Uva Wellassa University, Sri Lanka

Faculty of Science and Technology

Department of Computer Science & Technology

End Semester Examination –September/October 2013

CST 326-2 Distributed and Parallel Computing

Instructions for candidates:

Answer all five questions.

Time Allowed: Two (02) Hours.

1.

a. Write a small description about **four decades of computing.** (5 mark)

b. What is the **FLYNN's taxonomy of computer architecture** and explain each category by using simple diagrams. (5 mark)

c. **Shared memory machines** can be divided into two groups based upon **memory access time**. What are these two groups? Draw schematic diagram to explain these two. (5 mark)

d. **"Hybrid Distributed-Shared Memory** is used in largest and fastest computers in the world today". Explain this statement. (5 mark)

2.

a. What is **memory interleaving**? (3 mark)

b. Categorize a **topology-based taxonomy** for an interconnection network. (4 mark)

c. A given task can be divided into **n** number of equal subtasks that can be executed by one processor.

t_S - Execution time of the whole task using a single processor

t_m - Time taken by each processor

S(n) - Speedup factor

17

i. Write the expression for **time taken by each processor (t_m)** and **speedup factor (S(n))**. (2 mark)

ii. If **communication overhead is t_c**, then write the expression for **speedup factor (S(n)).** (2 mark)

iii. Think that fraction **f** of the given task is **not dividable** into concurrent tasks and **(1-f)** part is assumed to be **dividable** into concurrent tasks. Derive the expression for **time taken by each processor (t_m)** and then derive the expression for **speedup factor (S(n))** for this model. (6 mark)

iv. Derive the expression for **speedup factor (S(n))** again using **communication overhead (t_c)** for part **iii.** (3 mark)

3. **Parallel Random Access Machine (PRAM)** is one of the models that have been used extensively. In this model, communication cost and synchronization overhead are negligible.

a. Draw a simple schematic diagram for **PRAM model for parallel computations** and write the **task of active processors** on this model. (4 mark)

b. There are four different modes for read and write operations in a **PRAM**. List these four modes and simply describe the operation of each of these modes. (8 mark)

c. Assume that **x** is memory location; **x** is needed by all processors at a given time in a PRAM (Parallel Random Access Machine). Concurrent read by all processors can be performed in the **CREW** and **CRCW** cases in constant time, but in the **EREW**, this is not possible. Therefore, a **broadcasting mechanism** is used by **EREW**. Write down the **algorithm** for this broadcasting mechanism used by **EREW** model. (8 mark)

4. Network used in distributed systems is basically built from **Transmission Media, Hardware Devices, Network Interfaces** and **Software Components.**

   a. Explain **four** requirements for a computer network. (4 mark)

   b. Request-reply protocol is one of the protocols used for communication in a distributed system. This is based on a **trio** of communication primitives. What are these **three communication primitives**? (3 mark)

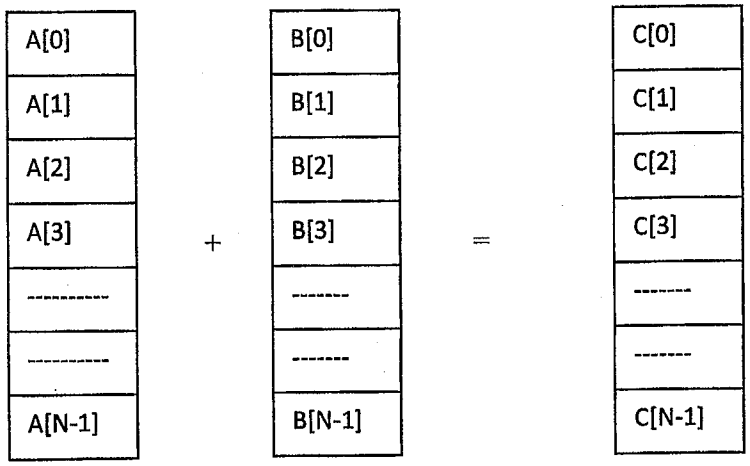
   c. Explain **operation** of each of these primitives as **methods (functions)** with **corresponding input parameters.** (9 mark)

   d. Support of the **operating system** is another considerable factor for **distributed system.** Explain the **two operating system concepts** used during the development of distributed systems. (4 mark)

5. We can use **Compute Unified Device Architecture (CUDA)** to expose GPU (Graphic Processing Unit) computing for general purpose.

   a. What are the main steps you have to follow (simple process flow of GPU computing) if you want to run some parallel code on CUDA enabled GPU? (3 mark)

   b. You have to add two arrays together **(Array A and B)** and get the answers to the new array **(Array C)** as following figure using CUDA enable GPU in parallel.



i. No of elements in each array is **N** and **number of threads per block is 128**. How many **blocks** have to be created in this task? (Assume that both **block dimensions** and **grid dimensions** are **1D**). (2 mark)

ii. Write the kernel function called "**__global__ void add(...)**" that **run on device**. You have to clearly indicate how to calculate each index of array using **blockIdx.x, blockDim.x** and **threadIdx.x** and how to get answers to the array. (8 mark)

iii. Write down **main c function body** of your program with **memory management part** and **launching kernel**. (7 mark)

Hints:
CUDA library functions:

- **cudaMalloc (void **devPtr, size_t size)** - Allocate memory on the device.
- **cudaMemcpy (void *dst, const void *src, size_t count, enum cudaMemcpyKind kind)** - Copies data between host and device.
- **cudaFree (void *devPtr)** - Frees memory on the device.