CSI 702 2006 Midterm Exam - Take-Home

The test is closed book, closed notes, no web access. You are encouraged to type the answers, but hand written notes are acceptable as well. The test should take no more than four hours to complete. Answers should be brief - most require only one paragraph. All work must be turned in by the end of class on May 10. Please use additional paper to answer these questions, and write your name at the top of each page. **Good luck!**

Honor Code Certification
Name :

Time started:

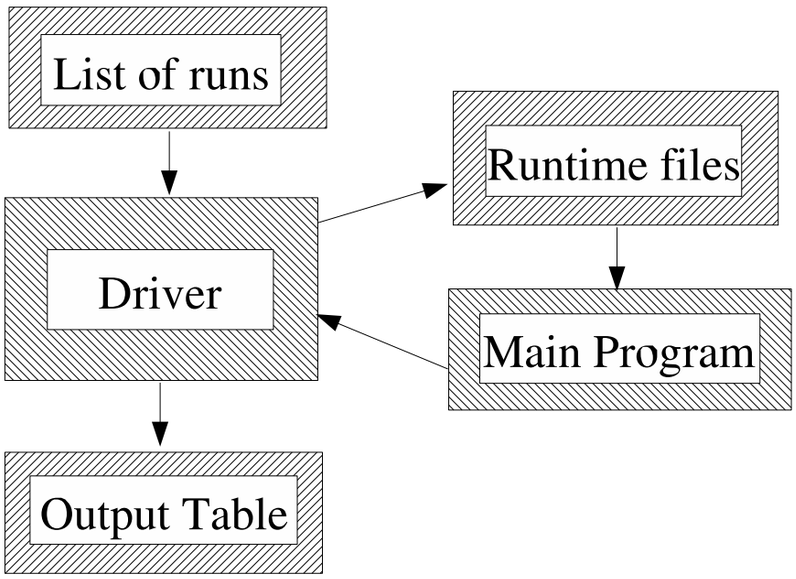
Time completed:

I certify that I have abided by the GMU honor code in taking this examination. The work on this exam is my own. I have received no assistance from other persons in completing this exam. I have not consulted any sources of information other than a calculator.

Signature:

1. OpenMP, MPI, GRIDS – In about a page, compare and contrast OpenMP, MPI, and Grids using examples of problems or programming patterns that would fit better on each of the three environments. Be concise. Use tables, diagrams and/or specific examples.

2. As part of your research at GMU, you have to use a simple code to calculate the goodness of fits of various models (see diagram below). Based on a list of runs, a driver program creates a set of run-time files that are used as input to a main program. The output of this main program is then processed by the driver code and stored in an output table. Each run is independent of every other run, and takes about 1 minute to complete on a typical 3GHz Intel/AMD machine single node machine. Describe the computers and software you would write or use to complete this project for each of the following three phases.



(a) In the class, we have talked about different types of parallel problem taxonomies. In general, how would you characterize the taxonomy need to solve this problem on parallel machines?

(b) For the first round of this project, you need to have 1000 runs. How would you complete these runs using your own computers and GMU resources? DO NOT WRITE CODE- just describe the general approach to this problem.

(c) Assume you have to complete 10,000 runs for the next phase of the project. How would you complete these runs using your own computers and GMU resources? DO NOT WRITE CODE- just describe the general approach to this problem.

(d) Assume you have to complete 100,000 runs for the next phase of the project. How would you complete these runs? DO NOT WRITE CODE- just describe the general approach to this problem.

3. You are given a project that requires you to solve both equations of gravity and hydrodynamics. The gravity uses direct summation of the contributions of each cell (a.k.a. Brute force) requires communications between all the cells in the system. The hydrodynamics code only requires communication between nearest neighbors. Assume you can separate the gravity calculation from the hydrodynamics calculation into two different steps. You can assume we are using a regular cubic three dimensional grid for this problem.

(a) For each grid cell i, assume there are 500 floating point calculations for each neighboring cell around i for the hydrodynamics calculation. Also assume that for every cell i in the grid there are 10 floating point calculations needed for each of the $n-1$ other cells to find the gravitational potential and force. Assuming a 3 dimensional grid and a fast Intel/AMD computer, what is the largest grid size you could reasonable expect to complete in one day?

(b) Assume that for each hydrodynamic calculation, you need to communicate 20 floating point variables across the boundaries in each direction for each cell. For a cubic volume in space that has 1000^3 cells and a computer with eight nodes, describe the best configuration spatial decomposition that minimizes the communication across processor boundaries. Explain how many messages will be sent, what size the messages will be, and how this answer might depend on the latency of the communication.

(c) Assume that for the gravity calculation, you need to communicate 4 floating point numbers between cells in both directions. Describe the communications requirements for the eight CPU and 1000^3 node configuration required above.

(d) Explain what you would do to make this code run faster for large problems.

4. Grids

(a) Diagram and briefly explain the essential elements for a computer that provides resources to a Grid. How does this differ in complexity from the resources already available through the Web processing model? In other words, how are these resources on a computer detected and used by the rest of the Grid?

(b) GMU has been considering developing a Grid for computing instead of buying either a large Beowulf cluster or a shared memory machine. In one page or less, explain why this would or would not be a better strategy for meeting the needs for high end computing on campus. Specifically, what problems could and could not be solved using this approach.

5. Based on your experience in this class, why is parallel programming harder than programming in serial? If this is true, why (and when) is parallel programming worth the effort? What changes in the future might make this easier for scientists?