ELE 549
Computer System Modeling

Bin Li
Dept. of Electrical, Computer & Biomedical Engineering
University of Rhode Island
Self-introduction

• Name

• M.S. or Ph.D. student

• Research topic and methodology
Transportation
Airport Security Lines
Amusement Parks
Common Features

• Customers experience a long wait 😞

• Reason???
  • Limited resources

• Solutions
  • Add more resources
  • Smart controls (algorithms)...

Cost vs. Performance
Contact Centers

Staff are all busy having a cup of coffee and a bit of a sit down ... please hold on, your call is important to us!
Computer Systems: Job Scheduling

Limit resources:
- CPU
- memory
- storage space
- I/O ports
- ...
Computer Systems: Web Servers

client 1

client 2

client 1000

"Get Page"

"Get Video"

"Get Audio"

Internet

Web Server
Queues Are Indeed Everywhere!
Common Features

• Systems with limited resources

• Smart scheduling algorithms

Cost vs. Performance
Limited resources: time, bandwidth, power
Cyber-Physical Systems

• Current intra-vehicular networks
  • 75 sensors and 75 switches
  • Requires 1,000 cables (45kg, 1km)
  • Low fuel efficiency
  • High cost and complexity of maintenance
  • $\sim 3 \times$ increases in sensors over 5 years

• Wireless solution
  • Limited resources: time, bandwidth, power
Mobile Edge Computing
Networking for Virtual/Augmented Reality
Big Data Processing

Climate Modeling

Deep Learning
How to Evaluate a System Design?

• Implementation and testbed/field deployment
  • Pros: high accuracy
  • Cons: Costly, difficult to repair/experiment in-field

• Simulations
  • Pros: can be accurate, given realistic models; broad applicability
  • Cons: can be slow, don’t always provide intuition behind results

• Analysis
  • Pros: Quick answers, provide insights
  • Cons: can be inaccurate or inapplicable
Course Goal

• Introduction to analytical tools which are needed to construct/analyze models of resource contention systems.
  • Computers
  • Networks

• It is a course on methodologies PLUS applications
Example 1: First-Come-First-Served (FCFS) server

Avg. arrival rate
\[ \lambda \text{ jobs/sec} \]
FCFS
Avg. service rate
\[ \mu \text{ jobs/sec} \]
Condition: \( \lambda < \mu \)

\( S \): job size (sec) = service requirement
\[ E[S] = \frac{1}{\mu} \]

Example:
- On average, job needs \( 3 \times 10^6 \) cycles
- Machine executes \( 9 \times 10^6 \) cycles/sec

Avg service rate
\[ \mu = 3 \text{ jobs/sec} \]
Avg size of job on this server
\[ E[S] = \frac{1}{3} \text{ sec} \]
Example 1: FCFS server (Cont’)

• **QUESTION**: If the arrival rate double, what service rate do you need to maintain the same delays for jobs?

(a) $2\mu$
(b) Less than $2\mu$
(c) More than $2\mu$
Example 2: Many slow or 1 fast?

\[ \lambda \text{ jobs sec} \rightarrow \begin{array}{c} \mu \\ \vdots \\ \mu \end{array} \]

\[ \rho = \frac{\lambda}{\mu} \]

\[ \lambda \text{ jobs sec} \rightarrow \begin{array}{c} \mu \\ \vdots \\ \mu \end{array} \]

\[ \rho = \frac{\lambda}{k \mu} \]

**QUESTION:** Which is better for minimizing E[T]?
Example 3: Job Scheduling

- FCFS (First-Come-First-Served, non-preemptive)
- PS (Processor-Sharing, preemptive)
- SJF (Shortest-Job-First, non-preemptive)
- SRPT (Shortest-Remaining-Processing-Time, preemptive)
- LAS (Least-Attained-Service-First, preemptive)

**QUESTION:** Which scheduling policy is best for minimizing $E[T]$?
Example 4: Dynamic Load Balancing

• Examples:
  • F5 Big-IP
  • Microsoft SharePoint
  • Cisco Local Director
  • Coyote Point Equalizer
  • IBM Network Dispatcher

**QUESTION**: What is a good dispatching policy for minimizing $E[T]$?
Example 4: Dynamic Load Balancing (Cont’)

• Round-Robin

• Join-Shortest-Queue

• Least-Work-Left

• ... ...
Time-Varying Arrivals

Call center

Web servers
Course Outline

• The formulation & analysis of these models relies on
  • Stochastic processes
  • Queueing theory
  • Statistics
  • Scheduling
  • Optimization
  • Game theory
  • Control theory
  • Machine learning
  • ....
Textbook

Other References


Grading

- Homework: 10%
- Midterm: 30%
- Final: 30%
- Project: 30%