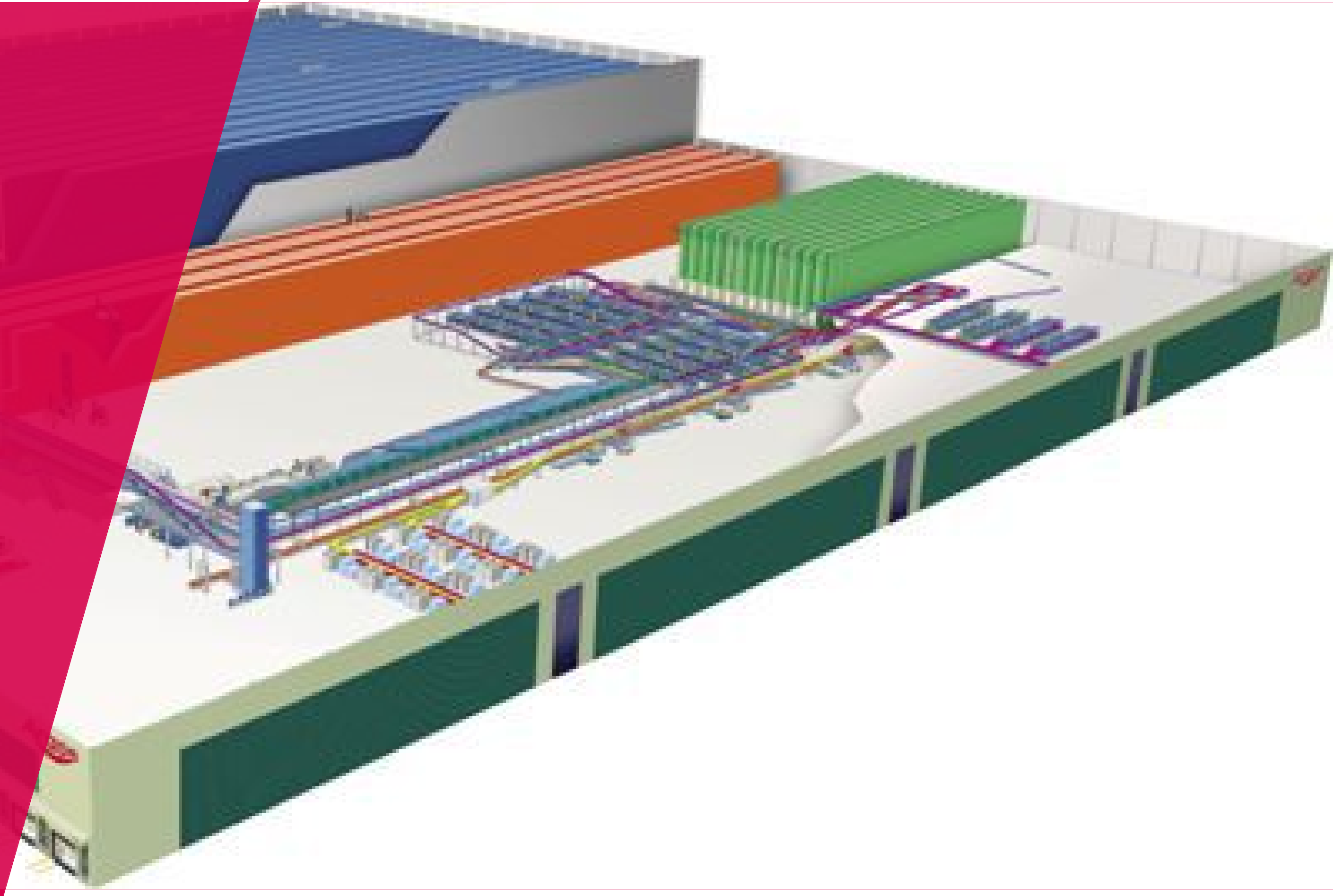
Technische Universiteit
Eindhoven
University of Technology

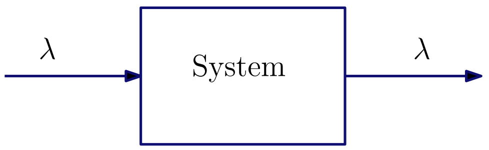
Facility Logistics Management

Ivo Adan



Fundamental relations: Little's law

- $E(L)$ is the mean number in the system
- $E(S)$ is the mean time spent in the system
- λ is the arrival rate (or departure rate or throughput)



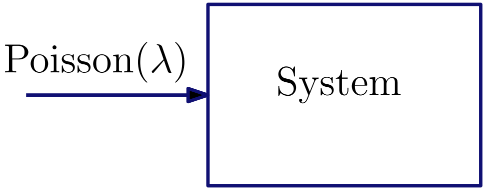
- Little's law

$$E(L) = \lambda E(S) \quad \text{or} \quad \lambda = \frac{E(L)}{E(S)}$$

- According to Little's law: Same throughput λ can be achieved with
  - Large number in system $E(L)$ and long mean time in system $E(S)$
  - Small number in system $E(L)$ and short mean time in system $E(S)$
- Question: What causes the difference?
- Answer: Variability!

Fundamental relations: PASTA Property of Poisson arrivals

- Poisson arrivals with rate λ



- PASTA Property: Poisson Arrivals See Time Averages

Fraction of arrivals that see the system in state A = Fraction of time the system is in state A

- Example: Q^a is number in queue on arrival, Q is number in queue at arbitrary point in time

$$P(Q^a = k) = P(Q = k), \quad k = 0, 1, 2, \ldots$$

$$E(Q^a) = \sum_{k=0}^{\infty} kP(Q^a = k) = \sum_{k=0}^{\infty} kP(Q = k) = E(Q)$$

Single Exponential server

- Poisson arrivals with rate λ and single server with Exponential service times with rate μ
- Stability: $\rho = \lambda/\mu < 1$
- p_k is long-run probability (or fraction of time) of finding k jobs in the system
- p_k is Geometric:

$$p_k = (1-\rho)\rho^k, \quad k = 0, 1, \ldots$$

- Mean performance

$$\begin{aligned} E(L) &= \sum_{k=0}^{\infty} k p_k = \frac{\rho}{1-\rho} \\ E(S) &= E(L)/\lambda = \frac{1}{1-\rho}\frac{1}{\mu} \\ E(Q) &= \sum_{k=1}^{\infty} (k-1) p_k = \frac{\rho^2}{1-\rho} \\ E(W) &= E(Q)/\lambda = \frac{\rho}{1-\rho}\frac{1}{\mu} \end{aligned}$$

Single Exponential server

- Poisson arrivals with rate λ and single server with Exponential service times with rate μ
- Stability: $\rho = \lambda/\mu < 1$
- Mean performance through PASTA and Little:

$$\mathrm{E}(S) = \mathrm{E}(L^a)\frac{1}{\mu} + \frac{1}{\mu}$$

where L^a is number on arrival

- PASTA: $\mathrm{E}(L^a) = \mathrm{E}(L)$

$$\mathrm{E}(S) = E(L)\frac{1}{\mu} + \frac{1}{\mu}$$

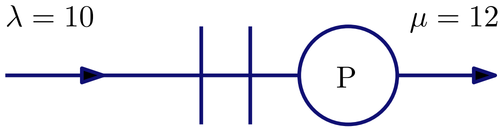
- Little's law: $\mathrm{E}(L) = \lambda E(S)$

$$\mathrm{E}(S) = \mathrm{E}(S)\rho + \frac{1}{\mu}$$

so

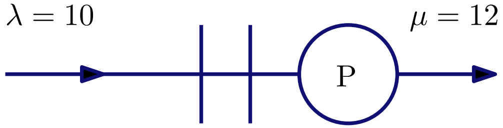
$$\mathrm{E}(S) = \frac{1}{1-\rho}\frac{1}{\mu}$$

Example: Manual order picking station



- Customer orders arrive at an order picker according to a Poisson stream with rate $\lambda = 10$ orders per hour
- It takes the order picker an Exponential time to collect all items for an order
- Mean pick time is 5 minutes, so pick rate is $\mu = 12$ orders per hour
- Question: What is the utilization of the picker?
- Answer: $\rho = \frac{\lambda}{\mu} = \frac{10}{12} = \frac{5}{6}$

Example: Manual order picking station



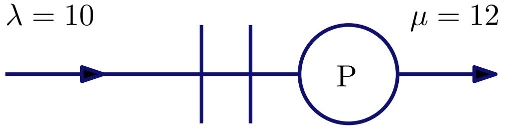
- Customer orders arrive at an order picker according to a Poisson stream with rate $\lambda = 10$ orders per hour
- It takes the order picker an Exponential time to collect all items for an order
- Mean pick time is 5 minutes, so pick rate is $\mu = 12$ orders per hour
- Question: What is the probability that there are at least 2 orders at the pick station?
- Answer:

$$p_n = (1 - \rho)\rho^n$$

so the probability of at least 2 orders at the pick station is

$$p_2 + p_3 + \cdots = 1 - p_0 - p_1 = 1 - (1 - \rho) - (1 - \rho)\rho = \rho^2 = \frac{25}{36}$$

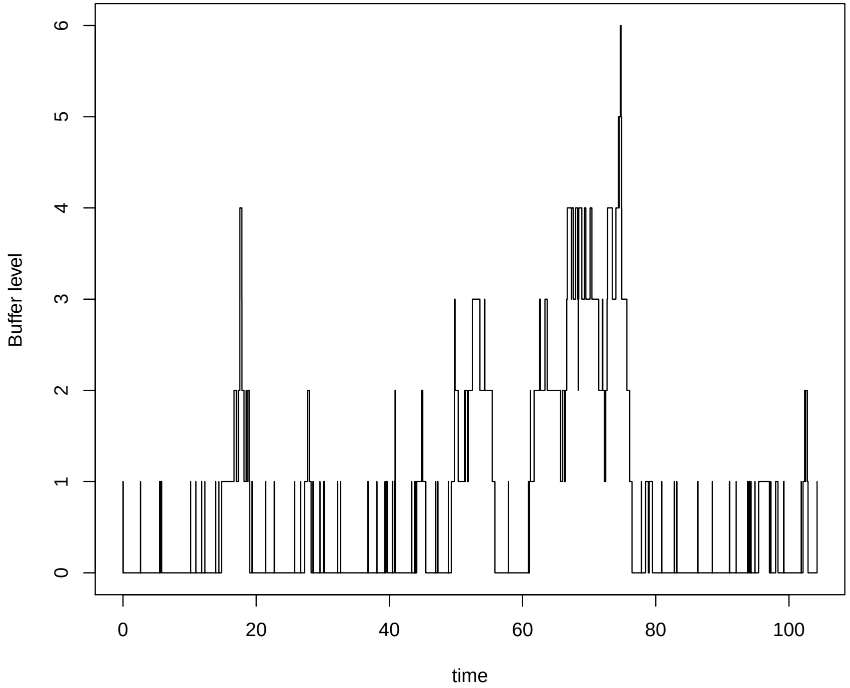
Example: Manual order picking station



- Customer orders arrive at an order picker according to a Poisson stream with rate $\lambda = 10$ orders per hour
- It takes the order picker an Exponential time to collect all items for an order
- Mean pick time is 5 minutes, so pick rate is $\mu = 12$ orders per hour
- Question: What is the mean number of orders at the pick station and what is the mean flow time?
- Answer:

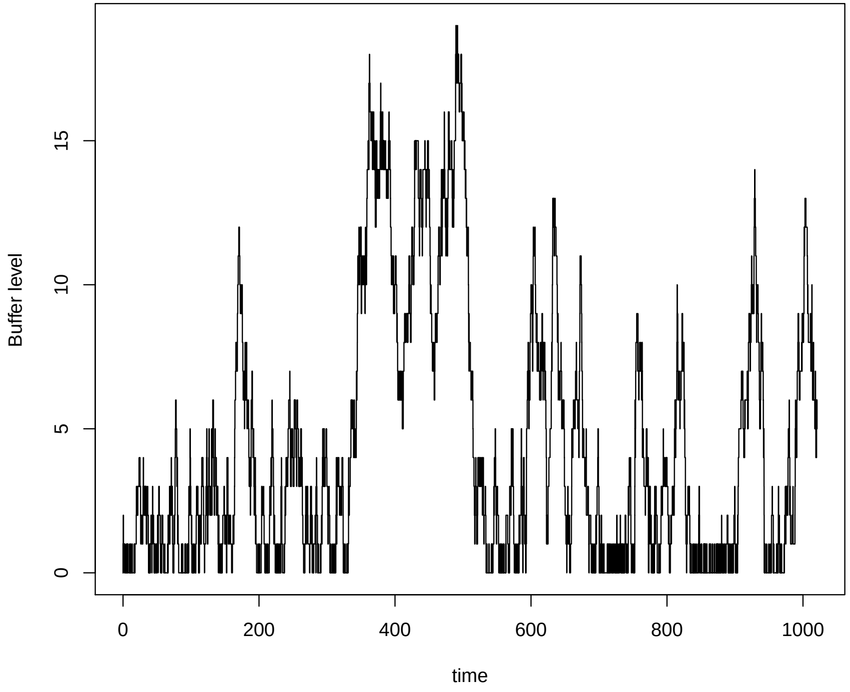
$$E(L) = \frac{\rho}{1-\rho} = 5, \quad E(S) = \frac{1}{1-\rho}\frac{1}{\mu} = \frac{1}{2}$$

Number in system over time



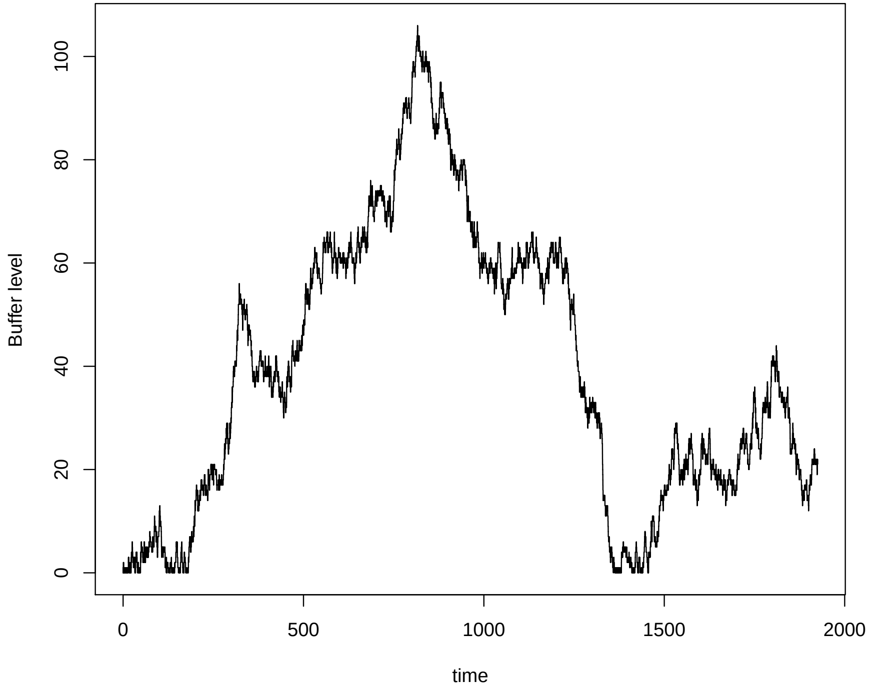
Single Exponential server, $\lambda = 1.0$, $1/\mu = 0.5$, $\rho = 0.5$

Number in system over time



Single Exponential server, $\lambda = 1.0$, $1/\mu = 0.9$, $\rho = 0.9$

Number in system over time



Single Exponential server, $\lambda = 1.0$, $1/\mu = 0.95$, $\rho = 0.95$

Single Exponential server

- Poisson arrivals with rate λ
- Single server with Exponential service times B, mean $E(B) = \frac{1}{\mu}$
- Stability: $\rho = \lambda E(B) < 1$
- Mean waiting time

$$E(W) = \frac{\rho}{1-\rho} E(B)$$

Single General server

- Poisson arrivals with rate λ
- Single server with General service times B, density $f_B(x)$, mean $\mathrm{E}(B)$, variance $\sigma^2(B)$
- Stability: $\rho = \lambda \mathrm{E}(B) < 1$
- Mean waiting time

$$\mathrm{E}(W) = \mathrm{E}(Q^a)\mathrm{E}(B) + \rho \mathrm{E}(R)$$

  where Q^a is number in queue on arrival and R is residual service time
- PASTA: $\mathrm{E}(Q^a) = \mathrm{E}(Q)$

$$\mathrm{E}(W) = \mathrm{E}(Q)\mathrm{E}(B) + \rho \mathrm{E}(R)$$

- Little's law: $\mathrm{E}(Q) = \lambda E(W)$

$$\mathrm{E}(W) = \frac{\rho}{1-\rho}\mathrm{E}(R)$$

- Question: What is $\mathrm{E}(R)$?

Mean residual service time

- X is randomly selected service time

$$P(x \leq X \leq x + dx) = f_X(x)dx = Cxf_B(x)dx$$

- C is normalizing constant

$$C = \frac{1}{\int_0^\infty x f_B(x)dx} = \frac{1}{E(B)}$$

-

$$f_X(x) = \frac{xf_B(x)}{E(B)}, \quad E(X) = \int_0^\infty x f_X(x)dx = \frac{1}{E(B)} \int_0^\infty x^2 f_B(x)dx = \frac{E(B^2)}{E(B)}$$

- Conclusion

$$E(R) = \frac{1}{2}E(X) = \frac{E(B^2)}{2E(B)} = \frac{1}{2}(1 + c_B^2)E(B)$$

- c_B is coefficient of variation of B

$$c_B = \frac{\sigma(B)}{E(B)}$$

Single General server

- Poisson arrivals with rate λ
- Single server with General service times B, mean $\mathrm{E}(B)$, variance $\sigma^2(B)$, cv $c_B = \sigma(B)/\mathrm{E}(B)$
- Stability: $\rho = \lambda \mathrm{E}(B) < 1$
- Mean waiting time

$$\mathrm{E}(W) = \frac{\rho}{1-\rho}\mathrm{E}(R) = \frac{1}{2}(1 + c_B^2)\frac{\rho}{1-\rho}\mathrm{E}(B)$$

Single General server with General arrivals

- General arrivals with inter-arrival times A, mean $\mathrm{E}(A)$, variance $\sigma^2(A)$, cv $c_A = \sigma(A)/\mathrm{E}(A)$
- Single server with General service times B, mean $\mathrm{E}(B)$, variance $\sigma^2(B)$, cv $c_B = \sigma(B)/\mathrm{E}(B)$
- Stability: $\rho = \mathrm{E}(B)/\mathrm{E}(A) < 1$
- Mean waiting time

$$\mathrm{E}(W) \approx \frac{1}{2}(c_A^2 + c_B^2)\frac{\rho}{1-\rho}\mathrm{E}(B)$$

- Mean waiting time $\mathrm{E}(W)$ separated into three terms $V \times U \times T$
  - Dimensionless Variability term $\frac{1}{2}(c_A^2 + c_B^2)$
  - Dimensionless Utilization term $\frac{\rho}{1-\rho}$
  - Mean service Time term $\mathrm{E}(B)$

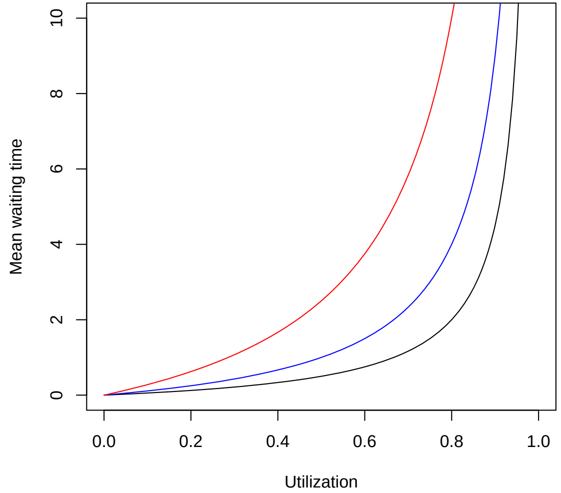
Single General server with General arrivals: Observations

$$E(W) \approx \frac{1}{2}(c_A^2 + c_B^2)\,\frac{\rho}{1-\rho}\,E(B)$$

- Poisson arrivals
  Approximation is exact for Poisson arrivals
- Insensitivity
  Mean waiting time only depends on mean and standard deviation of inter-arrival times and process times
- Heavy load
  As the utilization ρ tends to 1, then:
  - Mean waiting time $E(W)$ tends to ∞
  - Relative error of approximation tends to 0
  - Distribution of waiting time converges to Exponential

  So when the system operates close to $\rho = 1$, the waiting times are long and Exponential

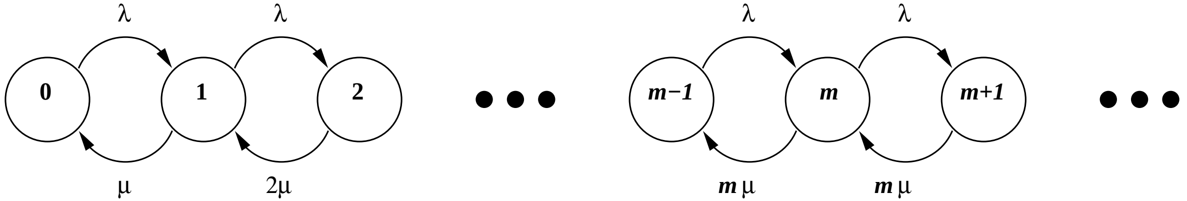
Single General server with General arrivals



Mean waiting time $\mathrm{E}(B)$ as function of utilization ρ for $c_A = 1$ and $c_B = 0, 1, 2$

Multi Exponential servers

- Poisson arrivals with rate λ
- m servers with Exponential service times with rate μ
- Stability: $\rho = \lambda/(m\mu) < 1$ where ρ is utilization of server
- p_i is long-run probability (or fraction of time) of finding i jobs in the system
- Flow diagram



- Balance equations: Flow from state i to $i-1$ = Flow from state $i-1$ to i
- Balance equations:

$$
\begin{aligned}
p_i i\mu &= p_{i-1}\lambda \quad i \le m \\
p_i m\mu &= p_{i-1}\lambda \quad i > m
\end{aligned}
$$

Multi Exponential servers

- Balance equations: $p_i i\mu = p_{i-1}\lambda \quad i \leq m$

$$p_i = p_{i-1}\frac{\lambda}{i\mu} = p_{i-2}\frac{1}{i(i-1)}\left(\frac{\lambda}{\mu}\right)^2 = \cdots = p_0\frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i = p_0\frac{1}{i!}(m\rho)^i$$

- Balance equations: $p_i m\mu = p_{i-1}\lambda \quad i > m$

$$p_i = p_{i-1}\frac{\lambda}{m\mu} = p_{i-2}\left(\frac{\lambda}{m\mu}\right)^2 = \cdots = p_m\left(\frac{\lambda}{m\mu}\right)^{i-m} = p_m\rho^{i-m} = p_0\frac{1}{m!}(m\rho)^m\rho^{i-m}$$

- p_0 is probability that system is empty and follows from normalization $p_0 + p_1 + \cdots = 1$

$$\frac{1}{p_0} = \sum_{i=0}^{m-1}\frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!}\frac{1}{1-\rho}$$

- Π_W is probability that all servers are busy (or the probability of waiting on arrival)

$$\Pi_W = p_m + p_{m+1} + \cdots = \frac{(m\rho)^m}{m!}\left((1-\rho)\sum_{i=0}^{m-1}\frac{(m\rho)^i}{i!} + \frac{(m\rho)^m}{m!}\right)^{-1} \approx \rho^{\sqrt{2(m+1)}-1}$$

Multi Exponential servers

- Poisson arrivals with rate λ
- m servers with Exponential service times, mean $E(B) = \frac{1}{\mu}$
- Stability: $\rho = \lambda E(B)/m < 1$
- Mean waiting time

$$E(W) = E(Q)\frac{E(B)}{m} + \Pi_W \frac{E(B)}{m}$$

- Little's law: $E(Q) = \lambda E(W)$

$$E(W) = E(W)\rho + \Pi_W \frac{E(B)}{m}$$

so

$$E(W) = \frac{\Pi_W}{1-\rho}\frac{E(B)}{m}$$

Multi Exponential servers

- Poisson arrivals with rate λ
- m servers with Exponential service times, mean $E(B) = \frac{1}{\mu}$
- Stability: $\rho = \lambda E(B)/m < 1$
- Mean waiting time

$$E(W) = \frac{\Pi_W}{1-\rho}\frac{E(B)}{m}$$

Multi General servers

- Poisson arrivals with rate λ
- m servers with General service times B, mean $E(B)$, variance $\sigma^2(B)$, cv $c_B = \sigma(B)/E(B)$
- Stability: $\rho = \lambda E(B)/m < 1$
- Mean waiting time

$$E(W) \approx \frac{\Pi_W}{1-\rho}\frac{E(R)}{m} = \frac{1}{2}(1+c_B^2)\frac{\Pi_W}{1-\rho}\frac{E(B)}{m}$$

- Π_W is probability of waiting in corresponding exponential system
  - Π_W is fairly insensitive to distribution of service time
  - Corresponding means with same mean service times

Multi General servers with General arrivals

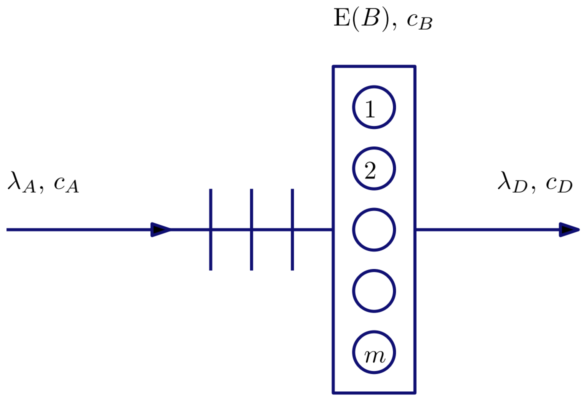
- General arrivals with inter-arrival times A, mean $\mathrm{E}(A)$, variance $\sigma^2(A)$, cv $c_A = \sigma(A)/\mathrm{E}(A)$
- m servers with General service times B, mean $\mathrm{E}(B)$, variance $\sigma^2(B)$, cv $c_B = \sigma(B)/\mathrm{E}(B)$
- Stability: $\rho = \lambda \mathrm{E}(B)/m < 1$
- Mean waiting time

$$\mathrm{E}(W) \approx \frac{1}{2}(c_A^2 + c_B^2)\frac{\Pi_W}{1-\rho}\frac{\mathrm{E}(B)}{m}$$

- Mean waiting time $\mathrm{E}(W)$ separated into three terms $V \times U \times T$
  - Dimensionless Variability term $\frac{1}{2}(c_A^2 + c_B^2)$
  - Dimensionless Utilization term $\frac{\Pi_W}{1-\rho}$
  - Mean service Time term $\frac{\mathrm{E}(B)}{m}$

Departure flow variability

- General arrivals with inter-arrival times A, rate $\lambda_A = 1/\mathrm{E}(A)$, coefficient of variation c_A
- m servers with General service times B, mean $\mathrm{E}(B)$, coefficient of variation c_B
- Departures with inter-departure times D, rate $\lambda_D = 1/\mathrm{E}(D)$, coefficient of variation c_D



- Question: What is relation between rate and variability of arrival flow and departure flow?

Departure flow variability

- Conservation of flow: $\lambda_D = \lambda_A$
- Variability in departures depends on variability in arrivals and process times
- Relative contribution of variability in arrivals and process times depends on the utilization of servers

$$\rho = \frac{\lambda_A \mathsf{E}(B)}{m}$$

- Single server ($m = 1$)

$$c_D^2 \approx (1-\rho^2)c_A^2 + \rho^2 c_B^2$$

- Multi servers ($m > 1$)

$$c_D^2 \approx 1 + (1-\rho^2)(c_A^2 - 1) + \frac{\rho^2}{\sqrt{m}}(c_B^2 - 1)$$

Departure flow variability

- Successive times between departures are typically not independent
- Assumption of independence is usually a reasonable approximation
- If
  - Arrivals are Poisson ($c_A = 1$)
  - Service times are Exponential ($c_B = 1$)

  then

  - Departures are also Poisson ($c_D = 1$)

Departure flow variability

$$c_D^2 \approx 1 + (1-\rho^2)(c_A^2 - 1) + \frac{\rho^2}{\sqrt{m}}(c_B^2 - 1)$$

- Question: Does the approximation for c_D make sense?
  - If $\rho \approx 1$ and $m = 1$, then server nearly always busy, so

$$c_D \approx c_B$$

  - If $\rho \approx 0$, then $\mathrm{E}(B)$ is very small compared to $\mathrm{E}(A)$, so

$$c_D \approx c_A$$

  - If $c_A = c_B = 1$, then $c_D = 1$ (as it should)