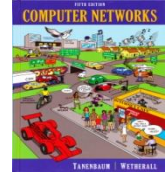


Lab Exercise – HTTP



Objective

HTTP (HyperText Transfer Protocol) is the main protocol underlying the Web. HTTP functions as a request–response protocol in the client–server computing model. A web browser, for example, may be the client and an application running on a computer hosting a website may be the server. The client submits an HTTP request message to the server. The server, which provides resources such as HTML files and other content, or performs other functions on behalf of the client, returns a response message to the client. The response contains completion status information about the request and may also contain requested content in its message body. A web browser is an example of a user agent (UA). Other types of user agent include the indexing software used by search providers (web crawlers), voice browsers, mobile apps, and other software that accesses, consumes, or displays web content. HTTP is designed to permit intermediate network elements to improve or enable communications between clients and servers. High-traffic websites often benefit from web cache servers that deliver content on behalf of upstream servers to improve response time. Web browsers cache previously accessed web resources and reuse them when possible to reduce network traffic. HTTP is an application layer protocol designed within the framework of the Internet protocol suite. Its definition presumes an underlying and reliable transport layer protocol and Transmission Control Protocol (TCP) is commonly used.

Step 1: Open the http Trace

Browser behavior can be quite complex, using more HTTP features than the basic exchange, this trace will show us how much gets transferred.

1. Open the trace file from here: <https://kevincurran.org/com320/labs/wireshark/trace-http.pcap>
You should then see a Wireshark screen like below.

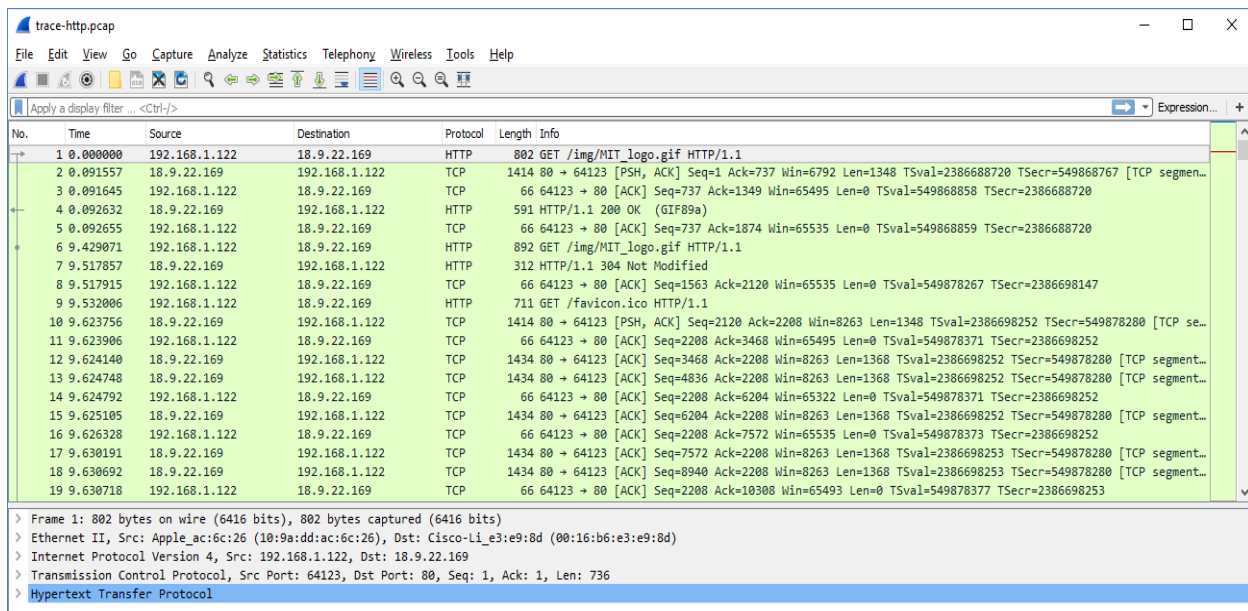


Figure 1: The Wireshark screen after opening the http trace file

Step 2: Inspect the Trace

2. To focus on HTTP traffic, *enter and apply a filter expression of “http”*.

This filter will show HTTP requests and responses, but not the individual packets that are involved. You should see a screen like below.

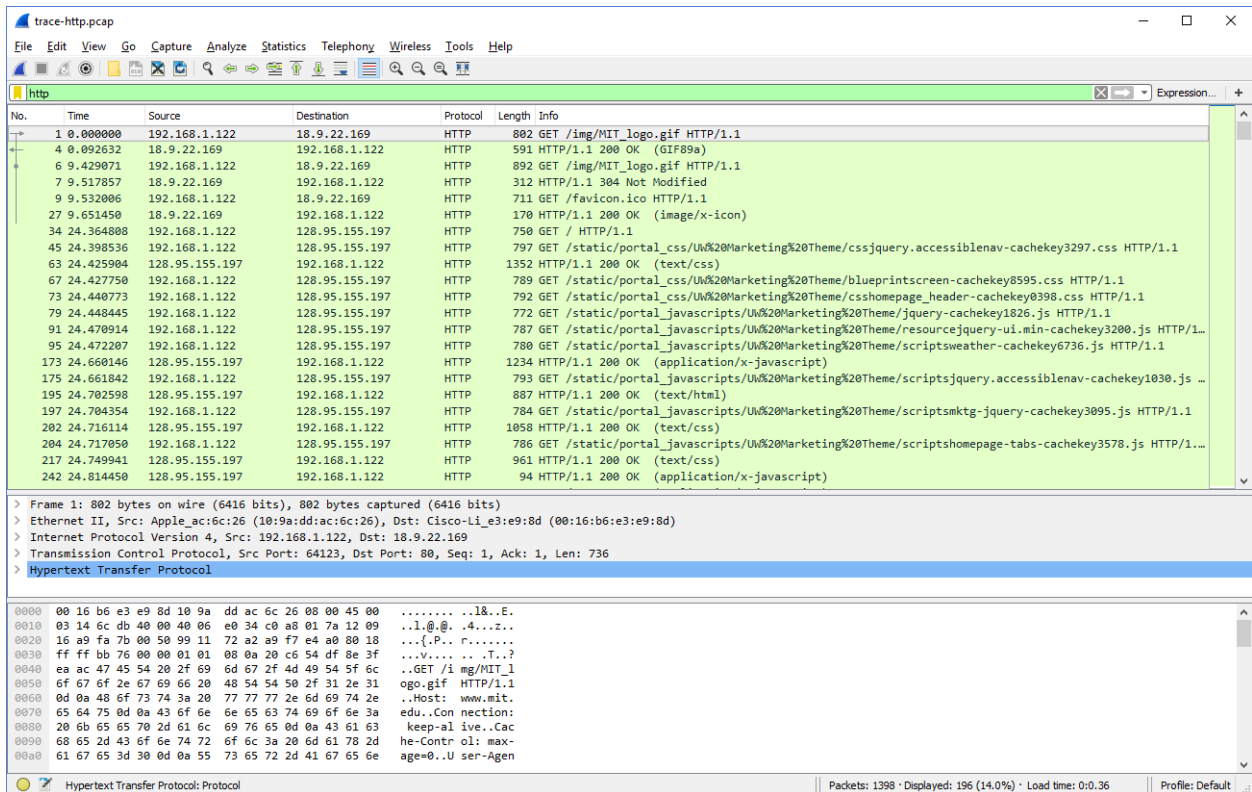
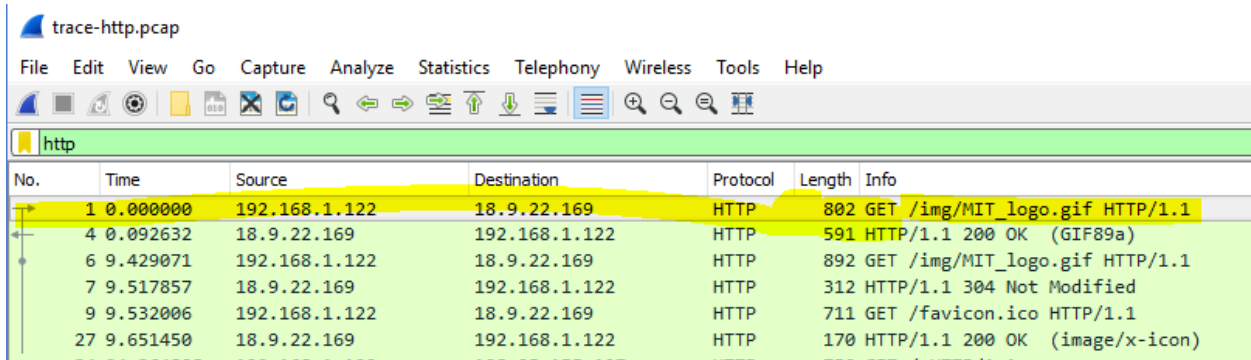


Figure 2: Trace of HTTP traffic

Recall that an HTTP response carrying content will normally be spread across multiple packets. When the last packet in the response arrives, Wireshark assembles the complete response and tags the packet with protocol HTTP. The earlier packets are simply TCP segments carrying data; the last packet tagged HTTP includes a list of all the earlier packets used to make the response. A similar process occurs for the request, but in this case it is common for a request to fit in a single packet. With the filter expression of “http” we will hide the intermediate TCP packets and see only the HTTP requests and responses.

3. Select the first GET in the trace and expand its HTTP block.

(This will be 802 GET /img/MIT_logo.gif HTTP/1.1 as shown below).



The screenshot shows the Wireshark interface with a packet capture of an HTTP GET request. The packet list pane shows several packets, with the first packet (No. 1) selected and expanded. The packet details pane shows the following information:

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	192.168.1.122	18.9.22.169	HTTP	802	GET /img/MIT_logo.gif HTTP/1.1
4	0.092632	18.9.22.169	192.168.1.122	HTTP	591	HTTP/1.1 200 OK (GIF89a)
6	9.429071	192.168.1.122	18.9.22.169	HTTP	892	GET /img/MIT_logo.gif HTTP/1.1
7	9.517857	18.9.22.169	192.168.1.122	HTTP	312	HTTP/1.1 304 Not Modified
9	9.532006	192.168.1.122	18.9.22.169	HTTP	711	GET /favicon.ico HTTP/1.1
27	9.651450	18.9.22.169	192.168.1.122	HTTP	170	HTTP/1.1 200 OK (image/x-icon)

Figure 3: The first GET packet in trace: get /img/MIT_logo.gif HTTP/1.1

This will let us inspect the details of an HTTP request.

4. Inspect the HTTP header by expanding the down arrow beside “Hypertext Transfer Protocol” In the middle pane (as shown below).

```
> Frame 1: 802 bytes on wire (6416 bits), 802 bytes captured (6416 bits)
> Ethernet II, Src: Apple_ac:6c:26 (10:9a:dd:ac:6c:26), Dst: Cisco-Li_e3:e9:8d (00:16:b6:e3:e9:8d)
> Internet Protocol Version 4, Src: 192.168.1.122, Dst: 18.9.22.169
> Transmission Control Protocol, Src Port: 64123, Dst Port: 80, Seq: 1, Ack: 1, Len: 736
v Hypertext Transfer Protocol
  > GET /img/MIT_logo.gif HTTP/1.1\r\n
    Host: www.mit.edu\r\n
    Connection: keep-alive\r\n
    Cache-Control: max-age=0\r\n
    User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_3) AppleWebKit/536.11 (KHTML, like Gecko) Chrom
    Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8\r\n
    Accept-Encoding: gzip,deflate,sdch\r\n
    Accept-Language: en-US,en;q=0.8\r\n
    Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.3\r\n
  > [truncated]Cookie: __gads=ID=8ad65dc86381f8d7:T=1303597795:S=ALNI_MYTf6iFAaVdJjXmft_V1_01Q1AyCg; __utm
  \r\n
  [Full request URI: http://www.mit.edu/img/MIT_logo.gif]
```

Figure 4: The Hypertext transfer protocol for the first http request

Observe that the HTTP header follows the TCP and IP headers, as HTTP is an application protocol that is transported using TCP/IP. To view it, select the packet, find the HTTP block in the middle panel, and expand it (by using the “v” expander or icon). This block is expanded in figure above.

Explore the headers that are sent along with the request. First, you will see the GET method at the start of the request, including details such as the path. Then you will see a series of headers in the form of tagged parameters. There may be many headers, and the choice of headers and their values vary from browser to browser. See if you have any of these common headers:

- Host. A mandatory header, it identifies the name (and port) of the server.

- User-Agent. The kind of browser and its capabilities.
- Accept, Accept-Encoding, Accept-Charset, Accept-Language. Descriptions of the formats that will be accepted in the response, e.g., text/html, including its encoding, e.g., gzip, and language.
- Cookie. The name and value of cookies the browser holds for the website.
- Cache-Control. Information about how the response can be cached.

The request information is sent in a simple text and line-based format. If you look in the bottom panel you can read much of the request directly from the packet itself.

5. *Select the response that corresponds to the first GET in the trace and expand its HTTP block (see below).*

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	192.168.1.122	18.9.22.169	HTTP	802	GET /img/MIT_logo.gif HTTP/1.1
4	0.092632	18.9.22.169	192.168.1.122	HTTP	591	HTTP/1.1 200 OK (GIF89a)
6	9.429071	192.168.1.122	18.9.22.169	HTTP	892	GET /img/MIT_logo.gif HTTP/1.1
7	9.517857	18.9.22.169	192.168.1.122	HTTP	312	HTTP/1.1 304 Not Modified
9	9.532006	192.168.1.122	18.9.22.169	HTTP	711	GET /favicon.ico HTTP/1.1
27	9.651450	18.9.22.169	192.168.1.122	HTTP	170	HTTP/1.1 200 OK (image/x-icon)

Figure 5: The response that corresponds to the first GET in the trace

The Info for this packet will indicate “200 OK” in the case of a normal, successful transfer. You will see that the response is similar to the request, with a series of headers that follow the “200 OK” status code. However, different headers will be used, and the headers will be followed by the requested content.

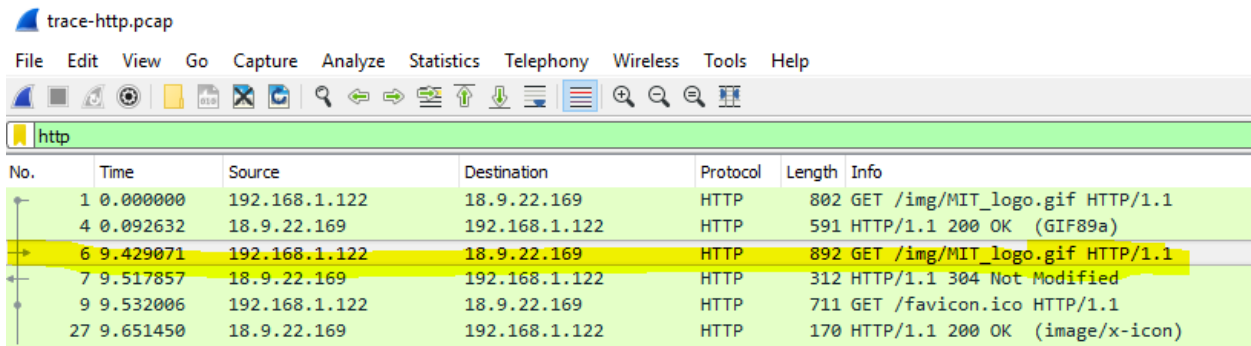
Examine the common headers such as:

- Server. The kind of server and its capabilities.
- Date, Last-Modified. The time of the response and the time the content last changed.
- Cache-Control, Expires, Etag. Information about how the response can be cached.

Step 3: Content Caching

The second fetch in the trace is a re-fetch of the first URL. This fetch presents an opportunity for us to look at caching in action, since it is highly likely that the image or document has not changed and therefore does not need to be downloaded again. HTTP caching mechanisms should identify this opportunity. We will now see how they work.

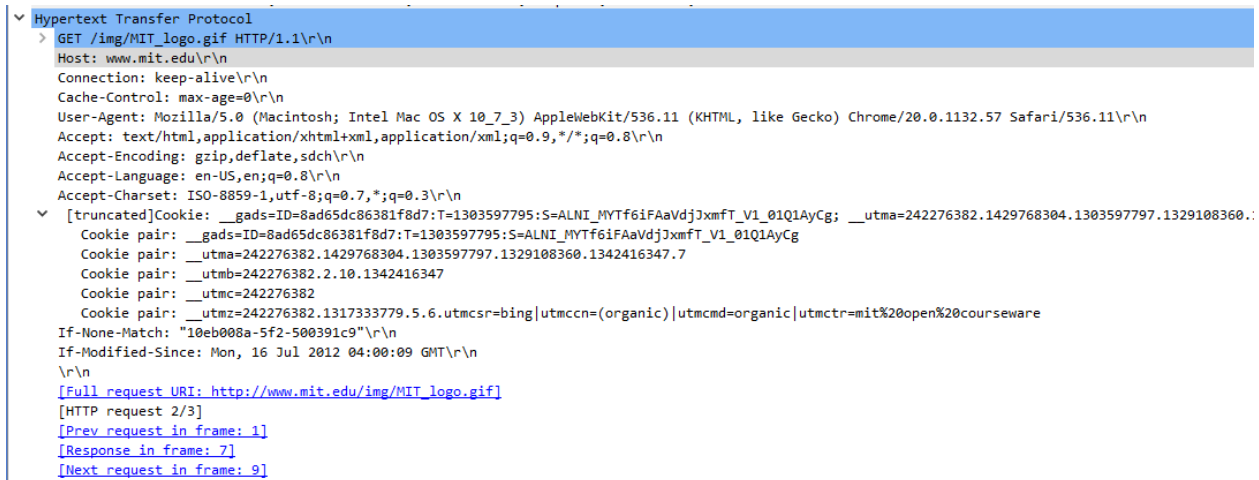
6. Click on the third line in the sample trace i.e. 892 GET /img/MIT_logo.gif HTTP 1.1



No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	192.168.1.122	18.9.22.169	HTTP	802	GET /img/MIT_logo.gif HTTP/1.1
4	0.092632	18.9.22.169	192.168.1.122	HTTP	591	HTTP/1.1 200 OK (GIF89a)
6	9.429071	192.168.1.122	18.9.22.169	HTTP	892	GET /img/MIT_logo.gif HTTP/1.1
7	9.517857	18.9.22.169	192.168.1.122	HTTP	312	HTTP/1.1 304 Not Modified
9	9.532006	192.168.1.122	18.9.22.169	HTTP	711	GET /favicon.ico HTTP/1.1
27	9.651450	18.9.22.169	192.168.1.122	HTTP	170	HTTP/1.1 200 OK (image/x-icon)

Figure 6: The second fetch - 892 GET /img/MIT_logo.gif HTTP 1.1

7. Expand its HTTP block (see below)



```
Hypertext Transfer Protocol
> GET /img/MIT_logo.gif HTTP/1.1\r\n
Host: www.mit.edu\r\n
Connection: keep-alive\r\n
Cache-Control: max-age=0\r\n
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_3) AppleWebKit/536.11 (KHTML, like Gecko) Chrome/20.0.1132.57 Safari/536.11\r\n
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8\r\n
Accept-Encoding: gzip,deflate,sdch\r\n
Accept-Language: en-US,en;q=0.8\r\n
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.3\r\n
[truncated]Cookie: __gads=ID=8ad65dc86381f8d7:T=1303597795:S=ALNI_MyTf6iFAAvdJjxmFT_V1_01Q1AyCg; __utma=242276382.1429768304.1303597797.1329108360.1342416347.7
Cookie pair: __gads=ID=8ad65dc86381f8d7:T=1303597795:S=ALNI_MyTf6iFAAvdJjxmFT_V1_01Q1AyCg
Cookie pair: __utma=242276382.1429768304.1303597797.1329108360.1342416347.7
Cookie pair: __utmb=242276382.2.10.1342416347
Cookie pair: __utmc=242276382
Cookie pair: __utmz=242276382.1317333779.5.6.utmcsr=bing|utmccn=(organic)|utmcmd=organic|utmctr=mit%20open%20courseware
If-None-Match: "10eb008a-5f2-500391c9"\r\n
If-Modified-Since: Mon, 16 Jul 2012 04:00:09 GMT\r\n
\r\n
[Full request URI: http://www.mit.edu/img/MIT_logo.gif]
[HTTP request 2/3]
[Prev request in frame: 1]
[Response in frame: 7]
[Next request in frame: 9]
```

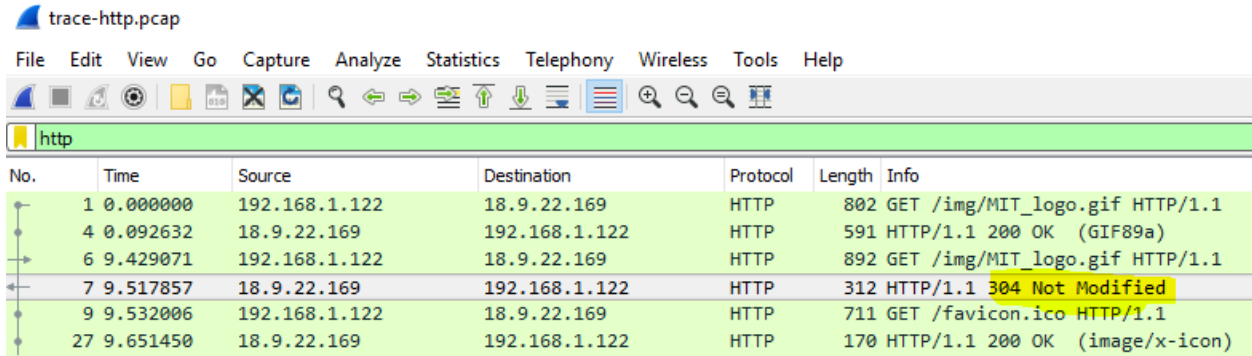
Figure 7: The HTTP block of 892 GET /img/MIT_logo.gif HTTP 1.1

Now find the header that will let the server work out whether it needs to send fresh content. The server will need to send fresh content only if the content has changed since the browser last downloaded it. To work this out, the browser includes a timestamp taken from the previous download for the content that it has cached. This header was not present on the first GET since we cleared the browser cache so the browser had no previous download of the content that it could use. In most other respects, this request will be the same as the first-time request.

This header is called “If-Modified-Since”, i.e., it asks the server to send the content if it has been modified since a given time (see the header below).

```
If-Modified-Since: Mon, 16 Jul 2012 04:00:09 GMT\r\n\r\n
```

8. Now select the response to the re-fetch, and expand its HTTP block.



The screenshot shows the Wireshark interface for a trace file named 'trace-http.pcap'. The packet list pane shows several HTTP packets. Packet 7 is selected and highlighted in yellow. The 'Info' column for packet 7 shows '304 Not Modified' in yellow. The 'Time' column for packet 7 is 9.517857. The 'Source' is 18.9.22.169 and the 'Destination' is 192.168.1.122. The 'Protocol' is HTTP and the 'Length' is 312. The 'Info' column also shows 'HTTP/1.1'.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	192.168.1.122	18.9.22.169	HTTP	802	GET /img/MIT_logo.gif HTTP/1.1
4	0.092632	18.9.22.169	192.168.1.122	HTTP	591	HTTP/1.1 200 OK (GIF89a)
6	9.429071	192.168.1.122	18.9.22.169	HTTP	892	GET /img/MIT_logo.gif HTTP/1.1
7	9.517857	18.9.22.169	192.168.1.122	HTTP	312	HTTP/1.1 304 Not Modified
9	9.532006	192.168.1.122	18.9.22.169	HTTP	711	GET /favicon.ico HTTP/1.1
27	9.651450	18.9.22.169	192.168.1.122	HTTP	170	HTTP/1.1 200 OK (image/x-icon)

Figure 8: The response to cached image

We can see that caching worked as expected, this response will not contain the content. Instead, the status code of the response is “304 Not Modified”. This tells the browser that the content is unchanged from its previous copy, and the cached content can then be displayed.

Note the timestamp value comes from the “Last-Modified” header of the most recent download of the content. It is a server timestamp for when the content last changed – it is not a timestamp according to the browser clock, and it is not a timestamp of the time of the download

Step 4: Complex Pages

Now we will examine the fourth fetch in the trace. This fetch was for a more complex web page that will likely have embedded resources. So the browser will download the initial HTML plus all of the embedded resources needed to render the page, plus other resources that are requested during the execution of page scripts. As we will see, a single page can involve many GETs.

- Click on no 34 which is 750 GET / HTTP/1.1 (as shown below).

27	9.651450	18.9.22.169	192.168.1.122	HTTP	170	HTTP/1.1	200	OK	(image/x-icon)
34	24.364808	192.168.1.122	128.95.155.197	HTTP	750	GET /	HTTP/1.1		
45	24.398536	192.168.1.122	128.95.155.197	HTTP	797	GET /static/portal_css/UW%20Marketing%20Theme/css/jquery.accessib..			

Figure 8: The large fourth fetch in http trace

- Summarize the GETs for this page by bringing up a HTTP Load Distribution panel. You will find this panel under “Statistics” and “HTTP”.

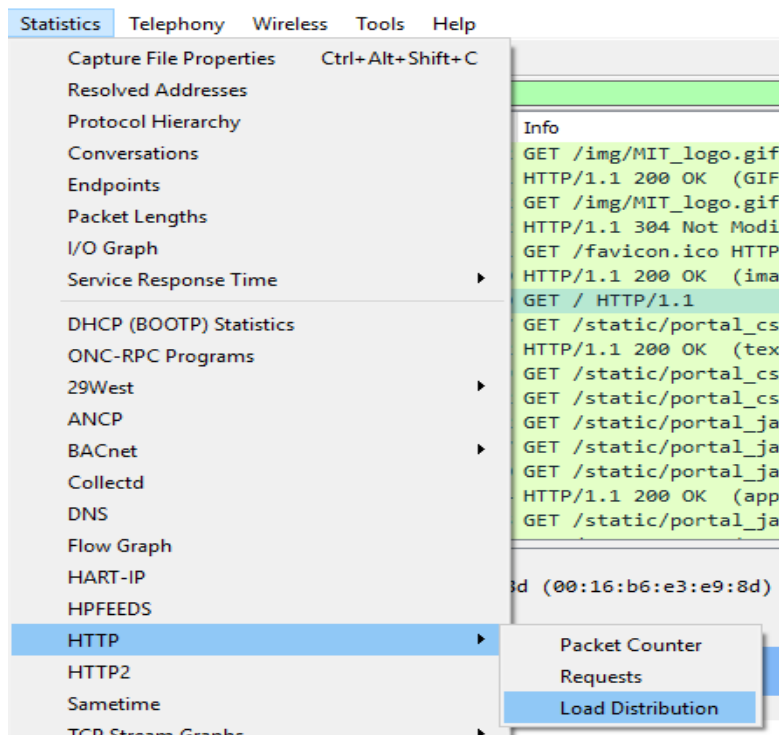


Figure 9: Statistics → HTTP → Load Distribution

Looking at this panel will tell you how many requests were made to which servers. This is typical, when surfing, chances are that your fetch will request content from other servers you might not have suspected to build the page. These other servers may include third parties such as content distribution networks, ad networks, and analytics networks. Our panel is shown below – the page fetch involved 98 requests to 5 different servers.

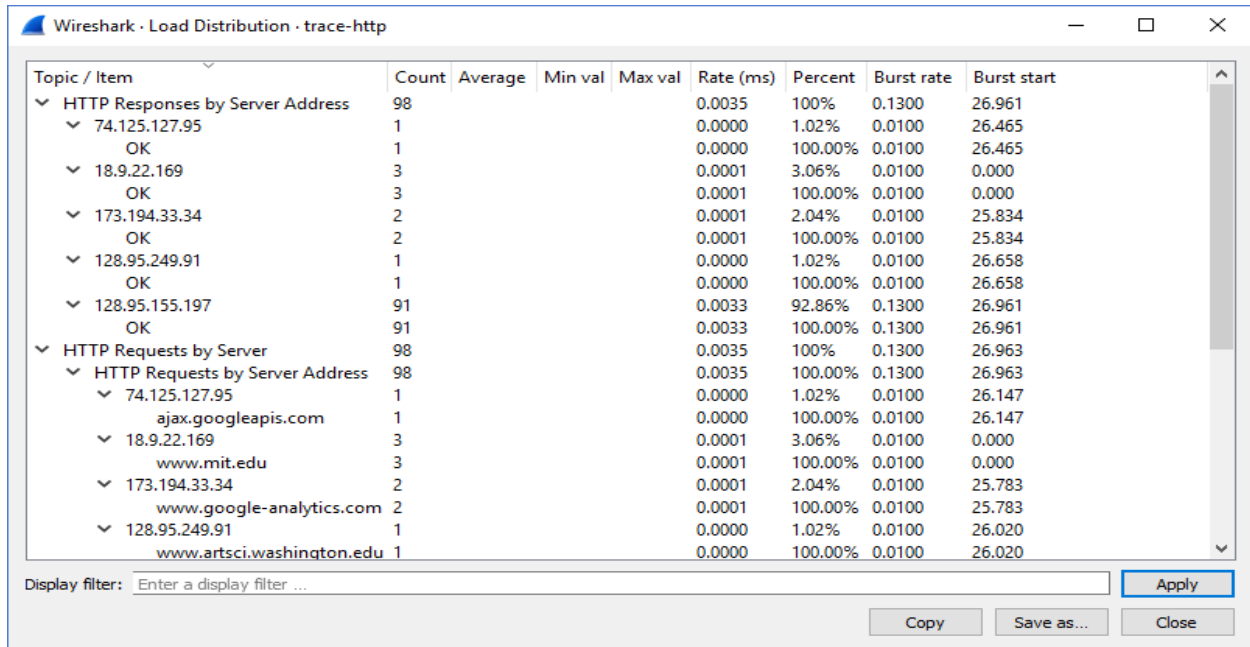


Figure 10: HTTP Load Distribution panel

11. For a different kind of summary of the GETs, bring up a HTTP Packet Counter panel.

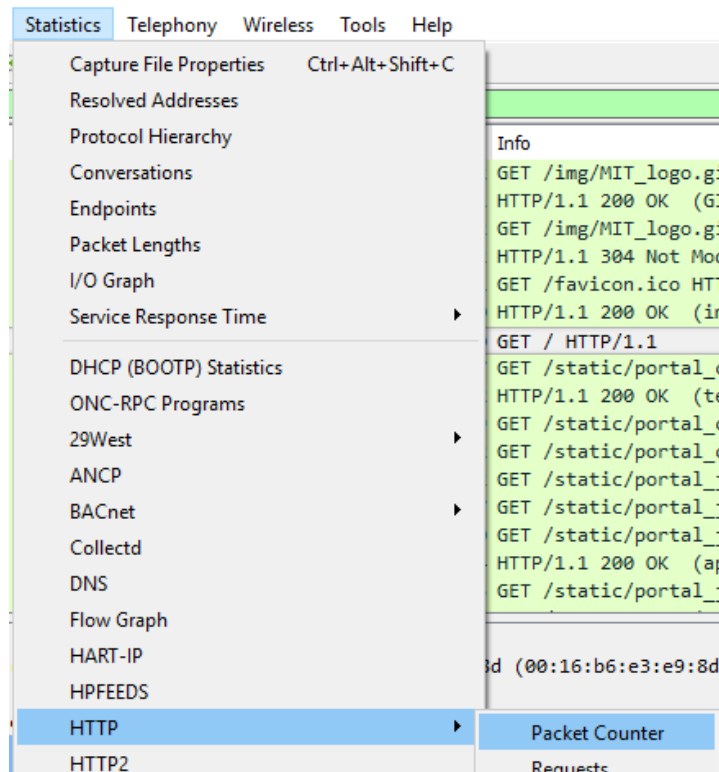


Figure 11: Statistics → HTTP → Packet Counter

You will also find this panel under “Statistics” and “HTTP”, and you should filter for the packets that are part of the fetch as before. This panel will tell you the kinds of request and responses. Our panel is shown in the figure below. You can see that it consists of GET requests that are matched by 200 OK responses and 304 Not Modified. However, there are a variety of other response codes that you might observe in your trace, such as when the resource is already cached.

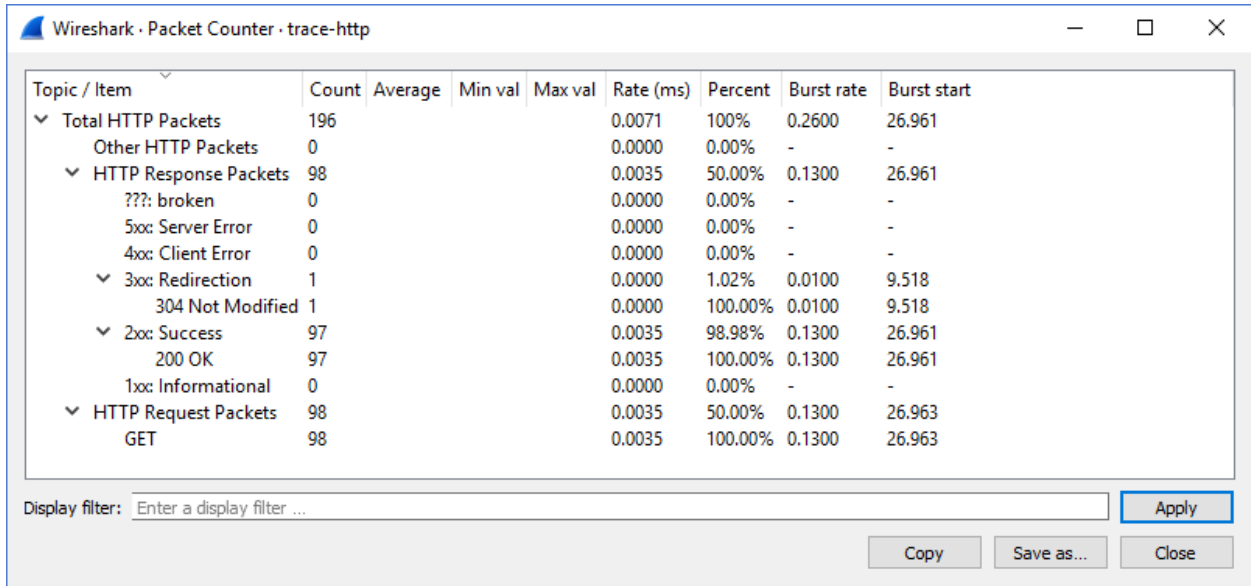


Figure 12: HTTP Packet Counter panel

You might be curious to know what content is being downloaded by all these requests. As well as seeing the URLs in the Info column, you can get a summary of the URLs in a HTTP Request panel under “Statistics” and “HTTP”. Each of the individual requests and responses has the same form we saw in an earlier step. Collectively, they are performed in the process of fetching a complete page with a given URL.

For a more detailed look at the overall page load process, use a site such as Google’s [PageSpeed](https://www.pagespeed.org/) or <https://www.webpagetest.org/>. These sites will test a URL of your choice and generate a report of the page load activity, telling what requests were fetched at what times and giving tips for decreasing the overall page load time.

The beginning of the “waterfall” diagram for the page load corresponding to a trace on www.ulster.ac.uk is shown in the next figure. After the initial HTML resource is fetched there are many subsequent quick fetches for embedded resources such as JavaScript scripts, CSS stylesheets, images, and more.

Waterfall View

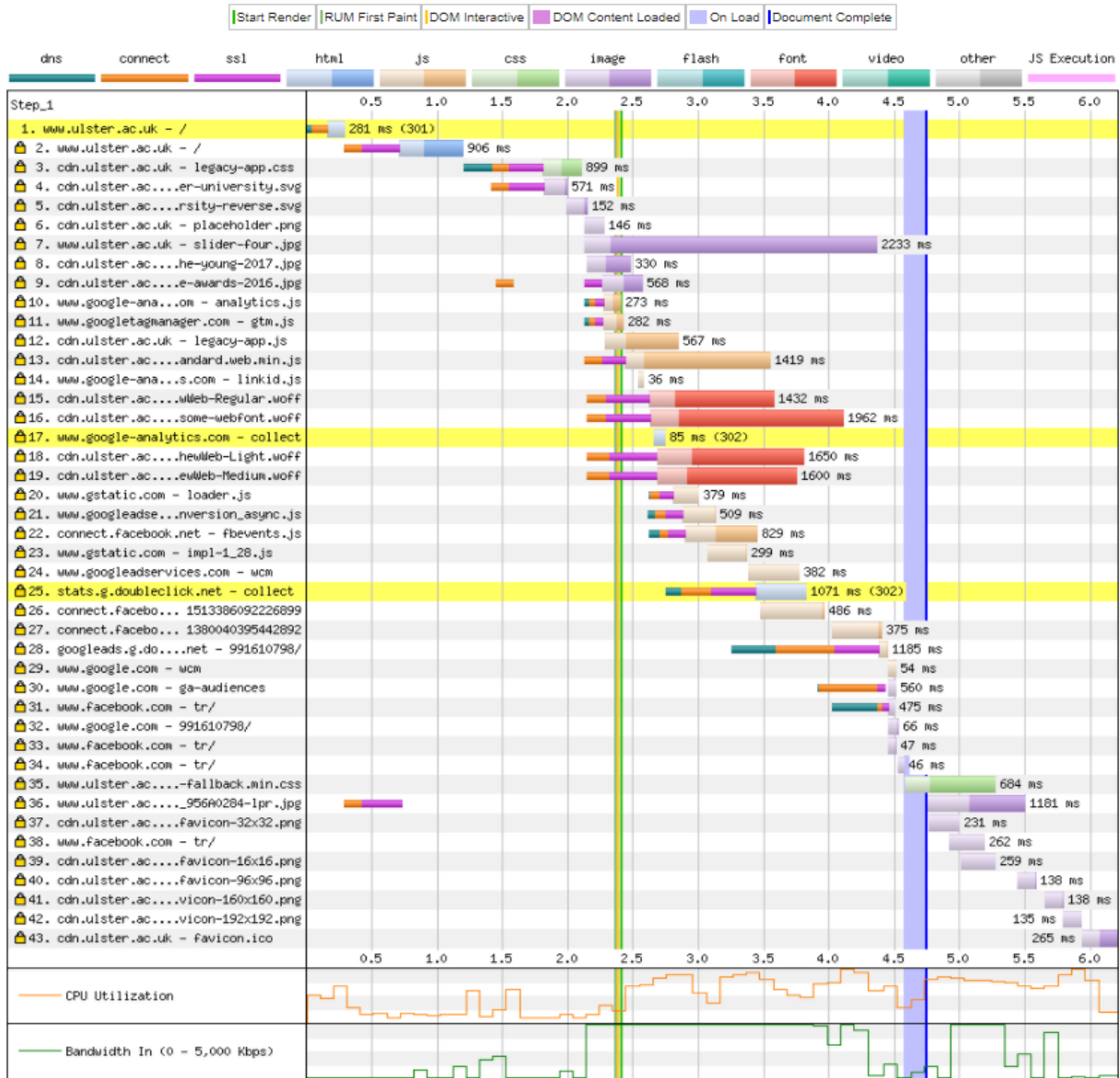


Figure 13: Start of waterfall graph for <https://www.ulster.ac.uk/> (from pageloadtest.org)

Tester: VM3-01-192.168.11.137
 First View only
 Test runs: 3

[Export HTTP Archive \(.har\)](#)
[Custom Metrics](#)

	Load Time	First Byte	Start Render	Visually Complete	Speed Index	Result (error code)	Document Complete			Fully Loaded		
							Time	Requests	Bytes In	Time	Requests	Bytes In
First View (Run 1)	4.736s	0.892s	2.400s	5.700s	4175	0	4.736s	35	1,542 KB	6.199s	43	1,657 KB

Colordepth	RUM First Paint	domInteractive	domContentLoaded	loadEvent
24	2.354s	2.380s	2.380s - 2.380s (0.000s)	4.578s - 4.737s (0.159s)

Figure 14: Summary stats for <https://www.ulster.ac.uk/> (from pageloadtest.org)