63-2-3

294 915

Western Management Science Institute

University of California • Los Angeles

$2.60

UNIVERSITY OF CALIFORNIA, LOS ANGELES

GRADUATE SCHOOL OF BUSINESS ADMINISTRATION

WESTERN MANAGEMENT SCIENCE INSTITUTE

September 28, 1962

Western Management Science Institute
Working Paper No. 17

Management Sciences Research Project
Research Report No. 79

QUEUING WHEN JOBS REQUIRE SEVERAL SERVICES WHICH NEED NOT BE SEQUENCED

By

Richard V. Evans

Acknowledgements. This work was supported partly by the Office of Naval Research under Task 047-003, and partly by the Western Management Science Institute under a grant from the Ford Foundation. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Summary

Several queuing models are considered in which a job requires a number of different services. The services are performed in any sequence, possibly simultaneously, by a number of servers. Steady state probabilities are found for two service jobs processed by two server systems. In the first system, the two servers work on the same job if they are working. In the second system, the servers may work on different jobs. If server one is working on job k, server two may be working on job k-1, k, or k+1 but not on k-2 or k+2. In these systems, the servers are specialists. In the third system, the servers may perform either of the two required services. If one has bogged down, the other server does both parts of each job.

(i)

1. Introduction

Queuing theory thus far has concentrated on the study of single server or parallel channel queues. Since the former is only a special case of the latter, study has focused on only one type of system element. What network theory has been developed concerns the processing of jobs requiring a sequence of services at different facilities. The purpose of this discussion is to focus attention on more general organization, in particular, the need to relax the serial nature of systems. The traditional models of production processes are the job shop and line. Both concepts emphasize the serial nature of operations; the former allows many sequences while the latter uses one fixed sequence. It is interesting to note that the concept of line production is associated with assembly lines historically. Perhaps a more natural model of an assembly type process is a pyramid where each stage of production combines the outputs of a number of different sub-processes.

The simplest possible case which provides at least some of the flavor of the more general organization is one in which a job requires a number of services by an equal number of servers with no sequencing restriction. Assuming that a job consists of n parts and that the system has n servers, the n servers are permitted to work simultaneously on the n parts of the job. Admitting the possibility of simultaneous work by the servers does not mean that this is necessary or desirable. A number of disciplines could be used to control such a system. At one extreme, one might require all n servers to be idle before a new job could start. At the opposite end of the spectrum of disciplines, one might have each server perform as much work as he could on every job in the system before becoming idle.

This class of systems introduces a second kind of queuing into the system, for not only are jobs waiting for processing but also there is an inventory of partially completed work. If the n servers are a sub-system of a larger processing system, there will be two inventories between successive stages of production corresponding to partially completed and fully completed jobs. The importance of the partially completed job inventory naturally varies greatly with the queue discipline which is used. In systems where each of the n servers is busy so long as there is any work that he can do, the inventory will be important if not excessive. The situation will be even more critical if priorities are used in the selection of the next job to be done by each server so that the different servers may process jobs in different orders.

Unfortunately, it appears that analytic work will not produce explicit results for any large number of these systems. Even the assumption of poisson input and negative exponential service times which assures that the steady state probabilities may be studied by solving systems of linear equations does not simplify the situation sufficiently. The major difficulty is that the state of the system becomes a two-dimensional quantity even in the simplest models. One approach to the study of these models is to show that solutions have a particular form involving some parameters which are the roots of a polynomial equation of high degree. The crucial point here is to show that the roots are distinct. Rather than follow this approach, this paper considers simple systems in which the parameters of the solution may be found explicitly. This facilitates the comparison of different types of systems which is so necessary in developing understanding of networks of queues.

2. One Job At A Time

The most restricted queue of the type considered requires that all n servers work on the same job if they work at all. This minimizes the partially completed jobs inventory and also the capacity of the system. For want of a better measure, the maximum load for which a steady state exists will be used as the measure of capacity. Having already assumed negative exponential service times, if one further assumes that the service rates are equal, he has a particularly simple system. The system is completely equivalent to a single channel system with a special negative exponential phase service process. The service process is not Erlang but a close relative. The relation of the job service time to the server service time is:

prob (job service time $\leq$ t) = [prob (server service time $\leq$ t)]^n

For the negative exponential with mean μ for each server this is:

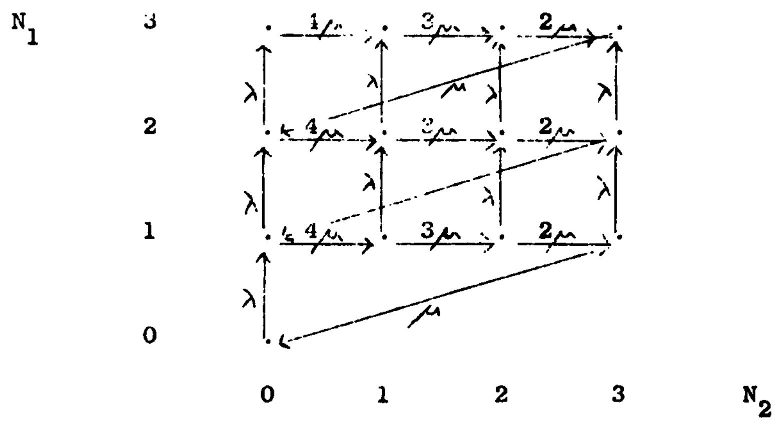
$$F(t) = [1-e^{-\mu t}]^n$$

using F(t) to denote the distribution of system service time (T). The expected service time is given by:

$$E(T) = \int_0^\infty tn(1-e^{-\mu t})^{n-1}\mu e^{-\mu t}dt = \sum_{i=1}^{n} \frac{1}{i\mu}$$

This suggests the rational for the construction of the equivalent phase service; n phases with the mean of the i'th phase $\frac{1}{(m+1-i)\mu}$ or with service rate $(m+1-i)\mu$. The first phase is the time to the completion of the first service regardless of which server finishes. The second phase is the time between the first and second completions, again regardless of which servers finish, etc. The first phase is then the minimum of n independent negative exponentials which is, of course, a negative exponential with mean $n\mu$. The second phase is similar due to the forgetful property of the negative exponential but only n-1 servers are functioning. The process continues until the last phase is the extra time required by the slowest server.

Analysis of the queue in this system presents no great difficulty. The situation is best understood from a picture. Let N_1 be the number of jobs in the system, including the one in service in conventional style, and N_2 be the number of services which have been completed. The state of the system is the two dimensional random variable (N_1, N_2). The diagram for definiteness is for a four-server system and only for the first three values of N_1 which is assumed to have no bound.



The normal queuing equations can easily be constructed for an m server system.

$$
\begin{pmatrix} dP_{n,0}(t)/dt \\ \\ \\ dP_{n,m}(t)/dt \end{pmatrix} = \lambda I \begin{pmatrix} P_{n-1,0}(t) \\ \\ \\ P_{n-1,m}(t) \end{pmatrix} + \begin{pmatrix} (-\lambda-m\mu) & & 0 \\ m\mu & (-\lambda-(m-1)\mu) & 0 \\ & (m-1)\mu & 0 \\ 0 & 0 & (-\lambda-\mu) \end{pmatrix} \begin{pmatrix} P_{n,0}(t) \\ \\ \\ P_{n,m}(t) \end{pmatrix}
$$

$$
+ \begin{pmatrix} 0 & 0 & 0 & 0 & \mu \\ 0 & 0 & 0 & 0 & 0 \\ & & & 0 & 0 \\ 0 & & & 0 & 0 \end{pmatrix} \begin{pmatrix} P_{n+1,0}(t) \\ \\ \\ P_{n+1,m}(t) \end{pmatrix}
$$

In steady state, the left side is, of course, the zero vector. Furthermore, by summing equations one easily verifies that:

$$
\mu P_{n+1,m} = \lambda \sum_{i=0}^{m-1} P_{n,i}
$$

using lower case p's for steady state probabilities. With this substitution and using P_n for the vector with components $p_{n,i}$, $i = \ldots m$.

$$
\lambda I P_{n-1} = \begin{pmatrix} m\mu & -\lambda & -\lambda & -\lambda \\ -m\mu & (+\lambda+(m-1)\mu) & 0 & 0 \\ 0 & -(m-1)\mu & (+\lambda+(m-2)\mu) & 0 \\ 0 & 0 & -(m-2)\mu & 0 \\ \\ 0 & 0 & 0 & (\lambda+\mu) \end{pmatrix} P_n
$$

This format of the recursion relationship corresponds to $\lambda_i p_i = \mu_{i+1} p_{i+1}$ for the simple queue which permits direct transitions only among neighboring states of the system. The relationship is in good form for numerical work, since the operations of matrix inversion and multiplication are readily available on computing machines.

Further work on the relationship would require that the matrix on the right side of the equation have an easily obtainable inverse in terms of the symbolic service and arrival rates. This step is really not important, however, for it is really the representation of the nth power of the inverse which is needed. If the eigen values of the inverse are all distinct, the representation of the nth power takes on a simple form. The eigen value problem is thus at the heart of finding a simplification of the relationship.

The same problem precisely is found if one uses a generating function approach. Let:

$$\pi_i = \sum_{n=0}^{\infty} p_{n,i} z^n \qquad \text{and} \qquad \pi^t = (\pi_0, \pi_1, \quad \pi_{m-1})$$

Then by multiplying the steady state equations by the appropriate power of z and summing one obtains:

$$\begin{pmatrix} \lambda z-\lambda-m\mu & 0 & 0 & 0 & \mu/z \\ m\mu & \lambda z-\lambda-(m-1)\mu & 0 & 0 & 0 \\ 0 & (m-1)\mu & \lambda z-\lambda-(m-2)\mu & 0 & 0 \\ & & & & \\ 0 & & & & \lambda z-\lambda-\mu \end{pmatrix} \pi = \begin{pmatrix} -m\mu p_{0,0} \\ +m\mu p_{0,0} \\ 0 \\ \\ 0 \end{pmatrix}$$

for z=1 the matrix on the left is singular. This difficulty can be repaired by replacing one equation by the sum of the equations divided by (z-1). The sum of the equation is:

$$[\lambda(z-1), \lambda(z-1), \ . \ . \ . \ , (\lambda(z-1) - \tfrac{\mu}{z}(z-1))] \ \pi = 0$$

or $\qquad [\lambda, \lambda, \ . \ . \ . \ , (\lambda - \frac{\mu}{z})] \ \pi = 0$

This is the same device used to reduce the original recurrence relationship involving three groups of probabilities to one involving only two groups. If the generating functions are to assist, one must solve:

$$\begin{pmatrix} \lambda & \lambda & \lambda & (\lambda - \frac{\mu}{z}) \\ m\mu & \lambda z-\lambda-(m-1)\mu & 0 & 0 \\ 0 & (m-1)\mu & \lambda z-\lambda-(m-2)\mu & 0 \\ & & & \\ 0 & 0 & 0 & \lambda z-\lambda-\mu \end{pmatrix} \pi = \begin{pmatrix} 0 \\ +m\mu p_{0,0} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

This requires the inversion of the matrix on the left. If one wants not only the moments but also the probabilities, partial fraction expansions will be required. This is the eigen value problem in the previous approach in a slightly different form.

Although it is possible that there is a reasonable representation of the eigen values as functions of the number of service phases, the functions have not been discovered. For the purposes of this paper since the same difficulty appears in variations of this system, attention will be focused on the simplest possible case, two servers. In this case, the generating function equations are:

$$\begin{pmatrix} \lambda & \lambda \ \mu/z \\ 2\mu & \lambda z-\lambda-\mu \end{pmatrix} \pi = \begin{pmatrix} 0 \\ +2\mu p_{0,0} \end{pmatrix}$$

After some algebra this becomes:

$$\pi = \begin{pmatrix} \dfrac{2\mu(\mu-\lambda z)p_{0,0}}{\lambda^2 z^2-\lambda(\lambda+3\mu)+2\mu^2} \\[2ex] \dfrac{2\lambda\mu z p_{0,0}}{\lambda^2 z^2-\lambda(\lambda+3\mu)_z+2\mu^2} \end{pmatrix}$$

As usual, the $p_{0,0}$ can be eliminated by $\pi_0(1)+\pi_1(1) = 1$. Thus:

$$2\mu^2 p_{0,0} = 2\mu^2-3\lambda\mu \quad \text{or} \quad p_{0,0} = 1-\frac{3\lambda}{2\mu}$$

Since this is a probability, it must be between zero and one, which restricts λ/μ to be less than 2/3. Using ρ for λ/μ, the expected number of jobs in the system, $E(N_1)$, is found from:

$$E(N_1) = \pi_0'(1)+\pi_1'(1)$$

where the prime denotes differentiation with respect to z. Simplifying the result gives:

$$E(N) = \frac{\rho(3-\rho)}{2-3\rho}$$

Since $\pi_0(1)$ and $\pi_1(1)$ are the marginal probabilities in N_2, the expectation of N_2 is just $\pi_1(1)$.

$$E(N_2) = \left(\frac{2\lambda\mu}{2\mu^2-3\lambda\mu}\right)\left(\frac{2\mu-3\lambda}{2\mu}\right) = \rho$$

This is, of course, both the probability that a phase of a job has been completed and the expected number of completed phases.

Finally, the p's may be obtained:

$$\begin{pmatrix} p_{n,0} \\ p_{n,1} \end{pmatrix} = (1-3/2\rho)\left(\frac{\lambda}{2\mu}\right)^n \begin{pmatrix} \frac{\lambda-\mu+\sqrt{\ }}{2\sqrt{\ }}e_1^{\,n} + \frac{-\lambda+\mu+\sqrt{\ }}{2\sqrt{\ }}e_2^{\,n} \\ \frac{\lambda+3\mu+\sqrt{\ }}{\sqrt{\ }}e_1^{\,n-1} + \frac{-\lambda-3\mu+\sqrt{\ }}{\sqrt{\ }}e_2^{\,n-1} \end{pmatrix}$$

where $\sqrt{\ } = \sqrt{\lambda^2+6\lambda\mu+\mu^2}$ and

$$e_1 = \frac{\lambda+3\mu+\sqrt{\ }}{2\mu} \qquad e_2 = \frac{\lambda+3\mu-\sqrt{\ }}{2\mu}$$

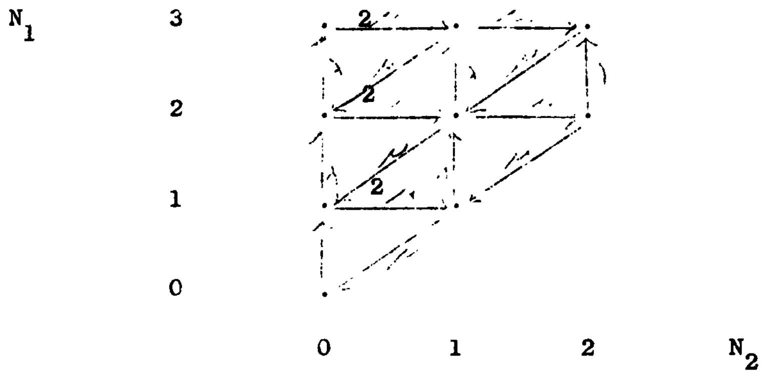
This is a rather awesome quantity of symbols. By substituting the values of e_1 and e_2, one can write these expressions in a form in which the square root has been eliminated. Thus, this representation is only a concise way of writing rational functions of polynomials of degree 2n+1 in λ and μ.

The comparison of these results with single server systems is reasonably straight forward. The capacity has been reduced to 2/3 of that of the single server queue with service rate μ or this same system with each server working as long as there are jobs waiting regardless of the progress of the other server. The same comment holds if one considers the single server system which processes both parts of the job. One might envision this type of system if the two servers could work co-operatively on the same item without interfering with each other. Such a system is an $E/E_2/1$ system. A strong word of warning needs to be injected before one comes to the conclusion that this form of organization is dominated in the sense that it must always produce more congestion. In real systems, it is generally far from true that division of a job into parts for processing by specialists corresponds to going from $E/E_2/1$ to the present system. The specialist doing half of the job often takes much less than one half the time. Thus, these results indicate some order of magnitude which the effects of specialization must have on the service time if congestion problems are not to change radically. Furthermore, this one measure of capacity does not tell the whole story about congestion.

3. Two Jobs At A Time

One would undoubtedly like to describe the general situation as a function of the maximum number of the jobs in the in-process inventory. If one is content to leave the results in terms of some parameters which are the roots of a polynomial equation, this can be accomplished. On the other hand, if no such parameters are allowed, the possibilities are severely restricted. In the former case, numerical work seems required to stimulate one's understanding of the system and the latter is the spirit of this paper. In the limiting case where the servers only stop work because of a zero queue of work of the appropriate type, the servers act independently. As far as the phases of jobs yet to be started, such a system looks like n independent queues. The capacity is obviously one. The size of the work in process, however, is not clear; for it cannot be studied by decomposing the system into n independent queues. At lease one special case can be solved in the sense that the distributions of the queue sizes can be found explicitly. This is the case of the two server system in which at most two jobs can be in process at one time. One important reason for this is that again summation of the equations will reduce the recurrence relation from one involving three groups of states to one involving only two groups.

Another way of stating this guarantee of easy solution is that it is possible to divide the states of the system into groups, which may be ordered, and that transitions from a higher group to a lower one are made only from one state in the higher group. That this is true in this case may be seen by reference to the state space diagram. Again N_1 is the number of jobs in the system and N_2 is the number of job parts which have been completed. Assuming equal service rates for the two servers, it is unnecessary to keep track of which server has completed which job. Only the first part of the diagram is shown, the same pattern is repeated for higher values of N_1.



It is no longer true that any grouping will have the required property. If one forms groups (n,0), (n,1), (n+1,2), the only state of the group from which one may move to a lower group is (n,1).

Again let P_n be the vector of steady state probabilities of the states in group n.

$$P_n = \begin{pmatrix} p_{n,1} \\ p_{n,0} \\ p_{n+1,2} \end{pmatrix}$$

For convenience, a rather odd ordering of the states has been used. The basic steady state equations are:

$$0 = \lambda P_{n-1} + \begin{pmatrix} -\lambda-2\mu & 2\mu & \mu \\ 0 & -\lambda-2\mu & 0 \\ 0 & 0 & -\lambda-\mu \end{pmatrix} P_n + \begin{pmatrix} 0 & 0 & 0 \\ \mu & 0 & 0 \\ \mu & 0 & 0 \end{pmatrix} P_{n+1}$$

There are, of course, appropriate modifications for the P_0 vector since states (0,1) and (1,2) do not exist. Also, it is not possible to have services in state (0,0), and state (1,1) allows only one server to be busy while in other states (n,1) both servers are busy. The special problems of the first few states are, of course, irrelevant to the fundamental question of how to simplify the basic recurrence relationship. By summing the general equations for all index values larger than n one finds:

$$\lambda p_{n,0} + \lambda p_{n,1} + \lambda p_{n+1,2} = 2\mu p_{n+1,1}$$

Using this $p_{n+1,1}$ can be eliminated from the general equations giving:

$$\lambda P_{n-1} = \begin{pmatrix} \lambda+2\mu & -2\mu & -\mu \\ -\lambda/2 & \lambda/2+2\mu & -\lambda/2 \\ -\lambda/2 & -\lambda/2 & \lambda/2+\mu \end{pmatrix} P_n$$

The inversion of a three by three by hand although tedious is nevertheless quite feasible. This gives: $P_n = RP_{n-1}$

where
$$R = \begin{pmatrix} \lambda/2\mu & \lambda/2\mu & \lambda/2\mu \\ \dfrac{\lambda(\lambda+\mu)}{2\mu^2(3\lambda+4\mu)} & \dfrac{\lambda^2+3\lambda\mu+4\mu^2}{2\mu^2(3\lambda+4\mu)} & \dfrac{\lambda(\lambda+3\mu)}{2\mu^2(3\lambda+4\mu)} \\ \dfrac{\lambda(\lambda+2\mu)}{2\mu^2(3\lambda+4\mu)} & \dfrac{\lambda(\lambda+4\mu)}{2\mu^2(3\lambda+4\mu)} & \dfrac{\lambda^2+4\lambda\mu+8\mu^2}{2\mu^2(3\lambda+4\mu)} \end{pmatrix}$$

The equations for the probabilities of the first few states are:

$$\begin{array}{lllll}
0 = & -\lambda p_{0,0}+\mu p_{1,1} & & & \\
0 = & -(\lambda+\mu)p_{1,1} & +2\mu p_{1,0} & +\mu p_{2,2} & \\
0 = & \lambda p_{0,0} & -(\lambda+2\mu)p_{1,0} & & +\mu p_{2,1} \\
0 = & & & -(\lambda+\mu)p_{2,2} & +\mu p_{2,1} \\
0 = & \lambda p_{1,1} & & & -(\lambda+2\mu)p_{2,1} \quad +\mu p_{3,1}
\end{array}$$

By subtracting the first from the second, the second, third and fourth equations can be made to follow the general recurrence relationship except that P_0 must be redefined. Let $P_0^* = (p_{0,0}, p_{0,0}, 0)^t$. Thus P_n may be expressed in terms of P_0^*.

$$P_n = R^n P_0^* \qquad n > 1$$

Unfortunately $p_{0,0}$ is still an unknown in this expression. Although further transformations of this relationship are possible, they do not appear to stimulate one's intuition. There are three real distinct eigen values of R, but they are not simple in form. The cubic equation which they satisfy does not factor easily.

The evaluation of $p_{0,0}$ and the mean number of jobs in the system require summation of the p's. For analytic work, generating functions appear to be somewhat easier than direct methods of attack. Again define

$$\pi_i = \sum_{n=0}^{\infty} p_{n,i} z^n \qquad \text{and} \qquad \pi = (\pi_0, \pi_1, \pi_2)^t .$$

From the original form of the steady state equations one obtains:

$$\begin{pmatrix} \lambda z-\lambda-2\mu & \mu/z & 0 \\ 2\mu & \lambda z-\lambda-2\mu & \mu/z \\ 0 & \mu & \lambda z-\lambda-\mu \end{pmatrix} \pi + \begin{pmatrix} 2\mu & 0 \\ -2\mu & +\mu z \\ 0 & -\mu z \end{pmatrix} \begin{pmatrix} p_{0,0} \\ p_{1,1} \end{pmatrix} = 0$$

Remembering that the first steady state equation is $\lambda p_{0,0} = \mu p_{1,1}$, it is possible to eliminate $p_{1,1}$. Furthermore, the problems at z = 1 can be eliminated by replacing one equation by the sum of the equations divided by (z-1). Completing both operations one obtains:

$$\begin{pmatrix} \lambda z-\lambda-2\mu & \mu/z & 0 \\ \lambda & \lambda-\mu/z & \lambda-\mu/z \\ 0 & \mu & \lambda z-\lambda-\mu \end{pmatrix} \pi = p_{0,0} \begin{pmatrix} -2\mu \\ 0 \\ +\lambda z \end{pmatrix}$$

From this one obtains:

$$\pi = \frac{p_{0,0}}{d} \begin{pmatrix} \mu(\mu-\lambda z)(2\lambda z-3\lambda-4\mu) \\ \lambda z\{2\mu(\lambda z-\lambda-\mu)-(\lambda z-\mu)(\lambda z-\lambda-2\mu)\} \\ \lambda^2 z^2(\lambda z-\lambda-3\mu) \end{pmatrix}$$

where

$$d = \lambda^3 z^3-2\lambda^3 z^2-5\lambda^2\mu z^2+\lambda^3 z+8\lambda\mu^2 z+5\lambda^2\mu z-3\lambda\mu^2-4\mu^3$$

At z=1, π gives the marginal probabilities in N_2 in terms of $p_{0,0}$.

$$\pi(1) = \frac{p_{0,0}}{\mu(4\mu-5\lambda)} \begin{pmatrix} (\mu-\lambda)(4\mu+\lambda) \\ 4\lambda\mu - 2\lambda^2 \\ 3\lambda^2 \end{pmatrix}$$

Summing these probabilities and equating the sum to one gives $p_{0,0}$.

$$p_{0,0} = \frac{4\mu-5\lambda}{\lambda+4\mu}$$

The condition that operations are all valid is that $p_{0,0}$ be a probability. Since the numerator is always less than the denominator, one need only require $4\mu-5\lambda > 0$ or $\lambda/\mu < 4/5$. Thus the capacity of this system is 4/5 compared to the 2/3 found for the previous system.

The expected value of N_1 can be found from $\pi_0'(1)+\pi_1'(1)+\pi_2'(1)$. Carrying out this process using $\rho = \lambda/\mu$, one finds:

$$E(N_1) = \rho\frac{(12+5\ )(2-\rho)}{(4+\rho)(4-5\rho)}$$

Inspection perhaps is not sufficient to relate this to the value found in the previous model; however, when plotted as in figure 1, the anticipated reduction is clear.

The expected number of completed job parts still in the system, $E(N_2)$ may be computed as before.

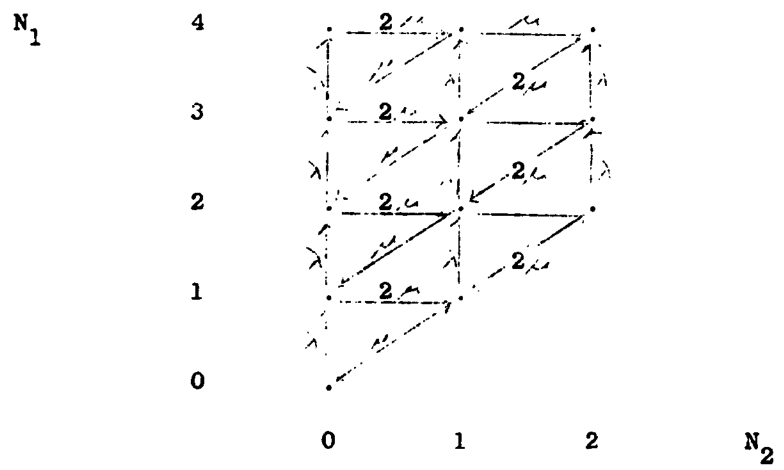
$$E(N_2) = 0\pi_0(1)+1\pi_1(1)+2\pi_2(1) = \rho\frac{(4+4\rho)}{4+\rho}$$

This is, of course, always greater than ρ which was the value in the previous system.

4. Unspecialized Servers

Another system which can be easily solved completely occurs if one changes the nature of the servers. Suppose the servers are not specialists, so that either server can process either part of a job. The algebra is only easy if one continues to consider only two-part jobs and two servers; thus, this restriction is maintained. In this case, as long as there is a part of a job waiting to be started, neither server is idle. This model undoubtedly is much less general than the previous ones in which the servers were specialists, but it provides some insight into congestion processes.

The definition of the state of the system variable is the same as in the previous section. In fact, the problem is only changed in that one transition rate, (n+1,2) to (n,1), has been increased from μ to 2μ since both servers are always busy if there is any work. This change introduces a symmetry into the state space diagram which was not previously present. Not only are the transition rates (n+1,1) to (n+1,2) and (n+1,1) to (n,0) equal but also the rates (n,0) to (n,1) and (n+1,2) to (n,1) are now equal. Thus, there is a symmetry in the role of (n,0) and (n+1,2). State Space Diagram:



The existance of state (0,0), however, prevents the symmetry from being perfect which would permit a reduction in the number of states to be considered.

The solution process follows that of the previous model. The recurrence relationship is:

$$0 = \lambda P_{n-1} + \begin{pmatrix} -\lambda-2\mu & 2\mu & 2\mu \\ 0 & -\lambda-2\mu & 0 \\ 0 & 0 & -\lambda-2\mu \end{pmatrix} P_n + \begin{pmatrix} 0 & 0 & 0 \\ \mu & 0 & 0 \\ \mu & 0 & 0 \end{pmatrix} P_{n+1}$$

Only the μ's in column three of the second co-efficient matrix have been changed. Summing equations this reduces to:

$$\lambda P_{n-1} = \begin{pmatrix} \lambda+2\mu & -2\mu & -2\mu \\ -\lambda/2 & (\lambda+4\mu)/2 & -\lambda/2 \\ -\lambda/2 & -\lambda/2 & (\lambda+4\mu)/2 \end{pmatrix} P_n$$

The matrix may be inverted giving:

$$P_n = \begin{pmatrix} \frac{\lambda}{2\mu} & \frac{\lambda}{2\mu} & \frac{\lambda}{2\mu} \\ \frac{\lambda^2}{8\mu^2} & \frac{4\lambda\mu(\lambda+2\mu)+\lambda^3}{8\mu^2(\lambda+2\mu)} & \frac{\lambda^2(\lambda+4\mu)}{8\mu^2(\lambda+2\mu)} \\ \frac{\lambda^2}{8\mu^2} & \frac{\lambda^2(\lambda+4\mu)}{8\mu^2(\lambda+2\mu)} & \frac{4\lambda\mu(\lambda+2\mu)+\lambda^3}{8\mu^2(\lambda+2\mu)} \end{pmatrix} P_{n-1}$$

For this case, the eigen values are reasonably concise expressions and distinct; so the recurrence relationship may be put in a form that facilitates expressing P_n in terms of P_1.

$$P_n = TDT^{-1}P_{n-1}$$

where

$$T = \begin{pmatrix} -1+\frac{r}{\lambda} & -1-\frac{r}{\lambda} & 0 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \quad D = \begin{pmatrix} \frac{\lambda(\lambda+4\mu+r)}{8\mu^2} & 0 & 0 \\ 0 & \frac{\lambda(\lambda+4\mu-r)}{8\mu^2} & 0 \\ 0 & 0 & \frac{\lambda}{\lambda+2\mu} \end{pmatrix}$$

$$T^{-1} = \begin{pmatrix} \frac{\lambda}{2r} & \frac{\lambda}{4r}+\frac{1}{4} & \frac{\lambda}{4r}+\frac{1}{4} \\ \frac{-\lambda}{2r} & \frac{-\lambda}{4r}+\frac{1}{4} & \frac{-\lambda}{4r}+\frac{1}{4} \\ 0 & +\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \quad P_{n-1}$$

using $\surd$ for $\sqrt{\lambda(\lambda+8\mu)}$.

Now the equations for the first few states are again special.

$$
\begin{array}{lllllll}
0 = & -\lambda p_{0,0} & +\mu p_{1,1} & & & \\
0 = & & -(\lambda+\mu)p_{1,1} & +2\mu p_{1,0} & +2\mu p_{2,2} & \\
0 = & \lambda p_{0,0} & & -(\lambda+2\mu)p_{1,0} & & +\mu p_{2,1} \\
0 = & & & & -(\lambda+2\mu)p_{2,2} & +\mu p_{2,1}
\end{array}
$$

By subtracting the first equation from the second, one makes these initial equations like the general ones as far as the P_n and P_{n+1} terms are concerned. Define P_0^* as $(p_{0,0}, p_{0,0}, 0)^t$. The recurrence relations can now be written as $P_n = TD^nT^{-1}P_0^*$. The reader unfamiliar with this result for the nth power of a matrix can verify that $(TDT^{-1})^2 = TDT^{-1}TDT^{-1} = TD^2T^{-1}$. An induction on n completes the result. Raising a diagonal matrix to the nth power merely requires raising the elements to the nth power. The expression may be further simplified by performing the multiplication and rearranging the result in the form:

$$
P_n = p_{0,0}\begin{pmatrix} \frac{3\lambda+\surd}{4\surd} & \frac{-3\lambda+\surd}{4\surd} & \frac{1}{2} \\ \frac{3\lambda+\surd}{4\surd}\left(\frac{-\lambda+\surd}{\lambda}\right) & \frac{-3\lambda+\surd}{4\surd}\left(\frac{-\lambda-\surd}{\lambda}\right) & 0 \\ \frac{3\lambda+\surd}{4\surd} & \frac{-3\lambda+\surd}{4\surd} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} [\frac{\lambda}{8\mu^2}(\lambda+4\mu+\surd)]^n \\ [\frac{\lambda}{8\mu^2}(\lambda+4\mu-\surd)]^n \\ (\frac{\lambda}{\lambda+2\mu})^n \end{pmatrix}
$$

This result could, of course, have come from generating function analysis. This type of analysis has its problems although it is more mechanical. Using generating functions, one never finds the transformation which puts the recurrence relationship into simple form. Before finding the transformation, it was hoped that it might provide heuristics for use in other problems. The role of the symmetry noted previously is, of course, displayed in the eigen vectors which are the columns of matrix of the transformation T. Other than this, little insight seems possible from this example.

The generating functions are easily obtained from the last result because of the geometric structure. The generating functions are a useful method for summing the probabilities to evaluate $p_{0,0}$ and finding the expected value of N_1 simultaneously. To find the generating functions directly from the steady state equations, one may proceed as in the previous model. Again let π be the vector of generating functions $\pi_i(z)$, $i = 0, 1, 2$. The original form of the steady state equations gives:

$$\begin{pmatrix} \lambda z-\lambda-2\mu & \mu/z & 0 \\ 2\mu & \lambda z-\lambda-2\mu & 2\mu/z \\ 0 & \mu & \lambda z-\lambda-2\mu \end{pmatrix} \pi + \begin{pmatrix} 2\mu & 0 \\ -2\mu & \mu z \\ 0 & -\mu z \end{pmatrix} \begin{pmatrix} p_{0,0} \\ p_{1,1} \end{pmatrix}$$

Again $p_{0,0} = p_{1,1}$ and the sum of these equations can be divided by z-1. Thus, one must solve:

$$\begin{pmatrix} \lambda z-\lambda-2\mu & \mu/z & 0 \\ \lambda z & \lambda z-\mu & \lambda z-2\mu \\ 0 & \mu & \lambda z-\lambda-2\mu \end{pmatrix} \pi = \begin{pmatrix} -2\mu \\ 0 \\ \lambda z \end{pmatrix} \quad p_{0,0}$$

The result is:

$$\pi = \frac{p_{0,0}}{(\lambda+2\mu-\lambda z)(+4\mu^2-(\lambda^2+4\lambda\mu)z+\lambda^2 z^2)} \begin{pmatrix} 2\lambda^2\mu z^2-3\lambda^2\mu z-8\lambda\mu^2 z+4\lambda\mu^2+8\mu^3 \\ \lambda z(4\mu-\lambda z)(\lambda+2\mu-\lambda z) \\ \lambda^2 z^2(3\mu+\lambda-\lambda z) \end{pmatrix}$$

Summing the components of π evaluated at 1 and equating the sum to zero gives $p_{0,0}$.

Using $\rho = \lambda/\mu$

$$p_{0,0} = \frac{2-2\rho}{2+\rho}$$

All of these operations are legitimate if $\rho < 1$ or the system capacity is one.

The sum of the derivatives of the generating functions evaluated at one gives:

$$E(N_1) = \frac{\rho}{1-\rho} \left(\frac{12-\rho-2\rho^2}{8+4\rho}\right)$$

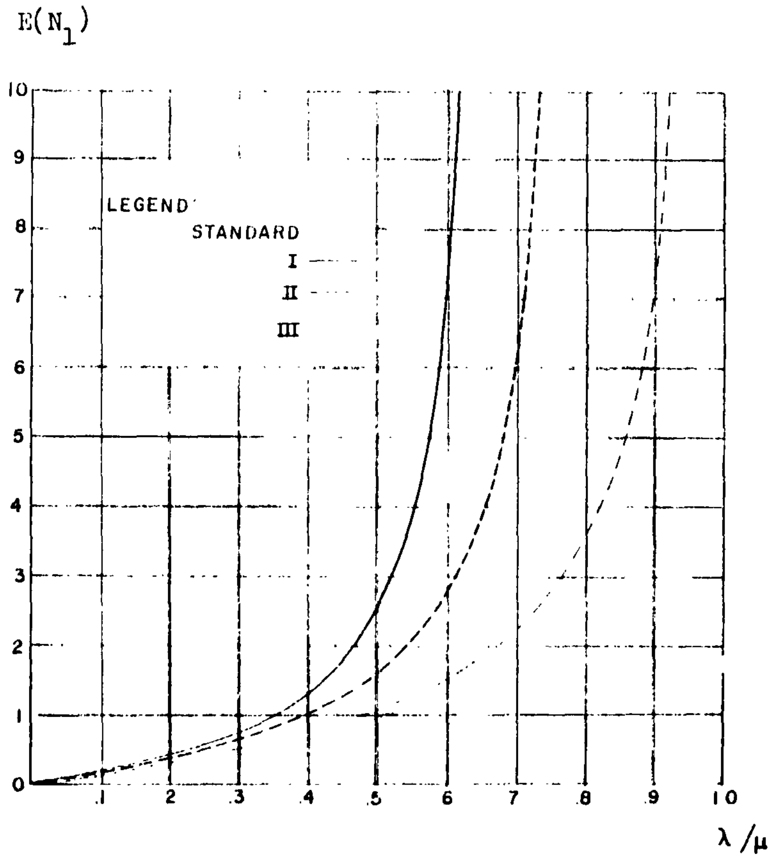


Figure 1. The expected number in the systems as a function λ/μ

This measure of system performance is interesting when compared to the single server system with load ratio = / . At low loads, the single server produces less congestion since $(12-\rho-2\rho^2)/(8+4\rho)$ approaches 3/2 as ρ approaches 0. This is the effect of each job having two phases and the job remaining in the system until the longer phase has been completed. This is precisely what one anticipates. Since $(12-\rho-2\rho^2)/(8+4\rho)$ approaches 3/4 as ρ approaches 1, the situation is reversed at high loads, and the two server system has less congestion. This result appears to be less intuitive, at least among those colleagues who have been bold enough to hazard a guess. The heuristic explanation of this phenomenon is the crucial nature of the job requiring excessive service under heavy load. In the single server case, there is no output from the system while such a job is in service. On the other hand, in the two-phase, two-server case, complete blockage requires that two very long phases occupy the servers simultaneously.

Compared to the two previous models, this model shows a marked reduction in congestion as expected. Figure 1 shows the expected number of jobs in the system for all three models discussed here and the standard single channel system.

As far as the number of completed job phases is concerned, one must first find the marginal probabilities in N_2 which are given by $\pi_1(1)$.

$$E(N_2) = \frac{\lambda(4\mu-\lambda)}{2\mu(\lambda+2\mu)} + \frac{6\lambda^2}{4\mu(\lambda+2\mu)} = \rho$$

This is again the minimum size of this type of inventory found in the first model. It is interesting that as far as $E(N_2)$ is concerned, the increase in the effective service rate tending to reduce congestion just balances the effect of increasing the maximum number of jobs in the system.

5. Conclusions

The specific results have been commented on as they were developed and need not be summarized. A word of caution does need to be added. Congestion is but one of many measures of system performance and perhaps of negligible importance in many cases. Furthermore, when one makes comparisons among models as has been done here, he compares them naturally with respect to the characteristics which happen to be under study. Among the models discussed here, there are qualitative differences of the greatest importance in any real design problem. Thus, one must be cautious in saying that one system is always preferable to another. What a theoretical investigation such as this hopes to explore is how strong is the preference of one over another. For example, there are undoubtedly many situations in which the increase in service rate due to specialization is so great that the system of model two is far preferable to that of model three as far as congestion is concerned. Thus, none of the systems here can be ruled out of design questions on the basis of what is reported here alone. Perhaps as the study of networks of queues continues, there will emerge dominant forms of organization. This time has not yet come.

The study of these systems was begun not merely because of their interest but rather because they are examples of a general class of queuing networks. This class is that for which the state of the system variable has more than one dimension and only one component is allowed to have infinitely many values. This permits the discussion of the states in groups. This sort of system results when one has a network system with only a finite number of waiting spaces internally but an unlimited number externally. The internal limitations may be either physical or the result of control rules. Also implicit in the concept of this class was the requirement that a group recurrence relationship exist at least from some point on. The order for the groups is, of course, provided by values of the one queue permitted to become infinite. The pressing question is how to reduce the infinite problem to one of solving a small number of linear equations for the first few states not following the general pattern and then to apply the recurrence relationship to produce the tails of the distribution. This is precisely what was done in all three models. The general form of the coefficient matrix of the steady state equations was

$$\begin{pmatrix} A^* & B^* & C^* & 0 & 0 & 0 \\ 0 & A & B & C & 0 & 0 \\ 0 & 0 & A & B & C & 0 \\ 0 & 0 & 0 & A & B & C \\ & & & & & \end{pmatrix}$$

The matrixes A*, A, B, and C are all square. The fundamental step in solution was to put this into a truncated form which gives the probabilities up to a scale factor.

A*	B*	C*
	A	B**

If one can truncate the system here, he can do it at any level with the same B**. In the cases studied, B** could easily be produced. One may think in terms of purely formal methods of eliminating C from the last group of equations, or he can provide an interpretation for B** which can assist in finding it. The entries of B** are the usual probabilities of single-step transitions plus some additional probabilities. The latter may be interpreted as the probability that the state of the system starting from one state in group n, goes to one of the states in n+1 and eventually comes back to another state in group n, not having been in group n in the meantime. Note that the structure assumed in the form of the steady state equations implies that the only communication between states in group n+1 and group n-1 is through states in group n. Thus, since the state of the system first moves to group n+1, the restriction that no value in group n is assumed in the interval is equivalent to saying that the state of the system remains in group n+1 or higher. The length of the time interval required for these transitions is immaterial. Another way to think about this is to say that a new truncated process is considered which agrees with the original one when the state of the system takes a value in group n or a lower group. When the state of the system in the original process is in groups higher than n, the new process is not defined and time is measured only when the process exists. Clocks are stopped when a transition from n to n+1 occurs and started when the opposite occurs. The two processes have the same steady state probabilities up to a scale factor. If transitions from group n+1 to group n can only occur from one state in group n+1 say (n+1,i), then the probabilities which convert B to B* are products of probabilities of leaving group n times the conditional probability of going from (n+1,i) to a particular state in group n given that a downward transition occurs from (n+1,i). These are almost immediately available from the state space diagram. In the last two models, they were of the form (1/2) since it was equally likely to go from (n+1,1) to (n+1,2) or to (n,0). If there had been two or more states in n+1 from which it was possible to move to states in group n, it would have been necessary to evaluate these probabilities considering the possible histories of any length starting from a state in group n and ending at another state in the group. This opens Pandora's box, for in general it represents a difficult counting problem typical of combinational problems.

The other difficulty in the algebra of these problems was putting the probabilities in the form of sums of geometric terms. This provides a severe limitation on analytic work since it involves finding roots of polynomials whose degree is equal to the group size. Unfortunately, in more complex systems this becomes confounded with the truncation problem or perhaps might as well be attacked simultaneously.

Thus, the amount of understanding of queuing networks which can be gained from purely analytic work appears to be quite limited. On the other hand, there certainly is a set of systems slightly more complicated than those considered here for which reasonable methods of numerical solution may be found.

W. Feller, An Introduction To Probability Theory And Its Applications, Second Edition, Wiley, New York, 1957.

D. R. Cox and W. L. Smith, Queues, Wiley, New York, 1961.

J. Reordan, Stochastic Service Systems, Wiley, New York, 1962.

BASIC DISTRIBUTION LIST FOR UNCLASSIFIED TECHNICAL REPORTS

PREPARED UNDER RESEARCH TASK NR 047-003

(CONTRACT nonr 233(73)

Head, Logistics and Mathematical Statistics Branch Office of Naval Research Washington 25, D.C.	3 copies
Commanding Officer Office of Naval Research Branch Office Navy # 100 Fleet Post Office New York, New York	2 copies
ASTIA Document Service Center Arlington Hall Station Arlington 12, Virginia	10 copies
Technical Information Officer Naval Research Laboratory Washington 25, D.C.	6 copies
Commanding Officer Office of Naval Research Branch Office 346 Broadway New York 13, N.Y. Attn: J. Laderman	1 copy
Commanding Officer Office of Naval Research Branch Office 1030 East Green Street Pasadena 1, California Attn: Dr. A. R. Laufer	1 copy
Institute for Defense Analyses Communciations Research Division von Neumann Hall Princeton, New Jersey	1 copy

Professor Russell Ackoff
Operations Research Group
Case Institute of Technology
Cleveland 6, Ohio 1

Professor K. J. Arrow
Serra House
Applied Mathematics &
Statistics Lab.
Stanford University
Stanford, California 1

Professor G. L. Bach
Carnegie Institute of Technology
Planning and Control of Industrial
Operations
Schenley Park
Pittsburgh 13, Pennsylvania 1

Professor A. Charnes
Mathematics Department
The Technological Institute
Northwestern University
Evanston, Illinois 1

Mr. L. W. Cohen
Department of Mathematics
University of Maryland
College Park, Maryland 1

Professor Donald Eckman
Director, Systems Research Center
Case Institute of Technology
Cleveland, Ohio 1

Dr. Lawrence E. Fouraker
Department of Economics
The Pennsylvania State University
State College, Pennsylvania 1

Dr. David Gale
Department of Mathematics
Brown University
Providence 12, Rhode Island 1

Mr. Murray Geisler
The RAND Corporation
1700 Main Street
Santa Monica, California 1

Professor L. Hurwicz
School of Business Administration
University of Minnesota
Minneapolis 14, Minnesota 1

Dr. James R. Jackson
Management Sciences Research Project
University of California
Los Angeles 24, California 1

Professor Samuel Karlin
Department of Mathematics
Stanford University
Stanford, California 1

Dr. C. E. Lemke
Mason Laboratory
Department of Mathematics
Rensselaer Polytechnic Institute
Troy, New York 1

Dr. W. H. Marlow
Logistics Research Project
The George Washington University
707 22nd Street, N.W.
Washington 7, D.C. 1

Professor Oskar Morgenstern
Econometric Research Program
Princeton University
92 A Nassau Street
Princeton, New Jersey 1

Dr. R. Radner
Department of Economics
University of California
Berkeley, California 1

Professor Stanley Reiter
Department of Economics
Purdue University
Lafayette, Indiana 1

Mr. J. R. Simpson
Bureau of Supplies & Accounts (Code W31)
Navy Department
Washington 25, D.C. 1

Professor J. Wolfowitz
Department of Mathematics
Lincoln Hall
Cornell University
Ithaca 1, New York 1

Professor A. W. Tucker
Department of Mathematics
Princeton University
Princeton, New Jersey 1

Professor G. B. Dantzig
Inst. of Engineering Research
University of California
Berkeley 4, California 1

MAILING LIST

American Bosch Division
American Bosch Arma Corp.
3664 Main Street
Springfield, Massachusetts
Attn: Mr. R. K. Balkeslee 1

Mr. H. T. Amrine, Head
Dept. of Industrial Engineering
Purdue University
Lafayette, Indiana 1

Applied Mathematics Laboratory
David Taylor Model Basin
Washington 7, D.C. 1

Dr. E. Leonard Arnoff
Director of Operations Research
Ernst & Ernst
1356 Union Commerce Bldg.
Cleveland 14, Ohio 1

Dr. Kenneth J. Arrow
Department of Economics
Stanford University
Stanford, California 1

Dr. Max Astrachan
Logistics Department
The RAND Corporation
1700 Main Street
Santa Monica, California 1

Professor David H. Blackwell
Department of Statistics
University of California
Berkeley 4, California 1

Dr. Frederick Bock
Armour Research Foundation
Technology Center
Chicago 16, Illinois 1

Dean L. M. K. Boelter
Department of Engineering
University of California
Los Angeles 24, California 1

Professor Glen Camp
Case Institute of Technology
Cleveland 6, Ohio 1

Chief, Bureau of Ships
Code 732 (A. E. Smith)
Navy Department
Washington 25, D.C. 1

Bureau of Supplies and Accounts
Code OW
Department of the Navy
Washington 25, D.C. 1

Commanding Officer
U. S. Naval Ammunition Depot Earle
Red Bank, New Jersey 1

Commanding Officer
Attn: Mr. Jules Vichness
Transportation Research & Development Command
Fort Eustis, Virginia 1

Professor R. W. Conway
Department of Industrial and Engineering Administration
Cornell University
Ithaca, New York 1

Professor G. Debreu
Cowles Commission for Research in Economics
Box 2125
Yale Station
New Haven, Connecticut 1

Dr. William Dorfman
CEIR
1200 Jefferson Davis Highway
Arlington 2, Virginia 1

Mr. H. F. Dickie
General Electric Company
Production Control Services
Room 2401, 570 Lexington Avenue
New York 22, New York 1

Director
National Security Agency
Fort George G. Meade, Maryland
Attn: C3141 (Room 2C087) 1

Director
Operations Evaluation Group
Office of Chief of Naval Operations,
OPO3EG
Navy Department
Washington 25, D.C. 1

Mr. Louis Doros
Management Engineering Staff
Bureau of Yards and Docks
Navy Department
Washington 25, D.C. 1

Engineering Library
University of California
Los Angeles 24, California
Attn: Mrs. Johanne E. Tallman 1

Professor Merrill M. Flood
Mental Health Research Institute
University of Michigan
Ann Arbor, Michigan 1

Professor George E. Forsythe
Computation Center
Stanford University
Stanford, California 1

Cornell Aeronautical Laboratory, Inc.
4455 Genessee Street
Buffalo 21, New York
Attn: Joseph P. Desmond, Librarian 1

Mr. Murray A. Geisler
The RAND Corporation
1700 Main Street
Santa Monica, California 1

General Electric Company
FPD Technical Information Center
Building 100
Cincinnati 15, Ohio 1

Mr. Ezra Glaser
Section 12.05
National Bureau of Standards
Washington 25, D.C. 1

Dr. R. E. Gomory
IBM Research Center
Box 218
Yorktown Heights, New York 1

Headquarters USAF
Attn: Mr. Edward G. Bunaway
Chief, AFASC-6F
Washington 25, D.C. 1

Alan J. Hoffman
I.B.M.
Yorktown Hgts., New York 1

Professor Richard S. Rudner
Department of Philosophy
Michigan State University
East Lansing, Michigan 1

Professor C. C. Holt
School of Industrial Administration
Carnegie Institute of Technology
Pittsburgh 13, Pennsylvania 1

Professor W. Grant Ireson
Industrial Engineering Department
Stanford University
Stanford, California 1

Mr. Gersaon Cooper, V.P.
Dunlap and Associates, Inc.
429 Atlantic Street
Stanford, Connecticut 1

Professor Neil H. Jacoby
School of Business Administration
University of California
Los Angeles 24, California 1

Mr. P. D. O'Donnell, Mgr.
Manufacturing Controls Hdqts. Dept.
Westinghouse Electric Corporation
3 Gateway Center 17-E
Pittsburgh 22, Pennsylvania 1

Mr. R. N. Lehrer
Professor & Chairman
Department of Industrial Engineering
The Technological Institute
Northwestern University
Evanston, Illinois 1

Professor S. B. Littauer
Department of Industrial Engineering
Columbia University
409 Engineering Building
New York 27, New York 1

Dr. Clifford J. Maloney
U. S. Army Chemical Corps
Biological Laboratories
Fort Detrick
Frederick, Maryland 1

C. H. Masland & Sons
Carlisle, Pennsylvania
Attn: Ast. Controller 1

Mr. Joseph Mehr
Ass't Planning Officer
U. S. Naval Training Device Center
Port Washington, L.I., New York 1

Mr. H. D. McLoughlin
7709 Beland Avenue
Los Anglees 45, California 1

Mr. F. F. Middleswart
E. I DuPont DeNemours & Co., Inc.
Engineering Department
Wilmington 98, Delaware 1

Mr. Marvin O. Miller
Code 250A
San Francisco Naval Shipyard
San Francisco 24, California 1

Dr. Richard A. Miller
4071 West 7th Street
Fort Worth 7, Texas 1

Mr. Michael Montalbano
Kaiser Steel Corporation
300 Lakeside Drive
Oakland 12, California 1

Dr. Jack Moshman
Corp. for Economic & Industrial Research
1200 Jefferson Davis Highway
Arlington 2, Virginia 1

Mr. G. Nadler
Dept. of Industrial Engineering
Washington University
St. Louis, Missouri 1

Professor R. R. O'Neill
Department of Engineering
University of California
Los Angeles 24, California 1

President
Naval War College
Attn: Mahan Library
Newport, Rhode Island 1

Mr. Joseph Romm
Staff Director
Vulnerability Analysis Division
OASD (I&L) Rm. 3D831, Pentagon
Washington, D.C. 1

Mr. R. W. Rutledge
c/o Colonial Sugar Refining Co. Ltd.
1 O'Connell Street
Sydney, AUSTRALIA 1

Dr. Melvin E. Salveson
16 Parish Road
New Canaan, Connecticut 1

Professor Andrew Schultz, Jr.
Sibley School of Mechanical Eng.
Dept. of Industrial & Eng. Administration
Cornell University
Ithaca, New York 1

Dr. Alexander Silverman
2930 Cathedral of Learning
University of Pittsburgh
Pittsburgh 13, Pennsylvania 1

Professor H. A. Simon
Head, Dept. of Industrial Administration
Carnegie Institute of Technology
Schenley Park
Pittsburgh 13, Pennsylvania 1

Dr. Otto J. M. Smith
University of California
Division of Electrical Engineering
Berkeley 4, California 1

Mr. Thomas P. Styslinger
Senior Staff Industrial Engineer
525 William Penn Place
Room 1824
Pittsburgh 30, Pennsylvania 1

Mr. C. B. Sung
Mechanical Development Department
Research Laboratories Division
Bendix Corporation
10½ Northwestern Highway
Southfield (Detroit) Michigan 1

Superintendent
U. S. Naval Postgraduate School
Monterey, California
Attn: Library 1

The John Hopkins University Library
Acquisitions Department
Baltimore 18, Maryland 1

Professor R. M. Thrall
University of Michigan
Department of Mathematics
274 W. Engineering Building
Ann Arbor, Michigan 1

Dr. C. B. Tompkins, Director
Numerical Analysis Research
University of California
405 Hilgard Avenue
Los Angeles 24, California 1

Professor Eduardo Valenzuela
P. O. Box 6646
Caracas, Venezuela 1