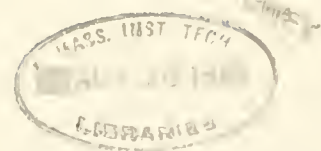


HD28
.M414
no. 2017-88



WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

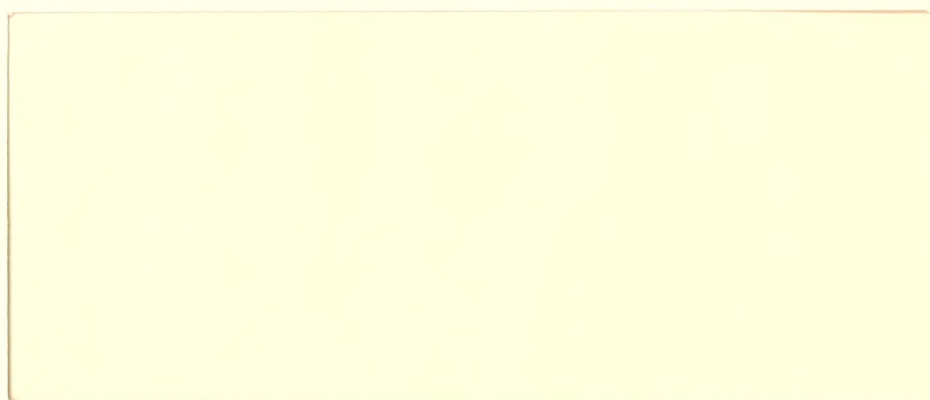
**ORDERING TANDEM QUEUES
IN HEAVY TRAFFIC**

Lawrence M. Wein

*Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139*

#2017-88

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139



ORDERING TANDEM QUEUES
IN HEAVY TRAFFIC

Lawrence M. Wein

*Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139*

#2017-88

MIT LIBRARIES
AUG 1 388
RECEIVED

ORDERING TANDEM QUEUES IN HEAVY TRAFFIC

Lawrence M. Wein

Sloan School of Management, Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

We consider the queueing system design problem of finding the best order of two or more single-server stations in tandem. Customers arrive to this system according to a renewal process, and each station has a different general service time distribution. The objective is to minimize expected equilibrium customer delay. An existing heavy traffic approximation is used to reduce the design problem to a directed travelling salesman problem. Using an interpolation scheme, we combine the heavy traffic results with existing light traffic results to analyze the design problem for a two-station tandem system with Poisson arrivals.

1. Introduction.

Consider a tandem queueing system with K single-server stations, each of which has an infinite capacity waiting room. Customers arrive according to a renewal process and each customer is served once at each station, with the order of the stations being the same for each customer. The interarrival times have finite mean λ^{-1} and finite squared coefficient of variation (variance divided by the square of the mean) c_0^2 . The service times at station $k = 1, \dots, K$ are independent and identically distributed random variables with finite mean μ_k^{-1} and finite squared coefficient of variation c_k^2 . The sequence of interarrival times and the sequence of service times at each station are assumed to be mutually independent.

Customers are served FIFO (first-in first-out) at each station. Also, it is assumed that the traffic intensity $\rho_k = \lambda/\mu_k < 1$ for $k = 1, \dots, K$, so that the system is stable. We consider the design problem of finding the order of the K stations that minimizes the expected equilibrium sojourn time per customer (or equivalently, the expected equilibrium customer delay or the expected equilibrium number of customers in the system).

The primary motivation for this problem stems from the design of a production line. In many cases, the set of tasks (either fabrication or assembly) that need to be performed for each job can be done in any order. Probably the most common example, as mentioned in Greenberg and Wolff (1987), is the insertion of components on a circuit board.

If each station has a deterministic service time distribution or each has an exponential service time distribution, then it is well known (Friedman (1965) and Weber (1979), respectively) that the sojourn time distribution is unaffected by the order of the stations. In the general case, however, the departure process from queues are not renewal processes, and the problem becomes much more difficult to analyze. Tembe and Wolff (1974) and Pinedo (1982a,b) obtained results for systems with deterministic and non-overlapping (i.e., ordered with probability one) service time distributions. More recently, Greenberg and Wolff (1987) consider two-station systems with Poisson arrivals in the light traffic limit, i.e., as the arrival rate λ goes to zero. They examine the interesting case where the service times of a particular customer at the different stations are not independent.

The only paper in the literature that suggests what to do in the general case is Whitt (1985). Applying the approximation methods for networks of queues developed in Whitt (1983), he describes the solution in several special cases and obtains four simple heuristic design principles. Although the design principles perform well in some cases, Whitt suggests that it is desirable to calculate the approximate value of the expected sojourn time for each of the $K!$ permutations, and to choose the order with the lowest value.

Our results are most closely related to that of Whitt. Using an approximation based on

results from heavy traffic theory developed by Reiman (1984) and Harrison and Williams (1987), we reduce the problem of ordering queues in series to a directed traveling salesman problem (TSP). The distance from city i to city j in the TSP corresponds to the customer delay incurred at station j by placing station i directly in front of station j . This value is $c_i^2/(\mu_j - \lambda)$, which has a clear intuitive meaning: in designing a tandem queueing system, one should attempt to precede the bottleneck stations with stations that have a low squared coefficient of variation of its service time distribution. This makes sense in heavy traffic, since the squared coefficient of variation of the interarrival times to a station will be very close in value to the squared coefficient of variation of the service times at the preceding station, and the lower that this value is, the less congestion will occur at the station. Since many efficient algorithms exist for solving TSP's (see, for example, Held and Karp (1970,1971)) our procedure offers a computationally less expensive alternative to calculating all $K!$ permutations using simulation or the Queueing Network Analyzer (QNA) package (Whitt (1983)). Furthermore, our results are consistent with all four heuristic design principles proposed in Whitt (1985), and, for the two-station case, our results and Whitt's are similar, but not identical.

Using the interpolation scheme proposed by Reiman and Simon (1987), we combine the heavy traffic results with Greenberg and Wolff's light traffic results to obtain the expected customer sojourn time for a two-station tandem system with Poisson arrivals. This interpolation approximation is appropriate for all values of the traffic intensity of the system. From this approximation, the optimal order of servers is found when one server has an exponential service time distribution and the other has a general service time distribution.

In manufacturing settings, jobs are often sent back to a station to be reworked if the operation performed on the job at the station was not successful. To accommodate this situation, the heavy traffic approximation is generalized to the case where immediate feedback of customers is allowed at each station; that is, a customer completing service at

station k will go to the next station with probability p_k , and will return to the end of the queue at station k with probability $1 - p_k$. The problem of ordering queues where feedback is allowed also reduces to a TSP.

The remainder of this paper is organized as follows. In Section 2, the Brownian model proposed by Harrison and Williams (1987) is used to calculate the approximate expected number of customers present in equilibrium for K stations in series. These results are used to reduce the design problem to a directed TSP. In Section 3, we compare our results with those obtained by Whitt (1985). In Section 4, the interpolation approximation is presented for a two-station system with Poisson arrivals. In Section 5, the heavy traffic approximation is generalized to include probabilistic feedback of customers.

2. The Brownian Approximation

In this section the Brownian model proposed by Harrison and Williams (1987) will be used to calculate the approximate expected number of customers present in equilibrium for K stations in series. This model is a refinement of the heavy traffic approximation of a generalized Jackson network (a Jackson network with general, rather than exponential, interarrival and service time distributions) derived by Reiman (1984). The Brownian approximation requires that the total load imposed on each station is approximately equal to its capacity. More precisely, we assume that there exists a large integer n such that

$$\max_{1 \leq k \leq K} \sqrt{n} |1 - \rho_k| < 1. \quad (1)$$

As a canonical example, one may think of ρ_k being between 0.9 and 1.0 for each station k , in which case $n = 100$ satisfies the *balanced heavy loading* condition (1).

In the Brownian approximation of the K -station tandem queueing system described in Section 1, the primary process of interest is the *scaled vector queue length* process

$Q^* = (Q_k^*)$ defined by

$$Q_k^*(t) = \frac{Q_k(nt)}{\sqrt{n}}, \quad t \geq 0, \text{ for } k = 1, \dots, K, \quad (2)$$

where $Q_k(t)$ is the number of customers queued and in service at station k at time t , and n is the large integer specified in (1). Under the balanced heavy loading condition (1), Harrison and Williams (1987) show that the scaled queue length process Q^* is well approximated by a process Z that has a unique stationary distribution. From equation (6) of that paper, it can be shown that this stationary distribution has a product-form density function if and only if

$$c_0^2 = c_1^2 = \dots = c_{K-1}^2. \quad (3)$$

It has been suggested in Harrison and Williams (1987) that the stationary distribution of Z may be approximated by this product form distribution even when (3) does not hold exactly, and this suggestion is incorporated into our approximation scheme. Under this assumption, Harrison and Williams show that the expected number of customers at station k in equilibrium is

$$\frac{\sigma_k^2}{2(\mu_k - \lambda)} \text{ for } k = 1, \dots, K, \quad (4)$$

where it follows from equation (27) of Reiman (1984) and equation (2.23) of Harrison and Williams (1987) that, for queues in series,

$$\sigma_k^2 = \lambda c_{k-1}^2 + \lambda c_k^2, \text{ for } k = 1, \dots, K. \quad (5)$$

It can be seen from equations (4)-(5) that our problem is to find the permutation $\pi(1, \dots, K)$ that minimizes

$$\sum_{k=1}^K \frac{c_{k-1}^2 + c_k^2}{\mu_k - \lambda}. \quad (6)$$

Since $\sum_{k=1}^K c_k^2 / (\mu_k - \lambda)$ is unaffected by the ordering of the queues, this problem is equivalent to a travelling salesman problem with $K + 1$ cities indexed by $k = 0, \dots, K$, where

city zero represents the outside of the queueing system and city k represents station k for $k = 1, \dots, K$. The asymmetric distance matrix $d = (d_{ij})$ for the TSP is defined by

$$d_{ij} = \begin{cases} 0 & \text{if } j = 0, \\ c_i^2 / (\mu_j - \lambda) & \text{otherwise.} \end{cases} \quad (7)$$

As mentioned in Section 1, the distance from city i to city j corresponds to the expected customer delay incurred at station j by having station i directly precede station j .

3. Comparison to Whitt's Results

In this section, we compare our results to some of those obtained by Whitt (1985). For the special case of two stations in series, Whitt shows that the stations should be ordered so that $\delta_1 \leq \delta_2$, where the quantity δ_k is defined by

$$\delta_k = (1 - \rho_k)(c_k^2 - c_0^2). \quad (8)$$

Going through the analogous calculations using equations (4)-(5), it follows that the stations should be ordered so that $\mu_1 \delta_1 \leq \mu_2 \delta_2$. Thus if the two queues have equal means, the two approximations schemes suggest the same ordering. When the two stations have nearly equal means, which is quite common in practice, the two schemes will often suggest the same ordering of the queues; this occurs in Example 1 of Whitt (1985).

To examine how much improvement in performance is possible, let us define $T(i, j)$ as the expected customer delay for the two-station tandem system, where station i precedes station j . From equations (4)-(5), it follows that

$$\frac{T(2, 1) - T(1, 2)}{T(2, 1)} = \frac{(c_0^2 - c_1^2)(\mu_1 - \lambda) + (c_2^2 - c_0^2)(\mu_2 - \lambda)}{(c_0^2 + c_2^2)(\mu_1 - \lambda) + (c_1^2 + c_2^2)(\mu_2 - \lambda)}. \quad (9)$$

We now consider three cases where it is desirable to have station 1 before station 2, in which case the quantity in (9) represents the relative improvement in performance by switching to the better ordering. If $\mu_1 = \mu_2$ and $c_1^2 \leq c_2^2$, then the relative improvement is

$$\frac{c_2^2 - c_1^2}{c_0^2 + c_1^2 + 2c_2^2}. \quad (10)$$

This quantity decreases as c_0^2 increases, and approaches an upper bound of 0.5 as $c_1^2 \rightarrow 0$ and $c_0^2 \rightarrow 0$; this upper bound is consistent with equation (11) in Whitt (1985). If $c_0^2 \leq c_1^2 = c_2^2 = c^2$ and $\mu_1 \leq \mu_2$, then the relative improvement is

$$\frac{(c^2 - c_0^2)(\mu_2 - \mu_1)}{c_0^2(\mu_1 - \lambda) + c^2(\mu_1 + 2\mu_2 - 3\lambda)}. \quad (11)$$

This quantity also decreases as c_0^2 increases and approaches an upper bound of 0.5 when $c_0^2 = 0$ and $\lambda \rightarrow \mu_1$. Thus, in both these cases, a 50% improvement in performance is maximal. When $c_0^2 \geq c_1^2 = c_2^2 = c^2$ and $\mu_1 \geq \mu_2$, then the relative improvement is

$$\frac{(c_0^2 - c^2)(\mu_1 - \mu_2)}{c_0^2(\mu_1 - \lambda) + c^2(\mu_1 + 2\mu_2 - 3\lambda)}. \quad (12)$$

This quantity decreases as $c^2 \rightarrow c_0^2$. As $c^2 \rightarrow 0$, this quantity approaches $(\mu_1 - \mu_2)/\mu_1(1 - \rho_1)$, which increases as ρ_1 increases and as the service rates of the stations become more imbalanced.

Whitt suggests the following heuristic design principles for K stations in series:

(P1) If $\mu_1 = \mu_2 = \dots = \mu_K$, then $c_1^2 \leq c_2^2 \leq \dots \leq c_K^2$ is desirable.

(P2) If $c_0^2 \leq c_1^2 = c_2^2 = \dots = c_K^2$, then $\mu_1 \leq \mu_2 \leq \dots \leq \mu_K$ is desirable.

(P3) If $c_0^2 \geq c_1^2 = c_2^2 = \dots = c_K^2$, then $\mu_1 \geq \mu_2 \geq \dots \geq \mu_K$ is desirable.

(P4) If $c_0^2 \leq c_1^2 \leq c_2^2 \leq \dots \leq c_K^2$, and $\mu_1 \leq \mu_2 \leq \dots \leq \mu_K$, then the order is desirable.

Whitt proves that (P1)-(P3) holds under his approximation scheme, but principle (P4), which includes (P1) and (P2) as special cases, remains a conjecture. We now prove that (P1)-(P3) are consistent with our approximation procedure, and that principle (P4) is true under our procedure.

Proposition 1. *Principles (P1)-(P3) are consistent with our approximation scheme.*

Proof. If $\mu_1 = \dots = \mu_K$, then the expected customer delay in equilibrium is proportional to, by equations (4)-(5),

$$c_0^2 + 2 \sum_{k=1}^{K-1} c_k^2 + c_K^2. \quad (13)$$

This is clearly minimized by choosing

$$c_K^2 = \max_{1 \leq k \leq K} c_k^2, \quad (14)$$

which is consistent with (P1).

If $c_1^2 = \dots = c_K^2 = c^2$, then the average customer delay in equilibrium is proportional to

$$\frac{c_0^2 + c^2}{\mu_1 - \lambda} + 2c^2 \sum_{k=2}^K \frac{1}{\mu_k - \lambda}. \quad (15)$$

If $c_0^2 \leq c^2$, one would choose

$$\mu_1 = \min_{1 \leq k \leq K} \mu_k \quad (16)$$

to minimize customer delay, and if $c_0^2 \geq c^2$, one would choose

$$\mu_1 = \max_{1 \leq k \leq K} \mu_k. \quad (17)$$

Equations (16) and (17) are consistent with principles (P2) and (P3), respectively. ■

Proposition 2. *Principle (P4) holds under our approximation scheme.*

Proof. Let the K stations be indexed by $k = 1, \dots, K$, and let $a_k = c_k^2$ for $k = 0, \dots, K$ and let $b_k = \mu_k - \lambda$ for $k = 1, \dots, K$. Let $a_0 \leq \dots \leq a_K$ and $b_1 \leq \dots \leq b_K$. An ordering of the stations is denoted by a permutation $\pi = (\pi(1), \dots, \pi(K))$, where station k is placed in position $\pi(k)$, for $k = 1, \dots, K$. Thus, we need to prove that the identity permutation, which is defined by $\pi(k) = k$ for $k = 1, \dots, K$, offers the least expected delay.

The proof is by induction on K , the number of stations. By (6), this permutation is desirable for $K = 2$ if

$$\frac{a_0}{b_1} + \frac{a_1}{b_2} \leq \frac{a_0}{b_2} + \frac{a_2}{b_1}. \quad (18)$$

This holds, since

$$b_1(a_1 - a_0) + b_2(a_0 - a_2) \leq (a_1 - a_0)(b_1 - b_2) \leq 0. \quad (19)$$

Now suppose (P4) holds for K stations, and let us consider a system with $K + 1$ stations. It follows by (6) and our inductive hypothesis that any permutation π_1 of the

$K+1$ stations that does not have $\pi_1(2) \leq \dots \leq \pi_1(K+1)$ achieves no smaller expected delay than the permutation π_2 that has $\pi_2(1) = \pi_1(1)$ and $\pi_2(2) \leq \dots \leq \pi_2(K+1)$. Similarly, the inductive hypothesis implies that the identity permutation of $K+1$ stations offers no larger expected delay than any permutation π such that $\pi(K) = K$ and $\pi(K+1) = K+1$.

Therefore, the only two permutations left to consider are $\pi = (K+1, 1, \dots, K)$ and $\pi = (K, 1, \dots, K-1, K+1)$. The identity permutation offers a smaller expected delay than $(K+1, 1, \dots, K)$ if

$$\frac{a_0}{b_1} + \frac{a_K}{b_{K+1}} \leq \frac{a_0}{b_{K+1}} + \frac{a_{K+1}}{b_1}. \quad (20)$$

This is true, since

$$b_{K+1}(a_0 - a_{K+1}) + b_1(a_K - a_0) \leq (a_0 - a_K)(b_{K+1} - b_1) \leq 0. \quad (21)$$

The identity permutation offers a smaller expected delay than $(K, 1, \dots, K-1, K+1)$ if

$$\frac{a_0}{b_1} + \frac{a_{K-1}}{b_K} + \frac{a_k}{b_{K+1}} \leq \frac{a_0}{b_K} + \frac{a_K}{b_1} + \frac{a_{K-1}}{b_{K+1}}. \quad (22)$$

This inequality holds if

$$b_K b_{K+1}(a_0 - a_K) + b_1 b_{K+1}(a_{K-1} - a_0) + b_1 b_K(a_K - a_{K-1}) \leq 0. \quad (23)$$

Reexpressing $(a_0 - a_K)$ by $(a_0 - a_{K-1} + a_{K-1} - a_K)$, it follows that (23) holds, since

$$b_{K+1}(a_0 - a_{K-1})(b_K - b_1) + b_K(a_{K-1} - a_K)(b_{K+1} - b_1) \leq 0. \blacksquare \quad (24)$$

4. An Interpolation Approximation

Using the interpolation scheme proposed by Reiman and Simon (1987), we combine the heavy traffic results of Section 2 with Greenberg and Wolff's light traffic results to obtain the expected customer sojourn time for a two-station tandem system with Poisson



arrivals. This interpolation procedure yields an approximation of the expected customer sojourn time for *all* values of the traffic intensity of the system. From this approximation, we derive the optimal order of servers when one server has exponentially distributed service times and the other has a general service time distribution.

Consider a two-station tandem queueing system with Poisson arrivals of rate λ . Let S_k be the service time for a random customer at station k , for $k = 1, 2$. We assume that S_1 and S_2 are independent, and let S_k have finite mean μ_k^{-1} and finite squared coefficient of variation c_k^2 . Let G be the distribution of S_2 and let the random variable S_{2e} have the distribution

$$P\{S_{2e} \leq t\} = \mu_2 \int_0^t 1 - G(s) ds. \quad (25)$$

The expected customer sojourn time at the first station is easily obtained from the Pollaczek-Khintchine formula. Let $f(\lambda)$ denote the expected customer sojourn time at the second station when the arrival rate to the system is λ . Greenberg and Wolff (1987) show that $f(0) = \mu_2^{-1}$ and

$$f'(0) = \mu_1^{-1} E[(S_2 - S_1)^+] + \mu_2^{-1} E[(S_{2e} - S_1)^+]. \quad (26)$$

Consider a normalized version of $f(\lambda)$ defined by $F(\lambda) = (\mu_2 - \lambda)f(\lambda)$. Then $F(0) = 1$ and

$$F'(0) = \mu_2 \mu_1^{-1} E[(S_2 - S_1)^+] + E[(S_{2e} - S_1)^+] - \mu_2^{-1}. \quad (27)$$

By the results in Section 2, if condition (1) holds then

$$\lim_{\lambda \rightarrow \mu_2} F(\lambda) = \frac{c_1^2 + c_2^2}{2}. \quad (28)$$

The interpolation method approximates $F(\lambda)$ by a second degree polynomial $\hat{F}(\lambda)$, and then approximates $f(\lambda)$ by $\hat{f}(\lambda) = (\mu_2 - \lambda)^{-1} \hat{F}(\lambda)$. By Reiman and Simon (1987), there exists a unique second order polynomial $\hat{F}(\lambda)$ that satisfies $\hat{F}(0) = 1$ and equations (27)-(28) (with F replaced by \hat{F}). This leads to the approximation of $f(\lambda)$ by

$$\hat{f}(\lambda) = (\mu_2 - \lambda)^{-1} \left(\left(\frac{c_1^2 + c_2^2}{2\mu_2^2} - \mu_1^{-1} E[(S_2 - S_1)^+] - \mu_2^{-1} E[(S_{2e} - S_1)^+] \right) \lambda^2 \right.$$



$$+ (\mu_2 \mu_1^{-1} E[(S_2 - S_1)^+] + E[(S_{2e} - S_1)^+] - \mu_2^{-1}) \lambda + 1 \Big). \quad (29)$$

Equation (29) and the Pollaczek-Khintchine formula can be used to compare the expected customer sojourn time for the two different orderings of two-station tandem queueing systems. We now derive the optimal order of servers in the special case where one server has exponential service times with mean μ^{-1} and the other server has a general service time distribution G . Let $G^*(\mu)$ be the Laplace transform of the general distribution and let the general service times have mean γ^{-1} and squared coefficient of variation c^2 . A similar theorem in light traffic was proved by Greenberg and Wolff (1987).

Proposition 3. *Expected customer sojourn time is minimized by putting a general server in front of an exponential server if and only if*

$$G^*(\mu) \leq \left(\frac{\gamma}{\mu + \gamma} \right) \left(\frac{2\mu - \lambda(c^2 + 1)}{2(\mu - \lambda)} \right). \quad (30)$$

Proof. As in Greenberg and Wolff, let system A be a tandem queueing system with Poisson arrivals, where the first server has exponential service times and the second server has general service times. Let system B be a queueing system with the servers in the opposite order. Let the exponential service times be denoted by X and the general service times by S . Then the expected sojourn time at the second server in system A is the same as the expected sojourn time at the first server in system B, since both servers receive Poisson arrivals. By equation (29), system A has a smaller expected total sojourn time than system B if and only if $(\mu - \lambda)^{-1}$ is less than

$$(\mu - \lambda)^{-1} \left(\left(\frac{c^2 + 1}{2\mu^2} - (\gamma^{-1} + \mu^{-1}) E[(X - S)^+] \right) \lambda^2 + \left(\left(\frac{\mu}{\gamma} + 1 \right) E[(X - S)^+] - \mu^{-1} \right) \lambda + 1 \right). \quad (31)$$

Since $E[(X - S)^+] = \mu^{-1} G^*(\mu)$, rearranging terms gives our result. ■

This result is consistent with three existing results. If both servers are exponential, then $c^2 = 1$ and $G^*(\mu) = \gamma/(\mu + \gamma)$, and the order of service does not matter, as was shown



in Weber (1979). If one server has constant service times and the other has exponentially distributed service times, then $c^2 = 0$ and $G^*(\mu) = e^{-\mu/\gamma}$. Since

$$e^{\mu/\gamma} \geq \frac{\mu + \gamma}{\gamma} \geq \frac{\mu + \gamma}{\gamma} \left(\frac{2(\mu - \lambda)}{2\mu - \lambda} \right), \quad (32)$$

it follows from (30) that the deterministic server should be first, which is a special case of a result by Tembe and Wolff (1974). Finally, if one takes $\lambda \rightarrow 0$ in equation (30), Proposition 3 reduces to a light traffic theorem of Greenberg and Wolff (1987).

5. Tandem Queues With Probabilistic Feedback

In this section, the model described in Section 1 is generalized to allow for immediate probabilistic feedback of customers. A customer completing service at station $k = 1, \dots, K$ will return to the end of the queue at station k with probability $1 - p_k$ (independent of previous history), and will move on to the next station (or possibly exit the system) with probability p_k . For ease of notation, we will define $p_0 = 1$. Now the traffic intensity at station k is $\rho_k = \lambda/p_k\mu_k$ for $k = 1, \dots, K$, and we assume that these values satisfy the balanced heavy loading condition (1). The expected number of customers at station k in equilibrium is again given by equation (4), where σ_k^2 is now defined by

$$\sigma_k^2 = \lambda(p_{k-1}(c_{k-1}^2 - 1) + 2 - p_k) + \lambda p_k c_k^2, \text{ for } k = 1, \dots, K. \quad (33)$$

When there is no feedback, then $p_k = 1$ for $k = 1, \dots, K$, and equation (33) reduces to equation (5). Since $\sum_{k=1}^K \lambda p_k c_k^2 / 2(\mu_k - \lambda)$ is unaffected by the ordering of the stations, the problem of ordering these stations again reduces to a travelling salesman problem, where the asymmetric distance matrix $d = (d_{ij})$ is now defined by

$$d_{ij} = \begin{cases} 0 & \text{if } j = 0, \\ p_i(c_i^2 - 1) + 2 - p_j / (\mu_j - \lambda) & \text{otherwise.} \end{cases} \quad (34)$$

Thus the expected customer delay incurred by having station i directly precede station j decreases as p_j increases, and increases with p_i only if $c_i^2 > 1$.

REFERENCES

- [1] Friedman, H. D., "Reduction Methods for Tandem Queueing Systems," *Oper. Res.*, 13 (1965), 121-131.
- [2] Greenberg, B. S. and R. W. Wolff, "Optimal Order of Servers for Tandem Queues in Light Traffic," to appear in *Management Sci.*, (1987).
- [3] Harrison, J. M. and R. J. Williams, "Brownian Models of Open Queueing Networks with Homogeneous Customer Populations," to appear in *Stochastics*, (1987).
- [4] Held, M. and R. M. Karp, "The Traveling-Salesman Problem and Minimizing Spanning Trees," *Oper. Res.*, 18 (1970), 1138-1162.
- [5] Held, M. and R. M. Karp, "The Traveling-Salesman Problem and Minimizing Spanning Trees: Part II," *Math. Programming*, 1 (1971), 6-25.
- [6] Pinedo, M., "On the Optimal Order of Stations in Tandem Queues," in *Applied Probability-Computer Science: The Interface*, Vol.II, R. L. Disney and T. J. Ott (Eds.), Birkhauser, Boston, (1982a), 307-325.
- [7] Pinedo, M., "Minimizing the Expected Makespan in Stochastic Flow Shops," *Oper. Res.*, 30 (1982b), 148-162.
- [8] Reiman, M. I., "Open Queueing Networks in Heavy Traffic," *Math. Oper. Res.* 9 (1984), 441-458.
- [9] Reiman, M. I. and B. Simon, "An Interpolation Approximation for Queueing Systems with Poisson Input," (1987).
- [10] Tembe, S. V. and R. W. Wolff, "The Optimal Order of Service in Tandem Queues," *Oper. Res.*, 24 (1974), 824-832.
- [11] Weber, R. R., "The Interchangeability of Tandem $M/M/1$ Queues in Series," *J. Appl. Prob.* 16 (1979), 690-695.



- [12] Whitt, W., "The Queueing Network Analyzer," *Bell System Tech. J.*, 62 (1983), 2779-2815.
- [13] Whitt, W., "Best Order for Queues in Series," *Management Sc.* 31 (1985), 475-487.

Date Due

FEB 19 1989

JUN 15 1989

NOV 29 1991

JAN 14 1994

MIT LIBRARIES



3 9080 005 358 814

