Optimal Network Architectures for On-Demand Services

Ran Oz, BigBand Networks, Inc. (rano@bigbandnet.com)
John B. Carlucci, BigBand Networks, Inc. (johnc@bigbandnet.com)
John Horrobin, Cisco Systems, Inc. (jhorrobin@cisco.com)

1 Introduction

1.1 Background

The Cable TV service portfolio has migrated from just broadcast TV to the introduction of premium pay channels, then pay-per-view, onto near video-on-demand, and now video-on-demand and Web browsing on the TV. Incremental introduction of new services reveals a distinctively evolutionary path of increasing personalization of offerings. Digital technology has been instrumental in this evolution. Successful roll out of new services, while maintaining existing services' popularity, depends on the network progressing to support the increasing variety of applications, media, session types, subscriber profiles and access devices involved.

Today's cable network infrastructure, with its high capacity and two-way capability, has the fundamental ability to support On-demand services; however, a number of obstacles must be addressed to optimize. This paper explores changes in the role of the Multimedia-Service router and provisioning network in enabling the migration to a fully On-demand network. The authors begin with a high-level description of the problems facing On-demand services in today's network, and then explore the system modifications to address these issues.

1.2 Problem Statement

While today's cable network infrastructure with its high capacity and two-way capability can support On-demand services, a number of obstacles block optimization for mass-market rollout. Hardwired cabling and static allocation of bandwidth prevent customization of services on a per fiber node basis. As node sizes have decreased, the network has not responded to allow control of services on a fiber node by fiber node basis. Additionally, static allocation of bandwidth forces the operator to perform traffic engineering based on peak loading, even when traffic levels across nodes, across services and across time are highly variable, largely unpredictable and uncorrelated. The result is underutilized bandwidth and foregone service deployments, often concurrently.

A new service cannot be deployed without expanding bandwidth capacity or reducing capacity dedicated to another service at a significant operational cost. Detailed traffic engineering analysis, including unreliable uptake projections must be performed to try to predict the viability of a new service. Another shortcoming of today's networks is the inability to support the variety of QOS requirements and transport strategies (MPEG transport, IP, multicast, unicast, broadcast) needed to deliver a range of services. Additional requirements are introduced by the need to implement flow control in support of the capabilities of various access devices ranging from scaled down consumer electronics to high capability PCs.

Current VOD systems have low title/stream ratios (title capacity is a fraction of stream capacity). This is well suited for movie-on-demand services with "lumpy" demand (for example, Top 10 movies represent 80% of demand). As the selection of titles grows with the introduction of time-shifted broadcast TV and other new On-demand services, demand becomes less "lumpy". This leads to high storage requirements but also more propagation capacity throughout the network. This is particularly problematic in decentralized systems with servers located in the hubs.

Finally, operators have made significant capital investment in the HFC infrastructure and digital modulation. Millions of digital STBs have already been deployed, and millions more are being rushed into the field to meet customer demand for digital broadcast services. To leverage the operators' investment, service delivery to legacy customer premises equipment must be preserved. In addition to continuing to support legacy digital deployments, the network must support roll out of cost-effective consumer equipment that provides new services. It must also feed advanced multimedia content that takes advantage of the new functionality possible through emerging high-end consumer equipment. Overall, steep network challenges are presented by the increasing diversity that must be supported across a great range of considerations.

1.3 Proposed Solution

Changes in network architecture are needed to better support the migration to an On-demand service environment. For popular on-demand content, with "lumpy" demand, distributed hub servers will be used. Servers will cache frequently requested material close to the subscriber's home. For some On-demand services, such as niche on-demand content, storing titles in centralized servers and streaming across the network on-demand will become more viable than replicating titles in hub servers.

Edge routing and multimedia processing capability in the network maximize service flexibility, optimize bandwidth utilization, preserve legacy investment, and provide an evolutionary path for new service rollouts. This routing and processing capability can be located in either the headend or the hub. Whether a centralized server or a hub server delivers the program, the multimedia processing needs to be supported.

Regardless of the underlying network transport, multimedia processing offers unique options. Content formatting standards allow transcoding into new formats to support different client devices. The elastic nature of the content permits dynamic bandwidth variation to address congestion. Statistical multiplexing and rate adaptation are needed to yield optimal bandwidth usage for narrowcast services.

The On-demand network seamlessly supports blending of other services such as in-band data delivery, broadcast and pay-per-view services. A centralized bandwidth manager controls access of these services to network resources. The bandwidth manager is responsible for resource allocation as well as maintaining and enforcing the policies for resource allocation. Policies must accommodate not only the required bandwidth range and service type requested, but must also account for media type, subscriber, and access device. By basing the resource allocation on actual rather then expected or peak resource requirements, this approach eases the operational and planning burden of new service rollouts.

In order to support multiple services over a single network, and to accommodate the introduction of new services, the On-demand network will provide a common platform capable of interfacing to legacy systems and supporting new services with standards-based protocols and QoS mechanisms.

2 On-Demand System Reference Architecture

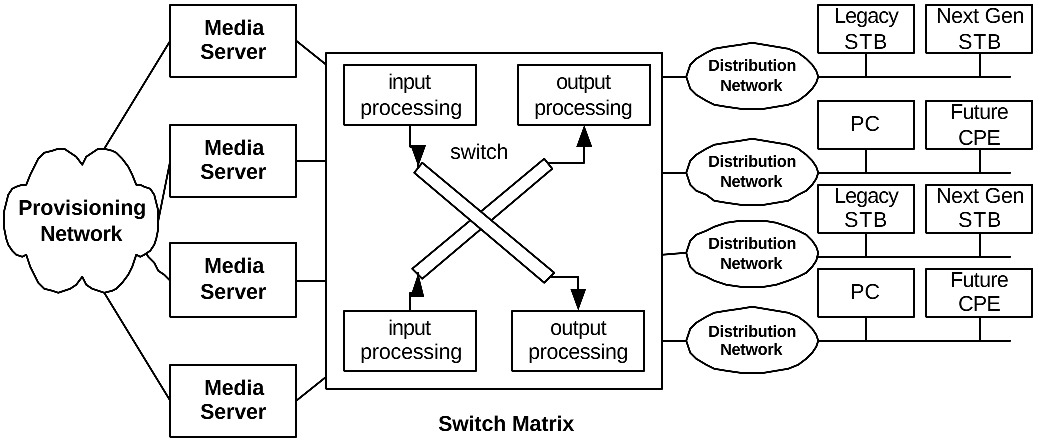

Figure 1: On-demand Reference Architecture

An On-demand system reference architecture is depicted above in Figure 1 [1]. The system is composed of the following components:

?? Provisioning Network – delivers content to the media servers.
?? Distribution Network – delivers content to consumer devices.
?? Media Servers – provide On-demand content.
?? Switching Matrix – provides a path from each media server to each distribution network.
?? Set-top boxes, PCs, other CPEs – consumer devices provide a termination point for On-demand media and adapt it for the TV.

The media servers and switching matrix may be centralized in the headend or distributed to various hub sites. In the next section, we explore the Switch Matrix and its function in an optimal on-demand network.

3 Multimedia-Service Router and Headend Architecture

To allow an operator to successfully compete, the network architecture for On-demand services must be scalable, flexible, efficient, and evolutionary. Scalability and flexibility enable new, high-volume service deployments. Efficiency and evolution are required to support the current services while adding new services, such as VOD, in-band IP data delivery, and interactive television. The headend needs better switching and media processing capabilities to realize this network.

Switching Matrix

The switching matrix, found in the headend or hub, is a key element of the On-demand model. When located in the headend, it impacts the largest number of nodes while easing the operational burden. When located in the hub, it allows for grooming and manipulation of content on a hub basis as well as reduces network loading by caching popular content at the hub site. The basic requirements of an on-demand system are delivery of broadcast-quality content, two-way communications, stream control, and maintainable streaming media quality of service. In addition to enabling On-Demand service by supporting these requirements, the switching matrix must also assure and optimize the delivery of other services such as broadcast video and interactive television.

---

[1] *OpenCable™ Architecture* by Michael Adams

A number of classical approaches have been used to realize the On-demand switching matrix, each with its own weaknesses[1]. The content replication approach becomes unwieldy as the number of titles made available to the viewer increases. The massively parallel server approach limits the ability to integrate existing broadcast and new data services. The ATM switch approach introduces additional overhead to each video and data packet. Additionally, ATM performance improvements are not keeping pace with other switching technologies. The pure IP routing approach would require modifications to legacy digital video systems to provide the proper interfaces.

A promising switching alternative is the Multimedia-Service Router that provides a hybrid of MPEG2 and IP switching capabilities. This combination of switching capabilities brings the benefits of packet switching technology for data and video to the On-demand architecture.

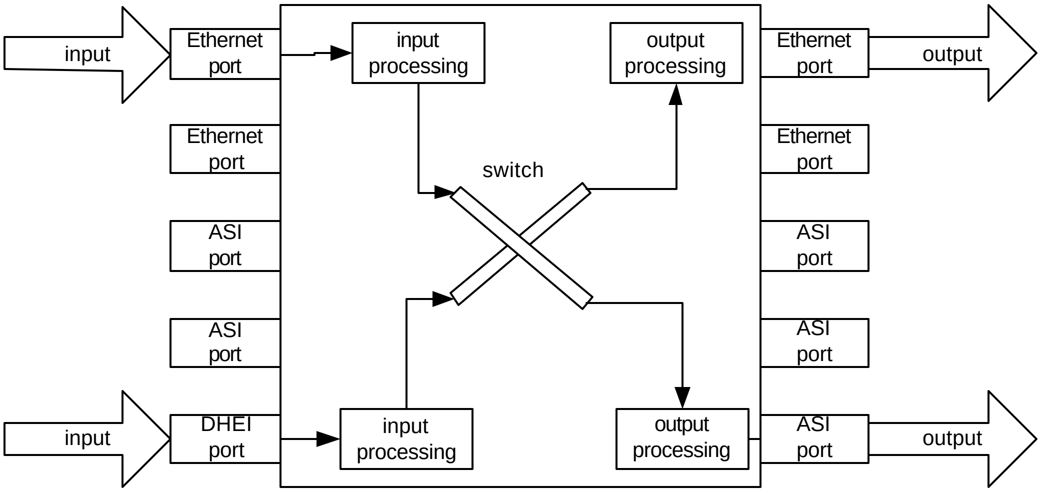

Figure 2 Multimedia-Service router

Multimedia-Service Router

The Multimedia-Service router must provide a number of key functions. The router allows any media source (e.g. VOD server) to be connected to any client device (e.g. digital STB). Selective assignment of a service to a QAM frequency balances the traffic between the sources and the clients across QAM channels. In addition, this provides an efficient reliability model in which a single redundant source backs up multiple media sources. Finally, the ability to blend services within the same routed output, as in synchronized delivery of content from multiple servers, allows new services to be tried without sacrificing support for existing services.

While functioning as a switch, it also performs a bridging operation between two heterogeneous networks connecting the core backbone network to the highly distributed HFC plant. The network architect can change the backbone network, to incorporate technologies such as Gigabit Ethernet, without impacting HFC plant. The optical backbone doubles its efficiency every 10 months, without requiring HFC plant modifications. This junction also provides an entry point for Internet based streaming content into the HFC network. As content passes through the Multimedia-Service router, it can be transcoded before delivery to legacy STBs.

The Multimedia-Service routers act as a content aggregation fabric combining, synchronizing, filtering and re-mapping content from a number of sources such as VOD servers and interactive TV servers, satellite and digital terrestrial receivers, and locally encoded channels all around a network. In the On-demand architecture, the router also performs a range of multimedia processing. Statistical multiplexing and rate adaptation allow for variable bit rate video services from a mix of sources to be efficiently carried. Video

---

[1] *OpenCable™ Architecture* by Michael Adams

format conversion allows delivery of streaming video services based on MPEG-4 and other video streaming formats to legacy MPEG-2 digital STBs.

The Multimedia-Service router supports all the necessary features of an On-demand architecture.

Efficient:
The Multimedia-Service router enhances network efficiency in a number of ways. It avoids transport overhead by switching the most popular forms of network traffic, including MPEG-2 and IP, in their native forms. It allows maximum bandwidth utilization by performing statistical multiplexing and rate adaptation for On-demand and broadcast services, increasing the number of VOD sessions per channel. It allows carriage of a mix of services with complementary loading profiles within the same QAM channel and the mix of services gets varied over time and across nodes. Beyond this, the mixing of services also supports the establishment of hybrid services such as t-commerce, as in electronic commerce transactions being complemented by the on-demand promotional video.

A centralized bandwidth manager coordinates the resource allocation and policy enforcement needed by the various services. The bandwidth manager assures new offerings can come on line without inadvertently depriving existing ones. This eliminates the need for resource negotiation between services.

Scalable:
A single deployed switch can support a user base distributed across the entire HFC system, keeping the cost and complexity of entry-level systems low. Additional routers are deployed and distributed to support the demand brought on by adoption of new services. Increasing use of Multimedia-Service routing has an offsetting benefit in increasing utilization efficiency of distribution network transmission equipment such as QAM modulators and up-converters, and thus decelerating the need to scale up use of those components. Rapid advances in switching performance, in the MPEG and IP domain, provide increased switch density to scale with service demand. Additional On-demand servers can be incorporated to support alternative titles without needing to expand the router infrastructure.

Flexible:
The Multimedia-Service router's multimedia formatting/transcoding capability allows video to be provided to and from the routers in a variety of digital formats. This supports the repurposing of Internet video content to legacy STBs through MPEG-2 conversions as well as the delivery of broadcast video programs to the PC. This transcoding will allow the network to respond to innovations in customer premises equipment.

Further flexibility is found in the location of the router. While maintaining centralized bandwidth management control, it can be placed in both the headend and hub portions of the network. Router distribution is driven by the location of video and other content sources, with more frequently accessed content being located closer to the subscriber. Multimedia-Service routers located at the hub site can operate on both headend- and hub-based content as it moves through the hub site to the subscriber.

Evolutionary:
As more efficient transport and compression technologies are deployed elsewhere in the network, the MPEG-2 switching capability and format conversion will continue to support existing digital video deployments. Multiple generations of fiber optic backbone equipment can be deployed while the Multimedia-Service router enables bread and butter video services to continue to be delivered. The operator is not forced to undertake major changes in distribution infrastructure or client devices to deploy the On-demand services. The HFC infrastructure elements carrying services to the home are insulated from service and backbone network changes.

The Multimedia-Service router is an integral part of the On-demand service network.

4 New On-Demand Services

The flexibility of the Multimedia-Service router to process digital video in a variety of formats, including MPEG-2 and MPEG-4, enables the MSO to deploy a variety of On-demand services to legacy digital STBs as well as next-generation STBs. In addition to traditional On-demand services such as VOD and time-shifted broadcast TV, MSOs can offer new classes of programming with streaming media technologies. Programming from streaming media web sites, such as scheduled and live webcasts, can supplement the traditional On-demand service offerings. With the proliferation of streaming media web sites (currently in excess of 10,000 sites), MSOs can greatly increase the variety of programming choices available to subscribers.

Streaming media services require a fraction of the bandwidth of MPEG-2 streams, which enables MSOs to support a greater number of On-demand subscribers. This efficiency is critical when considering the migration from broadcast services to On-demand services. Streaming media services can fulfill demand for a personalized TV experience while conserving network capacity.

MSOs can differentiate their streaming media services by providing targeted content tailored to TV, service packages to match a variety of subscriber buying patterns, and improved quality of service compared to accessing streaming media content via Internet access services.

These new On-demand services will require advanced network architectures, as described in the following section.

5 On-Demand Network Architectures

This section discusses emerging technologies that will enable MSOs to deploy provisioning and distribution networks that maximize the benefits of the Multimedia-Service router and enable the introduction of new On-demand services.

As depicted in Figure 1, the provisioning network delivers content from the content provider to the media servers located in the headend or hubs. The provisioning network requirements of traditional On-demand services are dramatically different than the requirements of streaming media services. Therefore, we will discuss the requirements of each type of On-demand service individually.

Traditional On-Demand Services

For VOD service, content is typically pre-loaded on media servers located in the headend or the hubs and subsequently accessed on-demand by subscribers. In this case, the content can be distributed from the content provider using standard file transfer mechanisms on an IP network. Since the VOD content is updated relatively infrequently (days or weeks), the content can be scheduled for delivery over the provisioning network during off-peak hours, so the network capacity and performance requirements can be attained with best effort service, and QoS mechanisms are not required.

For time-shifted broadcast TV service, the broadcast content is typically captured directly from the broadcast feeds by the media servers. Alternatively, the content may be captured by a central server and uploaded to the media servers in the same manner as VOD content.

Streaming Media Services

For On-demand streaming media services such as live webcasts, and when distribution of the content is restricted by the content provider, the provisioning network will have to support real-time delivery of the content. The media servers in the headend or hubs may provide caching functions, or the caching function may be provided deeper within the provisioning network so that no caching is done in the headend or hubs. In this case, the provisioning network takes the form of a Content Delivery Network (CDN), which is specifically designed for managing and sourcing streaming media services on IP networks. A CDN is comprised of the following key components:

?? Content Distribution & Management: load and manage content at edge servers.
?? Content Routing: redirects subscriber to best site.
?? Content Switching: load balances traffic across servers.
?? Content Edge Delivery: delivers streaming media from edge servers.
?? Intelligent Network Services: enables traditional IP services at a content level, including QoS.

A CDN supports the delivery streaming media content as multicast or unicast services. Since the CDN streams content in real-time over an IP network, QoS mechanisms must be in place to ensure the content is delivered in accordance with latency and throughput requirements. Two QoS mechanisms used to ensure that streaming media content is given appropriate priority while traversing the IP network are Differentiated Services (DiffServ) and Resource Reservation Protocol (RSVP). In addition, the Real-time Transport Protocol (RTP) is used to map the streaming media content into IP datagrams, and the Real-time Transport Control Protocol (RTCP) is used to provide feedback from the receivers to the sender regarding the performance of the streaming media service. Also, application-level signaling between the client and server is facilitated using the Real-time Streaming Protocol (RTSP).

MSOs can deploy their own CDN, and establish peering relationships with other CDNs to extend the reach of the streaming media services. Alternatively, MSOs can establish connectivity to CDNs of other service providers. To facilitate interoperability across CDNs and to reduce network and STB resource requirements, MPEG-4 is the recommended codec and file format for streaming media applications. For legacy STBs that can not support MPEG-4 decoding functions, the Multimedia-Service router can transcode MPEG-4 streams into MPEG-2 streams, allowing subscribers with legacy STBs to receive streaming media services.

The technologies described above for delivering streaming media services can also be applied to delivering traditional On-demand services in real-time over the distribution network. Using these technologies, an IP network can be deployed as a common platform for supporting On-demand services as well as cable modem and cable telephony services.

6 Conclusion

The Multimedia-Service router supports many applications in the evolving On-demand architecture. The rate adaptation capability allows rate changing to become an import new ingredient in efficient delivery of On-demand services. By performing flow control and multimedia format conversion, the router enables a range of customer premises equipment. The ability to dynamically blend services within a common QAM channel optimizes delivery of video and data content needed to enable the interactivity and program delivery of on-demand services. The Multimedia-Service router will be an important element in enabling the migration to an architecture rich in On-demand services.

The On-demand network supports traditional On-demand services such as VOD and time-shifted broadcast TV as well as new On-demand services, including streaming media. A Content Delivery Network (CDN) is incorporated into the On-demand network for managing and sourcing streaming media services on IP networks. The key technologies for delivering streaming media services can also be applied to traditional On-demand services so that a single integrated network can be deployed.