Lab 2: HTTP and Throughput

(100 points)

Platform: This entire lab is to be run on the Mininet VM. The lab computers are available for all of these problems.

Submission:

Important Announcement about Gradescope submissions:

- When submissions are made on Gradescope, the page/region for each problem MUST be marked.
- We will not grade submissions that are not marked. If you do not know how to do this, ask for help.
- If a submission does not have the region/page marked, you will be asked to resubmit.
- The recorded date for the assignment will be the resubmission date, and late days will be counted.

One PDF file for this assignment is to be submitted directly on Gradescope. Naming convention of the file: [YourCruzID].pdf.

For all questions that require a screenshot:

- Make sure that a **date timestamp** is visible next to your results.
- No credit will be given for screenshots without a timestamp.

References: Chapters 1 and 2, Computer Networking: A Top Down Approach

- Section 1.4: Delay, Loss and Throughput in Packet-Switched Networks
- Section 2.2: The Web and HTTP
- RFC 2616: Hypertext Transfer Protocol -- HTTP/1.1

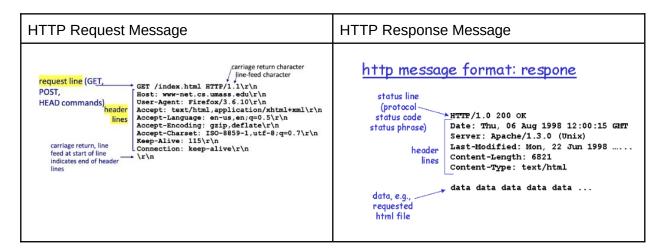
Introduction to HTTP

Before beginning this lab, it is very helpful to read Section 2.2 of the text.[1]

The HTTP protocol is the *lingua franca* of the Web. HTTP specifies methods and headers so that web clients can issue requests for web objects and receive responses from web servers. HTTP methods provide a set of actions to be performed on web resources. You will observe several of these methods in this assignment. HTTP headers are extra helpful information included in an HTTP message sent by either the client or the server. Headers are specified as name:value pairs. You can search the vast number of HTTP headers online.

Having gotten our feet wet with the Wireshark packet sniffer in Lab1, we're now ready to use Wireshark to investigate protocols in operation. In this lab, we'll explore several aspects of the HTTP protocol: the basic request/response interaction between client and server, HTTP message formats, retrieving HTML files with multiple embedded objects, and Conditional HTTP requests. Additionally, in the last problem, we will take a look at link transmission rates and bottleneck links.

Please review class notes and the textbook to review the format of the messages exchanged. A sample client request and server response are shown below:



Part A - HTTP: Hypertext Transfer Protocol

1. [35 pts] The Basic HTTP request/response messages

a. [10 points] In the following problem, match the HTTP headers to their respective definitions and usages. This exercise will help deepen your understanding of how each header contributes to the HTTP communication process, as detailed in RFC 9110

Table 1: Match the Header (on the left) with its Definition/Usage (on the right) by filling in the "Answer Here" column **(No Need To Memorize Header Definitions)**

	Header	Match ← →	Answer Here	Definition/Usage
Α	"Accept-Encoding"		С	Carries credentials to authenticate the requestor.
В	"Accept"		О	Contains data from the server to be stored by the client.
С	"Authorization"		A	Describes the compression formats that can be decoded.
D	"Cache-Control"		В	Details preferred content types for the response.
Е	"Connection"		I	Directive to the client to store the cookie.
F	"Content-Encoding"		D	Directs how responses are to be stored by caches.
G	"Content-Length"		P	Information about the client software making the request.
Н	"Content-Type"		Т	Numeric indication of the outcome of the server's attempt to fulfill the request.
I	"Cookie"		K	Only send the response if the content has been modified since this date/time.
J	"Host"		E	Options for managing the current connection state.
K	"If-Modified-Since"		L	Parameters for a persistent connection.
L	"Keep-Alive"		F	Specifies the encoding used on the data.
М	"Last-Modified"		R	Text description accompanying the status code, explaining the status.
N	"Server"		J	The domain name of the server and the TCP port number.
0	"Set-Cookie"		G	The exact byte length of the HTTP body.
Р	"User-Agent"		Н	The media type of the body of the request/response.
Q	"Request Version"		N	The name of the server software responding to the request.
R	"Response Phrase"		Q	The protocol version used by the client in making the request.
S	"Request Method"		M	The time at which the content was last changed.
Т	"Status Code"		S	The type of action the client wants the server to take on a resource.

b. **[10 points]** Now let's explore HTTP by downloading a very simple HTML file - one that is very short and contains no embedded objects.

Do the following:

- First, open up the web browser in the Mininet VM.
- Clear the cache in the browser^[2] (go to the Chromium browser setting, search Clear browsing data, select everything for all time, and clear data).
- Start up Wireshark. Set the display filter so that only HTTP messages will be displayed. (We're only interested in the HTTP protocol here and don't want to see the clutter of all captured packets).

Then, start a packet capture using the interface **eth0**.

- Go back to the browser and enter: http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html
 http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html
 <a href="http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/HTTP-wireshark-labs/H
- Stop the Wireshark packet capture, ignore the GET /favicon.ico packet.

Answer the questions in the following tables:

Table 2: In Wireshark, use the Client Request Packet to answer the following. All answers are visible in the packet.

Network and Transport Layer (TCP/IP)					
Questions:	Answer (Field Detail from Wireshark)				
Example: What is the destination port number?	Example Solution (from Wireshark): Destination port: http (80)				
What is the Client IP Address?	10.0.2.15				
What is the Client port number?	38456				
Application Layer (HTTP)					
Request Line (Initial Line of HTTP Request) Questions:	Answer (Elements of the Request Line from Wireshark)				
What is the Method used for the request?	get				
What is the URI requested by the client?	/wireshark-labs/HTTP-wireshark-file3.html				
What HTTP version is used by the Client?	1.1				
HTTP Headers (Following the Request Line) Questions:	Answer (Header Name: Value Pair from Wireshark)				
What is the hostname?	Gaia.cs.umass.edu				
What is the Client connection management preference?	keep-alive				
What is the Client browser and OS information?	Mozilla/5.0 (X11; Linux i686) AppleWebKit/537.36 (KHTML, like				

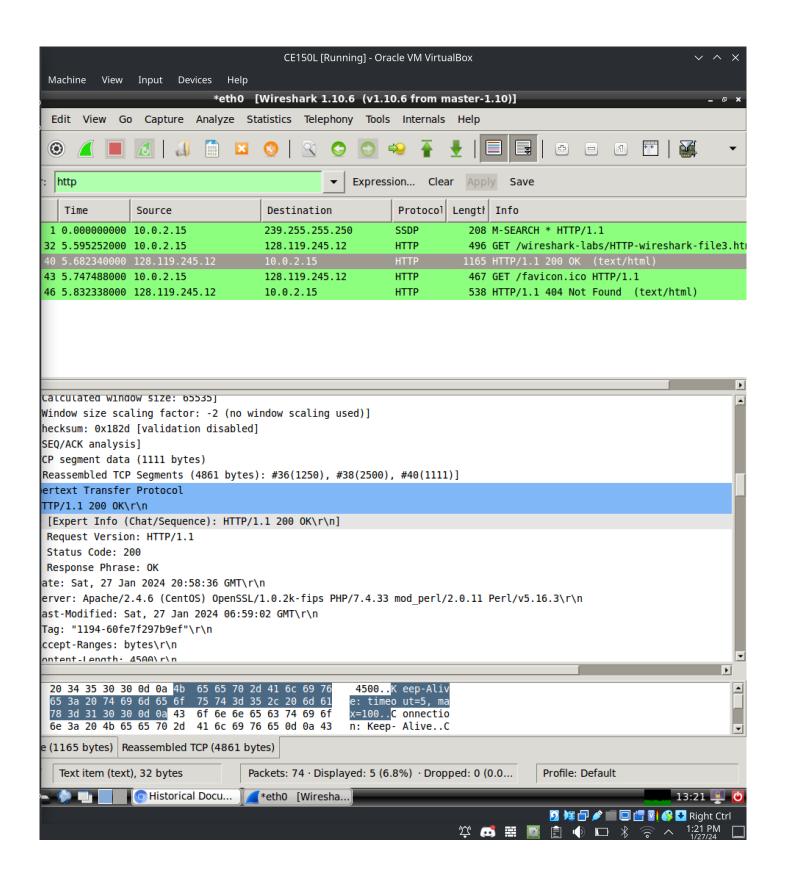
	Gecko) Ubuntu Chromium/63.0.3239.84 Chrome/63.0.3239.84 Safari/537.36	
I What types of content are accepted by the Citebra	text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8	
What compressed data does the Client accept?	gzip, deflate	

Include a timestamped screenshot, and circle the client IP Address and port number.

c. [10 points] Table 3: In Wireshark, use the Server Response Packet to answer the following. All answers are visible in the packet.

Network and Transport Layer (TCP/IP)						
Questions:	Answer (Field Detail from Wireshark)					
What is the Server IP address?	128.119.245.12					
What port number is the Server listening on?	80					
Application Layer (HTTP)						
Status Line (Initial Line of HTTP Response) Questions:	Answer (Elements of the Status Line from Wireshark)					
What HTTP version is used by the Server?	1.1					
What is the code that indicates the result of the Client's request?	200					
What is the status phrase that accompanies the above code?	OK					
HTTP Headers (Following the Status Line) Questions:	Answer (Header Name: Value pair from Wireshark)					
What is the date and time of the server response?	Sat, 27 Jan 2024 20:58:36 GMT					
What is the server software information?	Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips PHP/7.4.33 mod_perl/2.0.11 Perl/v5.16.3					
What is the length of the requested file?	Content length: 4500					
What response does the Server provide regarding the management of the network connection?	Keep alive					
What type of content is being sent by the Server?	text/html; charset=UTF-8					

Include a **timestamped screenshot** and circle the date.



d. [5 points] Protocol Layers:

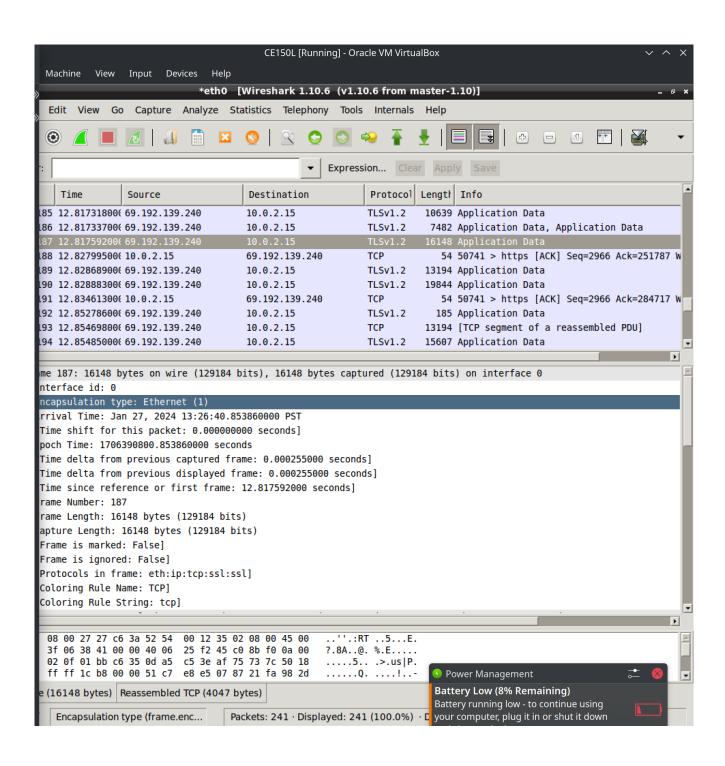
For the following problem, here's a reference to HTTPS: https://datatracker.ietf.org/doc/html/rfc2818

Now do a new Wireshark packet capture for "https://www.wellsfargo.com/" without using the https://www.wellsfargo.com/" without using the https://www.wellsfargo.com/"

In the Wireshark captures of both "https://www.wellsfargo.com/" and "http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file3.html", examine the different layers (e.g., application, transport) of network protocols.

- How do the application and transport layers interact differently in HTTP and HTTPS? One big thing is that https communicates over port 443 while http operates at port 80, second is https uses http it does not operate it's own protocol rather it just secures http
- If the bank transmitted a password over HTTP, would it be visible in a
 Wireshark capture? If a password was shared through http you would be able
 to capture It. How does this visibility differ if transmitted when using HTTPS?
- Https scrambles everything sent through it which makes it very difficult to unscramble the password without knowing the pattern that the local browser and website made.
- List all the protocols that make these two captures **different**. Focus on identifying protocols that contribute to security in the HTTPS capture.
- Https uses tls to secure it, it has http/2

Include a **timestamped screenshot** and circle at least one of the protocols listed.



2. [12 pts] Experiment with the wget command

Start by reading up on the wget command. Determine which command option to use with wget so that the <u>server returns only the HTTP headers and not the actual</u> <u>requested file.</u> Use the option to load this URL: <u>www.httpforever.com</u>

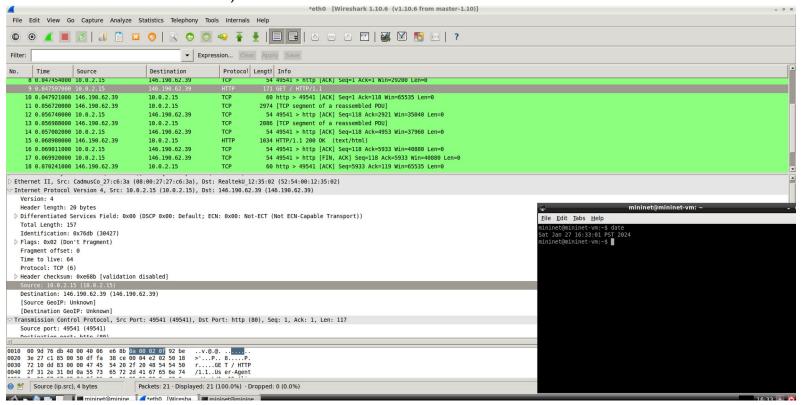
Run Wireshark to capture the packets exchanged during the wget execution.

- a. What is the full command you used?
- b. wget --server-response www.httpforever.com

<u>Use the waet terminal output</u> to answer the following questions:

- c. Server Response
 - Does the server have the requested file? yes How do you know? Next to OK it says text
 - When was the web page last modified?Last-Modified: Wed, 22 Mar 2023 14:54:48
 GMT\r\n How is this information used? Most likely if you cached parts of the website it can check if it was edited and if you need to pull all the info again
 - What type of *HTTP connection* is used? Keep ALive Explain what it means. It uses one tcp connection to send and receive info
 - What is the size (in bytes) of the base HTML page?56

Include a <u>single timestamped screenshot</u> and circle each question above (there should be 4 items circled).



Now look at your Wireshark capture:

Apply a filter in Wireshark and find the HTTP packet from your client to the web server.

Right-click on that packet and select "Follow TCP Stream". You should be able to see all the packets transferred on that TCP connection.

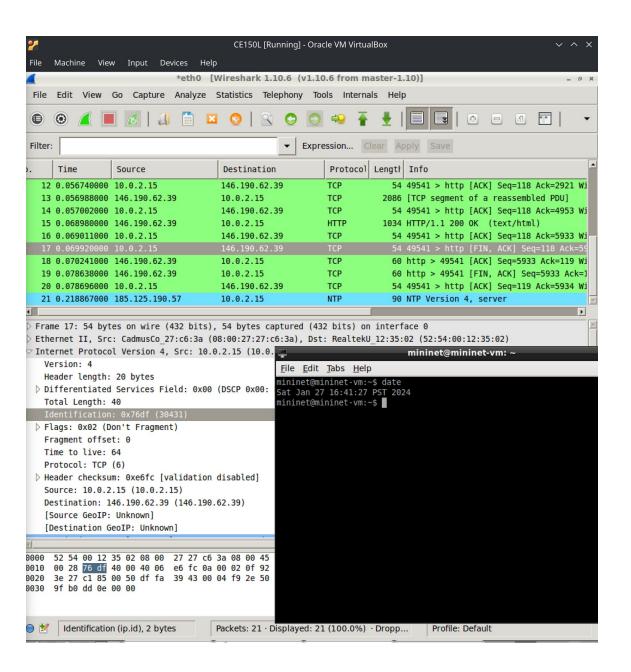
d. Client Request:

- According to Wireshark, which HTTP method is used for the client request? GETIs this what you expected? Yes because Im only requesting headers Why or why not?
- Explain the HTTP method you observe.
- o Get, It request the info from the server

e. Server:

- What is the status code of the response message from the server?
 What does it mean? 200 it means OK
- On what port number is the server listening for HTTP requests?
 Explain why this port is being used.80 becase it is an http request
- Which Transport layer protocol is used for these requests?
- tcp
- f. Explain the concept of a TCP stream in the context of loading a webpage.
- g. The TCP stream allows a two way communication between client and server that can allow data and images to travel safely between the two

Include a <u>timestamped screenshot</u> of the Follow TCP Stream with circles around fields that were used in the above guestions/answers.



3. [16 pts] Looking at Embedded Objects

Let's look at what happens when your browser downloads a file with references to several embedded objects, i.e., a file that includes references to other objects, accessible by hyperlink, that are stored on a server(s).

Do the following:

- Enter http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file4.html in your browser to display a short HTML file with two embedded images.
- Start a Wireshark capture.
- These two images are referenced in the base HTML file. The images themselves
 are not contained in the HTML file; instead, the URLs for the images are
 referenced using HTML. As discussed in the textbook (and class), your browser
 retrieves these objects from the indicated websites.
- Stop the Wireshark packet capture and filter for http packets.
- Ignore the /favicon.ico request/response in Wireshark.

a. [10 points] Table 4:

Find the response packet associated with the first GET request in Wireshark.				
Questions:	Answers:			
Describe the purpose of the very first GET request and the corresponding response message.	The Purpose of the first get and response was to get the html text to load the page, that being said images do not get sent in this			
Was it successful? Explain how you know.	Yes, because if you go to the html data in the captured package you can see the entire web page data			
If successful, how many bytes of the requested file does the Server send?	942			
Copy the lines from the base HTML page that reference the two images (Locate the raw data in this packet and look for the tags).	Reference 1: <img 2:<img="" \n="" height="82" reference="" src="http://gaia.cs.umass.edu/pearson.png" width="140"/> \n			
Enter each URL separately into your browser, what do you see?	URL 1: image of the book URL 2: the pearson logo			
From the Wireshark trace, how many HTTP GET request messages did your browser send in total? (Ignore /favicon.ico in Wireshark)	3 in totoal			

Embedded images: Using Wireshark, identify the two HTTP client requests for the embedded images and their corresponding server responses.				
	The 8e cover small jpg is hosted on a different site as you can see that the destination address if differentt			
Which image is hosted on a different server?	The 8e cover small jpg is hosted on a different site			

b. **[6 points]** Draw a timing diagram to show the sequence of interactions between the client and server when loading

"http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file4.html".

Your diagram should include:

- The client (your web browser), one server hosting the HTML file, and the other server that hosts one of the embedded images.
- The initial GET request for the HTML file and the server's response.
- Subsequent GET requests for each embedded image and the server response.
- Use arrows to indicate the direction of each request and response.
- There should be a clear order of interactions, labeling each step to show the progression from the initial request to the complete rendering of the web page with all embedded images.

4. [7 pts] HTTP Conditional GET

Recall from Section 2.2.5 that most web browsers perform object caching and thus often perform a conditional GET when retrieving an HTTP object. Before performing the steps below, make sure your browser's cache is empty^[2].

Now do the following (follow instructions closely):

- Start up your web browser, and make sure your browser's cache is cleared.
- Start up Wireshark and listen on the eth0 interface.

1st HTTP GET:

Enter this URL into your browser:

http://gaia.cs.umass.edu/wireshark-labs/HTTP-wireshark-file2.html

Your browser should display a very simple five-line HTML file.

2nd HTTP GET:

Quickly enter the same URL into your browser again (or simply select the refresh button on your browser)

• Stop the Wireshark packet capture and filter to display **HTTP** packets. (*Ignore the /favicon.ico request/response in Wireshark*.)

Questions:

- a. Why might your browser issue a Conditional GET?
- b. If it is not sure if it has updated since it cached the data,

1st HTTP GET and server response:

c. Is the first HTTP GET request from your browser to the server a Conditional GET? Explain your answer referencing Wireshark. No because its the first time it is caching the data as we cleared it earlier (there is no previous get request). However the second should be.

2nd HTTP GET and server response:

- d. Now inspect the contents of the second HTTP GET request from your browser to the server.
 - Describe how the second HTTP GET message is different from the first one.
 - It responds with a 304 message meaning nothin has been modified
 - What is the "IF-MODIFIED-SINCE:" header and its value, and why is it used?
 - It compares your cached data with its the If-modified-since data and makes the decision on whether to use cached or get new data
 - What is the date and its value, what does it mean? It means the date of the get request

Part B - Measuring Performance in Mininet

In this section, we continue our study of network performance and focus on throughput. We make a distinction between transmission rate (often observed at the sender) and throughput. Throughput is the rate data is delivered to the receiver and is specifically observed at the Application Layer.

Note: Remember that rates are measured in bits per second and when the numbers are too large we use notation such as Gpbs (gigabits per second) and Mbps (megabits per second). Typically these rates are stated in bits (not bytes). $G = 10^9$ and $M = 10^6$

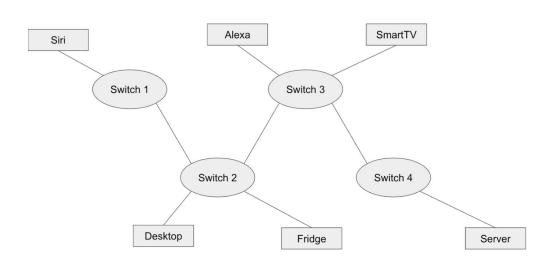


Figure1:

5. [30 pts] Exploring Transmission Rate

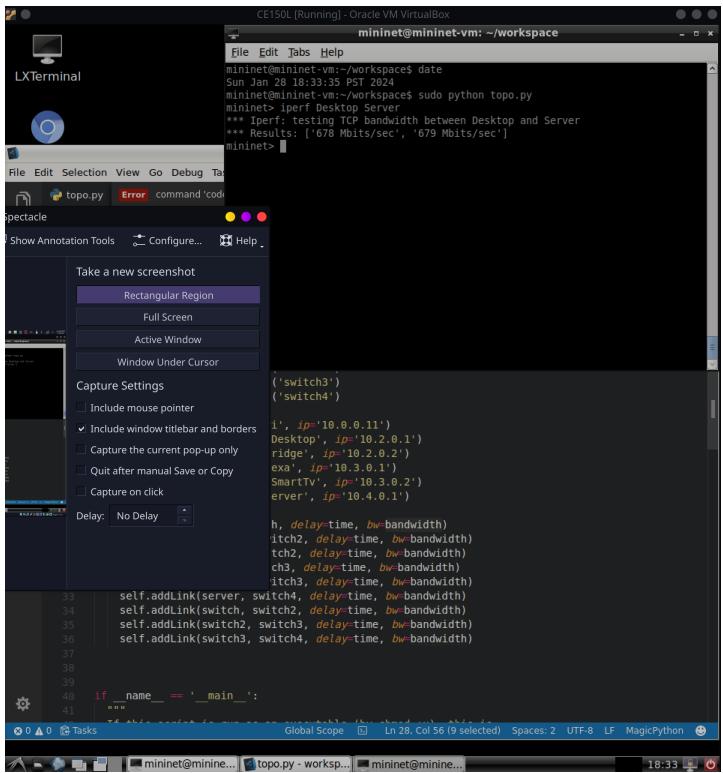
a. Research the **iperf** tool and describe it in your own words.

Using **iperf** on your **topology from Lab1 (Figure 1)**:

b. <u>Restricted bandwidth:</u> The bandwidth of the links can be altered by adding a bw= parameter to addLink (which takes its argument as a number representing the bandwidth in Mbps). Set the bandwidth of all links to 1 Gbps.

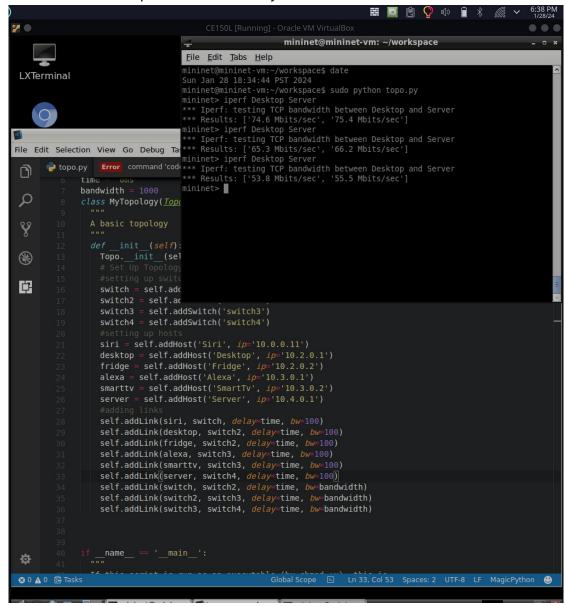
Run the iperf command between the Desktop and Server in the Mininet command line and screenshot the output. Explain the results. What is the meaning of the rates displayed and are they what you expected?

Include a **timestamped screenshot** of the results.



c. <u>Changing link bandwidth</u>: Change the bandwidth of edge (host-to-switch) links to **100 Mbps**. Run **iperf** 3 times between the Server and Desktop and find the average throughput seen by the Desktop. Are your results what you expected? Explain.

Average is 65.6, yes I expected this, we limited bandwith between all devices to routers as 100 mbps so there is no way for it to exceed said bandwith.



Include a **timestamped screenshot** of the results.

d. <u>Cross-traffic:</u> We will now generate simultaneous traffic between Desktop-Server and SmartTV-Server. Configure "Server" as an iperf server using iperf -s. Then generate iperf traffic from Desktop and SmartTv for 5 seconds as clients using:

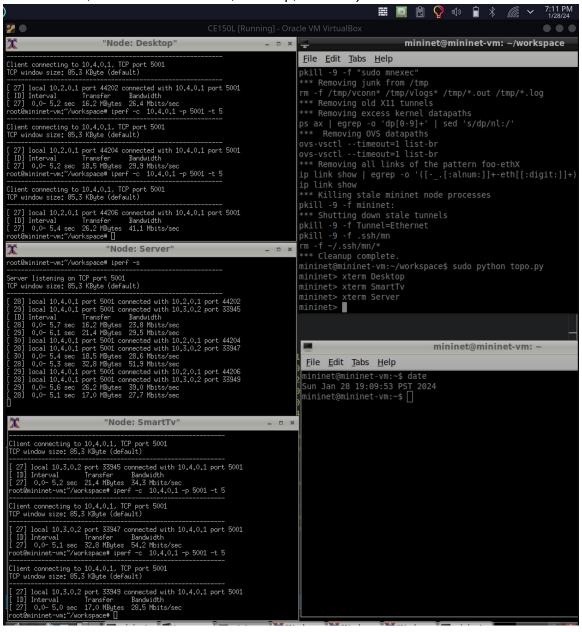
```
iperf -c <server IP> -p <server port> -t <seconds>
```

Be sure to run both iperf -c commands simultaneously. Repeat 3 times and

average the results.

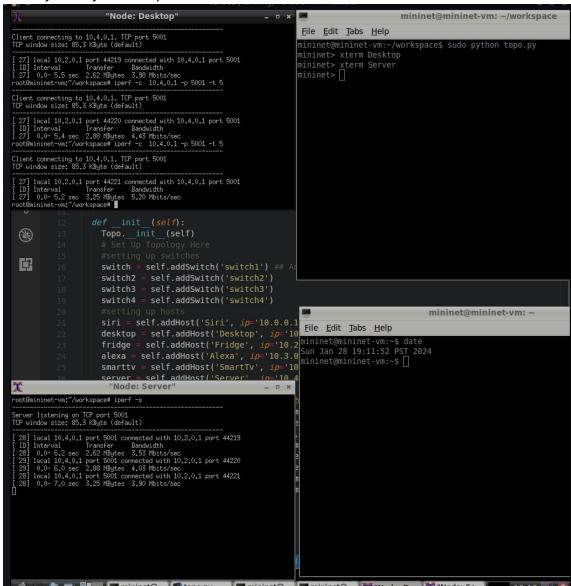
Which links were shared by the two traffic flows? Explain what you expected to happen. Discuss your observations after running iperf and if they differ from your initial expectation.

Include a <u>timestamped screenshot</u> of the results (this should have 4 terminal windows, 1 mininet, and 3 Xterm - Server, Desktop, SmartTv).



e. <u>The bottleneck link:</u> In this problem, we will observe the network with a bottleneck link. Change the bandwidth between Switch2-Switch3 to **5 Mbps**. Now run **iperf** between Server and Desktop 3 times and take the average of the results.

Discuss your results and how they have changed. Are your results what you expected? Why or why not? Explain.



Average is 4.5mbps and yes this is expected as even though the bandwith between the switch and devices are 100 mbps limiting the switches still slows it down.

Take a <u>timestamped screenshot</u> and circle the lowest value for the bandwidth observed.

f. Make a table comparing the output in 5(b), 5(c), 5(d), and 5(e). Discuss your results.

Problem	Bandwidth (s2-s3)	Bandwidth (host-switch)	Source of Data (Desktop/SmartTv)	Destination (Server)	Throughput	Rate at Sender
5(b)	1 Gbps	1 Gbps	Desktop	Server	678	679

5(c)	1 Gbps	100 Mbps	Desktop	Server	64.57	65.7
5(d)	1 Gbps	100 Mbps	Desktop	Server	32.47	30.47
5(d)	1 Gbps	100 Mbps	SmartTv	Server	39	36.37
5(e)	5 Mbps	100 Mbps	Desktop	Server	4.54	3.82

Screenshots Table:

Timestamped Screenshot Check-off: (just for you to keep track of your progress)

Problem #	Screenshot		
1b, c, d			
2b, e			
5b, c, d, e			

Footnotes:

- [1] References to figures and sections are for the 8th edition of our text, *Computer Networks, A Top-down Approach*, 8th ed., *J.F. Kurose and K.W. Ross, Addison-Wesley/Pearson, 2020.*
- ^[2] See https://www.howtogeek.com/304218/how-to-clear-your-history-in-any-browser/ for instructions on clearing your browser cache.
- ^[3] Hint: ideally, you should see an If-Modified-Since header since you've just downloaded this page a few seconds ago. However, depending on the browser you're using, and the format of the server's earlier response to your initial GET, your browser may not include an If-Modified-Since even if the document has been downloaded and cached. The Chrome browser is pretty good at regularly using If-Modified-Since. But Safari and Firefox are much more finicky about when to use If-Modified-Since. Life isn't always as easy in practice as it is in theory!