

EP2200 Queueing theory and teletraffic systems

Lecture 10 M/G/1 systems with vacation and priority

Viktor Fodor
KTH EES/LCN

Announcements

- Invited lecture: teletraffic theory at Ericsson, Dec. 1.
Queuing networks in the first part of the lecture
- Previous exam problems: under Examination on the web
- Do not forget to register for the exam

M/G/1 queues

- Recall:
 - Arrival process: Poisson
 - Service process: i.i.d service times
 - first and second moment to determine average performance measures
 - distribution for distribution of performance measures
- Pollaczek-Khinchin mean formulas
 - based on mean value analysis

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

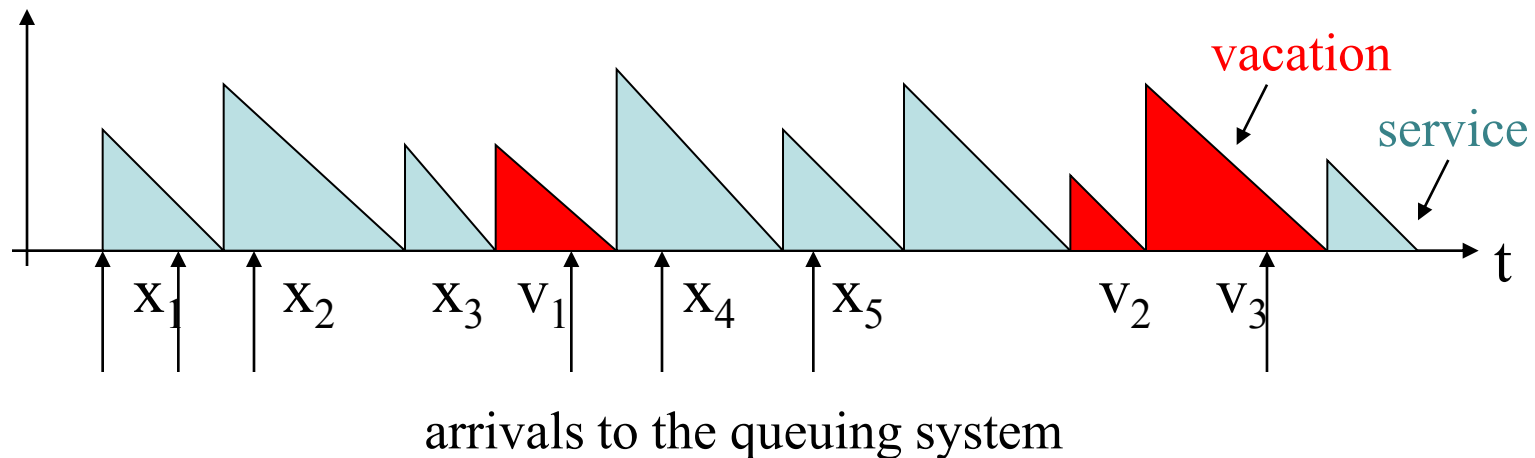
- Pollaczek-Khinchin transform equations
 - based on embedded MC

$$Q(z) = B^*(\lambda - \lambda z) \frac{(1-\rho)(1-z)}{B^*(\lambda - \lambda z) - z} \quad T^*(s) = B^*(s) \frac{s(1-\rho)}{s - \lambda + \lambda B^*(s)}$$

M/G/1 with vacation

- Vacation: the server is not available for a while after the system gets idle (empty)
 - there is no idle period, only vacation period
 - vacation periods: identically distributed, independent random variable, V
- Average performance measures
 - stability condition: $\lambda E[X] < 1$ – higher load $\rightarrow$ less vacation

$R_s(t)$, $R_v(t)$ {remaining service time, remaining vacation time}



M/G/1 with vacation – waiting time

$$E[W_k] = E[R_{s,k}] + (k-1)E[X], \quad k \geq 1 \quad (\text{waiting time for customer arriving at state } k \geq 1)$$

$$E[W_0] = E[R_{v,0}], \quad (k = 0)$$

$$W = E[W] = \sum_{k=0}^{\infty} p_k E[W_k] = p_0 E[R_{v,0}] + \sum_{k=1}^{\infty} p_k E[R_{s,k}] + \sum_{k=1}^{\infty} p_k (k-1)E[X]$$

$$\text{Def : } R_{s,0} = 0, \quad R_{v,k} = 0, \quad k \geq 1$$

$$R_s = \sum_{k=0}^{\infty} p_k E[R_{s,k}], \quad R_v = \sum_{k=0}^{\infty} p_k E[R_{v,k}]$$

$$W = N_q E[X] + R_s + R_v$$

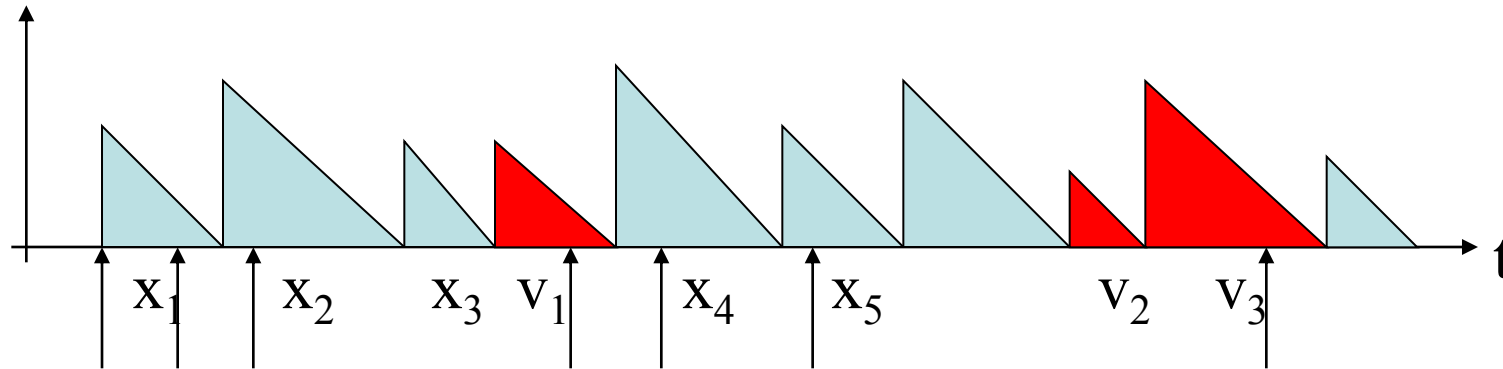
$$W = \lambda W E[X] + R_s + R_v$$

$$W(1-\rho) = R_s + R_v$$

$$W = \frac{R_s}{1-\rho} + \frac{R_v}{1-\rho} = \frac{\lambda E[X^2]}{2(1-\rho)} + \frac{R_v}{1-\rho}, \quad \left[\text{Recall M/G/1: } R_s = \frac{\lambda E[X^2]}{2} \right]$$

R_s : average remaining service time, averaged over time including vacation periods
 R_v : average remaining vacation time, averaged over time including service periods

$R_s(t), R_v(t)$ M/G/1 with vacation – waiting time



$$R_v = \frac{\sum_{i=1}^n \frac{1}{2} v_i^2}{T}, \quad \text{where} \quad T(1-\rho) = \sum_{i=1}^n v_i \Rightarrow \frac{1}{T} = \frac{(1-\rho)}{\sum_{i=1}^n v_i}$$

$$R_v = \frac{(1-\rho) \sum_{i=1}^n \frac{1}{2} v_i^2}{\sum_{i=1}^n v_i} = \frac{(1-\rho)}{2} \frac{\frac{1}{n} \sum_{i=1}^n v_i^2}{\frac{1}{n} \sum_{i=1}^n v_i} = \frac{(1-\rho)}{2} \frac{E[V^2]}{E[V]}$$

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

Calculate average remaining vacation time:

- consider the system for time T , within that vacation for $(1-\rho)T$
- n vacation periods, $n \rightarrow \infty$

The M/G/1 mean value analysis

Pollaczek-Khinchin mean formulas

Group work:

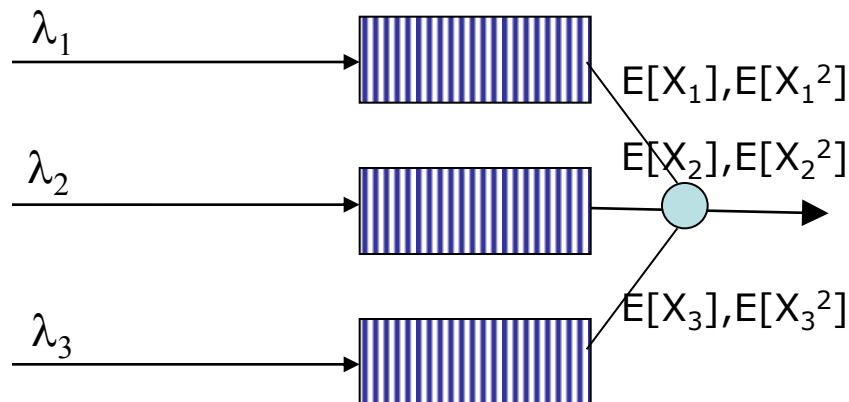
- Consider the following system:
 - Single server, infinite buffer
 - Poisson arrival process, 0.1 customer per minute
 - Service process: sum of two exponential steps, with mean times 1 minute and 2 minutes
 - Maintenance period starts whenever the system becomes idle, the maintenance takes exactly 1 minute.
 - Calculate the mean waiting time

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

(Similar M/G/1 without vacation
had $W=1\text{min}$)

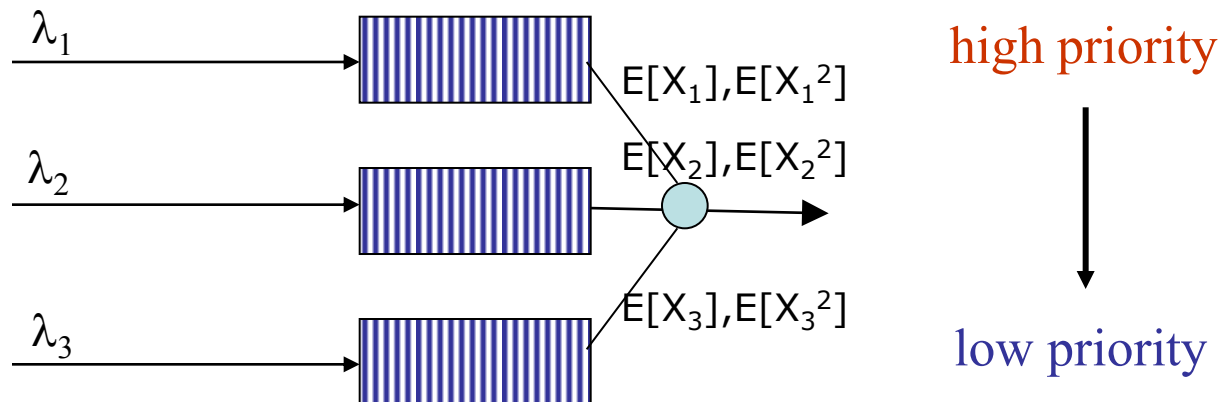
M/G/1 with priority

- M/G/1 queue
- K priority class:
 - Separate, infinite queue for each class, one server (multiclass system)
 - Poisson arrival in each class ($\lambda_i, \sum \lambda_i = \lambda$)
 - General service time distribution in each class ($E[X_i], E[X_i^2]$)
 - $E[X] = \sum \lambda_i / \lambda E[X_i]$
 - $E[X^2] = \sum \lambda_i / \lambda E[X_i^2]$,
 - Class 1 the highest priority



M/G/1 with priority

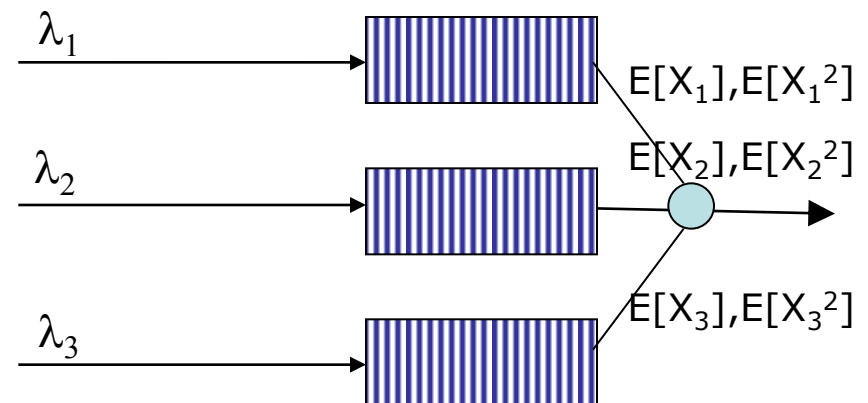
- Priority systems
 - Low priority customer selected only if high priority queues are empty
 - **Non-preemptive**: the service is completed even if higher priority customer arrives
 - **Preemptive**: the service is interrupted if higher priority customer arrives
 - **Resume**: the service continues from the point of interruption
 - **Non-resume**: the service starts from the beginning (not considered in this course)



M/G/1 with non-preemptive priority

- Derive mean performance parameters
- Waiting time for a customer of priority i =
Residual service time, R_s +
Service time of customers already waiting in queue i +
Service time of customers already waiting in queues $j < i$ (higher priority) +
Service time of customers arriving to queues $j < i$ while "our" customer is waiting

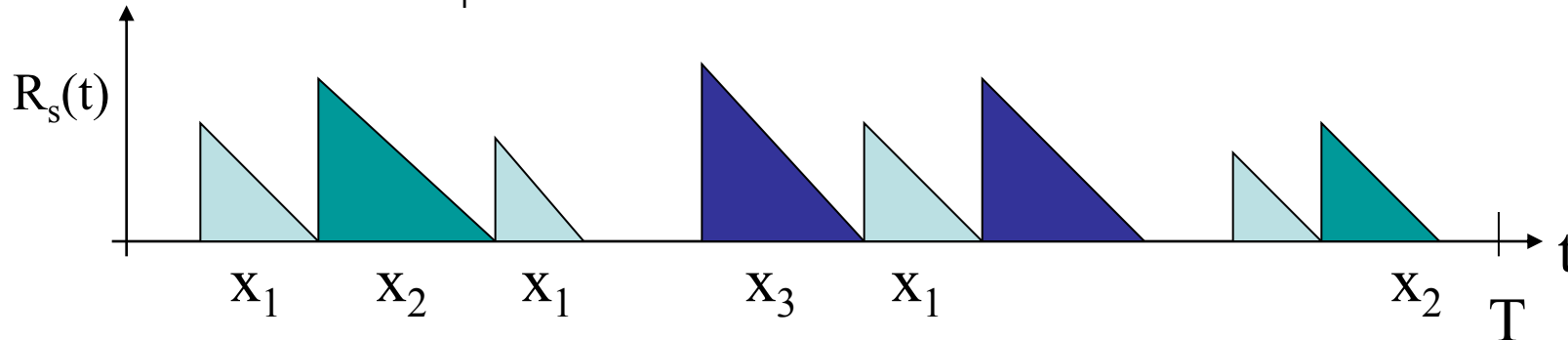
$$W_i = R_s + E[X_i]N_{q,i} + \sum_{j=1}^{i-1} E[X_j]N_{q,j} + \sum_{j=1}^{i-1} E[X_j]\lambda_j W_i$$



The M/G/1 mean value analysis

Pollaczek-Khinchin mean formulas

- We have to derive the average remaining service time:
 - n : number of services in a large T = number of Poisson arrivals: $n_i = \lambda_i T$
 - $T \rightarrow \infty$ and $n_i \rightarrow \infty$



$$R_s = E[R_s(t)] = \frac{1}{T} \sum_{i=1}^K \sum_{j=1}^{n_i} \frac{1}{2} X_{i,j}^2 = \frac{1}{2} \sum_{i=1}^K \frac{\lambda_i}{n_i} \sum_{j=1}^{n_i} X_{i,j}^2 = \boxed{\frac{1}{2} \sum_{i=1}^K \lambda_i E[X_i^2]}$$

$$W_i = R_s + E[X_i]N_{q,i} + \sum_{j=1}^{i-1} E[X_j]N_{q,j} + \sum_{j=1}^{i-1} E[X_j]\lambda_j W_i$$

- Express W_1 and W_2 !

M/G/1 with non-preemptive priority

- W_i, T_i general form:

$$W_i = \frac{R_s}{(1 - \sum_{j=1}^{i-1} \rho_j)(1 - \sum_{j=i}^K \rho_j)}, \quad R_s = \frac{1}{2} \sum_{i=1}^K \lambda_i E[X_i^2]$$

$$T_i = W_i + E[X_i]$$

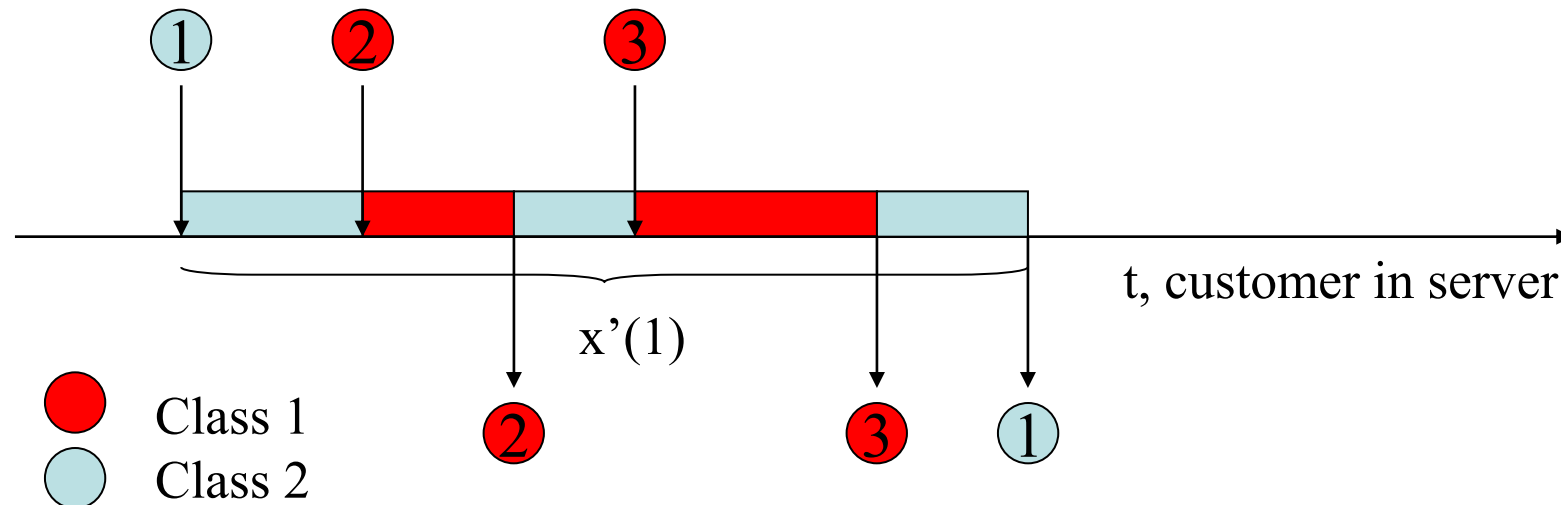
- Average waiting time:

$$W = \sum p_i W_i = \sum \frac{\lambda_i}{\lambda} W_i$$

- Comments:
 - W_i depends on W_j even if $i < j$ (through the residual service time)
 - Mean waiting time W can be decreased if shorter service gets priority
 - in multiclass systems average perf. measures are dependent on the service policy

M/G/1 with preemptive resume priority

- Service is interrupted if higher priority customer arrives
 - later the service continues from the point of interruption
- Derive mean performance parameters
- Now: lower class customer is invisible for higher class customers!
- Definition of service time of low priority customers (x'):
 - From the time of first entering the server until competition.



M/G/1 with preemptive resume priority

- Waiting time for a customer of priority i
= Time from arrival to the first service attempt =
residual service time, $R_{s,i}$ (now priority dependent) +
service time of customers already waiting in queue i +
service time of customers already waiting in queues $j < i$ (higher priority) +
service time of customers arriving to queues $j < i$ while “our” customer is waiting

$$R_{s,i} = \frac{1}{2} \sum_{j=1}^i \lambda_j E[X_j^2]$$

$$W_i = \frac{R_{s,i}}{(1 - \sum_{j=1}^{i-1} \rho_j)(1 - \sum_{j=1}^i \rho_j)}$$

Recall: non-preemptive case:

$$R_s = \frac{1}{2} \sum_{i=1}^K \lambda_i E[X_i^2]$$

- E.g., highest priority (class 1) customer?
- Comment: now the high priority customers do not “see” the lower priority ones!

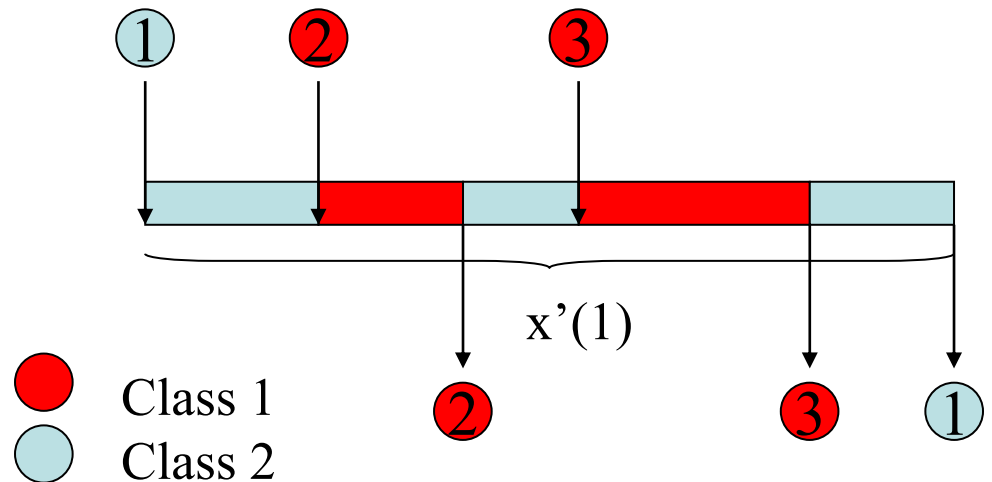
M/G/1 with preemptive resume priority

- Mean system time (T_i)?
 - Waiting time + service time including interruptions by arriving high priority customers

$$E[X'_i] = E[X_i] + \sum_{j=1}^{i-1} E[X_j] \lambda_j E[X'_i]$$

$$E[X'_i] = \frac{E[X_i]}{1 - \sum_{j=1}^{i-1} \rho_j}$$

$$T_i = \frac{R_{s,i}}{(1 - \sum_{j=1}^{i-1} \rho_j)(1 - \sum_{j=1}^i \rho_j)} + \frac{E[X_i]}{1 - \sum_{j=1}^{i-1} \rho_j}$$



- E.g., average service time for class 1 and for class 2 customer?

M/G/1 with vacation and with priorities

- Requirements for the exam:
- Derive and use P-K mean value formulas
- Understand the concept of remaining vacation time
- Understand the concept of remaining service time in priority systems
- Calculate expected remaining vacation and service times with different conditions (for all customers, for customers finding the system empty/busy)
- Understand the concept of service time in the preemptive priority system, calculate it for specific cases
- Exam: formula sheet is available.