COMP 25212

Two hours - online

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

System Architecture

Date: Monday 20th May 2019

Time: 14:00 - 16:00

This is an online examination. Please answer all FOUR Questions
Each question is worth 20 marks

This is a CLOSED book examination

The use of electronic calculators is permitted provided they
are not programmable and do not store text

[PTO]

Section A

1. **Caches**

a) Modern compute systems overwhelmingly use caches. Explain why caches are necessary and explain how they work. (3 marks)

b) Cache misses are classified in three categories: compulsory, capacity and conflict misses, based on their respective cause. For each type of cache miss, identify the cache parameters that have an effect on the occurrence of such a cache miss, then explain the effect and detail your reasoning about how increasing or decreasing each cache parameter is likely to impact cache behaviour.
Note: Only consider the effects of changing one parameter at a time, not combined effects. (7 marks)

c) Consider the following implementation of a dot product in a C-like syntax.

```
float X[256], Y[256], xy;

for (int i = 0; i < 256; ++i)
  xy = xy + X[i] * Y[i];
```

We will assume that the two vectors X and Y are stored consecutively in memory (i.e., Y[0] is stored immediately after X[255]) and that all scalar variables ("i" and "xy") are always in registers (i.e., they never occupy the cache). Each "float" value is stored on 32 bits.

Consider a system with a single level of cache split between instructions and data (Harvard architecture). The data cache has a capacity of one kilobyte (1024 bytes) and the cache line size is sixteen (16) bytes. This data cache is direct mapped.

i) What types of locality are present in this code? Explain your reasoning. (2 marks)
ii) How many data cache misses will occur during the execution of this program? Explain and detail your reasoning. (4 marks)
iii) What proportion of these data cache misses are compulsory misses? What proportion are capacity misses? What proportion are conflict misses? Explain your reasoning. (2 marks)
iv) How could the data cache configuration be amended to bring the total cache miss rate to 25%? Explain why. (2 marks)

2. **Virtualization and Storage**

a) What are the four main types of RAID systems commonly used today? Describe their mode of operation and explain what are their strengths and weaknesses. (8 marks)

b) Your company is currently relying on a RAID 5 system of four disks for storing business-critical data. This system is now reaching capacity and it has been proposed to move to a RAID 5 system composed of five disks. We will consider that each disk is identical in all respects, and in both configurations.

i) What percentage increase in effective storage capacity does this new system offer? Detail your reasoning. (3 marks)

ii) What impact does this upgrade have on the mean time to failure of the company's storage system? Assuming that disk failure is a uniformly distributed and independent event for each disk, quantify the difference in mean time to failure between the two systems and explain your reasoning. (3 marks)

iii) Is this choice of storage system appropriate for business-critical data? Explain why and, if not, what would you recommend? (2 marks)

c) Explain what is "live migration" and describe how it can be implemented using System Virtualisation. (4 marks)

Section B

3. **Processor Architecture**

a) What is pipelining in the context of processor design? What benefits does it provide? What issues arise from this architecture? What techniques can be used to Mitigate the impact of these issues? (12 marks)

b) What needs to be changed in an in-order processor to transform it into an out-of-order processor? What are the benefits and issues of performing instruction reordering in hardware? (8 marks)

4. **Multithreading / Multicore**

Consider a classic 5-stage pipelined, 8-core processor with 2-way multithreading and 3-way superscalar pipelines. The processor operates at 2GHz frequency. The memory subsystem features a local cache for each core and a shared main memory. Cache coherency is maintained through a MESI protocol with copy back.

a) What is this processor's peak IPC? (2 marks)

b) How many concurrent HW threads does it support? (2 marks)

c) What is the peak single-thread performance? (2 marks)

d) And the maximum computing throughput of the processor? (2 marks)

e) Describe the messages being sent through the bus, the cache state transactions and the actions in main memory for the following sequence of accesses to a shared variable 'total'. For the sake of simplicity, assume all the cache lines start in the invalid state. (12 marks)

```
core7:  LDR r2, total
core3:  LDR r1, total
core7:  STR r0, total
core5:  LDR r3, total
```