Lab 1: Network Fundamentals and Cloud Service Measurement

1. Objectives

- Master the network measurement tool: WireShark (Get Wireshark)
- Understand Internet traffic characteristics
- Understand the operations of a typical cloud-based storage service: Dropbox & Google Drive
- Compare Dropbox and Google Drive

2. Equipment Needs

- Computers
- Internet access

3. Experiments

3.1 Campus network traffic measurement

1) Download WireShark and study how to use it to capture packets, in particular, how to set filters to capture certain traffic or use built-in functions to display certain packets.
2) Go to NYU-Tandon cafeteria to capture packets in the wireless environment. Repeat the measurement in the morning, noon, and afternoon. For each time, capture the packets continuously for 10 minutes. (If you cannot make it to the NYU-Tandon campus, you can choose other public places with wifi access and decent amount of wifi users)
3) Analyze each measurement result and provide the following statistics.

Table 1 Campus Network Traffic Measurement

	Morning	Noon	Afternoon

Total number of packets captured	**6710**	**64020**	**62784**
Total number of bytes captured	**4391185**	**54208914**	**57148307**
Percentage of broadcast packets in packet numbers	127/6710 = **1.8927%**	131/64020 = **0.2046%**	131/62784 = **0.2087%**
Percentage of broadcast packets in bytes	21k/4391185 = **0.4782%**	22k/54208914 = **0.0406%**	21k/57148307 = **0.0368%**
Percentage of packets with transmission errors in packet numbers	0 errors = **0%** But 21 warnings including TCP windows full.	24/64020 = **0.0375%**	4/62784 = **0.0064%**

Question 1: How do you count the number of broadcast packets?

Answer 1: **By filtering with "eth.addr == ff:ff:ff:ff:ff:ff"**
In "Statistcs → Conversations", we look into the eth.addr B == ff:ff:ff:ff:ff:ff, which is the broadcast packets. On the same row, it shows packet number and bytes.

Ethernet · 9 | IPv4 · 65 | IPv6 · 3 | TCP · 67 | UDP · 173

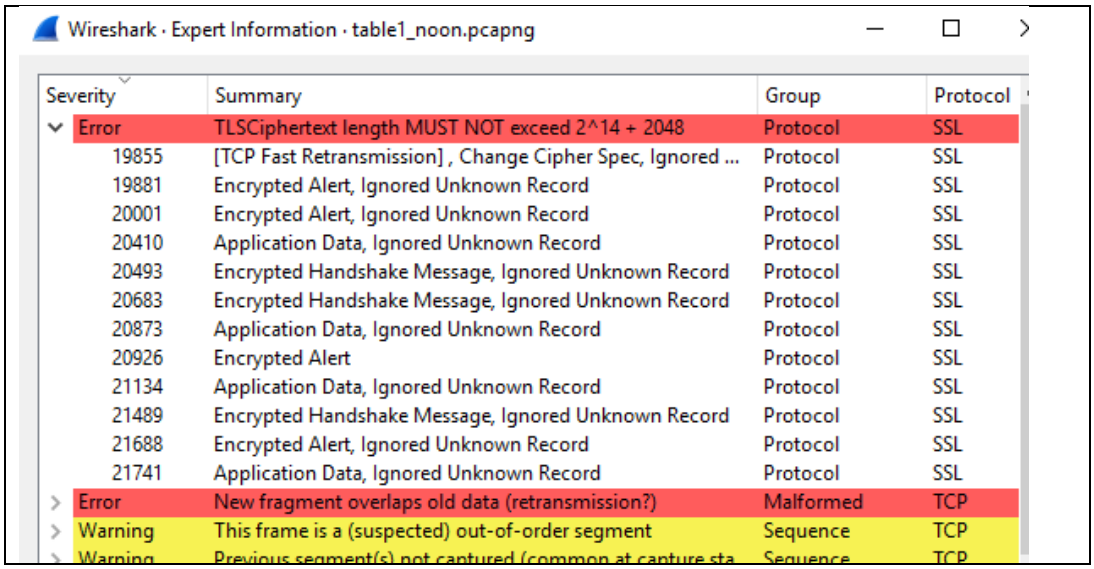
Address A	Address B	Packets	Bytes	Packets A → B	Bytes A → B	Packets B → A	Bytes B → A
00:00:5e:00:01:32	94:db:c9:41:3b:29	24,646	11 M	0	0	24,646	11 M
01:00:5e:00:00:fb	94:db:c9:41:3b:29	3	210	0	0	3	210
01:00:5e:00:00:fc	94:db:c9:41:3b:29	8	512	0	0	8	512
01:00:5e:7f:ff:fa	94:db:c9:41:3b:29	27	9206	0	0	27	9206
24:01:c7:25:4f:40	94:db:c9:41:3b:29	39,187	43 M	39,187	43 M	0	0
33:33:00:00:00:0c	94:db:c9:41:3b:29	7	5026	0	0	7	5026
33:33:00:00:00:fb	94:db:c9:41:3b:29	3	270	0	0	3	270
33:33:00:01:00:03	94:db:c9:41:3b:29	8	672	0	0	8	672
94:db:c9:41:3b:29	ff:ff:ff:ff:ff:ff	131	22 k	131	22 k	0	0

Question 2: How do you decide if a packet belongs to a transmission error?

Answer 2:
In Wireshark, it tells if a packet is error or not, for example, retransmission.

```
> Transmission Control Protocol, Src Port: 443, Dst Port: 52024, Seq: 1458535, Ack: 5895, Len: 90
v [Reassembly error, protocol TCP: New fragment overlaps old data (retransmission?)]
  v [Expert Info (Error/Malformed): New fragment overlaps old data (retransmission?)]
      [New fragment overlaps old data (retransmission?)]
      [Severity level: Error]
      [Group: Malformed]
```

To count the number of errors, we use "Analyze" → "Expert Information". It lists all the error packets.



3.2 Dropbox traffic measurement

1) Create a Dropbox account, and install the client software
2) Use Wireshark to capture the packets between the Dropbox client software and the cloud during the synchronization process (i.e., sync a file to dropbox), and understand the steps that the Dropbox client software takes to exchange data with the cloud.
3) Based on your measurement, fill out the following table to list ALL the servers the Dropbox client software interacted with. The order must be the same as what you observed.

Table 2 Dropbox-Cloud Interactions

Server domain name	Server IP address	Server's function	Amount of Traffic Exchanged
client.dropbox-dns.com	162.125.4.3	**Main contact window for clients' requests (forward requests, DNS to other service)**	180kB
d-sjc.v.dropbox.com	162.125.32.135	**Authorization** (User login)	19kB
bolt.v.dropbox.com	162.125.18.133	**Authorization** (Certificate, key exchange) **Open new session** (New session ticket)	17kB
block-debug.x.dropbox.com	52.73.220.123	**Detect errors on unsynchronized storage block**	57kB
api.dropbox-dns.com	162.125.4.7	**Handle API functions (to deal with unsynchronized storage block)**	9815B

edge-block-client-1.dropbox-dns.com	162.125.4.4	**File transfer/Storage**	11MB
Questions 1: How do you make sure the above servers/IPs are used for Dropbox but not other applications on your computer?			
Answer 1: **Firstly, the domain contained "dropbox". Moreover, before the dropbox starts, we records the existing flows. Therefore, we can be sure that these flows (servers/IPs) are not used for Dropbox. By filtering those IPs, we are sure that these IPs are for Dropbox.**			
Questions 2: How do you decide the function of each server contacted?			
Answer 2: **For the storage server, we upload a 10MB file to see which server received huge amount of application data.** **"client.dropbox-dns.com" is the very first one when dropbox starts and transfers application whenever other connection starts. Therefore, it handle where to pass the client's requests.** **After dropbox starts, we login to dropbox. Thus, we think that "d-sjc.v.dropbox.com" provide username/password login service.** **Then, "bolt.v.dropbox.com" exchanges the key and provides a new session ticket, so this should be the authorization services for other services.** **Next, we drag a 10 MB file into the folder, and "block-debug.x.dropbox.com" detects the unsynchronized errors. Then, some api is triggered (api.dropbox-dns.com) and starts to handle the unsynchronized storage. Then it triggers the upload to the storage server.**			

3.3 Comparison between Dropbox and Google drive

1) Place a computer A to one subnet, run Dropbox, then operate a computer B in a different subnet.
2) Create a shared Dropbox folder between computer A and computer B
3) Create a file in the shared folder on computer A, Dropbox should upload it to the Dropbox cloud, and mark the time to start upload as T1. Then wait for the file to be automatically downloaded to computer B, and mark the time the download is complete as T2.
4) Try different files with different sizes with Dropbox and repeat the same for Google drive, and fill out the following table. (Highly recommended to use text files)

Table 3 Comparison 1: Dropbox and GDrive

	Using Dropbox		Using Google drive	
Files	Time Consumed	Bytes Uploaded	Time Consumed	Bytes Uploaded
File-a: 1 MB	51.5515 - 47.0616 = **4.4899 sec**	1157k	37.5735-16.6377 = **20.9358 sec**	1099k
File-b: 10 MB	59.0848 - 49.0618 = **10.0230 sec**	12M	71.4062 - 34.4111 = **36.9954 sec**	10M
File-c: 100 MB	107.3291 – 5.3255 = **102.0035 sec**	117M	147.6797 – 40.7544 = **106.9253 sec**	110M
Question: How do you measure T1 and T2 to reach high accuracy?	Answer: **Firstly, we use "statistics → conversations" to double check which ip is the storage server.** **After getting the IP, we use flow graph and the IP to trace the file transfer flow. Then, we can find the first packet of the file transfer flow.** **We pick the first "Application data" as the first packet, and get the UTC**			

	time of the packet as T1. (We use UTC time to synchronize the time of two computers.) **Next, on computer B, we find the last ack packet to the "Application data" as the final packet of download flow. Again, we get the UTC time as T2.** **With T1 and T2, we can have accurate consumed time (with clock is automatically synchronized by OS).**
Question: Is there any difference between number of bytes uploaded for Dropbox and Google Drive? If so, why?	Answer: **Yes, the numbers of bytes uploaded for Dropbox and Google Drive are different. We think that Dropbox updates additional information of the file to the cloud, so that in Table 4, the duplication can be done on cloud. On the other hand, Google Drive upload only the file itself, so it don't have additional information to identify if the file is directly duplicated from an exist file. Thus, in Table 4, it needs to upload the full size of the duplicate file.**

4) Now copy a file, File-C in the above table, to create a duplicate file. It will be uploaded to dropbox. Repeat the operation and let it be uploaded to Google drive. Measure the number of bytes being uploaded in each operation. Fill out the following table.

Table 4 Comparison 2: Dropbox and GDrive

# of bytes uploaded to Dropbox	6071
# of bytes uploaded to Google drive	110MB
Why there is a difference between the above two numbers?	Answer: **When we duplicate the file on Google drive, it treats the duplicated file as a new file. Thus, it upload the full size of the file.** **On the other hand, when we duplicate the file on Dropbox, it detects that the file is duplicated from a exist file. Thus, only few commands are transferred to update the duplicated file on cloud server.** **Therefore, by uploading the duplicated file (Google drive) and by uploading some commands (Dropbox), the numbers of uploaded bytes are different.**

4. Reports

Please include your name, student ID, and the followings,

- Completed Tables 1-4.
- Wireshark capture files from each table.

We have zero tolerance to forged or fabricated data!! A single piece of forged/fabricated data would bring the total score down to zero.