First name: Christian
Last name: Pfligersdorffer
Date: April 18, 2004
Homework number: 1

Problem description: A problem comprises two components, A and B. A can be solved in 1000 seconds on computer C_1 and in 5000 seconds on computer C_2; B requires 4000 and 2000 seconds on C_1 and C_2, respectively. The two computers are connected by a 1000-km optical fiber link that can transfer data at 100 MB/sec with a latency of 10 msec. The two components can execute concurrently but must transfer 10 MB of data 10,000 times during execution. Is it cheapest (in terms of computational resources consumed) to solve the problem on computer C_1, on computer C_2, or on both computers together? Confirm your answer with an adequate analysis.

Problem solution: In an ideal multitasking environment, not counting the costs for process switching, the problem would take 5000 seconds on C_1, which is the sum of the execution times of the two components on this machine. Since C_2 seems to be a bit slower the problem would need at least 7000 seconds to be solved solely there. Now we have to take a look at the concurrent situtation.

Neglecting data transmission we get the concurrent computation time T_{comp} = 2000 seconds if subproblem A is solved on computer C_1 and subproblem B on computer C_2 (which is obviously the more sensible spreading). But the components *have* to exchange data, more precicely there are 10,000 blocks of 10 MB to be transferred between them.

At 100 MB/sec each block takes 100 msec plus latency time to cross the fiber link. Assuming the worst, i.e. the communication cannot be run concurrently (because the components are too heterogenous or the data link is not bi-directional), the communication time sums up to $T_{comm} = 10{,}000 \times (100 + 10) = 1{,}100{,}000$ msec, making 1100 seconds.

Adding the time values for computation and communication tells us the total execution time for the concurrent setup. Using our previous results we get $T_{comp} + T_{comm}$ = 2000 + 1100 = 3100 seconds.

Results: The analysis stated above clearly shows that in terms of time ressources a parallel computation of the problem will be the most efficient. Our calculations estimate the execution time in this case with 3100 seconds (A on C_1, B on C_2) whereas the computation on a single machine would last 5000 (all on C_1) or even 7000 seconds (all on C_2).

Discussion and comments: The estimated execution time for a single machine is probably too low since the costs for process switching are not taken into account.