COL334 Assignment 3

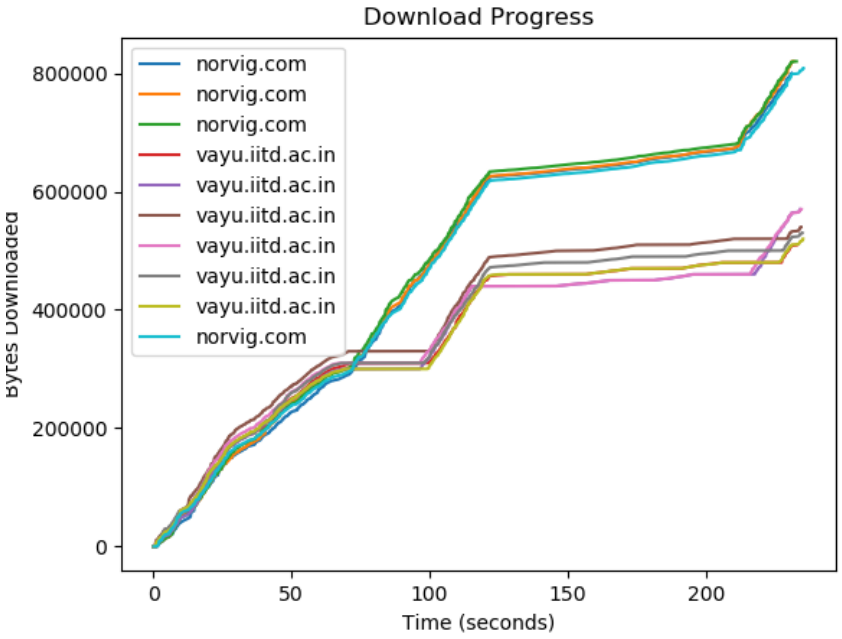
Dipen Kumar (2018CS50098)

November 25, 2020

1. Does your download time keep decreasing with more and more parallel TCP connections? Try to explain your finding.

It start decreasing and with increase in parallel TCP connections since we are introducing parallelism. It saturates once we reach our bottleneck speed. It also depends on the number of core in the computer. Higher the core more is the benefit from the parallel running threads. Hence it saturates at some value depending on the our internet speed and cores available in our computer. Download time then starts increasing because after optimal value it no longer benefit us from parallelism rather more TCP connections means more hand-shaking with increases the download time. Also frequently we may observe broken pipe exception because of longer waiting time of threads because all cores were already occupied and hence server had closed the connection which require to create TCP connection again and increases the download time. Optimal number of TCP connection (threads) to open was 6-10 depending on the present state of computer cores (different programs running at the moment) and internet speed.

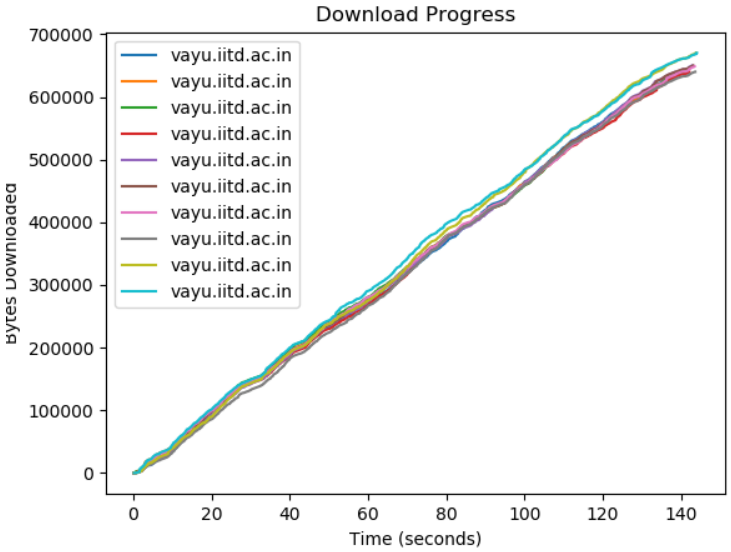
2. You can also log the download progress for each connection, and draw a graph that shows on the y-axis the number of bytes downloaded, and on the x-axis shows the time. You can check if some connections are faster than others, or some connections get stalled for some reason.



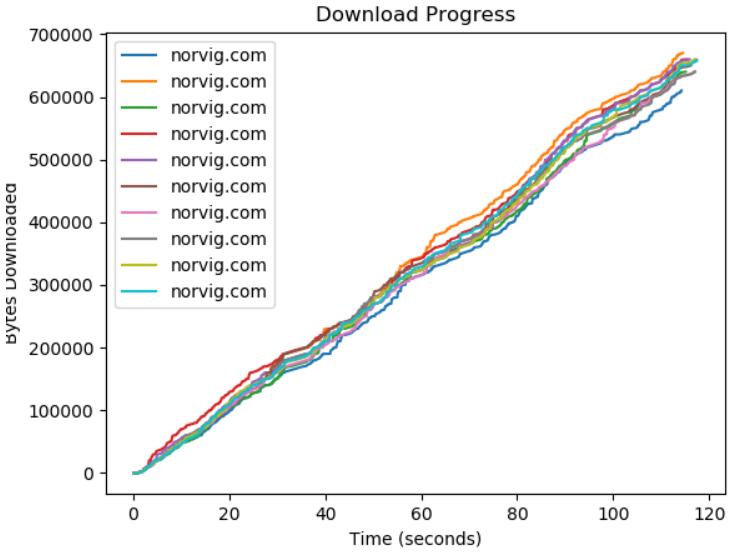
No, all the the TCP connection for same host(server) are same. Although it is sometimes possible depending on the system state (cores) that some thread may get stalled and may fall behind other threads but that is a stochastic process. Two different TCP connections differ in speeds when they are for two different host. For same host all are same with equally likely outcomes.

3. Does your download time reduce further if you download from different sources in parallel?

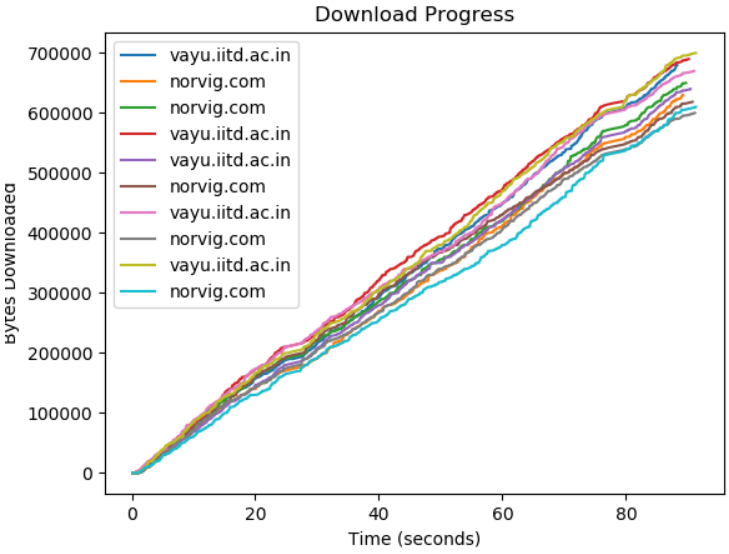
Download Time using 10 parallel TCP connections from vayu.iitd.ac.in was 146.4455 seconds.



Download Time using 10 parallel TCP connections from norvig.com was 119.7124 seconds.



Download Time using 5 parallel TCP connections each of vayu.iitd.ac.in and norvig.com was 93.8306 seconds.



Hence, Yes download time reduces further if we download from different sources in parallel.

4. What does this tell you about where bottlenecks lie?

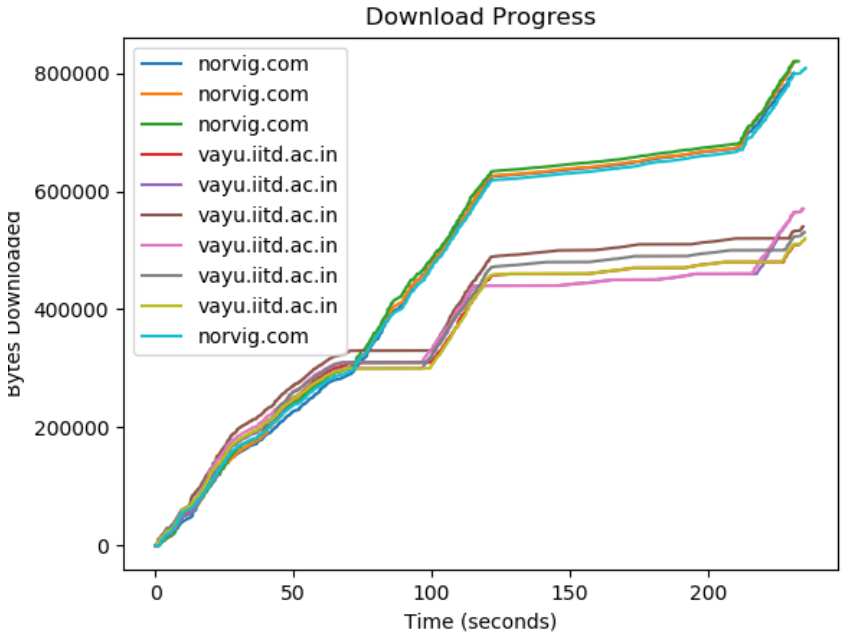
This says bottlenecks lie on server sides because same size file when allowed to download from two sources resulted in lesser time. That is download speed was increased i.e. we were getting data at effectively higher upload rate by combined server. Load on each server was reduced by another providing some data to client. upload rate combined (effective) $>$ upload rate (norvig) $>$ upload rate (vayu)

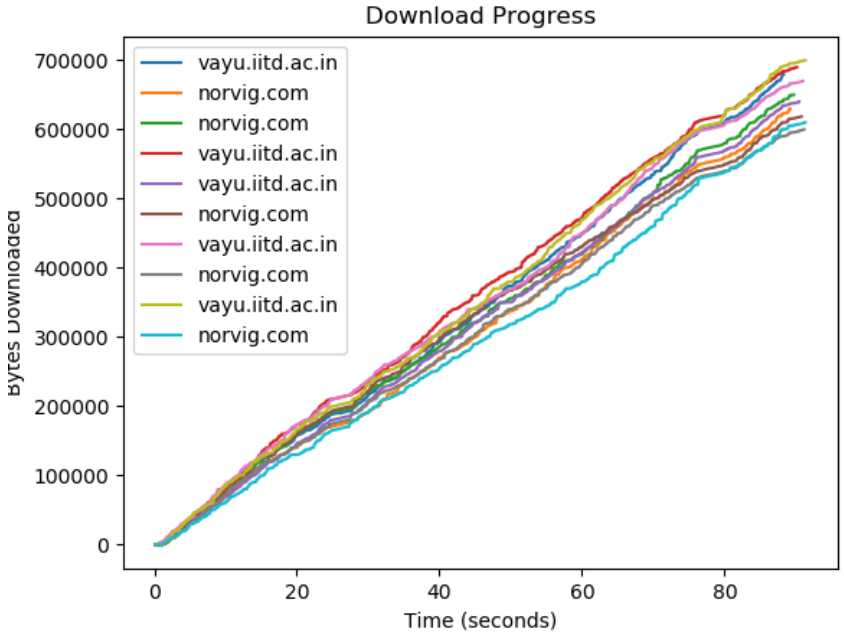
5. Is one server faster than the other?

Yes, norvig.com is faster than vayu.iitd.ac.in. It is clear from above data and graph discussed in question 3.

6. Is your program able to use this to download more from the faster server?

Yes, because whenever a thread completes a chunk download, it queries a synchronized object to pick up a new chunk. Chunks will thus get allocated to threads in a dynamic manner. The thread downloading with higher speed downloads more data. It is clear from below two graphs-


Download Progress
norvig.com
norvig.com
norvig.com
vayu.iitd.ac.in
vayu.iitd.ac.in
vayu.iitd.ac.in
vayu.iitd.ac.in
vayu.iitd.ac.in
vayu.iitd.ac.in
norvig.com
Bytes Downloaded
800000
600000
400000
200000
0
0
50
100
150
200
Time (seconds)


Download Progress
vayu.iitd.ac.in
norvig.com
norvig.com
vayu.iitd.ac.in
vayu.iitd.ac.in
norvig.com
vayu.iitd.ac.in
norvig.com
vayu.iitd.ac.in
norvig.com
Bytes Downloaded
700000
600000
500000
400000
300000
200000
100000
0
0
20
40
60
80
Time (seconds)