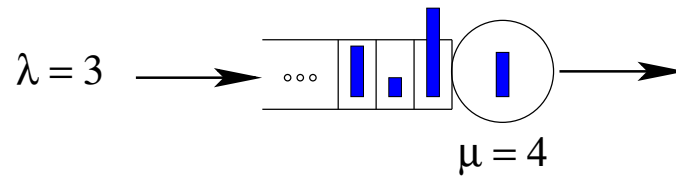


A single-server queue



Average Arrival Rate:

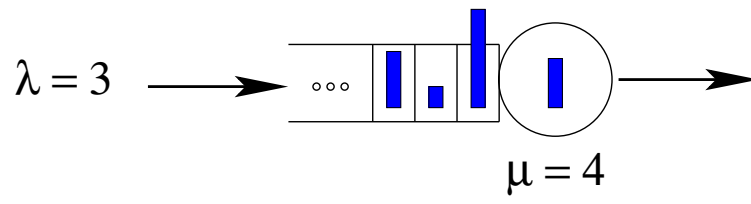
Interarrival Time:

Mean Interarrival Time:

Job Size (Service Requirement):

Mean Job Size:

Average Service Rate:



PEOPLE SPEAK

vs.

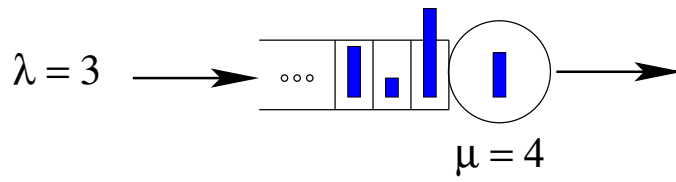
QUEUEING SPEAK

Examples from your work:

Common Performance Metrics

- Response Time, T :
- Waiting Time or Delay, T_Q
- Number of jobs in system, N
- Number of jobs in queue, N_Q

Stability



Question: What happens if $\lambda > \mu$?

Here's why:

$N(t)$ = Number jobs in system *at* time t

$A(t)$ = Number arrivals *by* time t

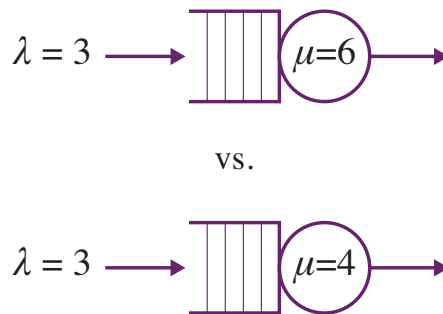
$C(t)$ = Number completions(departures) *by* time t

We will always assume $\lambda < \mu$. (Stability)

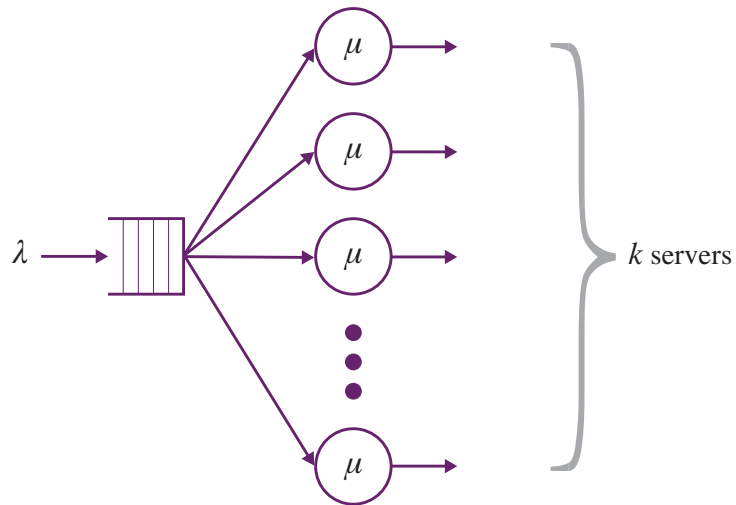
Throughput, X

Question: What is throughput?

Question: Which has higher throughput?



Throughput for Server Farm

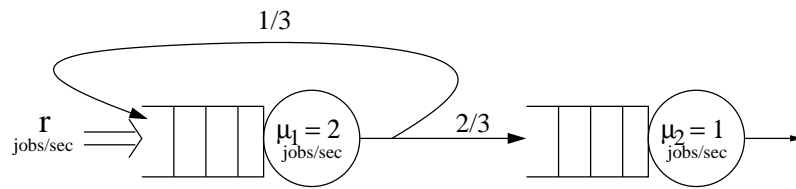


Question: What condition is needed for stability?

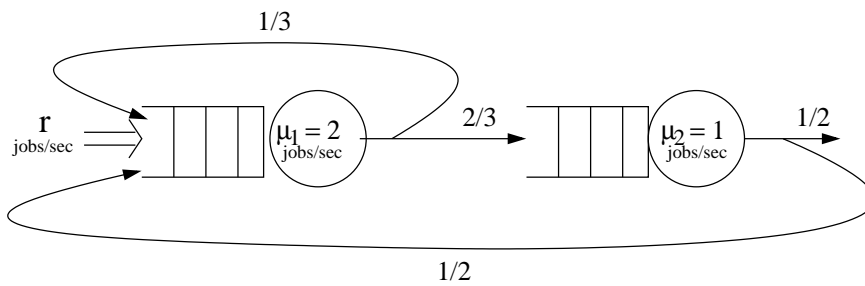
Question: What's the throughput? (assuming stability)

Throughput for Network of Queues

Question: What is the maximum outside arrival rate?

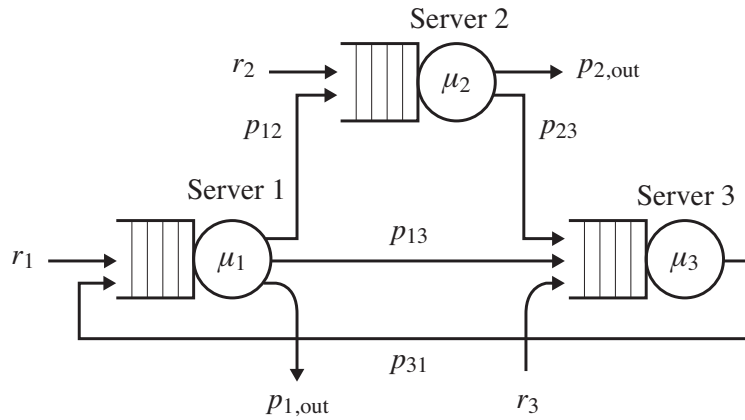


(a)



(b)

Throughput for Network of Queues



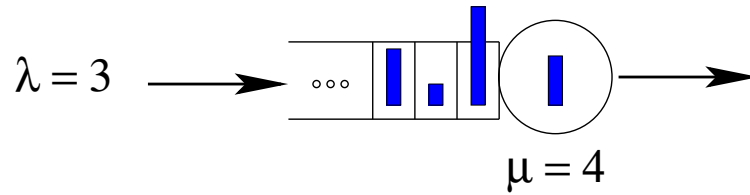
Question: What's the throughput of this system? (assume stability)

Question: What's the throughput of server i ?

Question: What do we need for stability of this system?

Device Utilization (Load)

When talking about “utilization,” we’re thinking of a single device.



Defn: Device **utilization**, a.k.a. **load**, is the long-run fraction of time that the device is busy.

Let $B(t)$ = total time server is busy during $[0, t]$.

Q: What is ρ ?

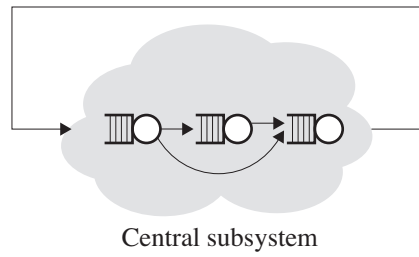
Example:

3 jobs/sec arrive on average.

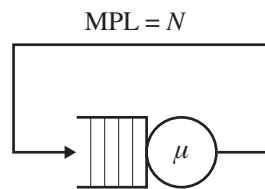
Each job contributes 1/4 sec of work on average.

Q: What is ρ ?

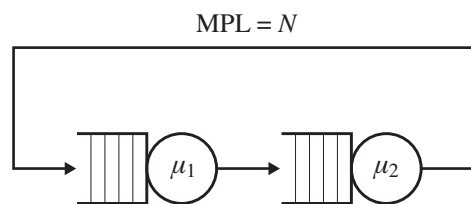
Closed System (Batch)



Question: What is throughput below?



Question: What is throughput below?



Closed System (Interactive)

