due to a series of distinct characteristics of audio/video data:

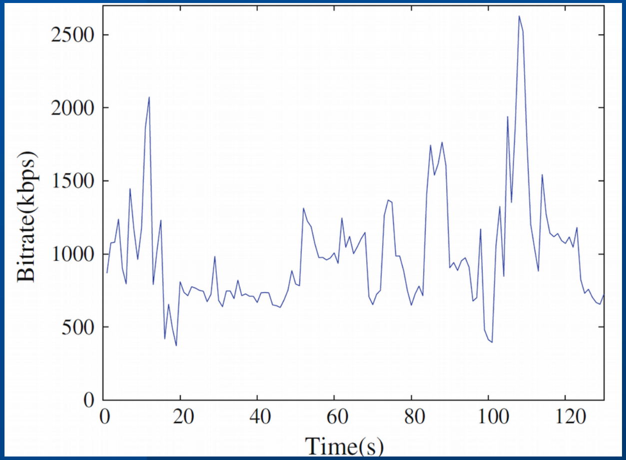
Voluminous and Continuous They demand high data rates, and often have a lower bound to ensure continuous playback. In general, a user expects to start playing back audio/video objects before they are fully downloaded.

Real-Time and Interactive They demand low startup delay and synchronization between audio and video for "lip sync". Interactive applications such as video conferencing and multi-party online gaming require two-way traffic, both of the same high demands.

Z.-N. Li, J. Liu and M. Drew, ***Fundamentals of Multimedia*****, Springer, 2nd edition, 2014**

Quality-of-Service for multimedia communications

Rate fluctuation The multimedia data rates fluctuate drastically and sometimes bursty. In a variable bit rate (VBR) video, the average rate and the peak rate can differ significantly, depending on the scene complexity.



MPEG-4 Video

Z.-N. Li, J. Liu and M. Drew, ***Fundamentals of Multimedia*, Springer, 2nd edition, 2014**

Quality-of-Service

QoS for multimedia data transmission depends on many parameters.

Bandwidth A measure of transmission speed over digital links or networks.
The data rate of a multimedia stream can vary dramatically, and both the average and the peak rates should be considered when planning for bandwidth for transmission.
Latency (maximum frame/packet delay) The maximum time needed from transmission to reception.
Packet loss or error A measure (in percentage) of the loss- or error rate of the packetized data transmission. The packets can get lost due to network congestion. They may also be delivered late or in the wrong order.
Jitter (or delay jitter) A measure of smoothness (along time axis) of the audio/video playback.
Technically, *jitter* is related to the variance of frame/packet delays.
Sync skew A measure of multimedia data synchronization. For a good *lip synchronization*, the limit of sync skew is 80 msec between audio and video.

Multimedia service classes

Here is a list of typical multimedia applications of different QoS demands:

Two-way traffic, low latency and jitter, possibly with prioritized delivery, such as voice telephony and video.
Two-way traffic, low loss and low latency, with prioritized delivery, such as e-commerce applications.
Moderate latency and jitter, strict ordering and sync, such as streaming video, web surfing and online gaming
No real-time requirement, such as downloading or transferring large files (movies).

Requirement on bitrate

Application	Speed requirement
Telephone	16 kbps
Audio conferencing	32 kbps
CD-quality audio	128-192 kbps
Digital music (QoS)	64-640 kbps
H.261	64 kbps-2 Mbps
H.263	<64 kbps
H.264	1–12 Mbps
MPEG-1 video	1.2–1.5 Mbps
MPEG-2 video	4–60 Mbps
MPEG-4 video	1–20 Mbps
HDTV (compressed)	>20 Mbps
HDTV (uncompressed)	>1 Gbps
MPEG-4 video-on-demand (QoS)	250–750 kbps
Videoconferencing (QoS)	384 kbps–2 Mbps

Tolerance on latency and jitter

Application	Average latency tolerance (msec)	Average jitter tolerance (msec)
Low-end videoconference (64 kbps)	300	130
Compressed voice (16 kbps)	30	130
MPEG NTSC video (1.5 Mbps)	5	7
MPEG audio (256 kbps)	7	9
HDTV video (20 Mbps)	0.8	1

Real-time Transport Protocol

Real-Time Transport Protocol (RTP) is designed for the transport of real-time data, such as audio and video streams. RTP resides in between the transport layer and the application layer, and bridges them for real-time multimedia transmission.

RTP's design follows two key principles, namely

- **application layer framing**, i.e., framing for media data should be performed properly by the application layer.
- **integrated layer processing**, i.e., integrating multiple layers into one to allow efficient cooperation.

RTP usually runs on top of UDP (User Datagram Protocol), which provides an efficient (albeit less reliable than TCP) connectionless transport service.

Real-time Transport Protocol

There are three main reasons for using UDP instead of TCP.

- First, TCP is a connection-oriented transport protocol; hence, it is more difficult to scale up in a multicast environment.

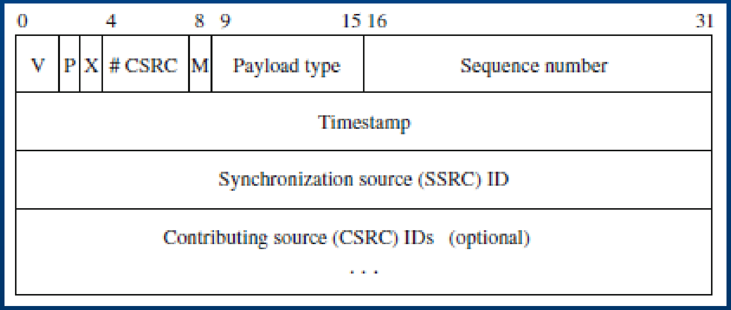
- Second, TCP achieves its reliability by retransmitting missing packets. Multimedia data transmissions is loss-tolerant and perfect reliability is not necessary. The late arrival of retransmitted data may not be usable in real-time applications.

- Last, the dramatic rate fluctuation (sawtooth behavior) in TCP is often not desirable for continuous media.

Real-time Transport Protocol

RTP introduces the following additional parameters in the header of each packet:

- **Payload type** indicates the media data type as well as its encoding scheme.
- **Timestamp** is the most important mechanism of RTP. The timestamp records the instant when the first octet of the packet is sampled.
- **Sequence number** is to complement the function of time stamping. It is incremented by one for each RTP data packet sent.
- **Synchronization source (SSRC) ID** identifies the sources of multimedia data.
- **Contributing Source (CSRC) ID** identifies the source of contributors, such as all speakers in an audio conference.



RTP Control Protocol

RTP Control Protocol (RTCP) is a companion protocol of RTP.
It monitors QoS in providing feedback to the source on quality of data transmission and conveys information about the participants of a multicast session. RTCP also provides the necessary information for audio and video synchronization, even if they are sent through different packet streams.

RTCP provides a series of typical reports:

- **Receiver report (RR)** provides quality feedback.
- **Sender report (SR)** provides information about the reception of RR, number of packets/bytes sent, and so on.
- **Source description (SDES)** provides information about the source.
- **Bye** indicates the end of participation.
- **Application-specific functions (APP)** provides for future extension of new features.

RTP / RTCP

RTP and RTCP packets are sent to the same IP address but on different ports.
RTCP reports are expected to be sent by all participants, even in a multicast session which may involve thousands of senders and receivers.
Such traffic will increase proportionally with the number of participants.
Thus, to avoid network congestion, the protocol must include session bandwidth management, achieved by dynamically controlling the frequency of report transmissions.
RTCP bandwidth usage should generally not exceed 5% of total session bandwidth.
While RTCP offers QoS feedbacks, it does not specify how these feedbacks are to be used, but leaves the operations to the application layer.
The rationale is that the multimedia applications have highly diverse requirements, and therefore no single set of operations can satisfy all of them. Instead, each application should customize their own operations with the feedbacks to improve QoS.

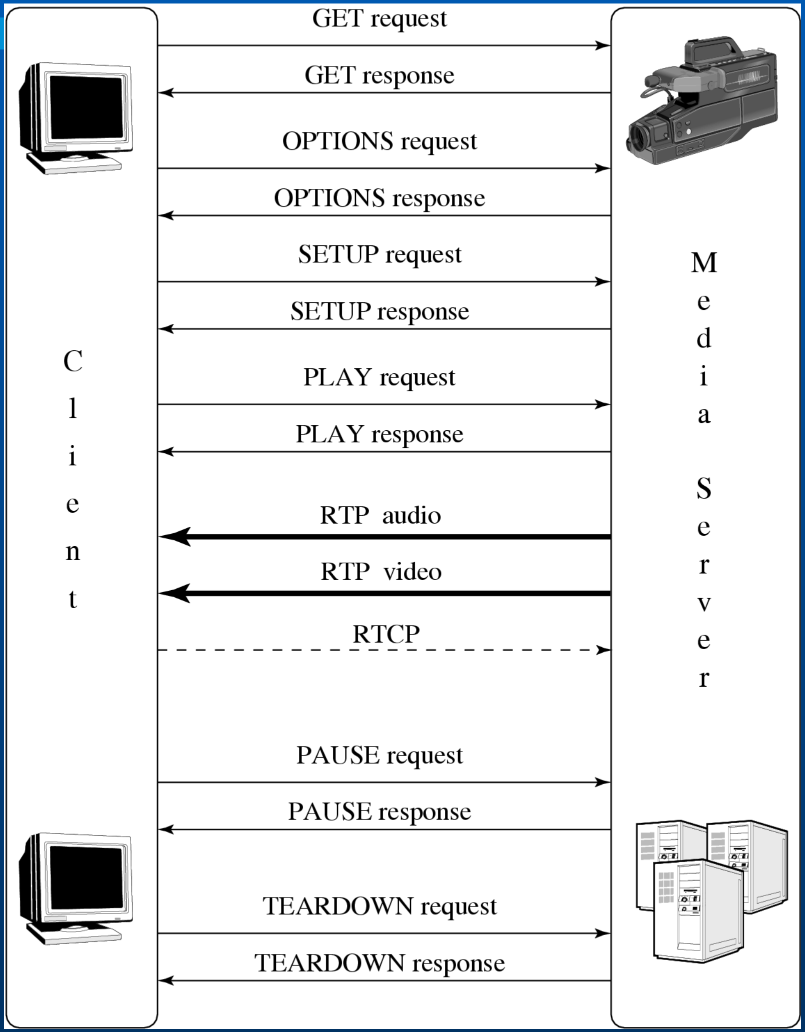
Real-time Streaming Protocol

The Real-Time Streaming Protocol (RTSP) is a signaling protocol to control streaming media servers and is used for establishing and controlling media sessions between end points.

Four typical RTSP operations:

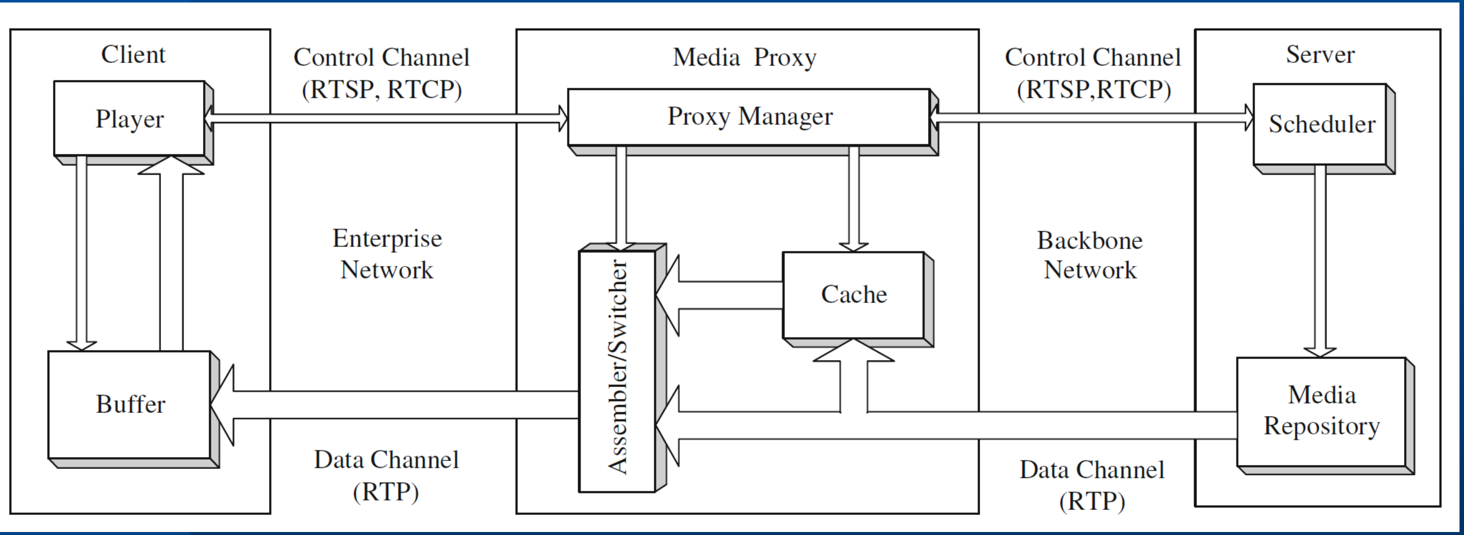
- **Requesting presentation description:** the client issues a DESCRIBE request to the Stored Media Server to obtain the presentation description.
- **Session setup:** the client issues a SETUP to inform the server of the destination IP address, port number, protocols, and Time-to-Live (for multicast).
- **Requesting and receiving media:** after receiving a PLAY, the server starts to transmit streaming audio/video data, using RTP.
- **Session closure:** TEARDOWN closes the session.

Real-time Streaming Protocol



Proxy caching

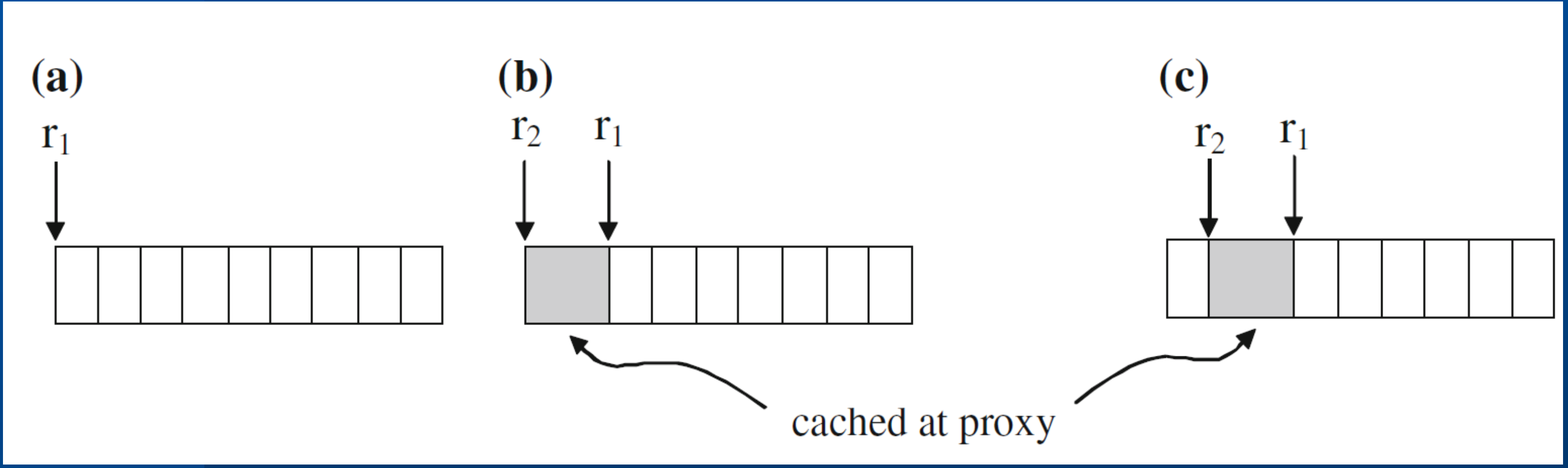
- Cache frequently used data at proxies close to clients
- Enhances the availability of objects
- Different from conventional web object caching:
  - An audio/video object is rarely updated
  - Caching each media object entirely at a proxy is hardly practical



Sliding-interval caching

A sliding interval of a media object is cached to facilitate consecutive accesses. Significantly reduce the network bandwidth consumption and start-up delay for subsequent accesses

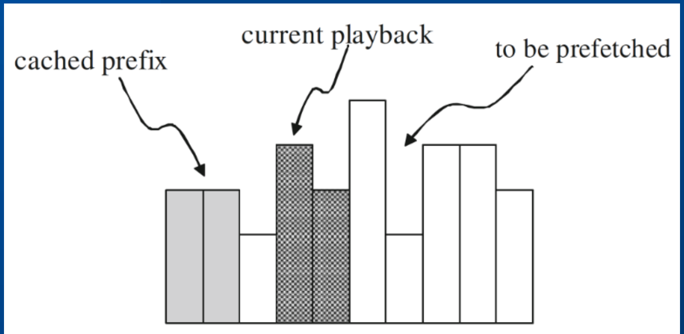
High disk bandwidth demands: double in the worst case due to the concurrent read/write operations

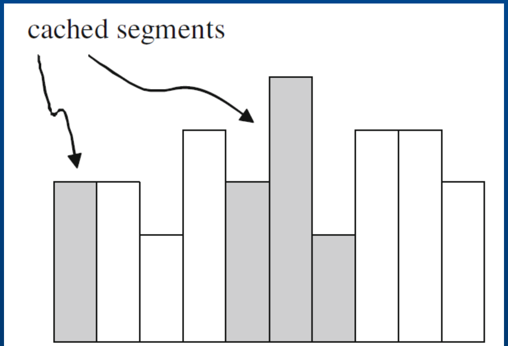


Z.-N. Li, J. Liu and M. Drew, ***Fundamentals of Multimedia*, Springer, 2nd edition, 2014**

Prefix caching and Segment caching

Cache the initial portion of a media object, called *prefix*, at a proxy
Then fetch the remaining portion, the *suffix*.
Reduce start up delay
Segment caching generalizes the prefix caching paradigm by partitioning a media object into a series of segments

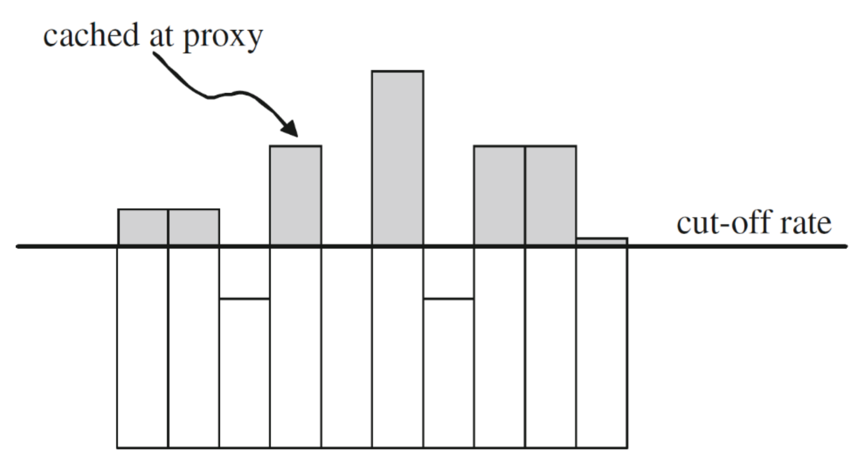




Z.-N. Li, J. Liu and M. Drew, ***Fundamentals of Multimedia*****, Springer, 2nd edition, 2014**

Rate-split caching

The three previous caching algorithms partition a media object along the time axis
Rate-split caching partitions a media along the rate axis
It is attractive for VBR streaming, as only a nearly constant rate
has to be delivered through the backbone network
Improves backbone network utilization for a QoS network with resource reservation



Z.-N. Li, J. Liu and M. Drew, ***Fundamentals of Multimedia*****, Springer, 2nd edition, 2014**

Content Distribution Networks

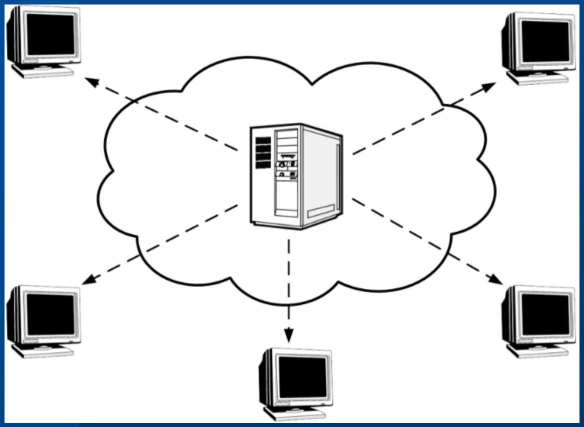
Caching is generally passive, in the sense that only if a user fetches an object, would the object be cached at a proxy. There will be no immediate benefit for the first user accessing an object.

A Content Delivery Network or Content Distribution Network (CDN) is proactive.

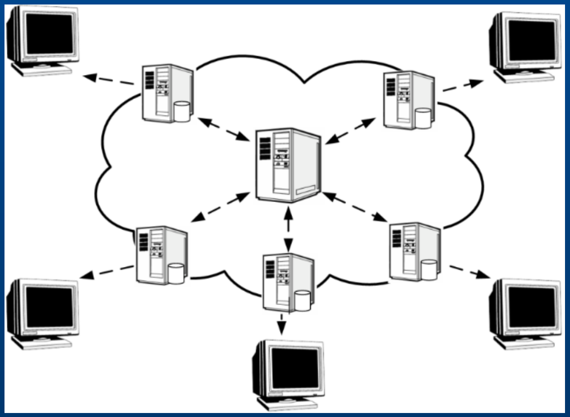
It is a large geo-distributed system of servers deployed in datacenters across the Internet; these servers replicate content from the origin server.

Today, CDNs serve a large fraction of the Internet data distribution.

Client/Server



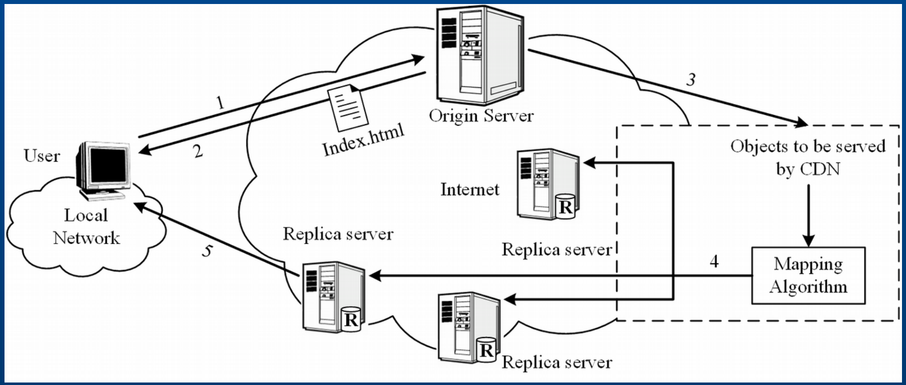
CDN



Z.-N. Li, J. Liu and M. Drew, ***Fundamentals of Multimedia*, Springer, 2nd edition, 2014**

Content Distribution Networks

- The mapping system is based on large amounts of historical and real-time data
- For performance optimization: the locations of the fewest hops or the highest server availability will be chosen
- For cost-optimization, the least-expensive locations can be chosen instead
- In real world both performance and cost-optimization considerations are combined



Z.-N. Li, J. Liu and M. Drew, ***Fundamentals of Multimedia*, Springer, 2nd edition, 2014**

Broadcast / multicast for heterogeneous users

The Internet’s intrinsic heterogeneity poses another challenge to multimedia multicast.

- Traditional end-to-end adaptation schemes: the sender adjusts its transmission rate according to some feedback from its receiver.
- In a broadcast/multicast environment: the traditional solution tends to be suboptimal, because there is no single target rate for a group of heterogeneous users.

It is thus necessary to use multi-rate multicast, in which the users in a multicast session can receive media data at different rates.

From the viewpoint of a media source, multi-rate streams can be produced via transcoding or scalable audio/video coding (layers).

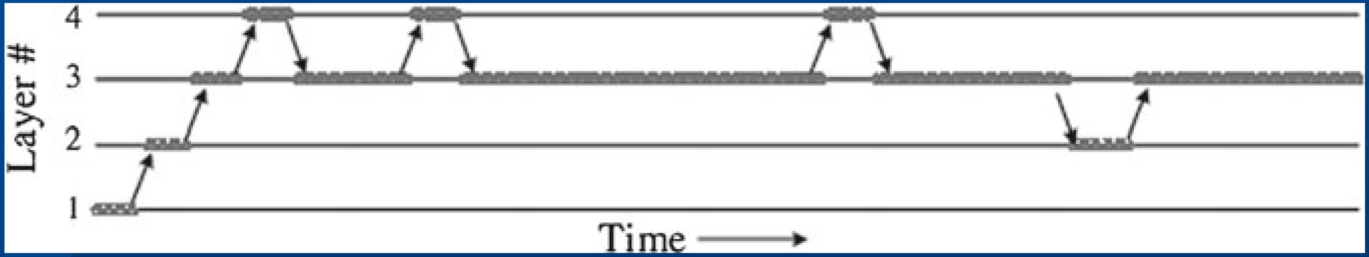
Broadcast / multicast for heterogeneous users

Multi-rate YouTube

FLV	320 x 240	100%
MP4	640 x 360	74%
FLV	640 x 360	66%
FLV	854 x 480	40%
MP4	320 x 240	26%

Layered multicast

- A user periodically joins a higher layer's group to explore the available bandwidth.
- If packet loss exceeds a tolerable threshold after the join-experiment, the user should leave the group; otherwise it will stay at the new subscription level.
- An exponential backoff ensures that the receiver will not be too aggressive in joining new layers and cause frequent congestion.



Z.-N. Li, J. Liu and M. Drew, ***Fundamentals of Multimedia*, Springer, 2nd edition, 2014**

Peer-to-peer video streaming

- Peer-to-peer (P2P) takes advantage of the ability of participating end-hosts, or *peers*, in a multicast group to contribute their uplink bandwidth. It was first brought to spotlight by the advent of Napster (1998) and Gnutella (2001).
- Later, the design philosophy in the highly popular BitTorrent software has converged with academic solutions in application layer multicast and a new generation of data-driven peer-to-peer streaming protocols on random mesh topologies emerged.
- A naive approach to distribute data without explicitly maintaining a structure is to use gossip algorithms. However, gossip cannot be used directly for video content distribution, because its random push may cause significant redundancy with the highbandwidth video.
- Thus, an important component of the data driven overlay is a scheduling algorithm, which strives to schedule the segments that must be downloaded from various partners to meet the playback deadlines.

HTTP for Streaming

- Hyper Text Transfer Protocol (HTTP) is used for streaming.
  - HTTP is generally firewall-friendly.
  - HTTP server resources are also widely available commodity.
  - Cost-effective using the existing web infrastructure to support HTTP streaming for massive audience.
- HTTP was not initially designed for streaming applications. The key to support streaming with HTTP is to break the overall media stream into a sequence of small HTTP-based file downloads.

HTTP for Streaming

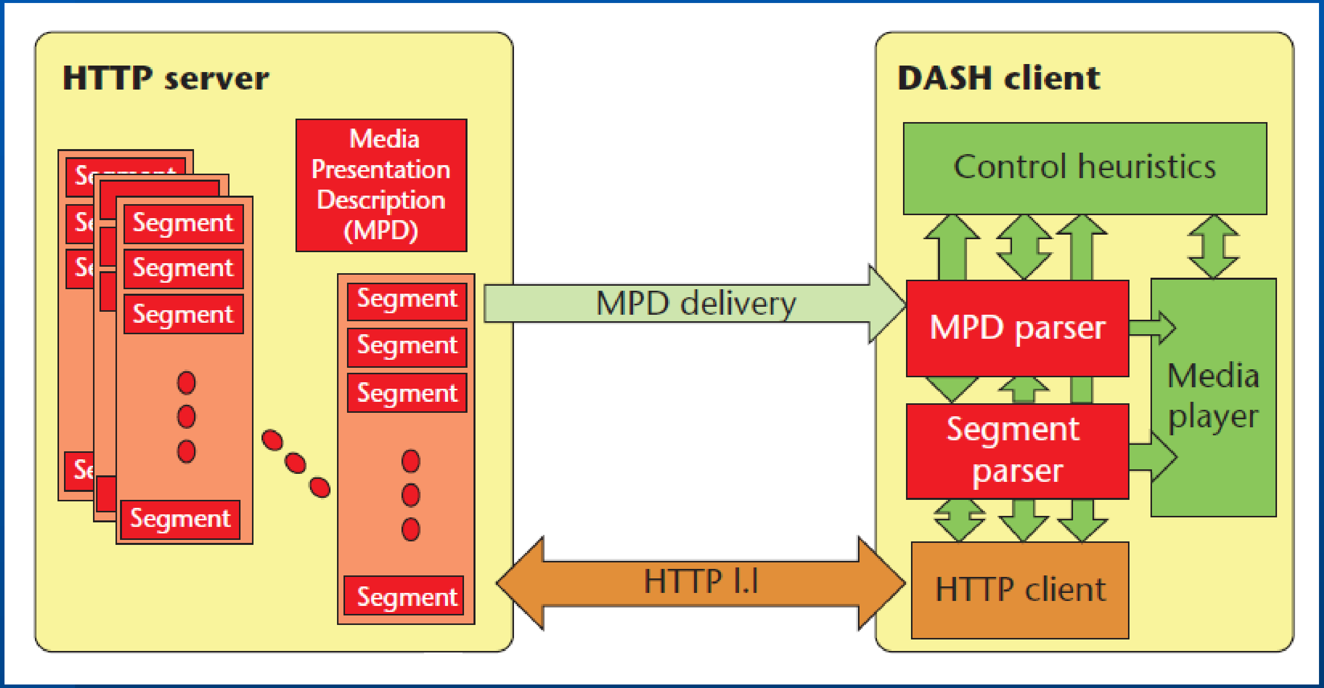
- HTTP does not maintain session states on the server.
  - Does not impose significant cost on server resources.
  - Quite different from RTP/RTCP/RTSP-based streaming.
  - Each client can keep a record of its playback progress.
  - The progressive download allows a client to seek to a specific position in the media stream by downloading the corresponding file.
- HTTP streaming has been implemented in commercial products.

  Netflix, Youtube, Hulu
- HTTP streaming is benefiting from the rapidly expanding capacity and dropping pricing of today's CDNs.

MPEG-DASH
Dynamic Adaptive Streaming over HTTP

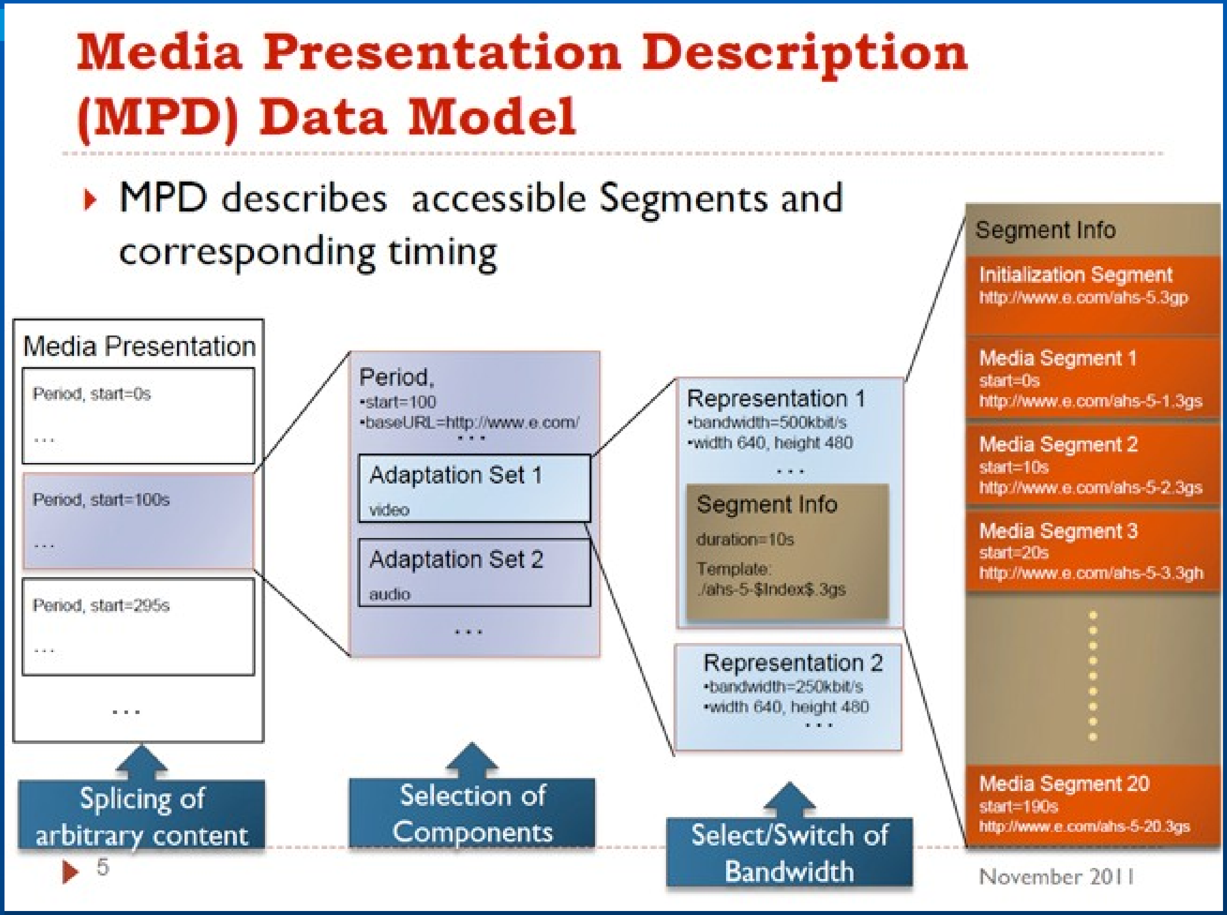
- The Dynamic Adaptive Streaming over HTTP (DASH) standard has been developed by the MPEG group for the following reasons.
  - The demand for standardization so that different devices can inter-operate.
  - The heterogeneous networks and devices also require the media streaming to be dynamical and adaptive.
- DASH became an international standard in November 2011
- DASH defines a set of implementation protocols across the servers, clients, and description files.
- In DASH, a video stream is encoded and divided into multiple segments.
  - Initialization segments that contain required information for initializing the media decoder.
  - Media segments that contain the media data and the stream access point.

MPEG-DASH

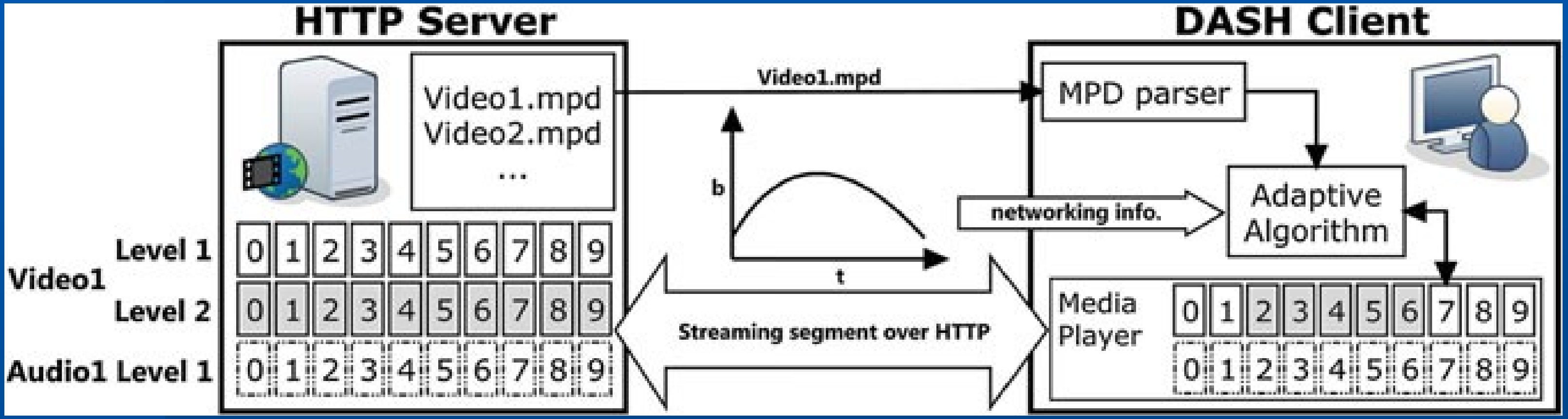


I. Sodagar, **The MPEG-DASH Standard for Multimedia Streaming over the Internet,** ***IEEE MultiMedia Magazine*****, Oct. 2011.**

MPEG-DASH


Media Presentation Description (MPD) Data Model
MPD describes accessible Segments and corresponding timing
Media Presentation
Period, start=0s
…
Period, start=100s
…
Period, start=295s
…
…
Period,
•start=100
•baseURL=http://www.e.com/
…
Adaptation Set 1
video
Adaptation Set 2
audio
…
Representation 1
•bandwidth=500kbit/s
•width 640, height 480
…
Segment Info
duration=10s
Template:
./ahs-5-$Index$.3gs
Representation 2
•bandwidth=250kbit/s
•width 640, height 480
…
Segment Info
Initialization Segment
http://www.e.com/ahs-5.3gp
Media Segment 1
start=0s
http://www.e.com/ahs-5-1.3gs
Media Segment 2
start=10s
http://www.e.com/ahs-5-2.3gs
Media Segment 3
start=20s
http://www.e.com/ahs-5-3.3gh
Media Segment 20
start=190s
http://www.e.com/ahs-5-20.3gs
Splicing of arbitrary content
Selection of Components
Select/Switch of Bandwidth
5
November 2011

MPEG-DASH



Type	Server	Client
Adobe adaptive streaming	Flash media server	Flash media player
Apple HTTP Live streaming	Generic HTTP servers	QuickTime/iOS player
Microsoft Live smooth streaming	Internet information services (IIS)	Silverlight player