Empirical Analysis of File Transfer Costs

- Gary Miller

Several tests were conducted to analyze the costs associated with using client and server sockets to transfer files of different sizes. the files used were pdf's of small (8,495 bytes), medium (98,576 bytes) and large (10,044,218 bytes) sizes. The cost of time associated was either file size was tested both locally and using two remote nodes. The first test that was conducted was finding the mean runtime to transfer the small file locally. After running the test five times, the mean runtime was determined to be .0122 seconds with a standard deviation of .00258 seconds, which is a relatively low cost. Next, the medium file was transfered five time and surprisingly had a mean runtime of just barely larger than the small file, equivalent to .0124 seconds with a standard deviation of .00134 seconds. Given the results from the first two tests, it would be expected that the large file size would take a similar amount of time as well, but this was not the case. the mean runtime for transferring the large file was approximately two and half times as costly with a mean runtime of .0272 seconds with a standard deviation of .00303 seconds.

Next, the same tests were conducted again, only this time the connection was established by using a client and server that were remotely connected over two different nodes. The results from this experiments were very unintuitive, due to the fact that both the small and medium files transferred faster over the remote connection, while the large file became exceptionally more costly. The mean runtimes for the small and medium files were .0102 and .0106 seconds with standard deviations of .00109 and .00114 seconds respectively. It is important to note that the small and medium files still had similar runtimes. The most interesting data point of the experiment was that the mean runtime of transferring the large file over a remote connection was .808 seconds with a standard deviation of .01534 seconds. This is very important to note because the cost of file transfer in this case was 30 times more costly than the remote connection.

The overall conclusion of the experiment was that small data transfers over a remote connection may be more effective over a remote connection, but as the data transfer gets larger it becomes many times more costly to do so. Though, it must be noted that transferring a file of about 10 megabytes took on average less than a second. This means that in terms of relativity the cost of this data transfer is much larger over a remote versus local connection, but in terms of actual costs, the cost is still relatively low.

Total	Execution Time (Seconds)	Standard Deviation	Files Size (bytes)
Small Local	0.0112	0.0025884	8495
Medium Local	0.0124	0.0013416	98576
Large Local	0.0272	0.0030331	10044218
Total	Execution Time (Seconds)	Standard Deviation	Files Size (bytes)
Small Remote	0.0102	0.0010954	8495
Medium Remote	0.0106	0.0011401	98576
Large Remote	0.808	0.0135462	10044218

Figure 1: Figure I