AD-A081 257 VIRGINIA POLYTECHNIC INST AND STATE UNIV BLACKSBURG --ETC F/G 12/1

QUEUEING NETWORKS.(U)

DEC 79 R L DISNEY N00014-77-C-0743

UNCLASSIFIED VTR-79-23 NL

1 OF 1

AD A081257

END

DATE FILMED

3-80

DDC

(12) LEVEL II

QUEUEING NETWORKS

Ralph L. Disney
Gordon Professor
Department of Industrial Engineering
and Operations Research
Virginia Polytechnic Institute
and State University
Blacksburg, Virginia 24061

TECH REPORT
VTR - 79-23
December 1979

This research was supported jointly by NSF Grant ENG77-22757 and by the Office of Naval Research Contract N00014-77-C-0743 (NR042-296).

Corrigendum

p. 2, Line 15 "moribind" should be "moribund".

p. 3, line 11, "in" should be "in-".

p. 5, line 6, "basic process" should be "basic processes".

p. 5, line 11, "exists" should be "exits".

p. 6, line 5 from bottom, "$[0,\infty)$" should be "$\{0,1,2,...\}$".

p. 9, line 12 from bottom, "Remembering" should be "Remember".

p. 9, line 3 from bottom, "out" should be omitted.

p. 10, line 2, "errormous" should be "enormous".

p. 11, line 10, insert "in the steady state" after "For fixed n," and change "i.i.d. random variables to independent random variables".

p. 17, line 6 from bottom, "family" should be "familiar".

p. 18, line 1 and 2, change the sentence "The busy period is in general..." to "The busy period sequence is also a sequence of i.i.d. random variables" and then delete the next sentence, "For the special case..."

p. 18, line 15, insert "an" between "to" and "empty".

p. 24, line 9, "sume" should be "sum".

p. 28, line 24, "research" should be "researched".

p. 31, Syski reference should be "Congestion" not "Conjection".

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report VTR 79-23	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Queueing Networks		5. TYPE OF REPORT & PERIOD COVERED Technical Rep[redacted]t, [redacted]
		6. PERFORMING ORG. REPORT NUMBER VTR-79-23
7. AUTHOR(s) Ralph L. Disney		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0743 NSF-ENG77-22757
9. PERFORMING ORGANIZATION NAME AND ADDRESS Dept. of Industrial Engineering & Operations Res. Virginia Polytechnic Inst. & State University Blacksburg, Virginia 24061		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR042-296
11. CONTROLLING OFFICE NAME AND ADDRESS Director, Statistics and Probability Program Mathematical & Information Sciences Division 800 N. Quincy St., Arlington, VA 22217		12. REPORT DATE Dec[redacted]79
		13. NUMBER OF PAGES 31
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 12) 35		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Queueing Networks
Survey

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The study of queueing networks is a rather recent addition to queueing theory. In this paper we will review about the last 20 years of developments in the area. Primary emphasis is placed on Jackson networks and the major contribution of Kelly. We discuss queue length processes, waiting time processes, busy period processes and departure processes. We will also note a few new studies that have appeared in the area of network flows. Topics that are not discussed in detail in the paper are briefly noted in the final section. The bibliography can serve as an introduction for further reading in the area.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406747

QUEUEING NETWORKS

Ralph L. Disney[1]

ABSTRACT. The study of queueing networks is a rather recent addition to queueing theory. In this paper we will review about the last 20 years of developments in the area. Primary emphasis is placed on Jackson networks and the major contribution of Kelly. We discuss queue length processes, waiting time processes, busy period processes and departure processes. We will also note a few new studies that have appeared in the area of network flows. Topics that are not discussed in detail in the paper are briefly noted in the final section. The bibliography can serve as an introduction for further reading in the area.

1. Introduction and Some Background

1.0 Introduction. In this paper we will briefly review some developments that have occurred over the last 20 years, in an area of applied mathematics called queueing network theory. The need for such applied work has been clear almost since the beginning of what is called queueing theory. Some of the work of A. K. Erlang (see Brockmeyer, et al. [1948]) seems to be pointing toward a study of interconnected systems of service centers. Certainly by the 1930's it was clear that such work was evolving (see chapters 7-10 in Syski [1960]). In most of these early studies, the developments appear to be closely tied to attempts to solve problems that occur principally in telephone switching systems.

Much of classical queueing theory (say up to about 1957 or 1960) was concerned with properties of Markov processes and except for names given to the parameters and processes it is difficult to separate out the peculiarities of queues from other Markov processes that were being used as models, for example,

[1]This research was supported jointly by NSF Grant ENG77-22757 and by the Office of Naval Research Contract N00014-77-C-0743 (NR 042-296). Distribution of this document is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

in biological modelling. In fact, many of the birth-death models of biological studies have interesting and useful counterparts in queueing theory. For this reason a carefully drawn history of queueing theory would be strained because of the interrelations with other fields. The useful thing to remember is that in its beginning phases the field was closely allied with attempts to solve real life problems.

The 1950's were perhaps the "golden age" of developments in the field. Under many impacts (including the introduction of more sophisticated and realistic models into problems that here to fore had been the domain of the industrial engineer) the field began to take on a life of its own. In much of the 1950-60 work and extending even to today, the work in the field veered more and more away from its applied bent. It was at this time that queueing theory started to be sufficiently arcane that many researchers were turned away from it. Easy queueing problems had been solved. Many papers were and had been published on special cases of birth-death processes. The field was moribind after about 1965-70.

Starting about 1965, the field of computer systems analysis began to delve into properties of time sharing and interconnected systems. Here the researchers encountered many of the same types of problems encountered earlier in telephone systems and production systems. It was natural, therefore, to use existing, known results. By 1970 and certainly by 1973 research into computers and computer systems had uncovered a large number of problems beyond the work in earlier queueing theory.

At the same time these developments in computer systems were taking place, new problems were occurring and old unsolved problems were becoming more pressing in telecommunication theory and production theory. New areas of application were evolving in military command and control modelling, disease modelling, military tactics modelling, public sector models for police fire department and medical emergency systems design and many others. All of these topics put demands on the queueing theory (and indeed other areas of probability and random processes) that theory was unable to handle. Systems (e.g. networks) of queues became areas of important study in many of the above application rather than single queues. Markov process theory did not provide a rich enough body of knowledge and new ideas evolved.

Largely because the applied topics put demands on existing theory that could not be accommodated quickly enough, much of applied work turned to Monte Carlo simulation of systems. But that placed a new load on the theory of Statistics as well as probability, random processes and queueing. Because of the relative quiescence of the 1957-70 period queueing theory has lost ground to applied topics.

1.1 Purpose. The purpose of this paper is to briefly summarize where we stand in the study of queueing network theory. Time and space do not permit any in depth discussion here. The bibliography should be consulted by the reader interested in "reading themselves into" the area. However, that reader should be forewarned that papers on these topics are spread over a large number of journals of the world. There is no one best place to look for research papers or to expect to follow diligently and thereby stay up with the field. For background into queueing theory the reader might consult Syski [1960] who does a masterful job summarizing work in queueing theory up to about 1960. Kleinrock's two volumes [1975; 1976] summarizes many known results in queueing theory and gives a nice discussion of many of the problems of queueing phenomena occurring in computer systems. There are no textbooks on the subject of queueing network theory though one can find some discussions concerning section 2, to follow, in most post 1975 texts. Material in section 4 is nicely presented but unfortunately it is in unpublished lecture notes by Prof. Frank Kelly. Material in section 5 coupled with the Disney [1975] reference therein is a fair review of those topics. Topics in section 6 have not been pulled together in any one place to the best of our knowledge so the reader is on his own there. References are provided herein but that list is far from complete.

1.2 Some Background, Notation and Symbolism. In order to embed our future discussion in the more classic field of queueing theory, we will present a brief review of a few ideas in basic queueing theory.

At an intuitive level queueing theory is concerned with problems arising in waiting lines. One supposes that there is something that can perform a needed service, a server. In the simpler cases it is assumed that the sequence of service times is a sequence of mutually independent random variables that are identically and non-negatively distributed. In queueing theory such a service process is called a G-type service process. In the theory of stochastic processes such a process is called a renewal process.

The demand for this service is usually modelled with the times between demands being the variables of interest. These times are assumed to be a sequence of independent, identically distributed random variables that are non-negative (another renewal process). Such an arrival process is called a GI-type arrival process. It is usual to assume that the arrival process and service time process are independent processes. Such a queueing system is called a GI/G/· queue.

Queueing or waiting occurs in such systems whenever an arrival occurs to find the server already engaged serving a previous arrival. If an arrival occurs and the server is not so engaged, that arrival immediately enters service, under the usual assumptions. (There are studies that do not include this assumption. They are called queueing with set up).

When service is completed, the next unit in the queue to be serviced is chosen and under the usual assumptions, that unit immediately goes into service. The rule used to choose the next unit to serve from among all those in the queue is called the queue discipline. The most common assumption in queueing theory is that the first unit in the waiting line is the first customer to be served. Such a discipline is called a first in-first out discipline. For other applications other disciplines have been considered. (In military studies the next unit to be served may be the one posing the most immediate threat. In a computer center jobs may be assigned to classes of importance - called priority classes - and the next job chosen is from among those with the highest priority. In production systems the next job to manufacture may be the one whose promised delivery date - the so called "due date" - is closest.)

There are four basic processes created by the interaction of service times, interarrival times and queue discipline that are of particular importance though others have been studied. The length of the waiting line, called the queue length process $\{N(t): t \geq 0\}$, is one of the basic processes of concern. The length of time from entry to exit of the queueing system $\{Q_n: n = 1,2,\cdots\}$ called the sojourn time process, is another of the basic process. The time from first entrance of a unit to an idle server until first entrance of the next unit to an idle server $\{B_n: n = 1,2,\cdots\}$ is called the busy cycle. The time between consecutive exists from the server $\{X_n: n = 1,2,\cdots\}$ is called the departure process.

Somewhat more formally we can define the problems as follows:

(1) Let $0 \leq T_0 < T_1 < T_2 \cdots$ be a sequence of real random variables representing the times at which an arrival occurs to a queueing system. Let $A_n = T_n - T_{n-1}$, $n = 1,2,\cdots$. Assume $\underset{\sim}{A} = \{A_n\}$ is a sequence of i.i.d. random variables (a renewal process). Call $\underset{\sim}{A}$ the arrival process to the queue.

(2) Let S_n be a non-negative real random variable. Let $S = \{S_n: n = 1,2,\cdots\}$ be a sequence of i.i.d. random variables. Call $\underset{\sim}{S}$ the service time process for the queue. We assume $\underset{\sim}{A}$ and $\underset{\sim}{S}$ are independent processes.

(3) Let $1_{[0,t]}(T_n)$ be an indicator random variable taking values 0 (or 1) if T_n does not (or does) occur in $[0,t]$.

(4) Define $N_A(t) = \sum_{n=0}^{\infty} 1_{[0,t]}(T_n)$. $\{N_A(t): t \geq 0\}$ is called the arrival counting process.

The four processes of basic concern to queueing theory can be defined formally in terms of the $\underset{\sim}{A}$ and $\underset{\sim}{S}$ process. Prabhu [1965] provides a formal definition of the queue length process. We will not reproduce that here but rather we assume that everyone has an intuitive understanding of such a process. The sojourn time process is closely related to the waiting time process for first in-first out disciplines which can be formally stated as

$$W_n = W_{n-1} + S_{n-1} - A_n, \quad \text{if } W_{n-1} + S_{n-1} - A_n > 0,$$
$$= 0 \quad , \quad \text{otherwise.}$$

Then $Q_n = W_n + S_n$ defines the sojourn time of the n^{th} arriving customer. The busy cycle process is the sequence of first passage times already discussed. If we define X_n as the time between the n and $(n-1)^{st}$ service completion then the departure process $\{X_n: n = 1,2,\cdots\}$ can be more formally given as

$$X_n = \begin{cases} S_n, & \text{if the } n^{th} \text{ unit leaves someone in the queue,} \\ I_n + S_n, & \text{if the } n^{th} \text{ unit leaves an empty queue.} \end{cases}$$

Here I_n is the foreward recurrence time in the arrival process measured from the time of departure of the $(n-1)^{st}$ unit. I_n, is called the idle time of the server.

1.3 Special Cases. The major thrust of queueing theory for much of its life has been to compute formulas for the above measures of effectiveness with special assumptions on the arrival process, service process, queue discipline or systems capacity. For future reference, we will record some results for one special case here. The following results as well as many of the special cases are discussed in most standard texts on the subject (for example, see Kleinrock [1975; 1976]).

If we assume that $N_A(t)$ is a Poisson process (with parameter λ as is usual) and S_n is exponentially distributed (parameter μ as is usual), the service center has 1 server, the queue discipline is first in-first out and the system's capacity is unlimited then the system is called an M/M/1 queueing system. A considerable amount is known about the above four processes (Kleinrock [1975]). For example,

(1) $\{N(t)\}$ is a Markov process with state space $[0,\infty)$.

(2) $Pr[N(t) = k]$ is known (but we'll not reproduce the results here. Essentially the probability is given by an infinite sum of Bessel functions of type - I i.e. Bessel functions of the second kind with imaginary arguments).

(3) $\lim_{t\to\infty} \Pr[N(t) = k] = (1-\rho)\rho^k$, $k = 0,1,2,\cdots$ for $\rho = \lambda/\mu$ (called the traffic intensity). Such a limiting distribution exists if and only if $\rho < 1$, otherwise the limit is identically 0 for all finite k. These limiting probabilities, when they exist, are called the steady state probabilities.

(4) $\{W_n\}$ is a random walk on the non-negative reals with a delaying barrier at 0. (see Feller [1963] or Borovkov [1976]).

(5) $\Pr[W_n \leq t]$ is known in principle (through the theory of random walks).

(6) $\lim_{n\to\infty} \Pr[Q_n \leq t] = 1-e^{-(\mu-\lambda)t}$, $t \geq 0$. This limit exists if and only if $\rho < 1$, otherwise the limit is 0 for all finite $t \geq 0$.

(7) The busy cycle process is a sequence of i.i.d. random variables whose distribution is known (see Cox and Smith [1961], for example).

(8) The departure process $\{X_n: n = 0,1,2,\cdots\}$ is, in principle, known for all n from (2).

(9) In the steady state (i.e. $t \to \infty$), when $\rho < 1$, $\{X_n\}$ is a sequence of i.i.d. random variables that are exponentially distributed with parameter λ. That is, the departure counting process is a Poisson process with the same parameter as the arrival process. (see Burke [1956] or Disney, et al. [1973]).

Results (1), (3), (9) are useful to our future discussion.

2. Jackson Networks

2.0 Introduction. It was recognized early in queueing theory that a theory of single server systems was not adequate. Indeed, at least by the 1930's (and before if one is not too picky) research in queueing had begun to explore systems of queues or what we shall call networks of queue. Erlang, about 1910 had studied systems where many identical servers all handled the same arrival process (the M/M/c model) in telephone systems. A good discussion of these and multiple server models is given in Syski [1960] especially chapters 5 and 6 as well as chapter 7 to 10.

2.1 Jackson Networks. A general model in this historic mold that has served almost as the definition of a queueing network was studied in a paper by J. R.

Jackson [1957]. These models and related problems have come to be called "Jackson networks." His 1957 model assumed:

(1) There are $1 \leq M < \infty$ service centers (Jackson did allow servicing center j to be comprised of more than 1 server. However, to keep the problem simple we assume throughout that each servicing center has only one server).

(2) These M servers are arbitrarily connected by arcs over which units travel instantaneously fast.

(3) To establish how units proceed through the network one defines p_{ij} to be the probability that a unit completing service at service center i proceeds to service enter j for its next service. Then for each i, $1 - \sum_{j=1}^{M} p_{ij}$ is the probability that a unit exiting node i leaves the system. The matrix $\underset{\sim}{P}$ whose elements are p_{ij} is called the <u>switching matrix</u> for the network. Note that these assumptions about switching imply that a unit leaving service center i chooses the next service center to visit without regard to any other condition. We call this switching behavior a <u>Bernoulli switch</u>.

(4) Service times, S_{ni}, $n = 1,2,\cdots$ at center i are i.i.d. random variables. These times are each exponentially distributed random variables with parameter μ_i. The sequences $\{S_{ni}\}_{i=1}^{M}$ are independent sequences.

(5) There may be many arrival processes but each one is a Poisson process (with parameter λ_i if the arrival process is to service center i). These arrival processes are independent of each other and of the network.

(6) All queue disciplines are first in-first out.

(7) All queue capacities are unlimited.

2.2 The Queue Length Process. The results that Jackson obtained were rather surprising. Jackson's major results were as follows:

Define $\underset{\sim}{N}(t)$ to be a vector whose j^{th} element, $N_j(t)$, $j = 1,2,\cdots,M$, is the queue length at the j^{th} service center at t.

Then

(1) $\{\underset{\sim}{N}(t): t \geq 0\}$ is a vector valued Markov process.

(2) $\lim_{t\to\infty} \Pr[N_1(t) = k_1, N_2(t) = k_2 \cdots N_M(t) = k_M]$

$= \lim_{t\to\infty} \Pr[N_1(t) = k_1]\Pr[N_2(t) = k_2] \cdots \Pr[N_M(t) = k_M]$.

That is, the queue length processes in these Jackson networks are asymptotically ($t\to\infty$), mutually independent.

(3) $\lim_{t\to\infty} \Pr[N_j(t) = k_j] = (1-b_j)b_j^{k_j}$, $j = 1,2,\cdots$ if and only if $b_j < 1$. Otherwise this limit is 0. Here $b_j = a_j/\mu_j$ and a_j satisfies the so called traffic equation

$$\underset{\sim}{a} = \underset{\sim}{\lambda} + \underset{\sim}{P}\underset{\sim}{a}.$$

$\underset{\sim}{a}$ is the M-vector of a_j, $\underset{\sim}{\lambda}$ is the M-vector of λ_i, $\underset{\sim}{P}$ is the switching matrix. We assume throughout that the traffic equation has a positive solution (or alternatively the maximum eigenvalue of $\underset{\sim}{P}$ is less than 1). In elemental form, this equation says simply that a_j is the rate of arrivals to service system j. This arrival rate is simply the sum of exogeneous arrivals to system j (λ_j) plus the sum of all arrivals to j from inside of the system $(\underset{\sim}{P}\underset{\sim}{a})_j$.

It is instructive to spend a moment looking at what Jackson said and did not say. Many papers have been written on Jackson networks that misinterpret and misuse his results. The major confusion occurs around the third result. Remembering that in the example of section 1.3 we showed that the queue length process was asymptotically ($t\to\infty$) geometrically distributed with a parameter $\rho = \lambda/\mu < 1$. If $\rho \geq 1$ the limit is zero. Also recall that λ was the parameter associated with the Poisson arrival process and μ was the parameter associated with the exponential service times.

Now result (3) above is also asymptotically geometrically distributed with a parameter a_j that is acting as an arrival rate and a parameter μ_j acting (is) as a servicing rate. This result led Jackson to state "This theorem (our results (2) and (3) above) says, in essence, that at least so far as steady states are concerned (out $t\to\infty$), this system with which we are concerned behaves <u>as if</u> its departments (our centers) were independent, elementary (i.e. M/M/1) systems ...". So far, so good. The <u>as if</u> emphasis is Jackson's and as it stands is innocuous enough. Jackson makes one more statement about his results that may have caused an errormous amount of confusion. He says "This conclusion is far from surprising in view of recent papers by E. J. Burke (that should be P. J. Burke) and E. Reich." It appears from his references that Jackson is here alluding to the result (9) of section 1.3 which is sometimes called "Burke's Theorem".

Confusion in queueing network literature has occurred by taking the "as if" of Jackson too seriously. Researchers seem to have taken this to mean that the queue length processes are independent and are created by M/M/1 subsystems. As a consequence there are papers in the queueing literature that study queueing networks one node at a time (the independence assumption) as M/M/1 queues (the assumption of Burke's theorem). There are other papers that study sojourn times in networks as sums of independent, exponential random varialbes (see section 1.3, result (6)). Unfortunately, many authors have shown that the flow of units within the network are, except for special network configurations such as trees, not Poisson processes and in fact, are not even renewal processes. Thus each service center in isolation is not only not an M/M/1 queueing system, it is not even a queueing system with a renewal arrival process. That is, for a time there was serious confusion in the literature as to what as if really meant. The problem is fairly well understood now. Section 5 and its references discuss the topic more thoroughly.

2.3 Waiting Times. Waiting times in Jackson networks is a rather unexplored area. Attention has recently turned toward that area but research here has only started. There are a few special cases that have been studied in detail principally in what are called tandem queues.

One can define a tandem, Jackson network as one having the structure of section 2.1 with the added features:

(1) There is only one arrival process and it occurs to the first server (in terms of the section 2.1 scheme $\lambda_1 = \lambda$, $\lambda_j = 0$, $j \neq 1$).

(2) $p_{ij} = 0$ unless $j = i + 1$ in which case $p_{ij} = 1$, $j = 1,2,\cdots M$. (see next

page) Since the only results, with which we are familiar, are concerned with first in-first out queue disciplines, we assume this discipline throughout. We need a bit more symbolism here. So define:

W_{ni} = the total time unit n spends waiting in service center i. Call this the <u>waiting time</u> of unit n in service center i.

Q_{ni} = the time spent by unit n in service center i. Call this the <u>sojourn time</u> of unit n in center i. $W_{ni} + S_{ni} = Q_{ni}$, $i = 1,2,\cdots M$, $n = 1,2,\cdots$.

The first result is from Reich [1957] which shows:

For fixed n, $\{Q_{ni}\}$ is a sequence of i.i.d. random variables each of which is exponentially distributed.

Thus, if Q_n is the sojourn time in the system for unit n, this result says simply that this sojourn time is the sum of independent, random variables

$$Q_n = Q_{n1} + Q_{n2} + \cdots + Q_{nM}.$$

The Q_{ni} may depend on μ_i. Nonetheless the distribution of Q_n can be obtained from simple convolution operations.

The next result is surprising. It comes from Burke [1964]. For the tandem queueing system he shows:

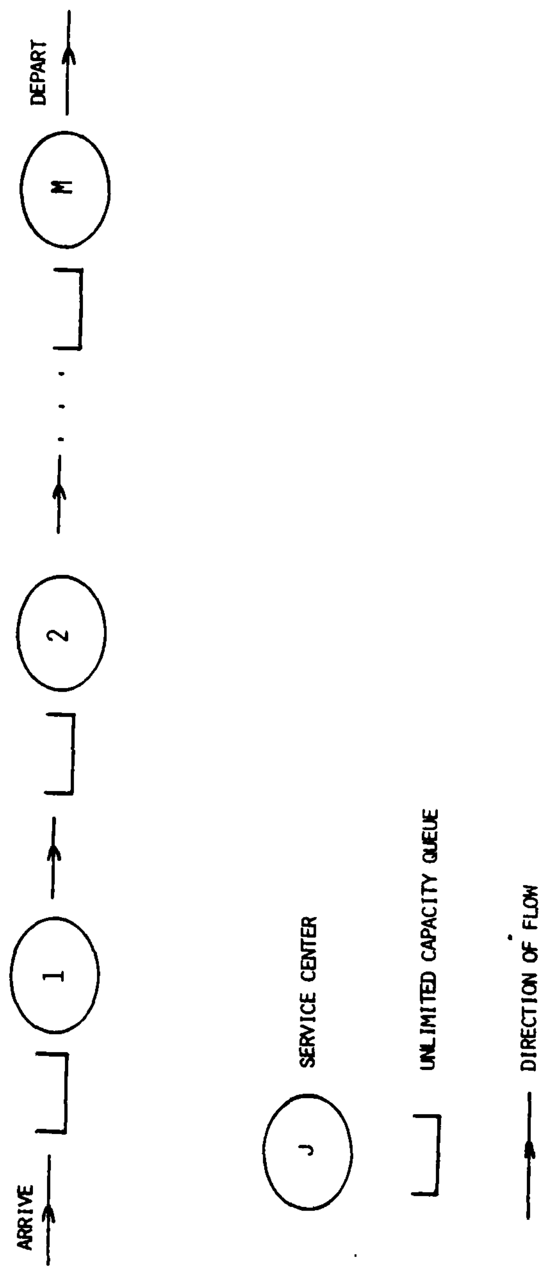
The sequence $\{W_{ni}\}$ for n fixed and $i = 1,2,\cdots M$ is a sequence of dependent random variables.

That is sojourn times at successive service centers in the network are independent but waiting times are not. It is curious that the sequence $\{W_{ni}\ i = 1,2,\cdots M\}$ for fixed n is a sequence of dependent random variables. But if to each term in this sequence we add a random variable, independent of $\{W_{ni}\}$ and exponentially distributed (the S_{ni}) to obtain a new sequence $\{Q_{ni} = W_{ni} + S_{ni}: i = 1,2,\cdots\}$ that sequence is a sequence of independent random variables for each n.

These waiting time problems exhibit another nasty property that has been proven in two distinct cases. Burke [1969] proved the first result. To expose it we must assume that a service center can have more than one server. It is

FIGURE 1

A TANDEM JACKSON NETWORK



usual to assume that all service times at a given center are mutually independent and all are exponentially distributed with parameter μ_j that depends only on the center (not on the server in the center - the servers are "identical"). With this set up, Burke considers a three center tandem queue with the first center having just one server, the second center having any finite number of servers, (> 1), and the third center having just one server. Otherwise, the problem is that of Reich discussed above. Burke is then able to show that for each n.

Q_{n1} and Q_{n2} are independent random variables.

Q_{n2} and Q_{n3} are independent ramdom variables.

Q_{n1} and Q_{n3} are <u>not</u> independent random variables.

This results is surprising. In this network one does not have the Reich mutual independence. In fact, one does not even have pairwise independence.

Simon and Foley [1979] have found a similar result to that of Burke in a different network. In their three service center network, there is one server at each center . $\lambda_1 = \lambda$, $\lambda_j = 0$, $j = 2,3,$. The new idea is that the network is not a tandem queueing network. Rather one has $p_{12} = p$, $p_{13} = 1-p$, $p_{23} = 1$, $p_{3j} = 0$, $j = 1,2,3$ in the structure of section 2.2 (see next page). Then they are able to show that:

Q_{n1} and Q_{n2} are independent random variables.

Q_{n2} and Q_{n3} are independent random variables.

Q_{n1} and Q_{n3} are <u>not</u> independent random variables.

One conjectures that what is happening here is that in both the Burke problem and the Simon - Foley problem a unit in queue 2 (in those examples) may be by-passed by units in server 1 that arrive to the system after unit n. In both cases then the queue length and hence the sojourn time at server 3 will depend on how many such units by-pass unit n which in turn depends on how long unit n was in server 1.

Following this line of thought, the following conjecture appears reasonable. If there is only one path connecting any two single service centers in

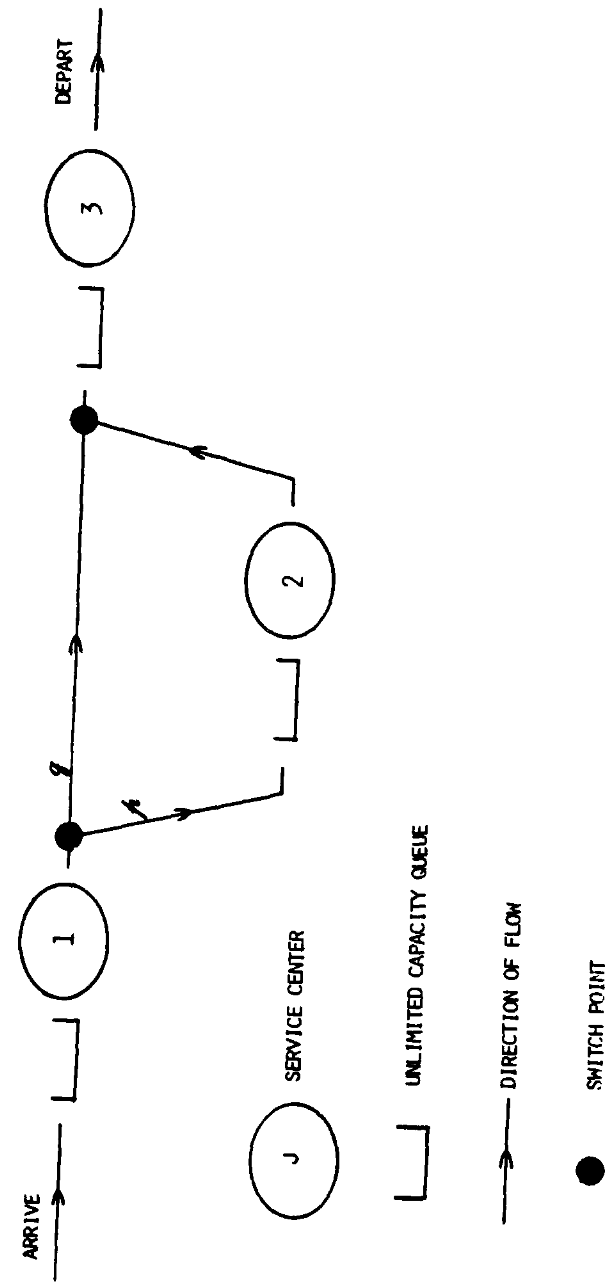


FIGURE 2

THE SIMON-FOLEY NETWORK

the network then one can use the result of Reich [1957] to determine the total sojourn time of a given unit through a Jackson network. If, on the other hand, there are multiple paths connecting single server service centers then the sojourn times of unit n at the service center at the start of these multiple paths and that at the termination of the multiple paths are dependent. Furthermore, if there can be multiple servers at any service center then the sojourn times at service centers feeding this multiple server service center and those being fed by it are dependent unless the multiple server center is first or last in the sequence.

Unfortunately, we do not know the nature of these dependencies. Neither do we know the sojourn times through either the Burke or Simon and Foley networks. This is an area in need of considerably more study.

There is yet one more problem with sojourn times in these Jackson networks. In perhaps the simplest non-trivial Jackson network one takes all of the assumptions of section 2.1 with the added assumption of single servers at each service center. The only alteration is to have just one service station with $p_{11} = p$ (see next page). We call this a queue with instantaneous, Bernoulli feedback. The definition of sojourn time needs a modification. So let:

Q_n^i = the sojourn time of unit n the i^{th} time that unit passes through the server. Then

Q_n = the sojourn time of unit n.

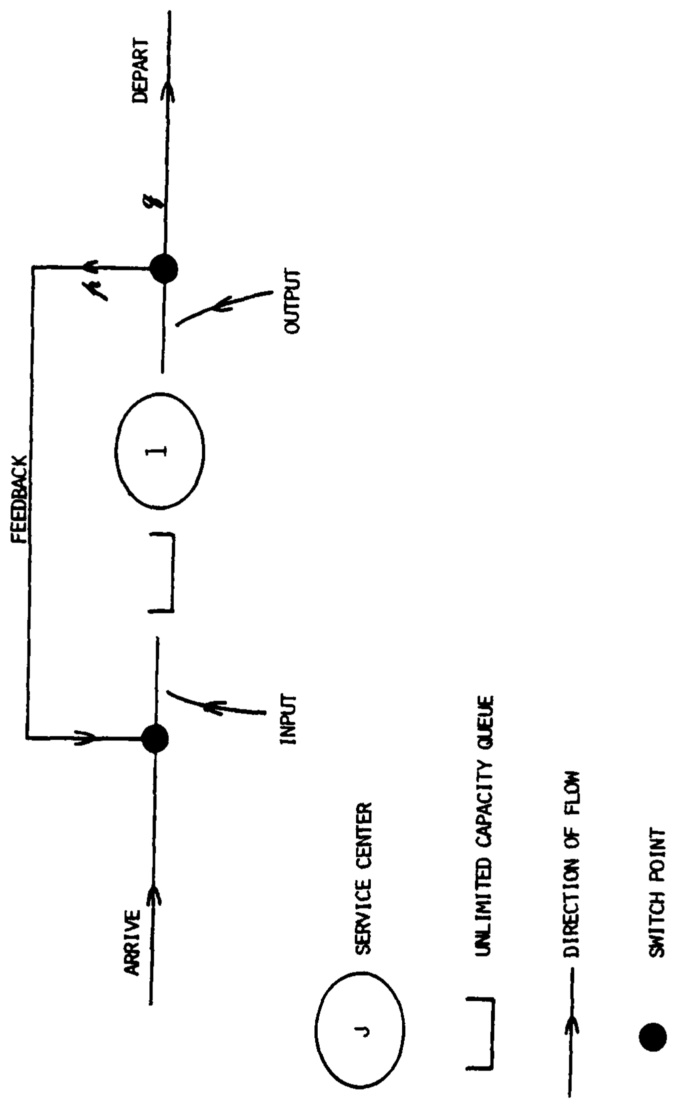
$$Q_n = Q_n^1 + Q_n^2 + \cdots + Q_n^k$$

where k is a geometrically distributed random variable.

It has been shown by Takacs [1963] and Disney and D'Avignon [1978] that for each fixed n, $\{Q_n^i, i = 1,2,\cdots\}$ is a Markov renewal process. Takacs and Disney and D'Avignon each solve the problem for slightly more general cases than Jackson networks using rather different methods. Once again as in the Burke and Simon and Foley examples the lack of independence seems to be coming from what we loosely called by-passing.

FIGURE 3

THE DISNEY-D'AVIGNON NETWORK



2.4 Summary of Results. In summary, unlike the Jackson queue length process results that have been obtained and are elegant in appearance, waiting time and sojourn time properties are largely unknown. For tandem queues studied by Reich the problem can be considered solved. For general, single server, Jackson networks that do not have multiple paths between any two service centers in the network, it appears that the Reich results can be used to obtain results easily. For more general networks of Jackson type, the sojourn time problem is largely unsolved. Because of its importance to many areas it is one of the most pressing problems in queueing network theory.

3. Busy Periods and Departure Processes

3.0 Introduction. The queue length process and sojourn times have been the major topics of interest to queueing theory since its beginning. While of some importance to systems analysis, the busy period and to a lesser extent (until rather recently) the departure process from single service queues have been of less importance. The same result is true for Jackson queueing networks as we shall see below.

3.1 The Busy Period. One can define two related processes of interest. One is called the busy period. The other is called the busy cycle. Both are first passage probabilities. The busy period is the time from first entrance of a unit to an empty queue to first exit of a unit that leaves the system empty. The busy cycle is the first return time from entrance of a unit to an empty queue to the next time of entrance to an empty queue. If the queue is ergodic such points occur infinitely often. If the queue is not ergodic such points, except possibly the first one, may not exist. Almost all studies of the busy cycle and busy period with which are are family, assume these entrance times occur infinitely often (see Cox and Smith [1961]).

In single server queueing theory in which $\{A_n\}$ is a sequence of i.i.d. random variables, $\{S_n\}$ is also a sequence of i.i.d. random variables and $\{A_n\}$ and $\{S_n\}$ are independent sequences, the sequence of busy cycles $\{B_n : n = 1,2,\cdots\}$ is a sequence of i.i.d. random variables if the origin of the time scale is taken as a point of arrival to an empty queue. The busy period is in general not a sequence of i.i.d. random variables because successive periods are not independent. For the special case in which $\{N_A(t)\}$ is a Poisson process, the busy period process as well as the busy cycle process are both sequences of i.i.d. random variables. General distributional results are known for the cases in which the two processes are sequences of i.i.d. random variables (for example, see Kleinrock [1975]).

Unfortunately, the related problems in queueing networks are in a much less well developed form. To the best of our knowledge there are no known results for busy periods or busy cycles for Jackson networks. The area could stand some study.

Stating the problem that needs to be considered is rather easy. Recalling from section 2.2 (item 1) that $\{\underset{\sim}{N}(t)\}$ is a vector valued Markov process, the busy cycle process is then simply the time from first entrance of an arriving unit to empty network to the next time of first entrance of an arriving unit to an empty network. In the busy period case the problem is to determine the time from first entrance to an empty network by an arriving unit to the time at which a departing unit leaves behind an empty network. We surmize that such points occur infinitely often for ergodic networks but as pointed out above we know of no results for the busy period or busy cycle of a Jackson queueing network.

3.2 Departure Processes. There are two problems here that lead to interesting questions related to our discussion of the queue length process and waiting time process. One problem is concerned with the nature of the departure processes from the Jackson network. The other problem is concerned with the nature of the departure processes from single centers in the network. To simplify our discussion we will assume the network is irreducible in the sense that the switching matrix $\underset{\sim}{P}$ is an irreducible matrix. More general cases have been studied. The interested reader is referred to Melamed [1979] and its references for these other cases.

When $\rho_i < 1$ for $i = 1,2,\cdots M$ every entering unit eventually leaves the system. Then from Melamed [1979] we have:

(1) In Jackson networks with single server service centers if j is a node from which departures from the network occur then the departure process from the network at this center is a Poisson process with parameter a_i.

(2) The collection of Poisson processes of departures from the network are mutually independent.

The first of these results is quite in keeping with the Burke theorem (section 1.3, item (9)). The second result is, at first glance, rather surprising. One would expect that the network itself imposed some dependencies on the departing processes.

When one turns to consider departure processes from individual service centers in the network things became a bit more complicated. For the time being we will continue the discussion within the framework of Jackson networks. Furthermore, we must distinguish two cases. If $p_{ii}^{(n)} > 0$ is defined as the n step transition from i to i in the switching process, then there is some path leading from service center i back to that service center. In this case we will say that service center i has <u>feedback</u>. Otherwise we will say service center i does not have feedback. The latter case we can dispose of quickly.

If service center i does not have feedback, the departure process is a Poisson process with parameter a_i.

The service center with feedback requires distinguishing two processes. In one process units leaving service center i will eventually return to i. In the other process, units leaving service center i will never return to i. Call the former process, the <u>feedback stream</u> and the latter the <u>departure stream</u>. (see figure 3). Then again from Melamed [1979] we have:

(1) The departure stream is a Poisson process.

(2) The feedback stream is not a Poisson process and in fact is not a sequence of i.i.d. random variables.

We can flesh out item (2) a bit more in the case $p_{ii} > 0$, that is feedback occurs in one step - the so called instantaneous feedback case. In that case the total process of units leaving the service center (called the output process) is a Markov renewal process whose transition functions are known. Then it has been shown by Disney, et al. [1979] that the output process is never a Poisson process nor is it a renewal process. However, the departure process is a Poisson process (parameter λ). The feedback process is not a renewal process. There is reason to believe that the departure process and the feedback process are not independent random processes but we know of no proof either way here.

These Melamed and Disney et al. results raise some interesting questions. In these Jackson networks as was noted in section 2.2 item 3, queue lengths at individual service centers act as if they were independent, M/M/1 queues. Yet if the network has feedback loops, the flow on those loops is not a Poisson process nor even a sequence of i.i.d. random variables. It is this property (i.e. the distinction between properties of the network and properties of the individual service centers in the networks) that seems to have created confusion in some applications of the Jackson network results both in the study of the queue length process and that of the waiting time process.

4. Extensions to the Jackson Network Theory of 1957

4.0 Introduction. Following his 1957 paper, Jackson next published a paper in 1963 in which he followed the basic ideas of the earlier paper. The new ideas were to allow the arrival processes to the network to be birth processes whose parameters could depend on the total number of units in the network. Similarly, the service time processes were death processes with parameters depending on the number of units at a given service center. In this way the queue length process $\{\underset{\sim}{N}(t)\}$ becomes a vector valued birth-death process.

4.1 Queue Length Processes in the Vector-valued Birth-Death Process. We will not reproduce the exact form of Jackson's 1963 results. They would require introducing a large amount of new symbolism and currently available work which we shall discuss later includes these results. However, it is important to summarize the findings of Jackson (at a cost of imprecision) because they have lead many others to explorations in queueing networks in an attempt to generalize the concept of a Jackson network.

In his 1963 paper, Jackson finds that the joint probability of the vector $\underset{\sim}{N}(t)$ is asymptotically ($t\to\infty$) geometric. The first term in this geometric distribution, representing the probability that the entire network is idle, however, cannot be factored (except in special cases) into a form that would imply that the individual queue length processes were independent.

Since these results appeared, considerable effort has been expended trying to determine approximations and easy ways to compute the initial term. Other research effort has been expended on exploring the "product form" (implied above) of the solution to many networks. There are no up to date summaries of the large amount of work. The Kleinrock [1975;1976] books are basic. The papers of Kelly [1976; 1978] and Schassberger [1977; 1978] trace some of the work.

4.2 A Generalization. In a 1976 paper, Kelly significantly generalizes the concept of Jackson network and provides important extensions to the concept of "product forms" of solutions. In an unpublished series of lecture notes (Kelly [1978]) he provides further insights and significantly broader applications of his study. We will follow his 1976 publication because it is accessible in the open literature.

Suppose that there can be I types of units entering the network. Units of type $i \in I$ enter the network as a Poisson process with rate $\nu(i)$ and pass through the servers according to the path

$$r(i,1)r(i,2)\cdots r(i,S(i))$$

before leaving the system. Thus at stage s ($s = 1,2,\cdots S(i)$) of his route, the unit is at queue $r(i,s)$.

Within each queue, the units are ordered so that there are units in positions $1,2,\cdots n_j$. n_j is the total number of units in queue j.

Each unit requires a random amount of service. This service time is an exponentially distributed random variable with mean 1.

There is a single server who supplies a total service effort at rate $\phi_j(n_j)$ in such a way that $\gamma_j(\ell,n_j)$ of this effort is directed to the unit in position ℓ. When this unit leaves the j^{th} service system, all units behind it move up a space.

When a unit arrives at service center j it moves immediately into location ℓ with probability $\delta_j(\ell, n_j+1)$. Units formerly occupying spaces ℓ, $\ell + 1,\cdots n_j$ are moved to spaces $\ell + 1$, $\ell + 2,\cdots n_j + 1$.

Such a structure is rather general for queueing network behavior. The queue discipline of the earlier Jackson networks has been considerably generalized. Arrival processes are still Poisson but note the arrival rate may depend on the "type" of the unit. "Type" may be associated with the path taken by the arrival simply by associating an arrival "type" with the path that arrival will follow. Service times here are also more general.

Another important aspect of the Kelly paper is its method of determining the limiting probability vector for the queue length random process. Whereas the earlier Jackson network paper of 1957 and 1963 proceeded from the structure of the problem to set up the usual limit form of the Kolmogorov equations of the Markov process $\{\underset{\sim}{N}(t)\}$, Kelly prefers to work with relations embodied in a reversed process. It appears that when such an approach can be made to work, detailed calculations are obviated. The Kelly lecture notes expand on this point and considers these methods for a wide variety of problems in multidimensional Markov processes, including the queueing networks of this 1976 paper.

Kelly starts with a stable, conservative, regular Markov chain (all of these networks discussed so far have these properties). Then it is well known to the theory of such processes that if i, j are vector valued states in these networks, then a solution to the equations

$$\Pi \mathbf{P} = 0$$

with $\Pi \geq 0$ and $\Pi \underset{\sim}{1} = 1$ is unique and

$$\Pi(j) = \lim_{t\to\infty} \Pr[\underset{\sim}{N}(t) = j]$$

(The elements of Π are the limiting probabilities - the so called "steady state" probabilities - of the Markov process).

Now it is known that if $\{N(t)\}$ is a Markov process in equilibrium (e.g. the initial vector for the process is Π) there then exists another process - called the reversed process, $\{N(-t)\}$, that is a Markov process in equilibrium. The reversed process has the same limiting probability vector, Π but its infinitessimal generator may not be that of $\{N(t)\}$. (If the two processes have the same infinitessimal generator then $\{N(t)\}$ is said to be reversible). In general if $q(i,j)$ is the (i,j) element of the infinitessimal generator of $\{N(t)\}$ and $q'(i,j)$ that of the infinitessimal generator of $\{N(-t)\}$ and if $q(i)$, $q'(i)$ are the corresponding diagonal elements of the infinitessimal generator of $\{N(t)\}$, and $\{N(-t)\}$ respectively, then we have

(4.2.1) $$\Pi(i)q(i,j) = \Pi(j)q'(j,i)$$

and

$$q(i) = q'(j).$$

(If the process is reversible these equations are called the equations of detailed balance).

The extremely useful result is that for the network set up by Kelly, one can find $q(i,j)$ and $q'(i,j)$ rather easily. This, along with (4.2.1) and the uniquess of Π as a probability vector then allows one to determine Π.

For his network, Kelly defines a two-tuple $c_j(\ell) = ((t_j(\ell), s_j(\ell))$ where $t_j(\ell)$ denotes the "type" of unit in position ℓ in queue j and $s_j(\ell)$, as previously defined, represents the position along its path reached by this ℓ^{th} unit in queue j. It is shown that for the vector

$$c_j = (c_j(1),\ c_j(2)\cdots c_j(n_j))$$

$$\underset{\sim}{C} = (c_1, c_2, \cdots c_M)$$

is an irreducible Markov process on a countable state space. The infinitessimal generator for this process and its reversed process is found. Then using the results on reversed processes stated above, Kelly shows that his network has the product form of solution which is determined up to a normalizing constant.

4.3 Another Generalization. Kelly generalizes his model one more step and in the process provides the basis for taking a major step out of the restrictive assumption of Poisson processes and exponential service times. Unfortunately, this step has a price to pay.

The first step is to generalize the service time assumption of section 4.2. Now instead of service times being exponentially distributed random variables, it is assumed that when at queue j, the unit that is at stage s of its path requires an amount of service that is the sume of Z(j,s) independent, identically distributed, exponential random variables (rate d(j,s)). That is, service times are now gamma distributed random variables with parameters Z(j,s) and d(j,s).

The other assumptions of section 4.2 are retained except that it is required that

(4.3.1) $$\delta_j(\ell, n_j + 1) = \gamma_j(\ell, j_j + 1).$$

For this problem the state space of the Markov process of interest must now become a three-tuple. A state now is defined by $t_j(\ell)$, $s_j(\ell)$, as in section 4.2, and $x_j(\ell)$ which denotes the phase of service currently occupied by the unit. Then, on this three-tuple space one can define a Markov process whose states are the three-tuples. This process is irreducible and the state space is countable. The process has a reversed process, its infinitessimal generator and that of the reversed process can be found. As before the properties of reversing are used and it is shown that the process has a product form of solution.

4.4 A Major Generalization. Were the Kelly results to stop here they would make interesting, perhaps useful, contributions to the theory of Jackson queueing networks. The restriction of Poisson arrival processes and either exponentially distributed or gamma distributed service times would preclude a wide spread use of the results (although much of the computer systems analysis literature finds these conditions to be reasonable for many computer studies). But more is available.

Under the conditions of the model of section 4.3, Kelly conjectures that his results go through for G-type service times (see section 1.2). The conjecture is based on a result of Whitt [1974] which shows that finite mixtures of gamma distributions are dense in the set of arbitrary non-negative distributions. Though Kelly does not prove that his model of section 4.3 extents to G-type servers, it is proven with the requisite care by Barbour [1976].

Finally, Kelly drops the Poisson arrival process assumption that has run through his work. Commenting on his models in our sections 4.2 and 4.3 he notes that one can allow these arrival processes to be birth processes whose parameter depends on the total number of units in the system. Recall that this step was made in the Jackson [1963] paper.

4.5 Comments. The Kelly work probably represents the state-of-the-art in the study of queueing networks originally arising out of the papers of Jackson. Work continues on these problems and the Kelly work will probably be mined for quite a while. If condition (4.3.1) could be removed from the Kelly model and still obtain computable equilibrium solution, one would have a major contribution to the theory of queueing networks.

Concerning waiting time, busy period, and busy cycle analysis we know of no results presently available. Concerning departure processes, Kelly presents some results on the departures from the network. These processes are independent, Poisson processes for the models studied in our sections 4.2 and 4.3.

It has been noted by many authors (for example, see Kelly [1976]) that since the Jackson limiting probability vector depends on the assumptions of the arrival process and service time process only through the expected values of these processes, similar results may well hold for more general arrival and service time processes. That is, results such as those obtained may be insensitive to the distributional assumptions of the model. Schassberger [1977; 1978] explores this topic in more detail and provides a bibliography for further reading.

5. Flow Processes

5.0 Introduction. Many writers, in recent years, have come to realize that the results of the 1957 Jackson paper are remarkable. Even though the results imply that the queue length processes act *as if* they are those of independent M/M/1 queues, they are, in fact, not. The result has lead to a collection of papers investigating the properties of the flow of units within the network. Such analysis is important if one is to gain an understanding of these flow processes, if one is to generalize the simple switching structure in Jackson networks, if one is to gain an understanding of waiting times and if one is to gain some understanding of sampling data produced by service centers in the network.

The only attempt to survey this overall field with which we are familiar is Disney [1975]. That paper is somewhat out of date by now. However, the bibliography therein is rather complete up to July, 1975.

5.1 A Few New Results. The study of flow processes in these networks consists primarily of the study of five subareas called: decomposition, recomposition, departures, feedback queues and queues with non-renewal arrivals. Few new results seem to have appeared since 1975 concerning the area of decomposition. Therefore we refer to the Disney [1975] paper for an up to date discussion on that topic. Recomposition studies, likewise, have received scant attention since 1975 as is also true for queues with non-renewal arrivals. We suggest that the interested reader consult the above review for work in this area.

There have been two major results on departure processes in Jackson networks and the networks of Kelly. In particular, Kelly shows that departures from his networks (both those discussed in section 4.2 and those discussed in section 4.3) are Poisson processes and that the several streams of departures from the service centers are mutually independent processes. (See Kelly [1976]). In Melamed [1979], the results of Kelly for Jackson [1957] networks are verified using quite different method. In addition Melamed shows that flows on arcs internal to Jackson network are Poisson processes on those arcs joining non-communicating classes of states in the switching process. On arcs within communicating classes, the flows are not Poisson processes and in fact are not even renewal processes.

In an attempt to study these flows in subsets of communicating service centers, Disney and D'Avignon [1978] and Disney, McNickle and Simon [1979] study a simple case in detail. This case is called a queue with instantaneous feedback (see p.16). The 1979 paper studies four random processes in these queues to show that the departure process from the system is a Poisson process if and only if service times are exponentially distributed random variables (see also Disney, Farrell and de Morais [1973] on this topic for single server queueing systems.) In case of renewal process service times the departure process is not only not a Poisson process, it is not a renewal process. It is shown that the process is a Markov renewal process and its transition functions are given. Flows internal to the network are never (for any service time process) Poisson processes (except for the trivial case in which there is no feedback). In general these internal flows are Markov renewal processes and in the single case studied by Disney, McNickle and Simon the transition functions for some of these Markov renewal processes are given.

The Disney and D'Avignon paper [1978] is a major study of queues with instantaneous feedback. In particular, the arrival process is allowed to be a Markov renewal process, service times may depend on the state of the arrival process and the switching probabilities may depend on queue length increments, previous switching decisions, types of units being switched and the amount of service time received by the customer being switched. This paper is probably the state-of-the-art for the study of queues with instantaneous feedback. It includes references to all work with which we were familiar in mid 1978. The details are far too complex to summarize here. We ask the interested reader to consult the paper.

Current research is being conducted on queueing processes as well as flow processes in queues with delayed feedback (see, for example Foley [1978]). Except as noted in Disney and D'Avignon [1978] almost all of this work is concerned with non-Jackson networks with emphasis on queue length processes and flow processes. We know of no results other than those already cited on the waiting time process.

6. Summary

6.0 Summary. We have attempted to review in a few pages, more than 20 years worth of research in the field of queueing network theory. To accomplish this in such a short space we have concentrated on two topics: Jackson networks and flow in networks. Our primary emphasis has been on the Jackson network results. In areas of application these are the results that are of primary importance. Under the pressing restrictions of time and space we have concentrated only on the basic Jackson work and the important Kelly work. While we have alluded to other work, we have by no means provided a definitive state-of-the-art survey. Considerable work has been done on Jackson networks in the past 15 years. This work is to be found principally in the literature of the computer scientist whose interests in these topics seems to have revitalized that field. We can only hope that the reader interested in a host of results and fascinating applications will consult this literature. The best starting point is probably the two volumes of Kleinrock [1975; 1976] and especially the interesting chapters 4,5,6 of volume II which present some of the basic queueing problems occurring in computer networks as well as an interesting discussion of trials and tribulations of applying known results to the design of a large scale system. Beyond that we can only suggest that the interested reader peruse the journals of computer science (e.g. J.A.C.M. or Acta Informatica) as these topics continue to be research and applied.

The study of flow processes in queueing networks is fragmented at present. There are many results. We have mentioned a few. There is much that we have not said and much that remains to be said. We have not discussed the interesting work referenced in Schassberger [1977; 1978] that is attempting to tie queueing theory into the more general field of stochastic point process theory. Indeed, it is our view that this link up is natural. For the study of flow processes in networks, it is natural to think of the network and its components as operators on random point processes. That view will probably provide greater generality and deeper insights into these flow processes than is now possible.

At present, results on the study of flow processes in queueing networks appear in print sporadically. Unfortunately, such results when they do appear are published in literature through the world. The journals Appl. Prob. and Adv. Appl. Prob. are necessary reading but not sufficient.

There are many other topics in these areas that we have not even mentioned. The useful computational work of Wallace [1974] and Neuts (for example see Neuts [1979]) has been left untouched. The study of closed networks has been nearly ignored (e.g. Gordon and Newell [1967]). The interesting network decomposition ideas of Courtois [1978] deserves attention both for its theory as well as its application potential. The concepts of approximations including diffusion approximations and heavy traffic approximation have not even been mentioned. One might consult Harrison [1978] to start into this area. It would seem that there is no end to such an enumeration. The study of queueing networks is an enormously large and diverse field. Our tutorial has at best "hit the high spots".

References

This list of references is intended to be a guide to the literature. If an author wrote two papers one a continuation of the other we only reference the latter under the knowledge that the first paper is included in the bibliography of the second. In this way the references can be used to get one started in the field. More extensive searching would then have to be done by the usual "follow your nose" principle.

1. Barbour, A. D. (1976), "Networks of Queues and the Methods of Stages," Adv. Appl. Prob., 8, 584-591.

2. Borovkov, A. A. (1976), Stochastic Processes in Queueing Theory, Springer-Verlag, New York.

3. Brockmeyer, E., Halstrom, H. L., and Jensen, A. (1948), "The Life and Work of A. K. Erlang," Trans. Danish Acad. Tech. Sci., Transactions No. 2.

4. Burke, P. J. (1956), "The Output of a Queueing System," Oper. Res., 4 699-714.

5. Burke, P. J. (1964), "The Dependence of Delays in Tandem Queues," Ann. Math. Stat., 35, 874-875.

6. Burke, P. J. (1969), "The Dependence of Sojourn Times in Tandem M/M/s Queues," Oper. Res., 17, 754-755.

7. Courtois, P. J. (1977), Decomposability: Queueing and Computer Systems Applications, Academic Press, New York.

8. Cox, D. R. and Smith, W. L. (1961), Queues, Chapman and Hall, London.

9. Disney, R. L., Farrell, R. L., and de Morais, P. R. (1973), "A Characterization of M/G/1/N Queues with Renewal Departures," Mgmt. Sci., 20, 1222-1228.

10. Disney, R. L. (1975), "Random Flow in Queueing Networks: A Review and Critique," Trans. Amer. Inst. Industr. Engr., 7, 268-288.

11. Disney, R. L., McNickle, D. C., and Simon, B. (1979), "The M/G/1 Queue with Instantaneous Bernoulli Feedback," (to appear).

12. Disney, R. L. and D'Avignon, G. R. (1978), "Some Problems of Queues with Feedback," paper presented at Colloquium on Point Processes and Queueing Theory, Keszthely, Hungary, Sept. 4-8, 1978. Also Tech. Report VTR 78-2, Dept. of Industrial Engineering and Operations Research, Virginia Polytechnic Institute and State University. To appear in Proceedings of Keszthely Conference.

13. Feller, W. (1966), An Introduction to Probability Theory and Its Applications, vol. 2, Wiley, New York.

14. Foley, R. D. (1979), "The M/G/1 Queue with Delayed Feedback," text of talk given at National O.R.S.A./T.I.M.S. Conference, New Orleans, La., April 30-May 2.

15. Gordon, W. J. and Newell, G. F. (1967), "Closed Queueing Systems with Exponential Servers," Oper. Res., 15, 254-265.

16. Gordon, W. J. and Newell, G. F. (1967), "Cyclic Queueing Systems with Restricted Length Queues," Oper. Res., 15, 266-278.

17. Harrison, J. M. (1978), "The Diffusion Approximation for Tandem Queues in Heavy Traffic," Adv. Appl. Prob., 10, (to appear).

18. Jackson, J. R. (1957), "Networks of Waiting Lines," Oper. Res., 5, 518-521.

19. Jackson, J. R. (1963), "Jobshop-like Queueing Systems," Mgmt. Sci., 10, 131-142.

20. Kelly, F. P. (1976), "Networks of Queues," Adv. Appl. Prob., 8, 416-432.

21. Kelly, F. P. (1978), Reversibility and Stochastic Networks, unpublished lecture notes, Dept. of Math., Cambridge Univ., Cambridge.

22. Kleinrock, L. (vol. 1, 1975, Vol. 2, 1976), Queueing Systems, Wiley Interscience, New York.

23. Melamed, B. (1979), "Characterizations of Poisson Traffic Streams in Jackson Queueing Networks," Adv. Appl. Prob., 11, 422-438.

24. Neuts, M. F. (1978), "Markov Chains with Applications to Queueing Theory, which have Matrix Geometric Invariant Probability Vector," Adv. Appl. Prob., 10, 185-212.

25. Prabhu, N. U. (1965), Queues and Inventories, Wiley, New York.

26. Reich, E. (1957), "Waiting Times when Queues are in Tandem," Ann. Math. Stat., 28, 768-773.

27. Schassberger, R. (Part I, 1977; Part II, 1978), "Insensitivity of Steady-State Distribution of Generalized Semi Markov Processes," Ann. Prob., (Part I) 5, 87-99, (Part II) 6, 85-93.

28. Simon, B. and Foley, R. D. (1979), "Some Results on Sojourn Times in Acyclic Jackson Networks," (to appear).

29. Syski, R. (1960), Introduction to Conjection Theory in Telephone Systems, Oliver and Boyd, Edinburgh.

30. Takacs, L. (1963), "A Single Server Queue with Feedback," Bell Syst. Tech. J., 505-519.

31. Wallace, V. L. (1974), "Algebraic Techniques for Numerical Solution of Queueing Networks," Math. Methods in Queueing Theory, Lecture Notes in Economics and Mathematical Systems, No. 98, Springer-Verlag, New York.

32. Witt, W. (1974), "The Continuity of Queues," Adv. Appl. Prob., 6, 175-183.

Acknowledgement

I would like to thank Robert D. Foley and Burton Simon for many helpful discussions in the preparation of this paper.

DEPARTMENT OF INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH
VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNVIERSITY
BLAKCSBURG, VIRGINIA 24061