COMP9334: Capacity Planning of Computer Systems and Networks

Course Review

UNSW
THE UNIVERSITY OF NEW SOUTH WALES

A/Prof Chun Tung Chou

CSE, UNSW

System performance is important

- Performance metrics: response time, waiting time, throughput
- Performance is determined by:
  - Workload
  - System parameters
- You can estimate system performance, without building the actual system, by using queueing models

Performance analysis techniques (1)

- Operational analysis (Week 2)
  - Measurements on the systems
  - Operational laws, in particular Little's Law
  - Key concept: Bottleneck
  - Upper bound on system performance

Performance analysis techniques (2)

- Need to identify the inter-arrival and service time distributions
- Queues with Poisson arrival
  - Exponential service time (Week 2)
    - Single or multiple servers: M/M/1 versus M/M/m
    - Infinite buffer or finite buffer: M/M/m versus M/M/m/m+k
  - General service time distribution (Week 4)
    - M/G/1. Key concept: residual service time
    - Priority queueing

Performance analysis techniques (3)

- Closed queueing networks with exponential service time
  - Markov chain analysis (Week 4)
    - Recipe: Identify state, transition probability, solve steady state probability, determine performance
  - Mean value analysis (Week 8)
    - Iterative method
    - $n = 0$ jobs $\rightarrow$ $n = 1$ job $\rightarrow$ $n = 2$ jobs $\rightarrow$ ...

Performance analysis techniques: some key points

- No universal analytical methods
  - Analytical solutions are only available for specific classes of queues
  - Upper bounds are only available for some general classes of queues
- Simulation can be used to determine general queueing problems

Simulation (1)

- How to do discrete event simulation?
- How to generate random events according to the specific inter-arrival time and service time distributions
  - Generating uniformly distributed pseudo-random numbers
  - Inverse transformation method

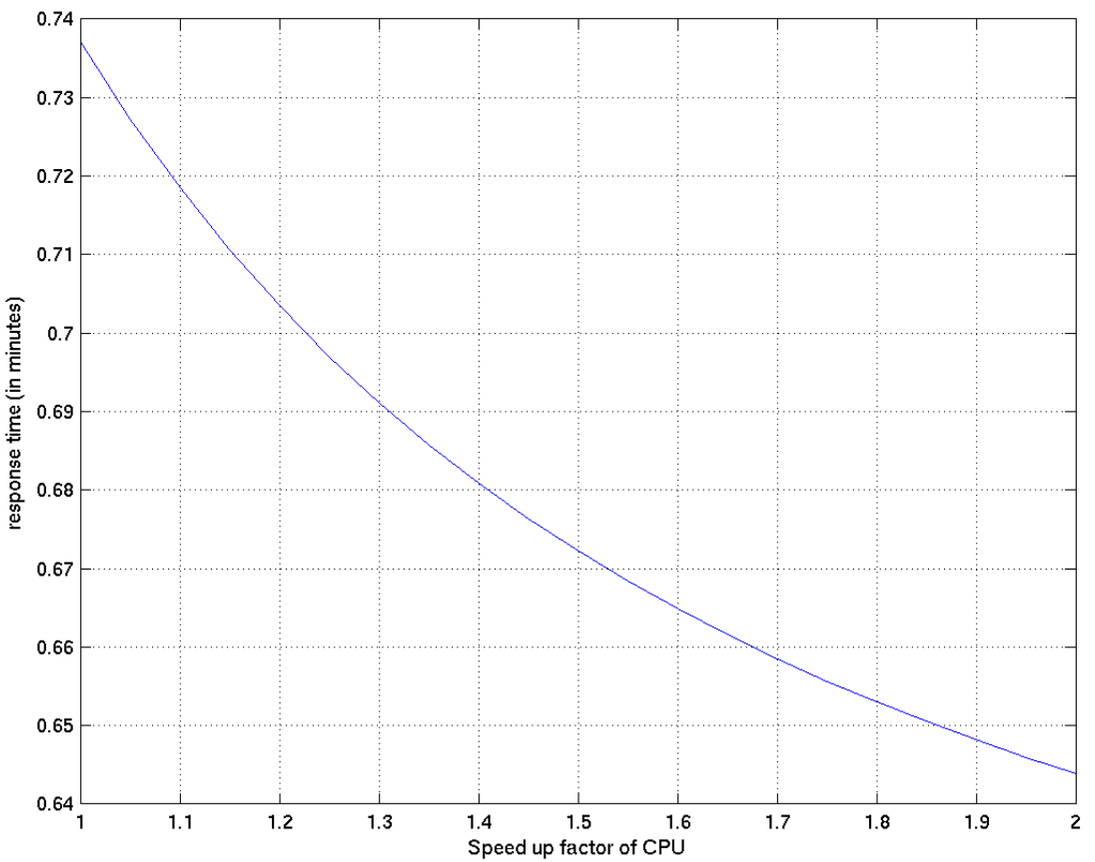
Simulation (2)

- Simulation is not just about writing correct simulation code (though it is important), it is also important to do sound statistical analysis on the simulation results obtained
- Transient removal
- Independent replications
- Confidence interval
- How to decide whether one system is better than the other using confidence interval?
  - Paired observations: Paired-t confidence interval
  - Approximate visual test

Capacity planning and performance analysis

- Solve the capacity planning problem by solving a number of performance analysis problems

- Example: Revision Problem: Week 4, Question2



Applications of queueing (Week 9)

- Web services
  - Fork-join queues
- Other applications
  - Determining a good multi-programming level
  - Power allocation in server farm
  - Server farm with set-up cost
- Processor sharing
  - If there are n jobs in a server, each job gets $\frac{1}{n}$ share of processing from the server

Integer programming (1)

- Linear programming (LP)
  - Real values for decision variables, linear in objective function, linear in constraints
- Integer programming (IP)
  - Some decision variables can only take integer values
  - Some decision variables can only take binary values, e.g. for making yes-or-no decisions

Integer programming (2)

- Applications of integer programming in network flow problems
  - Flow conservation constraints to ensure a unique path between two nodes in the network
- Example applications
  - Traffic engineering
  - Network design

Integer programming (3)

- Applications of integer programming in placement problems
  - Placement of wireless access points
  - Placement of controllers in software-defined networks (revision problem)
- Power of binary variables
  - Restricted range of values
  - Either-or constraints
  - Piecewise linear functions

Summary

- What you have learned through this course are fundamental techniques that can be applied to designing computer systems and networks to have good performance
  - We hope you have gained some skills from this course
  - We hope you have been trained to be performance-minded
- Due to the limited time and scope of this course, we cannot cover all techniques that have been developed in this field
- However, with the knowledge you have acquired from this course, you should have the foundation to learn more ...

Final exam

- Please check timetable to confirm date/time/venue
  - 3 hours + 10 minutes reading
- Open book
  - No programming questions
  - No extensive computation
- Question style similar to revision problems and assignment
  - Testing your understanding, not memorization
- Format
  - Answer 5 out of 7 questions
  - Each question can have multiple parts

Final exam (cont.)

- Show all equations, calculations and steps involved in problems
  - If you just write the final answer, you won't receive full marks even if it is correct
  - You may receive partial marks even if the final answer is incorrect
- What to bring
  - You must bring your own student ID, calculator, pencils, and erasers
  - You can bring textbook, lecture notes, handwritten/typed notes, sample problems/solutions, assignment problems/solutions
  - No electronics equipment (e.g. laptops, tablets) and watches are allowed

Preparations for final exam

- Make sure you understand all the concepts, techniques and examples discussed in the lectures
- Go through all the sample problems, assignment questions, etc for practice
- Misconception: Open-book exam means no preparation required
- Consultations: date and time to be announced
- Further questions
  - Post on Forum (Try to avoid last minute questions.)
  - E-mail cs9334@cse.unsw.edu.au

Parting messages

- Please complete the CATEI survey
  - Good/bad/more of this/less of that/what can be done better
- This course is different from many CSE courses ..
- Analytical and simulation methods are useful for many disciplines
- This world needs people with multiple skills. Important to find your talents and passions, but try to explore and learn as many different areas as you can.