



University
of Glasgow

Tuesday, 2 May 2017
2:00 pm – 4:00 pm
(2 hours)

DEGREES OF MSc, MSci, MEng, BEng, BSc, MA and MA (Social Sciences)

ADVANCED OPERATING SYSTEMS (M)

Answer 3 out of 4 questions

This examination paper is worth a total of 60 marks.

The use of calculators is not permitted in this examination.

INSTRUCTIONS TO INVIGILATORS: Please collect all exam question papers and exam answer scripts and retain for school to collect. Candidates must not remove exam question papers.

1. (a) Explain what is *Dennard scaling*. Outline how Dennard scaling interacts with Moore's law to affect the performance and power consumption of computing systems. Describe the impact the breakdown of Dennard scaling has on system design. [5]
- (b) Many modern multicore processor systems use a non-uniform memory access (NUMA) architecture. Explain in what way the memory is non-uniform, and describe how the non-uniform nature of the memory is visible to software running on such processors. [5]
- (c) Two alternative abstractions for concurrency are transactional memory with automatic roll-back and retry, and communication via message passing, where copies of immutable data are passed between processes that don't share memory. Which of these abstractions do you prefer as a basis for systems programming? Justify your answer, and explain your rationale for making this design choice. [10]
2. (a) One widely used approach to automatic memory management is to add reference counts to objects. The runtime system maintains these counts as the program allocates new objects and manipulates references. The memory allocated to an object is reclaimed when the reference count for that object is decremented to zero. Outline the main advantages and disadvantages of reference counting as a means of automatic memory management. [5]
- (b) The mark-sweep algorithm is one of the simplest garbage collection algorithms. Describe how the mark-sweep algorithm works. Outline three problems that limit the usefulness of this algorithm. [10]
- (c) Garbage collection and virtual memory can interact in ways that cause extremely poor system performance. Explain what is happening when this occurs, and why. [5]
3. (a) Recent years have seen concern that the network protocol stack in typical operating systems is becoming a bottleneck, since the performance of the network is increasing faster than the performance of the CPU. Give two reasons why this is the case. [2]
- (b) We discussed the `netmap` and `StackMap` APIs that provide alternatives to the traditional Berkeley Sockets network API. Outline the features and data structures of the `netmap` API. Explain why these features and data structures allow it to offer better performance than the Berkeley Sockets API. [8]
- (c) We discussed the paper by M. Eriksen on "Your Server as a Function" (Proc. ACM PLOS 2013). This describes the `Finagle` library used by Twitter to build scalable, high-performance, network services. Briefly outline how this library works. State whether you think the approach outlined in this paper is a good solution to the problem of building high-performance servers. Discuss and justify your answer. [10]
4. (a) It is possible to use a maximum utilisation test to determine whether a system of n independent periodic tasks can be scheduled in a pre-emptive manner on a single processor using the rate monotonic algorithm. What is the expression for the maximum utilization in such a system? What are the implications for scheduling if the utilisation of the system is greater than the maximum utilisation? [3]

- (b) How does the maximum utilisation for a system of rate monotonic tasks change as the relative deadlines of those tasks increase to be greater than their periods? [1]
- (c) An alternative to the maximum utilisation test is to perform time demand analysis of the behaviour of a system at its critical instants. Describe when the critical instants of a task occur, outline what time demand analysis is, and discuss how it can be used to determine if a system can be scheduled. [6]
- (d) When sporadic tasks are introduced into a priority-scheduled system of periodic tasks, it becomes necessary to incorporate an acceptance test into that system. Describe the purpose of an acceptance test, and why it is important for error handling. [5]
- (e) Are sporadic tasks incompatible with hard real-time systems? Discuss. [5]