Department of CSE, The University of Texas at Arlington

CSE5351/4351: Parallel Processing

Spring 2022

Homework Assignment 4

Assigned on March 28, 2022

Q1. (20 points) A scan is a generalization of reduction. Suppose there are n processes, and process $P(i)$ contains a value denoted as $a[i]$. Then after a scan, process $P(i)$ will get the result $a[i] + ... + a[n-1]$.

Show a short single-code program to compute the scan. The scan operation is realized in $\log n$ supersteps.
Show a figure to illustrate how your program computes the scan for $n = 8$. Follow the style in class notes.

Q2. (20 points) Refer to the paper entitled, "Isoefficiency: Measuring the Scalability of Parallel Algorithms and Architectures," by Grama, Gupta, and Kumar, and explain the efficiencies of matrix multiplication algorithms using row-wise partitioning and checkboard partitioning. Also, explain why the checkboard partitioning algorithm is more scalable on a hypercube parallel computer. A hypercube parallel computer is a non-shared distributed memory machine in which the interconnection network is a hypercube topology.

Q3. (20 points) The sequential time to multiply two *N by N* matrices is about $T_1 = cN^3$ seconds, where c is a constant. A parallel matrix multiplication scheme has a parallel execution time of $T_n = (cN^3)/n + bN^2/\mathrm{sqrt}(n)$ seconds on an n-node computer, where b is another constant. The first term denotes the computation time, and the second term the communication overhead.

(a) Find the fixed-workload speedup and comment on the results.
(b) Find the fixed-time speedup and comment on the results.

Q4. (20 points) Consider the two matrix multiplication schemes in (slide 40 of Part 3 Lecture). Assuming a fixed efficiency of $E = 0.25$, system B is more scalable than system A by the isoefficiency metric, for all values of $b, c > 0$. Assuming $b = 4c$, we have $W(A) = 1.33^3 n^{1.5}$ and $W(B) = n^{1.5}$.

How much more scalable is B over A under the above conditions?
See the parallel execution times in Table 3.25. Derive an expression for the ratio $T_n(B) / T_n(A)$, when .
Can system B be slower than system A? Explain why or why not.

Q5. (20 points)

The task graph shown in figure below represents an image processing application. Each bubble represents an inherently sequential task. There are 12 tasks: an input task, 10 computation tasks, and an output task. Each of the 12 tasks can be accomplished in 1 unit of time on one processor. The input task must complete before any computational tasks begin. Likewise, all 10 computational tasks must complete before the output task begins. The input task consumes the entire bandwidth of the input device. The output task consumes the entire bandwidth of the output device.

(a) What is the maximum speedup that can be achieved if this problem is solved on two processors? (Hint: Processors do not have to receive the message elements in order.)

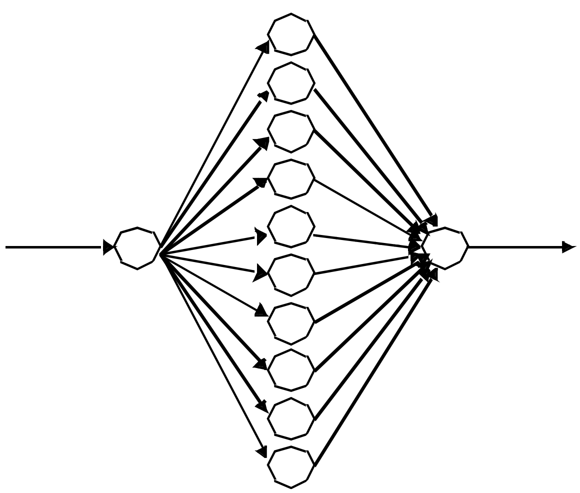
(b) What is an upper bound on the speedup that can be achieved if this problem is solved with parallel processors?

(c) What is the smallest number of processors sufficient to achieve the speedup given in part (b)?

(d) What is the maximum speedup that can be achieved solving five instances of this problem on two processors? Continue to assume that there is one input device and one output device.

(e) What is an upper bound on the speedup that can be achieved solving 100 instances of this problem with parallel processors? Continue to assume that there is one input device and one output device.

(f) What is the smallest number of processors sufficient to achieve the speedup given in part (e).



SUBMISSION: WHAT, WHEN & HOW

(1) Provide your answers on <u>in a Word file and email to the TA addison.clark@mavs.uta.edu.</u>

(2) This assignment is due on or before April 10, 2022