

Lecture 3b: Queuing Theory Exercises

PAMS'18

Zsolt István

zsolt.istvan@imdea.org

Exercise set 1

- A database server is serving user queries. On average each query is answered in 3ms. How many queries/s can we submit to the server without making it unstable?
- Instrumentation shows that the CPU core running an application is busy 80% of the time. The jobs we submit to this application take on average 1.5s to finish. What is the service rate of this application?
- A dual-disk RAID system is serving read requests of some fixed size arriving at the rate of 10000/s. It's service rate is 15000/s/disk. What is the expected number of jobs in the system?

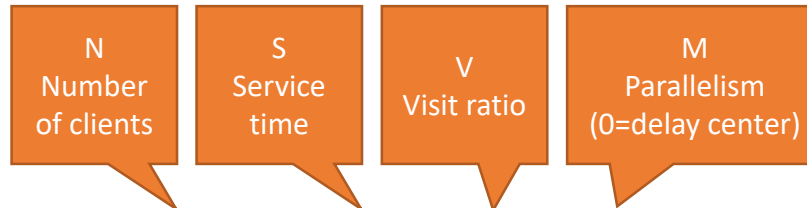
Exercise set 2

- We are benchmarking an application we see the following behavior:
 - If we use a single client and issue requests in a loop, it achieves 750ops/s
 - If the arrival rate is 500ops/s, the response time is 2ms
- What model would you use to describe this system?
- What are its parameters?

Exercise set 3

- A cloud-based server uses a dual core CPU, and based on our experiments it can serve at most 500req/s. If we measure the average ping time to this server, we see that it is 30ms. How can we model this system?
- What is the response time of this system with $N=2$ clients?

Octave example



```
>> [U, R, Q, X, G] = qncsmva (1, [0.03, 0.004], [1, 1], [0, 2])
```

U =

0.882353 0.058824

Utilization

R =

0.0300000 0.0040000

Response time

Q =

0.88235 0.11765

Jobs in "device"

X =

29.412 29.412

Throughput

```
>> [U, R, Q, X, G] = qncsmva (2, [0.03, 0.004], [1, 1], [0, 2])
```

U =

1.76471 0.11765

R =

0.0300000 0.0040000

Q =

1.76471 0.23529

X =

58.824 58.824

Octave example

