

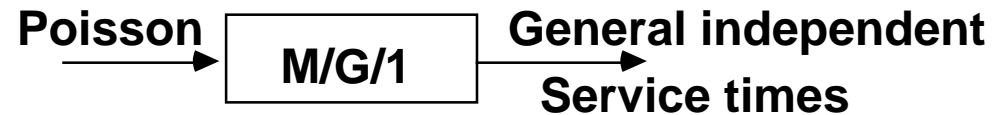
---

# M/G/1 Queues

Eytan Modiano  
MIT

# M/G/1 QUEUE

---



- **Poisson arrivals at rate  $\lambda$**
- **Service time has arbitrary distribution with given  $E[X]$  and  $E[X^2]$** 
  - **Service times are independent and identically distributed (IID)**
  - **Independent of arrival times**
  - **$E[\text{service time}] = 1/\mu$**
  - **Single Server queue**

# Pollaczek-Khinchin (P-K) Formula

---

$$W = \frac{\lambda E[X^2]}{2(1 - \rho)}$$

where  $\rho = \lambda\mu = \lambda E[X]$  = line utilization

From Little's formula,

$$N_Q = \lambda W$$

$$T = E[X] + W$$

$$N = \lambda T = N_Q + \rho$$

# M/G/1 EXAMPLES

---

- Example 1: M/M/1

$$E[X] = 1/\mu ; \quad E[X^2] = 2/\mu^2$$

$$W = \frac{\lambda}{\mu^2(1-\rho)} = \frac{\rho}{\mu(1-\rho)}$$

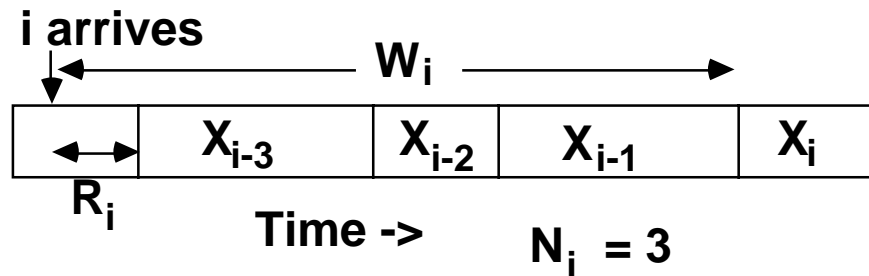
- Example 2: M/D/1 (Constant service time  $1/\mu$ )

$$E[X] = 1/\mu ; \quad E[X^2] = 1/\mu^2$$

$$W = \frac{\lambda}{2\mu^2(1-\rho)} = \frac{\rho}{2\mu(1-\rho)}$$

# Proof of Pollaczek-Khinchin

- Let  $W_i$  = waiting time in queue of  $i^{\text{th}}$  arrival  
 $R_i$  = Residual service time seen by  $i$  (i.e., amount of time for current customer receiving service to be done)  
 $N_i$  = Number of customers found in queue by  $i$



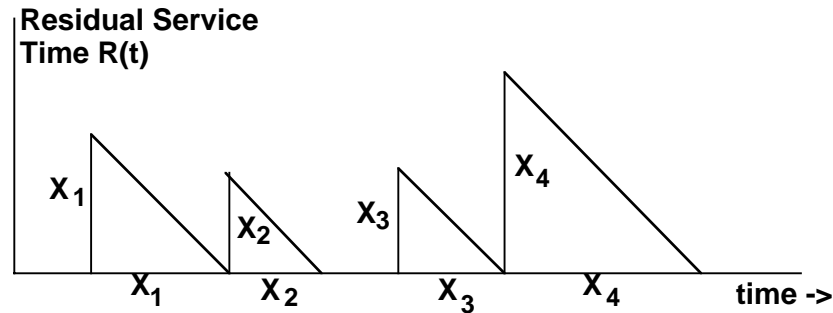
$$W_i = R_i + \sum_{j=i-N_i}^{i-1} X_j$$

- $E[W_i] = E[R_i] + E[X]E[N_i] = R + N_Q/\mu$ 
  - Here we have used PASTA property plus independent service time property
- $W = R + \lambda W/\mu \Rightarrow W = R/(1-\rho)$ 
  - Using little's formula

# What is R?

## (Time Average Residual Service Time)

---



Let  $M(t)$  = Number of customers served by time  $t$   
 $E[R(t)] = 1/t$  (sum of area in triangles)

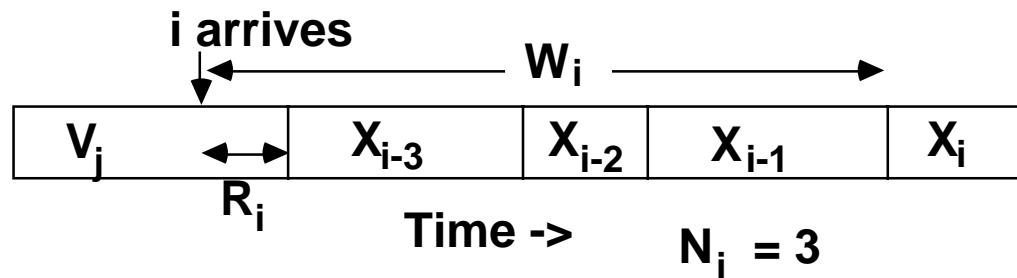
$$R_t = \frac{1}{t} \int_0^t R(\tau) d\tau = \frac{1}{t} \sum_{i=1}^{M(t)} X_i^2 = \frac{1}{2} \frac{M(t)}{t} \sum_{i=1}^{M(t)} \frac{X_i^2}{M(t)}$$

As  $t \rightarrow$  Infinity  $\frac{M(t)}{t}$  = average departure rate = average arrival rate

$$\frac{M(t)}{t} \sum_{i=1}^{M(t)} \frac{X_i^2}{M(t)} = E[X^2] \Rightarrow R = \lambda E[X^2]/2$$

# M/G/1 Queue with Vacations

- Useful for polling and reservation systems (e.g., token rings)
- When the queue is empty, the server takes a vacation
- Vacation times are IID and independent of service times and arrival times
  - If system is empty after a vacation, the server takes another vacation
  - The only impact on the analysis is that a packet arriving to an empty system must wait for the end of the vacation

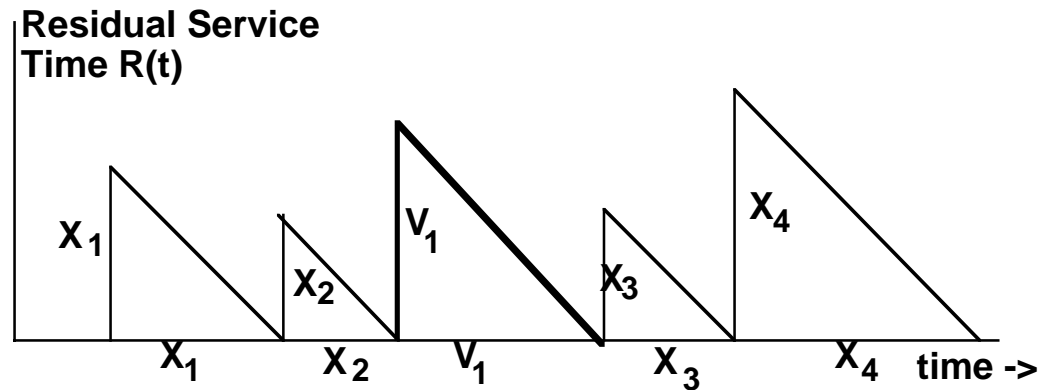


$$W_i = R_i + \sum_{j=i-N_i}^{i-1} X_j$$

$$E[W_i] = E[R_i] + E[X]E[N_i] = R + N_Q/\mu = R/(1-\rho)$$

# Average Residual Service Time (with vacations)

---



$$R = [R(t)] = \frac{1}{t} \int_0^t R(\tau) d\tau = \frac{1}{t} \left( \sum_{i=1}^{M(t)} \frac{X_i^2}{2} + \sum_{j=1}^{L(t)} \frac{V_j^2}{2} \right)$$

$$R = \lim_{t \rightarrow \infty} \frac{E[M(t)]}{t} \frac{E[X^2]}{2} + \frac{L(t)}{t} \frac{E[V^2]}{2}$$

- Where  $L(t)$  is the number of vacations taken up to time  $t$
- $M(t)$  is the number of customers served by time  $t$

# Average Residual Service Time (with vacations)

---

- As  $t \rightarrow \infty$ ,  $M(t)/t \rightarrow \lambda$  and  $L(t)/t \rightarrow \lambda_v = \text{vacation rate}$
- Now, let  $I = 1$  if system is on vacation and  $I = 0$  if system is busy
- By Little's Theorem we have,
  - $E[I] = E[\text{\#vacations}] = P(\text{system idle}) = 1 - \rho = \lambda_v E[V]$
  - $\Rightarrow \lambda_v = (1 - \rho)/E[V]$

- Hence,

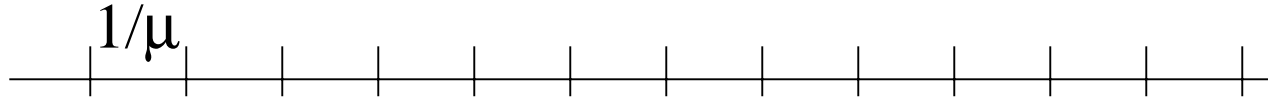
remember  $W = R/(1 - \rho)$

$$R = \lambda \frac{E[X^2]}{2} + \frac{(1 - \rho)E[V^2]}{2E[V]}$$

$$W = \lambda \frac{E[X^2]}{2(1 - \rho)} + \frac{E[V^2]}{2E[V]}$$

## Example: Slotted M/D/1 system

---



Each slot = one packet transmission time =  $1/\mu$

- Transmission can begin only at start of a slot
- If system is empty at the start of a slot, server not available for the duration of the slot (vacation)

- $E[X] = E[v] = 1/\mu$
- $E[X^2] = E[v^2] = 1/\mu^2$

$$W = \frac{\lambda / \mu^2}{2(1 - \lambda / \mu)} + \frac{1 / \mu^2}{2 / \mu} = \frac{\lambda / \mu}{2(\mu - \lambda)} + \frac{1 / \mu}{2}$$
$$= W_{M/D/1} + E[X] / 2$$

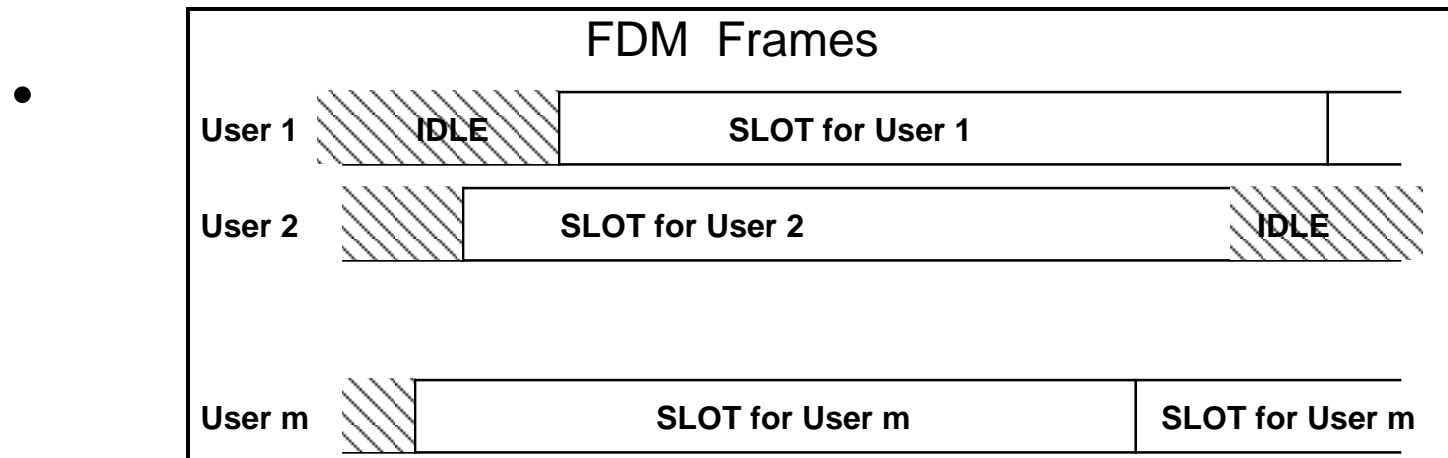
- Notice that an average of 1/2 slot is spent waiting for the start of a slot

# FDM EXAMPLE

- Assume  $m$  Poisson streams of fixed length packets of arrival rate  $\lambda/m$  each multiplexed by FDM on  $m$  subchannels. Total traffic =  $\lambda$

Suppose it takes  $m$  time units to transmit a packet, so  $\mu=1/m$ .

The total system load:  $\rho = \lambda$

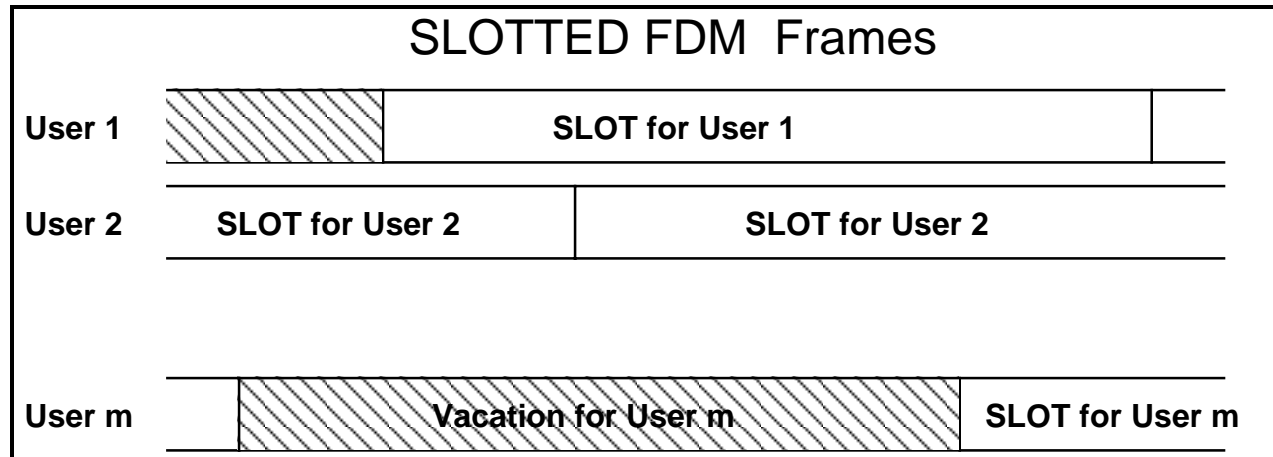


- We have an M/D/1 system  $\{ W=\lambda E[x^2]/2(1-\rho) \}$

$$W_{\text{FDM}} = \frac{(\lambda/m) m^2}{2(1-\rho)} = \frac{\rho m}{2(1-\rho)}$$

# Slotted FDM

- Suppose now that system is slotted and transmissions start only on  $m$  time unit boundaries.



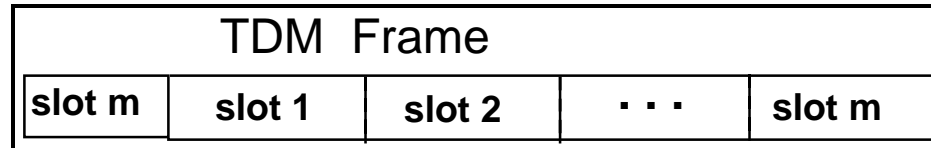
- This is **M/D/1 with vacations**
  - Server goes on vacation for  $m$  time units when there is nothing to transmit

$$E[V] = m; E[V^2] = m^2.$$

$$\begin{aligned} W_{\text{SFDM}} &= W_{\text{FDM}} + E[V^2]/2E[V] \\ &= W_{\text{FDM}} + m/2 \end{aligned}$$

# TDM EXAMPLE

---



- TDM with one packet slots is the same (a session has to wait for its own slot boundary), so

$$W = R/(1-\rho)$$

$$R = \lambda \frac{E[X^2]}{2} + \frac{(1-\rho)E[V^2]}{2E[V]}$$

$$W = \lambda \frac{E[X^2]}{2(1-\rho)} + \frac{E[V^2]}{2E[V]}$$

# TDM EXAMPLE

---

- Therefore,  $W_{TDM} = W_{FDM} + m/2$

Adding the packet transmission time, TDM comes out best because transmission time = 1 instead of m.

$$T_{FDM} = [W_{FDM}] + m$$

$$T_{SFDM} = [W_{FDM} + m/2] + m$$

$$T_{TDM} = [W_{FDM} + m/2] + 1$$

$$= T_{FDM} - [m/2 - 1]$$