COMP9334
Capacity Planning for Computer Systems and Networks

Week 3A: Revision problems

Note

- Some of these questions can be done by a calculator but some of them require laborious calculations that are best done by a computer software, e.g. Matlab, Octave, Python etc.

Question 1

- You have a computer system with a single CPU.
  - Both inter-arrival and service times are exponentially distributed.
  - The request only requires services at the CPU.
  - Each request only visits the CPU once.
  - A finished request will leave the system.
  - Mean arrival rate is 9 request/s
  - Mean service time required by a request at the CPU is 0.1s.
- What is the utilisation of the CPU?
- What is the mean response time?
- The utilisation is pretty high and you want to change the system. You can think of 3 alternatives.

Question 1 - Alternative 1

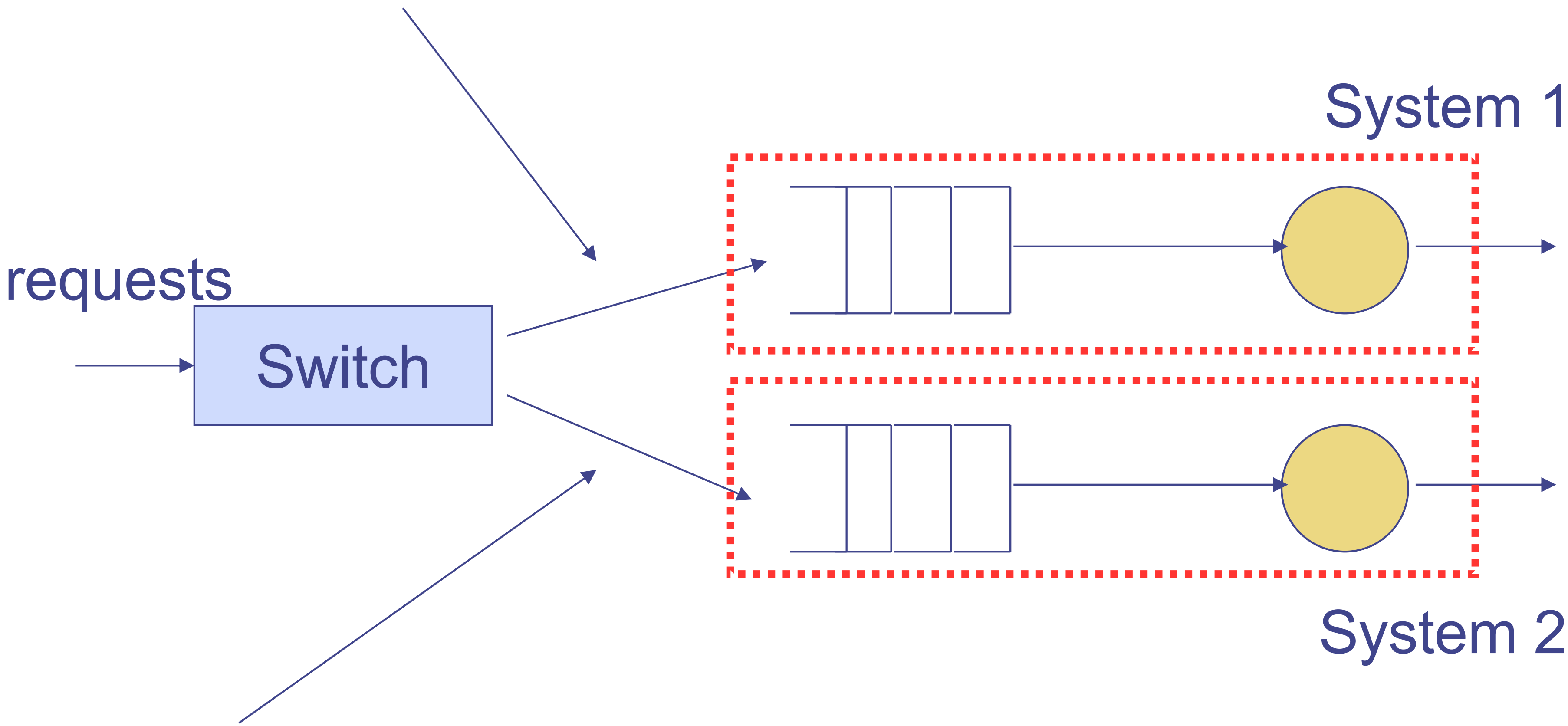
- Replace the existing CPU by one that is 2 times faster
- You may assume that the service time is inversely proportional to CPU speed.

Question 1 - Alternative 2

- Buy a system which is identical to the current one
- Put the two systems in parallel
- Add a switch in front of the system
- When a request arrives, the switch will randomly assign the request to one of the systems. On average, half of the request goes to each system
- (Pictorial representation on the next slide)
- Assume the switch requires negligible time

Question 1 - Alternative 2 (cont’d)

Half of the requests to system 1

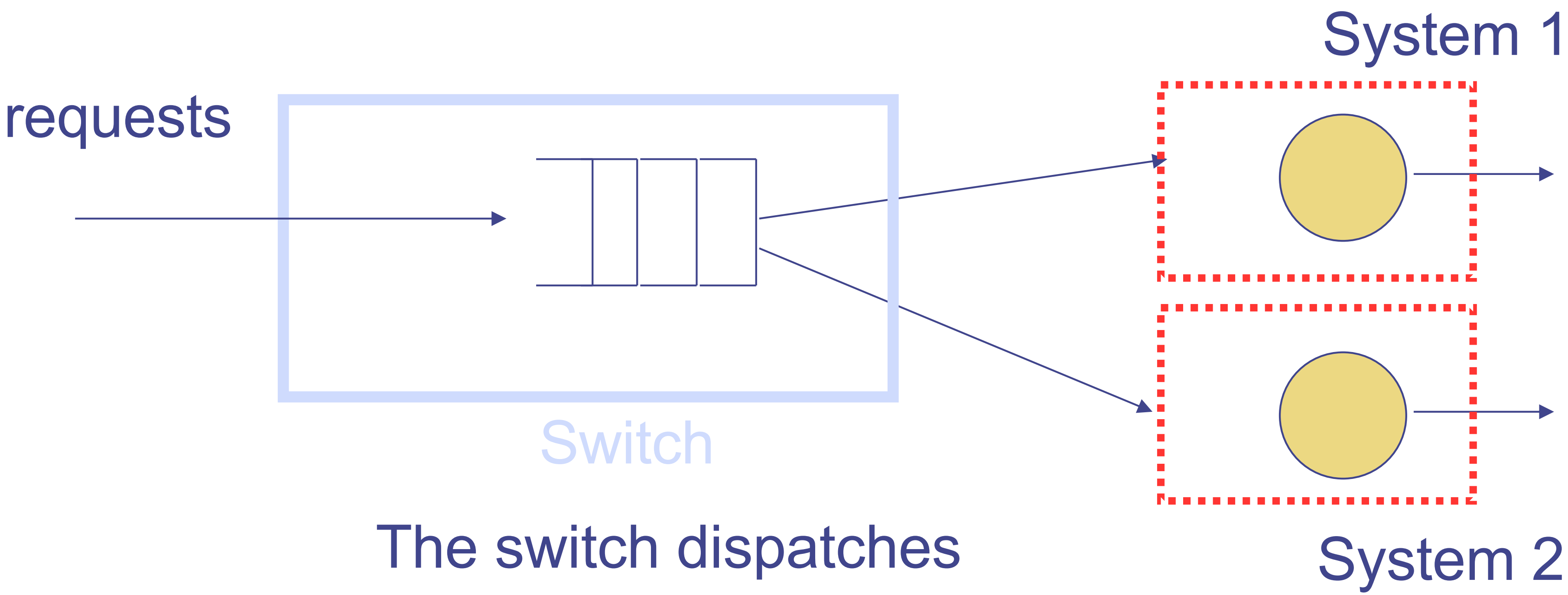


Half of the requests to System 2

Question 1 - Alternative 3

- Similar to Alternative 2, we buy a system which is identical to the current one and we also buy a switch
- However, we only maintain a queue at the switch
- If both systems are busy, the request waits at the switch; otherwise, the switch dispatches the request to any of the available systems
- Assuming that it takes negligible time for the switch to find out whether a system is idle
- (Pictorial representation on the next page)

Question 1 - Alternative 3 (cont’d)

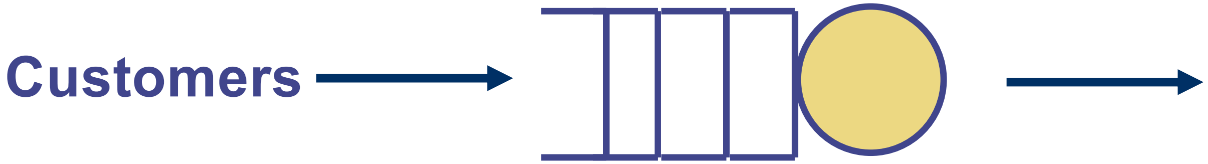


The switch dispatches the requests to the first available server.

Question 1

- Part (a): Calculate the resulting mean response time for each for the three alternatives
- Part (b): Repeat part (a) for a number of different mean arrival rates. Plot a graph of arrival rates against the mean response time.
- Part (c): What observations can you make from these calculations?
- Part (d): What is the best way to upgrade the system in terms of performance? However, the best way to upgrade in terms of performance may not be the best way to upgrade in terms of cost, why?

Question 2



- Consider a single server queue as shown above
- Part (a): Consider the situation
    - The inter-arrival time is a constant and is given by 1 second.
    - The service time required by each customer is always 0.5 second.
  - What is the mean waiting time per customer?
- Part (b): Consider the situation
    - The inter-arrival time is exponentially distributed with mean 1 second
    - The service time required by each customer is exponentially distributed with mean 0.5s
  - What is the mean waiting time per customer?
- Compare the answers of Parts (a) and (b). What conclusions can you draw?

Question 3

- An Internet Service Provider has 4 dial-up ports. Connection requests obey Poisson distribution with a mean arrival rate of 3 requests per hour. The session duration of each connection request is exponentially distributed with a mean of 1.5 hours. What is the probability that a connection request will be rejected?