P. E. Boudreau
J. S. Griffin, Jr.
M. Kac*

A Discrete Queueing Problem with Variable Service Times

Abstract: Methods from the theory of Markov chains are used to analyze a simple single-server queueing model. The model is of the sort that arises naturally in information-handling contexts, in that a discrete time basis is used, which matches the cyclic character of processors. Considerable generality is attained, in that no appeal is made to the exponential or other conventional forms for the probability distributions governing the number of arrivals per cycle and the service times.

The principal object of study is the queue length; the stationary distribution governing this quantity is calculated, along with various associated averages. The relation between the present method and the more usual continuous-variable method is illustrated by the derivation of some of the classical equations from a limiting case of our model.

Introduction

Consider a system in which transactions occur from time to time and are stored in a buffer, from which they are taken serially for processing. The time required to process a given transaction will be variable; furthermore, we will suppose the processor to operate on a cyclic (or discrete) basis, so that the time required to process a given transaction will be represented by a positive integer, namely the number of cycles during which the processor is occupied with that transaction. In this paper we will study the lengths of the queues that may be expected to form in the buffer.

A special case of this problem, namely that in which it is assumed that every transaction requires exactly two cycles for processing, has already been reported.[1] Also it may be mentioned that part of the present paper was presented to the American Mathematical Society.[2]

Since the literature in queueing theory includes several excellent books and other surveys, we need not comment at any length on the relation between this and other work. However, one aspect of this relation is shown in Sections 4.A-4.C, where some of the simple classical results are derived. Furthermore, we would like to direct attention to a recent paper by R. G. Miller[3] in which group arrivals are allowed, as they are here.

* The Rockefeller Institute.

1. The principal properties of the model

• *A. Formalization of the model*

We are concerned, then, with a system consisting of two parts: a buffer which accepts and stores transactions which originate externally, and a processor which from time to time takes a transaction from the buffer and processes it. We suppose that the processor is cyclic or discrete in character, in the sense that it will only accept a transaction for processing at the beginning of a cycle, and that it always requires an integral number of cycles to process a transaction. We suppose that only one transaction can be processed at a time, and thus it is convenient to say that at the beginning of a cycle the processor may be in either of two states, namely either available (if a transaction is not being processed) or occupied (if a transaction is being processed). Each transaction is to require a variable number of cycles for processing: let p_n be the probability that exactly n cycles are required to process a given transaction, where $p_1, p_2, \cdots, p_n, \cdots$ are to be any given probability distribution of the positive integers. During any cycle new transactions may be stored in the buffer: we let ρ_n be the probability that exactly n new transactions appear during any particular cycle, where $\rho_0, \rho_1, \cdots, \rho_n, \cdots$ is a given probability distribution of the non-negative integers.

We suppose that the system works according to the following rules:

1) If at the beginning of a certain cycle, say the k^{th}, there are transactions in the buffer, and the processor will be in the available state at the beginning of the $(k+1)^{st}$ cycle, then at the beginning of the $(k+1)^{st}$ cycle a transaction is transferred from the buffer to the processor. If this transaction requires n cycles for processing, then the processor will be in the occupied state at the beginning of the $(k+2)^{nd}$, $(k+3)^{rd}$, $\cdots$, $(k+n)^{th}$ cycles, and it will again be available at the beginning of the $(k+n+1)^{st}$.
2) If at the beginning of the k^{th} cycle there are no transactions waiting in the buffer, then no transaction will be moved to the processor at the beginning of the $(k+1)^{st}$ cycle, even if a new one should appear during the k^{th} cycle.
3) No transactions will be moved to the processor at the beginning of a cycle if the processor is occupied at the beginning of that cycle.

Thus, the processor can take a transaction from the buffer only at the beginning of a cycle, and then only if it is not still occupied with a prior transaction and only if there was a transaction already in the buffer at the beginning of the previous cycle.

The status of the system at the beginning of the k^{th} cycle may be specified quantitatively by the pair (M_k, E_k), where M_k is the number of transactions waiting in the buffer at the beginning of the k^{th} cycle, and E_k will be defined so as to describe the state of the processor at the beginning of the k^{th} cycle. Let X_k be the number of new transactions which occur during the k^{th} cycle; we assume that $X_0, X_1, \cdots, X_k, \cdots$ are identically distributed independent random variables and that their common distribution is written $\rho_0, \rho_1, \cdots, \rho_n \cdots$. This is the assumption that with probability ρ_n there occur exactly n new transactions during the k^{th} cycle. It is evident that

$$M_{k+1} = X_k + M_k - \xi_{k+1},$$

where ξ_{k+1} is to be 1 or 0 according as a transaction is or is not taken out of the buffer for processing at the beginning of the $k+1^{st}$ cycle. As to the availability of the processor, it is convenient to think of a counter being set as a transaction is transferred to the processor, the entry in this counter being the number of cycles required to process that transaction. The contents of the counter is to be diminished by 1 just before the beginning of each subsequent cycle, until it reads 1; thus at the beginning of a cycle a reading of the counter other than 1 is always the number of cycles, including the one then beginning, required to complete the processing of the current transaction. If the contents of the counter is 1 at the beginning of any cycle, then the processor will be in the available state at the beginning of the next cycle; and if not, then it will be in the occupied state. We let E_k be the contents of the counter at the beginning of the k^{th} cycle; these remarks may then be translated as follows:

If $E_k > 1$ then $E_{k+1} = E_k - 1$,
if $E_k = 1$ and $M_k = 0$ then $E_{k+1} = 1$, and
if $E_k = 1$ and $M_k > 0$ then $E_{k+1} = n$ with probability p_n.

Now we can define ξ_{k+1} to be 1 if $E_k = 1$ and $M_k > 0$, and to be 0 otherwise; thus strictly speaking ξ could be dispensed with, and (M_{k+1}, E_{k+1}) has been specified in terms of (M_k, E_k) and the distributions ρ and p.

Thus we perceive that the sequence (M_k, E_k), $k = 0, 1, 2, \cdots$ constitutes a Markov chain. Indeed, our interest is confined to the random variable M_k, and we introduce E_k precisely because the study of the pair (M_k, E_k) is simpler than the study of M_k alone; and the reason for this gain in simplicity is essentially that the two together constitute a Markov chain, while M_k by itself does not.

It is perhaps worthwhile to note that while we will borrow the language of the theory of Markov chains, our main argument is self-contained, in that it depends only on certain elementary theorems from analysis. Nowhere do we make use of a deep fact about Markov chains.

• *B. The probabilities of occurrence of the various states*

Let us denote by $P(\nu, \eta|\zeta, \xi)$ the probability that if $(M_k, E_k) = (\zeta, \xi)$, then $(M_{k+1}, E_{k+1}) = (\nu, \eta)$. The above remarks may be interpreted in this notation as follows, where we have adopted the convention that $\rho_\nu = 0$ for negative values of ν:

If $\xi > 1$ then

$$P(\nu, \eta|\zeta, \xi) = 0 \text{ if } \eta \neq \xi - 1,$$

and

$$P(\nu, \xi - 1|\zeta, \xi) = \rho_{\nu-\zeta} \text{ if } \eta = \xi - 1;$$

if $\xi = 1$ then

$$P(\nu, \eta|\zeta, 1) = \rho_{\nu-\zeta+1} p_\eta \text{ if } \zeta > 0,$$

and

$$P(\nu, 1|0, 1) = \rho_\nu .$$

These are the transition probabilities for our Markov chain.

Let $P^{(k)}(\nu, \eta)$ be the probability of the occurrence of the state (ν, η) on the k^{th} cycle, i.e.,

$$P^{(k)}(\nu, \eta) = \text{Prob}\{(M_k, E_k) = (\nu, \eta)\} .$$

Evidently there is the inductive relation

$$P^{(k+1)}(\nu, \eta) = \sum_{\zeta=0}^{\infty} \sum_{\xi=1}^{\infty} P(\nu, \eta|\zeta, \xi) P^{(k)}(\zeta, \xi),$$

so that if one knows $P^{(0)}(\nu, \eta)$ for all ν and η then in principle one may calculate $P^{(k)}(\nu, \eta)$ for all k, ν, and η. On making use of the values of the transition probabilities listed above, we find

$$P^{(k+1)}(v, \eta) = \sum_{\zeta=0}^{v} \rho_{v-\zeta} P^{(k)}(\zeta, \eta + 1)$$

$$+ \sum_{\zeta=1}^{v+1} p_\eta \rho_{v-\zeta+1} P^{(k)}(\zeta, 1) + \delta_{1\eta}\rho_v P^{(k)}(0, 1) ,$$

where $\delta_{1\eta}$ is the traditional Kronecker delta, i.e., the last term appears only when $\eta = 1$.

Now define

$$\pi_k(v) = \sum_{\eta=1}^{\infty} P^{(k)}(v, \eta) ;$$

since $P^{(k)}(v, \eta)$, is the probability that $M_k = v$ and $E_k = \eta$, we see that $\pi_k(v)$ is the probability that $M_k = v$, regardless of the values of E_k. Thus, $\pi_k(v)$ is the probability that there are exactly v transactions in the buffer on the k^{th} cycle. Our objective now becomes the determination of properties of the sequence of distributions π_k. A precise statement of our results appears in Sections 1.D and 1.E, but first we discuss a hypothesis which underlies all our arguments.

• C. *The load on the processor*

Let λ and μ be the means of the distributions ρ and p respectively:

$$\lambda = \sum_{v=0}^{\infty} v\rho_v$$

and

$$\mu = \sum_{\eta=1}^{\infty} \eta p_\eta ,$$

so that λ is the average number of new transactions occurring during a cycle, and μ is the average number of cycles required to process a transaction. *We shall assume throughout the present paper that* $\lambda\mu < 1$.

Intuitively, this is very reasonable. During the first N cycles, approximately λN transactions should occur; on the average the process time for a transaction is about μ cycles, so that if N is large and $\lambda\mu < 1$ the processor should be working about $\lambda\mu N$ of the first N cycles, and idle about $N - \lambda\mu N$ of the first N cycles. Thus the processor should be working with probability about

$$(1/N)\cdot\lambda\mu N = \lambda\mu$$

and idle with probability about

$$(1/N)\cdot(N - \lambda\mu N) = 1 - \lambda\mu .$$

A little more formally, we could say that the processor is working during the k^{th} cycle (of any one experiment) if $E_{k-1} > 1$ or else if $E_{k-1} = 1$ and $M_{k-1} > 0$, and is idle otherwise, i.e., if $E_{k-1} = 1$ and $M_{k-1} = 0$. It will appear during the course of our main argument that if $\lambda\mu < 1$ then

$$\lim_{N\to\infty} (1/N) \sum_{k=1}^{N} P^{(k-1)}(0, 1) = 1 - \lambda\mu .$$

On the other hand, if $\lambda\mu$ were to exceed 1, then after N cycles about λN transactions should have occurred, whereas only about N/μ should have been processed, so that about

$$\lambda N - \frac{1}{\mu} N = \frac{\lambda\mu - 1}{\mu} N$$

should be in the buffer. Thus M_k should grow approximately linearly with k. This also could be formalized, but seems to be clear enough as it stands.

Thus $\lambda\mu$ represents in some sense a load on the processor, and $1 - \lambda\mu$ represents an excess capacity; our hypothesis is that there be a positive excess capacity.

• D. *The average number of transactions in the buffer*

Let

$$m_k = \sum_{v=0}^{\infty} v\pi_k(v) ;$$

then m_k is the expectation of M_k, or the average number of transactions in the buffer on the k^{th} cycle. The variance of M_k, to be written v_k, is defined to be the expectation of $(M_k - m_k)^2$, or equivalently,

$$v_k = \sum_{v=0}^{\infty} v^2\pi_k(v) - m_k^{\ 2} .$$

On the face of it, of course, m_k and v_k depend on the initial distribution $P^{(0)}$, i.e., on the number of transactions in the buffer at the beginning of the process, even though this may only be given stochastically. Thus, if initially there were a large number of transactions in the buffer then at least during the early cycles m_k should be larger than if say there had been only a few. On the other hand, it would seem that in the long run the effect of the initial distribution should wear off; and indeed it does. We define the *average number of transactions in the buffer* to be

$$\lim_{N\to\infty} (1/N) \sum_{k=0}^{N} m_k ;$$

we will prove that, at any rate if the initial distribution is finite in a reasonable sense, and if $1 - \lambda\mu > 0$, then this limit exists, is independent of the initial distribution $P^{(0)}$, and in fact is equal to

$$\lambda + \frac{1}{2}\frac{1}{1 - \lambda\mu}(\lambda^2\mathrm{M} + \Lambda\mu) ,$$

where λ and μ are defined as above, and Λ and M are similar higher moments:

$$\Lambda = \sum_{v=0}^{\infty} v(v - 1)\rho_v$$

and

$$\mathrm{M} = \sum_{\eta=1}^{\infty} n(\eta - 1)p_\eta .$$

We will also prove under a stronger but still very general hypothesis that the means m_k themselves converge to this limit.

Somewhat analagous considerations hold for the variances. If

$$\mathscr{L} = \sum_{\nu=0}^{\infty} \nu(\nu-1)(\nu-2)\rho_\nu$$

and

$$\mathscr{M} = \sum_{\eta=1}^{\infty} \eta(\eta-1)(\eta-2)p_\eta$$

then

$$\lim_{N\to\infty} (1/N) \sum_{k=0}^{N} v_k$$

exists and is equal to

$$\Lambda + \lambda - \lambda^2 + \Delta^2 + \Delta$$
$$+ \frac{1}{3}\frac{1}{1-\lambda\mu}(\lambda^3\mathscr{M} + 3\lambda\Lambda\mathrm{M} + \mathscr{L}\mu),$$

where

$$\Delta = \frac{1}{2}\frac{1}{1-\lambda\mu}(\lambda^2\mathrm{M} + \Lambda\mu),$$

provided, however, that the average of $m_k{}^2$ is the same as the square of the average of m_k (which will be true, for example, if m_k converges). The same condition that insures the convergence of m_k insures also the convergence of v_k.

• *E. The limiting distribution*

It is known from the general theory of Markov chains that there will exist a *stationary* or *limiting* distribution, that is, there will be a set of quantities $\bar{P}(\nu, \eta)$ which is stationary in the sense that if $P^{(0)}(\nu, \eta) = \bar{P}(\nu, \eta)$ then also $P^{(k)}(\nu, \eta) = \bar{P}(\nu, \eta)$ for all k, and is limiting in the sense that, regardless of $P^{(0)}(\nu, \eta)$,

$$\lim_{k\to\infty} P^{(k)}(\nu, \eta) = \bar{P}(\nu, \eta)$$

for all ν and η. The interested reader may consult, for example, the book of Feller[4], or that of Chung[5]; however, we need make no explicit use of this theorem. Instead we will simply exhibit a certain distribution $\bar{P}$ and prove that it is stationary, and then, under a weak restriction on the distributions $\bar{P}$, p, and ρ we will prove directly the convergence.

If we let

$$\bar{\pi} = \sum_{\eta=1}^{\infty} \bar{P}(\nu, \eta),$$

it obviously must turn out that $\bar{\pi}$ is a limiting distribution for the sequence of random variables M_k. As it happens $\bar{\pi}$ is given by a very simple generating function, namely

$$\sum_{\nu=0}^{\infty} \bar{\pi}(\nu)z^\nu = (1-\lambda\mu)f(z)\frac{z-1}{z-g[f(z)]},$$

where f and g are the generating functions for the distributions ρ and p respectively:

$$f(z) = \sum_{\nu=0}^{\infty} \rho_\nu z^\nu \quad \text{and} \quad g(z) = \sum_{\eta=1}^{\infty} p_\eta z^\eta .$$

Naturally the mean and variance of the distribution $\bar{\pi}$ agree with the averages of m_k and v_k listed in Section 1.D.

2. The computation of the various averages

• *A. The generating functions*

We have already defined f to be the generating function of the distribution ρ: that is, we set

$$f(z) = \sum_{k=1}^{\infty} \rho_k z^k .$$

This series converges at least for $|z| < 1$, and of course defines an analytic function on this region. Similarly, we define, for each k and η, $F_\eta{}^{(k)}$ to be the generating function determined by $P^{(k)}(\nu, \eta)$ with ν running:

$$F_\eta{}^{(k)}(z) = \sum_{\nu=0}^{\infty} P^{(k)}(\nu, \eta)z^\nu .$$

We also define

$$F^{(k)}(z) = \sum_{\eta=1}^{\infty} F_\eta{}^{(k)}(z)$$

so that $F^{(k)}$ is the generating function of the distribution which governs M_k:

$$F^{(k)}(z) = \sum_{\nu=0}^{\infty} \pi^{(k)}(\nu)z^\nu .$$

Now the expectation m_k of M_k can be recovered from $F^{(k)}$: since

$$F^{(k)\prime}(z) = \sum_{\nu=0}^{\infty} \nu\pi^{(k)}(\nu)z^{\nu-1},$$

it follows that

$$m_k = F^{(k)\prime}(1).$$

Our first object will be to calculate the average number of transactions in the buffer as already defined, namely

$$L_1 = \lim_{N\to\infty} (1/N) \sum_{k=0}^{N} m_k . \tag{1}$$

We will do this by means of the following device. We form the function

$$\alpha(w) = \sum_{k=0}^{\infty} m_k w^k$$

for $0 \leqq w < 1$; then it is a fact of classical analysis that since each m_k is non-negative

$$L_1 = \lim_{w \to 1^-} (1-w)\alpha(w), \tag{2}$$

the existence of the limit (1) which defines L_1 being guaranteed by the existence of the limit (2). This is a theorem of the Tauberian type; a complete elementary discussion appears in Hobson's book on real variables,[6] and there is also a treatment in Widder's book.[7]

This interest in the function α leads us to consider the function

$$\Gamma(z, w) = \sum_{k=0}^{\infty} F^{(k)}(z) w^k,$$

because α can be recovered from Γ: evidently

$$\alpha(w) = \left.\frac{\partial \Gamma}{\partial z}\right|_{1,w}.$$

The evaluation of the limit L_1 of Eq. (1) will now consist in finding a more or less closed form for the function Γ and then computing

$$\lim_{w \to 1^-} (1-w) \left.\frac{\partial \Gamma}{\partial z}\right|_{1,w}.$$

B. A closed form for Γ

We saw in Section 1.B that the probability distribution $P^{(k+1)}$ is specified explicitly in terms of the distribution $P^{(k)}$; it is then more or less obvious that the functions $F^{(k+1)}$ should be expressible in terms of the functions $F^{(k)}$. In fact, as is easy to deduce,

$$zF_\eta^{(k+1)} = zfF_{\eta+1}^{(k)} + p_\eta f[F_1^{(k)} - P^{(k)}(0,1)] + \delta_{1\eta} zfP^{(k)}(0,1), \tag{3}$$

where here and hereafter we suppress unnecessary mention of z and w when they occur as arguments of the various functions which appear.

Define

$$G_\eta(z, w) = \sum_{k=0}^{\infty} F_\eta^{(k)}(z) w^k,$$

so that

$$\Gamma(z, w) = \sum_{\eta=1}^{\infty} G_\eta(z, w).$$

Then the recurrence relation (3) implies immediately that

$$zG_\eta = zwfG_{\eta+1} + p_\eta wf[G_1 - A(w)] + zF_\eta^{(0)} + \delta_{1\eta} zwfA(w), \tag{4}$$

where we have set

$$A(w) = \sum_{k=0}^{\infty} P^{(k)}(0,1) w^k.$$

The relations among the various functions G_η are surprisingly simple, and on making use of the fact that $\lim_{\eta \to \infty} G_\eta = 0$ one finds that they may be solved explicitly for each of the functions G_η in terms of the function A and f, the distribution p, and the initial distribution $P^{(0)}$ which appear as coefficients of the functions $F^{(0)}$. Specifically, on multiplying Eq. (4) by $(wf)^\eta$ and then summing over values of η between 1 and N, one finds that

$$zG_1 = z(wf)^N G_{N+1} + \sum_{\eta=1}^{N+1} p_\eta (wf)^\eta (G_1 - A) + zwfA + z \sum_{\eta=1}^{N+1} (wf)^{\eta-1} F^{(0)}. \tag{5}$$

Therefore on setting

$$\sigma(z, w) = \sum_{\eta=1}^{\infty} p_\eta [wf(z)]^\eta$$

and

$$\Phi(z, w) = \sum_{\eta=1}^{\infty} F^{(0)}(z)[wf(z)]^{\eta-1},$$

it appears that

$$(z - \sigma)G_1 = (zwf - \sigma)A + z\Phi$$

or

$$G_1 = \frac{(zwf - \sigma)A + z\Phi}{z - \sigma}.$$

Having obtained G_1, one could of course now write each G_η in terms of A and Φ. However, it is more to our immediate purpose to note simply that on summing Eq. (4) over all values of η we obtain

$$z(1 - wf) \sum_{\eta=1}^{\infty} G_\eta = wf(1 - z)(G_1 - A) + z \sum_{\eta=1}^{\infty} F_\eta^{(0)}$$

from which it follows that

$$\Gamma = \frac{wf(1-z)}{(1-wf)(z-\sigma)} \Phi - \frac{wf(1-z)}{z-\sigma} A + \frac{1}{1-wf} \Theta,$$

where we have set

$$\Theta(z) = \sum_{\eta=1}^{\infty} F_\eta^{(0)}(z).$$

Now in principle we know the functions Θ and Φ, in that they depend only on the distributions ρ and $P^{(0)}$, so that further understanding of Γ depends only on further understanding of the function A. It is easy to see that for each value of w satisfying $0 \leqq w < 1$, the function $G_1(z, w)$ is analytic in z for $|z| < 1$; but $G_1(z, w)$ has been expressed as a fraction

$$G_1(z, w) = \frac{[zwf - \sigma(z, w)]A(w) + z\Phi(z, w)}{z - \sigma(z, w)},$$

and it follows that the numerator of this fraction must vanish whenever the denominator does, at least inside the unit disk. From this observation can be deduced all the information we need about $A(w)$.

Indeed, set

$$\phi_w(z) = z - \sigma(z, w) .$$

If z is real, so is $\phi_w(z)$; it is easy to see that

$$\lim_{z\to 1} \phi_w(z) = 1 - \sum_{k=1}^{\infty} p_k w^k > 0 .$$

On the other hand

$$\phi_w(0) = -\sum_{k=1}^{\infty} p_k(w\rho_0)^k < 0 ,$$

since the hypothesis $1 - \lambda\mu > 0$ precludes the possibility that ρ_0 vanishes. Therefore, ϕ_w has at least one real zero in the unit interval. By Rouché's theorem ϕ_w can vanish but once on the open unit disk, since z vanishes but once there and on the unit circle we have

$$|z| > |\sigma(z, w)| .$$

Therefore there is for each w between 0 and 1 a unique root $\theta(w)$ in the unit disk of the equation

$$\phi_w(z) = 0$$

and this root is real.

If we replace z with $\theta(w)$ in the numerator of the expression for G_1, the result must vanish identically. Thus it follows that

$$A(w) = \frac{\Phi[\theta(w), w]}{1 - wf[\theta(w)]} ,$$

where we have made use of the fact that $\sigma[\theta(w), w]$ may be replaced with $\theta(w)$. Thus knowledge of the function $\Gamma(z, w)$ now turns on knowledge of $\theta(w)$.

• C. *The behavior of the root $\theta(w)$*

The study of $\theta(w)$ is considerably simplified by the fact that our interest is confined to values of w near 1. Differentiation of the relation defining θ, namely

$$\theta(w) - \sum_{k=1}^{\infty} p_k\{wf[\theta(w)]\}^k = 0$$

reveals that, since $1 - \lambda\mu > 0$, $\theta'(w)$ is positive. It follows that θ increases monotonely, and therefore that $\lim_{w\to 1^-} \theta(w)$ exists, say

$$\lim_{w\to 1^-} \theta(w) = a .$$

Define for z restricted to the unit interval

$$h(z) = z - \sum_{k=1}^{\infty} p_k f(z)^k ;$$

then

$$h'(z) = 1 - \sum_{k=1}^{\infty} kp_k f(z)^{k-1} \cdot f'(z)$$

so that

$$h'(z) > 1 - \lambda\mu > 0 .$$

Furthermore it is evident that $h(0) < 0$ (since again $\rho_0 \neq 0$) and $h(1) = 0$; thus h increases monotonely to zero as z varies along the unit interval. On the other hand ϕ_w converges uniformly to h as w approaches 1 from the left, and therefore $h(a) = 0$. Therefore

$$\lim_{w\to 1^-} \theta(w) = 1 .$$

It follows that

$$\lim_{w\to 1^-} \theta'(w) = \frac{\mu}{1 - \lambda\mu} ,$$

and, on differentiating twice the relation defining θ, that

$$\lim_{w\to 1^-} \theta''(w) = \frac{1}{(1 - \lambda\mu)^3} [\mu^2\lambda + M + 2\lambda\mu^2(1 - \lambda\mu)] .$$

These remarks contain sufficient information about $A(w)$ to enable us to evaluate the limit L_1.

• D. *The average number of transactions in the buffer*

One now computes without difficulty that

$$(1 - w) \frac{\partial\Gamma}{\partial z}\bigg|_{1,w} = \frac{d\Theta}{dz}\bigg|_1 + \frac{w}{1 - \sigma(1, w)} \left\langle \left\{\frac{1 - w}{1 - wf[\theta(w)]}\right\} \Phi[\theta(w), w] - \Phi(1, w) + \lambda \frac{1 - \sigma(1, w)}{1 - w} \right\rangle .$$

The quantity in angular brackets approaches zero as w approaches 1 from the left, and therefore l'Hospital's rule applies. One thereby verifies that

$$\lim_{w\to 1^-} \left\{\frac{1 - w}{1 - wf[\theta(w)]}\right\}' = -\lambda\mu - \frac{1}{2}\frac{1}{1 - \lambda\mu} \{\lambda \mathrm{M} + \mu^2\Lambda\}$$

and then uses this together with the fact that

$$\frac{d\Theta}{dz}\bigg|_1 + \lambda \frac{\partial\Phi}{\partial w}\bigg|_{1,1} = \frac{\partial\Phi}{\partial z}\bigg|_{1,1}$$

to calculate

$$\lim_{w\to 1^-} (1 - w) \frac{\partial\Gamma}{\partial z}\bigg|_{1,w} = \lambda + \frac{1}{2}\frac{1}{1 - \lambda\mu} (\lambda^2 \mathrm{M} + \mu\Lambda) .$$

This establishes Eq. (1) of Section 2.A.

• E. *The average of the variances*

Suppose we let

$$u_k = \sum_{v=0}^{\infty} v^2 \pi^{(k)}(v) ,$$

so that

$$v_k = u_k - m_k^2 ,$$

where v_k is the variance of the distribution π_k, as defined in Section 1.D. Let

$$L_2 = \lim_{N\to\infty} (1/N) \sum_{k=0}^{N} u_k .$$

Then it is evident

$$u_k = F^{(k)''}(1) + F^{(k)'}(1) ,$$

and therefore we may obtain L_2 from Γ by the equation

$$L_2 = \lim_{w\to 1^-} (1-w) \left[\frac{\partial^2\Gamma}{\partial z^2}\bigg|_{1,w} + \frac{\partial\Gamma}{\partial z}\bigg|_{1,w} \right] .$$

By a procedure which is strictly analagous to that just used to evaluate L_1, but whose complexity is remarkable, one may evaluate this limit; one finds

$$L_2 = \Lambda + \lambda + \frac{1}{3}\frac{1}{1-\lambda\mu}(\lambda^3\mathscr{M} + 3\lambda\Lambda\mathrm{M} + \mu\mathscr{L})$$
$$+ 2\Delta^2 + (2\lambda+1)\Delta ,$$

where

$$\Delta = \frac{1}{2}\frac{1}{1-\lambda\mu}(\lambda^2\mathrm{M} + \mu\Lambda) .$$

It follows

$$L_2 - L_1{}^2 = \Lambda + \lambda - \lambda^2$$
$$+ \frac{1}{3}\frac{1}{1-\lambda\mu}(\lambda^3\mathscr{M} + 3\lambda\Lambda\mathrm{M} + \mu\mathscr{L}) + \Delta^2 + \Delta ;$$

we would like, of course, to identify $L_2 - L_1{}^2$ with the average of the variances, i.e., to assert

$$\lim_{N\to\infty} (1/N) \sum_{k=1}^{N} v_k = L_2 - L_1{}^2 ,$$

but this is incorrect unless

$$L_1{}^2 = \lim_{N\to\infty} (1/N) \sum_{k=1}^{N} m_k{}^2 .$$

This last statement is, however, correct if

$$L_1 = \lim_{k\to\infty} m_k ;$$

and this we will prove in Section 3.B for somewhat restricted ρ, p, and $P^{(0)}$.

• *F. The stationary distribution*

If for any reason one suspected that there might exist a stationary distribution (e.g. because one knew the theorem mentioned in Section 1.E), then one might very naturally suspect that it would be $\bar{P}$, where $\bar{P}(\nu, \eta)$ is defined by

$$\bar{P}(\nu,\eta) = \lim_{N\to\infty} (1/N) \sum_{k=0}^{N} P^{(k)}(\nu,\eta) .$$

We do not have the quantities $P^{(k)}(\nu, \eta)$ available except in the guise of their generating functions, but this is no obstacle: we put

$$\bar{F}_\eta(z) = \lim_{N\to\infty} (1/N) \sum_{k=0}^{N} F^{(k)}(z) ,$$

and it should be true that

$$\bar{F}_\eta(z) = \lim_{w\to 1^-} (1-w) G_\eta(z, w) .$$

Let us calculate this last limit. Returning to Eq. (5) of Section 2.B, we have

$$G_\eta = \frac{\Phi - (1-wf)A}{z-\sigma} \sum_{j=1}^{\infty} p_{j+\eta-1}(wf)^j$$
$$+ \sum_{j=1}^{\infty} (wf)^{j-1} F_{j+\eta-1}{}^{(0)} + \delta_{1\eta} wfA .$$

Therefore using the fact that

$$\lim_{w\to 1^-} (1-w)A(w) = 1 - \lambda\mu ,$$

we find

$$\lim_{w\to 1^-} (1-w)G_\eta(z,w)$$
$$= (1-\lambda\mu)\left[-\frac{1-f}{z-\sigma(z,1)} \sum_{j=1}^{\infty} p_{j+\eta-1} f^j + \delta_{1\eta} f \right] .$$

Thus we are led to conjecture that if we put $\bar{F}_\eta(z)$ equal to the right member of this equation and define $\bar{P}(\nu, \eta)$ in terms of $\bar{F}_\eta(z)$, i.e., if we let

$$\bar{F}_\eta(z) = \sum_{\nu=0}^{\infty} \bar{P}(\nu,\eta) z^\nu ,$$

then the distribution $\bar{P}$ should be stationary.

But now the proof that $\bar{P}$ so defined is in fact stationary is trivial. In Eq. (3) of Section 2.B we let $F^{(k)} = \bar{F}_\eta$ for each value of η; a little rearrangement reveals that it follows that also $F^{(k+1)} = \bar{F}_\eta$ for all η. Since the functions $\bar{F}_\eta$ are unchanged from cycle to cycle, it follows that the distribution $\bar{P}$ is also unchanged from cycle to cycle, i.e., is stationary.

Finally if as in Section 1.E we let

$$\bar{\pi}(\nu) = \sum_{\eta=1}^{\infty} \bar{P}(\nu,\eta) ,$$

it is easy to calculate the generating function for the distribution $\bar{\pi}$; we find

$$\sum_{\eta=1}^{\infty} \bar{F}_\eta(z) = (1-\lambda\mu) f(z) \frac{z-1}{z-\sigma(z,1)} .$$

On noting that if g is the generating function for the distribution p then

$$\sigma(z,1) = g[f(z)] ,$$

we see that we have the generating function for $\bar{\pi}$ announced in Section 1.E.

3. The question of convergence

• *A. The analyticity of the function $\theta(w)$*

In Section 2.D we proved that

$$\lim_{N\to\infty}(1/N)\sum_{k=0}^{N} m_k$$

exists, but we left open the question of whether or not the sequence m_k itself converged. Similarly, in Section 2.F we came very close to proving that

$$\lim_{N\to\infty}(1/N)\sum_{k=0}^{N} P^{(k)}(\nu,\eta)$$

exists, although it turned out to be unnecessary to examine this particular question there. Now we will prove that *if the distributions $P^{(0)}$, ρ, and p are well behaved, in the sense that there exists a real number r greater than 1 such that each of the series*

$$\sum_{\nu=0}^{\infty}\sum_{\eta=1}^{\infty} P^{(0)}(\nu,\eta)r^{\nu+\eta}$$

$$\sum_{\nu=0}^{\infty}\rho_\nu r^\nu$$

and

$$\sum_{\eta=1}^{\infty} p_\eta r^\eta$$

converges, then

$$\lim_{k\to\infty} m_k = L_1$$

$$\lim_{k\to\infty} v_k = L_2 - L_1{}^2$$

and

$$\lim_{k\to\infty} P^{(k)}(\nu,\eta) = \bar{P}(\nu,\eta)\,.$$

We should perhaps note that we have no reason to believe that any of these limits fail to exist, even if some of the series diverge for all r greater than 1. Indeed, we have already noted in Section 1.E that the third of these equations is true more generally; and results obtained by Kiefer and Wolfowitz[8,9] in a somewhat analogous situation lead one to suspect that the first two also are true more generally. On the other hand for our purposes this loss in generality is not serious, and the argument we give below does recommend itself by virtue of its simplicity.

The use of this added hypothesis will be to insure that certain functions which arise are analytic on regions larger than the unit disk, and hence that certain Maclaurin expansions converge at 1. To be somewhat more explicit, we will show, for example,

$$\sum_{k=0}^{\infty}(m_k - L_1)w^k$$

is analytic on a disk of the w-plane which is larger than the unit disk, and hence that the series converges for $w = 1$, and hence that $m_k - L_1$ converges to zero.

We begin by considering the function θ. In Section 2.B, $\theta(w)$ was defined for $0 \leqq w < 1$; but we shall see that $\theta(w)$ could just as well have been defined for w complex and $|w| < 1$, and that the above hypothesis permits the extension of θ to a neighborhood of the unit circle. Remembering that $\theta(w)$ was to satisfy the equation

$$z - \sigma(z, w) = 0\,, \tag{6}$$

where

$$\sigma(z, w) = \sum_{\eta=1}^{\infty} p_\eta[wf(z)]^\eta\,,$$

we see if $|w| < 1$, then for z on the unit circle we have

$$|z| > |\sigma(z, w)|\,,$$

and therefore, again by Rouché's theorem, for each such w the function $z - \sigma(z, w)$ vanishes at exactly one point inside the unit circle; define $\theta(w)$ to be that point, so that θ is defined throughout the open unit disk of the w-plane, and takes values in the open unit disk of the z-plane. That θ is analytic for these values of w follows from the implicit function theorem, for the hypotheses of that theorem are satisfied, in that if we set

$$\sigma(z, w) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_{ij}z^i w^j\,, \tag{7}$$

then the series on the right certainly converges for $|z| < 1$ and $|w| < 1$; and

$$\frac{\partial}{\partial z}[z - \sigma(z, w)] = 1 - \sum_{\eta=1}^{\infty}\eta p_\eta[wf(z)]^{\eta-1}\cdot wf'(z)\,,$$

where

$$\left|\sum_{\eta=1}^{\infty}\eta p_\eta[wf(z)]^{\eta-1}\cdot wf'(z)\right| \leqq \lambda\mu < 1$$

for $|z| < 1$ and $|w| < 1$, so that this partial derivative does not vanish in the region in question.

Now since $\sum_{\nu=0}^{\infty}\rho_\nu z^\nu$ and $\sum_{\eta=1}^{\infty} p_\eta z^\eta$ converge for $|z| < r$, where $r > 1$, and since if $|z| < 1$ then also

$$\left|\sum_{\nu=0}^{\infty}\rho_\nu z^\nu\right| < 1\,,$$

it follows that there is a positive number ε such that the series (7) actually converges for all z and w satisfying $|z| < 1 + \varepsilon$ and $|w| < 1 + \varepsilon$; thus it appears likely that θ can be extended to a slightly larger region of the w-plane. Let w_0 be a point on the unit circle; we ask what value is to be assigned to $\theta(w_0)$. Now $\theta(tw_0)$ is defined for $0 \leqq t < 1$, and lies within the closed unit disk; thus there is at least one accumulation point, say a, for $\theta(tw_0)$ as t approaches 1 from the left. By a continuity argument it follows that $a - \sigma(a, w_0) = 0$;

now invoking the implicit function theorem again, we find that there is an open disk U_0 with center w_0 throughout which there is defined an analytic function θ_0 which is unique subject to the conditions

$$\theta_0(w) - \sigma[\theta_0(w), w] = 0$$

and

$$\theta_0(w_0) = a \ .$$

It is easy to see that either $|a| < 1$, or else $a = 1$, and that the latter can occur only if $w_0{}^{\eta} = 1$ whenever $p_\eta \neq 0$. In the case $|a| < 1$, the radius of U_0 may be chosen so small that the values of θ_0 all lie within the unit circle of the z-plane; it then follows from the uniqueness of the solution to Eq. (6) subject to the conditions $|z| < 1$ and $|w| < 1$ that θ and θ_0 agree at points common to U_0 and the unit disk of the w-plane, and thus that θ_0 provides a genuine extension of θ. In the case $a = 1$ and $w_0 = 1$, for t real and less than but near 1, $\theta_0(t)$ lies inside the unit disk, so that θ and θ_0 must agree at least for these values of t; but this suffices to insure agreement throughout their common domain. In the case $a = 1$ but $w_0 \neq 1$, there is an integer k such that $w_0{}^k = 1$ and $p_\eta = 0$ unless k divides η. But in this case if u is any k^{th} root of unity and

$$z - \sigma(z, w) = 0$$

then also

$$z - \sigma(z, uw) = 0 \ ,$$

so that it follows, at least for $|w| < 1$, that $\theta(uw) = \theta(w)$; thus in this instance the possibility of extending θ to a neighborhood of w_0 may be inferred from the possibility of extending θ to a neighborhood of 1.

Therefore θ is analytic throughout the open unit disk, and has no singular point on the unit circle; hence θ may be extended analytically to some open disk with center at the origin and radius greater than 1. Furthermore, by contracting this disk slightly if necessary, we can insure that θ takes the value 1 at no point of this disk except 1 and the k^{th} roots of unity, where k is the greatest common divisor of those η for which $p_\eta \neq 0$.

• *B. The convergence of the means* m_k

It is now easy to see that, in the presence of the hypotheses stated at the beginning of Section 3.A,

$$\lim_{k\to\infty} m_k = L_1 \ .$$

Recall that

$$\Gamma(z, w) = \sum_{k=0}^{\infty} \sum_{\nu=0}^{\infty} \pi_k(\nu) z^\nu w^k \ ,$$

and therefore

$$\left.\frac{\partial\Gamma}{\partial z}\right|_{1,w} = \sum_{k=0}^{\infty} m_k w^k \ ;$$

hence at least for $|w| < 1$

$$\left.\frac{\partial\Gamma}{\partial z}\right|_{1,w} - \frac{L_1}{1-w} = \sum_{k=0}^{\infty} (m_k - L_1) w^k \ . \qquad (8)$$

We will see that the function on the left is analytic on a disk with center at the origin and radius greater than 1, and hence that the radius of convergence of the series on the right exceeds 1. It will follow that this series converges in particular at $w = 1$, and hence that its general term tends to zero:

$$\lim_{k\to\infty} (m_k - L_1) = 0 \ .$$

To establish the required analyticity, we note that

$$\left.\frac{\partial\Gamma}{\partial z}\right|_{1,w} - \frac{L_1}{1-w} = \frac{1}{1-w} \times \left[\left.\frac{d\Theta}{dz}\right|_1 + \frac{w}{1-\sigma(1,w)}\Psi(w) - L_1\right] , \qquad (9)$$

where

$$\Psi(w) = \frac{1-w}{1-wf[\theta(w)]}\Phi[\theta(w), w] - \Phi(1, w) + \lambda\frac{1-\sigma(1,w)}{1-w} \ . \qquad (10)$$

Now in order that the ratio of two analytic functions, say $\alpha(w)/\beta(w)$, be analytic at zero of the denominator, say w_0, it is sufficient that $\alpha(w_0) = 0$ but that $\beta'(w_0) \neq 0$; the application of this principle in turn to each of the fractions appearing on the right of Eqs. (9) and (10) reveals first that Ψ and then that

$$\left.\frac{\partial\Gamma}{\partial z}\right|_{1,w} - \frac{L_1}{1-w}$$

is analytic on a neighborhood of the closed unit disk. Thus the radius of convergence of the series on the right of Eq. (8) must exceed 1 and the convergence of the sequence m_k to L_1 is established.

This same argument applied to the expression for

$$\left.\frac{\partial^2\Gamma}{\partial z^2}\right|_{1,w}$$

establishes

$$\lim_{k\to\infty} \sum_{\nu=0}^{\infty} \nu^2 \pi_k(\nu) = L_2 \ ;$$

since we now have

$$\lim_{k\to\infty} m_k{}^2 = L_1{}^2 \ ,$$

it follows

$$\lim_{k\to\infty} v_k = L_2 - L_1{}^2 \ .$$

There seems to be no reason to give any of the details.

C. *Convergence of the distributions* $P^{(k)}$

An argument similar to that just used suffices to establish the convergence of the probabilities of occurrence of any particular state on successive cycles; however the situation is now complicated by the fact that we have to deal with the generating functions of the quantities involved rather than with a simple sequence of numbers. We shall confine ourselves to showing how to circumvent this particular difficulty, and for simplicity we will treat the distributions π_k which govern the random variables M_k rather than the distributions $P^{(k)}$. If we let

$$\phi(z) = (1 - \lambda\mu)f(z)\frac{z-1}{z-\sigma(z,1)}$$

and define $\bar{\pi}$ as the sequence of coefficients of the Maclaurin expansion of ϕ, namely

$$\phi(z) = \sum_{v=0}^{\infty} \bar{\pi}(v)z^v,$$

then it is our intention to show

$$\lim_{k\to\infty} \pi_k(v) = \bar{\pi}(v)$$

for each v.

We have

$$\Gamma(z,w) - \frac{1}{1-w}\phi(z) = \Delta(z,w),$$

where

$$\Delta(z,w) = \frac{wf(z)(z-1)}{z-\sigma(z,w)}\Psi(z,w) + \frac{\Theta(z)-\bar{\Theta}(z)}{1-wf(z)}$$

and

$$\Psi(z,w) = \frac{\Phi[\theta(w),w] - \bar{\Phi}[\theta(w),w]}{1-wf[\theta(w)]} - \frac{\Phi(z,w)-\bar{\Phi}(z,w)}{1-wf(z)};$$

a horizontal bar over a function here means that the function refers to the initial distribution being chosen to be the stationary distribution. One proves that there are two disks D and E in the z and w planes respectively, each with center at the origin and with radius exceeding 1, such that for each z of D the function $\Delta(z,w)$ is analytic in w at each point w of E: this is as before a matter of examining $\Delta(z,w)$ of each value of w for which a quantity in a denominator vanishes; since z is fixed, only derivatives with respect to w become involved. One must also check that $\bar{\Phi}(z,w)$ is analytic in w for each choice of z. Now evidently

$$\Delta(z,w) = \sum_{k=0}^{\infty} [F^{(k)}(z) - \phi(z)]w^k,$$

and therefore for each z and D this series converges in particular at $w = 1$; we have proved that the sequence of functions $F^{(k)}$ converges pointwise in D to ϕ.

To conclude that the coefficients $\pi_k(v)$ of $F^{(k)}$ converge to those of ϕ, namely $\bar{\pi}(v)$, it is necessary to see that the convergence of $F^{(k)}$ to ϕ is in fact uniform. But this is immediate, in that if E has radius r', where $r' > 1$, then

$$\limsup_{n\to\infty} |F^{(k)}(z) - \phi(z)|^{1/k} = \frac{1}{r'},$$

and it follows that there is a number $s > 1$ and an integer N such that if $k > N$ then

$$|F^{(k)}(z) - \phi(z)| < s^{-k}$$

for all $z \varepsilon D$.

Since D is larger than the unit disk, it follows

$$\phi'(1) = \lim_{k\to\infty} F^{(k)\prime}(1),$$

and it turns out that $\phi'(1) = L_1$; this would again establish the convergence of m_k to L_1. Similarly

$$\phi''(1) = \lim_{k\to\infty} F^{(k)\prime\prime}(1),$$

from which one may conclude that v_k converges to $L_2 - L_1{}^2$.

4. Relation to the continuous case

A. *Discrete approximations*

Our equations may be used to derive some of the classical theorems. To this end we consider now a continuous process, in that we suppose arrivals of transactions may occur at any instant, and that the process time is represented by a real-valued rather than an integer-valued random variable. We will approximate this system by a discrete system, invoke facts established above, and pass to an appropriate limit.

We will restrict our attention to the case of Poisson inputs, and we will take the average interarrival time to be τ. Intuitively, this means that the arrivals of transactions for processing are independent of one another, and in a time interval of length t about t/τ of them may be expected to arrive; precisely, this means that during a time interval of length t the probability that exactly n transactions arrive for processing is to be $(1/n!)(t/\tau)^n e^{-t/\tau}$. It will become apparent that this restriction is essential, because the approximation procedure to be used turns on the independence of events in successive arbitrarily small cycles.

The process time will be taken to be governed by a given distribution ϕ, that is, the probability that the time required to process a particular transaction does not exceed t is to be exactly $\phi(t)$. Thus ϕ must be such a function that $\phi(0) = 0$, ϕ is monotone nondecreasing, and $\lim_{t\to\infty} \phi(t) = 1$. Furthermore, we will assume that ϕ is differentiable, although this hypothesis could be relaxed. The probability that the process time for a given transaction lies between t and $t + \delta$ is $\phi(t + \delta) - \phi(t)$, and taking ϕ to be differentiable merely enables us to replace this quantity with $\phi'(\bar{t})\cdot\delta$ for a suitable choice of $\bar{t}$ between t and $t + \delta$.

Thus we have in mind the Markov process described by the pair (M_t, E_t), where t ranges over the non-negative real numbers; for each t, M_t and E_t are random variables with values in respectively the non-negative integers and the non-negative real numbers; M_t increases in unit steps according to a Poisson process, and decreases by one unit at any instant t for which $M_t > 0$ and $E_t = 0$; and E_t increases by a random amount governed by the distribution ϕ whenever $M_t > 0$ and $E_t = 0$, and decreases linearly with unit slope whenever it is positive.

The N^{th} discrete approximation is constructed as follows. We let a cycle have length $(1/N)$, so that cycles begin at times $0, (1/N), (2/N), \cdots$, and it follows

$$\rho_v^N = \frac{1}{v!}\left(\frac{1}{N\tau}\right)^v e^{-(1/N\tau)},$$

and therefore

$$f_N(z) = \sum_{v=0}^{\infty} \rho_v^N z^v = e^{(z-1)/N\tau}.$$

It follows

$$\lambda_N = f_N'(1) = \frac{1}{N\tau};$$

intuitively this means that on the average a transaction arrives for processing once every $N\tau$ cycles. The distribution governing the process time is taken to be

$$p_k^N = \phi(k/N) - \phi(k - 1/N);$$

it follows

$$\mu_N = \sum_{k=1}^{\infty} k p_k^N = \sum_{k=1}^{\infty} k\cdot\phi'(t_k^N)\cdot\frac{1}{N}$$

for suitably chosen points $t_1^N, t_2^N, \cdots$. It follows that

$$\lim_{N\to\infty} \lambda_N\mu_N = \lim_{N\to} \frac{1}{\tau}\sum_{k=1}^{\infty}\frac{k}{N}\phi'(t_k^N)\cdot\frac{1}{N} = \frac{1}{\tau}\int_0^\infty t\phi'(t)dt;$$

this integral may of course be identified as the average process time. The requirement $1 - \lambda\mu > 0$ imposed above must be satisfied here; that is, for sufficiently large N it must be true that $1 - \lambda_N\mu_N > 0$; this will evidently be so if

$$1 - \frac{1}{\tau}\int_0^\infty t\phi'(t)dt > 0, \quad \text{i.e., if} \quad \tau > \int_0^\infty t\phi'(t)dt.$$

Intuitively, this simply says that the average inter-arrival time should exceed the average process time.

B. *The average number of transactions waiting*

It is evident

$$\Lambda_N = f_N''(1) = \frac{1}{\tau^2 N^2}$$

and that

$$M_N = \sum_{k=1}^{\infty} k(k-1)p_k^N = \sum_{k=1}^{\infty} k(k-1)\phi'(t_k^N)\cdot\frac{1}{N};$$

therefore in the N^{th} approximation the average number of transactions waiting in the buffer is

$$L_1^N = \lambda_N + \frac{1}{2}\frac{1}{1-\lambda_N\mu_N}(\lambda_N^2 M_N + \mu_N\Lambda_N),$$

and therefore

$$\lim_{N\to\infty} L_1^N = \frac{1}{2}\left[1 - \frac{1}{\tau}\int_0^\infty t\phi'(t)dt\right]^{-1}\cdot\frac{1}{\tau^2}\int_0^\infty t^2\phi'(t)dt.$$

Similar considerations apply to the second moment.

C. *The stationary distribution*

We have the fact that

$$\sigma(z, 1) = \sum_{k=1}^{\infty} p_k f(z)^k;$$

therefore in the N^{th} approximation

$$\sigma_N(z, 1) = \sum_{k=1}^{\infty}\left[\phi(k/N) - \phi\left(\frac{k-1}{N}\right)\right]e^{k(z-1)/N\tau},$$

and it follows

$$\lim_{N\to\infty}\sigma_N(z, 1) = \int_0^\infty \phi'(t)e^{t(z-1)/\tau}dt.$$

Therefore the limit of the stationary distributions governing the number of words in the buffer in the various approximations has the generating function

$$\lim_{N\to\infty}(1 - \lambda_N\mu_N)\frac{z-1}{z - \sigma_N(z, 1)} = \left[1 - \frac{1}{\tau}\int_0^\infty t\phi'(t)dt\right]\left[z - 1\right] \times \left[z - \int_0^\infty \phi'(t)e^{t(z-1)/\tau}dt\right].$$

For the particular case of exponential service times, one chooses $\phi(t) = 1 - e^{-at}$, and on evaluating and re-arranging the various integrals one finds that in this case the above expression reduces to

$$\left(1 - \frac{1}{a\tau}\right)\left(1 + \frac{1}{a\tau - z}\right).$$

These results are of course well known. The general result was given by Kninchine in 1932, and is discussed at length in Riordan's recent book[10]; the particular case of exponential service times appears in Feller's book.[11] In order to see that our results indeed agree with those cited, it is necessary to note that our state 0 corresponds to their states 0 and 1 combined: intuitively, "the buffer is empty" may mean that there is no demand present ("the processor is also empty") or else there is one demand present ("a transaction is in the processor").

Finally we know that the stationary probability that the counter reads η and there are exactly v transactions in the buffer is the coefficient of z in the expression

$$(1-\lambda\mu)\left[\frac{1-f(z)}{z-\sigma(z,1)}\sum_{k=1}^{\infty}p_{k+\eta-1}f(z)^k+\delta_{\eta 1}f(z)\right].$$

Therefore we take as the N^{th} approximation to the stationary probability that $E_t \leqq \mu$ and there are v transactions in the buffer the coefficient of z^v in the expression

$$(1-\lambda_N\mu_N)\sum_{\eta=1}^{u_0}\frac{1-f_N(z)}{z-\sigma_N(z,1)}\sum_{k=1}^{\infty}p_{k+\eta-1}f_N(z)^k,$$

where u_0 is the largest integer which does not exceed u_N. In the same spirit as above we find the limit of this expression as N becomes large: after some rearrangement it turns out to be

$$\left[z-\phi(u)-\int_0^{\infty}\phi'(t+u)e^{t(z-1)/\tau}dt\right]$$
$$\times\left[1-\left(\frac{1}{\tau}\int_0^{\infty}t\phi'(t)dt\right)^{-1}\right]\left[z-\int_0^{\infty}\phi'(t)e^{t(z-1)/\gamma}dt\right]^{-1}.$$

For the particular case of exponential service times this is

$$\left(1-\frac{1}{a\tau}\right)\left[1+(1-e^{-au})\frac{1}{a\tau-z}\right].$$

References

1. P. E. Boudreau and M. Kac, *IBM Journal*, **5**, 132 (1961).
2. M. Kac, P. E. Boudreau, and J. S. Griffin, Jr., *Notices of the American Mathematical Society*, Abstract No. 61T-122.
3. R. G. Miller, Jr., " A Contribution to the Theory of Bulk Queues," *Journal of the Royal Statistical Society*, **B**, **21**, 320 (1959).
4. W. Feller, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, New York, 1958, 2nd ed., Vol. I.
5. K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, Springer, Berlin, 1960.
6. E. W. Hobson, *The Theory of Functions of a Real Variable*, Cambridge University Press, 1926, 2nd ed., Vol. II.
7. D. V. Widder, *The Laplace Transform*, Princeton University Press, 1946.
8. J. Kiefer and J. Wolfowitz, *Transactions of the American Mathematical Society*, **78**, 1 (1955).
9. J. Kiefer and J. Wolfowitz, *Annals of Mathematical Statistics*, **27**, 147 (1956).
10. J. Riordan, *Stochastic Service Systems*, John Wiley and Sons, New York, 1962.
11. W. Feller, op. cit., p. 415.

Received April 17, 1962