Analysis of Enterprise Media Server Workloads: Access Patterns, Locality, Content Evolution, and Rates of Change

Ludmila Cherkasova*, Member, IEEE,* and Minaxi Gupta

***Abstract*—Understanding the nature of media server workloads is crucial to properly designing and provisioning current and future media services. The main issue we address in this paper is the workload analysis of today's enterprise media servers. This analysis aims to establish a set of properties specific to the enterprise media server workloads and to compare them to well-known related observations about the web server workloads. We partition the media workload properties in two groups: *static* and *temporal*. While the static properties provide more traditional and general characteristics of the underlying media fileset and quantitative properties of client accesses to those files (independent of the access time), the temporal properties reflect the dynamics and evolution of accesses to the media content over time. We propose two new metrics characterizing the temporal properties: 1) the new files impact metric characterizing the site evolution due to new content and 2) the life span metric reflecting the rates of change in accesses to the newly introduced files. We illustrate these new metrics with the analysis of two different enterprise media server workloads collected over a significant period of time.**

***Index Terms*—Access patterns, content evolution, dynamics, enterprise media servers, locality, sharing patterns, static and temporal properties, workload analysis.**

I. Introduction

STREAMING media represents a new wave of rich Internet content. Recent technological advancements in video creation, compression, bandwidths, caching, streaming, and other content delivery technologies have brought audio and video together to the Internet as rich media. Products for still (JPEG) and motion (MPEG) pictures are also available in consumer markets. This enables potentially anyone to be a producer of rich media content that can be easily distributed and published over the Internet. There are predictions that rich media will add significantly to the user experience and will be the Internet's next "killer app."

Video from news, sports, and entertainment sites is more popular than ever. Media servers are being used for educational and training purposes by many universities. The use of the media servers in enterprise environment is catching momentum too. Enterprises are using more and more rich media to attract prospective customers, improve the effectiveness of online advertising, web marketing, customer interaction centers, collaboration, and training.

Recently, there have been several studies attempting to understand the multimedia workload characteristics. However, most of the studies are devoted to the analysis of workloads for educational media servers [1]–[3], [13], [14], [17]. One recent study [10] characterizes the workload of a media proxy of a large university. Our paper presents and analyzes the *enterprise media server workloads* based on the access logs from two different media servers at Hewlett-Packard Corporation. Both logs are collected over a long period of time (2.5 years and 1 year 9 months). The duration of the logs makes the studied workload unique and allows us to discover typical and specific client access patterns, media server access trends, and the dynamics and evolution of the media workload over time.

We partition the media workload properties in two groups: *static* and *temporal*. While static properties provide more traditional and general characteristics of the underlying media fileset and quantitative properties of client accesses to those files (independent of the access time), the *temporal* properties, studied in this paper, aim to reflect the dynamics and evolution of accesses to the media content over time.

Web workload studies have identified different types of *locality* in web traffic [4], [5] that strongly influence the traffic access patterns seen by the web servers. One goal of our analysis is to characterize the locality properties in media server workloads and to compare them with traditional web workloads characterization.

The other questions we address in this paper are tightly related to new trends observed in the evolution of the Internet infrastructure such as content distribution networks (CDNs) and overlay networks. CDNs are based on a large-scale distributed network of servers located closer to the edges of Internet for efficient delivery of digital content including various forms of streaming media. The emergence of CDNs has brought a new set of questions about the *client-side characterization.* Since CDNs deal with delivering content and services at the "edge," bandwidths available to clients and their access and viewing patterns are important considerations.

Access patterns and dynamics of the site have to be taken into account when making a decision about using caching or content distribution systems. For example, if the site is very dynamic, i.e., a large portion of the client requests are accessing new content (news web sites being a prime example), then CDNs are clearly a good choice to handle the load, because traditional caching solutions will be less efficient in distributing the load

Manuscript received April 4, 2002; revised December 4, 2003; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor M. Crovella.

L. Cherkasova is with Hewlett-Packard Laboratories, Palo Alto, CA 94303 USA (e-mail: cherkasova@hpl.hp.com).

M. Gupta is with the College of Computing, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: minaxi@cc.gatech.edu).

Digital Object Identifier 10.1109/TNET.2004.836125

due to the time involved in propagating the content through the network caches. Thus, the other question we address in this paper is how to characterize the *dynamics* and *evolution of accesses* at media sites. The first natural step is to observe the introduction of new files in the logs and to analyze the portion of all requests targeting those files. We define a *new files impact* metric that aims to characterize the site evolution due to new content. It is obtained by computing the ratio of the accesses targeting the new files over time. The definition of "new files" depends on the time scale at which information changes and might be different for different sites. We propose a second metric, called the *life span* metric, to measure the rate of change in the access pattern of the site.

We have developed a tool called **MediaMetrics** that characterizes a media server access profile and its system resource usage in both a quantitative and qualitative way. It extracts and reports information that could be used by service providers to evaluate current solutions and to improve and optimize relevant future components.

Key new observations from our analysis include the following.

1) Despite the fact that the two studied workloads had significantly different file size distribution (one set had well-represented groups of short, medium, and long videos, while the other set was skewed in long videos range), the client viewing behavior was similar for both sets: 77%–79% of media sessions were less than 10-min long, 7%–12% of the sessions were 10–30 min, and 6%–13% of sessions continued for more than 30 min. Additionally, this reflects the browsing nature of client accesses in the enterprise workloads under study. Similar browsing access patterns were observed in [2].
2) Most of the incomplete sessions (i.e., terminated by clients before the video was finished) accessed the initial segments of media files. The percentage of sessions with interactive requests (such as pause, rewind, or fast forward during the media session) was much higher for medium and long videos.
3) Like web workloads, both the media workloads exhibit a high locality of accesses: 14%–30% of the files accessed on the server account for 90% of the media sessions and 92%–94% of the bytes transferred and were viewed by 96%–97% of the unique clients.
4) While there was a significant number of files that were rarely accessed (16% to 19% of the files are accessed only once), these numbers are somewhat lower compared to web server workloads.
5) In tune with the findings in [2], we observed that the overall distribution of clients accesses to the media files does not follow a Zipf distribution. However, noteworthy is that the time scale plays important role in this approximation. We considered 1-month, 6-month, 1-year, and a whole log duration as a time scale for our experiments. For one workload, distribution of clients accesses to media files on a 6-month scale starts to fit Zipf-like distribution. While for the other workload, file popularity on a monthly basis can be approximated by Zipf-like distribution. For a longer time scale in the same workloads, the file access frequency distribution does not follow Zipfian distribution.
6) Accesses to the new files constitute most of the accesses in any given month. Also, the bytes transferred due to the accesses to new files are dominant in both workloads. It makes the access pattern of the enterprise media sites under study resemble the access patterns of the news web sites where most of the client accesses target new information. We introduce the *new files impact* metric to measure the site dynamics due to the new files. Moreover, we observed that for the studied enterprise media servers, the tendency of the number of accesses to be increasing or decreasing in nature is strongly correlated with the number of newly added files.
7) For both workloads, 51%–52% of the accesses to media files occur during the first week of their introduction. The first five weeks of a file's existence account for approximately 70%–80% of all the accesses. We define a *life span* metric to reflect the rates of change in accesses to the newly introduced files. The life span metric also reflects the timeliness of the introduced files. The lower rates of change in file accesses reflect that the media content on a site is less timely and has a more consistent access pattern over a longer period of time.

The remainder of the paper is organized as follows. We review related work in Section II. Section III briefly describes the sites we used in our study and provides a short description of the media server log formats. Section IV characterizes static properties of the media files. While Section V is devoted to temporal workload properties and introduces two new metrics to reflect them. Finally, Section VI presents conclusion and future work.

II. Related Work

While web server workloads have been studied extensively [4]–[6], [9], [11], relatively fewer papers have been written about multimedia workloads. Acharya *et al.* [1] characterized nonstreaming multimedia content stored on the web servers. In their later work [2], the authors presented an analysis of a 6-month trace data from the multicast Media on Demand (mMOD) system which had a mix of educational and entertainment videos. They observed a high temporal locality of accesses, a client preference to preview the initial portion of the videos, and that the rankings of video titles by popularity do not fit a Zipfian distribution.

Studies of client accesses to the MANIC system audio content [17] and the low-bit rate videos in the Classroom2000 system [14] analyze the accesses to educational media servers in terms of daily variation in server loads, distribution of media session durations, and client interactivity.

Extensive analysis of educational media server workloads is done in [3]. Their study is based on two media servers (eTeach and BIBS) in use at major public universities in the United States. The authors provide a detailed study of the client session arrival process; the client session arrival in the BIBS workload can be characterized as Poisson, while arrivals in the eTeach workload are closer to a heavy-tailed Pareto distribution. They also observed that the media delivered per session depends on the media file length. They discovered different client interactivity patterns for frequently and infrequently accessed files. In particular, each video segment is equally likely to be accessed for frequent files, while access frequency is higher for earlier segments in the infrequent videos.

While all the above papers used media server logs, the study by Chesire *et al.* [10] analyzed the media proxy workload at a large university. The authors presented a detailed characterization of session duration (most media streams were less than 10 min long), object popularity (78% of the objects were accessed only once), server popularity, and sharing patterns of the streaming media among the clients.

As the number of Internet users continues to grow and as the high-speed access methods become more ubiquitous, streaming media starts to occupy a sizable fraction of the Internet's bandwidth. A few recent papers [15], [16], [22] analyze the impact of streaming media on the Internet traffic and the performance of popular Internet real-time streaming technologies.

Our paper extends the existing work in a number of significant ways. To our knowledge, this paper is the first study of enterprise media server workloads. Our data is collected over a significant period of time, which makes it unique. The duration of the data allowed us to analyze dynamics and evolution of the media workload over time and to propose two new metrics to measure these properties.

III. Media Workloads Under Study

A. Data Collection Sites

We used access logs from two different servers.

- **HP Corporate Media Solutions server (HPC)** hosts diverse information about HP: video coverage of major events, keynote speeches, addresses and presentations, meetings with industry analysts, promotional events, product introduction, information related to software and hardware products, and demos illustrating the products usage. Additionally, it hosts some training and educational information. The logs cover almost 2.5 years of duration from the middle of November 1998 to the middle of April 2001. This site is supported by a media server cluster. For our analysis, we combined several access logs collected at this cluster. The HPC content is delivered by the Windows Media Server [23].
- **HPLabs Media server (HPLabs)** provides information about HP Laboratories; in particular, it hosts videos of monthly HPLabs-wide meetings, videos of prominent presentations, seminars, and meetings, some of the HP wide business related events, promotional materials, and some training and educational information. The logs cover 1 year and 9 months of duration from the middle of July 1999 to the middle of April 2001. It is an internal server available only for access to HP employees. The HPLabs content is delivered by the RealServer G2 [18].

B. Media Server Log Formats

The media access logs record information about all the requests and responses processed by the media server. Each line of the access logs provides a description of the user request for a particular media file.

TABLE I
Statistics Summary for the Two Sites

	HPC	HPLabs
Duration	29 months	21 months
Total sessions	666,074	14,489
Total Requests	1,179,814	NA
Unique Files	2,999	412
Unique Clients	131,161	2,482
Storage Requirement	42 GB	48 GB
Bytes Transferred	2,664 GB	172 GB

Windows Media Server and RealNetworks Media Server have different log formats, but the typical fields contain information about the IP-address of the client machine making the request, the time stamp of the request, the filename of the requested document, the advertised duration of the file (in seconds), the size of the requested file (in bytes), the elapsed time of the requested media file when the play ended (a play is ended prematurely when the client hits the stop button), the average bandwidth (kilobytes per second) available to the user while the file was playing, the number of bytes sent by the server, and the number of bytes received by the client (for more details on media log formats see [8]).

A client can pause, rewind, fast forward, or skip to a predefined point using a slide bar while viewing the requested media file. We define each such interaction as a "request." A *session* is a sequence of client requests corresponding to the access of a particular file. There can be *multiple requests* corresponding to the same session, due to client's interactivity.

Windows Media Server logs contain a separate entry for each client request. Thus, a single session may be comprised of multiple entries in the server access logs. Each log entry has a *start position*, the place where the client started viewing the file; *duration* is length of time the client watched the file for; and *client action* is pause/stop/rewind/fast/forward. This is useful information for the analysis of clients' interactive behavior during the media sessions.

RealServer log format allows for similar fields, but unfortunately, in the HPLabs server, the relevant options were not switched on. Thus, HPLabs workload contains only information about client sessions and not about client requests. There is one entry for each client session in these logs. As a result, the client interactivity data are not available for HPLabs workload.

C. Summary Statistics

The overall workload statistics for the HPC and HPLabs media servers are summarized in Table I.

In HPC, 471 files corresponded to live streams, while the others were stored content. Since most of the fields reflecting client and server activities are not defined or are not applicable to a multicast session (i.e., most of the fields are either "0" or "-" in the log's entry), we excluded the log records corresponding to live streams from further analysis.

A glance at the basic statistics shows that the HPC media server witnesses more activities and reaches larger client population than the HPLabs server. The HPLabs server clearly targets more a specific smaller research community at HP and as

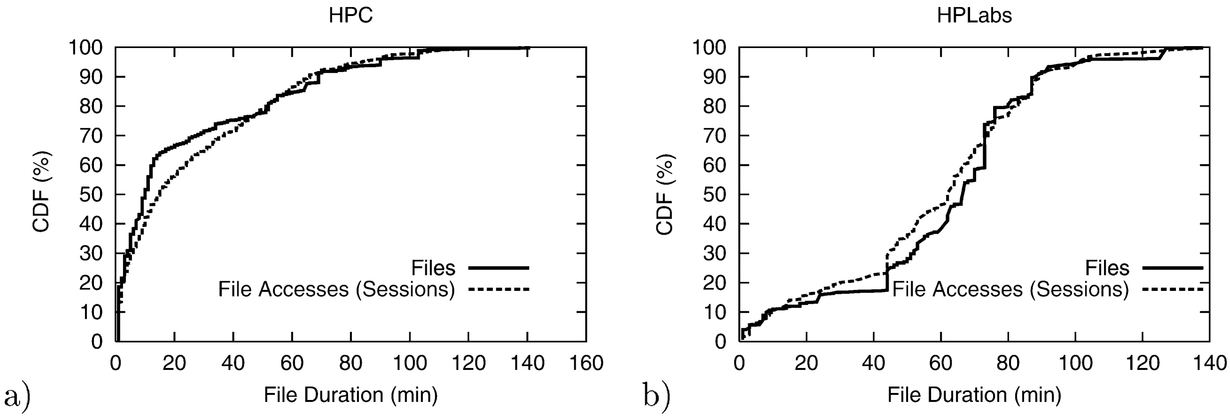

Fig. 1. CDF of file (video) durations and client sessions to them for (a) HPC and (b) HPLabs.

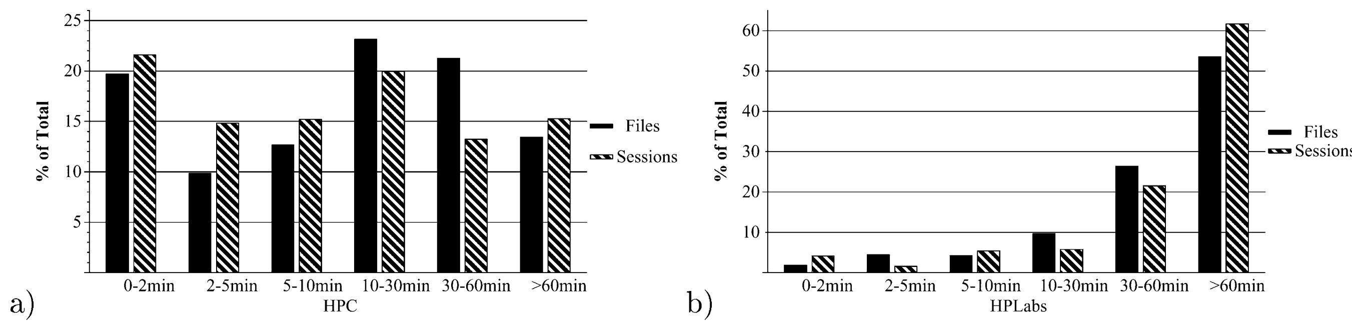

Fig. 2. Six classes of file durations and percentage of client sessions to each of them for (a) HPC and (b) HPLabs.

a result has a very different "modest" profile. The HPC workload represents a reasonably busy media server with 300–800 client sessions each week day and occasional peaks reaching 12 000 sessions. The HPLabs server witnesses a lighter load. By noticing this very obvious difference, it becomes even more interesting whether we can find common properties for the considered enterprise media workloads.

D. Static and Temporal Properties of Media Workloads

We partition media workload properties in two groups: *static* and *temporal*.

1) **Static properties** provide the characteristics of the underlying media file set, reflect the aggregate quantitative properties of client accesses, and present the properties of individual file accesses. Static properties include:
   a) file and session duration characteristics;
   b) file encoding bit rates and session bandwidths;
   c) characterization of completed, aborted, and interactive sessions;
   d) workload locality characterization;
   e) file access popularity;
   f) client characterization.
2) **Temporal properties** characterize the evolution of media site content and rate of changes of accesses to media content over time. To reflect temporal properties we introduce two new metrics:
   a) new file impact metric;
   b) file life span metric.

Sections IV and V describe the static and temporal properties of studied media workloads in more detail.

IV. Static Properties

A. File and Session Characteristics

In this section, we provide the basic characterization of the media files referenced in the logs and the corresponding client sessions.

The advertised duration of the media file reflects the total length of the video. First, we analyze the distribution of the durations of stored videos and the distribution of the client sessions to the corresponding files. Fig. 1 shows the *cumulative density function (CDF)* of the stored videos and the CDF of the corresponding accesses to them over the advertised media duration for both workloads.

To simplify further analysis, we created six classes of the videos based on duration, including three groups of short videos: 1) less than 2 min; 2) 2–5 min; 3) 5–10 min; one group of medium size videos: 4) 10–30 min; and two groups of long videos: 5) 30–60 min; and 6) longer than 60 min. Fig. 2 shows the percentage of stored videos for each of the six classes and the percentage of corresponding sessions to them.

Fig. 2(a) shows that for the HPC workload the content is well represented by videos of different durations: 42% of files belong to the short video group (less than 10 min), 23% of files are in the medium video group, and 34% of files belong to the long video group. The HPLabs workload is strongly skewed in favor of long videos as shown in Fig. 2(b); only 7% of videos are in the medium group, while 79% of files belong to the long video group.

Even though the two workloads have different distribution of short, medium, and long videos, it is interesting to note that the percentage of client accesses to the files in each of the duration

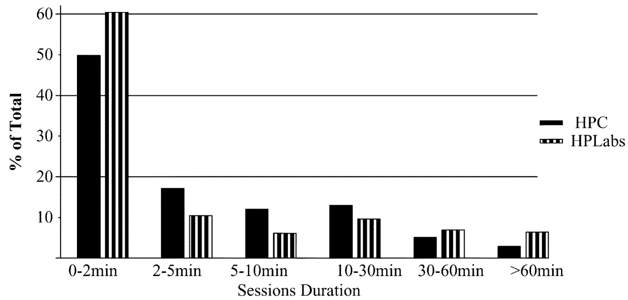

Fig. 3. Session duration characterization.

categories is proportional to the percentage of files in each of them for both workloads! This implies that each of the duration groups is equally likely to be accessed by clients. This property is very useful for synthetic workload generation, since it proposes a simple model of defining a media file duration distribution and percentage of corresponding client accesses to those files.

During the media session, the client can stop viewing or downloading the file by hitting the stop button before the video is finished; or browse the video through a sequence of pause, rewind, fast forward; or jump to specific sections of the video by using the slide bar. We define a session duration as the actual client viewing time. When we analyzed the actual duration for which clients viewed the videos, the statistics change dramatically for both workloads. As shown in Fig. 3, for both workloads, 50%–60% of the media sessions were less than 2 min long. The statistics presented in this graph reflect the overall client viewing time distribution, and it is not correlated with the actual distribution of the duration of the media files.

Another noteworthy fact from Fig. 3 is that in spite of the significant differences in the original file size distribution, the distribution of the actual durations for which the clients viewed the videos was similar for both logs: 77%–79% of the media sessions were less than 10 min long, 7%–12% of the sessions were 10–30 min, and only 6%–13% of sessions continued for more than 30 min. Given that the decision to abort the session is only partially influenced by the available bandwidth (we will show this analysis in more detail in Section IV-C), the observed access pattern highlights the browsing nature of the enterprise client accesses. The knowledge of the approximate percentage of "browsing" clients could help in estimating and predicting the short term load on the media server.

B. Media File Encoding Rates and Session Bandwidths

Both servers, HPC and HPLabs, had videos encoded at different rates. Table II presents the statistics on file encoding rates and their trends over time for both workloads. Videos stored at the HPC server had most of the files (59%) encoded at a 56-kb/s rate and lower. However, over the years, the trend showed that more files at the HPC site were encoded at a higher rate: for example, in 1999, only 4% of the videos were encoded at rates 128 kb/s and higher, while in 2001 this group of videos constituted 29% of the total. On the other hand, the HPLabs server had most of the files encoded at high bit rates; 84% of all the files were encoded at 128 kb/s and higher.

TABLE II
TRENDS IN FILE ENCODING RATES FOR BOTH WORKLOADS

Period	HPC		
Bit Rate	≤56kb/s	56-128kb/s	≥128kb/s
Files (1999)	74%	22%	4%
Files (2000)	56%	27%	17%
Files (2001)	53%	18%	29%
All files	59%	21%	20%
	HPLabs		
Bit Rate	≤56kb/s	56-128kb/s	≥128kb/s
Files (1999)	16%	7%	77%
Files (2000)	10%	5%	85%
Files (2001)	13%	2%	85%
All files	11%	5%	84%

TABLE III
TRENDS IN AVERAGE AVAILABLE BANDWIDTH PER SESSION FOR BOTH WORKLOADS

Period	HPC		
Bandwidth	≤56 kb/s	56-128kb/s	≥128kb/s
Sessions (1999)	57.8%	42%	0.2%
Sessions (2000)	40%	52%	8%
Sessions (2001)	36%	57%	7%
All Sessions	43%	51%	6%
	HPLabs		
Bandwidth	≤56kb/s	56-128kb/s	≥128kb/s
Sessions (1999)	71%	15%	14%
Sessions (2000)	79%	17%	4%
Sessions (2001)	78%	18%	4%
All Sessions	75%	16%	9%

Media access logs report the average bandwidth available to the user session while the file was playing. The term "average available bandwidth per session" is used in the description of the media log format, and it deserves an explanation. This is the number reported in the log for each client request retrieving a stored video from the server. Typically, this metric does not reflect the overall bandwidth available to the client. Since each video is encoded at a certain bit rate X kb/s, the required bandwidth to the client for an ideal transfer of the corresponding video is X kb/s. The network path between the client and the server may have a higher available bandwidth than X kb/s. However, only the consumed bandwidth per media session is reported in the log entry. Thus, for the case when the network path between the client and the server has an available bandwidth lower than X kb/s, the achieved bandwidth per session is correspondingly lower and this actually consumed bandwidth is reported in the log entry.

Table III presents the statistics on available session bandwidths during the different time periods of logs: years 1999, 2000, and 2001. Overall, the HPC media sessions had higher available bandwidths to the clients, with an increasing trend over the years. For example, in 2001, 64% of sessions had an average available bandwidth above 56 kb/s (we will call these sessions as *high-bandwidth sessions*). For the HPLabs workload, in 2001, the high-bandwidth sessions constituted only 22% of the total.

For the HPC workload, most of the file encoding bit rates and the average available bandwidth per session show a good alignment as shown in Fig. 4. Only the group of videos encoded at rates 128 kb/s and higher could not meet the requirements. While for the HPLabs workload, where the most of the files were encoded at 128 kb/s and higher, the gap between the demand and

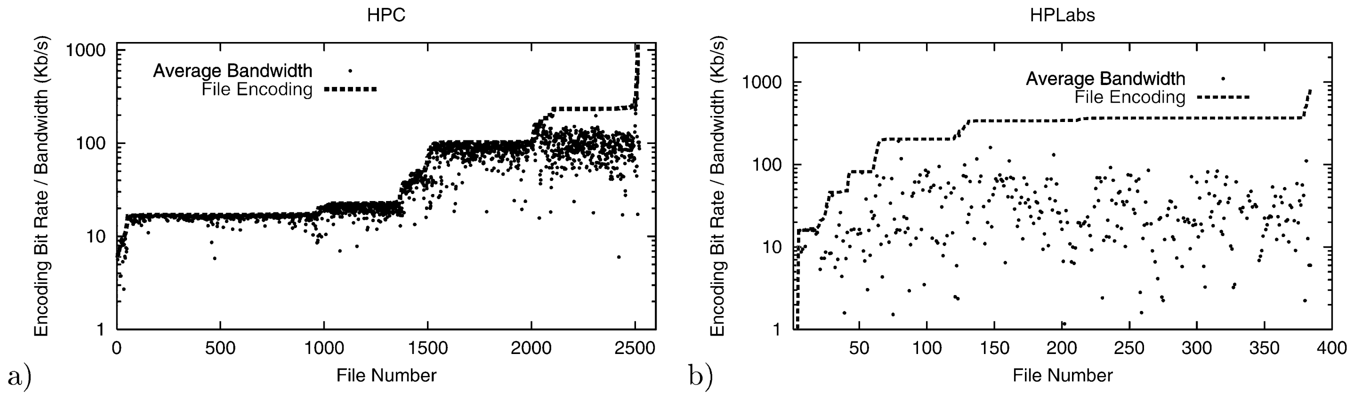

Fig. 4. File encoding bit rates and average available bandwidth per session to those files: (a) HPC and (b) HPLabs.

achieved session bandwidth is very high. Most of the sessions have a significant mismatch between the file encoding bit rates and the available bandwidths.

The information on the file encoding bit rate versus average achieved session bandwidth, provided by MediaMetrics, could be used by the service providers for choosing the appropriate bit rates during the file encoding.

Media access logs also report the number of bytes sent by the server and the number of bytes received by the client. Our tool uses this information to estimate the percentage of bytes lost during the file transfer and to implicitly judge the quality of service a client might have experienced.[1] This simple technique can produce useful results when data is transmitted over UDP, because in that case the difference in sent and received bytes reflects the percentage of the bytes lost on the way to the client. It might be less accurate when data is transferred over TCP because in the presence of congestion, the media server will retransmit part of the data to compensate for lost packets. If those packets were received by the client in time, then the difference in server sent-bytes and the client received-bytes will not result in a worse QoS. As mentioned earlier, the HPC data were transmitted using UDP, while the HPLabs data were transferred over TCP.

In the HPC workload, the percentage of "good quality" sessions with 0%–5% of byte loss was very high; 96.5% of low-bandwidth sessions and 97.1% of high-bandwidth sessions experienced the good quality. For the HPLabs workload, the corresponding numbers were much lower; only 64.6% of low-bandwidth sessions and 88.8% of high-bandwidth sessions experienced 0%–5% of byte loss per session.

C. Characterization of Completed, Aborted, and Interactive Sessions

We call a media session *completed* if during this session the entire media file was transmitted to the client.[2] For the HPC workload, 29% of sessions were completed, while for the HPLabs workload, completed sessions accounted for only 12.6% of all the sessions. Fig. 5(a) shows the CDF of all the completed sessions, where the X axis represents the advertised duration of the corresponding videos.

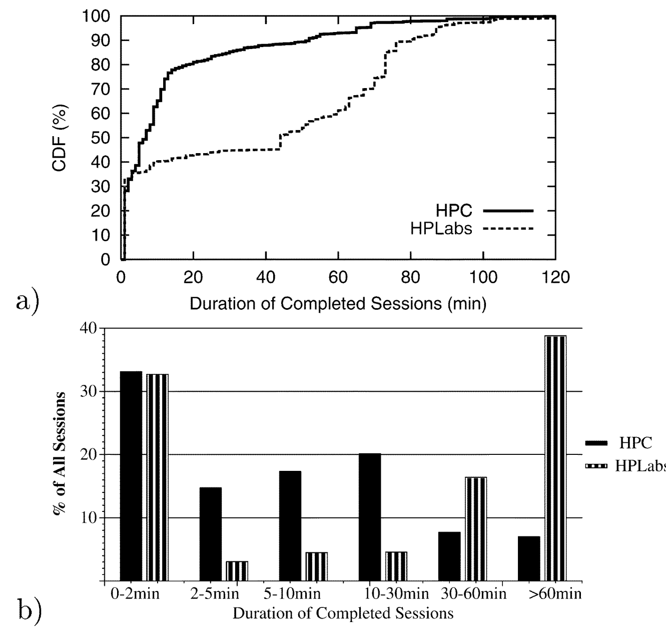

Fig. 5. (a) CDF of completed session by advertised duration of the corresponding video. (b) Simplified distribution of completed sessions for six duration classes.

Fig. 5(b) presents a simplified view of the same distribution of all the completed sessions via the six duration classes as defined in Section IV-A. Media sessions with duration under 2 min account for 33% of all the completed sessions for both workloads. While for the rest of the completed sessions, their durations reflect the corresponding distribution of media session durations specific to considered workloads as shown in Figs. 1 and 2.

A reasonable question to ask is whether the completed sessions had higher available bandwidths to the clients. In other words, were the aborted sessions interrupted because of poor available bandwidth?

Table IV presents the statistics on available bandwidths for completed, aborted, and all the sessions for both workloads. For HPC workload, the completed sessions had higher percentage of high-bandwidth sessions. However, the difference in the

[1]While the byte loss metric does not directly translate into the perception quality for the client, it is a useful indicator of networking conditions, where the high loss rate can be used as an alarming event about degraded quality of the viewed video. Typically, the discrepancies caused by the packet loss in the range of 0%–5% can be successfully concealed by the error correction algorithms implemented in the current media players [20], [21].

[2]We are able to determine whether a session was completed by comparing the size of stored video to the number of bytes received by the client.

TABLE IV
DISTRIBUTION OF AVAILABLE BANDWIDTH PER SESSION FOR BOTH WORKLOADS

	HPC		
Session Bandwidth	≤56kb/s	56-128kb/s	≥128kb/s
Completed sessions	33.9%	60.7%	5.4%
Aborted sessions	47.1%	47.4%	5.5%
All sessions	43.3%	51.1%	5.6%
	HPLabs		
Session Bandwidth	≤56kb/s	56-128kb/s	≥128kb/s
Completed sessions	77.3%	11.5%	11.2%
Aborted sessions	74.8%	15.8%	9.4%
All sessions	75%	15.5%	9.5%

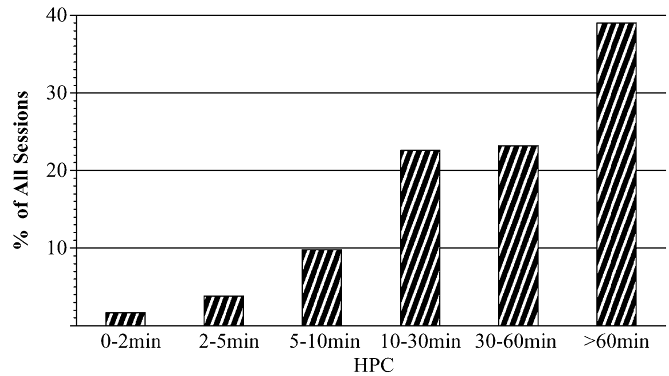

Fig. 6. HPC workload: percentage of sessions with interactive requests for videos of different duration classes.

achieved session bandwidths is not high enough to assert that the sessions were aborted because of the "poor bandwidth" conditions. For the HPLabs workload, the bandwidth characteristics of the completed and aborted sessions were similar, which suggests that the clients perhaps watch the video only while they are interested in the content of the video.

Another interesting observation is that most of the aborted sessions were accessing the initial segments of media files. The number of sessions which had incomplete accesses to any other segments of the file other than the beginning depend on the size of the video: less than 1.5% of sessions in short video group accessed any segment of the video other than the beginning, 2.4% of sessions in a medium video group, and 4%–7% of sessions in long video group. Clearly, such knowledge about the client viewing patterns may be beneficial when designing media caching strategies.

Windows Media Server log format has a separate entry for each client request (recall that a client session may consist of more than one request, where each request represents the client's interactivity). As a result, we were able to get information about client activities such as pause, rewind, and fast forward in the HPC workload. Unfortunately, similar data was not available for HPLabs workload because the relevant option was turned off in the RealServer used by HPLabs. Analysis of these fields for HPC logs produced several intuitive but interesting results. First of all, it revealed that 99.9% of the sessions with interactive requests were *high-bandwidth sessions* with available bandwidth greater than 56 kb/s. Second, the percentage of sessions that accessed medium and long videos had much higher interactivity.

Fig. 6 shows that only 15.3% of sessions with interactivity were for the short video group, 22.6% of interactive sessions were for the medium size videos, and 62.2% of sessions had client interactivity for the long video group. Such statistic helps in better understanding the clients' viewing behaviors.

D. Locality Characterization

In this section, we will revisit a previously identified invariant for the web server workloads. The authors in [5] identified that the web traffic exhibits strong concentration of references: "10% of files accessed from the server typically account for 90% of the server requests and 90% of the bytes transferred."

For the locality characterization of our logs, we used a table of all files accessed along with their frequency (number of times a file was accessed during the observed period) and the file sizes. This table is ordered in the decreasing order of frequency.

Fig. 7(a) shows the locality of reference for the media server access logs used in our study: 90% of the media sessions target 14% of the files for the HPC server and 30% of the files for the HPLabs server. This shows a high locality of client accesses, though lower than observed locality in the web workloads.

Fig. 7(b) shows the bytes transferred by the corresponding media sessions: 94% for the HPC site and 92% for the HPLabs site. The observed graphs for both workloads are remarkably similar. Fig. 7(c) shows locality of clients for both workloads. It can be interpreted in the following way: at the HPC server, 14% of the most popular files (responsible for 90% of the accesses) are accessed by 96% of clients. For the HPLabs site, 30% of the most popular files are viewed by 97% of the clients.

We also analyzed the locality in the workload from a different angle: what percentage of active storage did the most popular files account for? Here, the *active storage* set is defined by the combined size of all the media files accessed in the logs. As shown in Fig. 8(a), for both workloads, we observed a high active storage set locality: 80%–88% of all sessions are to the files that constitute only 20% of the total active storage set. Similarly, 82%–92% of all the transferred (most popular) bytes are due to files that constitute only 20% of the total active storage set, as can be seen in Fig. 8(b).

This type of analysis helps in estimating the storage requirements and the potential bandwidth savings when using optimizations for the popular portion of the media content. Since these metrics are normalized with respect to the site's active storage set, it allows us to compare different workloads and to identify the similarities inherent to those workloads, independent of the absolute numbers for storage in each workload.

We also analyzed whether the locality characterization of the workload significantly changes depending on the chosen time scales. We found that both workloads exhibit a high locality of client accesses independent of the duration (1-month, 6-month, or 12-month durations); 90% of the media sessions target 10%–30% of the files for the HPC server during each duration interval and 20%–35% of the files for the HPLabs server. This reflects that a high locality of client accesses is an inherent property of the enterprise media server workloads and is not impacted by the choice of the time scale.

Complementary to the characterization of the most frequently accessed files, it is useful to have statistics about the "opposites,"

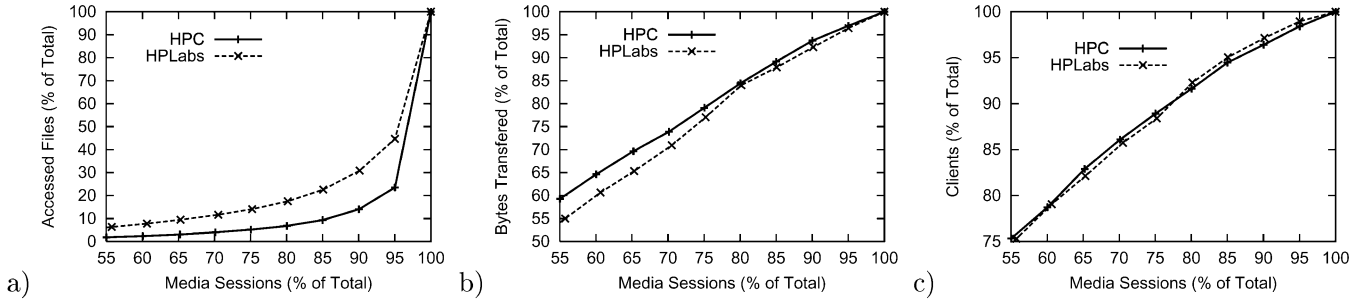

Fig. 7. Two workloads compared: (a) file set locality, (b) bytes-transferred locality, and (c) client set locality.

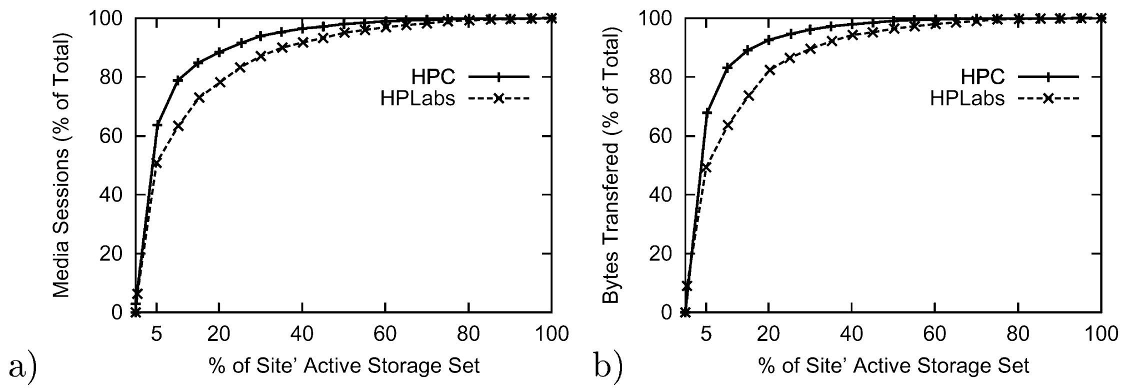

Fig. 8. Two workloads compared: (a) storage set locality and (b) bytes-transferred storage locality.

TABLE V
RARELY ACCESSED FILE STATISTICS

	Files Requested up to 1/5/10 times	Storage Requirements for Corresponding Files
HPC	16% / 38% / 47%	10% / 26% / 34%
HPLabs	19% / 45% / 59 %	17% / 39% / 52%

the percentage of the files that are requested only a few times and the percentage of active storage these files account for.

As Table V shows, 16%–19% of the files are accessed only once, and 47%–59% of the files are accessed less than ten times. These rarely accessed files account for quite a significant amount of storage: 34%–52% of total active storage set. These numbers, however, are somewhat lower compared to the web server workloads. For web server workloads, "onetimers" (files accessed only once) may account for 20%–40% of the files and the active storage.

The locality properties of the client references as well as the knowledge about the rarely accessed files are very important in designing the media proxy caching strategies and efficient content placement on distributed media servers and media proxies. As part of our future work in this direction, we intend to explore the temporal locality of client accesses as well as the degree of file sharing among the clients. We expect that this information will serve as a basis for using media delivery optimizations such as multicast.

E. File Access Popularity

Previous studies on web servers and web proxies [7] led to almost a universal consensus that the web page popularity follows a Zipf-like distribution, where the popularity of the ith most popular file is proportional to $1/i^{\alpha}$. For web proxies, the value of α is typically less than one, ranging from 0.64 to 0.83, while for web servers the typical value of α lies between 1.4–1.6. In [10], devoted to the analysis of a media proxy workload, the authors reported a Zipf-like distribution for the file access frequencies with $\alpha = 0.47$. In [3], the authors approximated the educational media server daily workloads using the concatenation of two Zipf-like distributions.

Since our workloads cover a significant period of time, we decided to investigate whether the file access frequencies exhibit the same behavior at different time scales. We considered 1-month, 6-month, 1-year, and the entire duration of the logs as the time scales for our experiments.

In order to characterize the distribution of the file access frequencies for our workloads, we ranked the files by popularity (i.e., the number of accesses to each file) and plotted the results on a log–log scale. Fig. 9(a) shows the file popularity over the entire duration of the logs. Both workloads exhibit very similar distributions: the HPLabs curve "follows" the HPC curve, but on a lower scale. This can be explained by almost two orders smaller number of accesses and files in the HPLabs workload. However, both of these curves are far from fitting a straight line of the Zipf-like distribution. Fig. 9(b) shows the file popularity for the HPC and HPLabs workloads for one year (year 2000) as well as 6-month intervals (the first half and the second half of year 2000). The HPC curves (both 1-year and 6-month) are still far from fitting a straight line of the Zipf-like distribution.

However, the 6-month curves for HPLabs fit reasonably well with the straight line of Zipf-like distribution when the first 15 to 20 files are ignored (in [7], authors make similar assumptions about ignoring the top 100 documents and a flat tail at the end

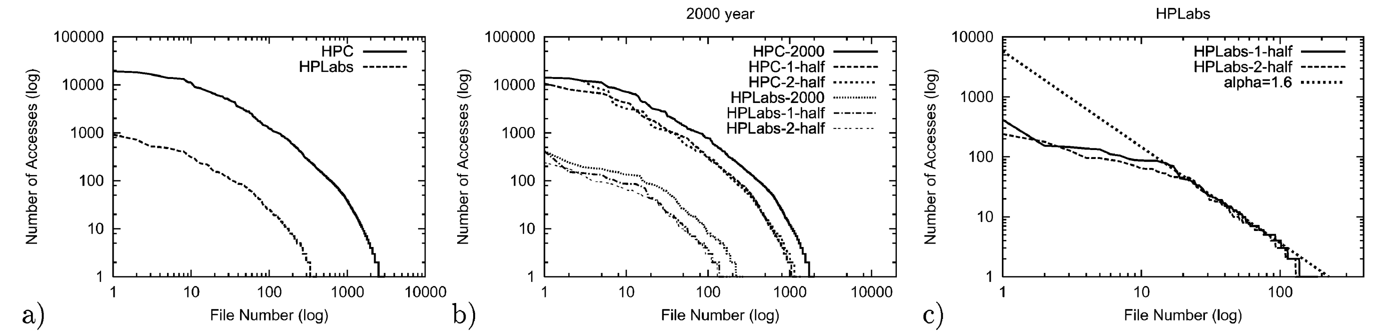

Fig. 9. File popularity distribution for both workloads: (a) over the entire duration of the logs, (b) over year 2000, and the first and the second 6 months of year 2000, (c) 6-month periods for the HPLabs workload with the corresponding Zipf-like function fitting.

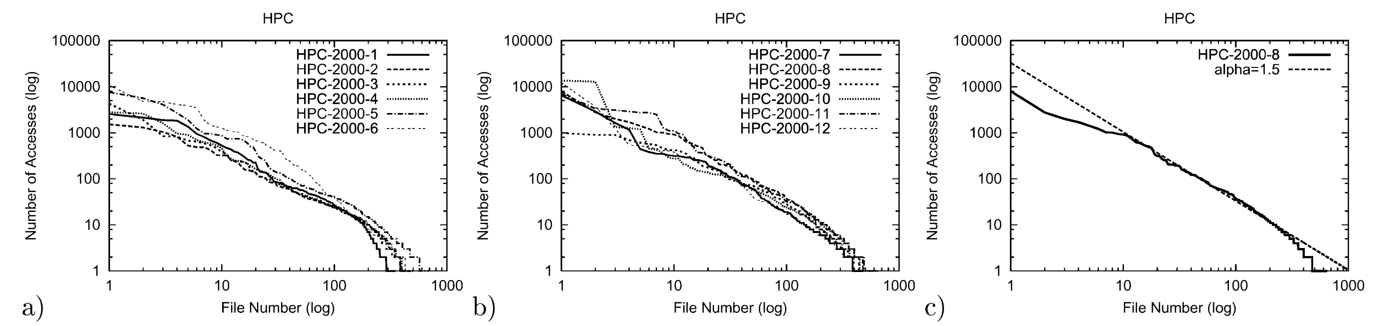

Fig. 10. File popularity distribution for the HPC workload: (a) monthly periods from 1 to 6 months, (b) monthly periods from 7 to 12 months, and (c) month 8 with the corresponding Zipf-like function fitting.

of the curve). A straight line on the log–log scale implies that the file access frequency is proportional to $1/i^{\alpha}$. Using a least square fitting, we obtained the values of α; for both the 6-month curves $\alpha = 1.6$ works very well. Fig. 9(c) shows the file popularity distribution for the HPLabs workload corresponding to the 6-month periods of 2000, approximated by a Zipf-like function $1/i^{\alpha}$ with $\alpha = 1.6$.

Finally, Fig. 10(a) and (b) shows the file popularity for the HPC workload on a monthly basis. Most of the monthly curves fit a straight line reasonably well when ignoring the first 10–15 files and last few files. For different months, the value of α ranges from 1.4 to 1.6. Fig. 10(c) shows the file popularity distribution for the HPC workload during August 2000, approximated by Zipf-like function $1/i^{\alpha}$ with $\alpha = 1.5$.

The observation that the file access frequencies for the media server workloads under study can be approximated by Zipf-like distribution is very useful for synthetic workload generation. An important fact to keep in mind is that the time scale plays an important role in this approximation.

In the recent work [19], where the authors used the same traces (HPC and HPLabs) for a synthetic media workload generator design, a new *generalized Zipf-like* distribution is proposed as a unified method to capture file popularity distributions of both Zipf-like and circular-curve shapes. For the technical details, we refer the readers to [19]. Here, we would like to discuss some intuition behind the generalized Zipf distribution. The reason that the original traces do not show perfectly straight lines at the heads of the curves is that there is little differentiation in the frequencies of the most popular files. This property is tightly related to the nonstationary popularity of media accesses (or the change of file popularity rank) defined by a file life span introduced and discussed in detail in Section V-B as well as the fact that a long-term trace can collect enough files with similar high popularities over time and, thus, these files should be considered as a group (equivalence class) where a group rank is a better reflection of the file popularities.

A Zipf-like distribution is a special case of the generalized Zipf distribution [19]. Thus, the generalized Zipf distribution can be used for the characterization of the file access frequencies in both the long-term and short-term media workloads.

F. Client Characterization

The high locality of accesses in studied media workloads implies that the popular files are widely accessed by many different clients as shown in Fig. 11(a) and (b). For the HPC workload, the first 70 files are accessed by more than 1000 unique clients, with some frequent files accessed by as many as 10 000–12 000 unique clients. (Note that for better viewability we used a log scale for file number/rank.)

For the HPLabs server, the degree of sharing is lower (it is expected, because of the smaller clients population), but for the most frequent files it is still very significant: the first 17 files are accessed by 113–341 unique clients. The sharing exhibited by the clients' access patterns is essential for designing an efficient caching infrastructure.

Our tool MediaMetrics provides information about the client clustering by associating them with various autonomous systems (ASs). It also reports the corresponding number of client sessions and percentage of bytes lost for those sessions for each AS. Since HPLabs logs only had HP's internal clients, they all belong to the same AS and the results of per AS analysis are not particularly interesting for this case. Here, we present some statistics about the HPC workload.

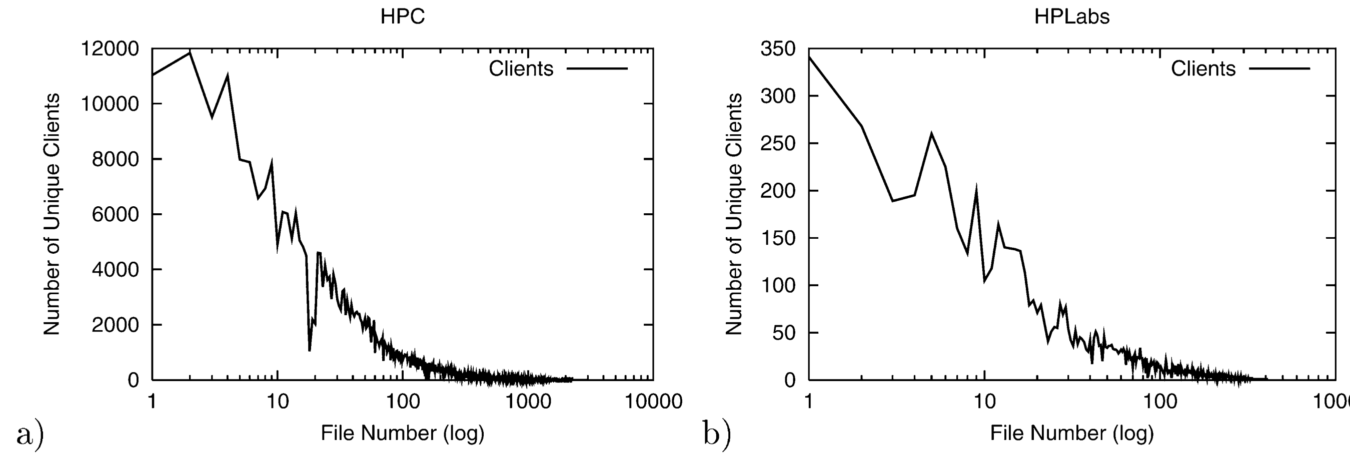

Fig. 11. Files sharing statistics: (a) HPC and (b) HPLabs.

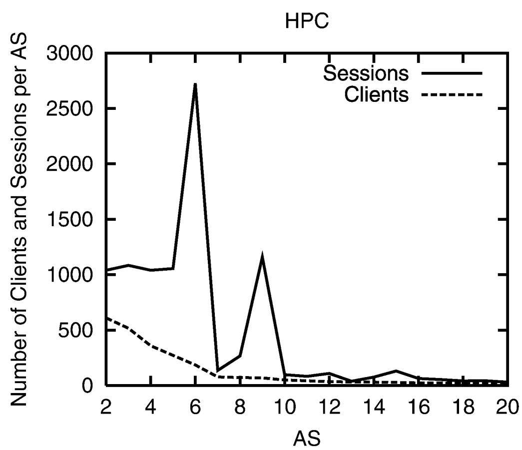

Fig. 12. Clients and their sessions per AS.

For HPC logs, the client population was spread across 200 different ASs, with 82% of the clients being HP internal clients. A total of 93% of all the sessions belonged to the internal HP clients. About 45% of all the ASs had just one client (with 1–16 sessions). To show the client clustering by ASs, we plotted 20 ASs with the largest number of sessions excluding the AS that represents HP.

Fig. 12 shows the number of clients and the corresponding number of sessions for the 20 ASs with the largest number of sessions. If we normalize this data, then the combined clients from the first ten ASs account only for 1.6% of clients population and 1.5% of all the sessions. Clearly, the client population profile was dominated by HP internal clients and their activities. For enterprise media servers, this might be a typical client characterization (while it obviously depends on the type of material hosted by the enterprise media server). Overall, with the spread of CDNs and overlay network technologies, the understanding of clients, the content they access, and their clustering will play an essential role in deciding efficient placement of edge servers and the content.

V. Temporal Properties

A. *New Files Impact in Dynamics and Evolution of Media Sites*

In this section, we investigate specific file access patterns reflecting the dynamics and evolution of accesses to the media content over time.

The first natural step is to observe the introduction of new files in the logs and to analyze the portion of all the requests targeting those files. We define a metric called *new files impact* to characterize the site evolution due to the new content, by computing the ratio of the accesses targeting these new files over time. Figs. 13(a) and 14(a) show the two curves for HPC and HPLabs workload, respectively. The curves show all the files which were accessed in a particular month and all the new files which were accessed in the same month.[3] We define a file as being *new* if it was never accessed before, based on the information in the access logs.

The HPC site has an explicit growth trend with respect to the total number of files accessed per month. A consistently steady number of new files is added to the site during each month. In general, the analysis of the HPC workload revealed the following growth trends: 1) the total number of sessions in each 6-month duration doubled over the duration of the logs and 2) the total number of unique clients accessing the media content in each 6-month duration also doubled over the duration of the logs.

The growth of the total number of files accessed each month for the HPLabs site is "negative." Since this was unexpected, we asked the team supporting this site whether there were specific reasons for the trend we observed. Specifically, we wanted to know if there was a significant number of new video files that "nobody watched," and hence the logs did not contain any information about them or if the new media content at that site had actually decreased over time. The media site support team explained that lately they had been adding only a limited number of new files because they were working on a transition plan to upgrade the entire site design and equipment. So, the "negative" trend in the addition of new files to the site was observed correctly.

Figs. 13(b) and 14(b) show the number of all the sessions per month and the number of sessions to the new files in the corresponding month for the HPC and HPLabs workloads, respectively. These graphs reflect that the accesses to the new files con-

[3]We used one month as a time unit in our study for this metric because it exhibited the observed trends in the most explicit way. Typically, a choice of a "correct" time scale is impacted by: 1) a time scale when new content is added (if new content is added every two months then the time unit should be two months or longer and 2) the life span of the files discussed in detail in Section V-B.

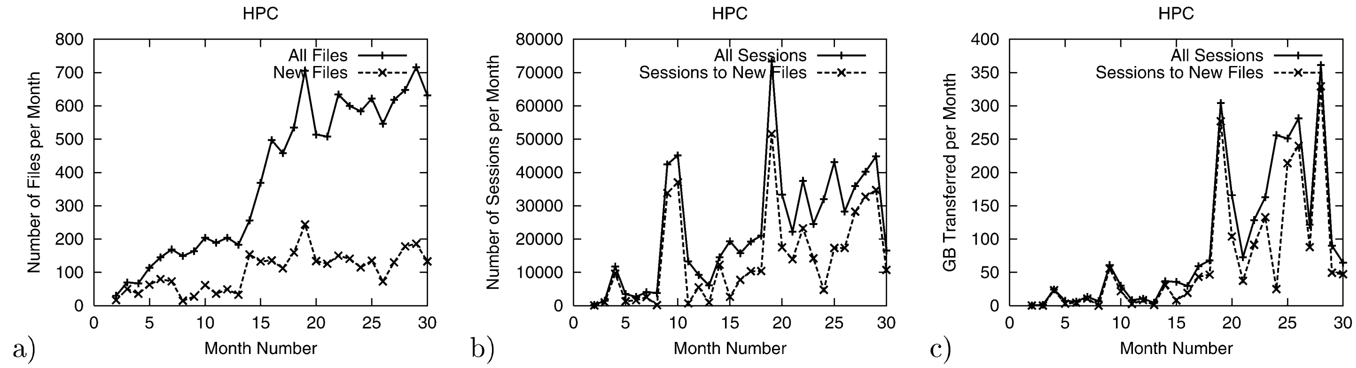

Fig. 13. HPC workload: (a) all and new files per month, (b) all sessions and sessions to new files per month, (c) all bytes transferred and bytes transferred due to the accesses to the new files per month.

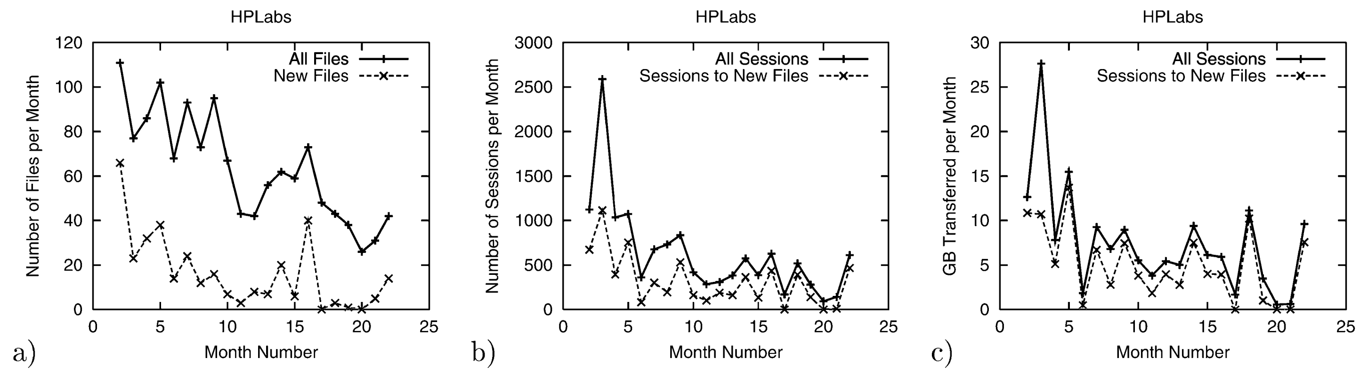

Fig. 14. HPLabs workload: (a) all and new files per month, (b) all sessions and sessions to new files per month, (c) all bytes transferred and bytes transferred due to to the accesses to the new files per month.

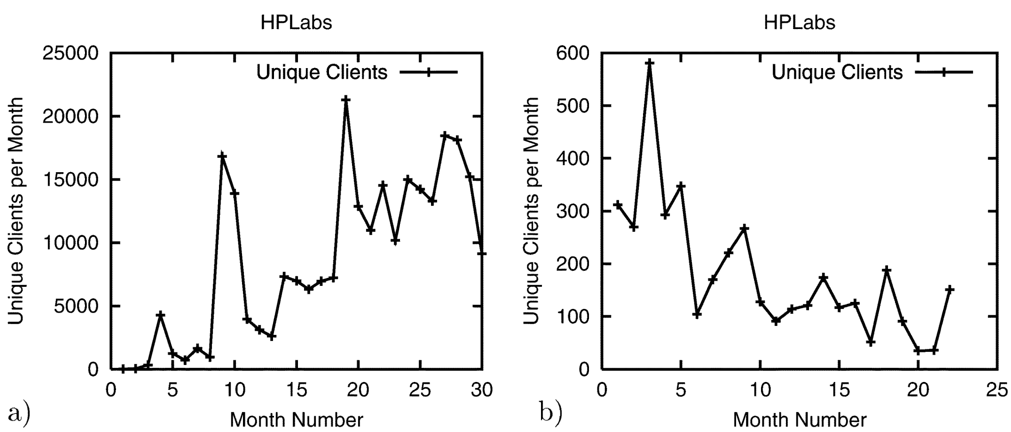

Fig. 15. Unique clients per month: (a) HPC and (b) HPLabs.

stitute the most or a very significant portion of all accesses, excluding a few months that were exceptions. Figs. 13(c) and 14(c) show very similar trends for the bytes transferred per month and the bytes transferred due to the accesses to new files. Since the number of new files added per month plays a crucial role in defining the site dynamics, evolution, and growth trends, evaluating the *new files impact* metric becomes very important.[4]

Figs. 15(a) and (b) show the number of unique clients accessing each of the HPC and HPLabs sites per month, respectively. Again, the trends in these graphs are correlated with the trends in the number of sessions to each site's new files. Thus, it appears that the client population at the enterprise media site is correlated to the amount of new information regularly added to the site.

In contrast, the enterprise *web sites* exhibits much more stability in terms of the accesses to the "old" documents. Only about 2% of the monthly requests are to the new files added that month as shown in [9]. The access pattern of the enterprise *media sites* seems to resemble the access pattern of *news web sites*, where most of the client accesses target the new regularly added information.

B. File Life Span

In this section, we attempt to answer the following question: how much does the popularity of a file and the frequency of

[4]In the analysis of the logs, we only derive the observations without knowing when and "how" the new files are added and when the old files are deleted from the site. The information on access patterns to the hosted content might be helpful in deciding whether it is the time to move and archive some of the current content or whether it is still actively accessed by the users.

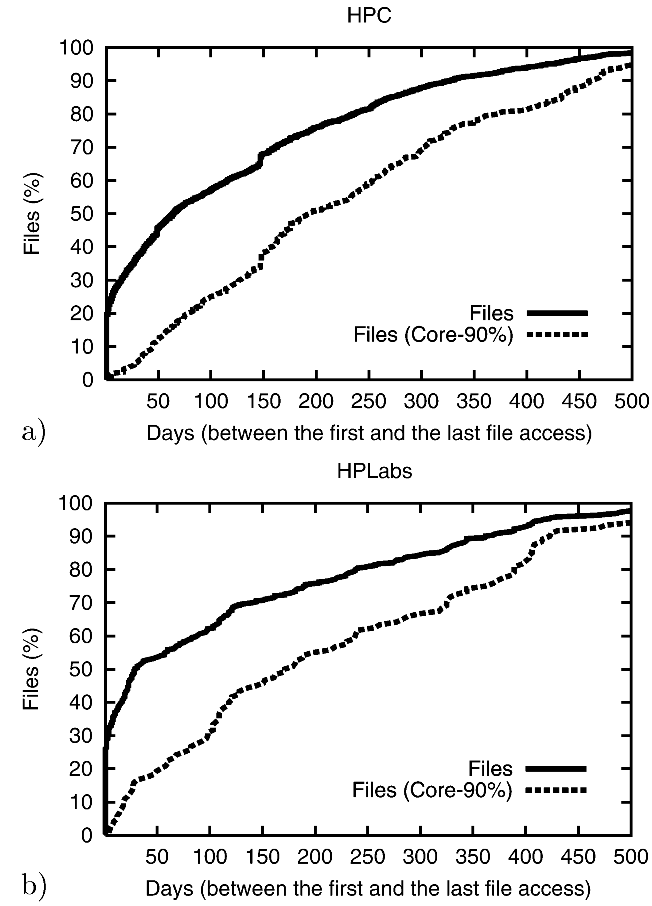

Fig. 16. Days between the first and last file accesses.

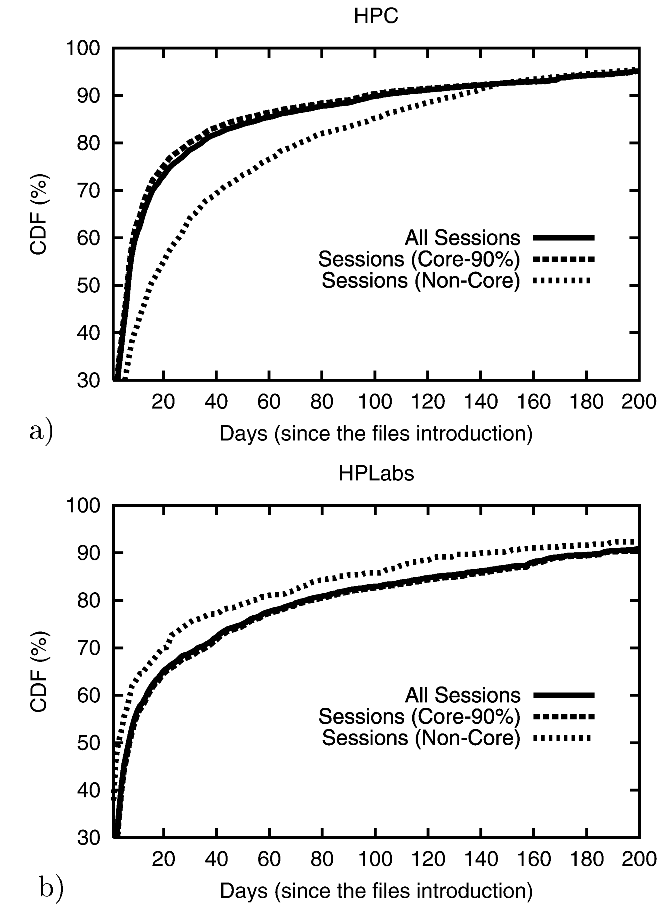

Fig. 17. Percent of sessions on days between the first and last file accesses.

file accesses change over time? The answer to this question is critical to designing prefetching or server-push algorithms, as well as for the design of efficient content distribution strategies in the CDNs for media content.

As shown in Section IV-D, the media workloads under study exhibit a high locality of references. We observed that 90% of the media server sessions target only 14%–30% of the files. Thus, a small set of files has a strong impact on the media site performance and its access patterns. We define *core 90%* as the set of most frequently accessed files that make up for 90% of all the media sessions. From the performance point of view, the attention of service providers should concentrate on efficient support for these *core* files because most of the accesses target the core files. Along with the understanding of the dynamics of all the files at the site, we would like to see whether the core files exhibit some specific properties.

We define a *life time* of a particular file to be the time between the first and the last access to this file in the given workload.

Fig. 16 shows the distribution of file *life times* for both workloads.The two curves in each of Fig. 16(a) and (b) represents a life time distribution for all the files and for the core 90% files. These graphs show that a high percentage of all the files have a short life time; files that "live" less than a month constitute 37% of all the files in the HPC workload and 50% of all the files in the HPLabs workload (this number is high in part because 16%–19% of all the files are accessed only once as reported in Section IV-D). A total of 73% of all the files for both workloads has a life time of less than six months. Only 10% of files for the HPLabs site and 8% for the HPC site live longer than a year. As for the most frequently accessed files, a much higher percentage of them live longer compared to the life time of all the files. The "short-lived" frequent files in the graphs are mostly the recently introduced files.

We define *a file life span* as the normalized distribution of file accesses (normalized with respect to the total number of file accesses) since the file's introduction at the site. This new property characterizes the file access rate over the file existence (file's life time) at the site.

In order to characterize the "change in access rate" specific to a particular subset of files at a site as well as to compare these rates for different groups of files, we introduce a *life span metric* which uses a CDF representation of the file life spans. In other words, for a particular subset of files, the life span metric is defined as a cumulative density function of accesses to these files since their introduction at a site. The introduced life span metric measures the change in access rates to the newly introduced files and reflects the timeliness of the accesses to them.

Fig. 17 shows the life span of the file's accesses for both workloads. The X axis represents the days since the introduction of the files and the Y axis represents the CDF of all the file accesses up to a particular day (relative to the total number of all the sessions over the entire duration of the logs).

For the HPC (HPLabs) workload, 52% (51%) of all the sessions occur during the first week of files' existence, 68% (61%)

of all the sessions occur during the two weeks of the files' existence, 74% (66%) during the three weeks of files' existence, 77% (69%) during the four weeks of files existence, and 80% (70%) during the five weeks of the files existence. Thus, the HPLabs site has the rates of change in file accesses lower than the HPC site.

The above statistics can be interpreted in a different way, reflecting the rates of change of accesses in a given workload: 52% (51%) of all the sessions occur during the first week of file existence, followed by only 16% (10%) of accesses during the second week, falling to 6% (5%) of accesses during the third week, and only 3% (1%) of the accesses for fourth and fifth weeks since the introduction of the files.

The life span of the core 90% files is almost identical to the life span of all the files. It is not surprising, because by definition the core 90% files are responsible for 90% of all the accesses to the site. The properties of the core 90% files have a major impact on the characteristics of the life span of the entire site. As for the rest of the files ("noncore" files), their properties are different for the HPC and the HPLabs workloads. For example, for the HPC workload, 70% of the sessions to the noncore files occur during first 42 days after the files' introduction, while for the HPLabs workload, 70% of corresponding sessions occur during the first 21 days after the introduction of the files.

The life span metric is a normalized metric. The files could have been individually introduced at different times. This metric reflects the rate of change in the file access pattern during the files' existence at the site. Moreover, the life span metric reflects the timeliness of the introduced files. A lower rate of change in file accesses reflects that the media information on a site is less timely and has a more consistent access pattern over a longer period of time. The life span metric can potentially interpolate the intensity of the client accesses to the new and the existing files for a future period of time.

We believe that the locality properties, the access patterns of newly introduced files, and their life spans are critical metrics in defining efficient caching infrastructures and content distribution strategies for CDNs.

VI. Conclusion and Future Work

Media server access logs are an invaluable source of information not only in extracting business related information but also for understanding traffic access patterns and system resource requirements of the media site. Our tool MediaMetrics is specially designed for system administrators and service providers to understand the nature of traffic at their media sites. Issues of workload analysis are crucial to properly designing a site and its support infrastructure, especially for large, busy media sites.

Our analysis aimed to establish a set of properties specific for the enterprise media server workloads and to compare them to the well-known related observations about the web server workloads. In particular, we observed a high locality of references in media file accesses for both workloads. Similar to the previous web workloads studies, our analysis of the video popularity distribution (collected over a relatively short time of 1–6 months) revealed that it can be approximated by a Zipf-like distribution with the α parameter in the range 1.4–1.6. The interesting new observation is that the time scale plays an important role in this approximation. For longer time scales in the same workloads, the file access frequency distribution does not follow a Zipfian distribution.

We propose two new metrics characterizing temporal properties of the media workloads. We introduce the new files impact metric for the media workloads and observe that in the studied workloads, the accesses to the new files constitute most of the monthly accesses as well as the bytes transferred due to the accesses to the new files account for most of the transferred bytes. Also, we observe that the trend in the growth of the site accesses directly depends on the number of the newly added files. We further define a life span metric to reflect the rates of change in accesses to the newly introduced files. For the studied workloads, 51%–52% of the accesses to the media files occur during the first week of their introduction. The access pattern of the enterprise *media sites* resembles the access pattern of the *news web sites*, where most of the client accesses target the new regularly added information. Understanding the new files impact and the life span metrics is important for efficient resource management and provisioning, especially for large busy sites.

Additionally, we also discovered some interesting facts about the client viewing behaviors. Despite the fact that the two studied workloads had significantly different file size distribution, the client viewing behavior was very similar for the both sets. We also found that the percentage of sessions with interactive requests was much higher for medium and long videos.

In our future work, we are planning to exploit the observed media workload properties for synthetic workload generation as well as in a design of the capacity planning tools for media sites.

Acknowledgment

The authors would like to thank their HP colleagues N. Lyons, W. Smallwood, B. Bausk, M. Karlsson, W. Tang, Y. Fu, J. Apostolopoulos, and S. Wee. They would also like to thank the anonymous referees for insightful questions and constructive criticism which helped in significantly improving the content and presentation of the paper.

References

[1] S. Acharya and B. Smith, "An experiment to characterize videos stored on the web," in *Proc. ACM/SPIE Multimedia Computing Networking Conf.*, Jan. 1998, pp. 166–178.

[2] S. Acharya, B. Smith, and P. Parnes, "Characterizing user access to videos on the world wide web," in *Proc. ACM/SPIE Multimedia Computing Networking Conf.*, San Jose, CA, Jan. 2000.

[3] J. M. Almeida, J. Krueger, D. L. Eager, and M. K. Vernon, "Analysis of educational media server workloads," in *Proc. 11th Int. Workshop Network Operating System Support for Digital Audio Video (NOSSDAV 2001)*, June 2001, pp. 21–30.

[4] V. Almeida, A. Bestavros, M. Crovella, and A. Oliviera, "Characterizing reference locality in the WWW," in *Proc. 4th Int. Conf. Parallel Distributed Information Systems (PFIS)*, 1996, pp. 92–106.

[5] M. Arlitt and C. Williamson, "Web server workload characterization: The search for invariants," in *Proc. ACM SIGMETRICS Conf.*, Philadelphia, PA, May 1996, pp. 126–137.

[6] P. Barford, A. Bestavros, A. Bradley, and M. Crovella, "Changes in web client access patterns: Characteristics and caching implications," Boston Univ., Boston, MA, Paper TR-1998-023, 1998.

[7] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *Proc. IEEE INFOCOM*, Mar. 1999, pp. 126–134.

[8] L. Cherkasova and M. Gupta, "Analysis of enterprise media server workloads: Access patterns, locality, dynamics, and rate of change," HP Lab. Rep., Palo Alto, CA, HPL-2002-56, 2002.

[9] L. Cherkasova and M. Karlsson, "Dynamics and evolution of web sites: Analysis, metrics and design issues," in *Proc. 6th Int. Symp. Computers Communications (ISCC'01)*, Hammamet, Tunisia, July 2001, pp. 64–71.

[10] M. Chesire, A. Wolman, G. M. Voelker, and H. M. Levy, "Measurement and analysis of a streaming media workload," in *Proc. 3rd USENIX Symp. Internet Technologies and Systems*, San Francisco, CA, Mar. 2001, pp. 1–12.

[11] F. Douglis, A. Feldmann, B. Krishnamurthy, and J. Mogul, "Rate of change and other metrics: A live study of the world wide web," in *Proc. USENIX Symp. Internet Technologies and Systems*, Dec. 1997, pp. 147–158.

[12] Content Networking. Inktomi Corp. [Online]. Available: http://www.inktomi.com

[13] L. He, J. Grudin, and A. Gupta, "Designing presentations for on-demand viewing," in *Proc. ACM 2000 Conf. Computer Supported Cooperative Work*, Philadelphia, PA, Dec. 2000, pp. 127–134.

[14] N. Harel, V. Vellanki, A. Chervenak, G. Abowd, and U. Ramachandran, "Workload of a media-enhanced classroom server," in *Proc. 2nd Annu. IEEE Workshop on Workload Characterization*, Austin, TX, Oct. 1999.

[15] D. Loguinov and H. Radha, "Measurement study of low-bitrate internet video streaming," in *Proc. ACM SIGCOMM Internet Measurement Workshop*, San Francisco, CA, Nov. 2001.

[16] A. Mena and J. Heidemann, "An empirical study of real audio traffic," in *Proc. IEEE INFOCOM*, Tel-Aviv, Israel, Mar. 2000, pp. 101–110.

[17] J. Padhye and J. Kurose, "An empirical study of client interactions with continuous-media courseware server," in *Proc. 8th Int. Workshop Network Operating System Support for Digital Audio Video (NOSSDAV)*, July 1998.

[18] (1998) Realserver Administration Guide—Realsyztem G2. RealNetworks, Inc. [Online]. Available: http://docs.real.com/docs/serveradminguideg2.pdf

[19] W. Tang, Y. Fu, L. Cherkasova, and A. Vahdat, "Medisyn: A synthetic streaming media service workload generator," in *Proc. 13th Int. Workshop Network Operating System Support for Digital Audio Video (ACM NOSSDAV)*, Monterey, CA, June 2003, pp. 12–21.

[20] S. Wenger, "H.264/AVC over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 645–656, July 2003.

[21] Y. Wang and Q. Zhu, "Error control and concealment for video communications: A review," *Proc. IEEE*, vol. 86, pp. 974–997, May 1998.

[22] Wang, M. Claypool, and Z. Zuo, "An empirical study of realvideo performance across the internet," in *Proc. ACM SIGCOMM Internet Measurement Workshop*, San Francisco, CA, Nov. 2001, pp. 295–309.

[23] Windows Media Services SDK Version 4.1. [Online]. Available: http://msdn.microsoft.com/workshop/imedia/windowsmedia/sdk/wmsdk.asp

Ludmila Cherkasova (M'92) is a Senior Scientist in the Internet and Computing Platforms Research Center at HPLabs, Palo Alto, CA. She joined Hewlett-Packard Laboratories in 1991. Before joining HPLabs, she was a Senior Researcher at the Institute of Computing Systems, Russia, and an Adjunct Associate Professor at Novosibirsk State University, Russia. Her current research interests include distributed systems, internet technologies and networking, performance measurement and monitoring, and performance analysis.

Dr. Cherkasova received a Certificate of Appreciation Award from IEEE Computer Society in 2002.

Minaxi Gupta received the B.Sc. degree from Bombay University, Bombay, India, and the M.Sc. degree from IIT Bombay, both in physics, and the M.S. degree in computer science from the Georgia Institute of Technology, Atlanta. She is currently working toward the Ph.D. degree at the Georgia Institute of Technology.

Her research interests include networking and distributed systems.