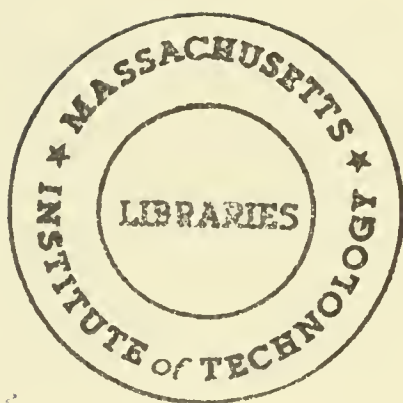


BASEMENT



HD28
.M414

29

DEWEY

no
3673-
94

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

**Heavy Traffic Analysis of a
Transportation Network Model**

William P. Peterson
Lawrence M. Wein

#3673-94-MSA

April 1994

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

**Heavy Traffic Analysis of a
Transportation Network Model**

William P. Peterson
Lawrence M. Wein

#3673-94-MSA

April 1994

M.I.T. LIBRARIES

MAY 17 1994

RECEIVED

HEAVY TRAFFIC ANALYSIS OF A TRANSPORTATION NETWORK MODEL

William P. Peterson ¹

Operations Research Center, M.I.T.

and

Lawrence M. Wein ²

Sloan School of Management, M.I.T.

Abstract

We study a model of a stochastic transportation system introduced by Crane. By adapting constructions of multidimensional reflected Brownian motion (RBM) that have since been developed for feedforward queueing networks, we generalize Crane's original functional central limit theorem results to a full vector setting, giving an explicit development for the case in which all terminals in the model experience heavy traffic conditions. We investigate product form conditions for the stationary distribution of our resulting RBM limit, and contrast our results for transportation networks with those for traditional queueing network models.

Keywords. Transportation networks, queueing, heavy traffic, reflected Brownian motion.

AMS 1991 subject classification. Primary: 90B22, 60K25. Secondary: 60F17.

April 11, 1994

¹Research supported by Middlebury College under the Faculty Leave Program.

²Research supported by National Science Foundation grant DDM-9057297.

1 Introduction and Summary

In this paper, we prove a heavy traffic limit theorem for a transportation system model introduced by Crane (1974). The model, depicted in Figure 1, is given a full mathematical formulation in Section 2. To summarize, the system under consideration consists of a linear network of $N + 1$ terminals, indexed $i = 0, 1, \dots, N$. The network is served by a fleet of M vehicles, each having fixed passenger capacity V . The fleet is partitioned into service groups S_1, \dots, S_N . Vehicles in group S_i are initially dispatched from terminal 0 and proceed directly to terminal i , where they pick up waiting passengers. They make successive stops at $i - 1, \dots, 1$ to discharge and pick up passengers, then return to terminal 0, where all remaining passengers disembark. This cyclic route is repeated indefinitely by each vehicle in the group. Terminals $i = 1, \dots, N$ each have an exogenous arrival stream of prospective passengers. Passengers arriving to terminal i require transportation to one of the terminals $i - 1, \dots, 0$. They join a single queue at i , and board arriving vehicles in a first-come first-served (FCFS) fashion subject to availability of seats. There is no bound on the number of passengers that can be queued at any station, and it is assumed that prospective passengers do not balk or renege. As Crane notes, this model can be applied to a public transportation system, such as a municipal subway line, in which there are no published schedules and passengers arrive without reservations. His paper appears to be the only previous attempt to analyze the performance of a dynamic stochastic network model of a public transportation system with limited vehicle capacity.

Crane presents functional central limit theorem results for his model, of the type obtained in the seminal papers of Iglehart and Whitt (1970) on queueing systems. Because the theory of what is now called multidimensional reflected Brownian motion was unknown at the time, Crane was able to explicitly describe his limit processes only for the case in which a single terminal in the network is critically loaded, in the sense that the exogenous passenger arrival rate at the terminal equals the rate at which vehicle capacity is provided there. Even in that case, the covariance manipulations presented are quite cumbersome. Our

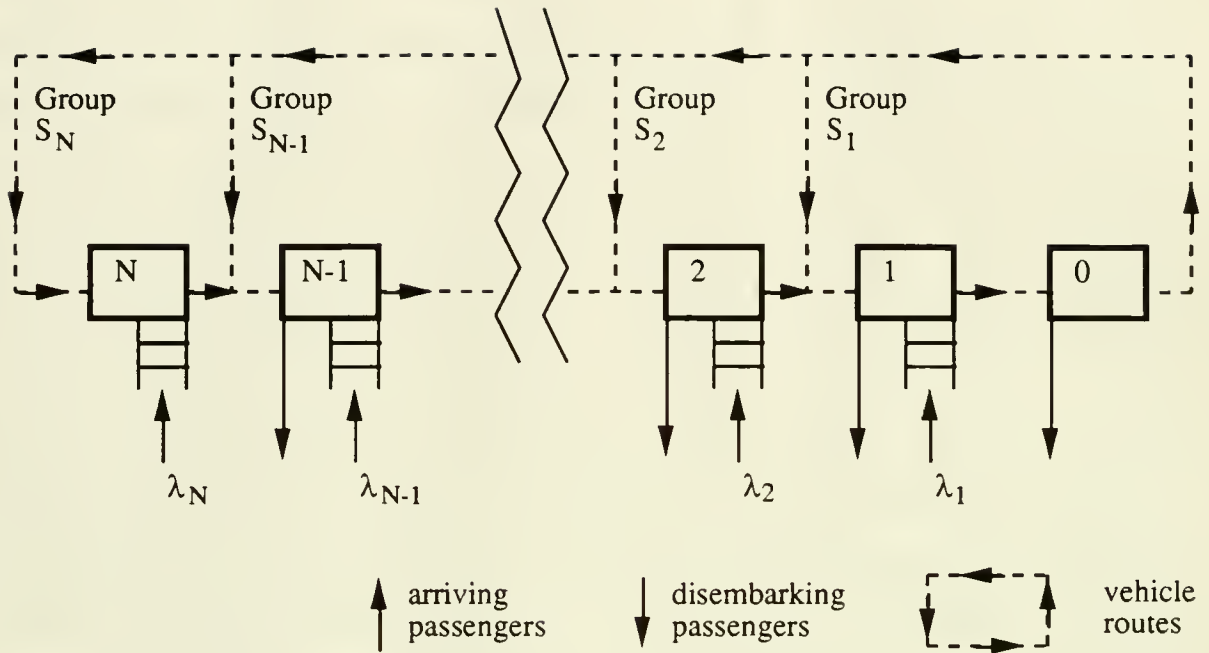


Figure 1: The transportation network.

contribution here is to consider a sequence of systems in which all terminals approach critical loading. This state of affairs, which corresponds to the usual *heavy traffic conditions* considered in the queueing literature, will be described precisely in Section 3. In Section 4, we state and prove a limit theorem for the vector process of queue lengths at the terminals, for which the limit process is a reflected Brownian motion (RBM) on the nonnegative orthant. The linear structure of the transportation network allows an inductive proof analogous to the treatment of feedforward queueing networks in Peterson (1991).

The theorem presented here justifies the use of what Harrison (1985) has called a *Brownian system model* for the network. The particular form of our limit process has some features that distinguish it from the Brownian models for traditional queueing systems, such as Reiman's (1984) limit result for generalized Jackson networks or the broader framework of

multiclass queueing networks described in Harrison and Nguyen (1993). In particular, in the form of the reflection matrix one can identify both customer routing effects (i.e., passenger destination information) and vehicle effects: the queueing models cited include only the former. We discuss this in the context of an example in Section 5. On a negative note, it will be seen there that except in artificial special cases, the RBM limit can never satisfy the conditions for a product form stationary distribution. We show in terms of the covariance structure of the RBM why this is the case. However, it should be noted that even though the stationary distribution cannot be found analytically, the numerical methods recently developed by Dai and Harrison (1992) for steady-state analysis of RBM on an orthant can be applied to analyze system performance in our model.

Our limit results are formulated in terms of weak convergence in product spaces D^d for appropriate dimension d , where $D = D[0, T]$ is the space of real-valued functions on $[0, T]$ that are right continuous and have left hand limits, under the Skorohod topology. Weak convergence will be denoted by \Rightarrow . All vectors are to be understood as column vectors, with primes used to indicate matrix and vector transpose.

2 The Model

We begin by summarizing the probabilistic structure of the model, following the treatment in Crane. Three basic phenomena need to be described: exogenous passenger arrivals to the system, passengers' choice of destination terminals, and circulation of vehicles in the network. For simplicity, it is assumed that there are no passengers in the network at time $t = 0$, and that all of the vehicles in the fleet begin their routes from terminal $i = 0$ at that time. For present purposes, these assumptions have no significant effect on our limit results.

For $i = 1, \dots, N$, set $u_i(0) = 0$ and for $l = 1, 2, \dots$ let the random variable $u_i(l)$ represent the l^{th} interarrival time in the exogenous passenger arrival stream to terminal i . Define the counting processes

$$A_i(t) = \max\{l \geq 0 : u_i(0) + u_i(1) + \dots + u_i(l) \leq t\} \quad (1)$$

for $t \geq 0$, so that $A_i(t)$ represents the number of exogenous passenger arrivals to terminal i over $[0, t]$.

Passenger destination selection is described by an $N \times N$ matrix $\Gamma = (\gamma_{ij})$. We assume that a passenger arriving to terminal i has destination terminal j with probability γ_{ij} , independent of the destination of any other passenger. As described in the last section, passengers always seek transportation to lower numbered terminals, so $\gamma_{ij} = 0$ for $j \geq i$; in other words, Γ is strictly lower triangular. Also, $1 - \sum_{j=1}^{i-1} \gamma_{ij}$ is the probability that a passenger originating at i has destination terminal 0. We introduce for each terminal i a sequence of i.i.d. N -dimensional random vectors $\chi_i(l) = (\chi_{i,1}(l), \dots, \chi_{i,N}(l))'$, where $\chi_{ij}(l)$ is equal to 1 if the l^{th} passenger in the arrival stream to i has destination j , and is zero otherwise. Observe that $E[\chi_i(1)] = \Gamma'_i$, where Γ_i denotes the i^{th} row of Γ . Next define the N -dimensional cumulative sum processes $U_i = \{U_i(r), r = 1, 2, \dots\}$ by

$$U_i(r) = \chi_i(1) + \dots + \chi_i(r). \quad (2)$$

Then the j^{th} component of $U_i(r)$, which we shall denote by $U_{ij}(r)$, gives the number of passengers among the first r arrivals to terminal i who have destination j .

Recall that vehicles $k = 1, \dots, M$ have been partitioned into service groups S_i . Consider a vehicle $k \in S_i$. We introduce sequences of random variables $\{v_{k,j}(l), l = 1, 2, \dots\}$ for $j = 0, 1, \dots, i$, where $v_{k,j}(l)$ represents the transit time for the vehicle's l^{th} trip along the link starting from terminal j ; this link goes to terminal $j - 1$ if $j \geq 1$, and to terminal i if $j = 0$. It is assumed that when a vehicle arrives at a station, any passengers it is carrying who are destined for that station disembark instantaneously, and any waiting passengers then instantaneously board subject to availability of space. With this understanding, we define

$$\begin{aligned} w_k(1) &= v_{k,0}(1), \quad \text{and} \\ w_k(l) &= v_{k,0}(l) + \sum_{j=1}^i v_{k,j}(l-1), \quad l = 2, 3, \dots \end{aligned}$$

Then the sequence $\{w_k(l), l = 1, 2, \dots\}$ gives the times between successive visits of the

vehicle to terminal i . Define the corresponding counting processes $\{C_k(t), t \geq 0\}$ by

$$C_k(t) = \begin{cases} \max\{l : w_k(1) + \cdots + w_k(l) \leq t\}, & \text{if } w_k(1) \leq t, \\ 0, & \text{if } w_k(1) > t. \end{cases} \quad (3)$$

Now recalling that each vehicle has fixed capacity V , we see that for $i = 1, \dots, N$,

$$G_i(t) = V \sum_{k \in S_i} C_k(t) \quad (4)$$

gives the total number of units of vehicle capacity (i.e., available passenger spaces on vehicles) provided at terminal i during $[0, t]$ by vehicles in service group S_i .

Two comments are in order here pertaining to the circulation of vehicles. First, as noted by Crane, the above framework can be used to accommodate a stochastic idle time at the terminals, by simply envisioning that such a delay has been incorporated in the most recent link transit time. Observe, however, that this does not allow the idle time to depend in any way on passenger data. Second, the model fails to capture any interaction between the travel times for different vehicles. This would become an issue if passing of vehicles were not permitted, as would be the case with a single track rail system, for example. In his dissertation, Crane (1971) outlines an approach to incorporating such dependencies into the model.

The counting processes $\{A_i(t), t \geq 0\}$ and $\{G_i(t), t \geq 0\}$, and the vector sum processes $\{U_i(r), r = 1, 2, \dots\}$ for $i = 1, \dots, N$, are assumed to be mutually independent. These are the basic building blocks for our system representation. We proceed somewhat differently from Crane, developing a streamlined description that facilitates the proof of our vector limit theorem. Let $Z_i(t)$ denote the number of passengers waiting at terminal i at time t , $E_i(t)$ denote the number of passengers who have boarded vehicles at terminal i during $[0, t]$, and $Y_i(t)$ denote the cumulative number of units of capacity that go unoccupied on vehicles departing from station i during $[0, t]$. The following is a recursive construction of representations for these processes. Let

$$X_N(t) = A_N(t) - G_N(t), \quad \text{and} \quad (5)$$

$$\begin{aligned}
X_j(t) &= A_j(t) - G_j(t) - \sum_{i=j+1}^N U_{ij}(E_i(t)) \\
&\quad - Y_{j+1}(t) + \epsilon_j(t), \quad j = N-1, \dots, 1,
\end{aligned} \tag{6}$$

where, for $j = N, \dots, 1$,

$$Y_j(t) = - \inf_{0 \leq s \leq t} \{X_j(s)\}, \tag{7}$$

$$Z_j(t) = X_j(t) + Y_j(t) \quad \text{and} \tag{8}$$

$$E_j(t) = A_j(t) - Z_j(t). \tag{9}$$

The idea here is to construct X_j as a *netput process*; that is, $X_j(t)$ represents the difference between the total number of passengers arriving to station j during $[0, t]$ and the cumulative number of units of vehicle capacity provided there over that time period. If one accepts this interpretation of (5)-(6), then (7) and (8) are standard representations from the queueing literature, and (9) is simple bookkeeping. To understand (5)-(6), observe that there are effectively three ways that vehicle capacity can be provided at j . The first is through arrival of vehicles in service class S_j , as accounted for in the basic process G_j . This is the only source of capacity at terminal $j = N$, so X_N can be defined in (5) entirely in terms of basic processes; this gets the recursion started. At terminals $j < N$, there are two additional sources of capacity represented in (6). Arriving vehicles from service classes S_i for $i > j$ provide capacity at j when they are carrying passengers whose destination is j , since these passengers immediately disembark there, which frees up space for passengers queued at j . Note that $U_{ij}(E_i(t))$ gives the number of passengers with destination j among the $E_i(t)$ who board vehicles at i during $[0, t]$. In addition, unused capacity on vehicles departing from terminal $j+1$ during $[0, t]$, as represented by $Y_{j+1}(t)$, will provide capacity when those vehicles reach j . The $\epsilon_j(t)$ is an error term, which is required because the vehicles whose available capacity is accounted for by $U_{ij}(E_i(t))$ and $Y_{j+1}(t)$ may not have reached terminal j by time t . Observe, however, that a vehicle can contribute to $\epsilon_j(t)$ only if that vehicle is in transit somewhere between terminals i and j for some $i > j$ at time t . Thus, at any time,

the error is certainly bounded by the total capacity of the fleet:

$$0 \leq \epsilon_j(t) \leq MV \quad \text{for all } t \geq 0, \quad j = 1, \dots, N-1. \quad (10)$$

Using this bound, the error will be shown to be negligible under the *heavy traffic rescaling* to be defined in the next section.

We conclude this section with representations for the queue length and departure processes at the terminals broken out by passenger destination. At each terminal i , let $Q_{ij}(t)$ be the number of customers waiting at terminal i whose destination is terminal j , for $j = 0, \dots, i-1$. Similarly let $D_{ij}(t)$ be the number of passengers who have departed from i whose destination is j . Then for $i = 1, \dots, N$ we have the following straightforward representations:

$$Q_{ij}(t) = U_{ij}(A_i(t)) - U_{ij}(E_i(t)), \quad j = 1, \dots, i-1, \quad (11)$$

$$Q_{i0}(t) = Z_i(t) - \sum_{j=1}^{i-1} Q_{ij}(t), \quad (12)$$

$$D_{ij}(t) = U_{ij}(E_i(t)), \quad j = 1, \dots, i-1, \quad \text{and} \quad (13)$$

$$D_{i0}(t) = E_i(t) - \sum_{j=1}^{i-1} D_{ij}(t). \quad (14)$$

The compound processes D_{ij} were already introduced, though not explicitly named, in discussing equation (6). Separating them out in this way will be convenient for developing the main theorem. The processes Q_{ij} will be treated in a corollary to that theorem, where it will be shown that their limit processes are given by deterministic fractions of the limit for the total queue length Z_i at each terminal i . This is a manifestation of the *state space collapse* phenomenon in heavy traffic results.

3 The Heavy Traffic Conditions

The precise statement of our results involves a sequence of systems, of the type described in Section 2, approaching conditions of perfect balance between passenger arrivals and vehicle capacity at each terminal. Hereafter, a superscript n will be used to index processes and

parameters associated with the n^{th} system in the sequence. We will allow the passenger interarrival and vehicle transit time distributions to vary, so that the basic processes $A^n \equiv (A_1^n, \dots, A_N^n)$ and $G^n \equiv (G_1^n, \dots, G_N^n)$ depend on n . For simplicity, however, we assume that the destination probability matrix Γ is fixed, which implies that the vector processes $U_i = (U_{i1}, \dots, U_{iN})$ for $i = 1, \dots, N$ do not vary with n .

We require that there exist sequences of N -dimensional vectors $\{\lambda^n\}$ and $\{\beta^n\}$ so that the scaled processes \tilde{A}^n and \tilde{G}^n defined by

$$\tilde{A}^n(t) = n^{-1/2}(A^n(nt) - \lambda^n nt) \quad (15)$$

$$\tilde{G}^n(t) = n^{-1/2}(G^n(nt) - \beta^n nt) \quad (16)$$

satisfy functional central limit theorems (f.c.l.t.'s). For example, if the component arrival processes A_i^n and the round trip counting processes C_k^n underlying the G_i^n in (4) are mutually independent renewal processes, then the required results can be established under some additional moment assumptions. By (2), $U_i(r)$ has a multinomial distribution for each r , so the classical Donsker theorem gives f.c.l.t.'s for the processes \tilde{U}_i^n defined by

$$\tilde{U}_i^n(t) = n^{-1/2}(U_i(\lfloor nt \rfloor) - \Gamma_i^n nt), \quad i = 1, \dots, N. \quad (17)$$

Therefore, recalling the assumed independence of A^n , G^n and the U_i 's, we have

$$(\tilde{A}^n, \tilde{G}^n, \tilde{U}_1^n, \dots, \tilde{U}_N^n) \Rightarrow (A^*, G^*, U_1^*, \dots, U_N^*) \quad (18)$$

as $n \rightarrow \infty$, where the processes on the right are $N+2$ independent, zero drift, N -dimensional vector Brownian motions, whose covariance matrices will be denoted by Σ_A , Σ_G and Σ_{U_i} , $i = 1, \dots, N$.

The definitions in (15)-(17) consist of a centering of key processes followed by the heavy traffic rescaling, in which time is scaled by n and space by $n^{-1/2}$. The convergence in (18) gives the following interpretation of the centering parameters: λ_i^n is the long run average arrival rate to station i , and β_i^n is the long run average rate at which units of vehicle capacity are provided by vehicles in S_i . Ignoring the rescaling for the moment, we introduce

the following centered versions of our basic processes. Define:

$$\begin{aligned}\hat{A}^n(t) &= A^n(t) - \lambda^n t, \\ \hat{G}^n(t) &= G^n(t) - \beta^n t \quad \text{and} \\ \hat{U}_i(t) &= U_i(\lfloor t \rfloor) - \Gamma_i' t.\end{aligned}\tag{19}$$

We can later construct the tilde processes in (15)-(17) from these by rescaling; for example, $\tilde{A}^n(t) = n^{-1/2} \hat{A}^n(nt)$. For now, as an intermediate step, we seek to rewrite the system state equations from Section 2 in terms of the centered processes. Anticipating the balanced loading conditions, we expect the long run departure rates to match the corresponding arrival rates, so the following defines a natural centering for the departure processes:

$$\hat{E}^n(t) = E^n(t) - \lambda^n t.\tag{20}$$

Recall from (13) that $D_{ij}^n(t) = U_{ij}(E_i^n(t))$, and observe that

$$\begin{aligned}U_{ij}(E_i^n(t)) &= \hat{U}_{ij}(E_i^n(t)) + \gamma_{ij} E_i^n(t) \\ &= \hat{U}_{ij}(E_i^n(t)) + \gamma_{ij} \hat{E}_i^n(t) + \lambda_i^n \gamma_{ij} t.\end{aligned}$$

This suggests the centering

$$\hat{D}_{ij}^n(t) = U_{ij}(E_i^n(t)) - \lambda_i^n \gamma_{ij} t,\tag{21}$$

where $\lambda_i^n \gamma_{ij}$ can be interpreted as the long run rate at which customers destined for terminal j board vehicles at terminal i . By solving (19)-(21) for the uncentered processes and substituting into (5)-(9), our system state equations become

$$X_N^n(t) = \hat{A}_N^n(t) - \hat{G}_N^n(t) + (\lambda_N^n - \beta_N^n)t, \quad \text{and}\tag{22}$$

$$\begin{aligned}X_j^n(t) &= \hat{A}_j^n(t) - \hat{G}_j^n(t) - \sum_{i=j+1}^N \hat{D}_{ij}^n(t) + \left(\lambda_j^n - \beta_j^n - \sum_{i=j+1}^N \lambda_i^n \gamma_{ij} \right) t \\ &\quad - Y_{j+1}^n(t) + \epsilon_j^n(t), \quad j = N-1, \dots, 1,\end{aligned}\tag{23}$$

where, for $j = 1, \dots, N$

$$Y_j^n(t) = - \inf_{0 \leq s \leq t} \{X_j^n(s)\},\tag{24}$$

$$Z_j^n(t) = \Lambda_j^n(t) + Y_j^n(t) \quad \text{and} \quad (25)$$

$$\hat{E}_j^n(t) = \hat{A}_j^n(t) - Z_j^n(t). \quad (26)$$

Consideration of the rescaling to follow leads to our *heavy traffic conditions*. We suppose that as $n \rightarrow \infty$,

$$\lambda^n \rightarrow \lambda > 0 \quad \text{and} \quad \beta^n \rightarrow \beta > 0 \quad (27)$$

in such a way that, for $j = 1, \dots, N$,

$$\sqrt{n} \left(\lambda_j^n - \beta_j^n - \sum_{i=j+1}^N \lambda_i^n \gamma_{ij} \right) \rightarrow \theta_j, \quad -\infty < \theta_j < \infty. \quad (28)$$

This requires that the quantity in parentheses in (28) converge to zero; it also specifies a common rate of convergence for all terminals j . Following our earlier discussion, $\lambda_i^n \gamma_{ij}$ can also be interpreted as the long run rate at which customers who originated at i disembark at terminal j . Therefore, we can see that (27)-(28) describe a sequence of systems approaching critical loading. In the limit, the exogenous passenger arrival rate to each terminal j is exactly balanced by the rate at which capacity is provided at j by vehicles in group S_j plus the rate at which capacity is made available there by passengers disembarking from vehicles in groups S_i for $i > j$. Note that by letting $\theta = (\theta_1, \dots, \theta_N)'$, we can restate (28) in vector form as

$$\sqrt{n} \left((I - \Gamma') \lambda^n - \beta^n \right) \rightarrow \theta. \quad (29)$$

It is interesting to compare our heavy traffic conditions with the stability conditions derived by Crane. He defines for each terminal an effective service rate μ_j , which in our notation would be given for the n^{th} system by

$$\begin{aligned} \mu_N^n &= \beta_N^n, \quad \text{and} \\ \mu_j^n &= \beta_j^n + \sum_{i=j+1}^N (\lambda_i^n \wedge \mu_i^n) \gamma_{ij} + [\mu_{j+1}^n - \lambda_{j+1}^n]^+, \quad j = N-1, \dots, 1. \end{aligned} \quad (30)$$

Here $a \wedge b \equiv \min\{a, b\}$ and $[a]^+ \equiv \max\{a, 0\}$. This set-up explicitly recognizes potential imbalances in loading. The $(\lambda_i^n \wedge \mu_i^n)$ acknowledges that the departure rate from terminal i is bounded by the total rate at which capacity is provided there. The $[\mu_{j+1}^n - \lambda_{j+1}^n]^+$ accounts

for any excess vehicle capacity at $j + 1$, which then becomes available capacity at j . Crane's stability conditions are

$$\lambda_j^n - \mu_j^n < 0. \quad \text{for } j = 1, \dots, N. \quad (31)$$

We will see that our heavy traffic conditions imply that $\lambda_j^n - \mu_j^n \rightarrow 0$. To explore this relationship completely, write $\mu^n = (\mu_1^n, \dots, \mu_N^n)'$ and define vectors θ^n by

$$\theta^n = (I - \Gamma')\lambda^n - \beta^n, \quad (32)$$

Next, introduce an $N \times N$ matrix S having ones on the main diagonal and negative ones on the first superdiagonal; that is

$$S_{ij} = \begin{cases} 1. & \text{if } j = i \\ -1. & \text{if } j = i + 1 \\ 0. & \text{otherwise} \end{cases} \quad (33)$$

Then we have the following.

Proposition 1 *The stability conditions (31) hold if and only if*

$$S^{-1}\theta^n < 0$$

where the inequality is to be interpreted component-wise.

Proof. First note that S^{-1} exists and is given by

$$(S^{-1})_{ij} = \begin{cases} 1. & \text{if } j \geq i \\ 0. & \text{otherwise} \end{cases}$$

If (31) holds, then for each j we have $\lambda_j^n \wedge \mu_j^n = \lambda_j^n$ and $[\mu_j^n - \lambda_j^n]^+ = \mu_j^n - \lambda_j^n$. Using this and (32), we can rewrite (30) as

$$\begin{aligned} \lambda_N^n - \mu_N^n &= \theta_N^n, \quad \text{and} \\ \lambda_j^n - \mu_j^n &= \theta_j^n + (\lambda_{j+1}^n - \mu_{j+1}^n), \quad j = N - 1, \dots, 1. \end{aligned} \quad (34)$$

In vector form, this says that $S(\lambda^n - \mu^n) = \theta^n$ and the stability condition (31) is $\lambda^n - \mu^n < 0$. Thus, (31) implies that $S^{-1}\theta^n < 0$.

Conversely, the scalar form of $S^{-1}\theta^n < 0$ is

$$\sum_{j=i}^N \theta_j^n < 0, \quad i = 1, \dots, N. \quad (35)$$

At $i = N$, this gives $\theta_N^n < 0$. This implies component N of (31), because $\theta_N^n = \lambda_N^n - \beta_N^n = \lambda_N^n - \mu_N^n$ by (32) and (30). Suppose inductively that (31) holds for components $j = k + 1, \dots, N$. Then, reasoning as above, we see that the equations (34) are valid for $j = k, \dots, N$. Adding these equations gives, after cancellation,

$$\lambda_k^n - \mu_k^n = \sum_{j=k}^N \theta_j^n. \quad (36)$$

This implies $\lambda_k^n - \mu_k^n < 0$ by (35). \square

In the notation just introduced, our heavy traffic condition (29) says $\sqrt{n}\theta^n \rightarrow \theta$. We will see later that $S^{-1}\theta < 0$ emerges as a necessary condition for positive recurrence of our RBM limit process. In light of Proposition 1, this means that the process must be achievable as a limit from a sequence of stable systems. Condition (29) also implies that $\theta^n \rightarrow 0$. Following the steps leading to (36) above, this gives $\lambda^n - \mu^n \rightarrow 0$ in Crane's set-up, as asserted earlier.

4 The Main Theorem

We are now ready to state our main theorem. Define rescaled versions of the vector processes from our system state description as follows:

$$\begin{aligned} \tilde{X}^n(t) &= n^{-1/2}X^n(nt), & \tilde{Y}^n(t) &= n^{-1/2}Y^n(nt), & \tilde{Z}^n(t) &= n^{-1/2}Z^n(nt), \\ \tilde{E}^n(t) &= n^{-1/2}\hat{E}^n(nt), & \tilde{\epsilon}^n(t) &= n^{-1/2}\epsilon^n(nt). \end{aligned}$$

Theorem 1 *If the the basic functional central limit theorems (18) and the heavy traffic conditions (27)-(28) hold, then*

$$(\tilde{X}^n, \tilde{Z}^n, \tilde{Y}^n, \tilde{E}^n) \Rightarrow (X^*, Z^*, Y^*, E^*).$$

The limit processes are specified component-wise by

$$X_N^*(t) = A_N^*(t) - G_N^*(t) + \theta_N t, \quad \text{and} \quad (37)$$

$$\begin{aligned} X_j^*(t) &= A_j^*(t) - G_j^*(t) - \sum_{i=j+1}^N \left(U_{ij}^*(\lambda_i t) + \gamma_{ij} E_i^*(t) \right) \\ &\quad + \theta_j t - Y_{j+1}^*(t), \quad j = N-1, \dots, 1 \end{aligned} \quad (38)$$

where, for $j = N, \dots, 1$,

$$Y_j^*(\cdot) \text{ is continuous and nondecreasing, and} \quad (39)$$

$$Y_j^*(\cdot) \text{ increases only when } Z_j^*(t) = 0, \quad (40)$$

$$Z_j^*(t) = X_j^*(t) + Y_j^*(t) \quad \text{and} \quad (41)$$

$$E_j^*(t) = A_j^*(t) - Z_j^*(t). \quad (42)$$

Remark 1. The key limit process is Z^* , the limit for the vector queue length process, which we can characterize as a multidimensional RBM by the following steps. Substituting (38) into (41) gives

$$\begin{aligned} Z_j^*(t) &= A_j^*(t) - G_j^*(t) - \sum_{i=j+1}^N \left(U_{ij}^*(\lambda_i t) + \gamma_{ij} E_i^*(t) \right) \\ &\quad + \theta_j t - Y_{j+1}^*(t) + Y_j^*(t). \end{aligned} \quad (43)$$

To rewrite this in vector form, introduce the vector process F^* defined by

$$F^*(t) = \sum_{i=1}^N U_i^*(\lambda_i t). \quad (44)$$

Then F^* is a zero drift vector Brownian motion with covariance matrix

$$\Sigma_F = \lambda_1 \Sigma_{U_1} + \dots + \lambda_N \Sigma_{U_N}.$$

Recalling the matrix S defined in (33), we can express (43) as

$$Z^*(t) = A^*(t) - G^*(t) - F^*(t) - \Gamma' E^*(t) + \theta t + S Y^*(t).$$

Substituting $E^*(t) = A^*(t) - Z^*(t)$, which is the vector form of (42), and collecting terms gives

$$(I - \Gamma') Z^*(t) = (I - \Gamma') A^*(t) - G^*(t) - F^*(t) + \theta t + S Y^*(t).$$

Because Γ is strictly upper triangular, $(I - \Gamma')$ is guaranteed to be invertible, so we can solve the preceding equation for $Z^*(t)$, obtaining

$$Z^*(t) = A^*(t) - (I - \Gamma')^{-1}(G^*(t) + F^*(t)) + \eta t + RY^*(t), \quad (45)$$

where

$$\eta = (I - \Gamma')^{-1}\theta \quad \text{and} \quad R = (I - \Gamma')^{-1}S. \quad (46)$$

Now setting

$$\xi(t) = A^*(t) - (I - \Gamma')^{-1}(G^*(t) + F^*(t)) + \eta t, \quad (47)$$

equation (45) becomes

$$Z^*(t) = \xi(t) + RY^*(t). \quad (48)$$

Observe that ξ is a vector Brownian motion process with drift vector η and covariance matrix

$$\Omega = \Sigma_A + (I - \Gamma')^{-1}(\Sigma_G + \Sigma_F)(I - \Gamma')^{-1}. \quad (49)$$

Therefore, equation (48), together with the properties (39)-(40) for Y^* , identifies Z^* as a multidimensional RBM with drift vector η , covariance matrix Ω and reflection matrix R .

Remark 2. The form of the reflection matrix $R = (I - \Gamma')^{-1}S$ has a nice interpretation. The first factor is attributable to customer effects; the second to vehicle effects. In a traditional queueing network, only the former would be present. The latter arise here because unused vehicle capacity at one terminal translates directly into available capacity at the next terminal downstream. In Section 5 we present a concrete example of these calculations.

Proof of Theorem. Given the recursive representations developed so far, the theorem can be proved in a straightforward manner by induction, starting at terminal $j = N$. By rescaling equation (22), we obtain the following expression for \tilde{X}_N^n :

$$\tilde{X}_N^n(t) = \tilde{A}_N^n(t) - \tilde{G}_N^n(t) + \sqrt{n}(\lambda_N^n - \beta_N^n)t.$$

It follows immediately from the basic f.c.l.t.'s for \tilde{A}_N and \tilde{G}_N , and the heavy traffic condition, that $\tilde{X}_N^n \Rightarrow X_N^* = A_N^* - G_N^* + \theta_N e$, where $\epsilon(t) \equiv t$. This establishes (37). Observe that X_N^*

is a one-dimensional Brownian motion process, and therefore continuous. It is easy to check by rescaling (7)-(9) that

$$\begin{aligned}\tilde{Y}_N^n(t) &= -\inf_{0 \leq s \leq t} \tilde{X}_N^n(t), \\ \tilde{Z}_N^n(t) &= \tilde{X}_N^n(t) + \tilde{Y}_N^n(t) \quad \text{and} \\ \tilde{E}_N^n(t) &= \tilde{A}_N^n(t) - \tilde{Z}_N^n(t).\end{aligned}$$

By the continuous mapping theorem, we have that $(\tilde{Y}_N^n, \tilde{Z}_N^n, \tilde{E}_N^n) \Rightarrow (Y_N^*, Z_N^*, E_N^*)$, jointly with the result for \tilde{X}_N^n , where

$$\begin{aligned}Y_N^*(t) &= -\inf_{0 \leq s \leq t} \{X_N^*(s)\}, \\ Z_N^*(t) &= X_N^*(t) + Y_N^*(t) \quad \text{and} \\ E_N^*(t) &= A_N^*(t) - Z_N^*(t).\end{aligned} \tag{50}$$

This expresses Y_N^* and Z_N^* in terms X_N^* via the familiar one-dimensional reflection mapping. Since X_N^* is continuous, the first equation here implies properties (39)-(40) for $j = N$. The above also verifies (41) and (42) at this terminal.

Now suppose by inductive hypothesis that the results of the theorem hold for terminals $N, N-1, \dots, j+1$, and consider the situation at terminal j . Rescaling equation (23) gives

$$\begin{aligned}\tilde{X}_j^n(t) &= \tilde{A}_j^n(t) - \tilde{G}_j^n(t) - \sum_{i=j+1}^N \tilde{D}_{ij}^n(t) \\ &\quad + \sqrt{n} \left(\lambda_j^n - \beta_j^n - \sum_{i=j+1}^N \lambda_i^n \gamma_{ij} \right) t - \tilde{Y}_{j+1}^n(t) + \tilde{\epsilon}_j^n(t),\end{aligned} \tag{51}$$

where

$$\tilde{D}_{ij}^n(t) = n^{-1/2} (U_{ij}(E_i^n(nt)) - \lambda_i^n \gamma_{ij} nt). \tag{52}$$

First note that our basic f.c.l.t.'s give $(\tilde{A}_j^n, \tilde{G}_j^n) \Rightarrow (A_j^*, G_j^*)$, and by the heavy traffic condition the deterministic factors multiplying t converge to θ_j . For the \tilde{D}_{ij}^n terms, note that $\tilde{E}_i^n \Rightarrow E_i^*$ for $i = j+1, \dots, N$, by inductive hypothesis. Combining this with the basic result for the \tilde{U}_{ij}^n , we can apply the standard Compound Process Functional Central Limit Theorem to (52) to obtain $\tilde{D}_{ij}^n \Rightarrow D_{ij}^*$, where

$$D_{ij}^*(t) = U_{ij}^*(\lambda_i t) + \gamma_{ij} E_i^*(t)$$

is a Brownian motion process. Also, the inductive hypothesis gives $\tilde{Y}_{j+1}^n \Rightarrow Y_{j+1}^*$, where the limit is continuous. Finally, rescaling (10) gives

$$0 \leq \sup_{0 \leq t \leq T} \tilde{\epsilon}_j^n(t) \leq n^{-1/2} MV$$

for any $T > 0$. It follows that $\tilde{\epsilon}_j^n \Rightarrow \zeta$, where $\zeta(t) \equiv 0$. Putting all these terms together in (51) gives $\tilde{X}_j^n \Rightarrow X_j^*$, with X_j^* as in (38). Furthermore, continuity in each term implies that X_j^* is itself continuous. Now the one-dimensional reflection mapping can be used as in (50) to get corresponding results for Y_j^* , Z_j^* and E_j^* , thereby establishing (39)-(42). \square

We conclude this section by recording a corollary which gives limit results for the detailed queue length processes Q_{ij}^n defined in (11)-(12). We define the usual rescalings

$$\tilde{Q}_{ij}^n(t) = n^{-1/2} Q_{ij}^n(nt).$$

Corollary 1 *Jointly with the results of Theorem 1, we have at terminals $i = 1, \dots, N$ that*

$$(\tilde{Q}_{i0}^n, \tilde{Q}_{i1}^n, \dots, \tilde{Q}_{i,i-1}^n) \Rightarrow (\gamma_{i0} Z_i^*, \gamma_{i1} Z_i^*, \dots, \gamma_{i,i-1} Z_i^*).$$

Remark. This is the state space collapse result mentioned in Section 2. It shows that in the limit, the queue lengths broken out by destination can be expressed as fixed multiples of the total queue length at each terminal. Thus, in the limit, each terminal can effectively be described in one dimension.

Sketch of proof. In fact, this result is already contained in Crane; what is missing there is the development of the vector limit process Z^* . The key to the corollary is to express the queue length processes (11) in terms of centered processes as

$$Q_{ij}^n(t) = \hat{U}_{ij}^n(A_i^n(t)) - \hat{U}_{ij}^n(E_i^n(t)) + \gamma_{ij}(\hat{A}_i^n(t) - \hat{E}_i^n(t)).$$

After rescaling, the difference between the first two terms will be seen to vanish in the limit. Since $\hat{A}_i^n - \hat{E}_i^n = Z_i^n$, the last term will give the desired result. See Crane for details. \square

5 A Simple Example

To illustrate the analysis presented so far in a concrete case, we consider a network which gives the simplest non-trivial example of our results. The network has $N = 2$ terminals, and is served by a single bus, which belongs to service group S_2 . Service group S_1 is empty, so we set $G_1(t) \equiv 0$. Observe that passenger destination information is completely determined by specifying the parameter γ_{21} . Thus, the building blocks for this model are the processes A_2, A_1, G_2 and U_{21} . Assume for simplicity that the arrival processes are independent. Then the basic limit processes A_2^*, A_1^*, G_2^* and U_{21}^* guaranteed by (18) are independent zero drift Brownian motions with variance parameters $\sigma_{A_2}^2, \sigma_{A_1}^2, \sigma_{G_2}^2$ and $\sigma_{U_{21}}^2 = \gamma_{21}(1 - \gamma_{21})$; here the expression for $\sigma_{U_{21}}^2$ follows from the binomial distribution. In what follows, we explicitly write out the components of the two-dimensional RBM Z^* given in (48) as the limit for the vector queue length process.

Following the development in Section 2, we have for this example

$$\Gamma = \begin{pmatrix} 0 & 0 \\ \gamma_{21} & 0 \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Then easy computations give

$$(I - \Gamma')^{-1} = \begin{pmatrix} 1 & \gamma_{21} \\ 0 & 1 \end{pmatrix}$$

and

$$R = (I - \Gamma')^{-1}S = \begin{pmatrix} 1 & -(1 - \gamma_{21}) \\ 0 & 1 \end{pmatrix}. \quad (53)$$

The basic Brownian motion processes described above enter into our calculations in vector form as

$$A^*(t) = \begin{pmatrix} A_1^*(t) \\ A_2^*(t) \end{pmatrix}, \quad G^*(t) = \begin{pmatrix} 0 \\ G_2^*(t) \end{pmatrix} \quad \text{and} \quad F^*(t) = \begin{pmatrix} U_{21}^*(\lambda_2 t) \\ 0 \end{pmatrix},$$

where F^* is defined as in (44). Following (47), the underlying Brownian motion ξ for our limit process is

$$\xi(t) = \begin{pmatrix} \theta_1 + \gamma_{21}\theta_2 \\ \theta_2 \end{pmatrix} t + \begin{pmatrix} A_1^*(t) - U_{21}^*(\lambda_2 t) - \gamma_{21}G_2^*(t) \\ A_2^*(t) - G_2^*(t) \end{pmatrix}.$$

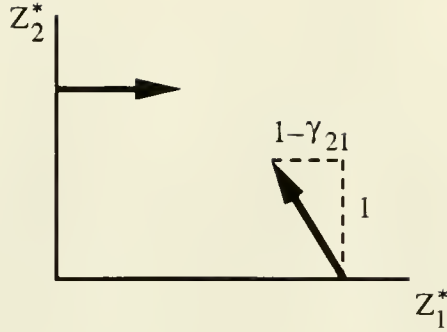


Figure 2: Directions of reflection for example network.

The drift vector η is apparent here, and the covariance matrix Ω given by (49) is easily calculated as

$$\Omega = \begin{pmatrix} \sigma_{A_1}^2 + \lambda_2 \gamma_{21}(1 - \gamma_{21}) + \gamma_{21}^2 \sigma_{G_2}^2 & \gamma_{21} \sigma_{G_2}^2 \\ \gamma_{21} \sigma_{G_2}^2 & \sigma_{A_2}^2 + \sigma_{G_2}^2 \end{pmatrix}. \quad (54)$$

On the interior of the nonnegative orthant \mathbb{R}_+^2 , the process $Z^* = \xi + RY^*$ behaves like an (η, Ω) Brownian motion. At the boundary $Z_j^* = 0$, it is subjected to an “instantaneous reflection,” which corresponds to a displacement in the direction given by the j^{th} column of R . This displacement, governed by the increase in Y_j^* , is of the minimum magnitude required to keep the process Z^* from leaving the orthant. The direction vectors are shown in Figure 2.

Hitting the boundary $Z_1^* = 0$ corresponds to unused capacity on a vehicle departing terminal 1. This has no effect on the “upstream” terminal 2; hence the direction of reflection is normal to the boundary. At the other boundary, the situation is more interesting. Hitting $Z_2^* = 0$ corresponds to unused capacity on a vehicle departing terminal 2. This will provide additional capacity to terminal 1, so we anticipate a negative impact on the queue there. As noted in Remark 2 following the main theorem, the result is the combination of two effects, corresponding to the factorization $R = (I - \Gamma')^{-1}S$. The *vehicle effect*, described by S , equates unused capacity at terminal 2 unit for unit with available capacity at terminal

1. This effect alone would suggest a 45-degree direction of reflection. To understand the *passenger effect*, recorded in $(I - \Gamma')^{-1}$, note that every occupied unit of capacity at terminal 2 represents an expected γ_{21} units of available capacity at terminal 1 due to disembarking passengers. This is foregone when $Z_2^* = 0$. Combining these two effects, it follows that the net decrease in the Z_1^* direction is $(1 - \gamma_{21})$ per unit of increase in the Z_2^* direction.

We now turn our attention to the equilibrium behavior of the RBM limit. Harrison and Williams (1987) describe conditions for an RBM process to have an equilibrium distribution whose density is a product of independent exponential densities. Consider an RBM Z^* on the nonnegative orthant having drift vector η , covariance matrix Ω and reflection matrix R , starting at $Z^*(0) = 0$. Then the condition $R^{-1}\eta < 0$ is necessary for positive recurrence. For our transportation model, it follows from (46) that $R^{-1}\eta = S^{-1}\theta$. Recall that at the end of Section 2, the condition $S^{-1}\theta < 0$ was seen to have a natural queueing theoretic interpretation. Harrison and Williams derive a skew symmetry condition which, together with the preceding condition, is necessary and sufficient for the product form result. When R has ones on the main diagonal, this skew symmetry condition is

$$2\Omega = RD + DR', \quad (55)$$

where $D = \text{diag}(\Omega)$ is the diagonal matrix constructed from the main diagonal of Ω . The reflection matrix for the transportation model always has ones on the diagonal because, as given by (46), it is the product of two upper triangular matrices, each having ones on the diagonal. With the data from (53) and (54) for our two station example, condition (55) reduces to the single off-diagonal equation

$$2\gamma_{21}\sigma_{G_2}^2 = -(1 - \gamma_{21})(\sigma_{A_2}^2 + \sigma_{G_2}^2). \quad (56)$$

This equation can only be satisfied if both sides are equal to zero, which leads to one of two uninteresting solutions. First, we could set $\sigma_{A_2}^2 = \sigma_{G_2}^2 = 0$. In other words, the basic processes at terminal 2 would have to be completely deterministic. Second, we could set $\gamma_{21} = 1$ and $\sigma_{G_2}^2 = 0$. In this case, all passengers boarding at terminal 2 have destination

terminal 1, and the vehicle circulation is deterministic. Furthermore, one can see that analogous problems will be encountered in general for an N terminal model. The off-diagonal condition relating terminals N and $N - 1$ will have exactly the form (56).

Given the lack of interesting solutions to the product form conditions, it is natural to ask what features of the transportation model make it differ from a conventional queueing network. The problem arises from the positive covariance between terminals 2 and 1, which results in the positive left hand side in equation (56). To understand the sign of the covariance, note that, away from the boundary, each unit of service by vehicles in group S_2 results in γ_{21} units of service capacity for terminal 1. Thus this unit of service effectively decreases the queue lengths at both terminals, which explains the positive covariance term. By contrast, in a traditional feedforward queueing network having the same basic topology, service completions at station 2 would decrease the queue there, but the portion of these customers routed to station 1 would increase the queue at that station. Hence the corresponding covariance term is negative in this case, and it becomes possible to satisfy the skew symmetry condition. Solutions for such models are covered by the results of Peterson.

References

- Crane, M.A. (1971). Limit theorems for queues in transportation systems. Ph.D. dissertation, Department of Operations Research, Stanford University.
- Crane, M.A. (1974). Queues in transportation systems II: an independently dispatched system. *J. Appl. Probab.* **11**, 145-158.
- Dai, J.G. and J.M. Harrison (1992). Reflected Brownian motion in an orthant: numerical methods for steady-state analysis. *Ann. Appl. Probab.* **2**, 65-86.
- Harrison, J.M. (1985). *Brownian Motion and Stochastic Flow Systems*. New York: John Wiley & Sons.

- Harrison, J.M. and R.J. Williams (1987). Multidimensional reflected Brownian motions having exponential stationary distributions. *Ann. Probab.* **15**, 115-137.
- Harrison, J.M. and V. Nguyen (1993). Brownian models of multiclass queueing networks: current status and open problems. *Queueing Systems Theory Appl.* **13**, 5-40.
- Iglehart, D.L. and W. Whitt (1970). Multiple channel queues in heavy traffic, I and II. *Adv. in Appl. Probab.* **2**, 150-157 and 355-364.
- Peterson, W.P. (1991). A heavy traffic limit theorem for networks of queues with multiple customer types. *Math. Oper. Res.* **16**, 90-118.
- Reiman, M.I. (1984). Open queueing networks in heavy traffic. *Math. Oper. Res.* **9**, 441-458.

MIT LIBRARIES



3 9080 00843941 3

BASEMENT

Date Due

DEC 23 1960

DEC 21 1960

Lib-26-67

