ECE G311: Communications Networks
"Theory and Analysis"

Fall Quarter 2003

Instructor: Prof. A. Bruce McDonald

Lecture Topic

# Introductory Analysis of M/G/1 Queueing Systems

Module Number One

## Steady-State System Behavior—The Recursion Approach

# Fundamental Queueing Theory

## Equilibrium Behavior of the M-G-1 Queueing System

### Using the Recursion Approach to Solve the P-K Formula

## Poisson Arrivals with General Service:

In this module students are introduced to a queueing system in which the arrivals are assumed to be Poisson, but, only the first and second of moments of the *general* service time distribution is known—the entire distribution need not be know in order to solve for the steady-state behavior of the system.

In the following analysis it is *assumed* that customers are served in a FIFO manner and that $X_i$ is the service time of the $i_{th}$ customer. The constraint is that the set of random variables (RVs) $(X_1, X_2, \ldots)$ are IID (mutually independent and identically distributed) *and* they are indenpendent of the interarrival times.

**Some Notation**

- $\lambda =$ Mean Customer Arrival Rate (Poission Process)
- $\mu =$ Mean Customer Service Rate (General Distribution)
- $\rho = \frac{\lambda}{\mu} =$ Average System Utilization
- $\overline{X} = E[X] = \frac{1}{\mu} =$ Average Service Time
- $\sigma =$ Variance of the Customer Service Times
- $\overline{X^2} = E[X^2] =$ Second Moment of the Service Time
- $\pi_i =$ Equilibrium Probability there will be $i$ Customers in the System

# The M/G/1 Queueing System

- MG1 is a Single Queueing System
- Customer Arrivals Characterized by a Poisson Process (independent of service times);
- Service Times are Characterized by an Arbitrary Distribution (independent of arrival process);
- Results are given by the (Pollaczek-Khinchin) P-K formulas:
- First: Find P-K formula for the *mean number in the system*;
- Using a transform equation the P-K formula is utilized to derive the equilibrium state probabilities;

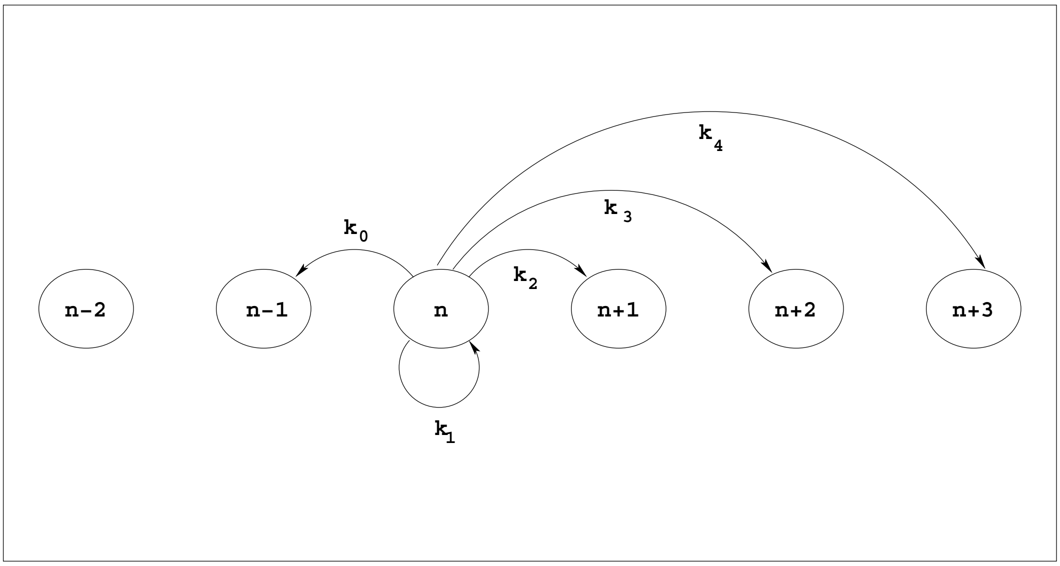


Figure 1: M/G/1 queueing system showing the imbedded Markov Chain.

**Recursion Approach:**

To find the *mean number* in the system start with a *difference equation* for the number in the system following the $i$th departure. The expected value of the square of the resulting expression is the **P-K mean value formula**—The *key assumption*, is that the state probabilities at departure instants are equal to the equilibriumm probabilities [1] This characteristic is the basis for the imbedded Markov Chain, as shown for departure instants, of the M/G/1 system in Figure-1.

[1] This observation is proven later.

## Markovian Queues in Equilibrium

### The Equilibrium Equations:

- Let $\pi$ represent a row vector of *equibrium* state probabilities, e.g.: $\pi = [\pi_0, \pi_1, \ldots, \pi_i, \ldots]$ where $\pi_i$ is the steady-state probability of the system being in state $i$; stated another way it is proportion of time the system spends in state $i$.
- Let $Q$ represent the infinitesimal generator matrix; the elements of the matrix are the infinitesimal *transition rates:* $q_{i,j}$ from state $i$ to state $j$ for the Markov process.
- Recalling that $\nu_i$ is the rate at which a CTMC departs state $i$ we have: $q_{i,j} = \nu_i p_{i,j}; p_{i,j}$ is the one step transition probability from state $i$ to state $j$. The transition probability matrix $P$ contains all the one-step transition probabilities.

**The task is to solve the *equations of motion* in equilibrium.**

## Equations of Motion

---

The equilibrium condition for an ergodic CTMC is expressed as follows (students: please be sure you understand this expression):

$$\pi Q = 0$$

Combined with the *normalization condition (a.k.a. the conservation of flow relation)*:

$$\sum_k \pi_k = 1$$

For the equilibrium case the method for solving these equtions is based on the observation that the probabilistic flow rate into a state must equal (or balance) the probabilistic flow rate out of that state; hence the equations of motion become the *flow balance equations*.

## Limiting Probabilities

Using graphical inspection the flow balance equations of the CTMC is easily deterimined and solved. Two important limiting probabilities are in the case of ergodic systems are

1. $p_k = \lim_{t\to\infty} Pr\{N(t) = k\}$
2. $r_k = Pr\{Arriving\ customer\ finds\ the\ system\ in\ state\ k\}$

One might be tempted to assume that $p_k = r_k$ in general; this is not true in all cases (students should think of a scenario where it is easy to show that this does not hold—e.g. D/D/1 system).

### PASTA

Our approach to solving the M/G/1 system is based largely on what we next show: that for any system with Poisson Arrivals $p_k = r_k$. In fact it is proven that for any stable queue with Poisson arrivals:

$$P_k(t) = R_k(t)$$

This expression characterizes the special property that *Poisson Arrivals See Time Averages: PASTA*.

## Proof of the PASTA Property

Assume that we have a system in which arrivals come according to a Poisson Process. $R_k(t)$ is the probability that a customer arriving at time $t$ finds the system in state $k$. Define $A(t, t+\Delta t)$ as the event that an arrival occurs during the interval $(t, t+\Delta t)$:

$$\begin{aligned} R_k(t) &= \lim_{\Delta t \to 0} Pr\{N(t) = k | A(t, t+\Delta t)\} \\ &= \lim_{\Delta t \to 0} \frac{Pr\{N(t) = k, A(t, t+\Delta t)\}}{Pr\{A(t, t+\Delta t)\}} \\ &= \lim_{\Delta t \to 0} \frac{Pr\{A(t, t+\Delta t) | N(t) = k\} Pr\{N(t) = k\}}{Pr\{A(t, t+\Delta t)\}} \end{aligned} \tag{1}$$

Due to the memoryless property we know that for a Poisson Process the event $A(t, t+\Delta t)$ must be independent of the number in the system at time $t$. Furthermore, based on stationary increments it must also be independent of $t$ as well. Thus, $Pr\{A(t, t+\Delta t) | N(t) = k\} = Pr\{A(t, t+\Delta t)\}$:

$$\begin{aligned} R_k(t) &= \lim_{\Delta t \to 0} \frac{Pr\{A(t, t+\Delta t)\} Pr\{N(t) = k\}}{Pr\{A(t, t+\Delta t)\}} \\ &= \lim_{\Delta t \to 0} Pr\{N(t) = k\} \\ &= P_k(t) \end{aligned}$$

# The M/G/1 Recursion for Departure Instants

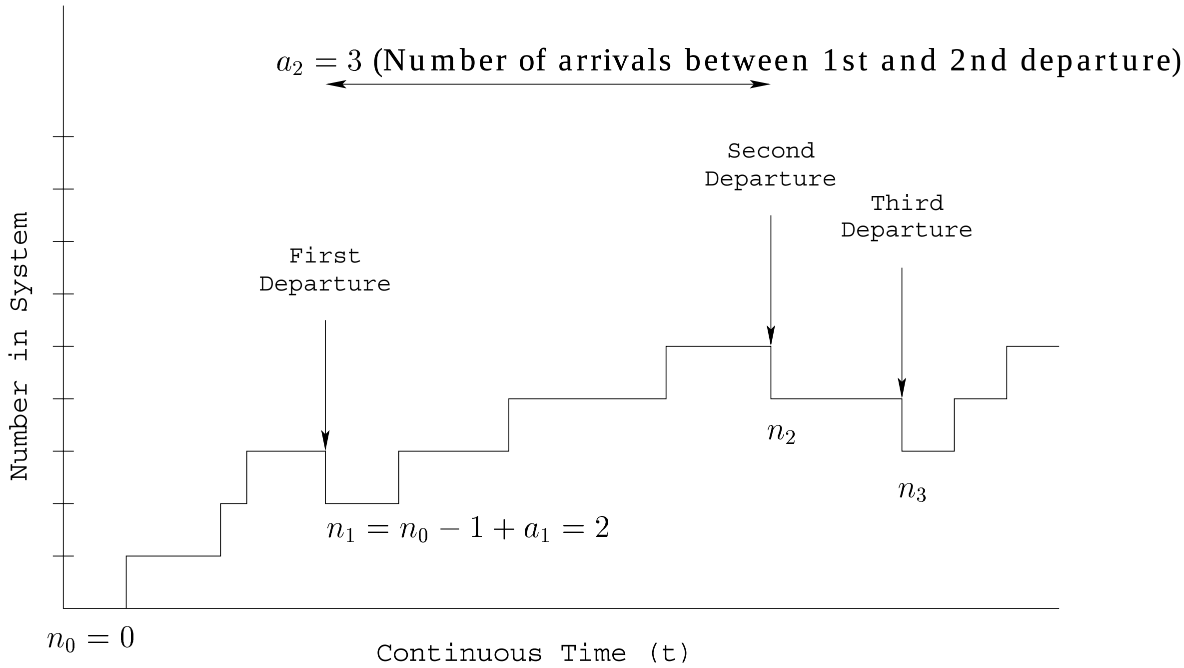


Figure 2: Sample path of M/G/1 system state at departure instants.

- Consider the system *"state"* at the departure instants;
- Define $n_{i+1}$ as the number in the system immediately following the $(i+1)$ st departure;
  1. $n_{i+1}$ enumerates the customers in the queue *and* in service;
  2. $n_{i+1}$ is also equal to the number of customers in the system after the ith departure minus 1 (since there **must** have been one-and-only-one departure, plus the number of customer arrivals between the ith and the $(i+1)$ st departure;
- Define $a_{i+1}$ as the number of customers that arrive to the system between the $i$ th and the $(i+1)$ st departure instants.

$$n_{i+1} = n_i - 1 + a_{i+1} \,, \qquad n_i > 0 \tag{2}$$

## Using the Embedded Markov Chain and PASTA

### Random Views at Departure Instants

Using the departure instants to evaluate system state it should be clear that the original system that requires a two-dimensional state space: $[N(t), X_0(t)]$ can be transformed into a system with a one-dimensional state space: $[N(t)]$. In effect, the departure instants represent a Markov-Chain since immediately upon service completion the amount of residual work must be zero. Hence, we can evaluate the system state as seen by arrivals that occur at these instants.

The idea is that the arrivals occur at *random instants* according to a memoryless process. The intuitive conclusion is that the stationary distribution of states seen by arrivals at specific times are representative of the distribution at any time. That is, we consider the observations to be equivalent to a random sampling of the system. Hence, arrivals at random departure instants are statistically the same as arrivals at any time.

## Special Case: The Empty Queueing System

**Empty Sytem:** The number in the system immediately following the departure of the $i+1$ $st$ customer [2] for the cases in which $n_i = 0$ will be equal to the number of arrivals during the *service time* of the arrival of the first customer to the empty system.

$$n_{i+1} = a_{i+1} \;, \qquad n_i = 0$$

The **unit step function** provides a nice way to express the separate recursions as a single equation as follows:

$$n_{i+1} = n_i - u(n_i) + a_{i+1}$$

Wherein the step function is defined as:

$$u(n_i) = \left\{ \begin{array}{ll} 1, & n_i > 0 \\ 0, & n_i = 0 \end{array} \right.$$

[2]Commonly refered to as the $i+1$ $st$ *departure instant*.

## The Second Moment of the State Equation

**Methodology:** Based on the recursion approach taken here the P-K formula can be found by, first, squaring both sides of the recursion equation, and, then, taking the expected values—in this case *second moments* because of the square.

$$n_{i+1}^2 = \{ \begin{array}{l} (n_i - u(n_i) + a_{i+1})^2 \\ n_i^2 + u(n_i)^2 + a_{i+1}^2 - 2n_i u(n_i) + 2n_i a_{i+1} - 2u(n_i)a_{i+1} \end{array}$$

Now take the expected value of both sides:

$$E[n_{i+1}^2] = E[n_i^2] + E[u(n_i)^2] + E[a_{i+1}^2] \quad (3)$$

$$-2E[n_i u(n_i)] + 2E[n_i a_{i+1}] - 2E[u(n_i)a_{i+1}] \quad (4)$$

**The P-K Formula:** The desired result is a useable equation for the steady-state number of customers in the queue, which, is simply $E[n_i]$. Hence, to find the equilibrium behavior of an M-G-1 queueing system the strategy is to *restate* the equation given above *term-by-term* and then to solve the simplified equation for $E[n_i]$.

**Objective:**

$$P - K\ formula \rightarrow E[n_i]$$

## Term-by-Term Simplification for P-K Evaluation Equivalent Moments

**Steady-State Equivalents:** The objective of this analysis is the characterize the steady-state behavior of an ergodic system. Under *equilibrium* conditions the $k$ th moment of any state variable, $S$, will be equal $\forall i, j \in$ steady-state: $E[S_i^k] = E[S_j^k]$, hence, the following two terms should be equal and cancel in the equation:

$$E[(n_{i+1})^2] = E[n_i^2] \tag{5}$$

## Term-by-Term Simplification for P-K Evaluation

## Properties of the Unit Step Function

**Product Equivalence:** Since the Unit Step is either equal to 0 or 1 it is clear that for any state variable $S$, state $i$, and powers $m, n$: $u(S_i)^m = u(S_i)^n$; therefore $u(n_i) = u(n_i)^2$, hence, the expected values will also be equal:

$$E[u(n_i)] = E[u(n_i)^2]$$

**Finding the expectation for $E[u(n_i)]$:** What is required is a more useful expression for $E[u(n_i)]$. Here, the theory of total probability and a very general queueing result for busy systems provide a straight forward solution:

$$\begin{aligned} E[u(n_i)] &= \sum_{n=0}^{\infty} u(n_i) Pr[n_i = n] \\ E[u(n_i)] &= \sum_{n=0}^{\infty} Pr[n_i = n] \\ E[u(n_i)] &= Pr[System\ is\ Busy] \end{aligned}$$

## Term-by-Term Simplification for P-K Evaluation
## Using System Busy Time to Eliminate the Unit Step Function

**Finding the System Busy Time:** In an ergodic system under steady-state it is easy to determine the probability of the system being busy by considering an arbitrary interval of time $\tau$ and using probability to characterize the fraction of the interval that there will be *one or more* customers in the system, i.e. the system will be busy. If $\tau_b$ is the busy time then:

$$\tau_b = \tau - \tau\pi_0$$

**Customer Service:** For *any* probability distribution for service times the mean service rate $\mu$ is known—it is the inverse of the first moment, or expected value of the service time for the given M-G-1 system: $\overline{X}$. The number of customers served in the busy interval $\tau_b$ is:

$$(\tau - \tau\pi_0)\mu$$

Global balance can be applied, which, reflects that on average the number of customers served in an arbitrary interval is equal to the number of customers that arrive in the same interval:

$$\begin{aligned} \lambda\tau &= (\tau - \tau\pi_0)\mu \\ \rho &= 1 - \pi_0 \end{aligned}$$

$$E[(u(n_i)^2] = E[u(n_i)] = Pr[System\ is\ Busy] = 1 - \pi_0 = \rho \tag{6}$$

# Term-by-Term Simplification for P-K Evaluation

## Additional Sundry Results

---

- Since $n_i u(n_i) = n_i$—the following simplification:

$$2E[n_i u(n_i)] = 2E[n_i] \tag{7}$$

- The number of customers that arrive to the system between the $i$ th and the $i+1$ st departure instants ($a_{i+1}$) is independent of the number in the system after the $i$ th departure system ($n_i$). Hence, by idependence the following simplification becomes possible:

$$E[u(n_i)a_{i+1}] = E[u(n_i)]E[a_{i+1}]$$

By Equation-6 it is shown that $E[u(n_i)] = \rho$, hence, it is possible to further simplify this equation by determining the value of $E[a_{i+1}]$ by returning to the original recurance relation (Equation-2 and taking the expectation of both sides:

$$E[n_{i+1}] = E[n_i] + E[u(n_i)] + E[a_{i+1}]$$

After rearranging terms and making all necessary substitutions the following result is obtained:

$$2E[u(n_i)a_{i+1}] = 2E[u(n_i)]E[a_{i+1}] = 2\rho_2 \tag{8}$$

- Finally, the same argument leads to similar simplification of the final term:

$$2E[n_i a_{i+1}] = 2E[n_i] = 2E[n_i]\rho \tag{9}$$

# Equilibrium Solution for the M-G-1 Queue
# The P-K Formula for the Mean Number in the Queue

The results derive for each term can be substituted into Equation-3, which can then be solved for the desired quantity: $E[n_i]$, which, is the expected value of the number in the system at the $i$ th departure instant:

$$E[n_i] = \frac{\rho + E[a_{i+1}^2] - 2\rho^2}{2(1-\rho)}$$

**Departure Instants = *Any* Instant:** Using the initial assumption that expected values at the departure instants are equivalent the the expected values at equilibrium, and, hence, *any* instant the following relation holds:

$$E[n_i] = E[n]$$

**Stationarity of the Poisson Process:** The final result follows by observing that a time homogeneous Poisson Process has both **independent increments** and *stationary increments*. Hence, the following relation holds:

$$E[a_{i+1}^2] = E[a^2]$$

**The P-K Formula:**

$$E[n] = \frac{\rho + E[a^2] - 2\rho^2}{2(1-\rho)} \tag{10}$$

## Equilibrium Solution for the M-G-1 Queue

## Some Useful Results from Probability and Statistics

---

The problem with Equation-10 is that knowledge of the second moment of the service distribution may not be readily available. Hence, a better expression would require knowledge of only three (3) system parameters:

1. The mean arrival rate
2. The mean service rate (or time)
3. The variance of the service time distribution

---

The required results are as follows [3]:

- $VAR[X] = E[X^2] - E[X]^2$
- $VAR[Y] = E[VAR[Y|X]] + VAR[E[Y|X]]$ [4]

---

The remaining details of the derivation are left for the student to work out. The important intermediate result (students should confirm) is:

$$E[a^2] = \rho + \lambda^2\sigma_s^2 + \rho^2$$

[3] $X$ and $Y$ are arbitrary RV

[4] Convince yourself of these two general results.

# The Pollaczek-Khinchin Mean Value Formula
# Results for the M-G-1 Queueing System

**Bringing Everything Together:** One form of the P-K mean value formula is given as follows:

$$E[n] = \frac{2\rho - \rho^2 + \lambda^2\sigma_s^2}{2(1-\rho)} \tag{11}$$

Some algebraic manupulation and application of **Little's Law** bring about some variations for of the same result:

$$E[n] = \rho + \frac{\rho^2 + \lambda^2\sigma_s^2}{2(1-\rho)} \tag{12}$$

$$= \rho + \frac{\lambda^2\overline{X^2}}{2(1-\rho)} \tag{13}$$

ECE 3511: Communications Networks
"Theory and Analysis"

Fall Quarter 2002

Instructor: Prof. A. Bruce McDonald

Lecture Topic

**Introductory Analysis of M/G/1 Queueing Systems**

Module Number Two

**Steady-State System Behavior—The Residual Work Approach**

# Fundamental Queueing Theory
# Equilibrium Behavior of the M-G-1 Queueing System
### Using Residual Service Times to Derive and Understand the
### *Pollaczek-Khinchin (P-K) Formula*

## Some Notation

- $W_i$ = Waiting time of the $i^{th}$ customer
- $R_i$ = Residual service time as seen by the $i^{th}$ customer
- $X_i$ = Service time of the $i^{th}$ customer
- $N_i$ = Number of waiting customers at the $i^{th}$ customer arrival
- $R$ = Mean residual waiting time ($R = \lim_{i\to\infty} E[R_i]$)
- $N_Q$ = Mean number of customers in the queue

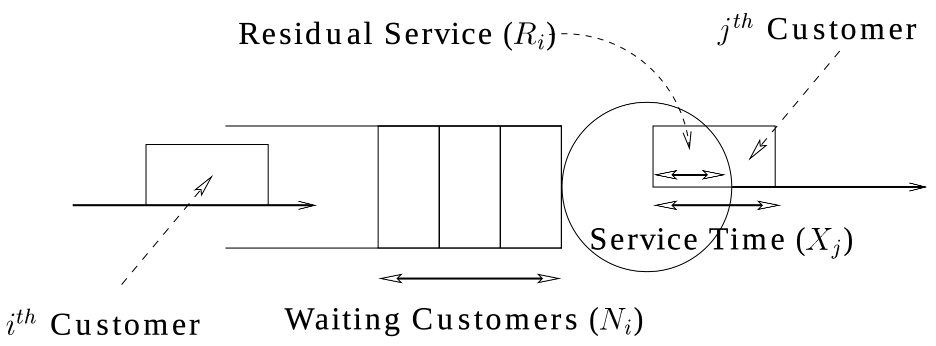


Figure 1: **Illustration of residual service seen by the $i^{th}$ customer.**

## What is the Residual Service Time?

## Definition of Residual Service Time

- If customer $j$ is being served when customer $i$ arrives, $R_i$ is the time remaining until $j$'s service is complete.
- If the system is empty then $R_i = 0$.

# Derivation of the P-K Formula

- Waiting time for the $i^{th}$ Customer:

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

- Take the expectation of both sides and condition $X_j$ on $N_i$:

$$\begin{aligned} E[W_i] &= E[R_i] + E[\sum_{j=i-N_i}^{i-1} X_j] \\ &= E[R_i] + E[\sum_{j=i-N_i}^{i-1} E[X_j|N_i]] \end{aligned} \quad (1)$$

- Recall from *Probability* that: $E[X] = E[E[X|Y]]$; Work this out to convince yourself;
- Furthermore, $N_i$ is independent of $X_k \forall k \in (i-1, i-N_i)$ , hence, Equation-1 can be re-written:

$$E[W_i] = E[R_i] + \overline{X} E[N_i]$$

- Take the limit of each side as $i \to \infty$:

$$\lim_{i \to \infty} E[W_i] = W = R + \frac{1}{\mu} N_Q \quad (2)$$

# Determining the Mean Residual Service Time

## Graphical Argument

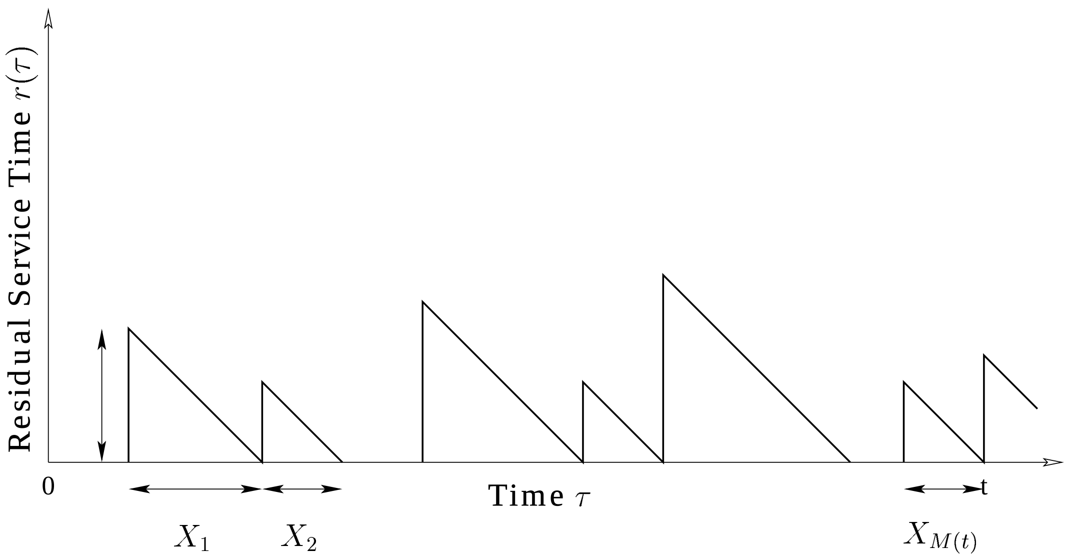


Figure 2: **Graphical Representation of Mean Residual Time.**

- Consider the time average of the residual service time, $r(\tau)$, during the interval $[0, t]$
- Let $M(t)$ be the number of *service completions* up to time $t$

$$\begin{aligned} \frac{1}{t}\int_0^t r(\tau)d\tau &= \frac{1}{t}\sum_{i=1}^{M(t)} \frac{1}{2}X_i^2 \\ &= \frac{1}{2}\frac{M(t)}{t}\frac{\sum_{i=1}^{M(t)} X_i^2}{M(t)} \end{aligned} \tag{3}$$

## Determining the Mean Residual Service Time Continued

- *Assuming* that the process is ergodic time averages can be replaced by ensemble averages!
- Determine the time average first;
- Take the limits of both sides of Equation-3 as $t \to \infty$ :

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t r(\tau)d\tau = \frac{1}{2}.\lim_{t\to\infty}\frac{M(t)}{t}.\lim_{t\to\infty}\frac{\sum_{i=1}^{M(t)} X_i^2}{M(t)}$$

- On the left is the *time-average* of the residual service time: $R$
- The $1^{st}$ limit on the right is the *time-average* of the service completion rate, which equals the arrival rate: $\lambda$ **(WHY?)**
- The $2^{nd}$ limit on the right is the *second-moment* of the service time: $E[X_i^2]$
- Based on the ergodic assumption:

$$R = \frac{1}{2}\lambda\overline{X^2} \tag{4}$$

# Evaluating the P-K Formula

---

- Substituting the result for the expected value of the residual work given in Equation-4 into Equation-2:

$$W = \frac{1}{2}\lambda\overline{X^2} + \frac{1}{\mu}N_Q$$

- But what is the value of the mean number in the queue ($N_Q$) seen by an arrival ?
- Based on the PASTA property defined earlier it is the *typical value* at an arbitrary instant—the expected value! Hence, by Little's Law:

$$\begin{aligned} N_Q &= \lambda W \\ W &= \frac{1}{2}\lambda\overline{X^2} + \rho W \\ &= \frac{\lambda\overline{X^2}}{2(1-\rho)} \end{aligned} \tag{5}$$

ECE 3511: Communications Networks
Theory and Analysis
Fall Quarter 2002

Instructor: Prof. A. Bruce McDonald

**Lecture Topic:**

| **Introductory Analysis of M/G/1 Queueing Systems** |
| --- |

**Module Three:**

**M-G-1 Examples Using the P-K Formula**

# Fundamental Queueing Theory

## Equilibrium Behavior of the M-G-1 Queueing System

### Some Examples from the P-K Mean Value Formula

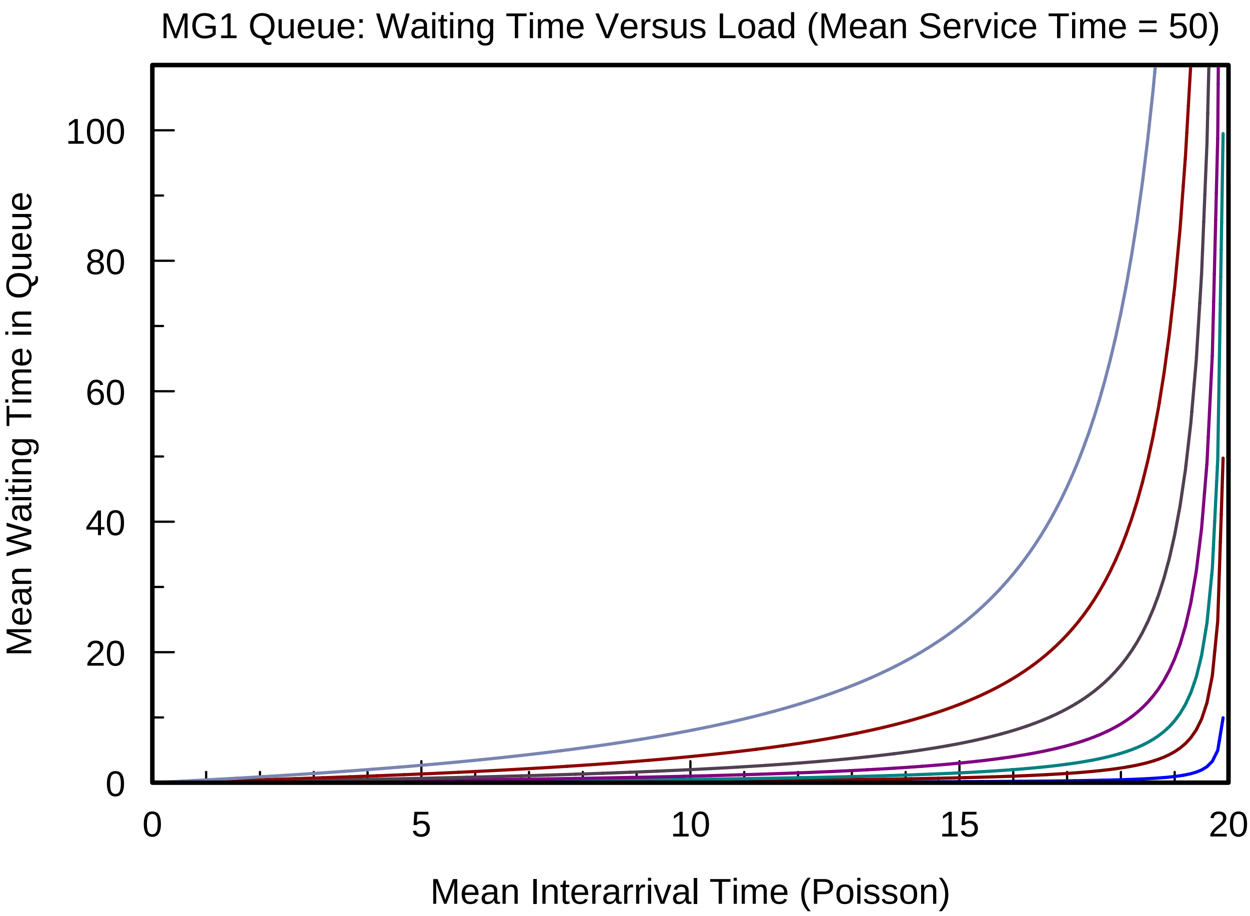


Figure 1: Effect of Incresing the *Variance* in Service Time with MG1 System.

- How is the mean waiting time determined?
- Try applying Little's Law

Using the recursion approach the *P-K Mean Value Formula* was derived for the number of customers in the system. Application of Little's Law and some algebraic manipulation transforms it into the more commonly used form that specifies the mean waiting time in the queue—the form that is derived directly when using the residual service time approach.

# Equilibrium Behavior of the M-G-1 Queueing System
# The MD1 Queueing System

## Deterministic Service Times: The M-D-1 Queue

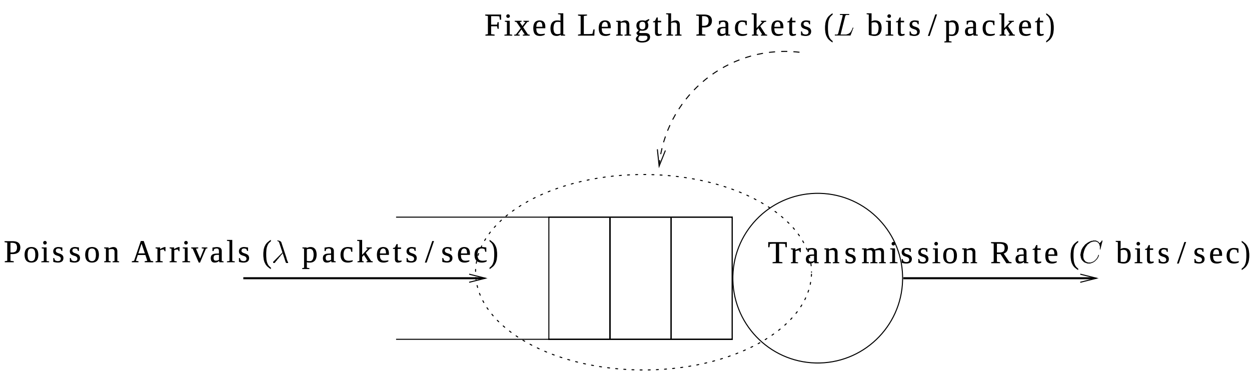


Figure 2: **Example of an MD1 System—Based on Fixed Packet Lengths.**

MG1 results provide a nice metholdology to gain insight regarding behavior of systems that use *fixed* packet lengths [1].

- Fixed length packets transmitted on a channel with a constant transmission rate will all be serviced in precisely the *same* time: That time is also the mean and the variance will be zero. Hence, for the MD1 system:

$$E[n] = \rho + \frac{\rho^2}{2(1-\rho)}$$

- Assuming IID Exponentially Distributed service times the P-K MVF reduces (as it should) to the well known MM1 result:

$$E[n] = \frac{\rho}{(1-\rho)}$$

[1] **NOTE:** Be **careful** not to confuse "service rate" with the transmission rate... Convince yourlself you understand the difference!

## Comparison of MD1 and MM1 Systems

With some algebraic manipulation one can show that the mean number in the system for an MD1 queue is *always* going to be less than the mean number in an MM1 queue given the same arrival rate and mean packet length (assuming the transmission rate is fixed). For the MD1 queue the P-K formula can be re-written as:

$$E[n]^{(\mathbf{MD1})} = \frac{\rho}{(1-\rho)} - \frac{\rho^2}{2(1-\rho)} = E[n]^{(\mathbf{MM1})} - \frac{\rho^2}{2(1-\rho)}$$

Under backlogged conditions, i.e., as $\rho \rightarrow 1$, it can be shown [2] that the delay for the MM1 system approaches twice that for the "equivalent" MD1 system.

**Why is this the case?**

[2] Students should work this out!

# Equilibrium Behavior of the M-G-1 Queueing System

## Delay Analysis of an ARQ System

## Analysis of delay in a *Go-Back-N* ARQ System

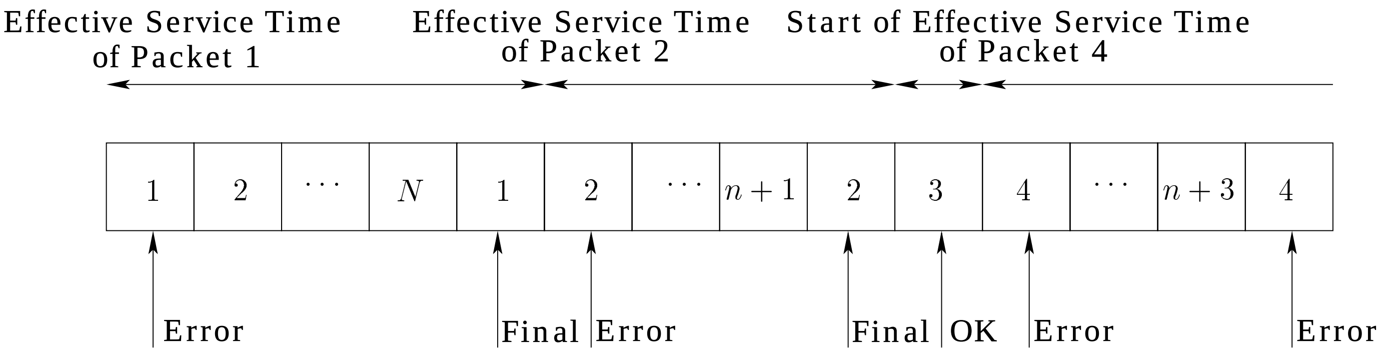


Figure 3: **Effective Service Times for Packets in an ARQ System** (Adapted from "Data Networks", Second Edition, Dimitri Bertsekas and Robert Gallager, 1992)

Consider the following *go-back-n* ARQ system:

- Packets are transmitted in frames that are *one* time unit long;
- There is a *maximum wait* time for an acknowledgement of $n-1$ frames before a packet is retransmitted;
- There are only *two* causes of packet retransmission:
  1. An error is detected at the receiver in frame $i$; the transmitter will transmit frames $i+1, i+2, \ldots, i+n-1$ (assuming there is a sufficient backlog of packets to send)—the corrupted packet will be retransmitted in frame $i+n$.
  2. The acknowledgement for the packet transmitted and received *without* error is not processed by the source by the completion of packet $i+n-1$. This may occur due to bit errors on the return channel, long propagation delays, or long frames (for piggybacked ACKs) in the return direction.
- For the current analysis assume that the probability of the second event is very small and can be ignored; thus only the probability of a *forward* packet received in error is considered.
- Assume that the probability a packet is rejected due to error is given by: $Pr[Packet\ Error] = p$ and is *independent* of all other packets.

## Delay in a *go-back-n* ARQ System (Continued)

- Packets arrive according to a Poisson Process with rate $\lambda$ (MG1)
- One way to analyze this system is to consider the *effective service time*
- How long does it actually take (on average) to correctly transmit a new packet across the link? **Hint:** What is the distribution of retransmissions?

$$Pr(There\ are\ k\ retransmissions) = (1-p)p^k$$

Define the time it takes for the $k+1$ transmissions to be the effective service time for the packet: $k$ transmissions involving a forward error and one correct transmission.

- Objective: determine steady-state behavior of the system.
- Based on the *model* proposed so far—what do we do now?

## Delay in a *go-back-n* ARQ System (Continued)

- An important assumption infers that the timer (for timeout and retransmission) will always trigger retransmission at the instant the window is exhausted;
- In Figure-3 it can be seen what this means (Assume constant backlog):
  1. Any packet transmitted after an *errored* packet will be discarded and retransmitted—error or not!
  2. Due to indenpendence this has no effect on the probability of future transmission errors.
  3. Only the receive status of the $1^{st}$ packet in a window of frames is significant with respect to delay—refer to this as a *primary* error and errors in succeeding packets within the same window as *secondary* errors.
  4. Consider only primary errors—how many packet transmission (time units) are used for a retransmission? $(n + 1)$ **WHY**
- The distribution of the *number* of primary errors is known;
- The number of time units required retransmissions due to primary errors is known;

## Bringing it All Together: Go-Back-N Analysis

The expected delay before receiving a packet correctly is equivalent to the product of the expected number of (primary) retransmissions and the time required for each retransmission: [3] [4]

- Denote the effective service time by the random variable $X$:

$$Pr\{X = 1 + kn\} = (1-p)p^k \quad k = 0, 1, \dots$$

- The first moment of the effective service time is:

$$\begin{aligned}
\overline{X} &= \sum_{k=0}^{\infty}(1+kn)(1-p)p^k \\
&= (1-p)(\sum_{k=0}^{\infty} p^k + n\sum_{k=0}^{\infty} kp^k) \\
&= (1-p)(\frac{1}{(1-p)} + \frac{np}{(1-p)^2}) \\
&= 1 + \frac{np}{(1-p)} \qquad (1)
\end{aligned}$$

- The second moment of the effective service time is:

$$\begin{aligned}
\overline{X^2} &= \sum_{k=0}^{\infty}(1+kn)^2(1-p)p^k \\
&= (1-p)(\sum_{k=0}^{\infty} p^k + 2n\sum_{k=0}^{\infty} kp^k + n^2\sum_{k=0}^{\infty} kp^k) \\
&= (1-p)(\frac{1}{(1-p)} + \frac{2np}{(1-p)^2}) + n^2\frac{(p+p^2)}{(1-p)^3} \\
&= 1 + \frac{2np}{(1-p)} + \frac{n^2(p+p^2)}{(1-p)^2)} \qquad (2)
\end{aligned}$$

[3]**NOTE:** Students must review probability and/or concrete math in order to ensure they can solve the go following derivations.

[4]**HINT:** The first summation is the geometric series—the next ones can be solved using differentiation.

## Final Result

- **Students: Now what??**

- All the information is now available to directly substitute values into the *P-K Formula*! **Work this out on your own...**

- The *P-K Formula* gives the *mean waiting time* in the queue...

- How can the mean time *in the system*—end-to-end latency be determined?

- Justify why only the *primary errors* were considered? Don't the packet transmissions that are dropped in the *go-back-n* windows affect the distribution of retransmissions??

## MG1 Queues with Busy Periods and Vacations

What about a system that is *non work-conserving*?
Again—residual work will provide an effective approach to solve this problem.

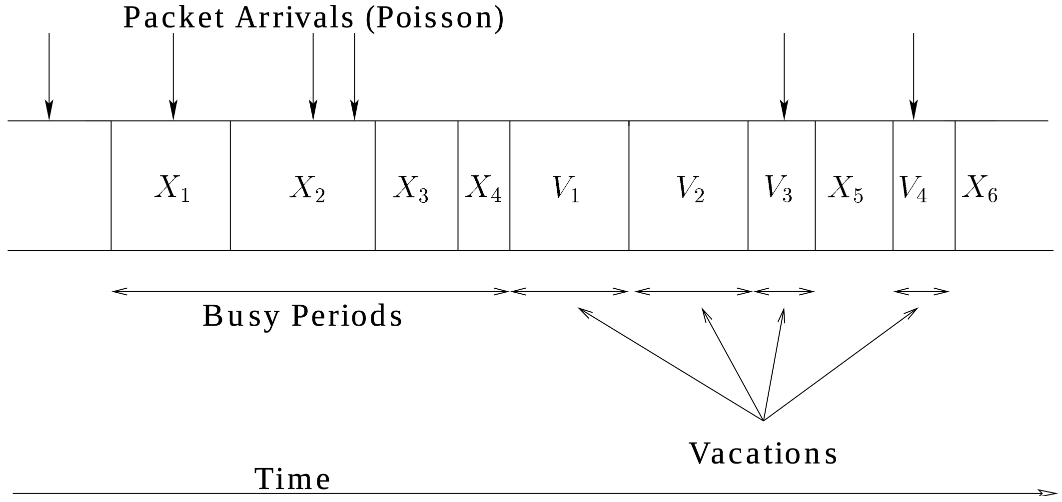


Figure 4: **Timing Diagram of an MG1 System with Vacations** (Adapted from "Data Networks", Second Edition, Dimitri Bertsekas and Robert Gallager, 1992)

- At the end of each *Busy Period* the MG1 Server "rests" for a random interval of time $V_j$ with first and second moments $\overline{V}$ and $\overline{V^2}$.
- The server cannot begin serving any customers that arrives during a vacation until the vacation has ended. Any customer arriving to an empty system during a vacation must wait until the end of the vacation to begin service.
- If the system remains empty on completion of any vaction, the server takes another vacation of duration $V_k$ that is IID with $V_j$.

## The M-G-1 Queueing System with Vacations

### Analysis Based on the *Residual Work* Approach

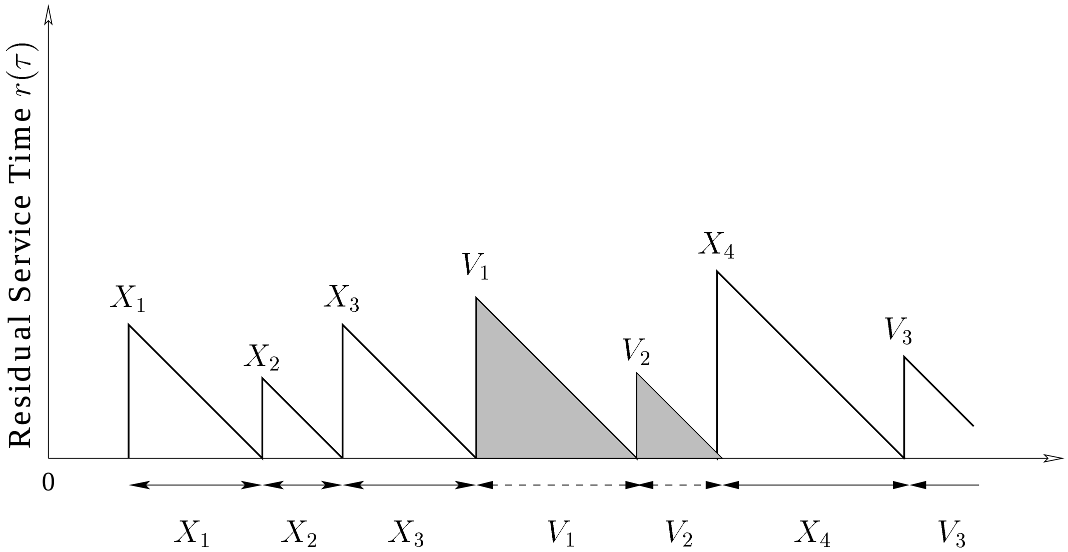


Figure 5: **Residual Service Times for the M-G-1 System with Vacations** (Adapted from "Data Networks", Second Edition, Dimitri Bertsekas and Robert Gallager, 1992)

- In a real network what are some possible *purposes* or *causes* for *vacations*?
- Based on the current model a vacation occurs $\Longleftrightarrow$ the system is empty at the end of a busy period.
- Does this have to be the case? Could M-G-1 still be used to model such a scenario? (e.g. take a vacation if there are on low priority packets in queue, etc.)
- How should the analysis be approached?

# The M-G-1 Queueing System with Vacations

## Analysis Based on the *Residual Work* Approach

- Follow the same (almost) approach as in the Residual Serivice derivation for the P-K Formula;
- In this case the *Vacation Completions* must be considered distinct from the *Service Completions*!
- The arrival process is Poisson—the set of random variables: $(V_1, V_2, \ldots)$ are the successive vacation times.
- What conditions are required on the distribution of the vacation times?

# The M-G-1 Queueing System with Vacations

## Analysis Based on the *Residual Work* Approach

Let $M(t)$ be the number of *service completions* up to time $t$ and let $L(t)$ be the number of *vacation completions* up to time $t$:

$$\begin{aligned} \frac{1}{t}\int_0^t r(\tau)d\tau &= \frac{1}{t}\sum_{i=1}^{M(t)}\frac{1}{2}X_i^2 + \frac{1}{t}\sum_{i=1}^{L(t)}\frac{1}{2}V_i^2 \\ &= \frac{M(t)}{t}\frac{\sum_{i=1}^{M(t)}\frac{1}{2}X_i^2}{M(t)} + \frac{L(t)}{t}\frac{\sum_{i=1}^{L(t)}\frac{1}{2}V_i^2}{L(t)} \end{aligned} \tag{3}$$

- Take the limits of both sides of Equation-3 as $t \to \infty$ :

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t r(\tau)d\tau = R = \frac{\lambda\overline{X^2}}{2} + \frac{(1-\rho)\overline{V^2}}{2\overline{V}}$$

- What are the arguments required to arrive at the solution for mean residual service time ($R$)?

  1. Assume that steady-state exists...
  2. As $t \to \infty$ argue that $\frac{M(t)}{t} \to \lambda$;
  3. As $t \to \infty$ argue that $\frac{L(t)}{t} \to \frac{(1-\rho)}{\overline{V}}$;
  4. **What else?**

- Must there be mutual independence amoung the vacation intervals?

- Must there be independence between the vacation intervals and the customer arrival and service times?