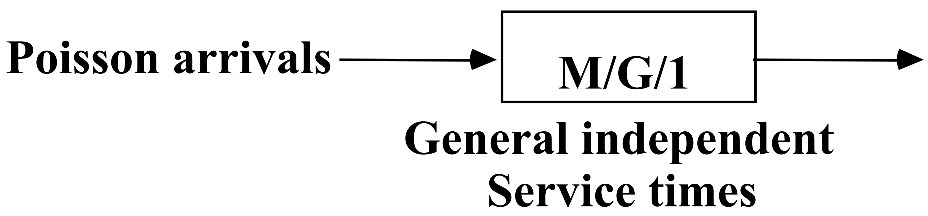
M/G/1 Queues

Eytan Modiano
Massachusetts Institute of Technology

M/G/1 QUEUE



- **Poisson arrivals at rate λ**
- **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[service time] = 1/μ**
  - **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 Theorem,

$$N_Q = \lambda W$$

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

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

M/G/1 EXAMPLES

- **Example 1: M/M/1**

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

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

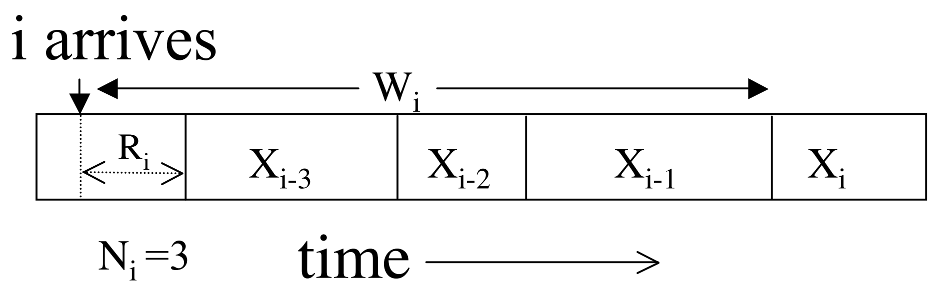
- **Example 2: M/D/1 (Constant service time $1/\mu$)**

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

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

Proof of Pollaczek-Khinchin (P-K) formula

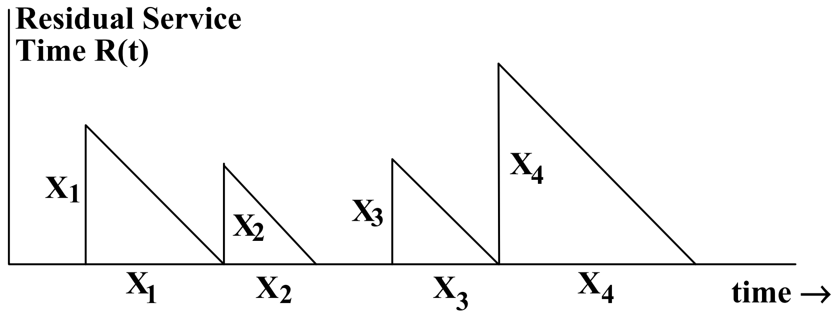
- **Let** **W_i = waiting time in queue of i^{th} arrival**
  R_i = Residual service time seen by i
  (i.e., amount of time for current customer receiving service to be done)
  N_i = Number of customers found in queue by i



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

- **$E[W_i] = E[R_i] + E[X]E[N_i] = R + N_Q/\mu$**
  - **Here we have used the PASTA property plus the 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)

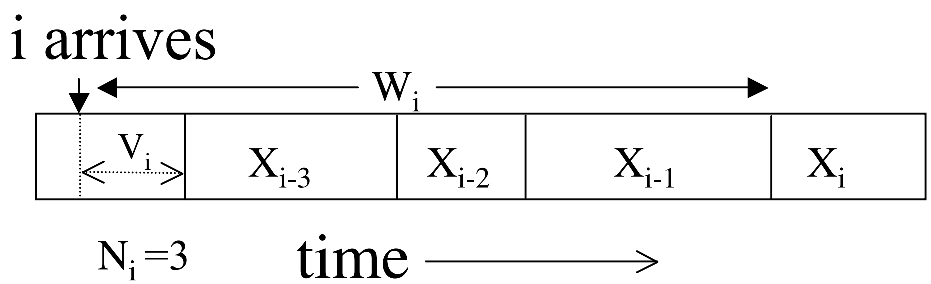
$$R_t = \frac{1}{t}\int_0^t R(\tau)d\tau = \frac{1}{t}\sum_{i=1}^{M(t)}\frac{X_i^2}{2} = \frac{M(t)}{2t}\sum_{i=1}^{M(t)}\frac{X_i^2}{M(t)}$$

As $t \to \infty$, $\frac{M(t)}{t}$ = **Average departure rate = Average arrival rate =** λ

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

M/G/1 Queue with Vacations

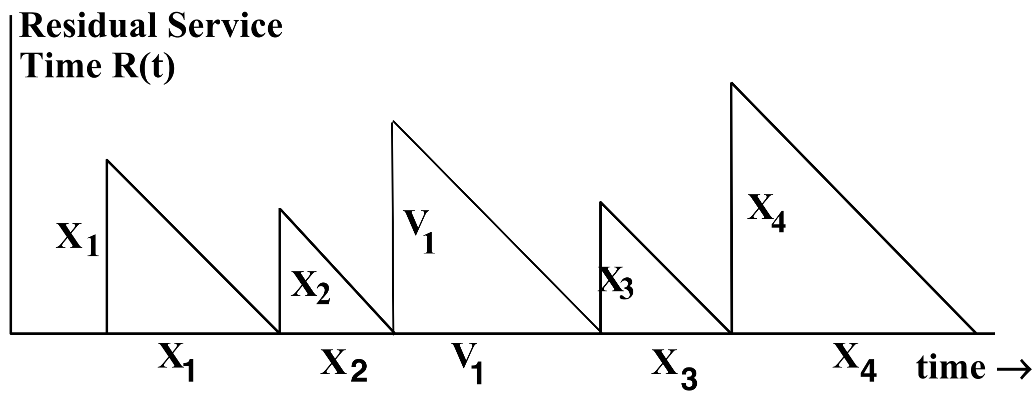
- Useful for polling and reservation systems (e.g., token rings)
- When the queue is empty, the server takes a vacation
- Vacation times are IID and independent of service times and arrival times
  - If system is empty after a vacation, the server takes another vacation
  - The only impact on the analysis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



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

$$E[W_i] = E[R_i] + E[X]E[N_i] = R + N_Q/m = R/(1-r)$$

Average Residual Service Time (with vacations)



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

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

- 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)

$$\text{As } t \to \infty, \quad \frac{M(t)}{t} \to \lambda \quad \text{and} \quad \frac{L(t)}{t} \to \lambda_v = \text{ vacation rate}$$

- Let I = 1 if system is on vacation and I = 0 if system is busy
- Little's Theorem ⇒

$E[I] = P(\text{system idle}) = 1-\rho = \lambda_v E[V]$

$\Rightarrow \lambda_v = (1-\rho)/E[V]$

- Hence,

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

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

Example: Slotted M/D/1 system

Each slot = one packet transmission time = 1/μ

- **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[X²] = E[v²] =** $1/\mu^2$

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

$$= W_{M/D/1} + E[X]/2$$

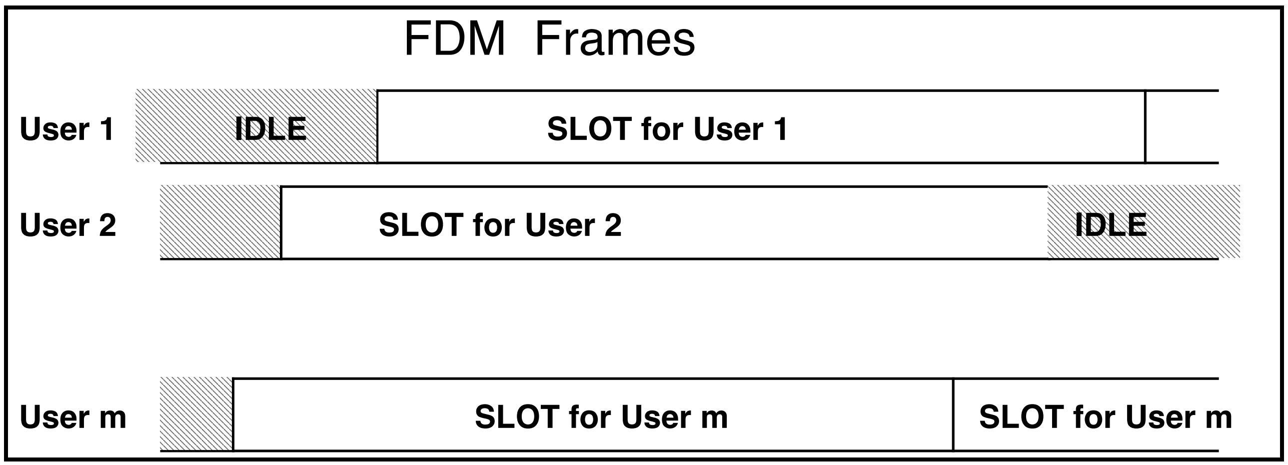
- **Notice that an average of 1/2 slot is spent waiting for the start of a slot**

FDM EXAMPLE

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

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

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

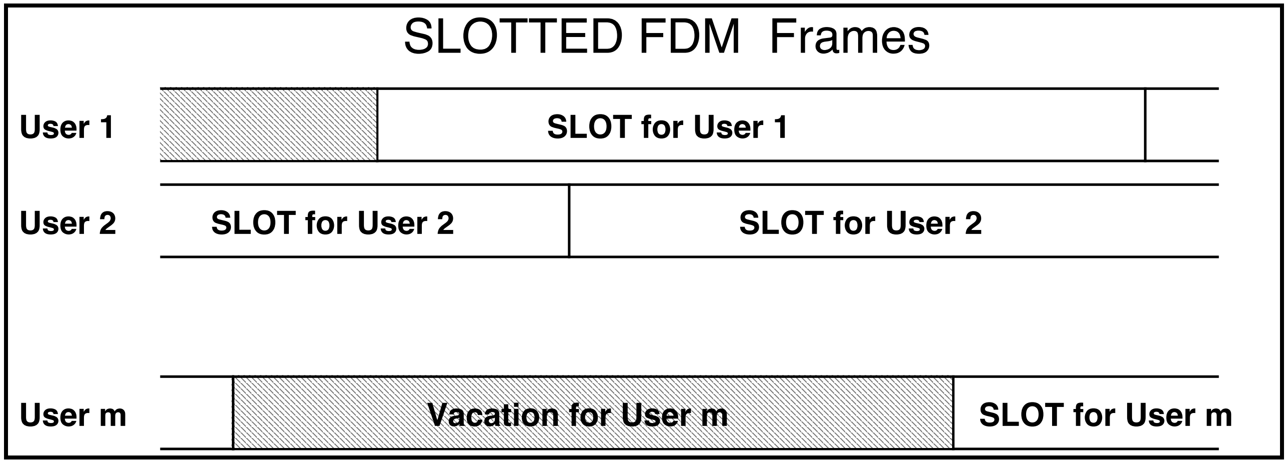


- We have an M/D/1 system $W=\lambda E[x^2]/2(1-\rho)$

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

Slotted FDM

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



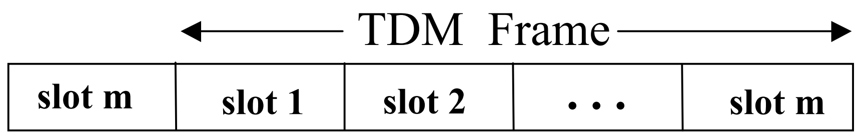
- **This is M/D/1 with <u>vacations</u>**
  – **Server goes on vacation for m time units when there is nothing to transmit**

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

$$W_{SFDM} = W_{FDM} + E[V^2]/2E[V]$$

$$= W_{FDM} + m/2$$

TDM EXAMPLE



- **TDM with one packet slots is the same**
  - **A session has to wait for its own slot boundary**

W = R/(1-ρ)

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

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

TDM EXAMPLE

- **Therefore,** $W_{TDM} = W_{FDM} + m/2$
- **Adding the packet transmission time, TDM comes out best because transmission time is 1 instead of m**

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

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

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

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