Video Lecture 1: Video Traffic Characterisation

7COM1030 – Multicast and Multimedia Networking

Dr. Xianhui Che (Cherry)
x.che@herts.ac.uk
School of Computer Science
University of Hertfordshire, UK

Topics

- Web Traffic Capture
- Case Study: YouTube Video Characteristics

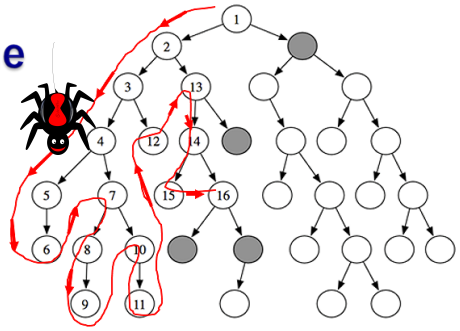
Network Traffic Capture

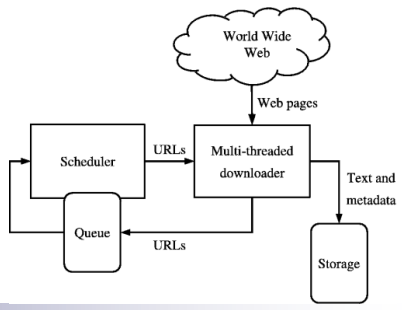
- Depending on the purpose of the traffic characterisation, there are a range of tools that can be used to capture network traffic, e.g.
  - OPNET ACE
  - WireShark
  - Customised web crawling tool
- The first two are more suitable for local or regional traffic capture. For traffic across world wide Internet domain, a common practice is to develop a customised web crawling tool.
- A web crawler (also known as a web spider or web robot) is a program or automated script which browses the World Wide Web in a methodical, automated manner.
- Legitimacy and ethical issues must be considered prior to traffic capture taking place.

Web Crawling Technique

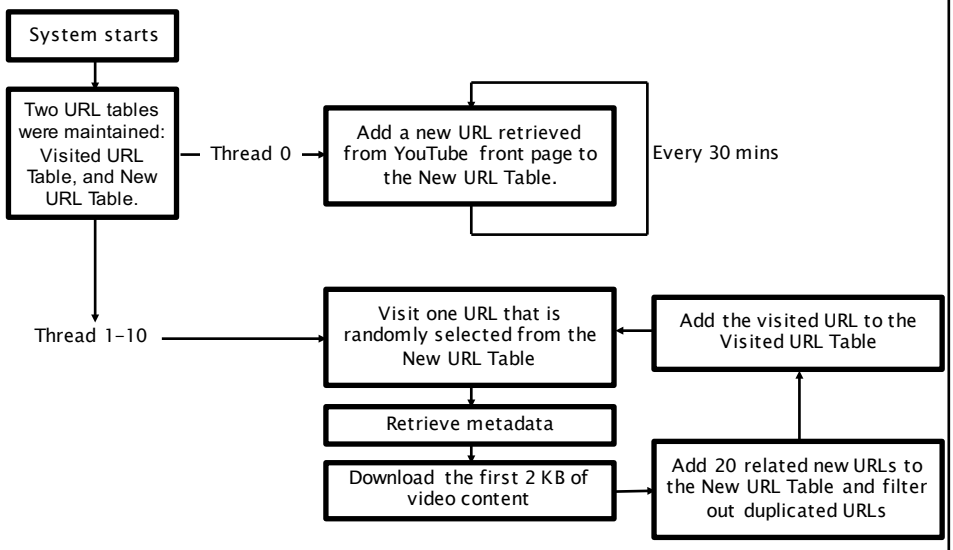
Take YouTube for instance:

1. The YouTube content item (that is, the video webpage) is linked to other content items that have similar titles, descriptions, or tags, chosen by the uploader.
2. A YouTube content item might have hundreds of YouTube content-related links, although the YouTube webpage only shows the top 20 related links at any given time.
3. Consequently, the relationship between YouTube videos can be considered as a directed spider web graph, where each video is a node on the graph and videos are linked to each other via the top 20 related links.
4. The web spider followed the recursive links among YouTube videos and captured a video dataset using a breadth-first search technique.





Flowchart for the Web Spider



HTTP Header for the First HTTP Transaction

- The first transaction acquires the YouTube metadata video information. YouTube only lets users view videos online; the videos can't be downloaded. Thus, the actual links for HTTP video streaming are encoded in YouTube's HTML and JavaScript pages and are updated periodically.

```
GET /watch?v=HBWncNcS5Mk HTTP/1.1
Referer: http://www.youtube.com/
Accept-Encoding: gzip, deflate
Accept: */*
Accept-Language: en-gb
User-Agent: Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.0;
WOW64; Trident/4.0; SLCC1; .NET CLR 2.0.50727; .NET CLR 3.5.30729;
.NET CLR 3.0.30729; .NET4.0C; .NET4.0E)
Host: www.youtube.com
Connection: Keep-Alive
Cookie: use_hitbox=72c46ff6cbcdb7c5585c36411b6b334edAEAAAAw;
VISITOR_INFO1_LIVE=c78fT-HK8a8; PREF=f1=50000000;
GEO=5221192b9aef326a6850871094454c5ccwsAAAAzR0I+aaM4Td4eNA==

HTTP/1.1 200 OK
Date: Thu, 26 May 2011 09:32:38 GMT
Server: wiseguy/0.6.10
X-Content-Type-Options: nosniff
Content-Encoding: gzip
Set-Cookie:
recently_watched_video_id_list=66287c502f23004450acaa2591fce586WwEA
AABzCwAAAEhCV25jTmNTNU1r; path=/; domain=.youtube.com
Set-Cookie: GEO=0aac78d7808c2f0d3fe34d4314cb37facwsAAAAzR0I
+aaM4Td4eNQ==; path=/; domain=.youtube.com
Expires: Tue, 27 Apr 1971 19:44:06 EST
Cache-Control: no-cache
Content-Type: text/html; charset=utf-8
Content-Length: 22256
```

HTTP Header for the Second HTTP Transaction

- The second HTTP transaction is responsible for downloading actual video content. Only the first 2 Kbytes of YouTube content is downloaded and stored. The length of the video is a critical parameter for content analysis, which only exists in the header of the HTTP download stream. Moreover, the audio and video encoding schema only exists within video content.

```
GET /videoplayback?sparams=id%2Cexpire%2Cip%2Cipbits%2Citag
%2Calgorithm%2Cburst%2Cfactor%2Coc
%3AUOhPS1RNU19FskNOOV9MR1dH&fexp=908607&algorithm=throttle-
factor&itag=34&ipbits=0&burst=40&sver=3&signature=CE548F12298389FC2
66DB38EA5DD14DC41ED1E9A.2DEE9BBAAE000F7A86D87283531475C5A3EA890EA&ex
pire=1306425600&key=yt1&ip=0.0.0.0&factor=1.25&id=1c15a770d712e4c9
HTTP/1.1
Referer: http://www.youtube.com/
Accept-Encoding: gzip, deflate
Accept: */*
Accept-Language: en-gb
User-Agent: Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.0;
WOW64; Trident/4.0; SLCC1; .NET CLR 2.0.50727; .NET CLR 3.5.30729;
.NET CLR 3.0.30729; .NET4.0C; .NET4.0E)
Host: v10.lscache2.c.youtube.com
Connection: Keep-Alive
Cookie: use_hitbox=72c46ff6cbcdb7c5585c36411b6b334edAEAAAAw;
VISITOR_INFO1_LIVE=c78fT-HK8a8; PREF=f1=50000000;
GEO=0aac78d7808c2f0d3fe34d4314cb37facwsAAAAzR0I+aaM4Td4eNQ==

HTTP/1.1 200 OK
Last-Modified: Fri, 02 Jul 2010 03:59:09 GMT
Content-Type: video/x-flv
Date: Thu, 26 May 2011 09:32:38 GMT
Expires: Thu, 26 May 2011 09:32:38 GMT
Cache-Control: private, max-age=22942
Accept-Ranges: bytes
Content-Length: 4349589
Connection: close
X-Content-Type-Options: nosniff
Server: gvs 1.0
```


HTTP Header for the Third HTTP Transaction

- The final transaction retrieves the viewing history for the relevant videos, e.g. daily visits to the video. The history indicates how the video's popularity has grown (or declined) and the lifespan of the YouTube video (for videos removed from YouTube). The view history is returned in the body of the HTTP response.

```
GET /insight_ajax?action_get_statistics_and_data=1&v=aZpD0btOZx8
HTTP/1.1
Accept: */*
Accept-Language: en-gb
Referer: http://www.youtube.com/watch?v=aZpD0btOZx8
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1;
Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET
CLR 3.0.30729; InfoPath.3; MS-RTC LM 8; .NET4.0C; .NET4.0E)
Host: www.youtube.com
Connection: Keep-Alive
Cookie: VISITOR_INFO1_LIVE=MSD8-W-tZoY;
__utma=27069237.1058702883.1301574126.1306314129.1306317488.4;
__utmz=27069237.1301648749.2.2.utmgclid=CIGf8bj7-
qccFUtC4Qod3mLzpw|utmccn=(not%20set)|utmcmd=(not%20set);
PREF=f1=50000000&fv=10.1.85;
dkv=ba89d11f83da4149215e846b87fe8ca0e3QEAAAAdGxpcGkx29xNMA==;
use_hitbox=72c46ff6cbcdb7c5585c36411b6b334edAEAAAAw;
GEO=42ab020acc2cb114520faaffc9d37dcbcwsAAAAzR0I+abNCTd4lxA==;
recently_watched_video_id_list=18c81a38b3bc582041d46059ab73a4b5WwYA
AABzCwAAAGFacEQwYnRPWng4cwsAAABoUTVNZFA5UlRoVXMLAAAAQUhWC3kwREhpVnN
zCwAAAFRwRFFacWZrT2tVcwsAAABOVFpFSE9ibUIwTXMLAAAAM0J2QUJuN0xfU28=;
MetrixLabp11265_uk_post=1

HTTP/1.1 200 OK
Date: Thu, 26 May 2011 10:09:01 GMT
Server: Apache
X-Content-Type-Options: nosniff
Expires: Tue, 27 Apr 1971 19:44:06 EST
Cache-Control: no-cache
Content-Length: 13307
Content-Type: text/xml; charset=UTF-8
```

Meta-Data of a YouTube Video

- After a video is uploaded and converted, YouTube randomly assigns it a unique 64-bit number, which is represented in a base-64 encoding algorithm by an 11-character alphanumeric ID. This unique ID can be used to check duplicate video downloading and remove redundancy.

Field Name	Field Value
YouTube ID	aZpD0btOZx8
Video Title	Super Mario
Video Category	Music
Content Length	4163902 (Bytes)
Content Resolution	34/640x360/9/0/115, 18/640x360/9/0/115, 5/320x240/7/0/0
Video Duration	64 (seconds)
Total Number of Views	36383401

Copyright and Access Issues

- Numerous debates and discussions are available with regard to the legality of downloading and the potential copyright infringement of YouTube videos.
- Naturally, YouTube does not want users to sidestep advertising by engaging in video downloads. However, avoiding advertising is not inherently illegal, in the same sense that it is legal to use in-browser advert blockers and skip ads on TV.
- Nevertheless, research conduct involving YouTube video downloading can be justified for two possible reasons:
  - First, each YouTube video content is not fully downloaded – only the first 2 Kbytes of content, because the header contains all necessary metadata infomation for this research.
  - Second, the research conducted complies with the fair use policy of the World Intellectual Property Organization (WIPO), which states: *"The fair use of a copyright work, for purposes such as criticism, comment, news reporting, teaching, scholarship, or research, is not an infringement of copyright."*

Topics

- Web Traffic Capture
- Case Study: YouTube Video Characteristics

Goals of YouTube Traffic Characterisation

- Traffic produced by YouTube has had a significant impact on both fixed and mobile networks.
- The study and evaluation of YouTube content features can benefit network traffic engineering by supporting the development of sustainable video delivery services and regulation of network traffic.
- Such evaluations are particularly useful to network operators who aim to refine and optimize existing cache algorithms to better adapt to YouTube video traffic patterns.

- In 2015, wired devices accounted for 52% of the IP traffic
- By 2020, wireless and mobile devices will take up 66% of the IP traffic
- Internet video (82% of the IP traffic by 2020)
- Virtual reality (x4 in 2015)
- Internet gaming (x7 in 2015)

Internet Traffic Prediction

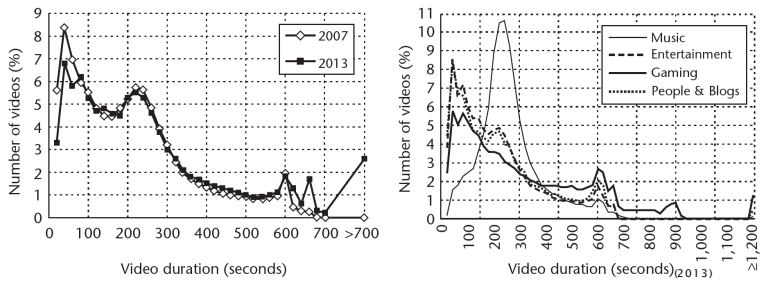
YouTube Video Category Distribution

- Following Google's acquisition of YouTube in 2008, several major aspects of the network and service framework were restructured, leading to changes in the user policy and service infrastructure.

	2007		2013	
Rank	Category	Percentage	Category	Percentage
1	Music	22.9	Music	22.8
2	Entertainment	17.8	Entertainment	16.0
3	Comedy	12.1	Gaming	8.5
4	Sports	9.7	People & Blogs	8.1
5	Film & Animation	8.4	Sports	8.0
6	People & Blogs	7.4	Comedy	5.9
7	Gaming	7.3	Film & Animation	5.9
8	News & Politics	4.3	How To & Style	5.1
9	Autos & Vehicles	2.5	News & Politics	4.6
10	Travel & Places	2.2	Cars & Vehicles	3.9
11	How To & DIY	2.0	Science & Technology	2.9
12	Pets & Animals	1.9	Education	2.9
13	—	—	Travel Events	2.2
14	—	—	Pets & Animals	1.8

YouTube Video Duration Distribution

- Generally speaking, YouTube videos are shorter than traditional media videos (such as films and TV programs).
- Most videos are less than 600 seconds in length. This is mainly due to the limit of 10 minutes imposed by YouTube in March 2006, which was increased to 15 minutes in July 2010. Partner users of YouTube and users with verified status can upload videos longer than the set limit.



The four peaks reflect various policies from YouTube.

The Four Peaks in the Statistics

1. The first peak is at 1 minute. Note that in 2007, approximately 21% of the videos were shorter than one minute; in 2013, 16% were shorter than one minute. YouTube has been viewed as an outlet for short videos since 2005, although the decrease in the percentage of short videos indicates that YouTube is gradually trying to cater to those wishing to upload longer videos as well.
2. The second and third peaks are consistent over the past five years. The second peak, which is within the range of 200 to 240 seconds, occurs because the Music category has been a long-standing popular category on YouTube, and the typical length of music videos is often within this range.
3. The third peak is near the duration of between 580 and 600 seconds due to the duration limit imposed by YouTube. Users often tend to divide long videos into several pieces with each fitting the boundary of 10 minutes.
4. The fourth peak is caused by the number of videos that exceed 700 seconds in length. In July 2010, YouTube raised the video uploading limit to 15 minutes, and five months later it allowed verified users to upload videos longer than 15 minutes. Users have clearly started to take advantage of this new facility.

YouTube Video Resolution Distribution

- The analysis of YouTube video resolutions is important because it not only gives a definitive indication of video qualities in the system, but also offers an indicative reflection of users' uploading capabilities.
- With the dramatic growth in smartphone use in recent years, YouTube started offering support for the MP4 format in 2007 for devices that do not offer Flash, such as Apple's iPhone and iPad. Furthermore, starting in March 2008, it permitted a wider range of resolutions. Such expansions of YouTube services illustrates the company's willingness to adapt to an evolving market.

1	320 × 240 (FLV)	100%
2	640 × 360 (MP4)	74%
3	640 × 360 (FLV)	66%
4	854 × 480 (FLV)	40%
5	320 × 240 (MP4)	26%
6	320 × 240 (others)	20%
7	1280 × 720 (total)	14%
8	1920 × 1080 (total)	3%

YouTube Video Resolution Rankings (2013)

When the same video content is being uploaded to YouTube servers, each unique upload is transcoded into a variety of formats and resolutions to support streaming requirements, so several sources (that is, files) are saved on the server, each corresponding to one resolution.

YouTube Video File Size Distribution

- YouTube's policy on the size limit of video files was 100 Mbytes when previous studies were carried out prior to 2008. The current file size limit is 2 GB for uploading via YouTube web or 20 GB if up-to-date browser versions are used*. Because of this policy change, the average file size has increased over the past few years.
- One important goal of investigating YouTube file sizes is to help network carriers with cache management. The average YouTube video file size is approximately 17.6 MB for a resolution of 640 x360 (FLV) and 6.5 MB for a resolution of 320x240 (FLV). Therefore, if 1 million YouTube videos were to be cached, the total disk space required for storage would be approximately 17.6 TB for resolution of 640x360 (FLV) and 6.5 TB for a resolution 320x240 (FLV).

File size	2007	2013	640 × 360 FLV (2013)	320 × 240 FLV (2013)
< 30 Mbytes	98.8%[1]	90.1%	84.5%	99.2%
30–100 Mbytes	1.0%[4]	6.7%	14.6%	0.8%
> 100 Mbytes	0.1%[2]	0.5%	0.9%	0.1%
Average file size	8.4 Mbytes[1] 9.8 Mbytes[4]	13.8 Mbytes	17.6 Mbytes	6.5 Mbytes

*Policy may have changed again to date.

YouTube Video Data Rate Distribution

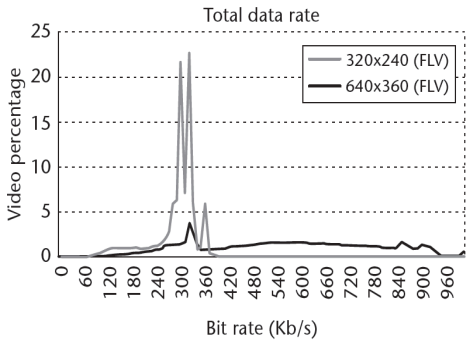
- YouTube video playback data rate can be derived from dividing the video file size by the video duration. Alternatively, the data rate can be easily observed from the FLV header retrieved from the metadata.
- The hex highlighted represents the meta-data for the total data rate. The first hex "0D" refers the length of ASCII code for "totaldatarate". The value of total data rate is in the format of IEEE 754-2008, IEEE Standard for Floating-Point Arithmetic.

```
00000000  46 4C 56 01 05 00 00 00  09 00 00 00 00 12 00 03  FLV.............
00000010  4B 00 00 00 00 00 00 00  02 00 0A 6F 6E 4D 65 74  K..........onMet
00000020  61 44 61 74 61 08 00 00  00 0F 00 08 64 75 72 61  aData.......dura
00000030  74 69 6F 6E 00 40 63 6E  14 7A E1 47 AE 00 09 73  tion.@cn.z.G...s
00000040  74 61 72 74 74 69 6D 65  00 00 00 00 00 00 00 00  tarttime........
00000050  00 00 0D 74 6F 74 61 6C  64 75 72 61 74 69 6F 6E  ...totalduration
00000060  00 40 63 6E 14 7A E1 47  AE 00 05 77 69 64 74 68  .@cn.z.G...width
00000070  00 40 74 00 00 00 00 00  00 00 06 68 65 69 67 68  .@t........heigh
00000080  74 00 40 70 00 00 00 00  00 00 00 0D 76 69 64 65  t.@p........vide
00000090  6F 64 61 74 61 72 61 74  65 00 40 73 6F 7C AA A6  odatarate.@so|..
000000a0  2B 60 00 0D 61 75 64 69  6F 64 61 74 61 72 61 74  +`..audiodatarat
000000b0  65 00 40 56 5C 2C F5 CF  77 DC 00 0D 74 6F 74 61  e.@V\,..w...tota
000000c0  6C 64 61 74 61 72 61 74  65 00 40 79 78 45 C0 57  ldatarate.@yxE.W
000000d0  66 36 00 09 66 72 61 6D  65 72 61 74 65 00 40 39  f6..framerate.@9
000000e0  01 A5 9D B5 F4 24 00 0A  62 79 74 65 6C 65 6E 67  .....$..byteleng
000000f0  74 68 00 41 5E 4F 99 80  00 00 00 00 0D 63 61 6E  th.A^O.......can
00000100  73 65 65 6B 6F 6E 74 69  6D 65 01 01 00 0A 73 6F  seekontime....so
00000110  75 72 63 65 64 61 74 61  02 00 20 42 41 44 43 32  urcedata.. BADC2
00000120  30 30 38 31 4D 48 31 33  30 32 32 31 36 39 35 32  0081MH1302216952
00000130  33 31 34 34 30 32 00 00  00 00 00 04 70 75 72 32  314402.....pur
00000140  6C 02 00 80 00 00 00 00  00 00 00 00 00 00 00 00  l...............
00000150  00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00  ................
00000160  00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00  ................
```

YouTube Video Data Rate Distribution

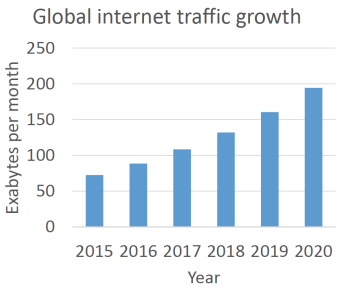
- More than 99.9 percent of the YouTube videos accessed contained FLV metadata specifying the content's total data rate, video data rate, and audio data rate. This indicates that virtually all YouTube videos are transmitted at a constant bit rate (CBR). Compared with a variable bit rate (VBR), which is suitable for high-quality video download and certain high bandwidth streaming environments, CBR is a more reliable choice for streaming videos than any bandwidth that users might have.

The peak implies that the most common data rate budget for today's users is approximately 320 Kbps.



Trend of Internet Traffic

Supply vs. Demand	*Mobile network*	*Content*



- Compound annual growth rate of 23% (Cisco 2016). Expected to saturate the lit fibre network by 2020

- In 2015, wired devices accounted for 52% of the IP traffic
- By 2020, wireless and mobile devices will take up 66% of the IP traffic

- Internet video (82% of the IP traffic by 2020)
- Virtual reality (x4 in 2015)
- Internet gaming (x7 in 2015)

Future network research in light of the challenges:

- Growing demand for internet capacity
- Hybrid, converged network consisting of wired and wireless networks, backbones and access networks
- Demanding user Internet content- latency and QOS issues
- Research collaboration is essential

Statistics from NDFIS – National Dark Fibre Infrastructure Service

Next Lecture

- Case Study: MMOG Traffic Analysis

Questions?

- **Email:** x.che@herts.ac.uk
- **Office:** LB218
- **Tel:** 01707 286206