A COMPARATIVE ANALYSIS OF IP STREAMING VIDEO VERSUS MPEG VIDEO OVER CABLE

Prashant P. Gandhi, Ph.D.[1]

Cisco Systems, Inc.

Abstract

Cable networks have been designed to carry broadcast digital video services for consumer entertainment. Recent popularity of cable high-speed data service indicates the possible use of the cable DOCSIS infrastructure to deliver streaming media services. This paper presents a high-level comparison of video delivery over MPEG and over IP transport mechanisms. Issues such as broadcast and point-to-point transmissions, quality of service (QoS) and set-top box support are discussed.

1. INTRODUCTION

The MPEG-2 standard for compression and transport of entertainment-quality video has been critical for offering digital television (DTV) services to cable consumers. The broadcast nature of MPEG transport allows its efficient delivery over a shared cable infrastructure. The quality and convenience of digital programming has made the DTV service quite popular in the cable community. Digital set-top boxes (STBs) capable of decoding MPEG-2 video are being deployed in growing numbers in consumer homes.

The two-way capable hybrid fiber coax (HFC) cable plant has also allowed offering of DOCSIS data service to consumers. Cable modem service for high-speed internet access has been making significant in-roads in internet-hungry homes. By the end of 2000, nearly 5.5 million US and Canadian homes subscribed to cable modem service [1]. Of these, 3 million subscribers were added in 2000 (with 1 million added in the 4th quarter alone). Appetite for broadband cable data service is also fueled by telecommuters that require high-speed access to corporate intranets and whose monthly subscription fees are subsidized by their employers.

Broadband has enabled the delivery of internet-based video (a la streaming media) to the PC. Growing number of websites are now providing streaming video at broadband bitrates, typically in the range 128 kb/s – 1 Mb/s. These are both broadcast (multicast) programs, such as television and radio channels, and on-demand programs such as news, sports, music, music videos, international, movie trailers and independent films. Also, Internet video is typically encoded in proprietary formats, such as Real Networks, Microsoft Windows Media Technology (WMT) and Apple QuickTime (QT).

With nearly 70 million cable-ready homes, the growth of digital TV and data services are expected to remain in heavy demand for the next several years. Clearly, these two services have traditionally addressed two different market segments. The DTV service provides television entertainment in the family room whereas high-speed data (HSD) service provides internet access to PC in the study.

[1] Prashant Gandhi (gandhi@cisco.com) is a Technical Leader in the Service Provider Line of Business at Cisco Systems.

This differentiating line between DTV and HSD services have now started to blur with the introduction of interactive television (iTV) STBs. These STBs, running iTV middleware software (e.g. from Liberate or Microsoft), integrate the television experience with internet-based broadband interactivity. This broadband interactivity is achieved by the integration of a DOCSIS cable modem with a digital STB.

Internet streaming media, as it stands today, is mostly viewed at little or no cost to consumers (other than the cost of HSD service). With broadband interactivity to the TV, there is an opportunity to provide IP streaming media services over DOCSIS. These services may be provided as part of a premium package or as part of additional subscription and/or pay-per-view fees. They can be delivered to a PC or a TV.

Niche content, generally not available on the television line-up, can be offered over cable's IP/DOCSIS network. Dynamic content, such as news and sports, international content, children's programming, etc. are examples of content that consumers may pay for to watch on demand. The value is in the flexibility of on-demand service in terms of convenience, choice and control. Portability of content can also be a value add – consumer should be able to watch a program on PC and transfer it to TV at a moment's notice.

This paper presents a comparison between MPEG and IP video and discusses some of the issues and challenges in deploying IP video to the TV.

2. CHARACTERISTICS OF MPEG AND IP VIDEO

MPEG Video [2,3]

In MPEG, the source of video is typically an MPEG-2 compressed video bitstream (in the form of packetized elementary stream or PES) that is sent to an MPEG transport stream (TS) multiplexor. The multiplexor takes one or more elementary streams and produces a single- or multi-program transport stream.

Each PES is packetized into 188-byte packets, with 184-byte payload and 4-byte header. Video, audio and private data streams are individually packetized. Video and audio streams are also time synchronized to help guide the decoder assembled a synchronized audio/video presentation (a la lip synchronization).

Each packetized stream is marked with a packet ID (PID). Audio/video/data PIDs of a specific program are sent separately as part of program specific information (PSI) tables. PSI tables include: program allocation table (PAT) with PID = 0 to provide map information of all transport stream programs, program map table (PMT) indicating PIDs for each program within the TS, conditional access table (CAT) with PID = 1 for PIDs of entitlement messages and tables that carry private data.

The MPEG TS is potentially encrypted and then protected via forward error correction (FEC) as per ITU J83 Annex B FEC. Quadrature amplitude modulation (QAM) is then applied to the FEC stream with 6 bits/symbol (QAM-64) or 8 bits/symbol (QAM-256).

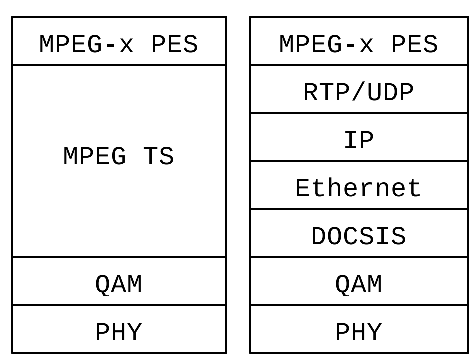


Figure 1: MPEG and IP Video Stack

As indicated earlier, MPEG TS packets do not carry any source or destination information. For a point-to-point video session (e.g. for on-demand video), PID information and entitlement information (for an encrypted stream) must be sent to the client. Also, for a point-to-point session, a control session needs to be established between client and video server. This is typically done via proprietary protocol or via MPEG's digital storage media – command and control (DSM-CC) protocol. Due to the complexity and interoperability issues with DSM-CC, there has been some movement to adopt the real-time streaming protocol (RTSP) for session control.

<u>IP Video</u>

With IP, both point-to-point (unicast) and point-to-multipoint (multicast) transmission is possible. This allows on-demand and broadcast video services to be offered on a single infrastructure. In addition, the IP infrastructure can be shared among multiple services, including data, telephony and video.

The video source for Internet protocol (IP) based transport can still be MPEG-1/2/4 PES. The PES stream is converted to a stream of IP packets which are then transported via the unreliable user datagram protocol (UDP). Each UDP/IP packetized video or audio stream is appended with real-time transport protocol (RTP) header to allow stream synchronization, time stamps and sequence numbering [4].

Typically, to avoid fragmentation, each RTP/UDP/IP packet size is less than one Ethernet frame payload (~1470 bytes). RTP/UDP/IP header overhead is at least 40 bytes (12 bytes for RTP, 8 bytes for UDP and minimum 20 bytes for IP). Clearly, the IP packet overhead is much larger than that for MPEG TS packets. However, for larger video packets (~1470 bytes), this overhead is not as significant.

Also, RTP packet header compression may be employed to reduce the 40-byte overhead to only several bytes [5]. Standardized mechanism of packetizing MPEG-1/2 video over RTP is specified in IETF RFC 2250 [6].

Real-time streaming protocol (RTSP) is employed to control a video session [7]. RTSP opens a separate connection (typically TCP/IP) between the client and video server. This is shown in Figure 2. RTSP allows session initiation, VCR-like session control such as play, pause, stop, etc., and session termination. RTP session parameters are communicated using RTSP. An optional communication channel via real-time control protocol (RTCP) can also be negotiated to allow the client to communicate additional information (e.g. packet loss) to the server. RTSP is also used to communicate multicast program information to the client.

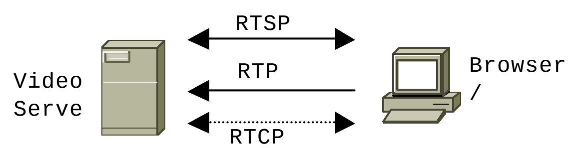


Figure 2: Client-server interactions for IP Video

Though it is important to support standards-based video for coding, transport and storage, today's Internet streaming media is dominated by proprietary systems of Real Networks, Microsoft and Apple. This is shown in Figure 3. Real Networks uses proprietary transport called Real Data Transport (RDT), proprietary G2 codec and proprietary file format (e.g. .rm, .ra). Microsoft WMT uses the proprietary control and transport protocol MMS and proprietary ASF file format. WMT does support standard MPEG-1/2 codecs as well as its own proprietary audio/video codecs. Apple QT uses the standards-based RTSP and RTP for control and transport and its file format has been adopted (with some variation) for storing MPEG-4 video.

Vendor	Product	Platform	Protocols	Codecs	File Formats
Real Network	Real	NT, Unix	RTSP, RDP	G2 (Proprietary)	RM
Microsoft	WMT	NT/W2K	MMS	Proprietary and MPEG-1/2	ASF
Apple	QuickTime	MacOS, Linux	RTSP, RTP	H.261, Sorenson (proprietary)	QT

Figure 3: IP Streaming Media Formats

However, its primary codec is proprietary (supplied by Sorenson).

Proprietary nature of IP video is not as critical on the PC as it is on the STB. STB's resources utilization (processing power, run-time memory, codesize, etc), player stability and frequent software upgrades are issues that need to be addressed for a scalable deployment of IP video to TVs.

Streaming Media Session Interactions

The following illustrates the interactions for establishing an on-demand streaming media session between a client and a Real Networks streaming media server.

1. User clicks on a web link: http://www.strartrek.com/ramgen/unimatrix-zero.rm
2. Real server sends a .ram file that contains an RTSP link (rtsp://www.startrek.com:554/voyager/unimatrix-zero.rm)
3. The browser forks off a Real player in a separate window (based on the MIME type received with the .ram file, e.g. x-pn-realaudio). The Real player connects with the server via RTSP (rtsp://www.startrek.com:554/voyager/unimatrix-zero.rm)
4. The server opens an RDT connection to the player and begins streaming.
5. The player sends an RTSP command to terminate the session.

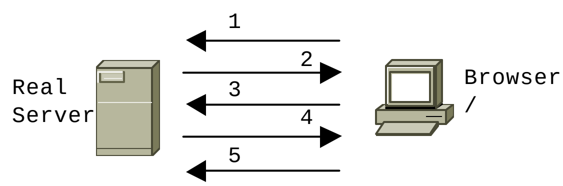


Figure 4: Client-server interactions for Real video

3. IP VIDEO DISTRIBUTION

Distribution of video over core and access network is necessary for delivery of IP video services. This distribution architecture is dependent on the type of video service offered. This is discussed below.

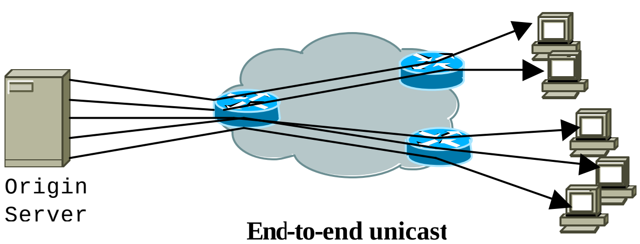
There are two main categories of video service: on-demand and live/scheduled. An on-demand program (such as video-on-demand or VoD) is receiver-controlled in that each receiver fully controls the streaming session. At any time, the receiver can initiate the program stream, control it via VCR-like functions (or "trick" modes) such as fast forward, rewind, pause and also terminate the program.

Live/scheduled programs, on the other hand, are source-controlled. A receiver can only tune to such a program (to listen to it, watch it, etc.) but has no ability to further control it. This is because live/scheduled programs are transmitted continuously by the server during pre-defined time intervals. The difference between a live versus scheduled program is primarily based on the source of content. For a live program, content source is a real-time event that is encoded and transmitted as it is happening; Content source of a scheduled program, on the other hand, is a pre-recorded event stored on a video tape, DVD/CD, film, etc.

There are multiple ways to transmit a video program over an IP network. These include unicast, multicast, splitting (or reflected unicast) and edge multicast (or reflected multicast).

Unicast

In a unicast transmission, each receiver receives its own stream from the media server (see figure below). If there are one thousand active receivers, then there are one thousand streams being served by the media server. Though unicast allows a receiver to fully control the stream, it quickly becomes unscalable due to the significant demand it places on network and server resources. For live/scheduled programs, unicast is clearly quite inefficient.

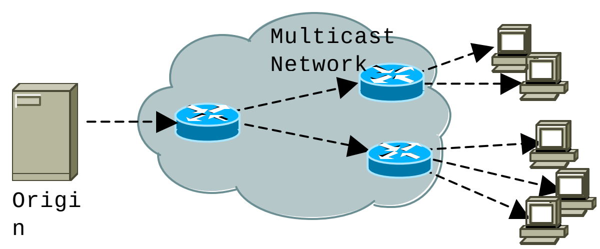


End-to-end unicast

Multicast

Multicast is an efficient Layer-3 mechanism for delivering a media program to multiple receivers. Here, the server sends out a single program stream to a multicast group[2]. The multicast-enabled IP network to which this server and the intended group of receivers are connected is responsible for delivering this stream. If the receivers are spatially distributed (in network sense), then the IP network replicates the stream as necessary to reach all receivers of that multicast group. Multicast is quite efficient for live/scheduled programs because (1) it requires minimum server resources – only one stream per program regardless of the number of receivers and (2) minimum stream replication in the network – at most one stream per link depending on the receiver distribution. This is shown in the figure below.

Note that the server continuously transmits each live/scheduled program stream. Receivers join and leave the program asynchronously using the Internet Group Multicast Protocol (IGMPv2, RFC 2236). The IP multicast network is responsible for building an efficient tree to distribute multicast traffic. Edge network devices are responsible for maintaining the group membership of receivers and replicating streams to them.



End-to-end multicast

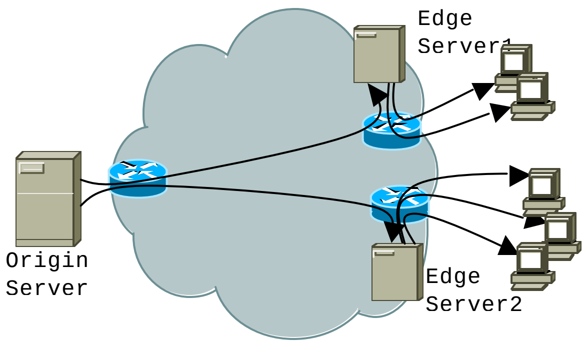
For an overview of IP multicast, visit http://www.cisco.com/warp/public/cc/pd/iosw/prodlit/ipimt_ov.htm. For a multicast quick-start configuration guide for routers, visit http://www.cisco.com/warp/public/105/48.html.

Splitting

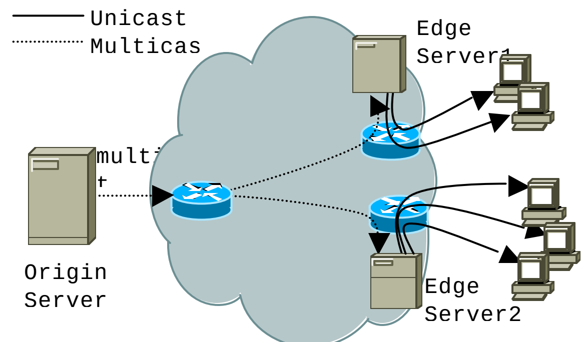
The term "splitting" is coined by Real Networks for an efficient delivery mechanism of live/scheduled programs over a non-multicast network. In IP multicast, stream replication occurs by IP network elements whereas, in splitting, replication is performed by video servers that are placed within the network. This is illustrated in the figure below.

Splitting allows tree-based distribution of unicast streams to the network edges. Each receiver

[2] IANA has assigned the Class D address 224.0.0.0 - 239.255.255.255 for multicast groups (i.e. destination IP addresses).



End-to-end Splitting



Edge Splitting

connects to an edge server close to it (as determined by the request routing mechanism). The edge receiver then determines a path from it to the origin server. The number of program streams served by the origin server equal the number of servers participating in that program at the next lower hierarchy. In the 2-level hierarchy shown in the figure above, both 2^{nd}-level (edge) servers are participating in streaming the program. Thus, the origin server is sending two unicast streams, one to each participating edge server. Each edge server splits the incoming stream into multiple unicast streams, one per each participating receiver.

Edge Multicast

Even with splitting, the edge server fan-out can be a concern when serving to a large community of users. Hence, it is beneficial to convert the edge portion of the non-multicast network to multicast. Streams traverse most of the network as unicast. The last-hop delivery of this stream, from edge to user community, is via multicast. As before, each receiver requests the edge network element to join to a multicast group via IGMPv2.

There are two mechanisms to deliver program streams over a non-multicast network: splitting and GRE tunnels. In a split-based distribution, each edge network element receives a multicast program from some edge server. Each such edge server receives the stream from the origin server through a tree-based server hierarchy.

Alternatively, the origin server continues to send a multicast stream. GRE tunnels are created within the non-multicast network portion to encapsulate multicast traffic.

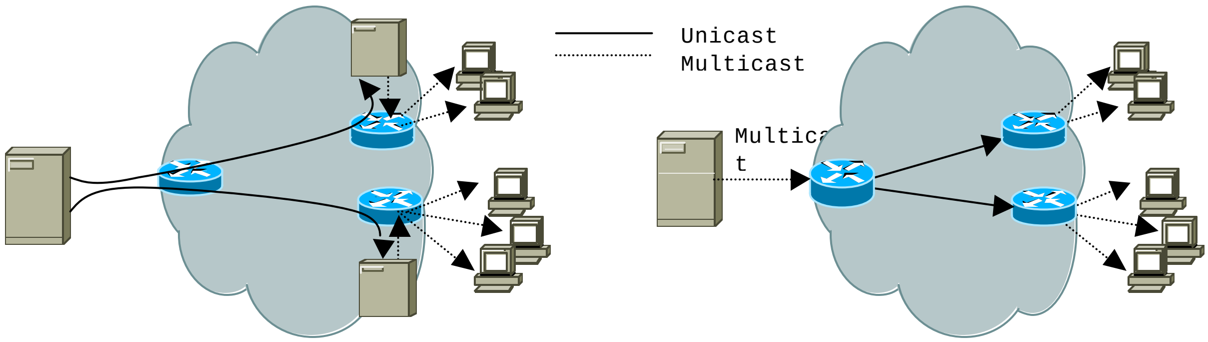
In either case, it is desirable to construct unicast delivery paths (trees or tunnels) dynamically, as per user demand. If no users are joined to a multicast program at a given network edge, then there is no need to send that stream to this edge.

4. IP VIDEO OVER CABLE NETWORK

Most two-way HFC networks today consists of broadcast video delivery via MPEG TS and HSD delivery via IP/DOCSIS. DOCSIS is assigned a separate 6MHz QAM.

Broadcast MPEG TS channels are statically assigned within a single QAM. For instance, 10 broadcast video channels, each having bitrate of 3.5Mb/s, can be multiplexed in a single QAM-256 (with aggregate bitrate of ~38Mb/s).

Peak concurrent IP video streams over DOCSIS QAM is the same as that for MPEG QAM for the same stream bitrate. One major issue with DOCSIS is that only one DOCSIS channel can be assigned to a cable modem. This limits the number of subscribers that can access video simultaneously. This is exacerbated by sharing of DOCSIS for HSD and potentially voice telephony. Dynamic channel change (DCC) will



Core splitting, Edge multicast **Unicast tunnel in the core, Edge multicast**

be a critical feature for scalable deployment of data, voice and video services over DOCSIS.

Additionally, like MPEG video, IP video requires many more downstream QAMs. This implies that QAM density on cable modem termination systems (CMTSs) needs to increase so that the economics of IP video delivery become reasonable.

Detailed traffic engineering analysis for multimedia services over HFC is presented in [12].

Content Distribution

In a typical cable network, the regional head-end and local hubs form a two-level content distribution hierarchy that is well suited for streaming media delivery. High-demand content is located as close to the subscriber as possible to reduce bandwidth demands of streaming media in the backbone network. Less popular content is available at the headend on an on-demand basis.

Each local hub contains a cluster of streaming media servers, caches and storage devices, collectively referred to as the edge media server. and can intelligently balance the load across these servers for optimum performance and reliability. The headend contains a core media server which typically has a several times bigger cache and storage system. In a hierarchical streaming media architecture, unicast streams are generated by edge media servers whereas multicast streams are generated by the core media server. The headend also contains content acquisition, management and distribution systems. New content is brought in via satellite, encoded (if required) and stored in the core media server. Content management and distribution system allows a cable operator to centrally manage media content, to set policies for cached and stored content and to distribute content reliably to local hubs.

Centralized vs Distributed Architecture

In a centralized architecture, a core media server in the headend serves all cable subscribers. Two issues need to be considered in this situation: server bandwidth utilization and backbone fiber ring bandwidth utilization. To understand server bandwidth utilization, consider Table 1 below which depicts media storage and server bandwidth requirements as a function of stream rate. For instance, approximately 600 Mb/s of server bandwidth is required to serve 2000 simultaneous streams at 300kb/s.

Figure 4 shows the fiber ring bandwidth utilization as a function of data and streaming media penetration rates (same as the ones used in Tables 1-3). Both OC-48 at 2.5 Mb/s and OC-192 at 10 Mb/s rates are considered for the fiber ring. It is clear that the scenario of a centralized web cache and a centralized media cache consumes over 60% the OC-48 ring bandwidth at low penetration rates whereas this scenario is

	Stream Rate (kb/s)				
	300	**500**	**700**	**1000**	**1500**
Storage Rate (MB/Hr)	135	225	315	450	675
Storage (GB)					
100 Hrs	14	23	32	45	68
200 Hrs	27	45	63	90	135
500 Hrs	68	113	158	225	338
1000 Hrs	135	225	315	450	675
2000 Hrs	270	450	630	900	1350
Server BW (Mb/s)					
100 Streams	30	50	70	100	150
200 Streams	60	100	140	200	300
500 Streams	150	250	350	500	750
1000 Streams	300	500	700	1000	1500
2000 Streams	600	1000	1400	2000	3000
Streams/Link					
Ethernet (10-Mb/s, 40%)	13	8	5	4	2
Ethernet (100-Mb/s, 70%)	233	140	100	70	46
Ethernet (1-Gb/s, 70%)	2333	1400	1000	700	466
OC-3 (155 Mb/s, 95%)	490	294	210	147	98
OC-12 (650 Mb/s, 95%)	2058	1235	882	617	411

Table 1: Media storage, server bandwidth and number of streams served per interface link for different stream rates

acceptable for OC-192. In the case of OC-48, a fully centralized architecture works only at low penetration, a partially centralized architecture (centralized web cache and hierarchical streaming caches) works at medium penetration and a fully hierarchical architecture is required at high penetration. Of course, as stream rate increases, a fully hierarchical solution is necessary.

For hierarchical streaming media architecture, some key requirements for core and edge media servers are discussed below.

Edge Media Server: At low penetration (10% to 15% of HHP), an edge media server with interface bandwidth of OC-3 or fast ethernet is sufficient.one to two thousand hours of stream storage and ability to serve several hundred streams simultaneously is sufficient. . For higher penetration (over 40% HHP), however, the edge server must be able to scale to aggregate interface bandwidth of up to an OC-12 link (about 600 Mb/s). Stream storage capacity of 50 GB initially or about 370 hours and should scale to to 200 GB or about 1850 hours at 300kb/s.

Core Media Server: The core server serves multiple edge servers (during cache misses) and also is the source for all multicast transmissions. . Link bandwidth of 300 Mb/s to 1 Gb/s is required, with ability to scale to multiple gigabits/sec for higher stream rates. The core media server storage capacity should be 2x to 5x more than an edge server, i.e. in the range 100 GB to 1000 GB.

Quality of Service

A critical consideration when designing streaming media networks is quality of service (QoS). To deliver a consistent, high-quality user experience, the network must maintain a given stream’s specified latency and throughput requirements. Though end-to-end QoS for streaming content is generally difficult to guarantee over the public internet, it becomes more manageable over a DOCSIS-based private cable network. In the HFC network segment, from local hub to subscriber STB or cable modem, QoS is

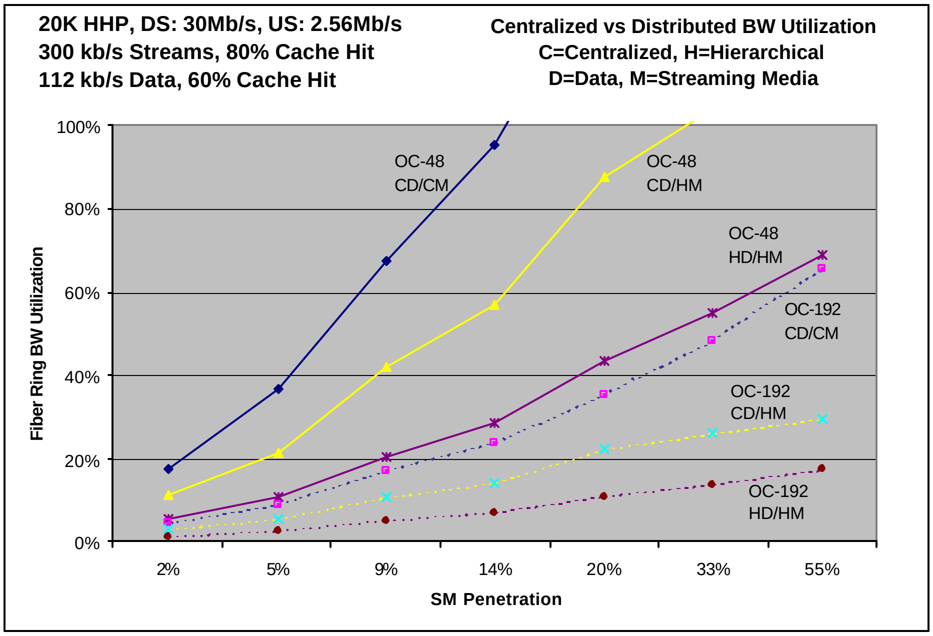


Figure 4: Fiber ring bandwidth utilization for centralized and hierarchical topologies for data and streaming media

achieved with dynamically provisioned service flows specified under DOCSIS 1.1.

On the optical backbone network segment, from the headend to the hub, QoS is achieved using either Differentiated services (Diff-serv) or Resource Reservation Protocol (RSVP) or thorough a combination of both [9,10]. Diff-serv classifies IP packets into a few aggregated classes using the type service (ToS) bits or diff-serv code points (DSCP). QoS-enabled routers and switches can then shape traffic using intelligent queuing (e.g. class-based weighted fair queuing or CBWFQ) to queue video streams as preferential class traffic ahead of other best-effort class traffic.

In addition to Diff-serv, RSVP signaling may be used to reserve bandwidth on a per-data-flow basis through the edge HFC network. RSVP signaling is used to dynamically initiate, configure and terminate DOCSIS 1.1 service flows from CMTS to cable modem. See [8] for the use of RSVP to set up voice telephony sessions under the PacketCable DQoS specification.

Set-top Boxes

A key challenge in delivering IP video to TVs is the availability of a suitable STB. As indicated earlier, the three streaming media formats used today are of proprietary nature. If decoding were to be done in software, significant CPU resources and run-time memory may be needed. Also, integration of the streaming video player with the STB middleware GUI is required.

Some next-generation STBs may contain programmable video processors (e.g. Philips TriMedia processor) that support multiple codecs in firmware. This will allow the support of MPEG-x and popular proprietary codecs in an economical manner.

Conditional access system (CAS) and digital rights management (DRM) are also necessary to protect copyrighted content and to provide differentiated video services. An open CAS/DRM mechanism for IP video is required for a cost-effective solution.

REFERENCES

[1] Broadband Net access nearly 8 million strong, Corey Grice, CNET News.com February 28, 2001, http://news.cnet.com/news/0-1004-200-4983492.html

[2] H. Benoit, *Digital Television*, John Wiley, 1997.

[3] Michael Adams, OpenCable Architecture, Cisco Press, 2000.

[4] RFC 1889, RTP: A Transport Protocol for Real-Time Applications, Internet Engineering Task Force (IETF), January 1996.

[5] RFC 2502, Compressing IP/UDP/RTP Headers for Low-Speed Serial Links, Internet Engineering Task Force (IETF), February 1999.

[6] RFC 2250, RTP Format for MPEG1/MPEG2 Video, Internet Engineering Task Force (IETF), January 1998.

[7] RFC 2326, Real-Time Streaming Protocol (RTSP), Internet Engineering Task Force (IETF), April 1998.

[8] PacketCable Dynamic Quality of Service Specification, http://www.packetcable.com

[9] RFC 2205, Resrouce ReSerVation Protocol (RSVP) – Version 1 Functional Specification, Internet Engineering Task Force (IETF), September 1997.

[10] RFC 2475, Diff-serv, Internet Engineering Task Force (IETF).

[11] Ivy Lui and Prashant Gandhi, Streaming Media Opportunity for Cable, Cisco white paper, December 1999, http://www.cisco.com/warp/public/cc/so/cuso/sp/stmda_wp.pdf

[12] John Chapman, Multimedia Traffic Engineering for HFC Networks-White Paper, Cisco Systems, 1999. http://www.cisco.com/warp/public/cc/so/cuso/sp/hfcn_wp.pdf