Workload Study of a Media-Rich Educational Web Site

Yang Liu
Department of Computer Science
University of Calgary
Calgary, AB, Canada
liu15@ucalgary.ca

Carey Williamson
Department of Computer Science
University of Calgary
Calgary, AB, Canada
carey@cpsc.ucalgary.ca

ABSTRACT

Modern educational Web sites often feature a rich assortment of linked media content. In this paper, we present a workload study of such an educational Web site hosted at the University of Calgary. Three main insights emerge from our study. First, educational Web sites can generate large volumes of Internet traffic, even when the number of users is limited. Second, network usage is highly influenced by course-related events, such as midterms and finals. Third, the approach used by the site for displaying videos can have adverse impacts on user experience and network traffic. We demonstrate these effects with active measurement of different Web browser and video player implementations.

Categories and Subject Descriptors

C.2.0 [**Computer Systems Organization**]: Computer-Communication Networks—*General*; J.7 [**Computer Applications**]: Computers in Other Systems

General Terms

Measurement, Performance

Keywords

Workload Characterization; Educational Web Site; Technology Enhanced Learning; Media; Video; HTTP

1. INTRODUCTION

A recent trend in academia is toward open educational resources. Publicly-accessible educational resources facilitate the development of world-wide on-line education through the sharing of scientific knowledge. For example, the worldwide OpenCourseWare (OCW)[1] site offers free on-line courses from noted universities, including Massachusetts Institute of Technology (MIT) and Yale University. Furthermore, these educational Web sites may have a large effect on network usage. For OpenCourseWare, a report from MIT[2] shows that MIT OCW was visited 2,385,654 times by 1,367,228 unique visitors in April 2015.

The University of Calgary (U of C) hosts multiple Web sites that share educational resources as well. After studying all incoming and outgoing network traffic at our university, we found that one particular astrophysics Web site (ISM)[3], generated a lot of network traffic. This site is hosted by the Department of Physics and Astronomy at the U of C.

The ISM site studies the Inter-Stellar Medium (i.e., the gas and dust in between the stars) in astrophysics. The site is created and maintained by a U of C professor. Apart from a brief introduction about the Inter-Stellar Medium and some corresponding research, the ISM site mainly provides educational materials with linked rich media content for three courses, including one Astronomy course (ASTR 209) and two Astrophysics courses (ASPH 213, ASPH 503). Among the courses, ASTR 209 and ASPH 213 were offered in Winter 2015 semester, with 400 U of C students registered in ASTR 209. During our four-month observation period (January 1, 2015 to April 29, 2015), the ISM site generated an average of 70 GB of data traffic volume every day. This volume was surprising, given the relatively small user community of 400 students.

The primary motivation for our study is a desire to understand how students use educational Web sites. Our goals are to measure the network traffic, characterize usage patterns, and identify performance issues.

There are three main insights that emerge from our study. First, the network traffic generated by this site is surprisingly large (about 70 GB per day), given the relatively small user community. Second, network usage is highly influenced by course-related events, such as midterms and final exams. Third, the approach adopted by the site for displaying the lecture videos has undesirable effects on user experience and network traffic. We close our paper with an active measurement study demonstrating better ways to share educational videos across the Internet.

2. RELATED WORK

Earlier work on Web workload characterization focused on Web client characterization [7], Web servers [2], and Web proxies [11]. Arlitt *et al.* [2] analyzed Web server workloads

[1]Wikipedia, OpenCourseWare, `https://en.wikipedia.org/wiki/OpenCourseWare`

[2]`http://ocw.mit.edu/about/site-statistics/monthly-reports/MITOCW_DB_2015_04.pdf`

[3]Star Formation & Molecular Astrophysics at the U of C, `http://ism.ucalgary.ca/`

WWW'16 Companion, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2890461.

and identified ten common properties. Crovella *et al.* [6] identified self-similarity in World Wide Web (WWW) traffic. Sedayao *et al.* [13] analyzed fundamental properties of WWW traffic patterns.

Recent Web workload studies focused on modern Web traffic [9], such as the traffic of Web 2.0 sites. Callahan *et al.* [4] analyzed Web workload evolution from a longitudinal perspective. Butkiewicz *et al.* [3] analyzed how complex Web pages are. Lin *et al.* [10] studied the on-line map application traffic on Web 2.0 sites. Cha *et al.* [5] studied the traffic of several user-generated content video Web sites. Gill *et al.* [8] studied the workload characteristics of YouTube.

There is limited literature about workload characterization of educational Web sites, and most studies of educational Web sites focus on pedagogy or psychology. Sheard *et al.* [14] found that inferring students' learning behaviors by analyzing their educational Web site usage patterns is feasible. Almeida *et al.* [1] analyzed educational media server workloads at two universities in 2000. They studied workload characteristics such as session arrival rate and session inter-arrival time distributions. They also found that traditional caching strategies do not work for streaming media.

3. METHODOLOGY

The dataset for our study was collected using passive network traffic measurement. At the U of C, the edge routers on the campus backbone connect the campus network to the Internet. We used a traffic monitor (Dell, 2 Intel Xeon E5-2690 CPUs, 64 GB RAM, 5.5 TB storage, CentOS 6.6 x64, Endace DAG 8.1SX card) to collect a mirror of all packet-level traffic flowing between the campus and the Internet. To process the traffic flows, we used Bro [12], configured to generate logs hourly. Since the ISM site is an HTTP server, we focus solely on the HTTP traffic. The Bro logging system can record detailed information of HTTP request-response headers, such as IP address and user agent.

We analyzed the traffic logs for a four-month period from January 1, 2015 to April 29, 2015, covering the Winter 2015 semester at U of C. In this semester[4], lectures began on January 12, and ended on April 15, with final exams from April 18 to 29. There was a reading week with no lectures from February 15 to 22. There was an outage of the logging system on April 11, with no logs recorded on that day.

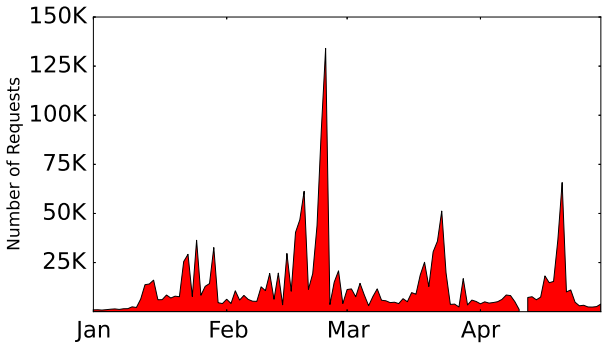
Due to the placement of our network monitor, we can only observe the ISM Web traffic generated by off-campus users. The on-campus traffic does not traverse the edge routers, and therefore is not seen. Analyzing the server-side logs of the ISM site for a complete view remains as future work.

4. WORKLOAD CHARACTERISTICS

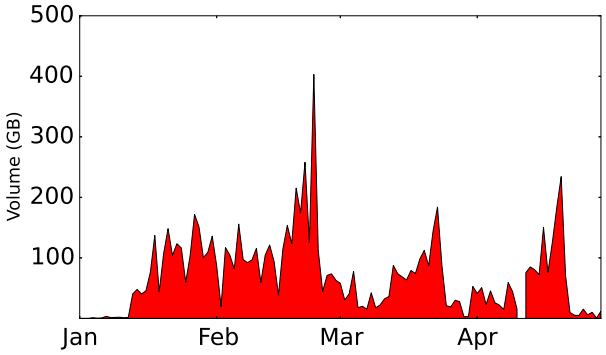
The ISM site is hosted and maintained by a U of C professor. The professor posts his research information and lecture materials on this site. Preliminary analysis of its network traffic indicates that most of the requests are generated for obtaining educational resources of the three courses, namely ASTR 209, ASPH 213, and ASPH 503. Most of the traffic is for ASTR 209, including lecture videos for the 400 U of C students registered in the course.

4.1 ISM Site Overview

[4]Academic Schedule 2014-2015, `http://www.ucalgary.ca/pubs/calendar/archives/2014/academic-schedule.html`



(a) HTTP Request Count Per Day



(b) Data Volume (GB) Per Day

Figure 1: HTTP Requests and Data Volume Per Day

A summary of the ISM site traffic is shown in Table 1. There are around 1.5 million requests in total during the semester, with an average of 13,000 requests per day. The daily ISM site traffic is illustrated in Figure 1. Note there are three obvious surges in the traffic over the four months, related with students' studying patterns. The surge in late-February aligns with the first midterm in ASTR 209 (February 24), while the subsequent surges align with the second midterm (March 24) and the final exam (April 21).

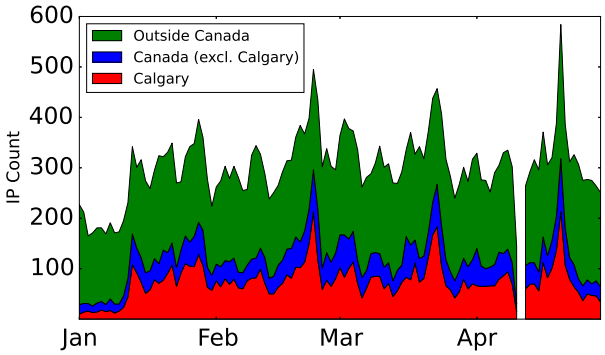
4.2 IP Analysis

During the four months of observation, 9,720 unique IP addresses visited the ISM site. Figure 2 shows the daily count of unique IP addresses. The amplitude of the surges is comparatively smaller than the request traffic and data volume in Figure 1, because the primary users are the students in the course, who are frequent repeat visitors.

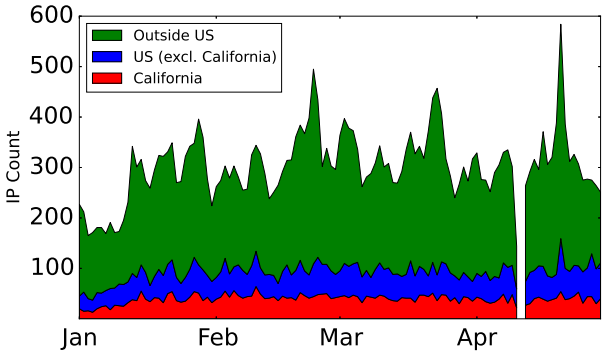
The geolocation analysis for all the IPs visiting the ISM site shows that visitors were from 101 different countries, though about half of those countries (55) generated fewer than 100 requests in four months. Most of the requests come from Canada (88.24%) and the USA (7.91%). Within Canada, Alberta surpasses all other provinces with 1.2 million requests (97.64%). Furthermore, 1.1 million requests (93.07%) came from Calgary, dominating all other cities in Alberta. The USA traffic distribution is more dispersed, with California accounting for 39,627 requests (32.85%). Many of the USA requests are generated by Google and Apple, for indexing Web content. Figure 2 shows how many unique

Table 1: Statistical Characteristics of the ISM Site (Jan 1/15 to Apr 29/15)

Total Reqs	Avg Reqs/day	Total GB	Avg GB/day	Uniq URLs	Uniq IPs
1,583,339	13,305	8,483	71.29	10,563	9,720



(a) From Calgary, Canada, and Elsewhere



(b) From California, USA, and Elsewhere

Figure 2: Number of Daily Unique IP Addresses

IPs visited the ISM site each day from Calgary, Canada, California, and USA.

4.3 HTTP Requests and Responses

The HTTP request-response statistical results are shown in Table 2. About 99.5% of the HTTP requests use the GET method, while 0.5% are HEAD requests. Furthermore, 7,285 HEAD requests (94.01%) were generated by Apple's iTunes application to check the existence of some resources or whether the RSS (Rich Site Summary) file on the ISM site was updated.

For HTTP response status codes, "206 Partial Content" is the topmost one accounting for around 60% of the requests, while "200 OK" is second at 32%. This result is quite different from most Web sites, where "200 OK" responses dominate [2]. This situation is primarily caused by students frequently requesting pieces of large-size files (e.g., videos), and by user agent behaviors.

4.4 URL Analysis

There are 10,563 different URLs on the ISM site requested in the four-month observation period. The most popular requested URL is "`ASTR209 - Lec8 - Feb 5, 2015.mov`" with 153,410 requests and 267 GB of data volume. Course materials (especially linked media content) are popular among

Table 2: HTTP Method and HTTP Status Code

(a) HTTP Method

HTTP Method	Reqs	Pct.
GET	1,575,574	99.51%
HEAD	7,749	0.49%
OPTIONS	11	0.00%
POST	5	0.00%

(b) HTTP Status Code

Status Code	Reqs	Pct.
206 Partial Content	927,733	58.59%
200 OK	507,358	32.04%
304 Not Modified	79,064	4.99%
404 Not Found	47,372	2.99%

the URLs. In addition, the large-size videos and PDF files generate voluminous data traffic.

There are 46 different file types observed in our trace. Table 3 shows that video file type `Video/QuickTime` accounts for the most requests (29.78%) and data volume (60.35%), while `Video/MP4` accounts for 10.26% of the requests and 36.06% of the data volume. Static HTML files are popular in terms of requests, but have minimal contribution to data volume. Course materials (e.g., homework and slides) are primarily provided in PDF format. As expected, these files contribute to a lot of the network traffic.

4.5 User Agent

We use the on-line user agent database provided by "User Agent String.Com"[5] to identify operating system and user agent information for our trace. The most popular user agent is "AppleCoreMedia" with 701,507 requests (44.3%). This agent is implemented in Apple's products (e.g., iPad, iPhone, and Mac) for dealing with on-line video files. Firefox (18.6%), Chrome (14.8%), Safari (10.9%), and Internet Explorer (3.3%) are the next most popular user agents. Among user agents labeled "crawler", we found that "Googlebot" from Google accounts for about half of the traffic (15,734 requests, 49.46%), and "Bingbot" from Microsoft ranks second with 8,031 (25.25%) requests.

For operating systems, Apple's products (59.0%) dominate, with Microsoft Windows (33.0%) second. To be specific, iPhone OS (iOS) accounts for 41.8% of the total requests, and OS X represents 17.2%. For Windows, 14.8% of the total requests were generated by Windows 7, 13.5% by Windows NT, and 3.0% by Windows 8. Android only accounts for 1.88% of the total requests.

We list the top 5 popular versions of these operating systems in Table 4, and the top 4 browsers in Table 5. These results suggest a technology-savvy set of students, with very recent OS and browser versions.

[5]User Agent String.Com, `http://www.useragentstring.com/`

Table 3: Top 5 Most Frequently Requested File Types

File Type	Rank	Total Reqs	Pct.	Rank	Total GB	Pct.
`Video/QuickTime`	1	532,883	29.78%	1	5,159	60.35%
`Application/PDF`	2	250,244	13.99%	3	284	3.33%
`Video/MP4`	3	183,636	10.26%	2	3,082	36.06%
`Text/HTML`	4	177,506	9.92%	6	3	0.03%
`Image/PNG`	5	144,361	8.07%	5	4	0.05%

Table 4: Top 5 OS Versions

iPhone OS		OS X	
Version	Pct.	Version	Pct.
8.1.3	25.67%	10.6.8	17.45%
7.1.1	21.63%	10.10.2	17.31%
8.1.1	13.48%	10.9.5	11.40%
8.0.2	10.24%	10.10.1	11.28%
7.0.2	9.78%	10.8.3	7.34%
Windows		Android	
Version	Pct.	Version	Pct.
Win 7	44.94%	4.4.2	49.66%
Win NT	40.87%	4.4.4	19.04%
Win 8	9.24%	5.0.1	7.72%
Win Vista	2.82%	5.0.2	3.14%
Win XP	1.29%	4.2.1	2.65%

Table 5: Top 5 Versions of Popular Browsers

Firefox		Chrome	
Version	Pct.	Version	Pct.
35.0	40.17%	40.0.2214.115	15.97%
36.0	29.20%	40.0.2214.111	12.01%
37.0	18.32%	42.0.2311.90	9.60%
34.0	4.54%	41.0.2272.118	9.29%
33.0	2.12%	41.0.2272.101	8.41%
Safari		Internet Explorer	
Version	Pct.	Version	Pct.
8.0	28.76%	11.0	61.88%
8.0.3	12.14%	10.0	18.93%
7.0	9.30%	7.0	9.22%
8.0.2	9.27%	8.0	5.01%
8.0.4	6.53%	9.0	2.70%

4.6 Course-Related Events

The ISM site traffic is heavily influenced by course-related events. As mentioned earlier, the surges in traffic (Figure 1) and unique IPs (Figure 2) are mainly caused by the scheduled exams of the course ASTR 209. Therefore, we take a closer look at the network usage related with the course events in this subsection.

By identifying the names of the requested URLs, we find that ASTR 209 accounts for 77.8% (1,231,339) of requests and 99.4% (8,434 GB) of data volume, while ASPH 213 (120,351 reqs, 46 GB) and ASPH 503 (2,613 reqs, 0.2 GB) generate minor traffic. The course ASPH 503 was not offered in Winter 2015, while ASTR 209 and ASPH 213 were both available in Winter 2015. There are some HTTP requests retrieving files of ASPH 213, however, the ASTR 209 traffic dominates in both requests and data volume.

The course materials in the ISM server are organized with meaningful names. For example, "`AST209/Entries/2015/1/28_Course_Notes_files/Part2_e&m.pdf`" is a course note file for ASTR 209, and "`AST209_Midterm1_info_files/FormulasheetMidterm1.pdf`" is midterm reviewing material for ASTR 209. Therefore, by extracting information from the requested URLs, we classify the course-related requests into 6 categories: "Video", "Course Notes", "Midterm" (midterm exam materials), "Outline" (course outline), "Homework", and "Final" (final exam materials).

Figure 3 shows the daily HTTP requests and data volume traffic for the 6 categories. Since the values in "Video" and "Course Notes" are much larger than the other four, we use two rows of figures with different y-axis scales to clearly display the trend of each category. Our observations are:

1) Outline and homework materials are popular near the beginning of the course. Students use these to acquire a general understanding of the learning outcomes for the course.

2) Videos (linked media content) account for most of the requests and data volume for the ISM site. Students rely more heavily on the videos for the first midterm exam than they do for the second midterm or the final exam.

3) Course notes are the primary materials for students to study for the midterms and the final. For the second midterm and the final, course notes receive as many requests as the videos do.

4) The popularity of midterm exam materials increases dramatically before the midterms and the final, indicating that students' reviewing period is relatively short. The surge of final exam materials before the final exam leads to a similar conclusion.

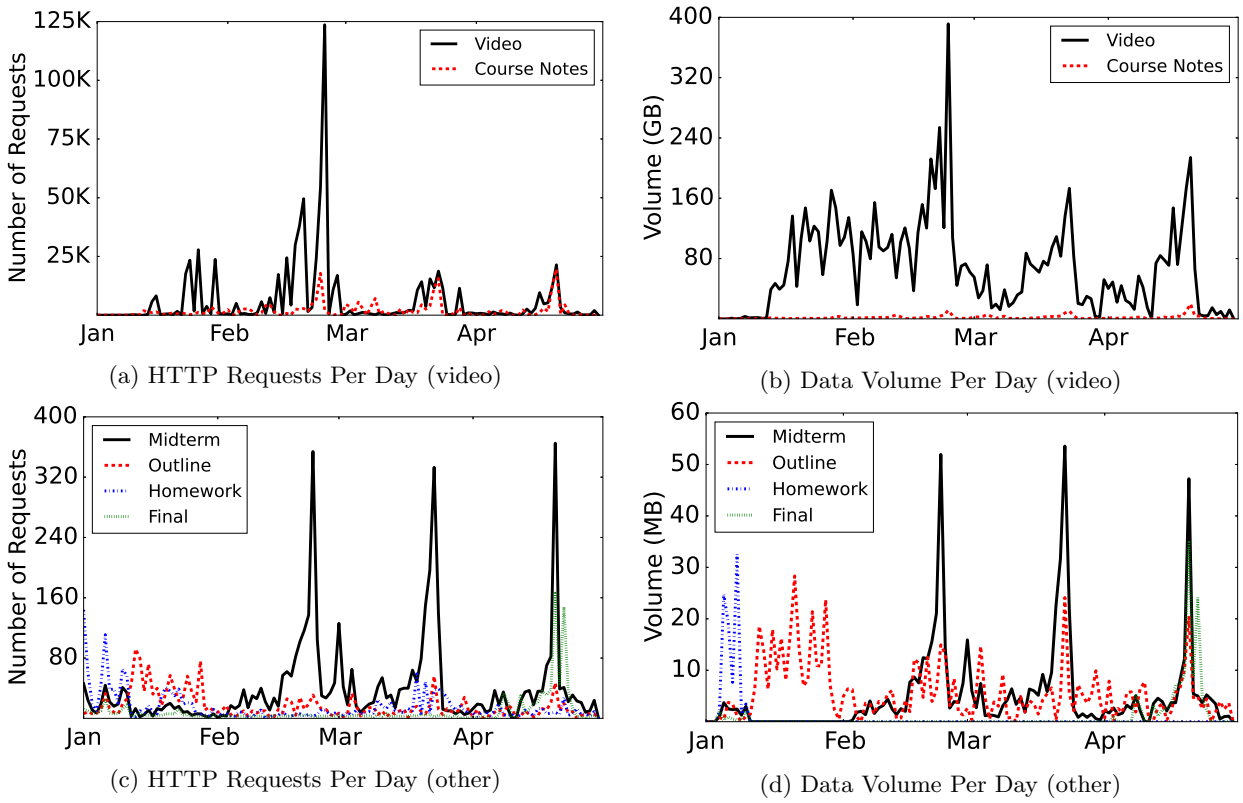
These network traffic trends align with the course events, which in turn indicates that real-world events heavily influence the Web usage.

4.7 Video-Related Traffic Analysis

Figure 3 shows the daily video-related requests and data volume. Clearly, most requests and data volume are caused by video requests. In fact, video requests triggered 716,519 requests (45.3%) and 8,241 GB of data volume (97.1%).

We analyzed the HTTP transaction duration and response size values of all the video requests. For all video requests, 98.1% of HTTP transaction duration values are shorter than 10 seconds, and 94.6% of response size values are smaller than 5 MB. In other words, short HTTP transaction durations and small response sizes dominate the video HTTP transactions, from the prevalence of HTTP partial content request-responses.

Further analysis indicates that the dominant response sizes are 64 KB (35.4% of requests), 128 KB (12.1%), and 256 KB (5.2%). This phenomenon is caused by user agents when fetching large (video) files from a server that supports partial GET requests.



(a) HTTP Requests Per Day (video)

(b) Data Volume Per Day (video)

(c) HTTP Requests Per Day (other)

(d) Data Volume Per Day (other)

Figure 3: HTTP Requests and Data Volume Per Day for the 6 Categories

5. WEB BROWSER EXPERIMENTS

In this section, we use active measurements to evaluate the effects of video displaying implementations. By analyzing the HTML source code of the ISM site, we find that it uses the progressive download technique, with which users can playback or fast forward video. However, they cannot view video that is not yet downloaded by the browser. This implementation of video streaming in the ISM site is inconvenient for users, and inefficient for the network. We further explore this issue using browser comparison experiments.

We deployed an Apache HTTP server on a PC, with "Accept-Ranges: bytes" enabled by default. The configuration of our server is essentially the same as the ISM server. We limit client bandwidth[6] to 10 Mbps, using the Apache "mod_ratelimit" module[7] configurations. The Web server and clients all run in the same PC (localhost), thus network issues are eliminated. One lecture video ("ASTR209 - Lec4 - Jan 22, 2015.mp4") is downloaded from the ISM site as a sample to deploy on our server. We experiment with four server-side video playing implementations, tested with the latest versions of Firefox, Chrome, Safari, and Internet Explorer:

Case 1) The video file is served as a static file in the server. This is the simplest way for delivering video files.

Case 2) The video file is embedded as an HTML "<object>" element, with its attribute "type" set to `Video/QuickTime`. This is implemented exactly the same as the ISM site.

Case 3) The video is displayed by the HTML5 "<video>" tag. This approach is a standard way to embed a video in a Web page, but was not feasible before HTML5.

Case 4) The video is displayed by MPEG-DASH implementation with Dash.js support. This approach needs to process the video and generate the MPD file beforehand. Dash.js requires Media Source Extensions (MSE) support in the browsers.

The results are shown in Table 6. The browser names and versions are listed in the leftmost column. We use "Static File", "Object Element", "HTML5 Video Tag", and "MPEG-DASH" to represent the four implementations. The column "Play" shows whether the video is able to be played in that condition, and "Forward" shows whether the video can be forwarded to any point (versus the user having to watch from the start and wait for the video to be downloaded).

Table 6 shows that the static file approach supports all the browsers except IE, since IE downloads the video file by default instead of invoking its internal video player. The HTML object element implementation used by the ISM site works for only two of the browsers. Furthermore, the browser uses the QuickTime plug-in video player to decode the video file in the object element approach, which doesn't support fast forward. The HTML5 video tag implementation is the only approach fully supporting the fast forward function in all the browsers.

There are four current commercial implementations of adaptive streaming, including Dynamic Adaptive Stream-

[6]List of countries by Internet connection speeds, `https://en.wikipedia.org/wiki/List_of_countries_by_Internet_connection_speeds`

[7]Apache Module mod_ratelimit, `http://httpd.apache.org/docs/2.4/mod/mod_ratelimit.html`

Table 6: Browser Support for the Four Video Playing Implementations

Browser	Static File		Object Element		HTML5 Video Tag		MPEG-DASH	
	Play	Forward	Play	Forward	Play	Forward	Play	Forward
Chrome (V44)	Yes	Yes	**No**	N/A	Yes	Yes	Yes	Yes
Safari (V8)	Yes	Yes	Yes	**No**	Yes	Yes	Yes	Yes
Firefox (V39)	Yes	Yes	Yes	**No**	Yes	Yes	**No**	N/A
IE (V11)	**No**	N/A	**No**	N/A	Yes	Yes	Yes	Yes

ing over HTTP (MPEG-DASH), Adobe Dynamic Streaming for Flash, Apple HTTP Adaptive Streaming, and Microsoft Smooth Streaming. MPEG-DASH is the only international standard widely supported by most HTTP servers. Safari supports this since "V8", IE since "V11", and Firefox only has partial support. The advantage of DASH is to provide best quality videos based on user's network speed. The disadvantage is the computation and storage resources used for compressing the videos in various bit-rates beforehand.

Analysis of the server-side Apache logs from our experiments shows that:

1) Chrome first generates a GET request for the video file. Then it generates a GET request with "bytes=0-" to test whether partial GET is supported. When the user clicks in the progress bar where the video is not already downloaded, Chrome aborts the previous GET request and generates a new partial request. IE behaves the same as Chrome does.

2) Firefox and Safari generate a GET request for the whole video file at first like Chrome. Then they both generate a series of partial GET requests with small-size responses, when the user forwards the video. However, the number of requests varies, and the range values are not always contiguous or monotonic. The popular response size values in Section 4.7 are primarily caused by Firefox and Safari.

In summary, the video streaming approach used by the ISM site is inconvenient for viewers, and inefficient for the network. In today's Internet, backward-compatible HTML5 approaches are suggested for improving the video viewing experience on educational Web sites.

6. SUMMARY AND CONCLUSIONS

We studied the workload of an educational Web site over a four-month period of observation. We analyzed the HTTP traffic thoroughly with passive measurements, and performed an active measurement study of video displaying approaches. Our conclusions are presented as follows:

1) Educational Web sites can generate large volumes of Internet traffic, even with a small number of users.

2) Network usage is influenced by course-related events.

3) The approach adopted by an educational Web site for displaying videos affects user experience and network traffic. Both can be improved using an HTML5 implementation.

7. ACKNOWLEDGMENTS

Financial support for this work was provided by NSERC and the University of Calgary. The authors thank the ASTR 209 students and instructor for generating lots of interesting network traffic.

8. REFERENCES

[1] J. Almeida, J. Krueger, D. Eager, and M. Vernon. Analysis of Educational Media Server Workloads. In *Proceedings of the ACM NOSSDAV*, pages 21–30. ACM, 2001.

[2] M. Arlitt and C. Williamson. Internet Web Servers: Workload Characterization and Performance Implications. *IEEE/ACM Transactions on Networking (ToN)*, 5(5):631–645, October 1997.

[3] M. Butkiewicz, H. Madhyastha, and V. Sekar. Understanding Website Complexity: Measurements, Metrics, and Implications. In *Proceedings of ACM IMC*, pages 313–328, Berlin, Germany, November 2011.

[4] T. Callahan, M. Allman, and V. Paxson. A Longitudinal View of HTTP Traffic. In *Proceedings of PAM*, pages 222–231, Zurich, Switzerland, April 2010.

[5] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. Analyzing the Video Popularity Characteristics of Large-scale User Generated Content Systems. *IEEE/ACM Transactions on Networking (TON)*, 17(5):1357–1370, October 2009.

[6] M. Crovella and A. Bestavros. Self-similarity in World Wide Web Traffic: Evidence and Possible Causes. In *Proceedings of the ACM SIGMETRICS*, pages 160–169, Philadelphia, PA, USA, May 1996.

[7] C. Cunha, A. Bestavros, and M. Crovella. Characteristics of WWW Client-based Traces. Technical report, BUCS-1995-010, Computer Science Department, Boston University, July 1995.

[8] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. YouTube Traffic Characterization: A View from the Edge. In *Proceedings of the ACM IMC*, pages 15–28, San Diego, California, USA, October 2007.

[9] S. Ihm and V. Pai. Towards Understanding Modern Web Traffic. In *Proceedings of the ACM IMC*, pages 295–312, Berlin, Germany, November 2011.

[10] S. Lin, Z. Gao, and K. Xu. Web 2.0 Traffic Measurement: Analysis on Online Map Applications. In *Proceedings of the ACM NOSSDAV*, pages 7–12, Williamsburg, Virginia, USA, June 2009.

[11] A. Mahanti, C. Williamson, and D. Eager. Traffic Analysis of a Web Proxy Caching Hierarchy. *IEEE, Network*, 14(3):16–23, May 2000.

[12] V. Paxson. Bro: A System for Detecting Network Intruders in Real-time. *Computer networks*, 31(23):2435–2463, December 1999.

[13] J. Sedayao. World Wide Web Network Traffic Patterns. In *Compcon'95.'Technologies for the Information Superhighway', Digest of Papers.*, pages 8–12, San Francisco, California, USA, March 1995.

[14] J. Sheard, J. Ceddia, J. Hurst, and J. Tuovinen. Inferring Student Learning Behaviour from Website Interactions: A Usage Analysis. *Education and Information Technologies*, 8(3):245–266, 2003.