File Transfer Experiment

Chapter 1: Recognition of and statement of the problem

Problem statement

The increasingly widespread availability of bandwidths enables the transfer of a large volume of file. Several data-intensive applications have their own specific requirements for data transfer services. These applications vary from bulk data transfer to high throughput interactive system. File transfer has become a crucial mean for transferring data between two hosts. Its file transfer systems rely on transport-level protocol, namely TCP and UDP, to move data from their source to their destination. In the real world, some applications can use standard File Transfer Protocol (FTP) that is TCP-based Data Transfer. However, there still are some applications that TCP-based Data Transfer cannot support well but UDP-based Data transfer can. Therefore, we construct this experiment to find out which protocol is better and what factors affect the throughput in the case of downloading.

Objectives

1. To find out which main factor or interaction of factors are the most influential on the response of interest. This phase is called “Factor screening and characteristic”.
2. To find the levels of important factors that result in desirable response values after the system has been characterized. This phase is called “Optimization”.
3. To find the best recommendation to obtain the best requirement for File Transfer.

Note: we assess with statistical significance level (α) compare to 95%.

Chapter II: Selection of the response variable

The response of File Transfer is measured by using verbose option of either FTP command or TFTP command. The value is shown in the terminal after the whole operation is done.

*Note:

Response Variable	Unit	Measurement Precision / Accuracy	Relationship to Objective
Throughput	MB/s	Measured by using verbose option of either FTP command or TFTP command.	The greater value of throughput, the less time use in file transfer.

Chapter III: Choice of factors levels and range

Factors

Potential design factors classification

Design factors	Held-constant factors	Allowed-to-vary factors
File size	Command used to observe the throughput responses	
Transport protocol	Computer used	
Host location	Network SSID	
Type of connection (last hop)		
Time of day		

Nuisance factors classification

Controllable factors	Uncontrollable factors	Noise factors
	Network congestion (e.g. BitTorrent)	

Factor levels and range

Due to our main objective is related to factor screening, we decide to take each factor in two levels (low level and high level) with broad range.

Factor	Level	
	Low (-)	**High (+)**
File size	Small (100 KB)	Large (5 MB)
Transport protocol	TCP (ftp)	UDP (tftp)
Host location	In-campus	Out-of-campus
Type of connection (last hop)	LAN	Wireless
Time of day	Afternoon (1 PM)	Evening (6 PM)

Factor details

Control variable

Control Variables	Measurement Precision / Accuracy	Predicted Effects
File size	Size of file that transfer across the network. We use 100 KB file (The maximum recommended total of all the elements on a single web page) as a small file and 5 MB file (size of three-minute MP3 audio at a very high bitrate (256kpbs)) as a large file.	Transferred file size should not affected throughput of the network.
Transport protocol	Use two standard transport protocols which are TCP and UDP. We use *ftp* command for TCP protocol and *tftp* command for UDP protocol.	TCP protocol should give higher throughput response.
Host location	We use server in CAST LAB as the in-campus host and use a virtual machine on Microsoft Azure as the out-of-campus host.	In-campus server should give higher throughput response.

Type of connection (last hop)	We use two types of the last hop connection which are LAN and Wireless connection.	LAN connection should give higher throughput response.
Time of day	We measure the response in 2 different times which are 1 PM as afternoon period and 6 PM as evening period.	The throughput response in the evening should higher that in the afternoon.

Held Constant Factors

Held Constant Factors	Measurement Precision / Accuracy	Settings	Predicted Effects
Command used to observe the throughput responses	We use the same command to transfer files using TCP and UDP protocol.	Use command *get <file>* to transfer *<file>*	None
Computer used	Use the same computer to measure all responses	N/A	None
Network SSID	Use the same network SSID to gather all responses.	Use *KMUTT-Secure* network	Negligible

Nuisance Factors

Nuisance Factors	Measurement Precision / Accuracy	Settings	Predicted Effect
Network congestion (e.g. BitTorrent)	Network congestion can change the throughput response. To mitigate the effects of network congestion, all samples were ran in the same time.	N/A	Negligible

Chapter IV: Experimental design

As we will study the effect of five factors and their interactions on a response and each of them has only two levels, two-level fractional design is appropriate for our experiment. Hence, we run full factorial experiment of two-level fractional design with two replications. We performed totally 32 runs with 16 different treatment combinations and 2 replications for each combination. The detail of the experiment are described in the following section.

Design the experiment

The factors tested in this experiment are as follows:

Variable	Name	Low level (-)	High level (+)
A	File size	Small (100 KB)	Large (5 MB)
B	Transport protocol	TCP (ftp)	UDP (tftp)
C	Host location	In-campus	Out-of-campus
D	Type of connection (last hop)	LAN	Wireless
E	Time of day	Afternoon (1 PM)	Evening (6 PM)

Table 4.1: The 2^5 fractional factorial design

Treatment combination	Factor				
	A	B	C	D	E
1	-	-	-	-	-
2	+	-	-	-	-
3	-	+	-	-	-
4	+	+	-	-	-
5	-	-	+	-	-
6	+	-	+	-	-
7	-	+	+	-	-
8	+	+	+	-	-

9	-	-	-	+	-
10	+	-	-	+	-
11	-	+	-	+	-
12	+	+	-	+	-
13	-	-	+	+	-
14	+	-	+	+	-
15	-	+	+	+	-
16	+	+	+	+	-
17	-	-	-	-	+
18	+	-	-	-	+
19	-	+	-	-	+
20	+	+	-	-	+
21	-	-	+	-	+
22	+	-	+	-	+
23	-	+	+	-	+
24	+	+	+	-	+
25	-	-	-	+	+
26	+	-	-	+	+
27	-	+	-	+	+
28	+	+	-	+	+
29	-	-	+	+	+
30	+	-	+	+	+
31	-	+	+	+	+
32	+	+	+	+	+

Equipment

Equipment	Quantity	
1. iMac for downloading files from local area and remote area	1	unit
2. Ethernet cable (LAN)	2	meters
3. Local server	1	unit
4. Virtual machine on Microsoft Azure	1	unit
5. Files 5.1 Small file (100 KB) 5.2 Large file (5 MB)	 1 1	 file file

Chapter V: Performing the experimental

Procedure

1. Prepare all materials.
2. Setup the local server
   - 2.1 Install vsftp
     - 2.1.1 Enter *sudo apt-get install vsftpd* into command line.
     - 2.1.2 Open up configuration file of vsftpd and configure as follows:

       anonymous_enable = NO
       local_enable = YES
       write_enable = YES
       chroot_local_user = YES
     - 2.1.3 Restart *vsftpd* service with this following command: *sudo service vsftpd restart*
   - 2.2 Install tftp
     - 2.2.1 Enter *sudo apt-get install xinetd tftpd* into command line.
     - 2.2.2 Create /etc/xinetd.d/tftp and put this entry.

       service tftp
       {
       protocol = udp
       port = 69
       socket_type = dgram
       wait = yes
       user = root
       server = /usr/sbin/in.tftpd
       server_args = -s /tftpboot
       disable = no
       }
     - 2.2.3 Create a folder */tftpboot* , which match in server_args of previous step.

       sudo mkdir /tftpboot
       sudo chmod -R 777 /tftpboot
     - 2.2.4 Restart *xinetd* service with this following command: *sudo service xinetd restart*
3. Repeat step 2 with the virtual machine on Microsoft Azure.
4. Perform the experiment in the afternoon period starting from 1:00 PM.
5. Use ethernet cable (LAN) to connect the local network.
6. Type *ftp 10.35.30.76* through command line and then provide the authorized FTP credentials.
7. Use TCP protocol to transfer small file with the following commands
   - 7.1 *verbose*
   - 7.2 *get small_file*
8. Observe the throughput response.
9. Repeat step 7-8 to observe the 2nd replication.
10. Repeat step 7-9 using the large file instead of the small file.
11. Type *tftp 10.35.30.76* through command line.
12. Repeat step 7-10 to observe the response from UDP protocol.

13. Repeat step 6-12 with the out-of-campus server on Microsoft Azure.
14. Disconnect LAN wire and use wireless to connect local network.
15. Repeat step 6-12 with the wireless connection.
16. Repeat the experiment in the evening period starting from 6:00 PM.
17. Collect result of each sample and record to experimental result
18. Analyze the result and draw a conclusion of this experiment.

Experimental results

Run	Factor					Throughput (MB/s)	
	A	B	C	D	E	Replication I	Replication II
1	-	-	-	-	-	7.92723	8.294236
2	+	-	-	-	-	12.740198	11.890852
3	-	+	-	-	-	0.341333	0.512
4	+	+	-	-	-	0.400219875	0.361577875
5	-	-	+	-	-	0.670228	0.63556608
6	+	-	+	-	-	1.83501	1.688207
7	-	+	+	-	-	0.0140274	0.013837875
8	+	+	+	-	-	0.014368	0.013357625
9	-	-	-	+	-	2.06569	1.65675
10	+	-	-	+	-	2.013266	1.88744
11	-	+	-	+	-	0.146286	0.128
12	+	+	-	+	-	0.1648705	0.145232125
13	-	-	+	+	-	0.481741	0.50899968
14	+	-	+	+	-	0.834478	1.04858
15	-	+	+	+	-	0.0131282	0.01422225
16	+	+	+	+	-	0.011103125	0.010619625
17	-	-	-	-	+	7.59169	7.853834

18	+	-	-	-	+	12.006195	12.16348
19	-	+	-	-	+	0.341333	0.341333
20	+	+	-	-	+	0.4032985	0.4032985
21	-	-	+	-	+	0.52606	0.52957184
22	+	-	+	-	+	1.68821	1.772093
23	-	+	+	-	+	0.0152836	0.016
24	+	+	+	-	+	0.014644875	0.01457975
25	-	-	-	+	+	1.95035	2.495611
26	+	-	-	+	+	2.93601	2.327839
27	-	+	-	+	+	0.170667	0.2048
28	+	+	-	+	+	0.16181725	0.1814145
29	-	-	+	+	+	0.481137	0.49475584
30	+	-	+	+	+	1.25829	1.174405
31	-	+	+	+	+	0.0144225	0.014027375
32	+	+	+	+	+	0.01389575	0.010435625

Average throughput response = 1.8446 MB/s

Chapter VI: Statistical analysis and conclusion

After collecting data from the experiment, we use Minitab software package to check the model adequacy, analyze the data, and perform the residual analysis. Then, we will formulate the empirical model to describe the relationship between the important factors and the response variable.

Firstly, we perform full factorial of a two-level fractional experiment. The results and the model gathered from Minitab shown below.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	31	687.527	22.178	740.71	0.000
Linear	5	383.668	76.734	2562.73	0.000
A	1	9.864	9.864	329.42	0.000
B	1	184.969	184.969	6177.54	0.000
C	1	116.540	116.540	3892.17	0.000
D	1	72.277	72.277	2413.90	0.000
E	1	0.019	0.019	0.62	0.438
2-Way Interactions	10	237.068	23.707	791.75	0.000
A*B	1	9.826	9.826	328.16	0.000
A*C	1	1.962	1.962	65.54	0.000
A*D	1	5.317	5.317	177.58	0.000
A*E	1	0.054	0.054	1.79	0.190
B*C	1	95.023	95.023	3173.56	0.000
B*D	1	64.766	64.766	2163.03	0.000
B*E	1	0.017	0.017	0.58	0.452
C*D	1	59.790	59.790	1996.86	0.000
C*E	1	0.006	0.006	0.21	0.653
D*E	1	0.307	0.307	10.24	0.003
3-Way Interactions	10	63.703	6.370	212.75	0.000
A*B*C	1	1.929	1.929	64.42	0.000
A*B*D	1	5.275	5.275	176.18	0.000
A*B*E	1	0.038	0.038	1.27	0.268
A*C*D	1	3.013	3.013	100.62	0.000
A*C*E	1	0.002	0.002	0.06	0.802
A*D*E	1	0.003	0.003	0.09	0.762
B*C*D	1	53.081	53.081	1772.77	0.000
B*C*E	1	0.006	0.006	0.21	0.653
B*D*E	1	0.239	0.239	7.99	0.008
C*D*E	1	0.117	0.117	3.91	0.057
4-Way Interactions	5	3.088	0.618	20.63	0.000
A*B*C*D	1	2.994	2.994	100.00	0.000
A*B*C*E	1	0.000	0.000	0.00	0.969
A*B*D*E	1	0.015	0.015	0.50	0.483
A*C*D*E	1	0.002	0.002	0.06	0.808
B*C*D*E	1	0.077	0.077	2.56	0.119
5-Way Interactions	1	0.001	0.001	0.03	0.866
A*B*C*D*E	1	0.001	0.001	0.03	0.866
Error	32	0.958	0.030		
Total	63	688.486			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.173038	99.86%	99.73%	99.44%

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		1.8446	0.0216	85.28	0.000	
A	0.7852	0.3926	0.0216	18.15	0.000	1.00
B	-3.4001	-1.7000	0.0216	-78.60	0.000	1.00
C	-2.6988	-1.3494	0.0216	-62.39	0.000	1.00
D	-2.1254	-1.0627	0.0216	-49.13	0.000	1.00
E	0.0340	0.0170	0.0216	0.79	0.438	1.00
A*B	-0.7837	-0.3918	0.0216	-18.12	0.000	1.00
A*C	-0.3502	-0.1751	0.0216	-8.10	0.000	1.00
A*D	-0.5765	-0.2882	0.0216	-13.33	0.000	1.00
A*E	0.0579	0.0290	0.0216	1.34	0.190	1.00
B*C	2.4370	1.2185	0.0216	56.33	0.000	1.00
B*D	2.0119	1.0060	0.0216	46.51	0.000	1.00
B*E	-0.0329	-0.0165	0.0216	-0.76	0.452	1.00
C*D	1.9331	0.9666	0.0216	44.69	0.000	1.00
C*E	-0.0196	-0.0098	0.0216	-0.45	0.653	1.00
D*E	0.1385	0.0692	0.0216	3.20	0.003	1.00
A*B*C	0.3472	0.1736	0.0216	8.03	0.000	1.00
A*B*D	0.5742	0.2871	0.0216	13.27	0.000	1.00
A*B*E	-0.0487	-0.0244	0.0216	-1.13	0.268	1.00
A*C*D	0.4339	0.2170	0.0216	10.03	0.000	1.00
A*C*E	-0.0109	-0.0055	0.0216	-0.25	0.802	1.00
A*D*E	0.0132	0.0066	0.0216	0.30	0.762	1.00
B*C*D	-1.8214	-0.9107	0.0216	-42.10	0.000	1.00
B*C*E	0.0196	0.0098	0.0216	0.45	0.653	1.00
B*D*E	-0.1223	-0.0611	0.0216	-2.83	0.008	1.00
C*D*E	-0.0855	-0.0428	0.0216	-1.98	0.057	1.00
A*B*C*D	-0.4326	-0.2163	0.0216	-10.00	0.000	1.00
A*B*C*E	0.0017	0.0009	0.0216	0.04	0.969	1.00
A*B*D*E	-0.0307	-0.0153	0.0216	-0.71	0.483	1.00
A*C*D*E	0.0106	0.0053	0.0216	0.24	0.808	1.00
B*C*D*E	0.0692	0.0346	0.0216	1.60	0.119	1.00
A*B*C*D*E	0.0073	0.0037	0.0216	0.17	0.866	1.00

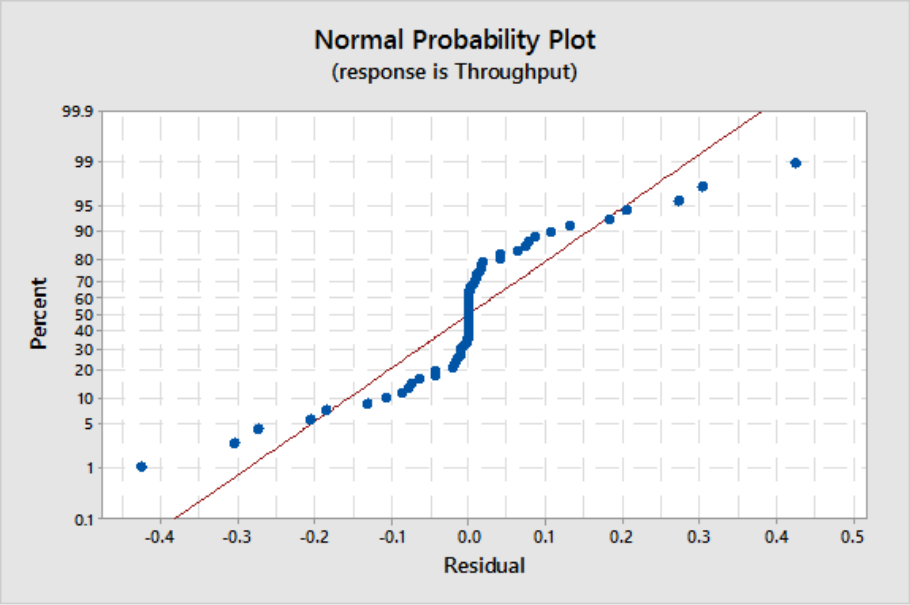
Regression Equation in Uncoded Units

$$\begin{aligned}\text{Throughput} = {} & 1.8446 + 0.3926\,A - 1.7000\,B - 1.3494\,C - 1.0627\,D + 0.0170\,E \\ & - 0.3918\,A*B - 0.1751\,A*C - 0.2882\,A*D + 0.0290\,A*E + 1.2185\,B*C \\ & + 1.0060\,B*D - 0.0165\,B*E + 0.9666\,C*D - 0.0098\,C*E + 0.0692\,D*E \\ & + 0.1736\,A*B*C + 0.2871\,A*B*D - 0.0244\,A*B*E + 0.2170\,A*C*D \\ & - 0.0055\,A*C*E + 0.0066\,A*D*E - 0.9107\,B*C*D + 0.0098\,B*C*E \\ & - 0.0611\,B*D*E - 0.0428\,C*D*E - 0.2163\,A*B*C*D + 0.0009\,A*B*C*E \\ & - 0.0153\,A*B*D*E + 0.0053\,A*C*D*E + 0.0346\,B*C*D*E \\ & + 0.0037\,A*B*C*D*E\end{aligned}$$

Checking the adequate assumptions

We have checked the model adequacy of the model above and we found that there are something unusual with the residual plot of the model as shown below.

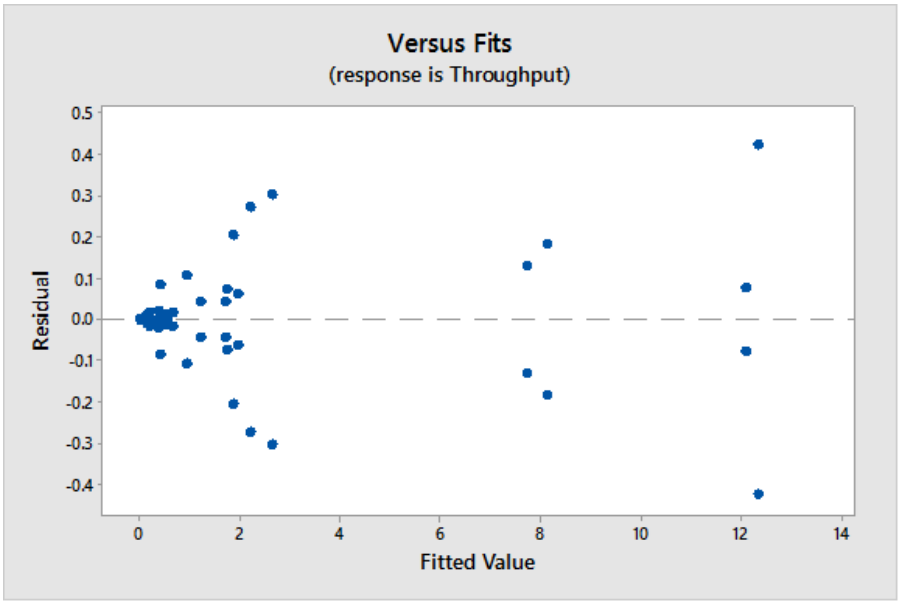
- Normal Probability Plot of Residuals for Throughput



From the graph above, we see that residuals plot are not fall along the straight line. Thus, there is something unusual with the model.

- Residuals vs Fits for Throughput

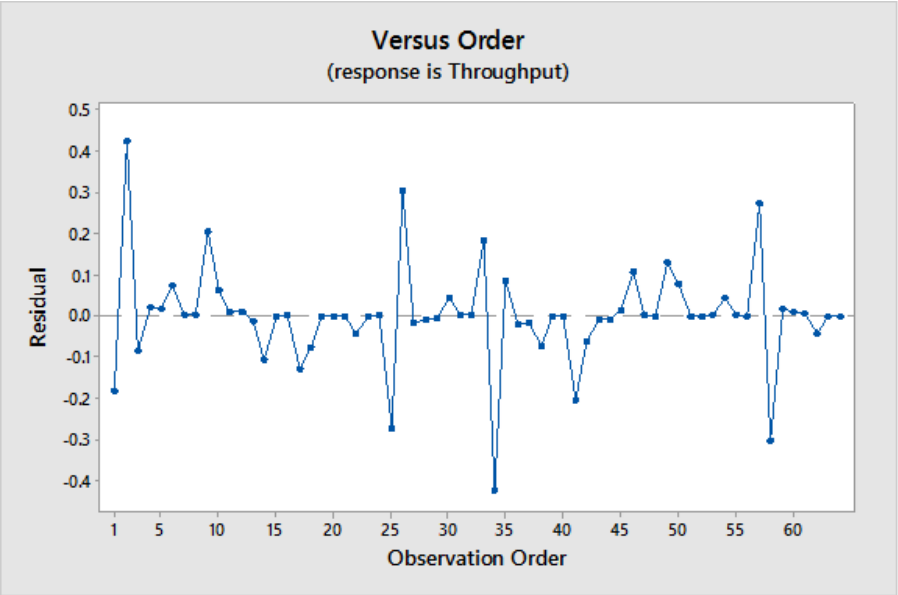
The variance of the observations in each treatment should be equal. We can check this assumption with Residuals versus fits plot.



From the graph above, the plot of residuals is not in a random pattern and reveal some recognizable patterns. Thus, the data are structured and it may related to other variables. That is, the constant variance assumption is unsatisfied.

- Residuals vs Order for Throughput

ANOVA requires that the observations should be randomly selected from the treatment population. They should not be affected by time, startup period or other observations. We can check this assumption by using Residuals versus order plot.



From the graph above, the plot of the residuals seem doesn't be in a random pattern. Thus, the independence assumption is unsatisfied.

Responses Transformation

After we check the model adequacy, we found that the model is unsatisfied. Hence, we transformed the responses using `Box-Cox transformation  λ = 0` or logarithm of the response. Then, we use ANOVA to analyze the experimental result by using Minitab program.

The ANOVA result is as follows:

Factorial Regression: Throughput versus A, B, C, D, E

```
Method

Box-Cox transformation  λ = 0

Analysis of Variance for Transformed Response

Source              DF   Adj SS   Adj MS   F-Value  P-Value
Model               31  307.057    9.905   1052.63    0.000
```

Linear	5	293.140	58.628	6230.50	0.000
A	1	1.240	1.240	131.74	0.000
B	1	199.318	199.318	21181.89	0.000
C	1	84.298	84.298	8958.49	0.000
D	1	8.222	8.222	873.73	0.000
E	1	0.063	0.063	6.67	0.015
2-Way Interactions	10	12.830	1.283	136.35	0.000
A*B	1	1.675	1.675	178.01	0.000
A*C	1	0.266	0.266	28.26	0.000
A*D	1	0.167	0.167	17.71	0.000
A*E	1	0.028	0.028	3.00	0.093
B*C	1	6.112	6.112	649.48	0.000
B*D	1	0.730	0.730	77.61	0.000
B*E	1	0.002	0.002	0.17	0.679
C*D	1	3.694	3.694	392.55	0.000
C*E	1	0.007	0.007	0.78	0.384
D*E	1	0.150	0.150	15.89	0.000
3-Way Interactions	10	0.995	0.099	10.57	0.000
A*B*C	1	0.647	0.647	68.81	0.000
A*B*D	1	0.052	0.052	5.54	0.025
A*B*E	1	0.020	0.020	2.14	0.153
A*C*D	1	0.005	0.005	0.49	0.487
A*C*E	1	0.005	0.005	0.50	0.485
A*D*E	1	0.002	0.002	0.21	0.651
B*C*D	1	0.212	0.212	22.53	0.000
B*C*E	1	0.011	0.011	1.21	0.279
B*D*E	1	0.014	0.014	1.54	0.224
C*D*E	1	0.026	0.026	2.75	0.107
4-Way Interactions	5	0.070	0.014	1.49	0.222
A*B*C*D	1	0.006	0.006	0.59	0.446
A*B*C*E	1	0.009	0.009	0.93	0.343
A*B*D*E	1	0.017	0.017	1.76	0.194
A*C*D*E	1	0.022	0.022	2.33	0.137
B*C*D*E	1	0.017	0.017	1.82	0.187
5-Way Interactions	1	0.023	0.023	2.40	0.131
A*B*C*D*E	1	0.023	0.023	2.40	0.131
Error	32	0.301	0.009		
Total	63	307.358			

Model Summary for Transformed Response

S	R-sq	R-sq(adj)	R-sq(pred)
0.0970042	99.90%	99.81%	99.61%

Coded Coefficients for Transformed Response

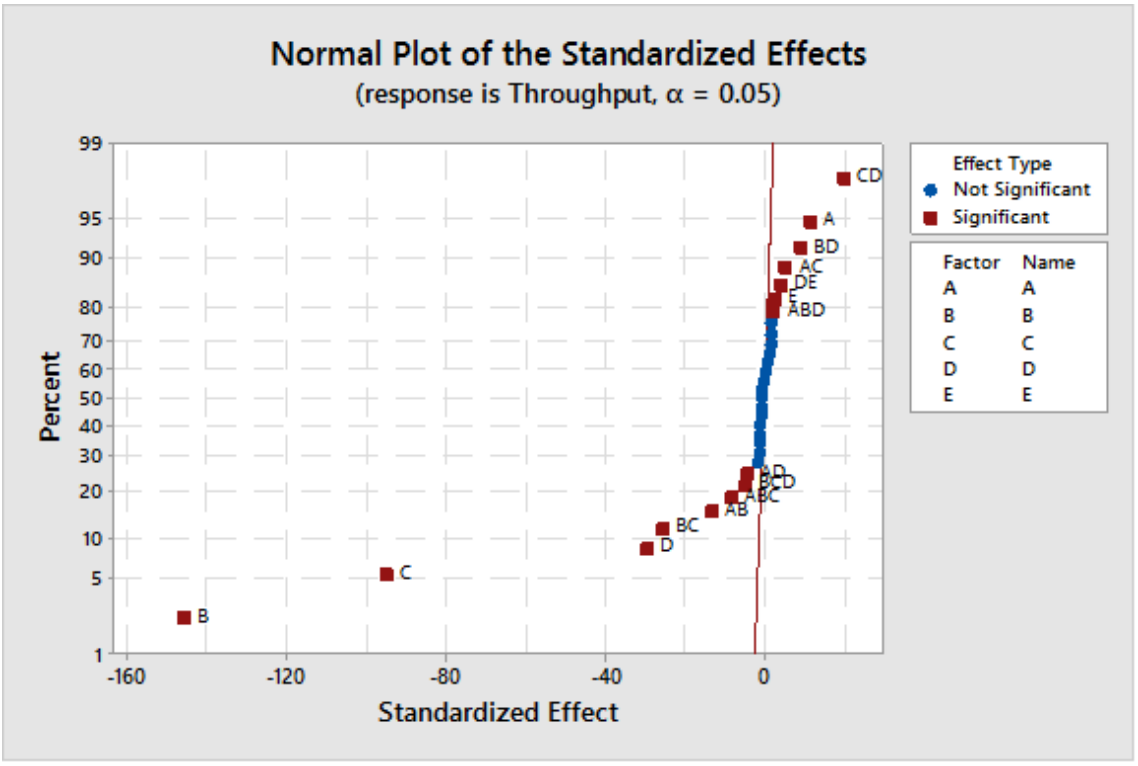
Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		-1.0816	0.0121	-89.20	0.000	
A	0.2784	0.1392	0.0121	11.48	0.000	1.00
B	-3.5295	-1.7647	0.0121	-145.54	0.000	1.00
C	-2.2953	-1.1477	0.0121	-94.65	0.000	1.00
D	-0.7168	-0.3584	0.0121	-29.56	0.000	1.00
E	0.0626	0.0313	0.0121	2.58	0.015	1.00
A*B	-0.3236	-0.1618	0.0121	-13.34	0.000	1.00
A*C	0.1289	0.0645	0.0121	5.32	0.000	1.00
A*D	-0.1020	-0.0510	0.0121	-4.21	0.000	1.00
A*E	0.0420	0.0210	0.0121	1.73	0.093	1.00
B*C	-0.6180	-0.3090	0.0121	-25.48	0.000	1.00
B*D	0.2136	0.1068	0.0121	8.81	0.000	1.00
B*E	0.0101	0.0051	0.0121	0.42	0.679	1.00
C*D	0.4805	0.2402	0.0121	19.81	0.000	1.00

C*E	-0.0214	-0.0107	0.0121	-0.88	0.384	1.00
D*E	0.0967	0.0483	0.0121	3.99	0.000	1.00
A*B*C	-0.2012	-0.1006	0.0121	-8.30	0.000	1.00
A*B*D	0.0571	0.0285	0.0121	2.35	0.025	1.00
A*B*E	-0.0355	-0.0177	0.0121	-1.46	0.153	1.00
A*C*D	-0.0171	-0.0085	0.0121	-0.70	0.487	1.00
A*C*E	0.0171	0.0086	0.0121	0.71	0.485	1.00
A*D*E	-0.0111	-0.0055	0.0121	-0.46	0.651	1.00
B*C*D	-0.1151	-0.0576	0.0121	-4.75	0.000	1.00
B*C*E	0.0267	0.0134	0.0121	1.10	0.279	1.00
B*D*E	-0.0300	-0.0150	0.0121	-1.24	0.224	1.00
C*D*E	-0.0402	-0.0201	0.0121	-1.66	0.107	1.00
A*B*C*D	-0.0187	-0.0094	0.0121	-0.77	0.446	1.00
A*B*C*E	-0.0233	-0.0117	0.0121	-0.96	0.343	1.00
A*B*D*E	-0.0322	-0.0161	0.0121	-1.33	0.194	1.00
A*C*D*E	0.0370	0.0185	0.0121	1.53	0.137	1.00
B*C*D*E	-0.0327	-0.0164	0.0121	-1.35	0.187	1.00
A*B*C*D*E	0.0376	0.0188	0.0121	1.55	0.131	1.00

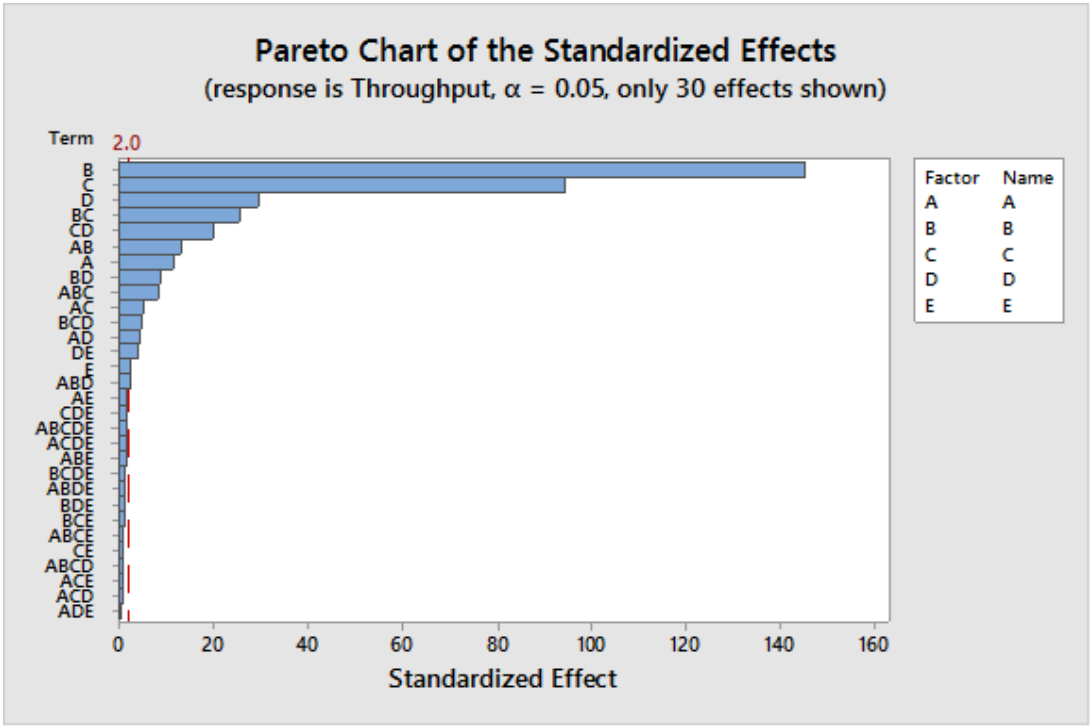
Regression Equation in Uncoded Units

$$\ln(\text{Throughput}) = -1.0816 + 0.1392\,A - 1.7647\,B - 1.1477\,C - 0.3584\,D + 0.0313\,E - 0.1618\,A*B + 0.0645\,A*C - 0.0510\,A*D + 0.0210\,A*E - 0.3090\,B*C + 0.1068\,B*D + 0.0051\,B*E + 0.2402\,C*D - 0.0107\,C*E + 0.0483\,D*E - 0.1006\,A*B*C + 0.0285\,A*B*D - 0.0177\,A*B*E - 0.0085\,A*C*D + 0.0086\,A*C*E - 0.0055\,A*D*E - 0.0576\,B*C*D + 0.0134\,B*C*E - 0.0150\,B*D*E - 0.0201\,C*D*E - 0.0094\,A*B*C*D - 0.0117\,A*B*C*E - 0.0161\,A*B*D*E + 0.0185\,A*C*D*E - 0.0164\,B*C*D*E + 0.0188\,A*B*C*D*E$$

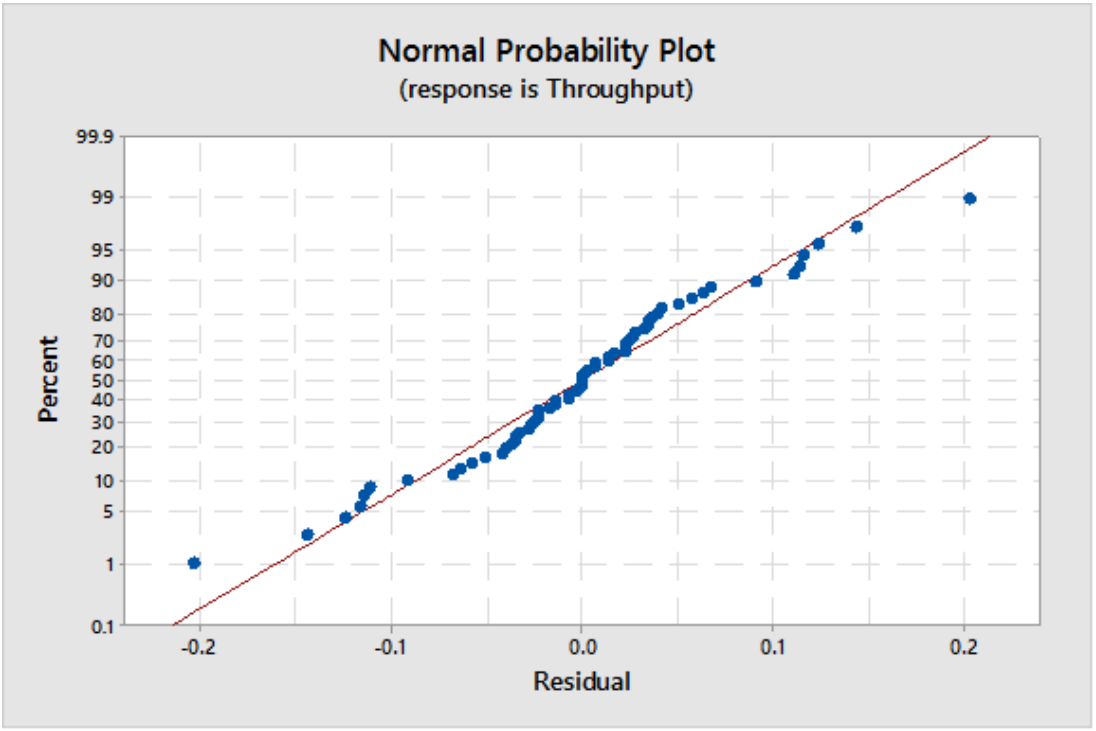
The normal plot of the standardize effects



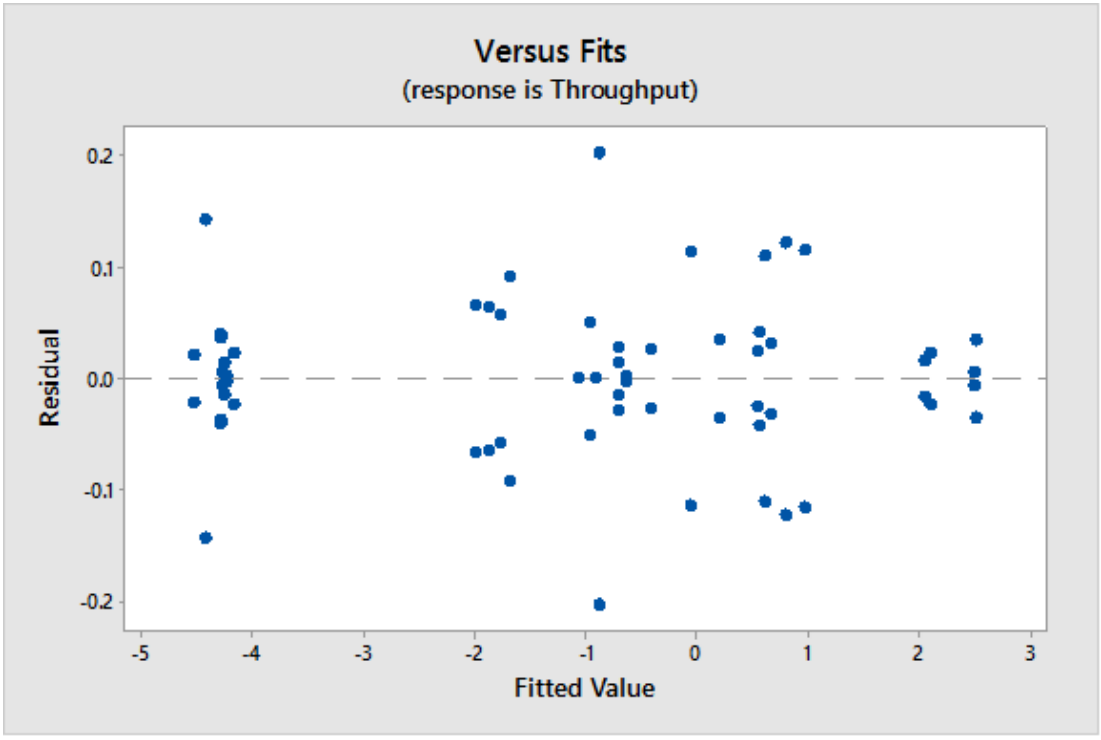
The Pareto chart of the standardize effects



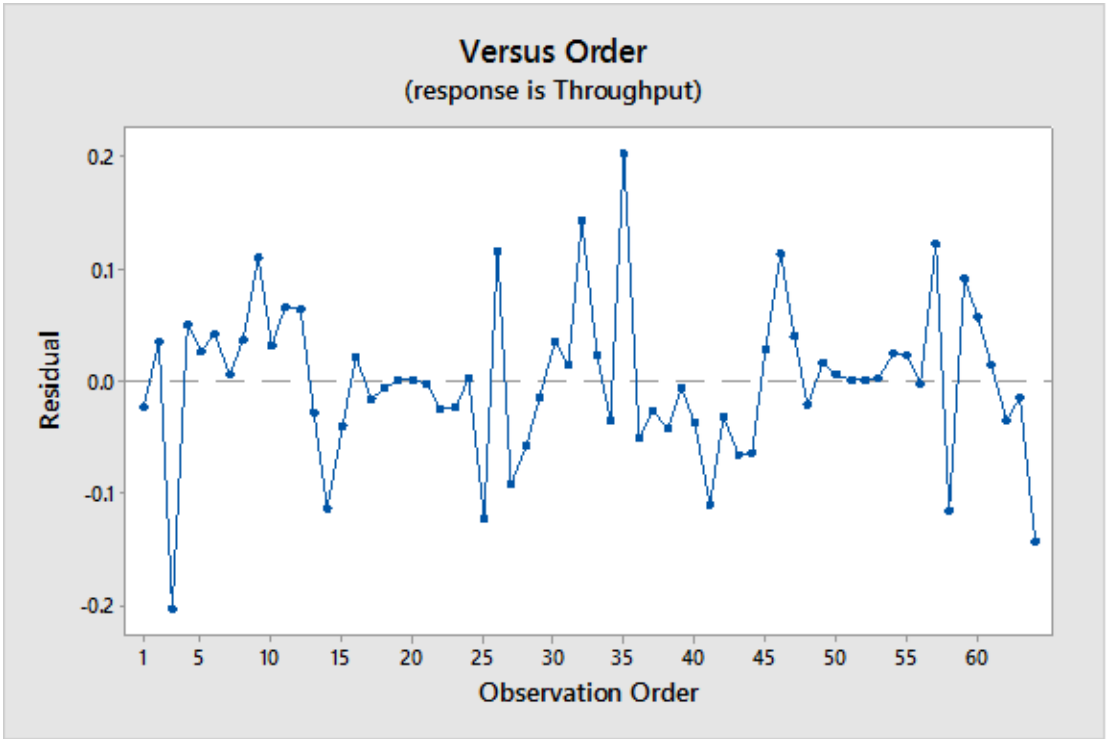
The normal probability plot of the residuals for throughput



Residuals vs Fits for Throughput



Residuals vs Order for Throughput



After we check the model adequacy, we found that it satisfies all of assumptions (i.e. Normality assumption, Equality of variance assumption, and Independence).

Analysis of the result

From the ANOVA result, there are five main effects, seven of second-order interaction effects and three of third-order interaction effect that are statistically significant at significant level = 0.05 as follows: (Note: Alternatively, the analysis can be considered from Pareto chart.)

- Main effect B: Transport protocol
- Main effect C: Host location
- Main effect D: Type of connection (last hop)
- Main effect A: File size
- Main effect E: Time of day
- Interaction effect BC: Transport protocol * Host location
- Interaction effect CD: Host location * Type of connection
- Interaction effect AB: File size * Transport protocol
- Interaction effect BD: Transport protocol * Type of connection
- Interaction effect AC: File size * Host location
- Interaction effect AD: File size * Type of connection
- Interaction effect DE: Type of connection * Time of day
- Interaction effect ABC: File size * Transport protocol * Host location
- Interaction effect BCD: Transport protocol * Host location * Type of connection
- Interaction effect ABD: File size * Transport protocol * Type of connection

There are several methods to analyze the factor effects to the response. First, we focus on the normal plot. This plot reveals whether factors affect to the response. The interpretation of this plot are as follows:

Position

Factors that fall in the left-hand side

- When the factor level is low, the response value is high.
- When the factor level is high, the response value is low.

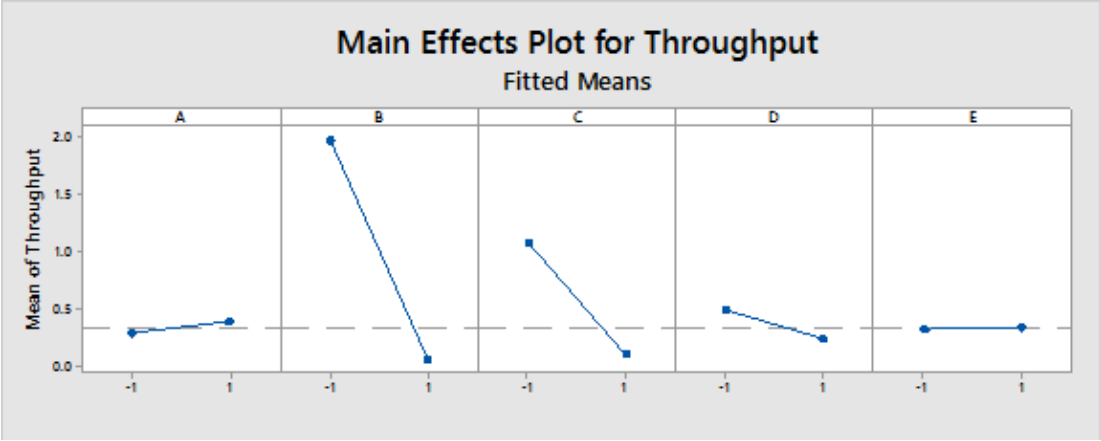
Factors that fall in the right-hand side

- When the factor level is low, the response value is low.
- When the factor level is high, the response value is high.

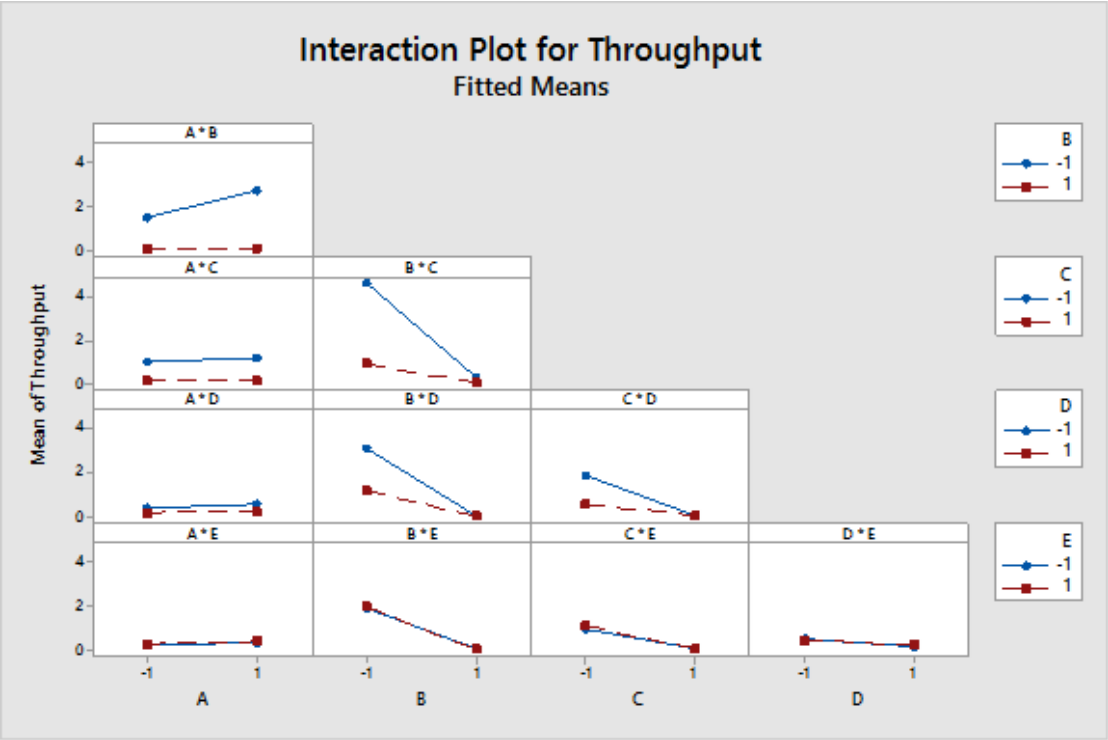
Distance

The farther factors from the normal line, the more strongly effect to the response. For example, the meaning of factor B (Transport protocol) that falls in the left-hand side and very far from the normal line is that when factor level is low, it strongly affects to the response in the way that gives high response.

There is another method to analyze how each factor effect to the response by plot the main or interaction effect against the response as follows:



The graph above is the plot of main effect, it shows the response value in each level of factors. If the slope of graph is high, that factor strongly affects to the response (significant).



The graph above also shows about the interaction effect.

- If the slope of two graphs are same (parallel), that interaction effect doesn't affect to the response.
- If their slopes are different, their interaction effect are significant.

Conclusion from the Normal plot and the plot of main effect give the same result as analysis of variance.

Analysis compare to hypothesis

According to factor detail section in Chapter III, we predicted the effect of each factor before performing the experiment. The table below shows a summary of comparison between predicted effects and actual effects from the experiment.

Control Variables	Predicted Effects	Actual Effects
File size	Transferred file size should not affected throughput of the network.	File size significantly affect the response. Large file would give higher throughput response.
Transport protocol	TCP protocol should give higher throughput response.	TCP protocol give higher throughput response.
Host location	In-campus server should give higher throughput response.	In-campus server give higher throughput response.
Type of connection (last hop)	LAN connection should give higher throughput response.	LAN connection give higher throughput response.
Time of day	The throughput response in the evening should higher that in the afternoon.	The throughput response in the evening higher that in the afternoon.

Chapter VII: Conclusion

Conclusion

The increasingly widespread availability of bandwidths enables the transfer of a large volume of file. Several data-intensive applications have their own specific requirements for data transfer services. These applications vary from bulk data transfer to high throughput interactive system. File transfer has become a crucial mean for transferring data between two hosts. Its file transfer systems rely on transport-level protocol, namely TCP and UDP, to move data from their source to their destination. In the real world, some applications can use standard File Transfer Protocol (FTP) that is TCP-based Data Transfer. However, there still are some applications that TCP-based Data Transfer cannot support well but UDP-based Data transfer can. Therefore, we construct this experiment to find out which protocol is better and what factors affect the throughput in the case of downloading.

From the experimental result, the best combination for that give highest throughput is to transfer large file from local server using TCP protocol with LAN connection. However, this experiment reveals that the time of day is almost not significantly difference. Thus, throughput response at the afternoon and evening period might not significantly difference.

The following table is the throughput response using the best factor combination.

File size	Transport protocol	Host location	Type of connection (last hop)	Time of day	Throughput (MB/s)
Large (5 MB)	TCP	Local server	LAN	Afternoon / Evening	12.74

We gain the knowledge from this experiment and it will be advantage for our daily life. That is, the throughput response depends mainly on the transport protocol, host location and type of connection (last hop). Thus, the right protocol and with the LAN connection used in file transfer will respond high throughput. Likewise, due to the experimental result shows that time of day which can have different number of connection might not significantly affect the throughput response.

In conclusion, there are both right and wrong suggestions about throughput response in the file transfer process. The result of our experiment clearly reveals that if you desires to obtain the highest throughput, we recommend you to use the best solution from the result of our experiment.

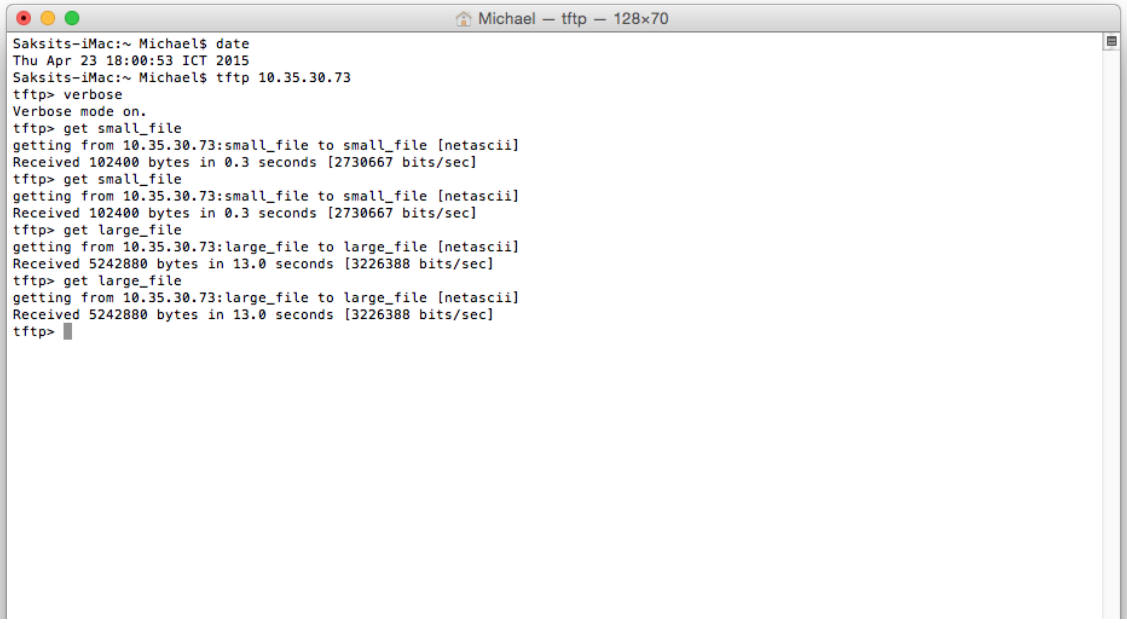
Chapter VIII: Appendix

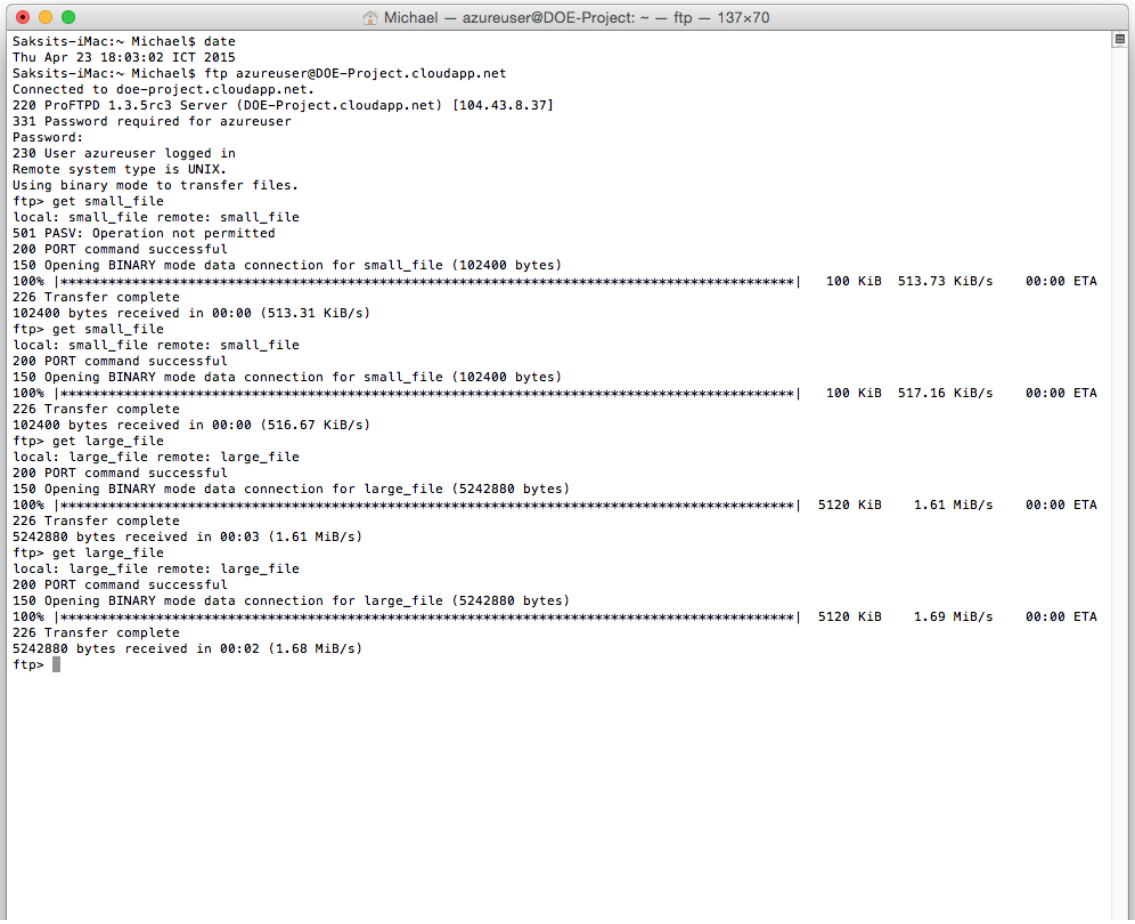
Experimental results (Throughput response value of 16 runs with 2 replications)

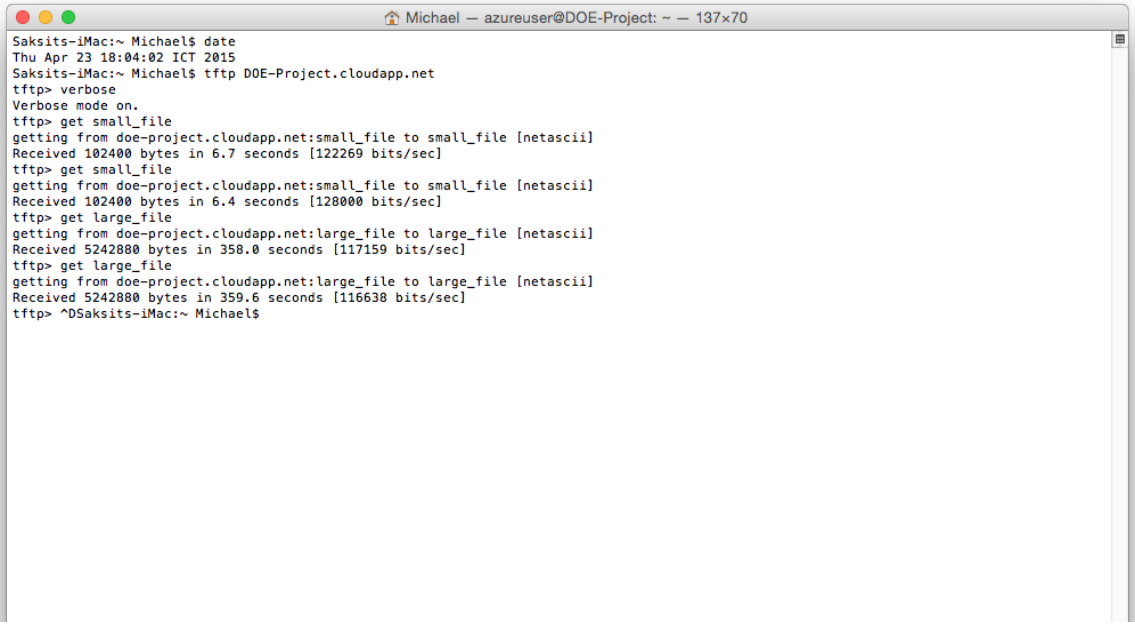
Run	Throughput (MB/s)	
	Replication I	Replication II
1	7.92723	8.294236
2	12.740198	11.890852
3	0.341333	0.512
4	0.400219875	0.361577875
5	0.670228	0.63556608
6	1.83501	1.688207
7	0.0140274	0.013837875
8	0.014368	0.013357625
9	2.06569	1.65675
10	2.013266	1.88744
11	0.146286	0.128
12	0.1648705	0.145232125
13	0.481741	0.50899968
14	0.834478	1.04858
15	0.0131282	0.01422225
16	0.011103125	0.010619625
17	7.59169	7.853834
18	12.006195	12.16348
19	0.341333	0.341333
20	0.4032985	0.4032985

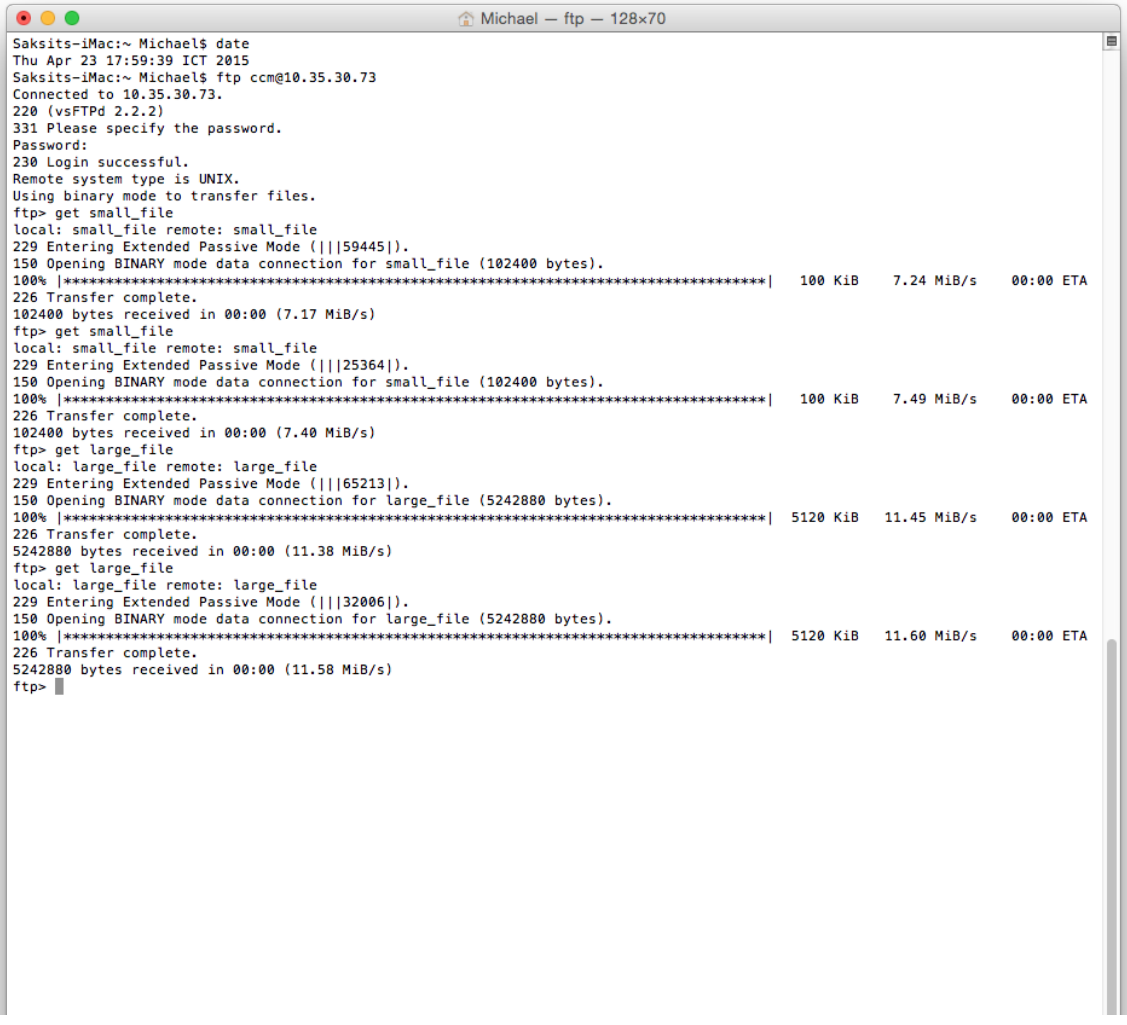
Run	Throughput (MB/s)	
	Replication I	Replication II
21	0.52606	0.52957184
22	1.68821	1.772093
23	0.0152836	0.016
24	0.014644875	0.01457975
25	1.95035	2.495611
26	2.93601	2.327839
27	0.170667	0.2048
28	0.16181725	0.1814145
29	0.481137	0.49475584
30	1.25829	1.174405
31	0.0144225	0.014027375
32	0.01389575	0.010435625

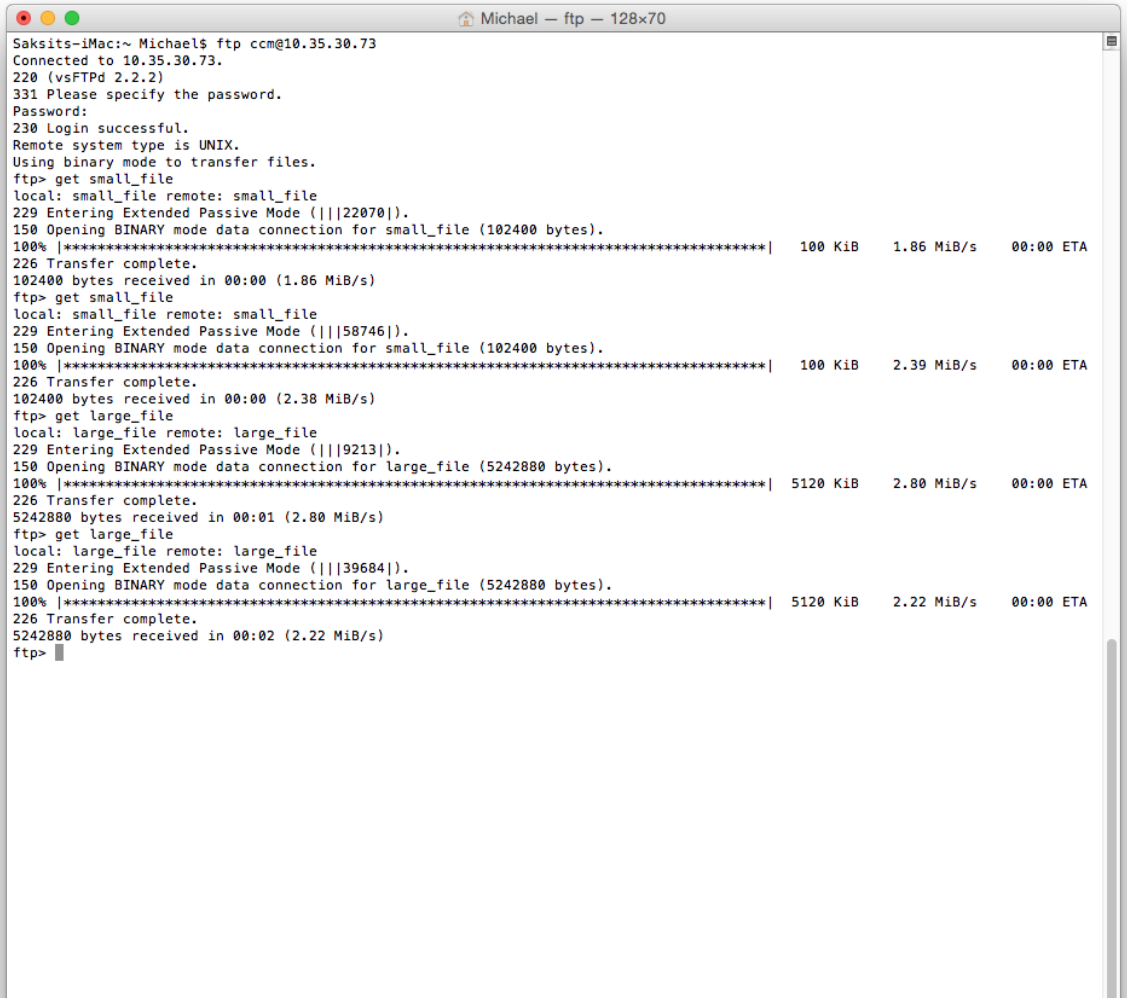
Examples of Terminal snapshot

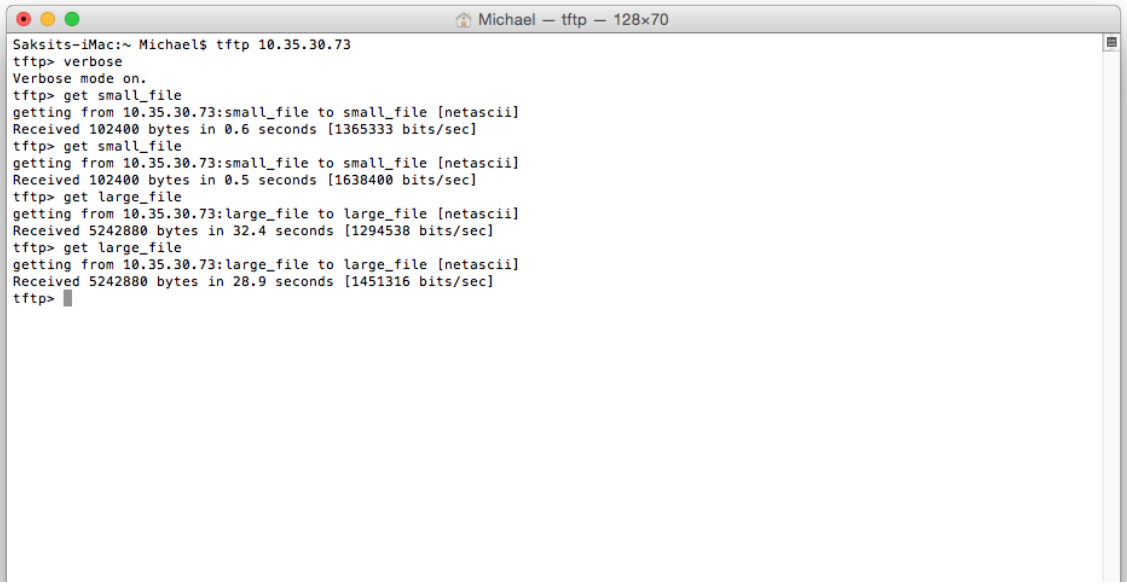


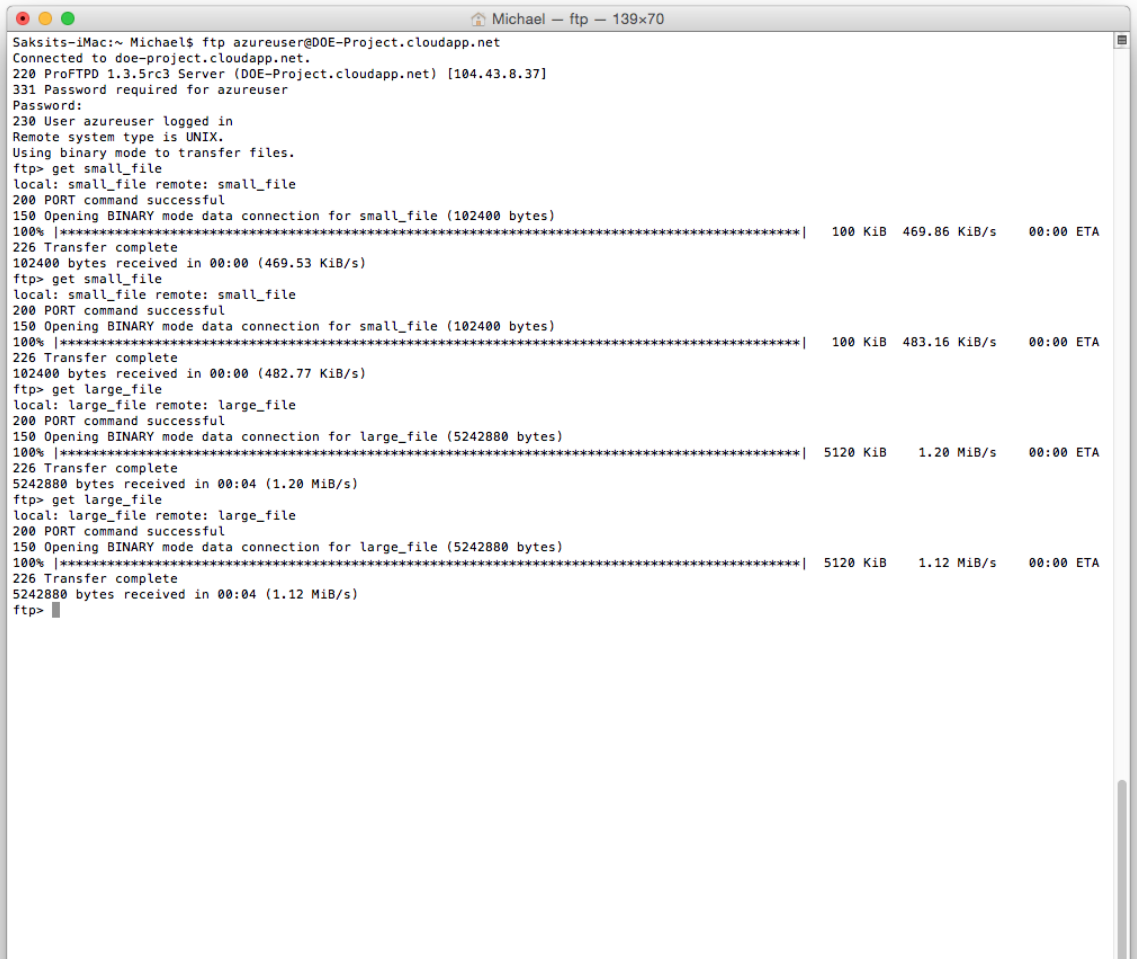


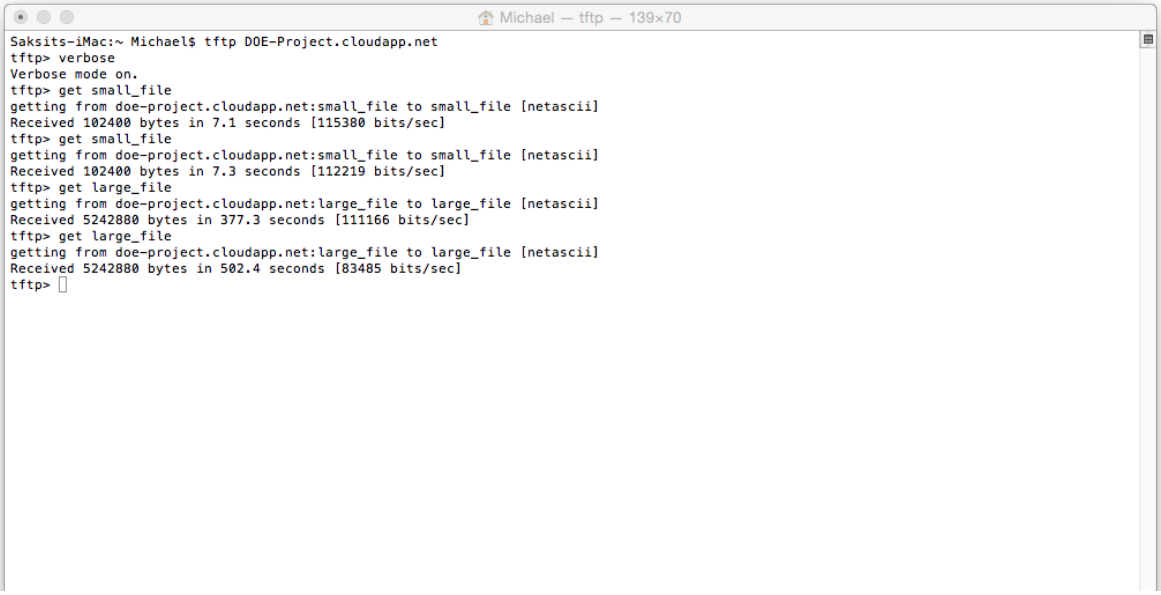

Michael — azureuser@DOE-Project: ~ — 137×70
Saksits-iMac:~ Michael$ date
Thu Apr 23 18:04:02 ICT 2015
Saksits-iMac:~ Michael$ tftp DOE-Project.cloudapp.net
tftp> verbose
Verbose mode on.
tftp> get small_file
getting from doe-project.cloudapp.net:small_file to small_file [netascii]
Received 102400 bytes in 6.7 seconds [122269 bits/sec]
tftp> get small_file
getting from doe-project.cloudapp.net:small_file to small_file [netascii]
Received 102400 bytes in 6.4 seconds [128000 bits/sec]
tftp> get large_file
getting from doe-project.cloudapp.net:large_file to large_file [netascii]
Received 5242880 bytes in 358.0 seconds [117159 bits/sec]
tftp> get large_file
getting from doe-project.cloudapp.net:large_file to large_file [netascii]
Received 5242880 bytes in 359.6 seconds [116638 bits/sec]
tftp> ^DSaksits-iMac:~ Michael$

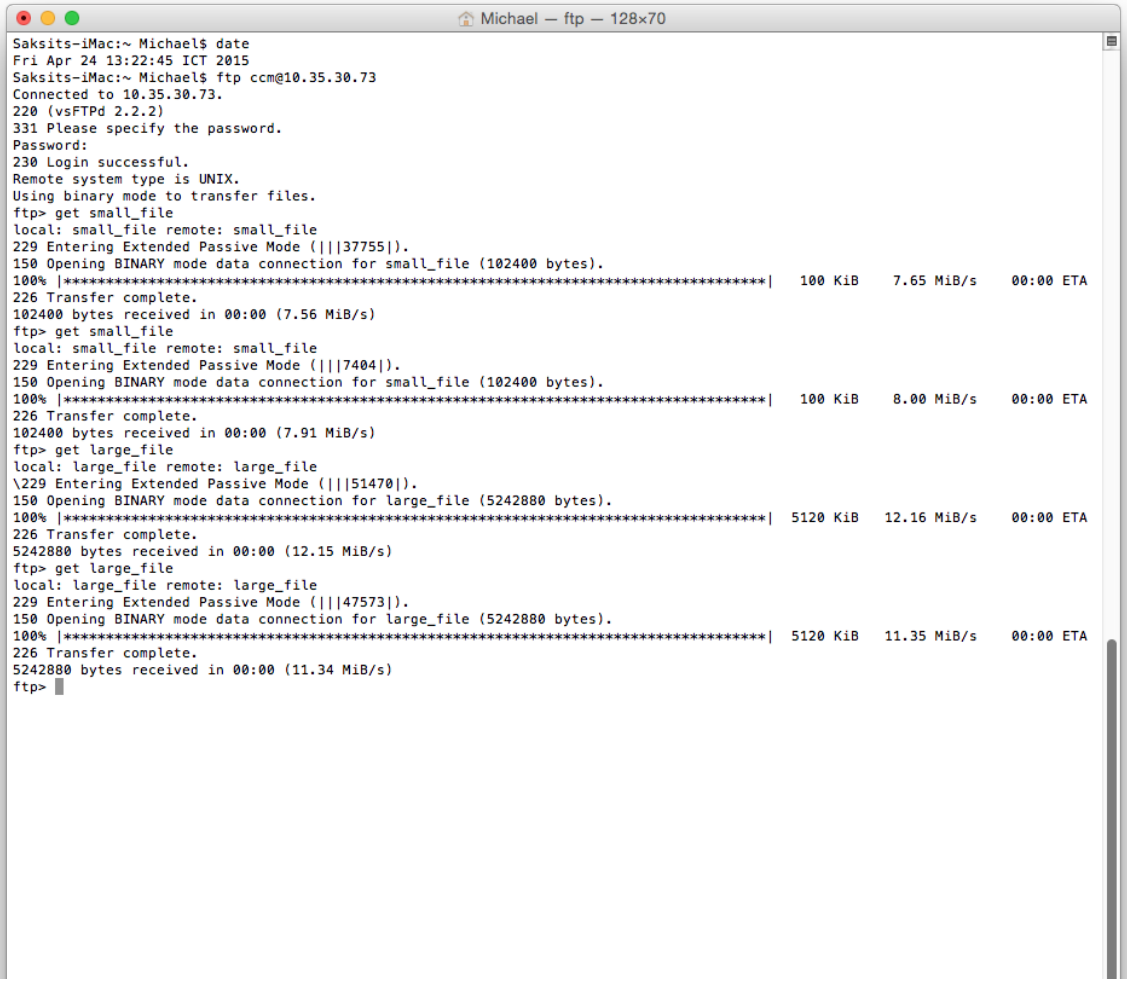

Michael — ftp — 128×70
Saksits-iMac:~ Michael$ date
Thu Apr 23 17:59:39 ICT 2015
Saksits-iMac:~ Michael$ ftp ccm@10.35.30.73
Connected to 10.35.30.73.
220 (vsFTPd 2.2.2)
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> get small_file
local: small_file remote: small_file
229 Entering Extended Passive Mode (|||59445|).
150 Opening BINARY mode data connection for small_file (102400 bytes).
100% |*****| 100 KiB 7.24 MiB/s 00:00 ETA
226 Transfer complete.
102400 bytes received in 00:00 (7.17 MiB/s)
ftp> get small_file
local: small_file remote: small_file
229 Entering Extended Passive Mode (|||25364|).
150 Opening BINARY mode data connection for small_file (102400 bytes).
100% |*****| 100 KiB 7.49 MiB/s 00:00 ETA
226 Transfer complete.
102400 bytes received in 00:00 (7.40 MiB/s)
ftp> get large_file
local: large_file remote: large_file
229 Entering Extended Passive Mode (|||65213|).
150 Opening BINARY mode data connection for large_file (5242880 bytes).
100% |*****| 5120 KiB 11.45 MiB/s 00:00 ETA
226 Transfer complete.
5242880 bytes received in 00:00 (11.38 MiB/s)
ftp> get large_file
local: large_file remote: large_file
229 Entering Extended Passive Mode (|||32006|).
150 Opening BINARY mode data connection for large_file (5242880 bytes).
100% |*****| 5120 KiB 11.60 MiB/s 00:00 ETA
226 Transfer complete.
5242880 bytes received in 00:00 (11.58 MiB/s)
ftp>

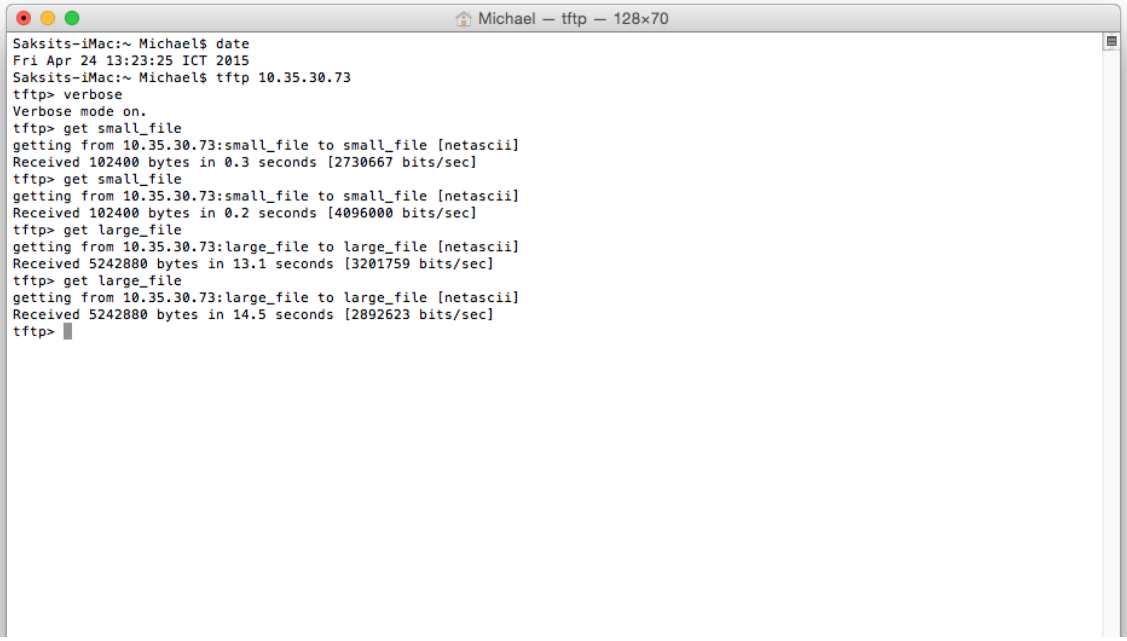

Michael — ftp — 128×70
Saksits-iMac:~ Michael$ ftp ccm@10.35.30.73
Connected to 10.35.30.73.
220 (vsFTPd 2.2.2)
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> get small_file
local: small_file remote: small_file
229 Entering Extended Passive Mode (|||22070|).
150 Opening BINARY mode data connection for small_file (102400 bytes).
100% |*****| 100 KiB 1.86 MiB/s 00:00 ETA
226 Transfer complete.
102400 bytes received in 00:00 (1.86 MiB/s)
ftp> get small_file
local: small_file remote: small_file
229 Entering Extended Passive Mode (|||58746|).
150 Opening BINARY mode data connection for small_file (102400 bytes).
100% |*****| 100 KiB 2.39 MiB/s 00:00 ETA
226 Transfer complete.
102400 bytes received in 00:00 (2.38 MiB/s)
ftp> get large_file
local: large_file remote: large_file
229 Entering Extended Passive Mode (|||9213|).
150 Opening BINARY mode data connection for large_file (5242880 bytes).
100% |*****| 5120 KiB 2.80 MiB/s 00:00 ETA
226 Transfer complete.
5242880 bytes received in 00:01 (2.80 MiB/s)
ftp> get large_file
local: large_file remote: large_file
229 Entering Extended Passive Mode (|||39684|).
150 Opening BINARY mode data connection for large_file (5242880 bytes).
100% |*****| 5120 KiB 2.22 MiB/s 00:00 ETA
226 Transfer complete.
5242880 bytes received in 00:02 (2.22 MiB/s)
ftp>

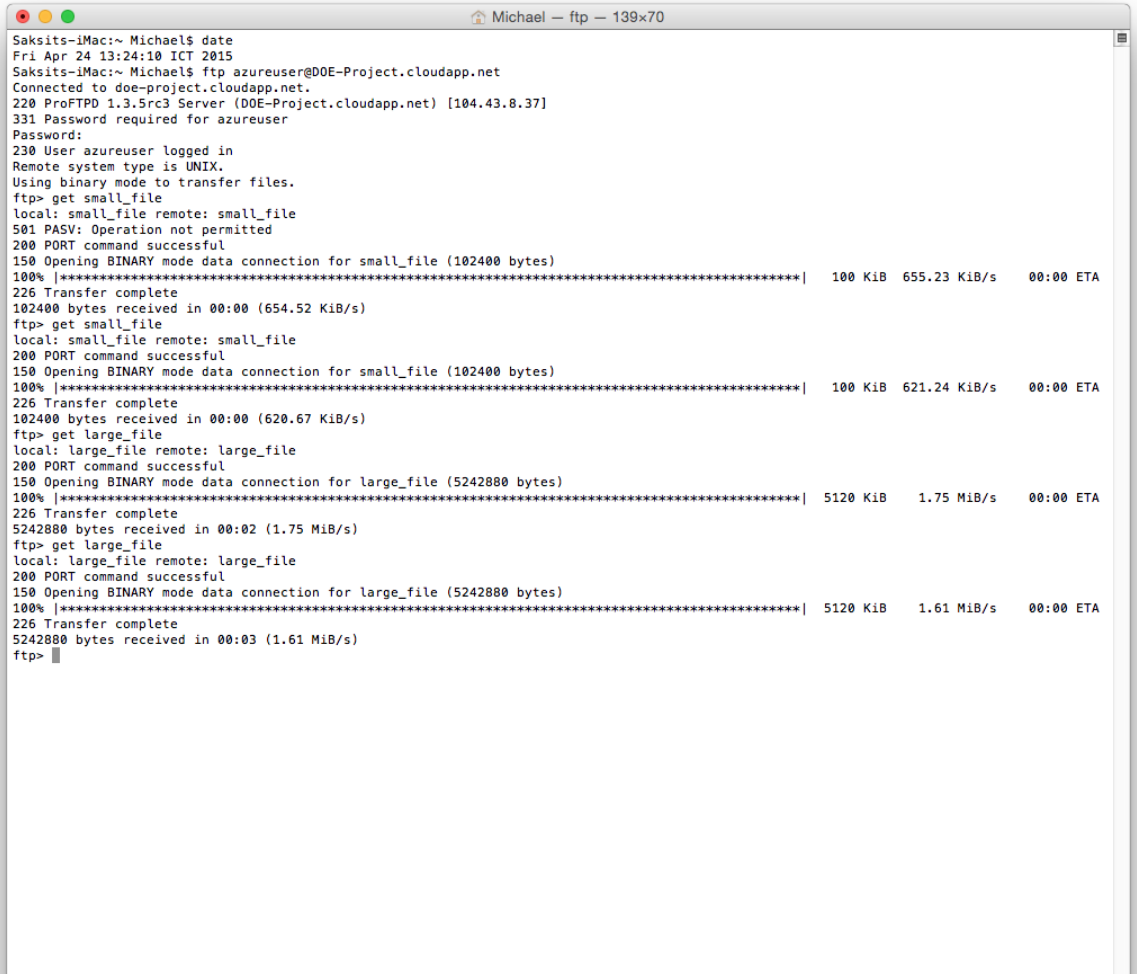

Michael — tftp — 128×70
Saksits-iMac:~ Michael$ tftp 10.35.30.73
tftp> verbose
Verbose mode on.
tftp> get small_file
getting from 10.35.30.73:small_file to small_file [netascii]
Received 102400 bytes in 0.6 seconds [1365333 bits/sec]
tftp> get small_file
getting from 10.35.30.73:small_file to small_file [netascii]
Received 102400 bytes in 0.5 seconds [1638400 bits/sec]
tftp> get large_file
getting from 10.35.30.73:large_file to large_file [netascii]
Received 5242880 bytes in 32.4 seconds [1294538 bits/sec]
tftp> get large_file
getting from 10.35.30.73:large_file to large_file [netascii]
Received 5242880 bytes in 28.9 seconds [1451316 bits/sec]
tftp>

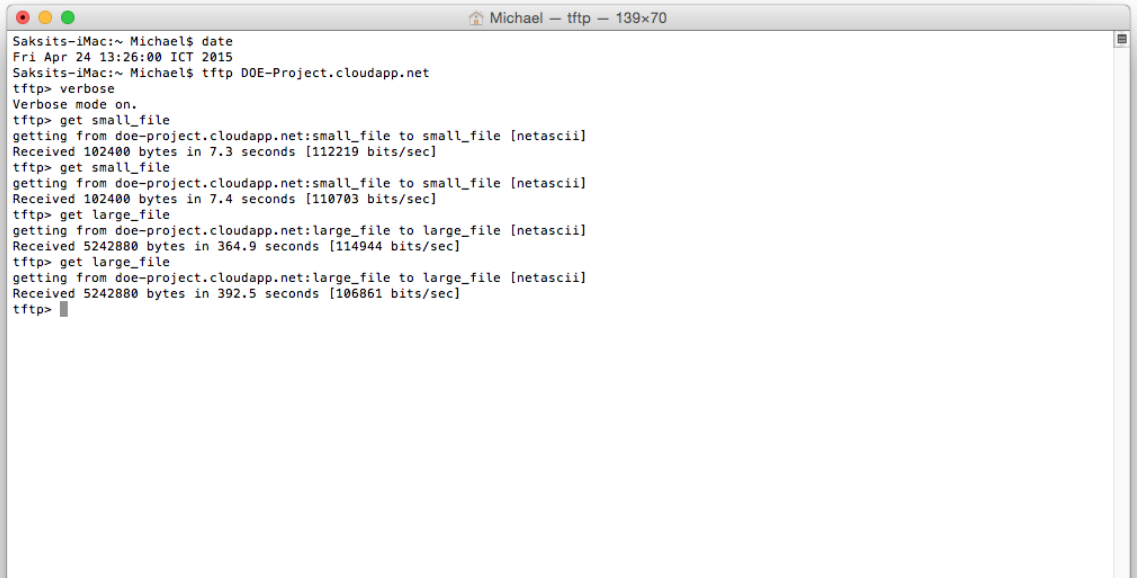

Michael — ftp — 139×70
Saksits-iMac:~ Michael$ ftp azureuser@DOE-Project.cloudapp.net
Connected to doe-project.cloudapp.net.
220 ProFTPD 1.3.5rc3 Server (DOE-Project.cloudapp.net) [104.43.8.37]
331 Password required for azureuser
Password:
230 User azureuser logged in
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> get small_file
local: small_file remote: small_file
200 PORT command successful
150 Opening BINARY mode data connection for small_file (102400 bytes)
100% |**********| 100 KiB 469.86 KiB/s 00:00 ETA
226 Transfer complete
102400 bytes received in 00:00 (469.53 KiB/s)
ftp> get small_file
local: small_file remote: small_file
200 PORT command successful
150 Opening BINARY mode data connection for small_file (102400 bytes)
100% |**********| 100 KiB 483.16 KiB/s 00:00 ETA
226 Transfer complete
102400 bytes received in 00:00 (482.77 KiB/s)
ftp> get large_file
local: large_file remote: large_file
200 PORT command successful
150 Opening BINARY mode data connection for large_file (5242880 bytes)
100% |**********| 5120 KiB 1.20 MiB/s 00:00 ETA
226 Transfer complete
5242880 bytes received in 00:04 (1.20 MiB/s)
ftp> get large_file
local: large_file remote: large_file
200 PORT command successful
150 Opening BINARY mode data connection for large_file (5242880 bytes)
100% |**********| 5120 KiB 1.12 MiB/s 00:00 ETA
226 Transfer complete
5242880 bytes received in 00:04 (1.12 MiB/s)
ftp>

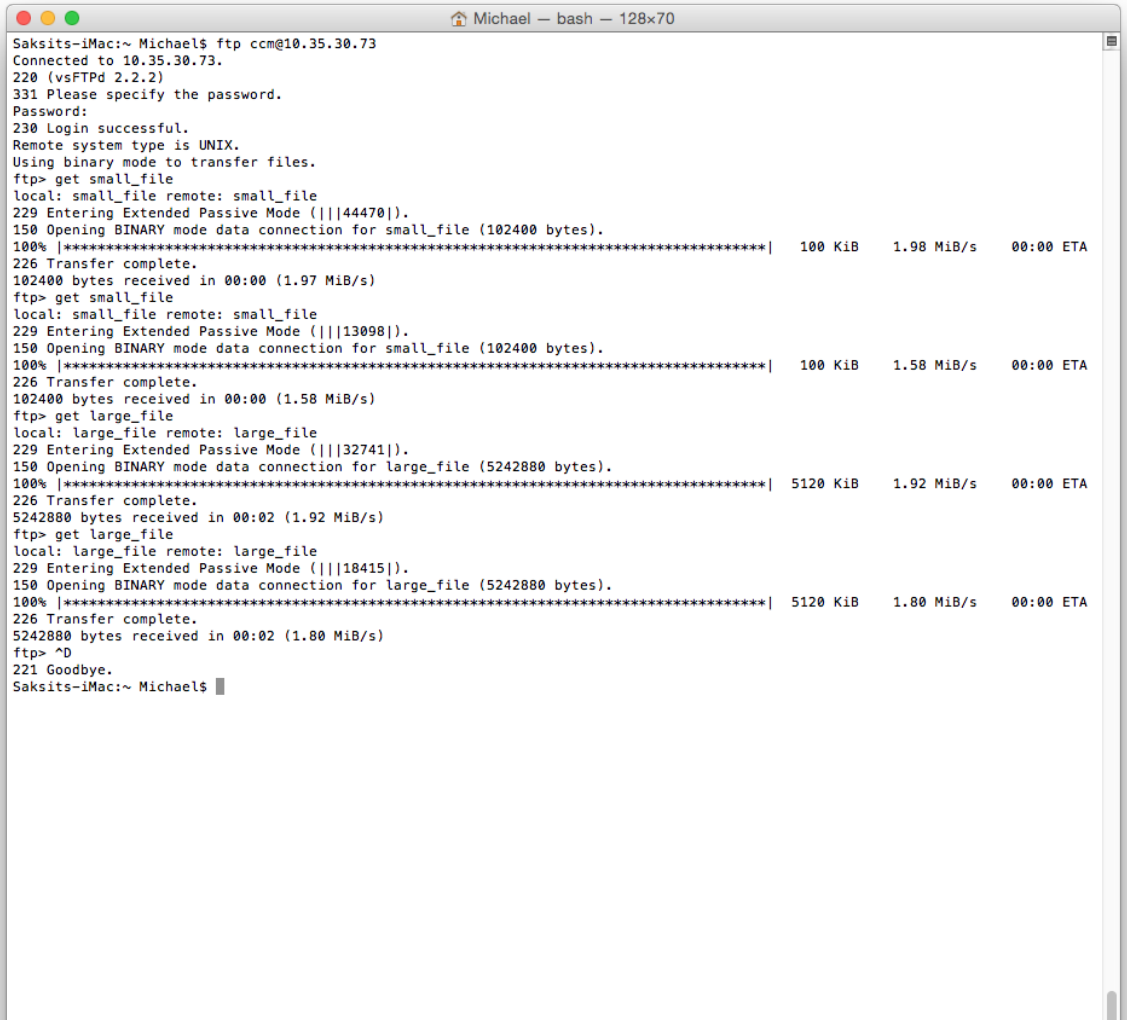

Michael — tftp — 139×70
Saksits-iMac:~ Michael$ tftp DOE-Project.cloudapp.net
tftp> verbose
Verbose mode on.
tftp> get small_file
getting from doe-project.cloudapp.net:small_file to small_file [netascii]
Received 102400 bytes in 7.1 seconds [115380 bits/sec]
tftp> get small_file
getting from doe-project.cloudapp.net:small_file to small_file [netascii]
Received 102400 bytes in 7.3 seconds [112219 bits/sec]
tftp> get large_file
getting from doe-project.cloudapp.net:large_file to large_file [netascii]
Received 5242880 bytes in 377.3 seconds [111166 bits/sec]
tftp> get large_file
getting from doe-project.cloudapp.net:large_file to large_file [netascii]
Received 5242880 bytes in 502.4 seconds [83485 bits/sec]
tftp>

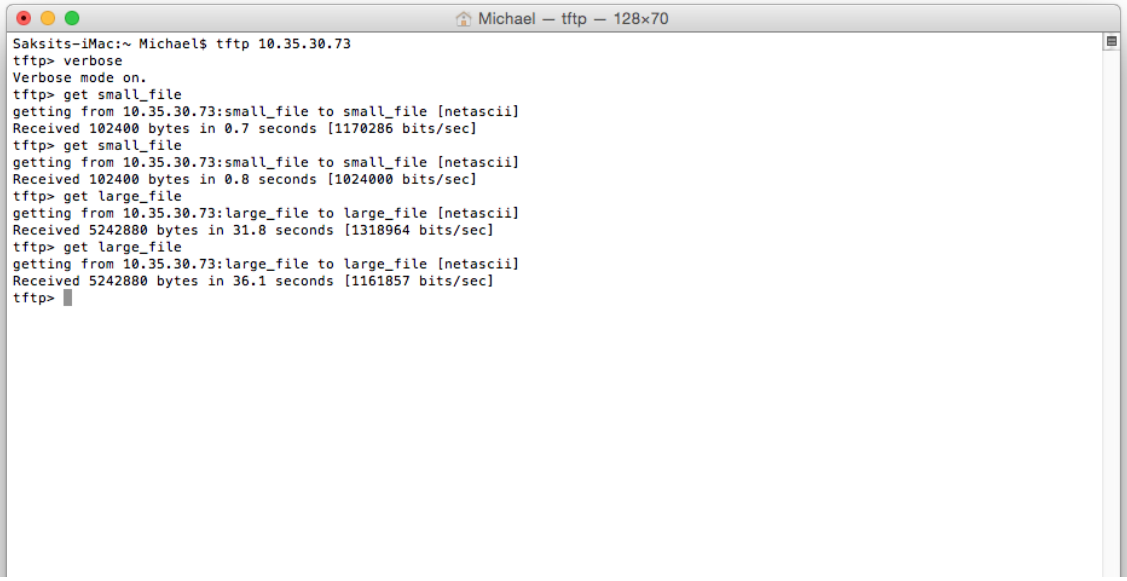

Michael — ftp — 128×70
Saksits-iMac:~ Michael$ date
Fri Apr 24 13:22:45 ICT 2015
Saksits-iMac:~ Michael$ ftp ccm@10.35.30.73
Connected to 10.35.30.73.
220 (vsFTPd 2.2.2)
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> get small_file
local: small_file remote: small_file
229 Entering Extended Passive Mode (|||37755|).
150 Opening BINARY mode data connection for small_file (102400 bytes).
100% |*****| 100 KiB 7.65 MiB/s 00:00 ETA
226 Transfer complete.
102400 bytes received in 00:00 (7.56 MiB/s)
ftp> get small_file
local: small_file remote: small_file
229 Entering Extended Passive Mode (|||7404|).
150 Opening BINARY mode data connection for small_file (102400 bytes).
100% |*****| 100 KiB 8.00 MiB/s 00:00 ETA
226 Transfer complete.
102400 bytes received in 00:00 (7.91 MiB/s)
ftp> get large_file
local: large_file remote: large_file
\229 Entering Extended Passive Mode (|||51470|).
150 Opening BINARY mode data connection for large_file (5242880 bytes).
100% |*****| 5120 KiB 12.16 MiB/s 00:00 ETA
226 Transfer complete.
5242880 bytes received in 00:00 (12.15 MiB/s)
ftp> get large_file
local: large_file remote: large_file
229 Entering Extended Passive Mode (|||47573|).
150 Opening BINARY mode data connection for large_file (5242880 bytes).
100% |*****| 5120 KiB 11.35 MiB/s 00:00 ETA
226 Transfer complete.
5242880 bytes received in 00:00 (11.34 MiB/s)
ftp>

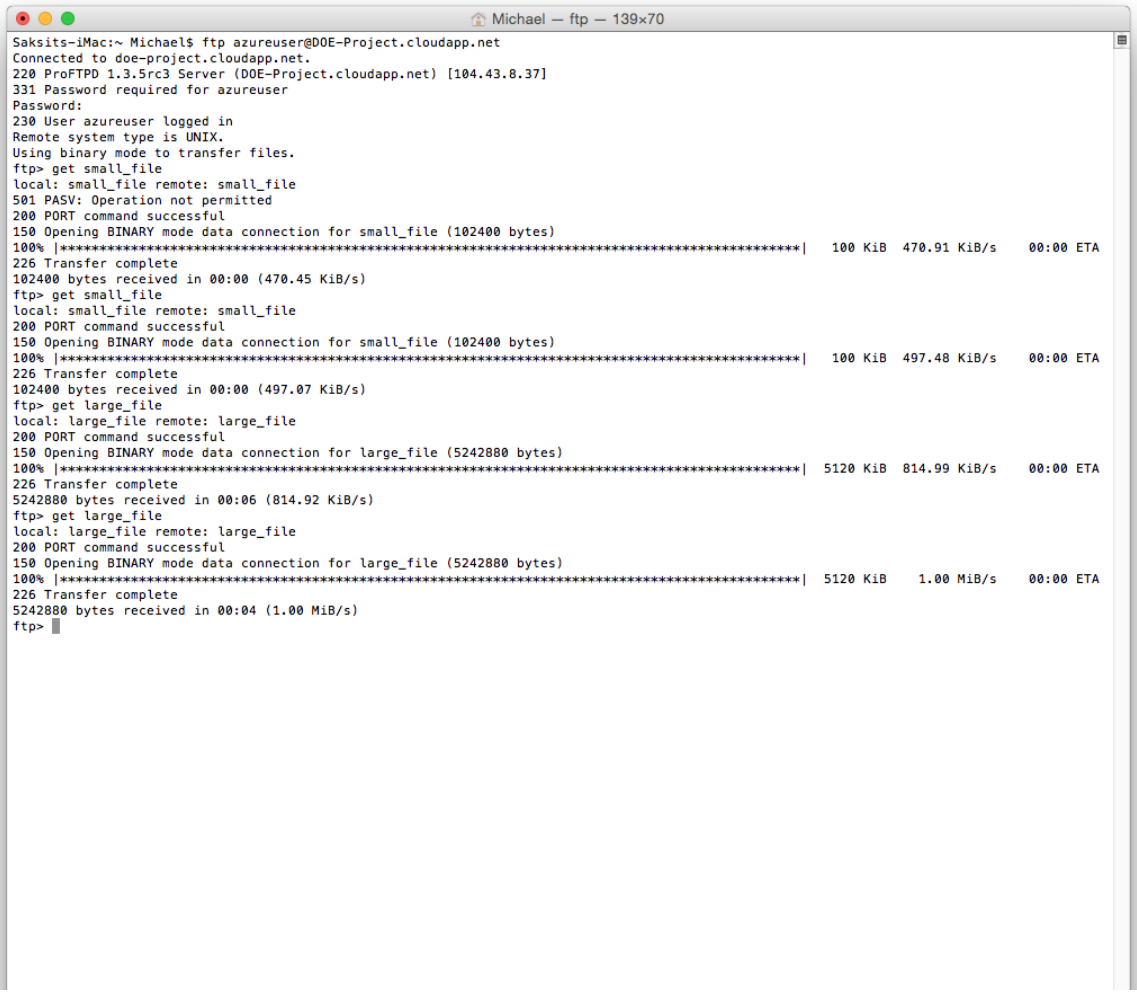

Michael — tftp — 128×70
Saksits-iMac:~ Michael$ date
Fri Apr 24 13:23:25 ICT 2015
Saksits-iMac:~ Michael$ tftp 10.35.30.73
tftp> verbose
Verbose mode on.
tftp> get small_file
getting from 10.35.30.73:small_file to small_file [netascii]
Received 102400 bytes in 0.3 seconds [2730667 bits/sec]
tftp> get small_file
getting from 10.35.30.73:small_file to small_file [netascii]
Received 102400 bytes in 0.2 seconds [4096000 bits/sec]
tftp> get large_file
getting from 10.35.30.73:large_file to large_file [netascii]
Received 5242880 bytes in 13.1 seconds [3201759 bits/sec]
tftp> get large_file
getting from 10.35.30.73:large_file to large_file [netascii]
Received 5242880 bytes in 14.5 seconds [2892623 bits/sec]
tftp>

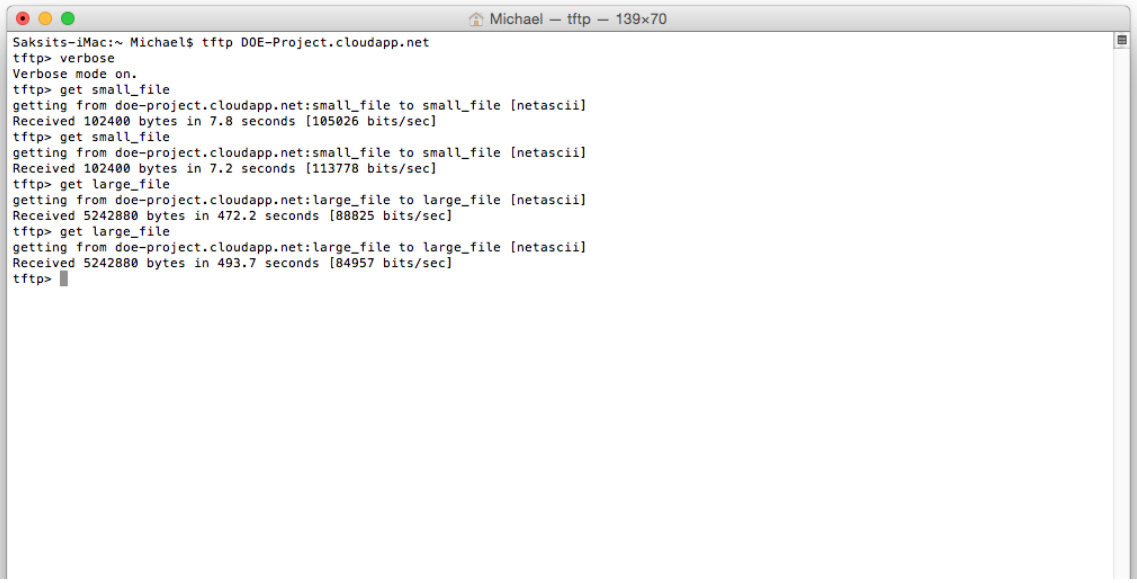

Michael — ftp — 139×70
Saksits-iMac:~ Michael$ date
Fri Apr 24 13:24:10 ICT 2015
Saksits-iMac:~ Michael$ ftp azureuser@DOE-Project.cloudapp.net
Connected to doe-project.cloudapp.net.
220 ProFTPD 1.3.5rc3 Server (DOE-Project.cloudapp.net) [104.43.8.37]
331 Password required for azureuser
Password:
230 User azureuser logged in
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> get small_file
local: small_file remote: small_file
501 PASV: Operation not permitted
200 PORT command successful
150 Opening BINARY mode data connection for small_file (102400 bytes)
100% |**********| 100 KiB 655.23 KiB/s 00:00 ETA
226 Transfer complete
102400 bytes received in 00:00 (654.52 KiB/s)
ftp> get small_file
local: small_file remote: small_file
200 PORT command successful
150 Opening BINARY mode data connection for small_file (102400 bytes)
100% |**********| 100 KiB 621.24 KiB/s 00:00 ETA
226 Transfer complete
102400 bytes received in 00:00 (620.67 KiB/s)
ftp> get large_file
local: large_file remote: large_file
200 PORT command successful
150 Opening BINARY mode data connection for large_file (5242880 bytes)
100% |**********| 5120 KiB 1.75 MiB/s 00:00 ETA
226 Transfer complete
5242880 bytes received in 00:02 (1.75 MiB/s)
ftp> get large_file
local: large_file remote: large_file
200 PORT command successful
150 Opening BINARY mode data connection for large_file (5242880 bytes)
100% |**********| 5120 KiB 1.61 MiB/s 00:00 ETA
226 Transfer complete
5242880 bytes received in 00:03 (1.61 MiB/s)
ftp>


Michael — tftp — 139×70
Saksits-iMac:~ Michael$ date
Fri Apr 24 13:26:00 ICT 2015
Saksits-iMac:~ Michael$ tftp DOE-Project.cloudapp.net
tftp> verbose
Verbose mode on.
tftp> get small_file
getting from doe-project.cloudapp.net:small_file to small_file [netascii]
Received 102400 bytes in 7.3 seconds [112219 bits/sec]
tftp> get small_file
getting from doe-project.cloudapp.net:small_file to small_file [netascii]
Received 102400 bytes in 7.4 seconds [110703 bits/sec]
tftp> get large_file
getting from doe-project.cloudapp.net:large_file to large_file [netascii]
Received 5242880 bytes in 364.9 seconds [114944 bits/sec]
tftp> get large_file
getting from doe-project.cloudapp.net:large_file to large_file [netascii]
Received 5242880 bytes in 392.5 seconds [106861 bits/sec]
tftp>


Michael — bash — 128×70
Saksits-iMac:~ Michael$ ftp ccm@10.35.30.73
Connected to 10.35.30.73.
220 (vsFTPd 2.2.2)
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> get small_file
local: small_file remote: small_file
229 Entering Extended Passive Mode (|||44470|).
150 Opening BINARY mode data connection for small_file (102400 bytes).
100% |*****| 100 KiB 1.98 MiB/s 00:00 ETA
226 Transfer complete.
102400 bytes received in 00:00 (1.97 MiB/s)
ftp> get small_file
local: small_file remote: small_file
229 Entering Extended Passive Mode (|||13098|).
150 Opening BINARY mode data connection for small_file (102400 bytes).
100% |*****| 100 KiB 1.58 MiB/s 00:00 ETA
226 Transfer complete.
102400 bytes received in 00:00 (1.58 MiB/s)
ftp> get large_file
local: large_file remote: large_file
229 Entering Extended Passive Mode (|||32741|).
150 Opening BINARY mode data connection for large_file (5242880 bytes).
100% |*****| 5120 KiB 1.92 MiB/s 00:00 ETA
226 Transfer complete.
5242880 bytes received in 00:02 (1.92 MiB/s)
ftp> get large_file
local: large_file remote: large_file
229 Entering Extended Passive Mode (|||18415|).
150 Opening BINARY mode data connection for large_file (5242880 bytes).
100% |*****| 5120 KiB 1.80 MiB/s 00:00 ETA
226 Transfer complete.
5242880 bytes received in 00:02 (1.80 MiB/s)
ftp> ^D
221 Goodbye.
Saksits-iMac:~ Michael$


Michael — tftp — 128×70
Saksits-iMac:~ Michael$ tftp 10.35.30.73
tftp> verbose
Verbose mode on.
tftp> get small_file
getting from 10.35.30.73:small_file to small_file [netascii]
Received 102400 bytes in 0.7 seconds [1170286 bits/sec]
tftp> get small_file
getting from 10.35.30.73:small_file to small_file [netascii]
Received 102400 bytes in 0.8 seconds [1024000 bits/sec]
tftp> get large_file
getting from 10.35.30.73:large_file to large_file [netascii]
Received 5242880 bytes in 31.8 seconds [1318964 bits/sec]
tftp> get large_file
getting from 10.35.30.73:large_file to large_file [netascii]
Received 5242880 bytes in 36.1 seconds [1161857 bits/sec]
tftp>


Michael — ftp — 139×70
Saksits-iMac:~ Michael$ ftp azureuser@DOE-Project.cloudapp.net
Connected to doe-project.cloudapp.net.
220 ProFTPD 1.3.5rc3 Server (DOE-Project.cloudapp.net) [104.43.8.37]
331 Password required for azureuser
Password:
230 User azureuser logged in
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> get small_file
local: small_file remote: small_file
501 PASV: Operation not permitted
200 PORT command successful
150 Opening BINARY mode data connection for small_file (102400 bytes)
100% |**********| 100 KiB 470.91 KiB/s 00:00 ETA
226 Transfer complete
102400 bytes received in 00:00 (470.45 KiB/s)
ftp> get small_file
local: small_file remote: small_file
200 PORT command successful
150 Opening BINARY mode data connection for small_file (102400 bytes)
100% |**********| 100 KiB 497.48 KiB/s 00:00 ETA
226 Transfer complete
102400 bytes received in 00:00 (497.07 KiB/s)
ftp> get large_file
local: large_file remote: large_file
200 PORT command successful
150 Opening BINARY mode data connection for large_file (5242880 bytes)
100% |**********| 5120 KiB 814.99 KiB/s 00:00 ETA
226 Transfer complete
5242880 bytes received in 00:06 (814.92 KiB/s)
ftp> get large_file
local: large_file remote: large_file
200 PORT command successful
150 Opening BINARY mode data connection for large_file (5242880 bytes)
100% |**********| 5120 KiB 1.00 MiB/s 00:00 ETA
226 Transfer complete
5242880 bytes received in 00:04 (1.00 MiB/s)
ftp>


Michael — tftp — 139×70
Saksits-iMac:~ Michael$ tftp DOE-Project.cloudapp.net
tftp> verbose
Verbose mode on.
tftp> get small_file
getting from doe-project.cloudapp.net:small_file to small_file [netascii]
Received 102400 bytes in 7.8 seconds [105026 bits/sec]
tftp> get small_file
getting from doe-project.cloudapp.net:small_file to small_file [netascii]
Received 102400 bytes in 7.2 seconds [113778 bits/sec]
tftp> get large_file
getting from doe-project.cloudapp.net:large_file to large_file [netascii]
Received 5242880 bytes in 472.2 seconds [88825 bits/sec]
tftp> get large_file
getting from doe-project.cloudapp.net:large_file to large_file [netascii]
Received 5242880 bytes in 493.7 seconds [84957 bits/sec]
tftp>

References

[1] Textbook: Montgomery, D. C. (2012). Design and Analysis of Experiments (8th Edition). Wiley.

[2] File Transfer Protocol. Retrieved from http://en.wikipedia.org/wiki/File_Transfer_Protocol

[3] How To Set Up vsftpd on Ubuntu 12.04. Retrieved from https://www.digitalocean.com/community/tutorials/how-to-set-up-vsftpd-on-ubuntu-12-04

[4] Trivial File Transfer Protocol. Retrieved from http://en.wikipedia.org/wiki/Trivial_File_Transfer_Protocol

[5] Configuring a TFTP Server for Linux. Retrieved from https://linuxlink.timesys.com/docs/linux_tftp

[6] Understanding file sizes from http://www.gn.apc.org/support/understanding-file-sizes