Multiple Channel Queues in Heavy Traffic. II: Sequences, Networks, and Batches

Donald L. Iglehart; Ward Whitt

Advances in Applied Probability, Vol. 2, No. 2 (Autumn, 1970), 355-369.

Stable URL:
http://links.jstor.org/sici?sici=0001-8678%28197023%292%3A2%3C355%3AMCQIHT%3E2.0.CO%3B2-W

Advances in Applied Probability is currently published by Applied Probability Trust.

Adv. Appl. Prob. **2**, 355–369 (1970)
Printed in Israel

MULTIPLE CHANNEL QUEUES IN HEAVY TRAFFIC. II: SEQUENCES, NETWORKS, AND BATCHES

DONALD L. IGLEHART[1], *Stanford University*
WARD WHITT[2], *Yale University*

1. Introduction

This paper is a sequel to [7], in which heavy traffic limit theorems were proved for various stochastic processes arising in a single queueing facility with r arrival channels and s service channels. Here we prove similar theorems for sequences of such queueing facilities. The same heavy traffic behavior prevails in many cases in this more general setting, but new heavy traffic behavior is observed when the sequence of traffic intensities associated with the sequence of queueing facilities approaches the critical value ($\rho = 1$) at appropriate rates.

As a second feature, we do not require that the basic sequences of interarrival times and potential service times associated with each queueing facility contain independent or identically distributed random variables. Thus we can regard our facility as just one facility within a complex network; some of the input streams may be outputs from other facilities. We also consider arrivals and service in batches. We assume the reader is familiar with our previous paper [7], which in addition to queueing notation means the weak convergence theory associated with the function space $D[0,1]$ as it appears in Billingsley (1968).

We assume each queueing system in the sequence of queueing systems has the same structure, that is, the same network of facilities with the same channels at each facility, but in general the sequences of random variables (interarrival times and potential service times) which are the basic data vary from queueing system to queueing system. Although each queueing system may be a complex network of facilities, we restrict our attention for the most part to a single

Received 28 July 1969.
[1] Research supported by Office of Naval Research Contract N00014-67-A-0112-0031 and National Science Foundation Grant GP-8790.
[2] Research supported by Office of Naval Research Contract Nonr-225(53) while visiting Stanford University in the Department of Operations Research during the 1968–1969 academic year.

facility within that network. It is possible to obtain limit theorems jointly for several characteristics associated with several different facilities within a network, but the limits are usually quite complicated. We determine such joint limits in Section 4, but we do not evaluate the distributions of the limits in detail.

For a specified facility in the nth system with r arrival channels and s service channels, let λ_i^n denote the arrival rate in the ith arrival channel and μ_j^n the maximum possible service rate in the jth service channel. We assume that these rates do not depend on time or the state of the system. These rates are typically defined as the reciprocal of the mean interarrival time and the reciprocal of the mean service time. Somewhat more generally, they can be defined as the constants in the translation terms of appropriate random functions in $D[0,1]$ necessary for weak convergence. In any case, $\lambda^n = \sum_{i=1}^r \lambda_i^n$ is the total arrival rate to the facility and $\mu^n = \sum_{j=1}^s \mu_j^n$ is the maximum total service rate of the facility. As a measure of congestion for the facility in question in the nth system, we define the traffic intensity $\rho^n = \lambda^n/\mu^n$.

While in [7] we investigated single multiple-channel queueing facilities in which $\rho \geqq 1$, now we investigate sequences of such facilities (perhaps within a network) when the associated sequence of traffic intensities $\{\rho_n\}$ approaches a limit greater than or equal to one. (The limiting behavior actually depends on $\lambda^n - \mu^n$ rather than λ^n/μ^n, but these are equivalent as long as $\lambda^n \to \lambda$ and $\mu^n \to \mu$, $0 < \lambda, \mu < \infty$.) When $\rho_n \to \rho > 1$ as $n \to \infty$, the limiting behavior is the same as for a single system with $\rho > 1$. When $\rho_n \to 1$ as $n \to \infty$, there are different cases depending upon the rate $\rho_n \to 1$ as $n \to \infty$. Under these various conditions, which we shall refer to as the various cases of heavy traffic, the systems are of course unstable (a proof of this fact is an easy by-product of our results). Our objective is to obtain functional central limit theorems (invariance principles) for the stochastic processes characterizing these systems after appropriately scaling and translating the processes. The entire discussion is carried out in the context of weak convergence of probability measures on the function space $D[0,1]$ (cf. [1]). These results provide useful descriptions of unstable queues and, perhaps, useful approximations of stable queues. Promising in this last regard is the limit obtained when ρ_n approaches one rather slowly from below.

As in [7], we investigate two different models: the standard system and the modified system. In both models the $r + s$ channels associated with the specified facility are independent. In particular, this prohibits cyclic behavior in a network, that is, the output of the facility cannot become part of the input. Furthermore, arriving customers form a single queue and are served in the order of their arrival without defections. The two models differ in their modes of operation for the service channels.

In the standard system a waiting customer is assigned to the first available service channel and the servers (servers ≡ service channels) are shut off when they are idle. Thus the classical $GI/G/s$ system is a special case of our standard system. In the modified system a waiting customer is assigned to the service channel that can complete his service first and the servers are not shut off when they are idle. For further discussion, see [2] or [7]. While the modified system is of some interest in its own right, we have introduced it, following Borovkov (1965), primarily as an analytical tool. Heavy traffic theorems are much easier to prove for the modified system than the standard system. Corresponding theorems are obtained for the standard system by defining the two systems in terms of the same basic data and showing the difference between the respective queueing processes is negligible in heavy traffic (cf. [7], Theorem 3.1). The difference approaches 0 as $n \to \infty$ for all values of ρ as long as we normalize by $n^{\frac{1}{2}}$. Moreover, the argument used in [7] applies to the more general situation discussed in this paper. Therefore, we shall only discuss the modified system and invoke the extension of Theorem 3.1 of [7] in order to obtain similar results for the standard system.

This paper is organized as follows. In Section 2 we obtain functional central limit theorems for the queue length process and the departure process under all cases of heavy traffic when the arrival and potential service processes have weak limits. Section 3 is devoted to a number of examples for which the hypotheses of Section 2 hold. Joint limits for several facilities in a network are discussed in Section 4. Finally, in Section 5 a few remarks are made about future research in the area of heavy traffic.

Limit theorems for sequences of queueing systems in heavy traffic were first proved by Kingman ((1961), (1965)). However, Prohorov (1963) was the first to consider all possible cases. Borovkov (1965) extended Prohorov's results to multiple channel queues with batches and Whitt (1968) extended Prohorov's results to weak convergence in $D[0, 1]$. This paper extends Borovkov's (1965) results to networks of multiple channel queues with batches and weak convergence in $D[0, 1]$. Also more processes are considered here and the proofs are simplified by applying the weak convergence theory associated with $D[0, 1]$. For further background and discussion, see [16]. In addition to [7] and [16], recent related work appears in Borovkov (1967), Gaver (1968), Iglehart and Kennedy (1970), and Prabhu (1970).

2. A sequence of modified queueing systems

We repeat that we shall only investigate sequences of modified queueing systems because exactly the same argument as we used in Theorem 3.1 of [7] implies sequences of standard queueing systems exhibit the same limiting behavior in heavy traffic. Moreover, this is true for all values of ρ (even $\rho < 1$) as long as we normalize by $n^{\frac{1}{2}}$.

Thus, we consider a sequence of modified queueing facilities with r arrival channels and s service channels. The arrival channels may be departure channels or part of departure channels from other facilities, but all $r + s$ channels associated with the facility in question are independent in each system.

The modified system differs from the standard system in two respects. First, the servers are not shut off when they become idle. With each server (and not with each customer, as is usually done) we associate a sequence of potential service times (random variables). If a server faces continued demand for service, then the actual service times of his successive customers are just these potential service times; but if there is no demand during any potential service time, then that potential service time is ignored and there is no actual service and no departure. After a server has begun working in the absence of demand, then the next demand will in general occur in the middle of some potential service time. Let the remaining portion of that potential service time be that next customer's actual service time.

The second difference in the modified system is that customers are served by the server who can complete the service first, which is not necessarily the first idle server. This means that customers will depart in the order they arrived. Moreover, every completion of a potential service time will generate an actual departure as long as there is a customer demanding service somewhere in the system. This property allows us to work directly with the net potential output process obtained by superimposing the potential outputs from the separate servers. This modified server system is of interest in its own right. For us, it is a device.

In each system assume that customers arrive and depart one at a time in each channel; interarrival times or service times of 0 can be used to represent batch processing. Let arriving customers join a single queue in front of the s servers or, equivalently, each customer immediately upon arrival can be assigned to one of s separate queues in front of the s servers. In this case, we look ahead and assign the customer to the server who would eventually serve him anyway.

Our basic data for each system are $r + s$ independent sequences of nonnegative random variables. (We make no i.i.d. assumptions for each sequence until the examples in Section 3.) Let $u(i,j,k)$ $[i = 1,2,\cdots;\ j = 1,2,\cdots;\ k = 1,\cdots,r]$ be the interarrival time between the $(i-1)$th and ith customer in the kth arrival channel of the jth queueing system. Let $v(i,j,k)$ $[\iota = 1,2,\cdots;\ j = 1,2,\cdots;\ k = 1,\cdots,s]$ be the ith potential service time in the kth service channel in the jth queueing system. Let all these random variables be defined on a common probability space $(\Omega, \mathscr{B}, P)$. We define counting processes associated with each channel in each system. Let

$$A^{kj}(t) = \begin{cases} \max\{m: u(1,j,k) + \cdots + u(m,j,k) \leqq t\}, & u(1,j,k) \leqq t \\ 0, & u(1,j,k) > t \end{cases}$$

for $t \geqq 0$, $1 \leqq k \leqq r$, $j \geqq 1$, and

$$S^{kj}(t) = \begin{cases} \max\{m: v(1,j,k) + \cdots + v(m,j,k) \leqq t\}, & v(1,j,k) \leqq t \\ 0, & v(1,j,k) > t \end{cases}$$

for $t \geqq 0$, $1 \leqq k \leqq s$, $j \geqq 1$. These processes represent the total number of arrivals or potential service times in the time interval $[0,t]$ in the appropriate channel and system. Because of the service discipline in the modified system, it is particularly easy to express the queue length process, $Q'(t)$, in terms of these basic counting processes. Throughout this paper all queue length processes count the customers being served as well as those waiting. We also place no upper bound on the number of waiting customers. Barriers corresponding to finite waiting rooms could be introduced here (cf. [16], page 111), but we shall not. For each $\omega \in \Omega$, $t \geqq 0$, and $j \geqq 1$,

$$Q'^{j}(t) = X^{j}(t) - \inf\{X^{j}(s),\ 0 \leqq s \leqq t\},$$

where

$$\begin{aligned} A^{j}(t) &= A^{1j}(t) + \cdots + A^{rj}(t), \\ S^{j}(t) &= S^{1j}(t) + \cdots + S^{sj}(t), \\ X^{j}(t) &= A^{j}(t) - S^{j}(t). \end{aligned}$$

and

Now define the (single) sequences of random functions in $D[0,1]$ induced by these stochastic processes

$$\begin{aligned} A_n^i &\equiv [A^{in}(nt) - \lambda_i^n nt]/n^{\frac{1}{2}}, \quad (1 \leqq k \leqq r), \\ A_n &\equiv [A^{n}(nt) - \lambda^n nt]/n^{\frac{1}{2}}, \\ S_n^j &\equiv [S^{jn}(nt) - \mu_j^n nt]/n^{\frac{1}{2}}, \quad (1 \leqq j \leqq s), \\ S_n &\equiv [S^{n}(nt) - \mu^n nt]/n^{\frac{1}{2}}, \\ X_n &\equiv [X^{n}(nt) - (\lambda^n - \mu^n)nt]/n^{\frac{1}{2}}, \\ Y_n &\equiv X^{n}(nt)/n^{\frac{1}{2}}, \\ Q_n' &\equiv Q'^{n}(nt)/n^{\frac{1}{2}}, \\ Q_n'' &\equiv [Q'^{n}(nt) - (\lambda^n - \mu^n)nt]/n^{\frac{1}{2}}. \end{aligned}$$

and

The constants λ_i^n and μ_j^n are to be chosen so that the random functions

A_n^i and S_n^j converge weakly in $D[0,1]$ as $n \to \infty$. When the sequences $\{u(i,j,k), i \geqq 1\}$ and $\{v(i,j,k), i \geqq 1\}$ are i.i.d. with finite expectation,

$$\lambda_i^n = 1/Eu(1,n,i)$$

and

$$\mu_j^n = 1/Ev(1,n,j).$$

We have normalized all our random functions by the usual normalizing constant $n^{\frac{1}{2}}$, but Lemma 1 and our main result, Theorem 1, hold for any normalizing constant yielding weak convergence.

Lemma 1. If $A_n^i \Rightarrow A^i$ $(1 \leqq i \leqq r)$ and $S_n^j \Rightarrow S^j$ $(1 \leqq j \leqq s)$, then $X_n \Rightarrow X$, where $X = A - S$, $A = \sum_{i=1}^r A^i$, and $S = \sum_{j=1}^s S^j$.

Proof. This is a simple generalization of Lemma 2.1 of [7]. Recall that the $r + s$ channels are independent in each system.

To proceed now to our functional central limit theorems for $Q'^j(t)$, we introduce the continuous mapping $f: D \to D$ which corresponds to an impenetrable barrier at the origin. For $x \in D$, f is defined by $f(x)(t) = x(t) - \inf_{0 \leqq s \leqq t} x(s)$, $0 \leqq t \leqq 1$. Let M be the constant linear function in D: $M(t) = ct$, $0 \leqq t \leqq 1$. Let M_n be the normalized translation term in X_n, that is, $M_n(t) = (\lambda^n - \mu^n)n^{\frac{1}{2}}t$, $0 \leqq t \leqq 1$. Notice that we make no explicit i.i.d. assumptions for each of the sequences $\{u(i,j,k), i \geqq 1\}$ and $\{v(i,j,k), i \geqq 1\}$.

Our principal result is

Theorem 1. Suppose $A_n^i \Rightarrow A^i$ $(1 \leqq i \leqq r)$ and $S_n^j \Rightarrow S^j$ $(1 \leqq j \leqq s)$ in D.

(a) If $(\lambda^n - \mu^n)n^{\frac{1}{2}} \to c$, $-\infty < c < +\infty$, then $Q_n' \Rightarrow f(X + M)$.
(b) If $(\lambda^n - \mu^n)n^{\frac{1}{2}} \to +\infty$, then $Q_n'' \Rightarrow X$.
(c) If $(\lambda^n - \mu^n)n^{\frac{1}{2}} \to -\infty$ and $\Pr\{X \in C\} = 1$, then $Q_n' \Rightarrow 0$.

Proof. (a) Theorems 4.4 and 5.1 of [1] imply that $Y_n \Rightarrow X + M$ since $X_n \Rightarrow X$ and $M_n \Rightarrow M$. However, $Q_n' = f(Y_n)$ so we can apply Theorem 5.1 of [1] again.

(b) It suffices to show $d(Q_n'', X_n) \Rightarrow 0$ and apply Theorem 4.1 of [1]. Observe that

$$d(Q_n'', X_n) \leqq \rho(Q_n'', X_n) = -\inf_{0 \leqq t \leqq 1} \{X^n(nt)/n^{\frac{1}{2}}\}.$$

For any t_0, $0 < t_0 < 1$,

$$-\inf_{0 \leqq t \leqq t_0} \{X^n(nt)/n^{\frac{1}{2}}\} \leqq -\inf_{0 \leqq t \leqq t_0} \{[X^n(nt) - (\lambda^n - \mu^n)nt]/n^{\frac{1}{2}}\}$$

and

$$-\inf_{t_0 \leqq t \leqq 1} \{X^n(nt)/n^{\frac{1}{2}}\} \leqq -\inf_{t_0 \leqq t \leqq 1} \{[X^n(nt) - (\lambda^n - \mu^n)nt]/n^{\frac{1}{2}}\} - (\lambda^n - \mu^n)nt_0/n^{\frac{1}{2}}$$

for $n > n_1$, where n_1 is chosen so that $\lambda_n > \mu_n$ for $n > n_1$. Suppose positive ε and η are given. Using Lemma 1 of [1], page 110 and the fact that $\Pr\{X(0)=0\}=1$, choose t_0 sufficiently small so that

$$\Pr\{-\inf_{0\leqq t\leqq t_0} X(t) > \varepsilon/2\} < \eta/4.$$

Next select m and n_2 so large that for $n > n_2$

$$\Pr\left\{-\inf_{0\leqq t\leqq t_0}[X^n(nt)-(\lambda^n-\mu^n)nt]/n^{\frac{1}{2}}] > \varepsilon/2\right\}$$
$$\leqq \Pr\{-\inf_{0\leqq t\leqq t_0}[X(t)] > \varepsilon/2\} + \eta/4,$$

and

$$\Pr\left\{-\inf_{t_0\leqq t\leqq 1}\{[X^n(nt)-(\lambda^n-\mu^n)nt]/n^{\frac{1}{2}}\} - (\lambda^n-\mu^n)n^{\frac{1}{2}}t_0 > \varepsilon/2\right\}$$
$$\leqq \Pr\left\{-\inf_{t_0\leqq t\leqq 1} X(t) > m\right\} \leqq \eta/2.$$

Then, for $n > n_1 \vee n_2$,

$$\Pr\left\{-\inf_{0\leqq t\leqq 1} X^n(nt)/n^{\frac{1}{2}} > \varepsilon\right\} < \eta.$$

Since

$$-\inf_{0\leqq t\leqq 1} X^n(nt)/n^{\frac{1}{2}} \geqq 0,$$

the proof is complete.

(c) Recall that $Q_n' = f(Y_n)$ and note that

$$\sup_{0\leqq t\leqq 1}\{[X^n(nt) - \inf_{t-\delta\leqq s\leqq t} X(ns)]/n^{\frac{1}{2}}\} \leqq w_{X_n}(\delta) \Rightarrow 0$$

as $n\to\infty$ and $\delta\to 0$. We use the fact that $\Pr\{X\in C\}=1$ to get C-tightness for $\{X_n\}$. Also for any positive ε and δ,

$$\inf_{0\leqq s\leqq t-\delta}\{X^n(ns)/n^{\frac{1}{2}}\} \geqq \inf_{0\leqq s\leqq t-\delta}\{[X^n(ns)-(\lambda^n-\mu^n)ns]/n^{\frac{1}{2}}\} + (\lambda^n-\mu^n)n^{\frac{1}{2}}(t-\delta)$$
$$\geqq X^n(nt)/n^{\frac{1}{2}}$$

for all $t, \delta \leqq t \leqq 1$, with probability greater than $1-\varepsilon$ for sufficiently large n. Hence, $Q_n' \Rightarrow 0$ as claimed.

In our model the queueing systems are empty at $t=0$, but Theorem 1 holds for other initial conditions as well. This is easily verified by applying Theorem 4.1 of [1].

We now investigate the departure process from this same facility. Again the

standard system exhibits the same limiting behavior as the modified system for all values of ρ when we normalize by $n^{\frac{1}{2}}$ (cf. [7], Lemma 4.1).

Let the sequence of departure processes for the sequence of modified systems be denoted by $\{\{D'^j(t), t \geqq 0\}, j \geqq 1\}$. From the definition of a departure process, $D'^j(t) = A^j(t) - Q'^j(t)$. From the definition of $Q'^j(t)$, we have

$$\begin{aligned} D'^j(t) &= A^j(t) - \{X^j(t) - \inf_{0 \leqq s \leqq t} X^j(s)\} \\ &= S^j(t) + \inf_{0 \leqq s \leqq t} \{A^j(s) - S^j(s)\}. \end{aligned}$$

Now define the random functions D_n' and D_n'' by

$$D_n' \equiv [D'^n(nt) - \mu^n nt]/n^{\frac{1}{2}}$$

and

$$D_n'' \equiv [D'^n(nt) - \lambda^n nt]/n^{\frac{1}{2}}, \ 0 \leqq t \leqq 1.$$

Also define the continuous function $g: D \times D \times D \to D$ by

$$g(x, y, z)(t) = y(t) + \inf_{0 \leqq s \leqq t} \{x(s) - y(s) + z(s)\}, \ 0 \leqq t \leqq 1,$$

for any $(x, y, z) \in D \times D \times D$.

Theorem 2. Suppose $A_n^i \Rightarrow A^i$ $(1 \leqq i \leqq r)$ and $S_n^j \Rightarrow S^j$ $(1 \leqq j \leqq s)$ in D.

(a) If $(\lambda^n - \mu^n)n^{\frac{1}{2}} \to c$, $-\infty < c < \infty$, then $D_n' \Rightarrow g(A, S, M)$.

(b) If $(\lambda^n - \mu^n)n^{\frac{1}{2}} \to +\infty$, then $D_n' \Rightarrow S$.

(c) If $(\lambda^n - \mu^n)n^{\frac{1}{2}} \to -\infty$ and $\Pr\{X \in C\} = 1$, then $D_n'' \Rightarrow A$.

We remark that (c) verifies a conjecture in [7].

Proof. (a) Since $D_n' = g(A_n, S_n, M_n)$, it only remains to apply the continuous mapping theorem ([1], Theorem 5.1).

(b) From Theorem 1(b), we know that $d(Q_n'', X_n) \Rightarrow 0$. Hence $d(D_n', S_n) \Rightarrow 0$ and we can apply ([1], Theorem 4.1).

(c) From Theorem 1(c), we know that $Q_n' \Rightarrow 0$. Hence, $d(D_n', A_n) \Rightarrow 0$.

All the results in Sections 5–9 of [7] also extend to the more general setting of this paper. Limit theorems when ρ is fixed at 1 extend to case (a), $(\lambda^n - \mu^n)n^{\frac{1}{2}} \to c$, $-\infty < c < \infty$. For example, in case (a), $Q_n^i \Rightarrow (\mu_i/\mu)f(X + M)$, where $\mu_i^n \to \mu_i$ and $\mu^n \to \mu$ as $n \to \infty$. Our limits for first passage times in Section 9 of [7] now would involve first passage times of Wiener processes with a drift. Since all the arguments are almost identical, we shall not discuss these topics further here.

3. Examples

We now mention several specific situations in which the previous theorems apply. Our first example is a generalization of the model considered in [7]. It contains sequences of $GI/G/s$ queues as a special case.

(1) *Sequences of i.i.d. random variables*

For each $j \geqq 1$, let the $r+s$ sequences $\{u(i,j,k), i \geqq 1\}$ $(1 \leqq k \leqq r)$ and $\{v(i,j,k), i \geqq 1\}$ $(1 \leqq k \leqq s)$ be independent sequences of non-negative i.i.d. random variables. Furthermore, for all j and k, assume that

$$0 < Eu(1,j,k) < \infty,$$

$$0 < Ev(1,j,k) < \infty,$$

$$0 < \sigma^2[u(1,j,k)] < \infty,$$

$$0 < \sigma^2[v(1,j,k)] < \infty,$$

$$\lim_{j\to\infty} Eu(1,j,k) = Eu(k),\ 0 < Eu(k) < \infty,$$

$$\lim_{j\to\infty} Ev(1,j,k) = Ev(k),\ 0 < Ev(k) < \infty,$$

$$\lim_{j\to\infty} \sigma^2[u(1,j,k)] = \sigma^2[u(k)],\ 0 < \sigma^2[u(k)] < \infty,$$

and

$$\lim_{j\to\infty} \sigma^2[v(1,j,k)] = \sigma^2[v(k)],\ 0 < \sigma^2[v(k)] < \infty.$$

We also need to further control the distribution of $u(i,j,k)$ and $v(i,j,k)$ as j varies. For this purpose, assume $E\{|u(1,j,k)|^{2+\delta}\}$ and $E\{|v(1,j,k)|^{2+\delta}\}$ are uniformly bounded in j and k for some positive δ. It is easy to see that this last condition implies that

$$\lim_{m\to\infty} \int_{|x|>m} x^2 dF_{u(1,j,k)}(x) = 0$$

and

$$\lim_{m\to\infty} \int_{|x|>m} x^2 dF_{v(1,j,k)}(x) = 0$$

uniformly in j and k (cf. [16] Chapter 5; [14], page 200; or [11], page 220), which in turn implies the standard Lindeberg condition for the normalized random variables

$$U(i,j,k) = \frac{u(i,j,k) - Eu(i,j,k)}{(j\sigma^2[u(i,j,k,)])^{\frac{1}{2}}}$$

and

$$V(i,j,k) = \frac{v(i,j,k) - Ev(i,j,k)}{(j\sigma^2[v(i,j,k)])^{\frac{1}{2}}},$$

that is,

$$\lim_{j\to\infty} \sum_{i=1}^{j} \int_{|x|>m} x^2 dF_{U(i,j,k)}(x) = 0$$

and

$$\lim_{j\to\infty} \sum_{i=1}^{j} \int_{|x|>m} x^2 dF_{V(i,j,k)}(x) = 0$$

for all positive m.

Now form the sequences of random functions $\{X_n^k, n \geqq 1\}$ $(1 \leqq k \leqq r+s)$ in $D[0,1]$ induced by the partial sums:

$$X_n^k(t) = n^{-\frac{1}{2}} \sum_{i=1}^{[nt]} [u(i,n,k) - Eu(1,n,k)],\ 0 \leqq t \leqq 1,\ (1 \leqq k \leqq r)$$

$$X_n^{r+k}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{[nt]} [v(i,n,k) - Ev(1,n,k)],\ 0 \leqq t \leqq 1,\ (1 \leqq k \leqq s).$$

Theorem 3.1 of Prohorov (1956) (cf. [11], page 220) implies that $X_n^k \Rightarrow \sigma[u(k)]\xi^k$ $(1 \leqq k \leqq r)$ and $X_n^{r+k} \Rightarrow \sigma[v(k)]\xi^{r+k}$ $(1 \leqq k \leqq s)$, where ξ^k $(1 \leqq k \leqq r+s)$ are $r+s$ independent Wiener processes. We can now apply Theorem 1 of [8] to obtain corresponding limit theorems for the counting processes. For this purpose, let

$$\lambda_k^j = 1/Eu(1,j,k)\ (1 \leqq k \leqq r),\ \mu_k^j = 1/Ev(1,j,k)\ (1 \leqq k \leqq s),\ \lambda^j = \sum_{k=1}^{r} \lambda_k^j,$$

$$\mu^j = \sum_{k=1}^{s} \mu_k^j,\ \alpha_k^j = (\lambda_k^{j3}\sigma^2[u(1,j,k)])^{\frac{1}{2}},\ \sigma_k^j = (\mu_k^{j3}\sigma^2[v(1,j,k)])^{\frac{1}{2}},$$

$$\alpha_k = \lim_{j\to\infty} \alpha_k^j,\ \sigma_k = \lim_{j\to\infty} \sigma_k^j,\ \text{and } \gamma^2 = \sum_{k=1}^{r} \alpha_k^2 + \sum_{k=1}^{s} \sigma_k^2.$$

With these assumptions, $A_n^i \Rightarrow \alpha_i \xi^i$ $(1 \leqq i \leqq r)$ and $S_n^j \Rightarrow \sigma_j \xi^{r+j}$ $(1 \leqq j \leqq s)$ and Theorems 1 and 2 may be applied to obtain limits for the various sequences of queueing processes.

Of particular interest as a possible approximation for stable queues is the case in which $(\lambda_n - \mu_n)n^{\frac{1}{2}} \to c$, $-\infty < c < 0$. Then $Q_n' \Rightarrow f(\gamma\xi + M)$ or $Q_n'/\gamma \Rightarrow f(\xi + M/\gamma)$. Notice that $\xi + M/\gamma$ is the ordinary Wiener process with drift c/γ. The process $f(\xi + M/\gamma)$ is a Wiener process with negative drift c/γ together with an impenetrable barrier at the origin. This process is completely described by the density function $f(t,y;y_0)$ obtained by taking the partial derivative with respect to y of $\Pr\{\sup_{0\leqq s\leqq t}\{\xi(s) - cs/\gamma\} \leqq y \,|\, \xi(0) = y_0\}$ (cf. Cox and Miller (1965), page 224):

$$\begin{aligned} f(t,y;y_0) = (2\pi t)^{-\frac{1}{2}}\{&\exp\{-(y - y_0 - ct/\gamma)^2/2t\} \\ &+ \exp\{[-4y_0ct/\gamma - (y + y_0 + ct/\gamma)]/2t\}\} \\ &+ 2c\gamma\exp\{2cy/\gamma\}\{1 - \Phi([y + y_0 + ct/\gamma]/t^{\frac{1}{2}})\}, \end{aligned}$$

where Φ is the standard normal distribution function. For further discussion of this example, see [2], [14], and [16], Chapter 5.

(2) *Deterministic flows*

Suppose that in the nth system customers arrive in the ith channel deterministically one at a time at regular intervals of length $1/\lambda_i^n$. If $A_n^i \equiv [A^{in}(nt) - \lambda_i^n nt]/n^{\frac{1}{2}}$, $0 \leqq t \leqq 1$, then $A_n^i \Rightarrow A = 0$ (the zero function), and the effect of this channel appears solely in the sequence of translation constants $\{\lambda_i^n\}$.

(3) *Exchangeable random variables*

The i.i.d. assumptions in Example 1 are by no means necessary. Heavy traffic limit theorems for sequences of queueing facilities having dependent sequences of interarrival times or potential service times are an immediate consequence of Theorems 1 and 2 here, Theorem 1 of [8], and Chapter 4 of [1]. For further discussion, see [16], Section 4.4. We now give an example.

Suppose that in the nth system exactly $n - 1$ customers arrive in the ith channel in the time interval $[0, n]$. Let the arrival times of these $n - 1$ customers be independent random variables uniformly distributed over the interval $[0, n]$. The interarrival times in each system may be obtained by looking at the differences between successive order statistics. They are exchangeable (not independent) random variables with mean 1 and variance $(n-1)/(n+1)$.

Let $\{X_n^i, n \geqq 1\}$ be the sequence of random functions induced in $D[0,1]$ by the partial sums

$$X_n^i(t) = n^{-\frac{1}{2}} \sum_{j=1}^{[nt]} (u_j^n - 1), \quad 0 \leqq t \leqq 1.$$

Theorem 24.2 of [1] implies that $X_n^i \Rightarrow \xi^\circ$, where ξ° is the Brownian Bridge. Theorem 1 of [8] implies that $A_n^i \Rightarrow \xi^\circ$ as well, where $A_n^i \equiv [A^{in}(nt) - nt]/n^{\frac{1}{2}}$, $0 \leqq t \leqq 1$. Other work on this queueing model has been done by Takács ((1967), page 125).

(4) *Batches*

Suppose customers arrive and are served in batches. Let b_i^{jk} be the ith batch size in the kth channel of the jth system. For each $j \geqq 1$, assume the $r + s$ sequences $\{b_i^{jk}, i \geqq 1\}$ are independent sequences of non-negative i.i.d. random variables with uniformly bounded $(2 + \delta)$th moments. Furthermore, assume $Eb_1^{jk} \equiv \beta_k^j \to \beta_k > 0$ and $\sigma^2[b_1^{jk}] \equiv \sigma_k^{j2} \to \sigma_k^2 > 0$ as $j \to \infty$. Let $\{X_r^k, n \geqq 1\}$ be the sequence of random functions induced by the double sequence of partial sums of the successive batch sizes in the kth channel, that is,

$$X_n^k(t) = n^{-\frac{1}{2}} \sum_{i=1}^{[nt]} (b_i^{nk} - \beta_k^n), \; 0 \leqq t \leqq 1.$$

Theorem 3.1 of [13] implies that $X_n^k \Rightarrow \sigma_k \xi$, where ξ is the standard Wiener process.

To be definite, suppose the kth channel is an arrival channel. Let $N^{jk}(t)$

be the counting process in the jth system recording the number of batches that have arrived in the kth channel in $[0,t]$. Let N_n^k be the associated random functions in D and assume that N_n^k is independent of X_n^k for each n. In addition, assume that $N_n^k \Rightarrow N$, where $N \in C$; $N_n^k \equiv [N^{nk}(nt) - \lambda_k^n nt]/n^{\frac{1}{2}}$, $0 \leqq t \leqq 1$; and $\lambda_k^n \to \lambda_k$, $0 < \lambda_k < \infty$. Then

$$A_n^k \Rightarrow \sigma_k \lambda_k^{\frac{1}{2}} \xi + \beta_k N,$$

where ξ and N are independent and

$$A_n^k \equiv \left[\sum_{i=1}^{N^{nk}(nt)} b_i^{nk} - \lambda_k^n \beta_k^n nt \right] / n^{\frac{1}{2}}, \quad 0 \leqq t \leqq 1$$

(cf. [6]). In the i.i.d. case, $N = (\tau_k^2 \lambda_k^3)^{\frac{1}{2}} \xi'$, where τ_k^2 is the limit as $j \to \infty$ of the variances of the times between events in $\{N^{jk}(t),\ t \geqq 0\}$. Then $A_n^k \Rightarrow A^k$, where $A^k = (\sigma_k^2 \lambda_k + \beta_k^2 \tau_k^2 \lambda_k^3)^{\frac{1}{2}} \xi$.

(5) *Split channels*

Suppose an arrival channel splits up into several separate arrival channels going into different facilities. Let successive customers in the main channel independently select the arrival channel leading to a specified facility with probability p^j in system j, $p^j \to p$, $0 < p < 1$. Let $A_n \equiv [A^n(nt) - \lambda^n nt]/n^{\frac{1}{2}}$, $0 \leqq t \leqq 1$, be the random functions induced by the counting processes associated with the main channel. Let B_n be the random functions induced by the counting processes associated with the tributary leading to the specified facility. Then

$$B_n(t) = \left[\sum_{i=1}^{A^n(nt)} \chi_i^n - \lambda^n p^n nt \right] / n^{\frac{1}{2}}, \quad 0 \leqq t \leqq 1,$$

where $\chi_i^n = 1$ if the ith customer in the main channel in the nth system selects the specified facility, and $\chi_i^n = 0$ otherwise.

If $A_n \Rightarrow A$, $\Pr\{A \in C\} = 1$, and $\lambda_n \to \lambda > 0$, then the double sequence version of the weak convergence theorem for random sums in [6] can be applied again to yield

$$B_n \Rightarrow p(1-p)\lambda \xi + pA,$$

where ξ is a Wiener process independent of A.

4. Joint limits for several facilities in a network

We now show how to obtain heavy traffic limit theorems jointly for stochastic processes associated with several different facilities in the same network. For this purpose, consider a network of five facilities linked together as follows:

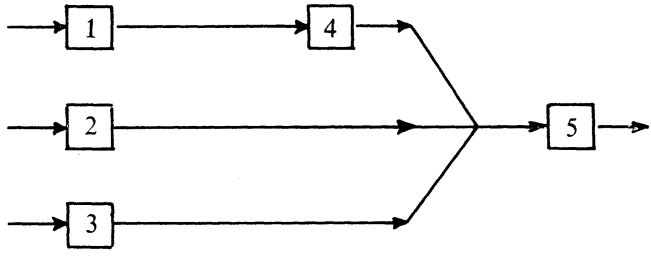

Figure 1

Let $A_n^i \equiv [A^{in}(nt) - \lambda_i^n nt]/n^{\frac{1}{2}}$, $0 \leqq t \leqq 1$, be the random functions induced by the net arrival processes into the ith facility, $1 \leqq i \leqq 3$. Let $S_n^j \equiv [S^{jn}(nt) - \mu_j^n nt]/n^{\frac{1}{2}}$, $0 \leqq t \leqq 1$, be the random functions induced by the net potential service processes of the jth facility, $1 \leqq j \leqq 5$. These eight random functions are the basic data. Assume that $(A_n^1, A_n^2, A_n^3, S_n^1, \cdots, S_n^5) \Rightarrow (A^1, A^2, A^3, S^1, \cdots, S^5)$ in the eight-fold product space of $D[0,1]$ with itself.

In order to determine heavy traffic limit theorems for $[Q'^1(t), \cdots, Q'^5(t)]$, we need to know which cases of heavy traffic prevail at each facility. The behavior at facilities 4 and 5 naturally depends on the behavior at the first three facilities. Suppose that

$$\gamma_1^n = (\lambda_1^n - \mu_1^n)n^{\frac{1}{2}} \to c_1, \quad -\infty < c_1 < \infty$$

$$\gamma_2^n = (\lambda_2^n - \mu_2^n)n^{\frac{1}{2}} \to +\infty,$$

$$\gamma_3^n = (\lambda_3^n - \mu_3^n)n^{\frac{1}{2}} \to -\infty,$$

$$\gamma_4^n = (\mu_1^n - \mu_4^n)n^{\frac{1}{2}} \to c_4, \quad -\infty < c_4 < \infty$$

and

$$\gamma_5^n = (\mu_4^n + \mu_2^n + \lambda_3^n - \mu_5^n)n^{\frac{1}{2}} \to c_5, \quad -\infty < c_5 < \infty.$$

Let $M_n^i \equiv \gamma_i^n t, 0 \leqq t \leqq 1$, $(1 \leqq i \leqq 5)$ and $M^i \equiv c_i t$, $0 \leqq t \leqq 1$, $(i = 1,4,5)$ be the corresponding random functions. Furthermore, assume $\Pr\{A_3 - S_3 \in C\} = 1$ Then, the continuous mapping theorem ([1], Theorem 5.1) implies that

$$(Q_n'^1, Q_n'^2, Q_n''^3, Q_n'^4, Q_n'^5) \Rightarrow (f(A^1 - S^1 + M^1), A^2 - S^2, 0,$$
$$f[g(A^1, S^1, M^1) - S^4 + M^4], f[S^2 + A^3 + g(g[A^1, S^1, M^1], S^4, M^4) - S^5 + M^5]).$$

Suppose that we are interested in the total number of customers in the network. The appropriate random functions are $Q_n = Q_n'^1 + Q_n'^2 + Q_n''^3 + Q_n'^4 + Q_n'^5$. Note that $Q_n''^3 = [Q_n'^3(nt) - (\lambda_3^n - \mu_3^n)nt]/n^{\frac{1}{2}}$, $0 \leqq t \leqq 1$. The continuous mapping theorem applied again yields

$$Q_n \Rightarrow f(A^1 - S^1 + M^1) + A^2 - S^2$$
$$+ f[g(A^1, S^1, M^1) - S^4 + M^4]$$
$$+ f[S^2 + A^3 + g(g[A^1, S^1, M^1], S^4, M^4) - S^5 + M^5].$$

Such limits are easy to determine, but obviously hard to evaluate in detail.

5. Future research

In [16], [7], and this paper we have conducted a fairly extensive investigation of queues in heavy traffic. Many of the problems posed in Chapter 10 of [16] are now solved. However, many interesting problems remain.

A major thrust of the heavy traffic research is the desire to find useful approximations for stable queues. The Wiener process with negative drift and an impenetrable barrier at the origin is a candidate obtained from our limit theorems (cf. Example 1 and [16], Sections 5.2 and 5.3). Further work is needed on rates of convergence (cf. [16], Section 4.3), and numerical comparisons (cf. Gaver (1968)).

Similar theorems for other queueing models should be proved. Queue disciplines other than first-come-first-served ought to be considered as well as other rules for assigning a customer to one of the servers (cf. [17]). Also, the arrival rates and the service rates should be allowed to depend on time and the state of the system. We have treated networks of queueing facilities, but a more detailed evaluation of the limits is needed. Furthermore, it still remains to consider networks which exhibit cyclic behavior, that is, part of the output from a facility may reappear in the input.

A significant feature of weak convergence in $D[0, 1]$ is that we have limit theorems for many functionals of our processes as well as limit theorems for the processes themselves by virtue of [1] Theorem 5.1. We have indicated how such additional results can be obtained in [16] Chapter 9, and [7] Section 9, but this property of weak convergence ought to be further exploited to treat problems of optimization and control.

References

[1] BILLINGSLEY, P. (1968) *Convergence of Probability Measures.* John Wiley and Sons, New York.

[2] BOROVKOV, A. A. (1965) Some limit theorems in the theory of mass service, II. *Theor. Probability Appl.* **10,** 375–400.

[3] BOROVKOV, A. A. (1967) On limit laws for service processes in multi-channel systems. *Siberian Math. J.* **8,** 746–763.

[4] COX, D. R. AND MILLER, H. D. (1965) *The Theory of Stochastic Processes.* John Wiley and Sons, New York.

[5] GAVER, D. P. (1968) Diffusion approximations and models for certain congestion problems. *J. Appl. Prob.* **5,** 607–623.

[6] IGLEHART, D. L. AND KENNEDY, D. P. (1970) Weak convergence of the average of flag processes. *J. Appl. Prob.* (To appear).

[7] IGLEHART, D. L. AND WHITT, W. (1969) Multiple channel queues in heavy traffic. I. *Adv. Appl. Prob.* **2.** 150–177.

[8] IGLEHART, D. L. AND WHITT, W. (1969) The equivalence of functional central limit theorems for counting processes and associated partial sums. Technical Report No. 5, Department of Operations Research, Stanford University.

[9] KINGMAN, J. F. C. (1961) The single server queue in heavy traffic. *Proc. Camb. Phil. Soc.* **57,** 902–904.

[10] KINGMAN, J. F. C. (1965) The heavy traffic approximation in the theory of queues. W. Smith and W. Wilkinson (eds.) *Proc. Symposium on Congestion Theory.* Univ. of North Carolina Press, Chapel Hill, 137–159.

[11] PARTHASARATHY, K. R. (1967) *Probability Measures on Metric Spaces.* Academic Press, New York.

[12] PRABHU, N. U. (1970) Limit theorems for the single server queue with traffic intensity one. *J. Appl. Prob.* **7,** 227–233.

[13] PROHOROV, YU. V. (1956) Convergence of random processes and limit theorems in probability theory. *Theor. Probability Appl.* **1,** 157–214.

[14] PROHOROV, YU. V. (1963) Transient phenomena in processes of mass service (in Russian). *Litovsk. Mat. Sb.* **3,** 199–205.

[15] TAKÁCS, L. (1967). *Combinatorial Methods in the Theory of Stochastic Processes.* John Wiley and Sons, New York.

[16] WHITT, W. (1968) *Weak Convergence Theorems for Queues in Heavy Traffic.* Ph. D. thesis, Cornell University. (Technical Report No. 2, Department of Operations Research, Stanford University.)

[17] WHITT, W. (1970) Multiple channel queues in heavy traffic. III: random server selection. *Adv. Appl. Prob.* **2,** 370–375.