CDNs for Personal Broadcasting and Individualized Reception

Sujata Banerjee, Jack Brassil, Amy Dalal, Sung-Ju Lee, Ed Perry, Puneet Sharma, and Andrew Thomas
Internet Systems & Storage Lab
Hewlett-Packard Laboratories
Palo Alto, CA 94304

Abstract

By putting the tools for media creation in the hands of the consumer, we anticipate exponential growth in the distribution and hosting of media on the Internet. We expect this growth to parallel that of HTTP traffic. We need to confront the issues related to this growth in the number of media sources and services. Significantly higher quality streaming and context-sensitive personalization of multimedia services will be key enablers of this whole technology area. Our vision differs from the delivery of video-on-demand services because we believe that there will be a vast number of sources of copyleft content. [1] In this paper we focus on the distributed discovery of content and the concurrent, context-aware manipulation of multiple media streams.

1 Introduction

We envision a future where Internet users will interact with applications, other humans and perhaps even objects in the environment, primarily via multimedia objects and streams. With emerging multimedia authoring tools and hardware, there is the potential for every Internet user to be a media content author and source, much like the prior WWW trend. We further envision a world where every user of a service can receive live, stored or interactive media content that is personalized, custom-made, context-aware and location-aware for that specific user or interest group. In this environment, service providers will be able to compete for customers on primarily their value-added service offerings. Examples of value-added services include advertisement insertion and commentary overlay in live or stored media streams. These services will require the manipulation of multiple streams sourced from different origins.

This vision differs from the conventional notion of mass dissemination of entertainment-quality video-on-demand, the goal of which is to deliver highly polished, copyrighted, rights-managed media at a price.

Content distribution networks (CDNs) improve performance and availability of web and some media content by pushing the content towards the network edges and providing replication and replica location services. Intelligent replica placement improves response times by serving content from replicas "closer" (in the topological sense) to consumers [3], thereby avoiding the congested backbone networks and network access (peering) points. Replica location services direct requests for objects to "nearby" replicas by layering redirections through DNS based on extensive measurement and monitoring of network performance. CDNs will continue to play a role in the future as a means to distribute the increased amount of content we envision.

First we highlight the technology drivers that we believe will make our vision possible. We then postulate some scenarios that differ from today's content authoring and delivery. We discuss the challenges that our vision presents us with in Section 2. Section 3 lists some novel research problems associated with this vision. We outline the CDN extensions necessary to support our vision in Section 4, followed by concluding remarks in Section 5.

1.1 Technology Drivers

Technology advances are paving the way for rapid growth in the creation and distribution of real-time multimedia streams. The consumer electronics industry is rapidly producing digital appliances to supplant the "old" analog equivalents. MP3 players are replacing portable cassette tape players, digital cameras are replacing film cameras, and digital video is challenging analog tape formats in the camcorder space. These digital devices are readily connected to the PC, which in turn acts as a bridge between the various devices and the Internet. Improvements in PC technology, particularly processor speed and disk space, have made it possible to store and edit digital video at a very attractive price point. The capabilities are already built into many high end personal computers and will soon be avail-

[1] According to the Webster online dictionary, copyleft is the copyright notice carried by GNU and other Free Software Foundation software, granting reuse and reproduction rights to all comers. See [7] for details.

able on all new PCs. The net result is that anyone with a PC and a digital camcorder will have the technology to be a cinematographer, editor and director of his or her own movie.

Advances in optical communications technology, including dense wavelength division multiplexing and emerging metropolitan area network systems, continue to increase Internet backbone and access transmission rates. The deployment of cable modem and DSL service has greatly increased the access bandwidth available to the home, with some five million homes wired in the U.S. [4].

Improvements in wireless technologies (both LAN and mobile telephony), video coding and miniaturization of audio and video capture technologies make possible a world where media can be streamed from anywhere into the Internet. This should open up a new world for the spontaneous creation and dissemination of media.

1.2 Scenarios

Media-centric services will dramatically and fundamentally change our future way of life. Below we take two seemingly familiar examples and expand them to demonstrate how we think future media services will be constructed.

• **Home Director**: Increases in available network bandwidth mean that future commercial Internet television stations will not face the channel (bandwidth) limitations faced by conventional broadcasters. For live broadcasts (e.g., news, sports), conventional broadcasters have historically created and transmitted a single “moderated,” or produced, program. Regardless of the number of cameras covering an event, at each instant only one is selected for broadcast, with all other camera feeds terminating in the local production room. With Internet backbone bandwidth plentiful, these alternate points-of-view could be transmitted in parallel with the primary, moderated stream, and each of these streams might find its own audience. Further, individual spectators could stream views from their mini-cameras using cellphones. The interaction between the remote audience and the event site could even direct camera activity. Alternatively, different points of view might be of interest to certain audiences (e.g., attention on specific players or positions for training purposes).

• **Sportscast Commentary**: American football fans in the U.S. have grown accustomed to seeing the electronic yellow “first down” line superimposed on the field in the television broadcast of games, or the commentators drawing “play analysis” that is superimposed on the image of the field. The speed-skating lanes in the Winter Olympics are superimposed with the flag of the country that the athlete represents. These are examples of value-added features to the original content that have tremendous support from the audience. However, production of these seemingly simple features is a big task. For example, a crew of four people (including one on the field) and five computers is dedicated to the “first down” line feature [17]. In the future, perhaps two renowned commentators will be brought in on the fly to discuss the game as it is happening and “draw” play analysis on the screen, when neither is actually at the game or the production site.

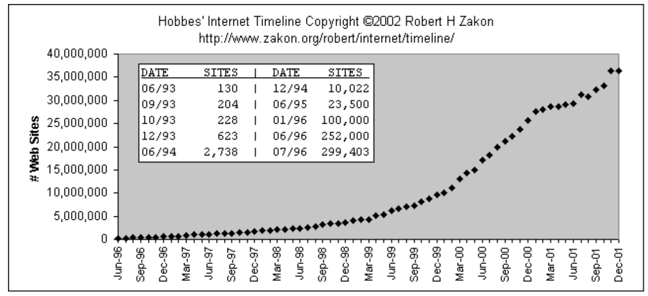

Figure 1: The growth in the number of World Wide Web sites, 1993-2001 [20].

In the next section, we provide some background on the research challenges alluded to in this section.

2 Elements of the Vision

We believe that enabling individuals to launch high quality streaming media from anywhere and context-sensitive personalization of multimedia services are key enablers of this whole technology area. To summarize, our vision rests on the following two pillars of thought:

• **Huge Number of Rich Media Sources**: The growth of multimedia on the Internet in some ways parallels the rise of the World Wide Web in the 1990s. Just as the Web increased the number of content “publishers,” as shown by the growth in the number of Web sites in Figure 1, we expect the number of media “producers” to rise as well.

It seems highly likely that there will be two camps: the entrenched copyright owners (such as the large studios) who will provide copyrighted movies and audio to large numbers of consumers; and the home users, who will take videos of their family, vacations, etc. (probably not for financial gain and without copyright concerns), and produce a substantial amount of content for consumption by a different demographic than the copyright owners. One family in a hundred producing a one hour home video per year (one million hours for the U.S. alone) will dwarf the output of the major studios. Finding and cataloging this material and making it available to its audience is a significant challenge.

• **Personalized, Context-Aware Media Streams**: Media today flows directly from the content creator to the audience without any intermediate manipulation. In the future we see this situation changing by the addition of value added content and services. These additions may be performed by a third party, and we are interested in supporting this model.

We can see examples of repurposed content today. An interesting example is "pop-up video" [18] that is shown in the U.S. on the VH1 television channel. The original content is a pop music video that provides the backdrop for speech bubbles that pop-up frequently during the video. The material in each bubble is some anecdote about the video, the artist, the location, the clothes and so on. In addition to modifying the video, the sound is augmented with a bubble popping noise or other appropriate jingle.

Mystery Science Theater 3000 [16] is another example of repurposed content, where movies are broadcast along with an additional overlay of cartoon characters in a cinema. The added value here is the lampooning of the feature by the cartoon characters who contribute a witty dialog for the duration of an otherwise dull movie.

Taking these two examples, we can postulate a similar phenomenon for media on the Internet. Moreover, the Internet is an open delivery mechanism where content can, at least in theory, be modified and republished by anyone. Today many sites exist that are devoted to televisions shows, offering plot summaries and character features. In the future these sites could repurpose content, with fans of the show adding their own audio or video commentary, or perhaps producing their own "directors' cuts."

Streams could also be customized based on personal profiles and preferences. Examples include insertion of location aware advertisements or specific camera feeds as described in the home director scenario in Section 1.2.

3 Novel Research Problems

We identify two important research areas and list the novel problems that need to be addressed for each. The two areas are (1) announcement and discovery of *personally interesting* media content, and (2) program manipulation.

To understand why these areas are so important to us, it is necessary to understand what we have chosen *not* to do. We are not building a video-on-demand system for movie and television content. A commercial system like this would most likely be *over-provisioned* with *dedicated resources* to ensure a good customer experience at possibly a high cost to the user. Further, in such a system, content manipulation would be forbidden, while announcement and discovery would be reduced to an electronic TV Guide. By starting with a large number of sources we have a very different problem. The sources will necessarily be widely distributed, and hence require a decentralized location service. Access to the media will be over a network with varying degrees of service requiring media assessment, adaptive protocols and media manipulation.

We acknowledge that there are several other important research areas that will have an impact on the viability of the services we envision. These include media encoding, quality of service, quality assessment, privacy, security, digital rights management and pricing mechanisms, and are out of scope of the paper. It is our belief that the applications we envision will first emerge in corporate intranets or "walled gardens" where these issues are less problematic.

3.1 Announcement and Discovery

In each of the scenarios described in Section 1.2 a *directory system* is the presumed basis for both announcing and discovering multimedia content. The challenges associated with this system are formidable. The system must scale to enable a vast audience to identify content of interest. Unlike traditional information discovery systems (e.g., Web search engines), the system must be capable of handling short duration, real-time communications. These sessions can be pre-scheduled, or occur spontaneously and be announced at the session start. The directory system should also be lightweight, consuming only small amounts of computing and bandwidth resources. Directory service users should have the benefit of powerful searching tools, yet search times must be kept low. Announcements of events should be sufficiently expressive to facilitate discovery, without putting undue burden on the announcer, who may in fact be little more than a camera connected to a communications link.

Further, the directory system might take on additional features when used in an enterprise setting. Here the directory could benefit from integration with other meeting directory services, such as that for Microsoft's NetMeeting. This integration could further extend to announcements for both Internet and conventional telephone system conferences. For example, an Internet voice conference call initiated by the Session Initiation Protocol (SIP) [10] might conceivably issue an INVITE message to our directory system for the purposes of a public announcement. Similarly, an announcement of a conventional telephone conference call, now often distributed semi-privately by mechanisms such as electronic mail, might also be listed. Indeed, if connected to a telephone switching system, the directory would in principle be able to announce any active connection in progress.

3.1.1 Related Work

Though research in conventional information discovery systems has been extensive, relatively little attention has been paid to systems capable of discovering live multimedia sessions. As an example of a conventional system, the Domain Name System (DNS) is a robust, heavily-used decentralized database primarily for name/address translation. DNS has a variety of desirable properties that we seek, including robustness, global scale, and caching. DNS queries are quite limited in nature, however, as is the data stored in the system (e.g., CNAMES, PTR). Moreover, even with advances such as dynamic DNS, management of system data is restricted, and propagating new records throughout the system can take hours, far too slow for our application.

The announcement of live multicast sessions on the MBone is achieved through the use of the Session Announcement Protocol (SAP) [9], and discovery is realized by client-based software 'session directory' tools such as *sd*, *sdr* and *rsdr*. SAP relies on the presence of IP multicast, and essentially implements a soft-state protocol; clients monitor an announcement channel, and perform all discovery operations (e.g., searching) locally. Announced sessions are described by the somewhat unexpressive Session Description Protocol (SDP) [8]. SAP/SDP was later extended to support a limited session directory hierarchy.

Though shown to be useful as a discovery system for small-scale, prescheduled MBone sessions, a SAP based system suffers certain limitations. Efforts to conserve bandwidth by making infrequent announcements caused relatively long startup delays for users joining the multicast group at program start time. The system was not intended for use by unscheduled sources, nor would it be effective for very short-lived sessions. SDP is also quite limited as a descriptive language for programming metadata.

A collection of research attempts have addressed these limitations. SDPng [12], a successor to SDP, is an XML-based session description language under consideration within the IETF. Also, tools to permit Web browsers to interact with directory information have tended to displace dedicated client-side session directory tools.

One proposal has specifically addressed a number of the scaling limitations within the SAP-based discovery system design. The Information Discovery Graph (IDG) [19] seeks to impose a hierarchical network of directory managers—each responsible for a semantically distinct session topic—to balance system load, facilitate searches and reduce system overhead.

3.1.2 Research Challenges

No existing directory service appears capable of coping with the scale and the dynamics of the media distribution system we are now considering. Among the difficult questions we seek to answer are:

• **Announcement Creation**: Each announcement must effectively describe its associated multimedia session with sufficient expressiveness such that a simple search by a potential recipient will discover the session. Yet suppose that there are two closely located sources, each unaware of the other, transmitting live video from the same event. How can we ensure that their independent announcements facilitate discovery? The challenge becomes greater if we consider sources with extremely limited data input capabilities (e.g., digital cameras). Can metadata not requiring human input, such as unique device identifiers or GPS coordinates, be used to facilitate content searches?

• **Access Control and Privacy**: The directory system we envision will be most powerful if it supports public, private and semi-private announcements. How can we ensure that the only parties to receive an announcement are those we desire? In addition, a receiver might desire personalized directory presentation according to pre-defined preferences. How will filtering of listings take place, and where should this filtering occur?

• **Directory System Architecture**: The architecture of the directory system must be scalable yet still enable rapid directory dissemination. Could CDNs be used to support directory services? Should such a system combine both *push* and *pull* distribution mechanisms? If so, how do we prevent a program directory listing from being stale? Will multicasting be part of the solution?

3.2 Program Manipulation

Future media streams will have audio, video, textual and markup components, and interesting multimedia programs will require several semantically related streams to be composed and customized. We define *program manipulation* as the control, modification and coordination of these streams.

3.2.1 Related Work

Defining the layout, interaction and temporal and semantic relationships among various media streams is very important for composable and customizable multimedia presentations. SMIL [1] (Synchronized Multimedia Integration Language) is a language developed by the SYMM (Synchronized Multimedia) Working Group of W3C to describe time-based structure for multimedia *documents*. CMIF (CWI Media Interchange Format) [5] also allows composition of multiple video and audio streams.

Various inter- and intra-stream synchronization algorithms have been proposed to smooth out the network delay jitter. A comparison of several such techniques has been presented in [11]. Several video compression standards such as MPEG-2 and MPEG-4 also allow for synchronization markers.

3.2.2 Research Challenges

We discuss key research problems in program manipulation.

• **Stream Manipulation**: Along the path between the original media and the final audience, the stream may pass through several processing steps to add value. These processing steps occur on *mediators* that perform the following functions:

- *Switching*: This component would take several incoming streams and generate a single outgoing stream. An example is switching between various camera feeds in the home director scenario described in Section 1.2.
- *Semantic Manipulation*: A simple example of semantic manipulation is inserting a logo into an existing stream. More complex examples include editing "pop-up" text or selecting an audio stream appropriate for a user's preferences.
- *Transcoding*: Transcoding transforms material from one format to another (e.g., MPEG-2 to MPEG-1), increases compression to allow transmission over a lower bandwidth link, or adds appropriate link level forward error recovery.
- *Synchronization*: Synchronization coordinates the arrival of multiple streams at clients to reduce the client buffering load (especially for mobile devices).

We believe that content distribution networks (such as Speedera and Akamai) are the pre-cursors to mediators. We argue that CDNs work for on-demand video, but that much of the research on optimal placement of mediators will be revisited because of the advanced processing and requirements for live media and multi-stream coordination.

Copyright also has an interesting architectural impact on how and where streams may be manipulated. In repurposing content a third party is producing a derived work, which may be a breach of copyright (Section 106(2) of the US Copyright Act does not allow unauthorized derived works). For example, a third party may take a video and create a commentary but be prohibited from mixing the two and then republishing it. On the other hand, it may be perfectly reasonable for the third party to pass a reference to the original work and for the commentary to the audience to be rendered in an appropriate browser.

• **Stream Markup**: Program manipulation requires enhancing the streams with metadata not only about their identities but also about structure and semantics. Cueing protocols [2] provide one mechanism for transporting metadata in-band with the stream.

Repackaging audio and video streams with additional markup allows us to personalize content. The markup allows the expression of the type of content that could be matched against the user's preferences to create a personalized experience of the media. We see this today in parental controls, but in the future we imagine a much wider expression of what is in the streams.

Stream markup will also aid synchronization among multiple streams with different origins. Though some video compression standards have proposed synchronization markers, we require a format-independent mechanism for stream synchronization.

• **Multi-Stream Program Composition**: The W3C standard SMIL [1] can be used for defining the layout and temporal structure of multimedia documents. Several SMIL-based presentation authoring tools are now available. Most of these presentations are limited to static (predefined) streams. For example, it is not possible to author a sportscast using SMIL where a viewer is interested in watching only the camera feed that has a particular player. Novel mechanisms are required to dynamically compose media presentations based on stream metadata and user preferences.

• **Navigation**: The ability to effectively navigate streams within multi-stream programs is key to good viewer experience. The single stream control protocols such as RTSP [15] are restricted to simple commands like *start*, *fast-forward*, *rewind*, *pause*, etc. Such primitives are not adequate for proposed program manipulation. Navigation among the different components of a multi-stream program will require new control primitives and protocols for exchanging these commands. For instance, new primitives are needed to express a viewer's intention to switch between different camera feeds in the home director scenario described in Section 1.2.

• **Personalization**: Users will need some way of expressing their preferences so that streams can be personalized for them. Stream metadata/cues will provide the mechanism for personalization, but how users specify their preferences is an open problem. We note that many useful services could be constructed by making the profile available publicly, and many more services would be enabled by including location information as part of a user's profile. However, history has shown us that there are privacy and security issues related to dissemination of user information.

4 CDNs for Enhanced Media Services

We believe CDNs are a natural choice for enabling our vision and have many characteristics to build upon: replication, replica management and replica location; the measurement and monitoring of network characteristics to assist in replica location and placement; and the assorted management infrastructure for dealing with a geographically distributed overlay network. In this section we discuss ways in which CDNs will evolve to support our vision.

• **Extensibility**: For content manipulation and personalization we need extensibility of the CDN service model. In addition to replicating static content we would like to tap into the *processing* as well as storage at the CDN nodes. The Open Pluggable Edge Services working group of the

IETF is taking the first step towards making CDNs truly extensible and flexible to support new services [14]. Ma *et al.* [13] have proposed such CDN extensibility to support value-added services such as watermarking and content adaptation.

• **Bidirectionality**: Today's CDNs unidirectionally push the content of a small number of large sites towards the network edges, closer to a very large audience. Personal broadcasting, on the other hand, will result in the injection of (perhaps a small amount of) content from considerably more sources in a bidirectional manner.

• **Edge Processing**: We envision extensive personalization and context-aware manipulation of streaming media and we see many benefits of doing this within the network. However, the processing requirements for such tasks may force us to reconsider many aspects of CDNs. For example, personalization could result in different caching behaviors at CDN nodes, while context aware manipulation could require load balancing of CPU resources.

• **Object Location**: Since CDNs provide replica location, we are faced with the problems of announcement and discovery of new media objects (that may be replicated). Although DNS is used in replica discovery, it does not appear to be a good mechanism for object location (Section 3.1.1).

• **Live Media**: To support our model of personal broadcasting, we need CDNs to better deliver live media. We also need better support for interactive applications.

5 Conclusions

Recent technology advances have made content creation accessible and affordable. There will be a vast number of content creators scattered across the Internet creating copyleft media. That the content is widely distributed and of temporal value has forced us to look at the discovery problem. That the content is copyleft allows us to be creative in the ways in which content is manipulated to create new works, perhaps by merging and mixing multiple streams. The scalability and closer-to-edge location of CDN nodes make content distribution networks an excellent choice for enabling personal broadcasting and individualized reception of rich media content.

References

[1] J. Ayars, *et al.* (editors), "Synchronized Multimedia Integration Language (SMIL 2.0)," *W3 Recommendation*, April 2001, `http://www.w3.org/TR/smil20`.

[2] J. Brassil and H. Schulzrinne, "Enhancing Internet Streaming Media with Cueing Protocols," *Proceedings of the IEEE INFOCOM'01*, Anchorage, AK, April 2001, pp. 95-103.

[3] P. Christy and J. Katsaros, "The 2000 Content Delivery and Distribution Report," *Internet Research Group*, February 2000.

[4] W. S. Ciciora, "Cable Modem Traffic Jam," *IEEE Spectrum*, June 2001 pp. 48-53.

[5] CWI Media Interchange Format, `http://www.cwi.nl`.

[6] Enhanced Content Specification, "Advanced Television Enhancement Forum," `http://www.atvef.com/library/ATVEF%20spec1_1a_Final.pdf`.

[7] GNU Project - Free Software Foundation (FSF), "What is Copyleft?," `http://www.gnu.org/copyleft/`.

[8] M. Handley and V. Jacobson, "SDP: Session Description Protocol," *IETF RFC 2327*, April 1998.

[9] M. Handley, C. Perkins, and E. Whelan, "Session Announcement Protocol," *IETF RFC 2974*, October 2000.

[10] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, "SIP: Session Initiation Protocol," *IETF RFC 2543*, March 1999.

[11] Y. Ishibashi and S. Tasaka, "A Comparative Survey of Synchronization Algorithms for Continuous Media in Network Environments," *Proceedings of the IEEE LCN'00*, Tampa, FL, November 2000, pp. 337-348.

[12] D. Kutscher, J. Ott, and C. Bormann, "Session Description and Capability Negotiation," *IETF Internet Draft*, `draft-ietf-mmusic-sdpng-03.txt`, November 2001, work in progress.

[13] W.Y. Ma, B. Shen, and J. Brassil, "Content Services Network: The Architecture and Protocols," *Proceedings of the WCW'01*, Boston, MA, June 2001, pp. 89-107.

[14] The IETF Open Pluggable Edge Services (OPES) Working Group, `http://www.ietf-opes.org`.

[15] H. Schulzrinne, A. Rao, and R. Lanphier, "Real Time Streaming Protocol (RTSP)," *IETF RFC 2326*, April 1998.

[16] SCIFI.com, Mystery Science Theater 3000, `http://www.scifi.com/mst3000`.

[17] C. Shih, "Drawing the line," *Mountain View Voice*, `http://www.mv-voice.com/morgue/2002/2002_02_08.sportvision.html`, February 2002.

[18] Spin The Bottle Dot Com, Pop Up Video, `http://www.popupvideo.com`.

[19] N. R. Sturtevant, N. Tang, and L. Zhang, "The Information Discovery Graph: Towards a Scalable Multimedia Resource Directory," *Proceedings of the IEEE WIAPP'99*, San Jose, CA, July 1999, pp. 72-79.

[20] R. Zakon, "Hobbes' Internet Timeline," `http://www.zakon.org/robert/internet/timeline/`.