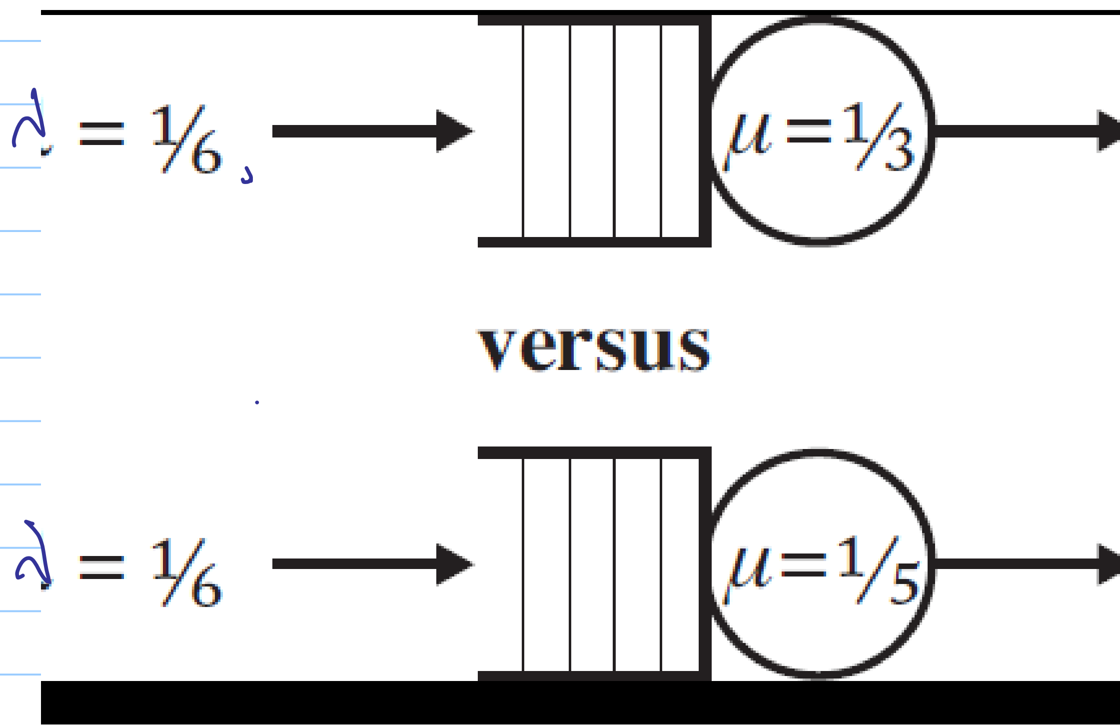
317 2016-01-21

Note Title 2016-01-21

2.5 [H] Throughput & Utilization



Which of these has the higher throughput?

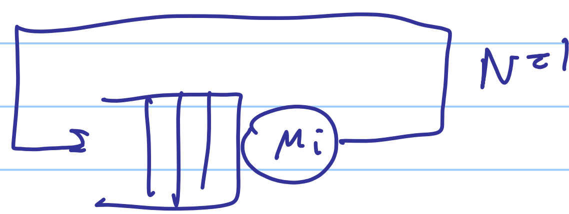
They are the same ($X = \lambda$)

Defn. Device utilization for device i, ρ_i, (ρ : rho) is the fraction of time that device i is busy.

$\rho_i = \frac{B}{\tau}$, where τ is a long time interval and B is the total busy time for device i within interval τ.

Device throughput for device i, X_i, is the rate of completions at device i (jobs/sec)

A $\boxed{\text{very}}$ special case



$\mu_i = X_i$
(no time in queue, no delays)

$$X_i = \frac{C}{\tau}$$, where τ is a long time interval, and C is the number of job completions within τ.

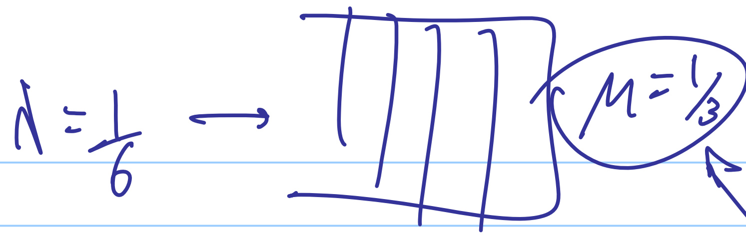
Relationship between device utilization ρ_i and device throughput, X_i

$$X_i = \frac{C}{\tau} = \frac{C}{B} \cdot \frac{B}{\tau} = \frac{C}{B} \cdot \rho_i = \mu_i \cdot \rho_i = \frac{1}{E[S_i]} \cdot \rho_i$$

μ_i: avg service rate

$E[S_i]$: exp. service time

$$\underset{1}{\rho_i} = \underset{\frac{\text{jobs}}{\text{sec}}}{X_i} \cdot \underset{\frac{\text{sec}}{\text{jobs}}}{E[S_i]}$$ Utilization law



What is the throughput?

$X = \rho\mu$

μ is given. What is ρ?

This is suspect (not proof!)

ρ = fraction of time the server is busy =

$$= \frac{\text{average service time required by a job}}{\text{average time between arrivals}} =$$

$$= \frac{1/\mu}{1/\lambda} = \frac{\lambda}{\mu} = \frac{\text{avg arrival rate}}{\text{avg service rate}}$$

proof is done on p. 100 [H] (Corollary 6.5)

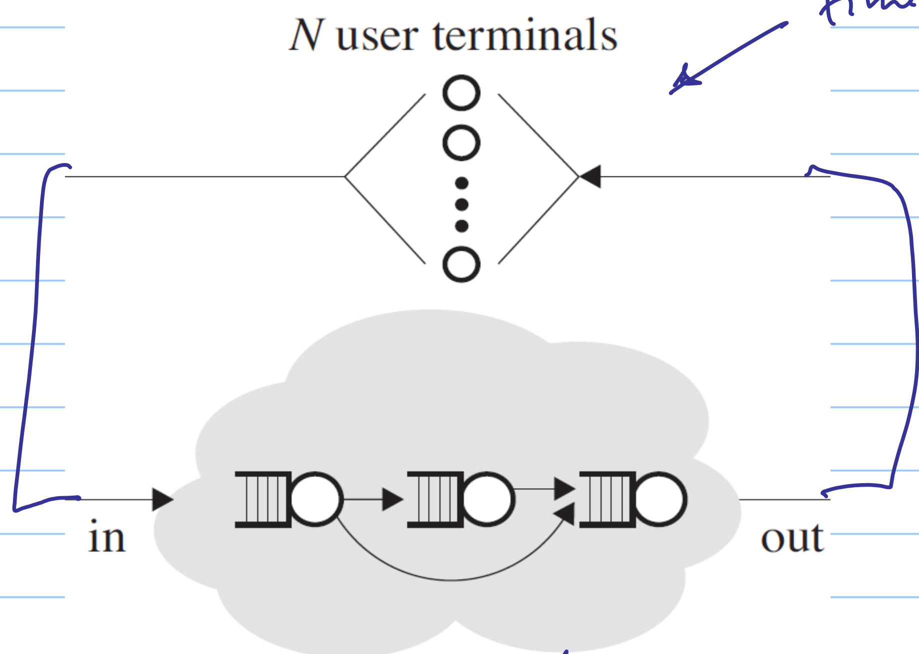
$X = \rho\mu = \frac{\lambda}{\mu}\mu = \lambda$. So, the throughput is not dependent on the service rate. The throughput for the simple one-queue, one-server open system  is the arrival rate. (!)

Closed Networks (2.6 [H])

Closed networks
- Interactive (Terminal-driven)
- Batch

2.6.1 Interactive (Terminal-driven) Systems

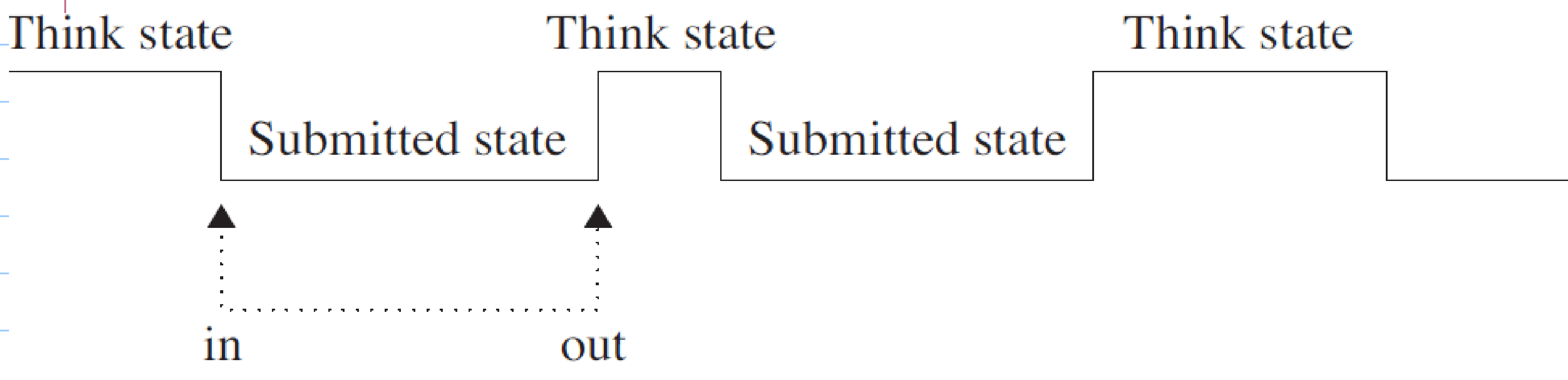
time spent here is the think time, Z



Central subsystem

Response time is time between "in and out"

An individual user (e.g., call center operator) alternates between think state & submitted state.

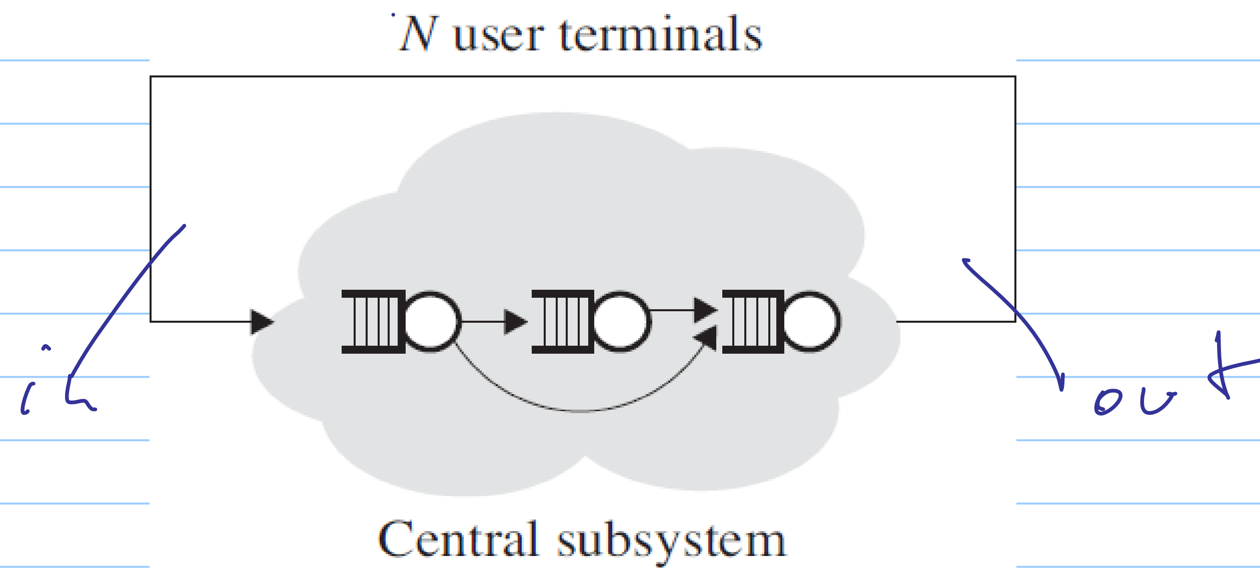


Time in system (for a job) is the sum of think time and response time.

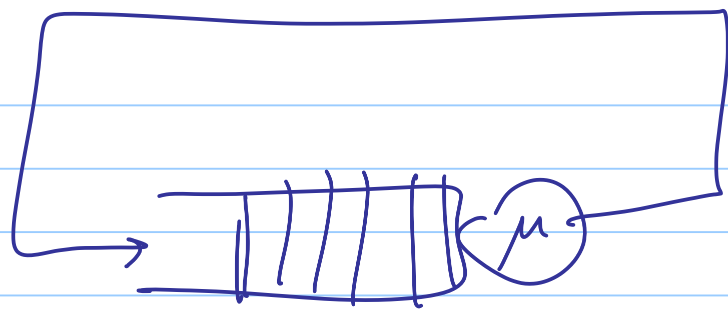
$$E[T] = E[R] + E[Z]$$

(time in system for a job) (between in and out) (i.e. in the central subsystem)

2.6.2 Batch Systems



MPL = N



Single-server closed batch system

$$\mu = X = \frac{1}{E[S]} = \frac{1}{E[R]}$$

Tandem system (a closed batch system)