THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 6, July-August, 1972
Printed in U.S.A.

Queues With Mixed Renewal and Poisson Inputs

By A. KUCZURA

(Manuscript received February 2, 1972)

In a queueing system with two independent input streams, as exists, for example, when first-routed and overflow traffic streams are offered to a common sender-group, the state of the system encountered by the two different types of customers upon their arrival will generally be different. Consequently, in a system where delayed customers wait for service, the service rendered to the individual streams may also be different.

The delay distribution in a single-server queue for each type of customer is derived under the assumption that one stream is Poissonian and the other is described by a renewal process. The difference in service received by the two streams is examined with the aid of numerical examples for two interarrival time distributions of the renewal stream. We show for two cases that a practical indicator of service received by the renewal customers is the coefficient of variation of their interarrival time distribution. If the coefficient is less than unity, then the renewal customers receive better service than the Poissonian customers. The converse is true when the coefficient exceeds unity.

The stationary distribution of the number of busy servers in an infinite-server system as seen by the two types of customers is also derived.

I. INTRODUCTION

The concept of a piecewise Markov process[1] is used to analyze two queueing systems the inputs of which are composed of two independent streams. One of the streams is Poissonian with intensity λ and the other (called a GI stream because of its General Independent Distribution of intervals between arrivals) is assumed to be a renewal process with intensity ν. We assume the service times of all the customers are independent and identically distributed according to an exponential distribution with mean μ^{-1}. Such models are denoted by GI + M/M/c in Kendall's notation, where the "M" refers to the Markovian character of Poissonian arrivals and of exponential service and c refers to the number of servers.

The state of the system (number of customers waiting and in service) seen by an arriving Poissonian customer will generally differ from that of the GI customer. Consequently, in a system where delayed customers wait for service such as in the GI + M/M/1 queue, the service received by the two types of customers will differ. Whether the GI customers receive better or worse service than the Poissonian customers depends on the variability of the interarrival times of the GI stream.

In Section II, we analyze the GI + M/M/1 queue. The delay distribution with order-of-arrival service for the two types of customers is derived for GI streams the interarrival time distributions of which have rational Laplace–Stieltjes transforms. The coefficient of variation of the interarrival time distribution is introduced as a practical measure of the variability properties of a stream of customers. Its usefulness in predicting service is evaluated through some numerical examples and its relation to the common measure called peakedness is examined. In Section III, in order to describe the intrinsic character of the two streams of customers as they would be observed in a system without delay, we derive the stationary distribution of the number of busy servers seen by the two types of arriving customers in a GI + M/M/∞ system.

II. THE GI + M/M/1 QUEUE

Let $Y(t)$ be the number of customers in the system (those waiting and in service) at time t. Since the GI stream is a renewal process and $Y(t)$ is Markovian between any two consecutive arrival epochs of the GI customers, $\{Y(t), t \geqq 0\}$ is a piecewise Markov process[1] with state space $\{0, 1, 2, \cdots\}$. The regeneration points are the arrival epochs of the GI customers. Thus the distribution of the length of the Markovian segments is given by $A(\xi)$, the interarrival time distribution of the GI customers. The regeneration matrix is given by

$$(p_{ij}) = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ & & \cdots & & \end{bmatrix} \tag{1}$$

The elements of this matrix are the transition probabilities across a regeneration point; that is, p_{ij} is the probability that immediately after regeneration the process is in state j, given that, immediately prior to regeneration, the process was in state i.

The Markov process operating within the segments is a birth-death process with birth rate λ and death rate μ, the same for all segments. The transition probability functions of this process are given in Ref. 2, page 13.

$$P_{ij}(t) = \rho^{\frac{1}{2}(j-i)} e^{-(\lambda+\mu)t} \Big\{ I_{i-j}(2t\sqrt{\lambda\mu}) + \rho^{-\frac{1}{2}} I_{i+j+1}(2t\sqrt{\lambda\mu}) + (1-\rho) \sum_{k=1}^{\infty} \rho^{\frac{1}{2}(k+1)} I_{i+j+k+1}(2t\sqrt{\lambda\mu}) \Big\},$$

where $\rho = \lambda/\mu$ ($\lambda/\mu < 1$) and $I_j(\xi)$ is the modified Bessel function

$$I_j(\xi) = \left(\frac{\xi}{2}\right)^j \sum_{k=0}^{\infty} \frac{1}{k!\,(j+k)!} \left(\frac{\xi}{2}\right)^{2k}.$$

The regeneration matrix (1), the distribution $A(\xi)$, and the transition functions $P_{ij}(t)$ determine the piecewise Markov process completely. We will first derive the distribution of the state of the system seen by an arbitrary GI arrival and then use the rate conservation principle to find the stationary distribution

$$q_j = \lim_{t\to\infty} P\{Y(t) = j \mid Y(0) = i\}, \qquad j = 0, 1, 2, \cdots, \quad \text{for all} \quad i, \tag{2}$$

which is the same as the distribution of the state of the system seen by a Poissonian arrival. Having found the two distributions, we can readily determine the individual delay distributions for an order-of-arrival service discipline.

2.1 *Delay Sustained by GI Customers*

Let $\{p_j\}$ be the stationary distribution of the Markov chain imbedded at points immediately preceding a GI arrival. If r_{ij} is the one-step transition probability from state i to state j, then

$$r_{ij} = \int_0^\infty P_{i+1,j}(\xi)\, dA(\xi), \qquad i, j = 0, 1, 2, \cdots.$$

The distribution $\{p_j\}$ satisfies the Chapman–Kolmogorov equations

$$p_j = \sum_{i=0}^{\infty} p_i r_{ij}, \qquad j = 0, 1, 2, \cdots,$$

and hence the solution to the linear system

$$p_j = \int_0^\infty \sum_{i=0}^\infty p_i P_{i+1,j}(\xi)\, dA(\xi), \qquad j = 0, 1, 2, \cdots, \tag{3}$$

subject to the normalization condition

$$\sum_{j=0}^\infty p_j = 1$$

is the stationary distribution $\{p_j\}$.

Let

$$p(u) = \sum_{i=0}^\infty p_i u^i$$

be the probability generating function of $\{p_j\}$. Multiplying (3) by u^j and summing over all j, we obtain

$$p(u) = \int_0^\infty \sum_{i=0}^\infty p_i \Gamma_{i+1}(u, \xi)\, dA(\xi), \tag{4}$$

where

$$\Gamma_i(u, \xi) = \sum_{j=0}^\infty P_{ij}(\xi) u^j$$

is the probability generating function of $P_{ij}(\xi)$, $j = 0, 1, 2, \cdots$. This function is not readily available. Its Laplace transform

$$\gamma_i(u, s) = \int_0^\infty e^{-s\xi} \Gamma_i(u, \xi)\, d\xi,$$

though, has the following form[2]:

$$\gamma_i(u, s) = \frac{u^{i+1} - (1 - u)\eta^{i+1}(s)/[1 - \eta(s)]}{u[s - h(u)]}$$

where

$$h(u) = \frac{1}{u}(1 - u)(\mu - \lambda u),$$

and

$$\eta(s) = \frac{\lambda + \mu + s - \sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu}}{2\lambda}.$$

We will transform the real integral

$$\int_0^\infty \Gamma_{i+1}(u, \xi)\, dA(\xi) \tag{5}$$

appearing in (4) into a complex integral involving $\gamma_{i+1}(u, s)$ and $\alpha(s)$, the Laplace–Stieltjes transform of $A(\xi)$.

If we set

$$dA(\xi) = e^{-a\xi}\, dB(\xi), \qquad a > 0,$$

then (5) becomes the Laplace–Stieltjes transform integral

$$\int_0^\infty e^{-a\xi}\Gamma_{i+1}(u, \xi)\, dB(\xi). \tag{6}$$

But the transform of a product is the complex convolution of the transforms of its factors and so (6) becomes

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \gamma_{i+1}(u, z)\beta(a - z)\, dz,$$

where c is any positive number, $i = \sqrt{-1}$, and

$$\beta(s) = \int_0^\infty e^{-s\xi}\, dB(\xi), \tag{7}$$

the Laplace–Stieltjes transform of $B(\xi)$. But

$$\beta(s) = \int_0^\infty e^{-\xi(s-a)}\, dA(\xi) = \alpha(s - a),$$

and we finally have

$$\int_0^\infty \Gamma_{i+1}(u, \xi)\, dA(\xi) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \gamma_{i+1}(u, z)\alpha(-z)\, dz. \tag{8}$$

Substituting for $\gamma_{i+1}(u, z)$ on the right-hand side of (8), and using the resulting identity in (4), we obtain the following integral equation for $p(u)$:

$$p(u) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \left[\frac{u^2p(u) - (1 - u)H(z)}{u[z - h(u)]}\right]\alpha(-z)\, dz, \tag{9}$$

where

$$H(z) = \frac{\eta^2(z)p[\eta(z)]}{1 - \eta(z)}.$$

We will evaluate the complex integral in (9) for the class of $\alpha(s)$, the members of which are rational functions; but first we note some properties of the integrand which suggest a contour to be used in applying the calculus of residues.

Since $P_{ij}(\xi)$ is a probability function, $\Gamma_i(u, \xi)$ is uniformly convergent for $|u| \leqq 1$ and $\gamma_i(u, z)$ is holomorphic for $|u| \leqq 1$ and $\Re(z) > 0$, where $\Re(z)$ denotes the real part of z. Hence the bracketed part of the integrand in (9) is holomorphic for $|u| \leqq 1$ and $\Re(z) > 0$ since it is the convergent series

$$\sum_{i=0}^{\infty} p_i \gamma_{i+1}(u, z).$$

Since $A(\xi)$ is a probability distribution function, $\alpha(z)$ is holomorphic for $\Re(z) > 0$. For $\Re(z) < 0$, $\alpha(z)$ may or may not be holomorphic. The predominant case is where $\alpha(z)$ is meromorphic in the half-plane $\Re(z) < 0$ and we shall address ourselves to this case.

Let $-z_1, -z_2, \cdots, -z_n$ be the poles of $\alpha(z)$. Since the poles of $\alpha(z)$ are in the left-half plane, the poles of $\alpha(-z)$ are in the right-half plane. Hence, the integrand in (9) is meromorphic in the right-half plane and we can use the residue theorem to evaluate the integrand over the contour consisting of the line $(c + iR, c - iR)$ and a semicircle of radius R in the right-half plane which connects $c - iR$ with $c + iR$. We choose c and R such that all the poles of $\alpha(-z)$ are interior to this contour.

Since $\alpha(z)$ is meromorphic, we can write

$$\alpha(z) = \frac{Q_m(z)}{(z + z_1)^{k_1}(z + z_2)^{k_2} \cdots (z + z_n)^{k_n}},$$

where k_j is the order of the pole at z_j, $Q_m(z)$ is a polynomial of degree m, and we assume that $m + 1 \leqq k_1 + k_2 + \cdots + k_n$. Two examples are:

(*i*) $A(\xi)$ is the gamma distribution with density

$$a(\xi) = \frac{k\nu}{(k-1)!}(k\nu\xi)^{k-1}e^{-k\nu\xi}, \qquad \xi > 0,$$

and transform

$$\alpha(z) = \left(\frac{k\nu}{k\nu + z}\right)^k; \tag{10}$$

(*ii*) $A(\xi)$ is the mixture of exponentials (hyperexponential)

$$A(\xi) = \sum_{i=1}^{n} a_i(1 - e^{-\nu_i \xi}), \qquad \xi > 0,$$

where $a_i > 0$, $\nu_i > 0$ and $a_1 + a_2 + \cdots + a_n = 1$, and transform

$$\alpha(z) = \sum_{i=1}^{n} \frac{a_i \nu_i}{\nu_i + z}. \tag{11}$$

If we now let $R \to \infty$, then the contour integral along the line $z = c$ tends to the desired limits and it can be shown that the integral along the semicircle tends to zero. Hence, using Cauchy's integral formula with

$$f(z) = \frac{u^2 p(u) - (1 - u)H(z)}{u[z - h(u)]} Q_m(-z),$$

we have

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{(-1)^k f(z)}{(z - z_1)^{k_1}(z - z_2)^{k_2} \cdots (z - z_n)^{k_n}} dz$$

$$= (-1)^{k+1} \sum_{j=1}^{n} \frac{1}{(k_j - 1)!} g_j^{(k_j-1)}(z_j), \qquad (12)$$

where $k = k_1 + k_2 + \cdots + k_n$ and

$$g_j(z) = \frac{f(z)}{(z - z_1)^{k_1} \cdots (z - z_{j-1})^{k_{j-1}}(z - z_{j+1})^{k_{j+1}} \cdots (z - z_n)^{k_n}}.$$

If all the poles are simple, such as in example (*ii*), then the integral in (12) is equal to

$$(-1)^{n+1} \sum_{j=1}^{n} g_j(z_j). \qquad (13)$$

If z_1 is the only pole and it is of order k_1, such as in example (*i*), then the integral in (12) is equal to

$$\frac{(-1)^{k_1+1}}{(k_1 - 1)!} f^{(k_1-1)}(z_1). \qquad (14)$$

Detailed analysis for the case $k = 2$ will be carried out later when we give numerical examples. The case of two simple poles will also be analyzed.

In general, after the integral has been evaluated, and the result substituted into (9), an equation in $p(u)$ results. This equation can be solved for $p(u)$. The values of the unknown function $H(z)$ at the poles z_j and the values of its derivatives may be determined by first applying the normalization condition $p(1) = 1$ and then, since $p(u)$ is holomorphic in $|u| \leq 1$, forcing the zeros of the numerator in the unit circle to coincide with the zeros of the denominator in the expression for $p(u)$. This procedure will be illustrated later by an example.

We return to the delay distributions. Let W be the delay, or waiting time, from the arrival instant until the beginning of service for an arbitrary GI customer. The GI delay distribution is defined by

$$D(t) = P[W \leqq t], \quad t \geqq 0.$$

If $D_j(t)$ is the conditional delay distribution, given that the arriving customer finds j other customers in the system, then

$$D(t) = \sum_{j=0}^{\infty} D_j(t)p_j .$$

Obviously, $D_0(t) = 1$. Since service is in order of arrival,

$$D_1(t) = 1 - e^{-\mu t},$$

and $D_j(t)$ for $j > 1$ is the convolution of j identical exponential distributions, each with mean μ^{-1}. If $\delta(s)$ is the Laplace–Stieltjes transform of $D(t)$, then

$$\delta(s) = p_0 + \sum_{j=1}^{\infty} p_j \int_0^{\infty} e^{-st}\, dD_j(t)$$

$$= p_0 + \sum_{j=1}^{\infty} p_j \left(\frac{\mu}{s+\mu}\right)^j,$$

since the transform of a convolution is the product of the transforms. Consequently we have

$$\delta(s) = p\left(\frac{\mu}{s+\mu}\right).$$

This equation can now be inverted to obtain the delay distribution for the GI customers.

2.2 *Delay Sustained by Poissonian Customers*

Let $E(t)$ be the delay distribution of Poissonian customers; that is, $E(t)$ is the probability that an arbitrary, arriving Poissonian customer will be delayed no more than t units of time. If $\epsilon(s)$ is its Laplace–Stieltjes transform, then, by the same argument as above,

$$\epsilon(s) = q\left(\frac{\mu}{s+\mu}\right),$$

where

$$q(u) = \sum_{j=0}^{\infty} q_j u^j$$

is the probability generating function of $\{q_j\}$, the distribution of state seen by an arbitrary, arriving Poissonian customer.

Using the rate conservation principle[1] we can write

$$\nu p_j + \lambda q_j = \mu q_{j+1}, \quad j = 0, 1, 2, \cdots, \tag{15}$$

where the left-hand side is the asymptotic rate of transition from j to $j + 1$ and the right-hand side is the asymptotic rate of transition from $j + 1$ to j. Multiplying both sides of (15) by u^j and summing over all j, we obtain the following relation between $p(u)$ and $q(u)$:

$$\nu p(u) + \lambda q(u) = \frac{\mu}{u}[q(u) - q_0].$$

Thus, if $p(u)$ has been found, then $q(u)$ is given by

$$q(u) = \frac{q_0 + u\sigma p(u)}{1 - \rho u},$$

where $\sigma = \nu/\mu$ and $\rho = \lambda/\mu$. Applying the normalization condition $q(1) = 1$, we get q_0, i.e.,

$$q_0 = 1 - \rho - \sigma,$$

and hence

$$q(u) = \frac{1 - \rho - \sigma + u\sigma p(u)}{1 - \rho u}. \tag{16}$$

In terms of the probabilities themselves, we solve (15) for $\{q_j\}$ and obtain

$$q_j = (1 - \rho - \sigma)\rho^j + \sigma \sum_{i=0}^{j-1} p_i \rho^{j-i-1}, \qquad j = 1, 2, \cdots .$$

2.3 *Some Relations Between the Two Results*

If l_1 is the mean of the distribution $\{p_j\}$ and l_2 is the mean of the distribution $\{q_j\}$, then using (16) and differentiating, we obtain

$$l_2 = \frac{\rho}{1 - \rho} + \frac{\sigma}{1 - \rho}(1 + l_1). \tag{17}$$

Note that the first term on the right-hand side of (17) is the mean number of customers in the system in the M/M/1 queue with traffic intensity ρ. Hence, the second term may be considered to be an increase in the mean number due to the presence of the GI customers.

Since the probability distribution functions $p(u)$ and $q(u)$ are related through (16), the transforms of the delay distributions are related through the following equation:

$$\epsilon(s) = \frac{s+\mu}{s+\mu-\lambda}(1-\rho-\sigma) + \frac{\nu}{s+\mu-\lambda}\,\delta(s). \tag{18}$$

Taking the inverse transform of both sides of this equation and rearranging the terms, we obtain

$$E(t) = 1 - \rho e^{-(\mu-\lambda)t} - \sigma\left\{1 + \mu\int_0^t [\rho - D(t-\xi)]e^{-(\mu-\lambda)\xi}\,d\xi\right\}. \tag{19}$$

If τ_1 is the mean delay of the GI customers and τ_2 is the mean delay of the Poissonian customers, then from (18) we have

$$\tau_2 = \frac{\rho}{\mu(1-\rho)} + \frac{\sigma}{\mu(1-\rho)}(1+\mu\tau_1).$$

Again, the second term on the right-hand side of this equation can be thought of as the added mean delay due to the presence of the GI customers.

2.4 *Measures of Variability of a Traffic Stream*

The delay suffered by an arbitrary customer in an input traffic-stream depends on how the customers' arrival epochs are distributed. Roughly, it can be said that the less "variation" in the arrival epochs, the better is the service received by the customers. We discuss two measures of this variation.

Let $A(\xi)$ be the interarrival time distribution of a traffic stream and μ_j the jth moment of $A(\xi)$:

$$\mu_j = \int_0^\infty \xi^j\, dA(\xi), \qquad j = 1, 2, \cdots .$$

We define V, the coefficient of variation of $A(\xi)$, by

$$V = \frac{\sqrt{\mu_2 - \mu_1^2}}{\mu_1}.$$

This measure is dimensionless and depends only on the properties of the stream itself.

Another measure which is used extensively in telephone traffic theory is defined with the aid of an infinite server system. The traffic stream is offered to an infinite number of servers with exponentially distributed service times. The ratio of the variance to the mean of the number of busy servers in statistical equilibrium is taken as a measure of the variation of the traffic stream. This number is called peakedness and is customarily denoted by Z.

We shall classify all traffic streams into two categories: smooth and peaked. A stream is smooth if $V < 1$ and peaked if $1 < V$.* The same dichotomy is effected by the inequalities $Z < 1$ and $1 < Z$; however, Z is a function of the stream's intensity and thus is not as convenient in measuring variability properties. For example, if $\alpha(s)$ is the Laplace–Stieltjes transform of $A(\xi)$, ρ is the stream's traffic intensity, and μ^{-1} is the mean holding time, then it follows (Ref. 3, Chapter 3) that

$$Z = \frac{1}{1 - \alpha(\mu)} - \rho. \tag{20}$$

Using Jensen's inequality,[4] one can show that for a fixed λ, the stream's peakedness (20) attains a minimum whenever $A(\xi)$ is the one-point distribution

$$A(\xi) = \begin{cases} 0, & \xi < \dfrac{1}{\lambda} \\ 1, & \dfrac{1}{\lambda} \leqq \xi. \end{cases}$$

In this case,

$$\alpha(s) = e^{-(1/\lambda)s},$$

at $s = \mu$,

$$\alpha(\mu) = e^{-1/\rho}$$

and with this substitution into (20) we see that Z can vary from $\frac{1}{2}$ (when $\rho = \infty$) all the way up to unity (when $\rho = 0$). In contrast, $V = 0$, independent of ρ. We conclude that Z is not a desirable measure of variation for smooth streams. We shall see later that for peaked traffic, Z turns out to be a good measure. Incidentally, the above argument also shows that with exponential holding times the minimum possible value which peakedness can attain is $\frac{1}{2}$.

2.5 *Examples*

We now give two examples: a case with GI being smooth and another with GI being peaked.

Example 1: GI Smooth. Let $\{X_i\}$ be a sequence of independent, exponentially distributed random variables with corresponding means $\{\theta_i\}$. Let the interarrival time distribution of the GI stream be given by the

* A Poissonian stream ($V = 1$) is considered the norm to which the relative properties of smoothness and peakedness are compared.

distribution of the sum

$$Y = \sum_{i=1}^{n} X_i ,$$

where $n > 1$. Then the GI stream is smooth since

$$V = \frac{\sqrt{\theta_1^2 + \theta_2^2 + \cdots + \theta_n^2}}{\theta_1 + \theta_2 + \cdots + \theta_n} < 1. \tag{21}$$

In particular, if all the means are the same, then Y has a gamma distribution and GI is E_n , an Erlangian stream. For this case, we have

$$V = \frac{1}{\sqrt{n}} ,$$

which, moreover, is also the minimum value of (21). For our numerical example, we take $n = 2$.

From (10) we have

$$\alpha(z) = \left(\frac{2\nu}{z + 2\nu}\right)^2$$

and hence (9) becomes

$$p(u) = -\frac{4\nu^2}{u}\frac{d}{dz}\left\{\frac{u^2 p(u) - (1 - u)H(z)}{z - h(u)}\right\} ,$$

with the derivative being evaluated at $z = 2\nu$. This follows from (14). Carrying out the differentiation, collecting terms, and solving for $p(u)$, we obtain

$$p(u) = \frac{R(u)}{T(u)} , \tag{22}$$

where

$$R(u) = 4\sigma^2\{\lambda\mu H_1 u^2 - [(\lambda - \mu - 2\nu)H_1 - H_2]u + \mu H_1\},$$

$$T(u) = \rho^2 u^3 - (4\sigma^2 + \rho^2 + 2\rho + 4\rho\sigma)u^2 + (2\rho + 4\sigma + 1)u - 1,$$

$$\rho = \frac{\lambda}{\mu} ,$$

$$\sigma = \frac{\nu}{\mu} ,$$

and

$$H_1 = H'(2\nu), \quad H_2 = H(2\nu).$$

It is easy to verify the following inequalities:

$$T(0) < 0, \quad T(1) > 0, \quad T(1/\rho) < 0.$$

Since $\rho < 1$ and $T(u)$ is a polynomial of third degree, these inequalities imply that $T(u)$ has three positive roots with one root being less than unity. Hence, if u_1, u_2, u_3 are the roots of $T(u)$, then $0 < u_1 < 1 < u_2 < u_3$. Since $p(u)$ is a holomorphic function of u for $|u| \leqq 1$, the root u_1 must also be a root of $R(u)$. Using this requirement and the normalization condition $p(1) = 1$ we can find H_1 and H_2.

Omitting the intervening algebra we see that

$$p(u) = \frac{(1 - \rho u_1 u)K}{(1 - \omega_2 u)(1 - \omega_3 u)}, \tag{23}$$

where

$$K = \frac{(1 - \omega_2)(1 - \omega_3)}{(1 - \rho u_1)},$$

and

$$\omega_2 = \frac{1}{u_2}, \qquad \omega_3 = \frac{1}{u_3}.$$

Note that p_0, the probability of not being delayed at all, is given by K. Recall that for the Poissonian customers this quantity was given by $q_0 = 1 - \rho - \sigma$.

Expanding (23) in powers of u, we obtain

$$p(u) = \frac{K}{(\omega_2 - \omega_3)} \sum_{k=0}^{\infty} [(\omega_2 - \rho u_1)\omega_2^k - (\omega_3 - \rho u_1)\omega_3^k]u^k,$$

and hence, the distribution $\{p_j\}$ is given by

$$p_j = \frac{K}{(\omega_2 - \omega_3)} (\omega_2 - \rho u_1)\omega_2^j - (\omega_3 - \rho u_1)\omega_3^j, \qquad j = 0, 1, 2, \cdots.$$

The mean of this distribution can be computed by differentiation. We have

$$p'(1) = \frac{\omega_2}{1 - \omega_2} + \frac{\omega_3}{1 - \omega_3} - \frac{\rho u_1}{1 - \rho u_1}.$$

The corresponding quantities for the Poissonian stream of customers can easily be obtained using differentiation and relation (16).

If $D(t)$ is the delay distribution for the GI customers, then, as we already have seen, its Laplace–Stieltjes transform is given by

$$\delta(s) = p\left(\frac{\mu}{s+\mu}\right).$$

From (23) we have

$$\delta(s) = \frac{K(s+\mu)[s+(1-\rho u_1)\mu]}{[s+(1-\omega_2)\mu][s+(1-\omega_3)\mu]}.$$

Inverting this transform, we obtain the delay distribution for the GI customers:

$$D(t) = 1 - A_2 e^{-(1-\omega_2)\mu t} + A_3 e^{-(1-\omega_3)\mu t}, \tag{24}$$

where

$$A_2 = \frac{\omega_2(1-\omega_3)(\omega_2-\rho u_1)}{(1-\rho u_1)(\omega_2-\omega_3)}$$

$$A_3 = \frac{\omega_3(1-\omega_2)(\omega_3-\rho u_1)}{(1-\rho u_1)(\omega_2-\omega_3)}.$$

Using (19) we can obtain $E(t)$, the delay distribution for the Poissonian customers. Performing the indicated integration, we have

$$E(t) = 1 - \rho e^{-(1-\rho)\mu t} - \sigma\left[1 + \frac{A_2}{\rho-\omega_2} - \frac{A_3}{\rho-\omega_3}\right]e^{-(1-\rho)\mu t}$$
$$+ \frac{\sigma A_2}{\rho-\omega_2}e^{-(1-\omega_2)\mu t} - \frac{\sigma A_3}{\rho-\omega_3}e^{-(1-\omega_3)\mu t}. \tag{25}$$

Figure 1 shows complementary delay distributions for the GI customers and the Poissonian customers in the E_2 + M/M/1 model. Note that E_2 customers receive better service since they arrive in a smoother stream ($V = 1/\sqrt{2}$). For the Poissonian stream we, of course, have $V = 1$. While significant at low traffic intensities, this advantage diminishes as the traffic intensity increases.

Example 2: GI Peaked. We can generate a peaked traffic stream using the interrupted Poisson process.[5] For a given σ and Z, the interarrival time distribution of such a stream is given by Ref. 5.

$$A(\xi) = k_1(1 - e^{-r_1\xi}) + k_2(1 - e^{-r_2\xi}),$$

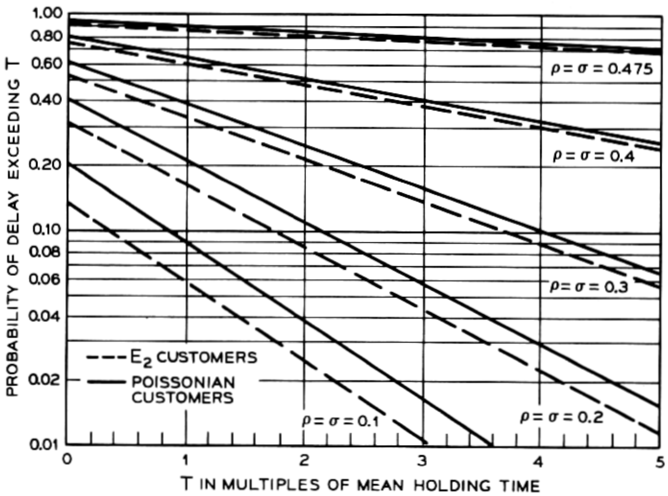

Fig. 1—Delay in the E_2 + M/M/1 queue.

where

$$r_1 = \tfrac{1}{2}\{\tau + \omega + \gamma + \sqrt{(\tau + \omega + \gamma)^2 - 4\tau\omega}\},$$
$$r_2 = \tfrac{1}{2}\{\tau + \omega + \gamma - \sqrt{(\tau + \omega + \gamma)^2 - 4\tau\omega}\},$$
$$k_1 = \frac{\tau - r_2}{r_1 - r_2},$$
$$k_2 = 1 - k_1,$$

and

$$\tau = \sigma Z + 3Z(Z - 1),$$
$$\omega = \frac{\sigma}{\tau}\left[\frac{\tau - \sigma}{Z - 1} - 1\right],$$
$$\gamma = \left(\frac{\tau}{\sigma} - 1\right)\omega.$$

The Laplace–Stieltjes transform of $A(\xi)$ is given by

$$\alpha(s) = \frac{k_1 r_1}{s + r_1} + \frac{k_2 r_2}{s + r_2}.$$

Carrying out the same steps as in the previous example, we see that the probability generating function of $\{p_i\}$ has the same form, that is,

$$p(u) = \frac{(1 - \omega_2)(1 - \omega_3)(1 - \rho u_1 u)}{(1 - \omega_2 u)(1 - \omega_3 u)(1 - \rho u_1)},$$

where

$$\omega_2 = \frac{1}{u_2}, \qquad \omega_3 = \frac{1}{u_3},$$

and u_1, u_2, u_3 are now the roots of the following polynomial:

$$\rho(\rho + k_1 r_1 + k_2 r_2)u^3 - [r_1 r_2 + k_1 r_1 + k_2 r_2 + (r_1 + r_2)\rho + 2\rho + \rho^2]u_2 + (r_1 + r_2 + 2\rho + 1)u - 1.$$

Hence, the delay distributions for the GI customers and the Poissonian customers have the same forms as those in Example 1, but with different values of u_1, ω_2, and ω_3.

Figure 2 shows the same information as Fig. 1 of the previous example. Note that the peaked stream with $Z = 3.0$ ($V = 2.3$) receives poorer service than the Poissonian stream. Again, this effect diminishes at higher traffic intensities.

Figure 3 shows the effect of peakedness and smoothness on the quality of service received by all customers. Holding the total traffic intensity constant ($\rho + \sigma = 0.8$) and comparing the results with the case when all the customers arrive in a Poissonian stream (GI = M),

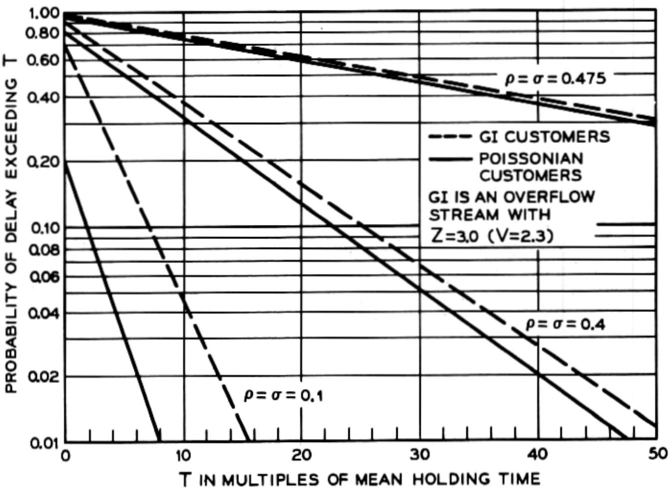

Fig. 2—Delay in the GI + M/M/1 queue when GI is an overflow stream.

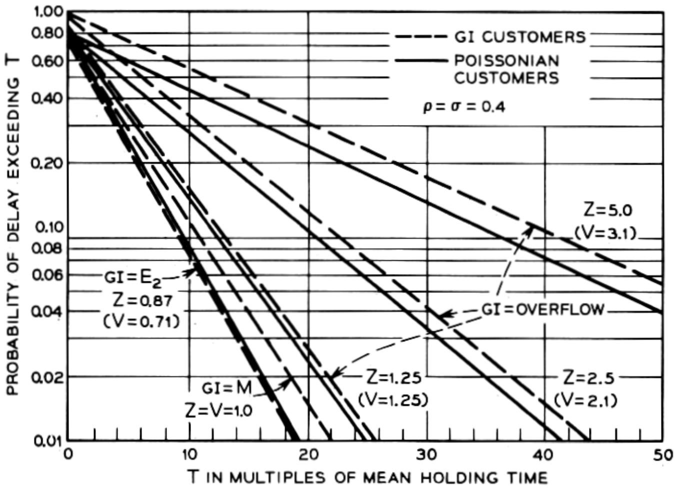

Fig. 3—The effect of peakedness on delay in the GI + M/M/1 queue.

we see that the service provided to *all* the customers either improves or deteriorates according to whether the GI stream is smooth or peaked.

Figure 4 shows the effect of mixing the traffic. Note that as the proportion of peaked traffic in the total offered traffic stream increases, the service deteriorates for everyone.

We now clarify a remark made earlier about the suitability of the peakedness factor Z for measuring variability properties of a stream. It was found numerically that the relation between Z and V in a peaked stream is monotone and nearly independent of the stream's intensity. In fact, by direct computation from $A(\xi)$, we have the two relations

$$Z = \frac{1 + V^2}{2} - \frac{V^2 - 1}{3(1 + V^2) + 2\sigma}$$

and

$$V = \left\{1 + \frac{2(Z - 1)}{1 - 1/(3Z + \sigma)}\right\}^{\frac{1}{2}}.$$

Figure 5 shows this relation for two streams, one with traffic intensity of 1 erlang and the other with 100 erlangs. Since the inverse relation is also monotone and nearly independent of the traffic intensity, the peakedness seems to be a suitable measure of the variational properties of a peaked stream.

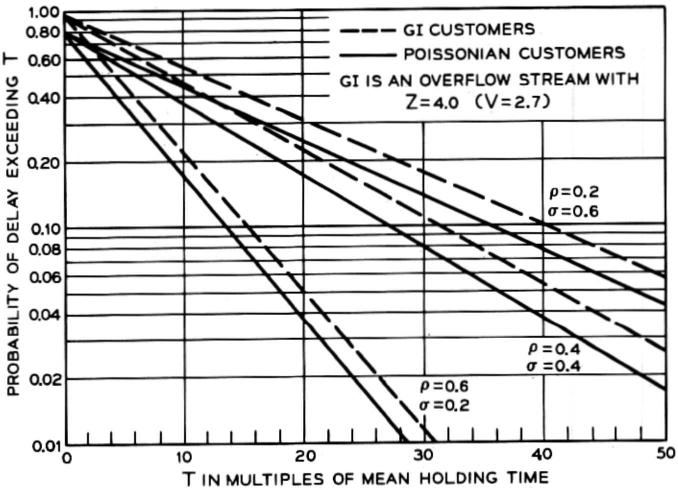

Fig. 4—Effect of mixture of input on delay in the GI + M/M/1 queue.

III. THE GI + M/M/∞ SYSTEM

Let $Y(t)$ be the number of customers in the system at time t. Since there are an infinite number of servers, $Y(t)$ is equal to the number of busy servers at time t. Again, we see that $\{Y(t),\ t \geqq 0\}$ is a piecewise Markov process. The regeneration points are the arrival epochs of the GI customers; the distribution of the length of Markovian segments is given by $A(\xi)$, the interarrival time distribution of the GI customers; and the regeneration matrix is given by (1). This is the same identification as that made in the GI + M/M/1 queue. The difference here is the Markov process operating within the segment.

The Markovian development of the process within the segments is governed by a birth-death process with birth rate λ, the arrival intensity of the Poissonian stream, and death rate $j\mu$ when the process is in state j, where μ^{-1} is the mean service time. The transition functions of this process are given by the transient solution to the M/M/∞ system (Ref. 2, page 24):

$$P_{ij}(t) = \sum_{k=0}^{j} \binom{i}{j-k} \frac{\rho^k}{k!} e^{-\rho g(t)} [1 - g(t)]^{i-k} g^{i-j+2k}(t), \tag{26}$$

where

$$g(t) = 1 - e^{-\mu t}, \qquad \rho = \frac{\lambda}{\mu}.$$

Let $\{p_i\}$ be the stationary distribution of the Markov chain imbedded at points immediately preceding a GI arrival. This distribution satisfies the Chapman–Kolmogorov eqs. (3). If $p(u)$ is the probability generating function of $\{p_i\}$, then from (4) we have

$$p(u) = \int_0^\infty \sum_{i=0}^\infty p_i G_{i+1}(u, \xi)\, dA(\xi),$$

where

$$G_i(u, \xi) = \sum_{j=0}^\infty P_{ij}(\xi) u^j.$$

But from Ref. 2 we have

$$G_i(u, \xi) = (1 - e^{-\mu\xi} + ue^{-\mu\xi})^i e^{-\rho(1-u)g(\xi)},$$

and we see that $p(u)$ satisfies the following integral equation:

$$p(u) = e^{-\rho(1-u)} \int_0^\infty r(u, \xi) p[r(u, \xi)] e^{-\rho(1-u)e^{-\mu\xi}}\, dA(\xi), \tag{27}$$

where

$$r(u, \xi) = 1 - e^{-\mu\xi} + ue^{-\mu\xi}.$$

Now let X_1 be the number of Poissonian customers and X_2 the number of GI customers in the system seen by an arbitrary GI arrival. Since the number of servers is infinite, X_1 and X_2 are independent. If $b(u)$ is the probability generating function of the distribution of X_2, then

$$p(u) = e^{-\rho(1-u)} b(u). \tag{28}$$

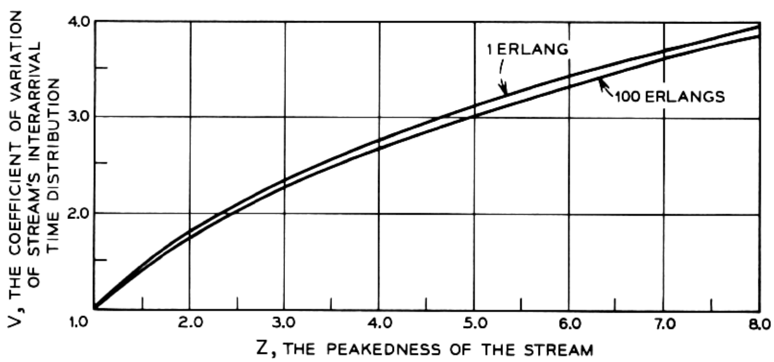

Fig. 5—Relation between V and Z in an overflow stream for two different means.

Since the arriving GI customers are sampling a Markov process according to a renewal process, they will see the stationary distribution of the process[1] and, hence, X_1 has the Poisson distribution

$$P\{X_1 = k\} = e^{-\rho} \frac{\rho^k}{k!}, \tag{29}$$

the stationary distribution of the number of customers in an M/M/∞ system.

Substituting (28) into (27) we obtain an equation for $b(u)$:

$$b(u) = \int_0^\infty r(u, \xi) b[r(u, \xi)] \, dA(\xi). \tag{30}$$

In solving for $b(u)$, it is convenient to consider the expansion of $b(u)$ about $u = 1$, rather than about $u = 0$. Hence, we set

$$b(u) = \sum_{j=0}^{\infty} d_j (u - 1)^j,$$

where

$$d_j = \frac{b^{(j)}(1)}{j!}.$$

If

$$b(u) = \sum_{j=0}^{\infty} b_j u^j$$

is the expansion of $b(u)$ about the origin, then the two sets of coefficients are mutually related through the equations

$$d_j = \sum_{k=j}^{\infty} \binom{k}{j} b_k$$

$$b_j = \sum_{k=j}^{\infty} \binom{k}{j} (-1)^{k-j} d_k \,. \tag{31}$$

Differentiating (30) j times and setting $u = 1$, we get

$$d_j = d_j \alpha_j + d_{j-1} \alpha_j \,, \qquad j = 1, 2, \cdots,$$

$$d_0 = 1,$$

where

$$\alpha_j = \alpha(j\mu) = \int_0^\infty e^{-j\mu\xi} \, dA(\xi).$$

Solving for d_j, we obtain

$$d_j = \prod_{i=1}^{j} \frac{\alpha_i}{1 - \alpha_i}, \qquad j = 1, 2, \cdots .$$

Thus, we have

$$\begin{aligned} p(u) &= e^{-\rho(1-u)} \sum_{j=0}^{\infty} d_j(u - 1)^j \\ &= e^{-\rho(1-u)} \sum_{j=0}^{\infty} b_j u^j, \end{aligned}$$

and the distribution $\{p_j\}$ is given by the convolution of $\{b_j\}$ with (29):

$$p_j = e^{-\rho} \sum_{k=0}^{j} b_{j-k} \frac{\rho^k}{k!}.$$

Now

$$q_j = \lim_{t \to \infty} P\{Y(t) = j\}, \qquad j = 0, 1, 2, \cdots ,$$

the distribution of the number of busy servers seen by an arbitrary Poissonian arrival, can be found using the rate conservation principle. If ν^{-1} is the mean of $A(\xi)$ and $\sigma = \nu/\mu$, then equating the asymptotic rate of transition out of the set of states $\{0, 1, 2, \cdots , j\}$ to the rate into that set we get

$$\sigma p_j + \rho q_j = (j + 1)q_{j+1}, \quad j = 0, 1, 2, \cdots . \tag{32}$$

If $q(u)$ is the probability generating function of $\{q_j\}$, then it follows from (32) that $q(u)$ satisfies the differential equation

$$q'(u) = \rho q(u) + \sigma p(u).$$

This equation has the general solution

$$q(u) = e^{\rho u} \sigma \int e^{-\rho \xi} p(\xi) \, d\xi.$$

Substituting for $p(\xi)$, carrying out the integration and using the normalization condition $q(1) = 1$ to determine the integration constant, we obtain

$$q(u) = e^{-\rho(1-u)} \left\{ 1 - \sigma \sum_{j=0}^{\infty} \frac{b_j}{j + 1} + \sigma \sum_{j=0}^{\infty} b_j \frac{u^{j+1}}{j + 1} \right\}.$$

Since $q(u)$ is the product of two probability generating functions, the distribution $\{q_j\}$ can now easily be determined by convolution. Another way of obtaining $\{q_j\}$ is by solving (32), i.e.,

$$q_j = \frac{\rho^j}{j!}\left\{q_0 + \sigma \sum_{i=0}^{j-1} \frac{i!\, p_i}{\rho^{i+1}}\right\}, \qquad j = 1, 2, \cdots,$$

$$q_0 = e^{-\rho}\left\{1 - \sigma \sum_{j=0}^{\infty} \frac{\rho^j}{j!} \sum_{i=0}^{j-1} \frac{i!\, p_i}{\rho^{i+1}}\right\}.$$

The means of the two distributions are given by

$$\sum_{j=1}^{\infty} jq_j = q'(1) = \rho + \sigma$$

$$\sum_{j=1}^{\infty} jp_j = p'(1) = \rho + \frac{\alpha(\mu)}{1 - \alpha(\mu)}.$$

We make two observations. First, the mean of $\{q_j\}$ is independent of the form of the interarrival time distribution of the GI customers; it depends only on ν^{-1}, the mean interarrival time. Second, using Jensen's inequality one can show that the mean of $\{p_j\}$ is minimized whenever the interarrival times of the GI customers are constant. For this special case we have

$$p'(1) = \rho + \frac{1}{e^{1/\sigma} - 1},$$

$$q'(1) = \rho + \sigma,$$

and hence

$$p'(1) < q'(1)$$

since

$$1 + \frac{1}{\sigma} < e^{1/\sigma}$$

for $\sigma > 0$.

REFERENCES

1. Kuczura, A., "Piecewise Markov Processes," to appear in SIAM J. Appl. Math.
2. Prabhu, N. V., *Queues and Inventories*, New York: John Wiley & Sons, 1965.
3. Takaćs, L., *Introduction to the Theory of Queues*, New York: Oxford University Press, 1962.
4. Feller, W., *An Introduction to Probability Theory and its Applications*, Vol. 2, New York: John Wiley & Sons, 1966.
5. Kuczura, A., "Loss Systems with Mixed Renewal and Poisson Inputs," unpublished work.