

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again -- beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.
5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International

300 North Zeeb Road

Ann Arbor, Michigan 48106 USA

St. John's Road, Tyler's Green

High Wycombe, Bucks, England HP10 8HR

7821907

ZIEGLER, CHAIM
ANALYTIC METHODS FOR DELAY ANALYSIS AT
PACKET-SWITCHED MERGER AND SEPARATION NODES,
CITY UNIVERSITY OF NEW YORK, PH.D., 1978

University
Microfilms
International 300 N ZEEB ROAD, ANN ARBOR, MI 48106

ANALYTIC METHODS FOR DELAY ANALYSIS AT PACKET-
SWITCHED MERGER AND SEPARATION NODES

by

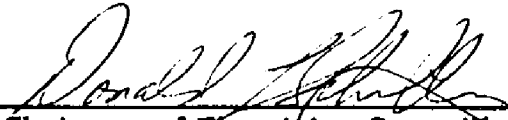
CHAIM ZIEGLER

A dissertation submitted to the Graduate Faculty in
Engineering in partial fulfillment of the requirements
for the degree of Doctor of Philosophy, The City
University of New York.

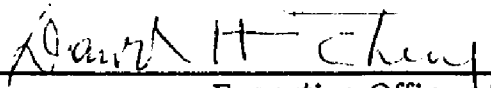
1978

This manuscript has been read and accepted for the Graduate Faculty in Engineering in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

6/1/78
date


Chairman of Examining Committee

6/1/78
date


Executive Officer

Prof. D. L. Schilling (mentor)
Prof. M. Anshel
Dr. A. Konheim
Prof. S. J. Oh
Prof. H. A. Rubin
Dr. J. Salz

Supervisory Committee

ABSTRACT

This dissertation concerns itself with the development of techniques for use in the analysis and modeling of packet switched computer communications networks.

Methods are developed for the analysis of the mean steady-state waiting time experienced by a packet at a merger node in networks of single server queues. Methods yielding exact results are derived for slow or dominant merger nodes. For fast merger nodes, the techniques developed yield results of varying quality. For some subcases exact results are obtained, for others a tight upper bound approximation is obtained and for yet other subcases results are obtained only for specific limiting cases.

Techniques developed for networks of single server queues are generalized and extended so as to apply to networks of multiple server queues. Methods are developed for use at dominant merger nodes in networks with parallel servers between adjacent stages. In addition, analysis methods to be used at separation and combined separation and merger nodes are derived.

Finally, a principle in delay decomposition is introduced. It is shown that the waiting time experienced by a packet may be decomposed into two basic components called self delay and interference delay.

ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude to Professor Donald L. Schilling for the confidence, encouragement, and guidance he afforded me during the course of this research. The many technical discussions we had and his valuable criticisms and suggestions are most appreciated.

The author also wishes to acknowledge the assistance of Prof. Michael Anshel in the formulation of this dissertation. The mathematical expertise which he imparted to me was of immeasurable value in the course of this research.

In addition, the author would like to thank the remaining members of his guidance committee; Dr. Jack Salz (Bell Laboratories) for obtaining reviews of much of the research contained herein, Dr. Alan Konheim (IBM) for his valuable suggestions in extending results of this research, Prof. Se Jeung Oh for the support and confidence he has shown in me, and Prof. Howard Rubin for his time and valuable friendship.

Finally, the author wishes to thank all his colleagues in room T502F, especially Mr. TuVia Apelewicz and Mr. Donald Ucci for their unwavering friendship and help which aided in smoothing the difficult periods that arose from time to time.

The excellent, professional typing of this dissertation was performed by Ms. Connie Engle and Ms. Maie Croner of New York University. Their efforts and professionalism in helping me meet my deadline are deeply appreciated.

The research contained in this dissertation was partially supported by the National Aeronautics and Space Administration under Grants NAS 9-13940, NSG-5013, and NAS 5038.

TABLE OF CONTENTS		Page
CHAPTER 1.	INTRODUCTION	
1.1	Queueing Theory in Computer- Communications Networks	1
1.2	Statement of the Problem	3
1.3	Summary of Prior Work	7
1.4	Summary of Results Obtained	11
CHAPTER 2.	NETWORKS OF SINGLE SERVER QUEUES	
2.1	Current Methodology	26
2.2	Principle of Delay Decomposition: Theory	33
2.3	Example - Exponentially Distributed Interarrival Times	47
2.4	Generalization to Networks with an Arbitrary Number of External Inputs	50
2.5	Networks with Fast Merger Nodes	54
2.6	Message Switching Networks	88
CHAPTER 3.	NETWORKS OF MULTIPLE SERVER QUEUES	
3.1	Networks with Parallel Servers	122
3.2	Delay Analysis at a Separation Node	149
3.3	Delay Analysis at a Combined Separation and Merger Node	160
CHAPTER 4.	CONCLUSION	
4.1	Conclusions	183
4.2	Suggestions for Future Work	196
APPENDIX A.	Derivation of Equation (2.5-17)	192
APPENDIX B.	Conjecture on the Extension of Results Obtained in Section 2.5, Case 2, Subcase 2.1	196
REFERENCES		204

1.1. QUEUEING THEORY IN COMPUTER-COMMUNICATIONS NETWORKS

In recent years, we have witnessed a phenomenal growth in research in the area of computer-communications networks. There have been many papers published on such varied, though interrelated, topics such as network analysis methods, topological designs, routing algorithms, and network protocols. Furthermore, we have seen these networks grow in type and complexity. We have come from basic centralized networks to complex distributed data networks such as the Advanced Research Projects Agency Network (ARPANET). More recently, we have seen the emergence of satellite and ground radio networks such as the ALOHA system. For a thorough review of the current state of the art in the various aspects of computer communications networks, the interested reader is referred to books by Abramson and Kuo [1], Chu [3], Davies and Barber [8], Green and Lucky [10], Kleinrock [11-13], Schwartz [23], and to the various papers present in the special issue of the IEEE Transactions on Communications [24].

In general, any computer-communications network may be viewed as an organized collection of resources which are shared by the various users of the network. Whenever a resource must be shared, it is clear that contention for use of the resource among the prospective users will occur. Indeed, it has been found that user demands on the

various resources of a network may be characterized as sporadic and bursty in nature. Thus, because of the unpredictable nature of the work demands placed on the resources, congestion at the resources will occur. As such, facilities must be provided so that users may queue for use of a given resource. In the performance evaluation of a computer network, these queues, and the delays they introduce, must be taken into consideration. Indeed, one of the major issues which concerns designers of these networks is the lack of a capability to effectively predict such performance measures as network response time, throughput, and resource utilization. Hence, it is clear that queueing theory and models play an important role in the performance modeling and evaluation of computer communications networks. Indeed, other network problems, such as routing, flow control, link capacity assignment, and concentrator placement rely heavily on queueing analysis. As an example, Kleinrock [11,13], has successfully formulated the problem of link capacity assignment as a problem in mathematical programming optimization based on average delay formulae derived from the analysis of queueing models of networks. Therefore, it is clear that queueing theory will continue as the mathematical tool to be used in the performance modeling and evaluation of computer communications networks.

1.2. STATEMENT OF THE PROBLEM

In this dissertation, it is our intent to concentrate on a specific problem which arises in the modeling and performance evaluation of computer communications networks; that is the problem of delay analysis at merger nodes present within computer communications networks.

Fig. 1.2-1 depicts a model for a computer network consisting of three stations which operates according to the following procedure. We assume that there are two classes of users that utilize the network. Users belonging to the first class enter the network at node N_1 where they receive some processing and are then transported to node N_3 where they receive additional processing. On the other hand, users belonging to the second class initially enter the network at node N_2 for processing after which they too journey to node N_3 for additional service. In our initial discussions, we assume that each station or node within the network will contain only a single service facility through which it will process all requests for service. Thus, node N_3 , of Fig. 1.2-1, must process users from both classes through its single service facility. Hence, users of the two classes are said to be merged as they are processed by the service station at node N_3 . Consequently, we refer to node N_3 as a merger node.

For our discussion, we shall consider each user or customer to enter the network to consist of a packet of bits that is to be transmitted between the various nodes within the network. Thus, our server, at a given station of the network, consists of a communications channel, having a certain capacity, over which one packet at a time can be transmitted in a bit by bit manner. In addition, we assume that all packets, regardless of the class they belong to, are of exactly the same length; that is, they all contain exactly the same number of bits. Hence, at a given node of the network, all users of that node's service facility require exactly the same amount of time to be processed. The amount of time that node N_i requires to process a packet is symbolized in Fig. 1.2-1 by α_i ; for example, node N_1 requires α_1 seconds to transmit a packet from itself to node N_3 and node N_3 has a packet service time of α_3 seconds. As for the arrival process of packets, from the external environment into the network, it is assumed to be probabilistic in nature following any general probability distribution such that the mean number of class i packets to arrive per second is given by λ_i , $i = 1, 2$. Because of the stochastic arrival of packets, a packet may arrive at a node and find the server busy with a packet that had arrived previously thus forcing this newly arrived packet to queue for service. Consequently, the

various nodes of the network must have space available for storing packets that arrive and cannot be serviced immediately. For our purposes, we assume that each node has infinite storage capability; that is, each node is assumed to have enough space available so as to store all packets that may arrive, never having to reject a packet for lack of storage space. Finally, we assume that each node, or queue, processes packets in the order of their arrival; that is, each queue operates according to a first-come first-served queueing principle.

As mentioned in the previous paragraph, because of the probabilistic nature of the packet arrivals into the networks, queues will be formed at the various nodes of the network. As such, the total time that a packet spends in the network consists of two basic components; namely waiting time and service time. The service time is the actual time that a packet spends being processed by the various channels over which it travels. On the other hand, waiting (or queueing) time, is that amount of time that a packet wastes as it is forced to queue for service at busy channels. It is in the derivation of methods to be used for the calculation of the waiting time that will be experienced by a packet at a merger node, such as node N_3 of Fig. 1.2-1, that we initially concentrate our efforts. The importance of having good analytic techniques for the calculation of this network characteristic lies

in the realization that these methods may lead to more efficient network designs so as to minimize the amount of time that a packet must waste in traversing its designated route through a given network.

1.3. SUMMARY OF PRIOR WORK

At this point, we will perform a qualitative review of present techniques that are used in the analysis of merger nodes. We shall leave the quantitative results, obtained using present methods, to be introduced in later sections as they become relevant.

Before reviewing methods presently used at merger nodes, let us look at the tandem queueing network depicted by Fig. 1.3-1. Here, we have an n stage network with a single external input packet stream. These tandem networks have been extensively studied by Rubin [17-19] and Friedman [9]. Qualitatively, their main result is that for tandem networks of constant server queues, the overall waiting time that would be experienced by a packet would be completely determined by the stage with the slowest service time. Thus, with respect to the calculation of the waiting time that would be experienced by a packet over the n stage network of Fig. 1.3-1, we could equivalently consider a single node network, as depicted by Fig. 1.3-2, whose service time is equal to the service time of the slowest stage of Fig. 1.3-1, that is, the service time of the single server of Fig. 1.3-2 is given by $\alpha_{\max} = \max(\alpha_1, \alpha_2, \dots, \alpha_n)$. With this result, one could then proceed to calculate the packet waiting at any specific stage of the tandem network of Fig. 1.3-1 by successively increasing the number of nodes considered in calculating the overall waiting time. As an example, if the input stream

of packets follows a Poisson distribution then node N_1 of Fig. 1.3-1 becomes a classical M/D/1 queue for which steady-state waiting time results have been well documented, (see Kleinrock [1], Ch. 5). If we now consider the subnetwork consisting of the first two stages of Fig. 1.3-1, nodes N_1 and N_2 , then the overall steady-state packet waiting time can be determined by considering an M/D/1 queue with service time equal to the maximum of α_1 and α_2 . After this calculation, the waiting time on a packet at node N_2 can be found by simply subtracting the waiting time experienced by a packet at node N_1 from the total waiting experienced by a packet over both N_1 and N_2 . In this manner, one could successively proceed through the entire tandem network and obtain exact results at each stage of the network. A more quantitative discussion of this procedure can be found at the beginning of Chapter 2.

Turning to networks that include merger nodes; such as, the network of Fig. 1.2-1, we find that current methods only yield approximate results for the steady-state waiting time experienced by a packet at a merger node; such as, node N_3 of Fig. 1.2-1.

The first approximation method to be discussed was developed by Kleinrock [11] for external inputs following a Poisson distribution. Referring to Fig. 1.2-1, we note that, for Poisson inputs, nodes N_1 and N_2 are simply M/D/1

queues whose characteristics are well known. However, the output stream of packets from each of these two nodes is no longer Poisson in nature. Thus, the input stream into node N_3 is not Poisson. It is here that we apply the "independence assumption," introduced by Kleinrock, and assume that the arrival process of packets into node N_3 is still Poisson. Consequently, node N_3 may also be viewed as M/D/1 queue, since the joint arrival process formed by two independent Poisson processes is also Poisson. In applying this "independence assumption," it has been found (see Rubin [20]) that results obtained are generally a loose upper bound to actual values. Indeed, this fact will become apparent when we introduce our techniques in later sections of this dissertation. Once again, we note that a more quantitative discussion of this approximation technique can be found in Sec. 2.1.

A second approximation method, which attempted to remove the independence assumption, at merger nodes, was introduced by Rubin [20]. In his method, Rubin would approximate the waiting time, on an arbitrary class packet at node N_3 , to consist of a sum of two parts; the first part being determined by the stochastic character of the class 1 arrival pattern into N_3 , with a deterministic like interference by the class 2 stream and the second part, in a similar manner,

by the stochastic nature of the class 2 arrival stream with a deterministic like interference by the class 1 packets. To calculate these two parts, the three node network would be transformed into two separate tandem networks of two nodes each; the first of which would consist of nodes N_1 and N_3 and the second of which would contain nodes N_2 and N_3 . These networks are depicted by Fig. 1.3-3. For the calculation of the first component of waiting time just mentioned, one used the network of Fig. 1.3-3a. In Fig. 1.3-3a, the class 1 packets still enter at node N_1 where they are processed by a channel, over which they are transmitted to N_3 , at a service rate of α_1 sec/packet. However, at N_3 , the service time has now been modified, to α_3' , so as to account for the interference by the class 2 packets. With this modification, one can then proceed to use methods for tandem networks, as described earlier, so as to calculate the waiting time on a class 1 packet at node N_3 of Fig. 1.3-3a. In a similar manner, Fig. 1.3-3b would be used in determining the second component of waiting time mentioned above. For a more complete and quantitative discussion of Rubin's method, the interested reader is referred to the technical paper (Rubin [20]) in which it was introduced. For purposes of this dissertation, we merely mention that results obtained using this technique are in general a lower bound to actual values. In addition, the tightness of this lower bound tends to vary significantly with different values of α_3 .

1.4. SUMMARY OF RESULTS OBTAINED

In this section, we outline the contents of the dissertation that follows. In addition, we shall qualitatively summarize the main results that we have obtained. The quantitative derivation and discussion of our new techniques and methods are left for the chapters that follow.

Chapter 2 begins, in Sec. 2.1, with a quantitative review of present methods used in network analysis. A thorough discussion of the independence assumption is contained therein and its application to both tandem networks and merger nodes explicitly shown. In addition, Rubin's methods, mentioned in the previous section, are also quantitatively discussed.

In Sec. 2.2, we develop techniques which allow for the exact calculation of the mean steady-state waiting time experienced by a packet at a merger node for cases in which the service time associated with the channel of the merger node is greater than or equal to the service times associated with all nodes through which the packets entering the merger node have already passed; for example, in Fig. 1.2-1, this implies the case where $\alpha_3 = \max(\alpha_1, \alpha_2, \alpha_3)$. The derivation of the technique begins with a comparative analysis of the two networks depicted by Figs. 1.4-1 and 1.4-2. In Fig. 1.4-1, we have a single node network which is being fed by two external packet streams. The arrival process of the

external packet streams is assumed to follow any general probability distribution. The service time of the single server at node N is α sec./packet. In Fig. 1.4-2, we have a three node network. We assume that the two input packet streams are exactly the same as those entering the network of Fig. 1.4-1. In addition, the service time of the channel at each of the three nodes within the network is α sec./packet. Through our discussion, we prove two interesting theorems with respect to equivalencies that exist between these two networks. The first theorem states that the lengths of the busy periods at nodes N of Fig. 1.4-1 and N_3 of Fig. 1.4-2 are identical; that is, busy period by busy period node N_3 of Fig. 1.4-2 will service exactly the same number of packets as node N of Fig. 1.4-1. The second theorem states that, for a given busy period of nodes N and N_3 , the total waiting time experienced by the packets of the given busy period at node N of Fig. 1.4-1 will be identical to the total waiting time that would be experienced by the same packets over the entire network of Fig. 1.4-2; that is, the sum of the waiting time experienced by the class 1 and class 2 packets at node N of Fig. 1.4-1 is identical to the sum of the waiting time experienced by the class 1 packets at nodes N_1 and N_3 and the class 2 packets at nodes N_2 and N_3 of Fig. 1.4-2, for any given busy period of nodes N and N_3 . Through simple division, these two theorems imply that the overall mean steady-state

waiting time experienced by an arbitrary class packet is identical in both networks. Consequently, we can now determine, exactly, the mean steady-state waiting time on an arbitrary class packet at node N_3 of Fig. 1.4-2, a merger node, by subtracting, appropriately weighted, the waiting time on class 1 packets at node N_1 and the waiting time on class 2 packets at node N_2 . Indeed, these results, together with application of Rubin's and Friedman's methods for tandem networks, lead to a technique for the exact determination of the mean steady-state packet waiting at node N_3 of Fig. 1.2-1 for cases in which $\alpha_3 = \max(\alpha_1, \alpha_2, \alpha_3)$.

In addition to the above mentioned results, we introduce a principle in delay decomposition. It is shown how the waiting time experienced by a packet at a node may be decomposed into two basic components which we refer to as self delay and interference delay.

In Sec. 2.3, we quantitatively illustrate the techniques developed in Sec. 2.2 through a specific example having external input streams following Poisson distributions.

Following this, in Sec. 2.4, we show that the techniques developed in Sec. 2.2 for networks with only two external inputs may be extended and generalized so as to apply to merger nodes with any number of external inputs so long as the service time of the merger node is greater than or equal to the service time of any prior node which feeds packets into the merger node.

Having established a technique for slow merger nodes, we turn our attention, in Sec. 2.5, to analysis methods for fast merger nodes; that is, for merger nodes whose packet service time is less than that of at least one of the channels which are feeding packets into the merger node. We show that, for the network of Fig. 1.2-1, we must decompose this problem into two separate cases; namely,

those in which $\alpha_3 > \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$ and those in which $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$.

This decomposition is necessitated by the fact that when

$\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$ it is impossible for node N_3 of Fig. 1.2-1 to

become unstable; that is, the packet waiting time can never become infinite. Indeed, it is proven that, for

$\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, the number of packets waiting for service can

never exceed one, thus implying that the waiting time on any packet at node N_3 can never exceed α_3 seconds. For the

case $\alpha_3 > \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, an approximation technique, yielding a

tight upper bound result, is developed. The technique consists of setting the service time of each channel feeding the merger node, whose service time is greater than the service time of the merger node, equal to the service time of the merger node; that is, in Fig. 1.2-1, set

$\alpha_i = \alpha_3$ for each $\alpha_i > \alpha_3$, $i = 1, 2$. The approximation technique is developed and then compared to existing methods and shown to provide vastly improved results. Turning back to the case $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, we find that we must further subdivide this case into two parts; namely, $\alpha_3 \leq \frac{1}{2} \min(\alpha_1, \alpha_2)$ and $\alpha_3 > \frac{1}{2} \min(\alpha_1, \alpha_2)$. For $\alpha_3 \leq \frac{1}{2} \min(\alpha_1, \alpha_2)$, we derive a method which yields exact results at N_3 . It is shown that these results apply at N_3 regardless of the distributions followed by the input packet streams. Furthermore, it is shown under which conditions these results may be extended to merger nodes with more than two input streams. As for the subcase $\alpha_3 > \frac{1}{2} \min(\alpha_1, \alpha_2)$, we derive some exact results which are applicable at the merger node under some specific and limiting conditions. These results are developed and a discussion of the difficulties encountered in attempting to remove the restrictions on these results is discussed.

In the final section of Chapter 2, Sec. 2.6, we relax the restriction that customers entering the network consist of single packet messages and allow for multiple packet messages. In addition, the nodes of the network are now assumed to operate in a message switching mode; such that, if a server, of a given node, begins to process the first packet from a given message it will proceed to process all packets belonging to that particular message before granting

service to packets from any other messages. In the section, it is shown how many of the methods, developed earlier for single packet messages, are directly extendable to merger nodes which process multiple packet messages in a message switching mode of operation.

In Chapter 3, we relax the restriction that each node of networks considered have only a single server and, instead, allow each node of the network to contain multiple servers with which to process packets. Allowing multiple server queues enables the discussion of networks containing parallel servers between adjacent stages. In addition, multiple server queues also permit the presence of separation nodes within networks; that is, nodes at which merged packet streams are decomposed into separate classes such that the various classes may be processed over different channels and transmitted to different destinations.

Sec. 3.1 begins with a quantitative review of Friedman's [9] results for tandem networks containing parallel servers between adjacent stages. The concept of nodal dominance is introduced and explained. Following this, networks with parallel servers, containing merger nodes are discussed. It is shown that techniques developed in Sec. 2.2 for slow merger nodes are directly extendable so as to allow for the exact waiting time analysis at dominant merger nodes.

In Sec. 3.2, we study analysis methods for separation nodes. Fig. 1.4-3 illustrates a network containing a separation node. In Fig. 1.4-3, class 1 and class 2 packets enter the network at node N_1 where they are merged and processed by a single channel, at a rate of α_1 sec./packet, over which they are transmitted to node N_2 . At N_2 the merged packet stream is separated; such that, the class 1 and class 2 packets are processed over separate channels. Several approximation techniques to be used at separation nodes are developed and compared for accuracy. An approach in which a separation node may be viewed as a non-separation node utilizing a preemptive priority discipline is introduced. At present, this approach, after some additional assumptions is shown to yield some useful approximation methods.

Finally, in Sec. 3.3, we study networks which contain combined separation and merger nodes. Fig. 1.4-4 depicts such a network. Class 1 and class 2 packets are merged at node N_1 and transmitted over a single channel to node N_3 . At N_3 , the merged stream is separated and proceeds over separate routes. Class 3 packets enter the network at N_2 and then proceed to N_3 . Upon arrival at N_3 , the stream of class 3 packets is merged with the stream of class 1 packets, that have been separated from the class 2 packets, and the combined stream of class 1 and class 3 packets is then

processed by a single server having service time α_3 sec./packet. It is shown how methods developed in Secs. 2.2 and 2.5 for pure merger nodes may be applied to combined separation and merger nodes so as to yield useful approximation methods. Furthermore, for the case $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, it is shown that the exact results derived in Sec. 2.5 for merger nodes also apply, as exact results, for combination nodes.

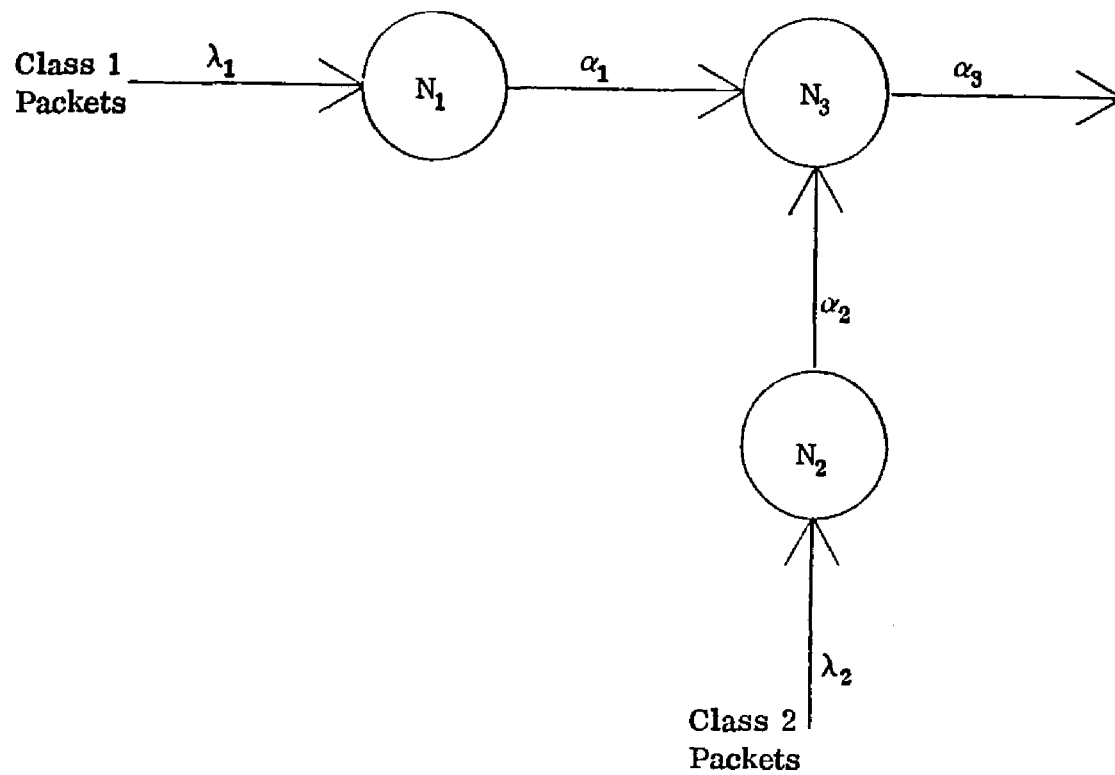


Fig. 1.2-1: A Three Node Network Having A Merger Node At N_3

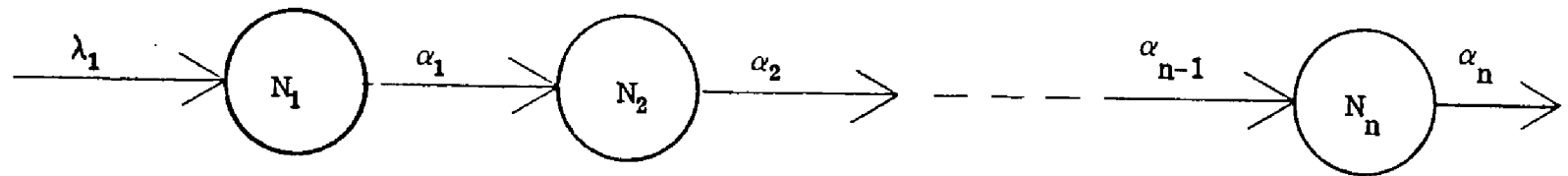


Fig. 1.3-1: An " n " Stage Tandem Queueing Network

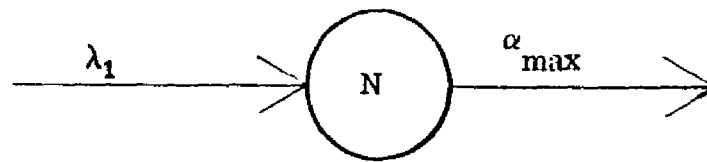
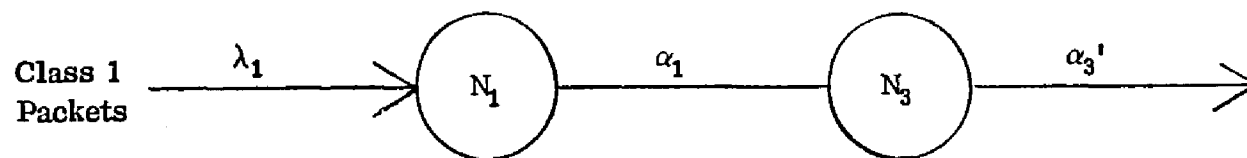
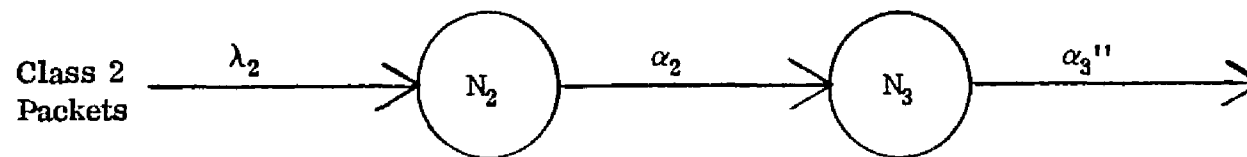


Fig. 1.3-2: A Single Node Network With Equivalent Packet Waiting Time As That
Of Fig. 1.3-1 When $\alpha_{\max} = \max(\alpha_1, \alpha_2, \dots, \alpha_n)$



(a)



(b)

Fig. 1.3-3: The Two Tandem Networks Used In Rubin's Approximate Technique For A Merger Node

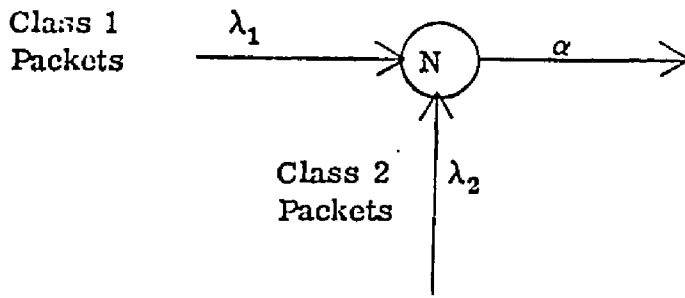


Fig. 1.4-1: A Queuing Network Consisting of a Single Server Queue and Two External Input Streams

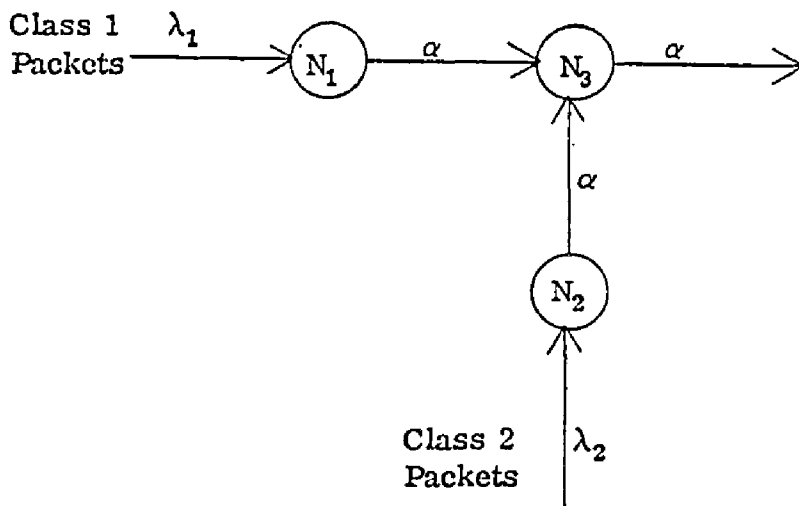


Fig. 1.4-2: A Queuing Network Consisting of Three Single Server Queues and Two External Input Streams

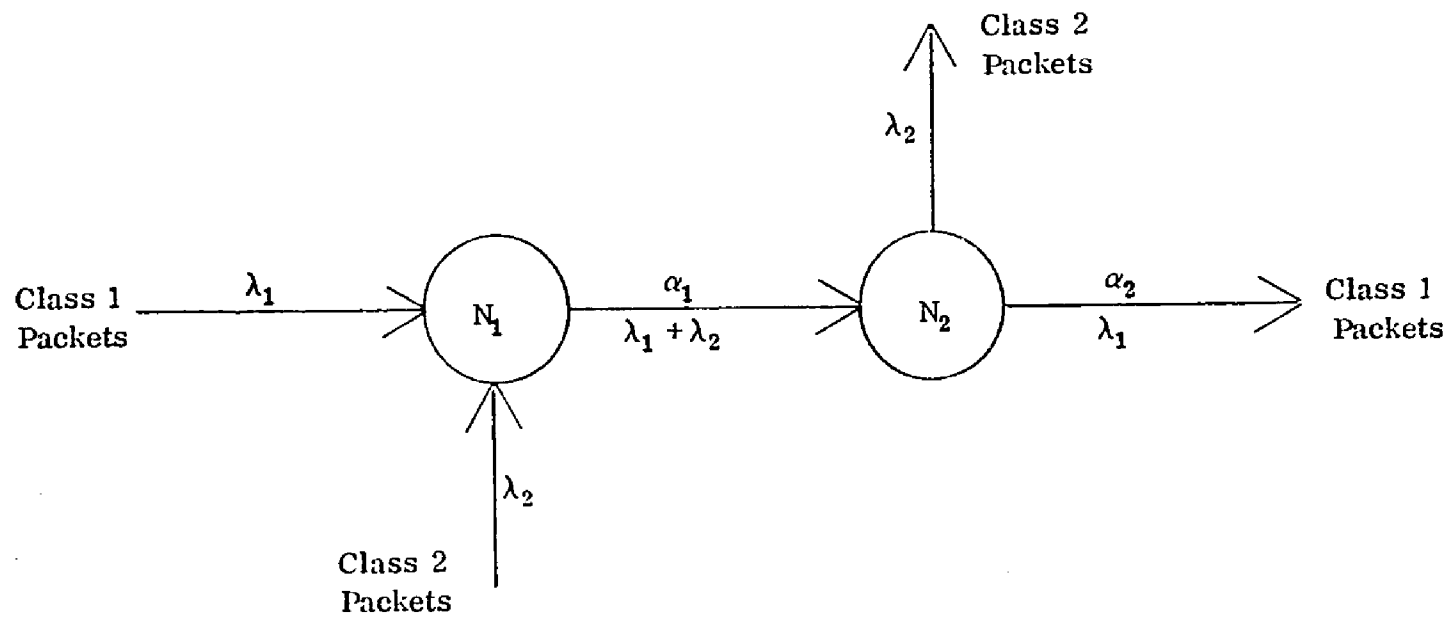


Fig. 1.4-3 : A Queueing Network Consisting of a Merger Node Followed by a Separation Node

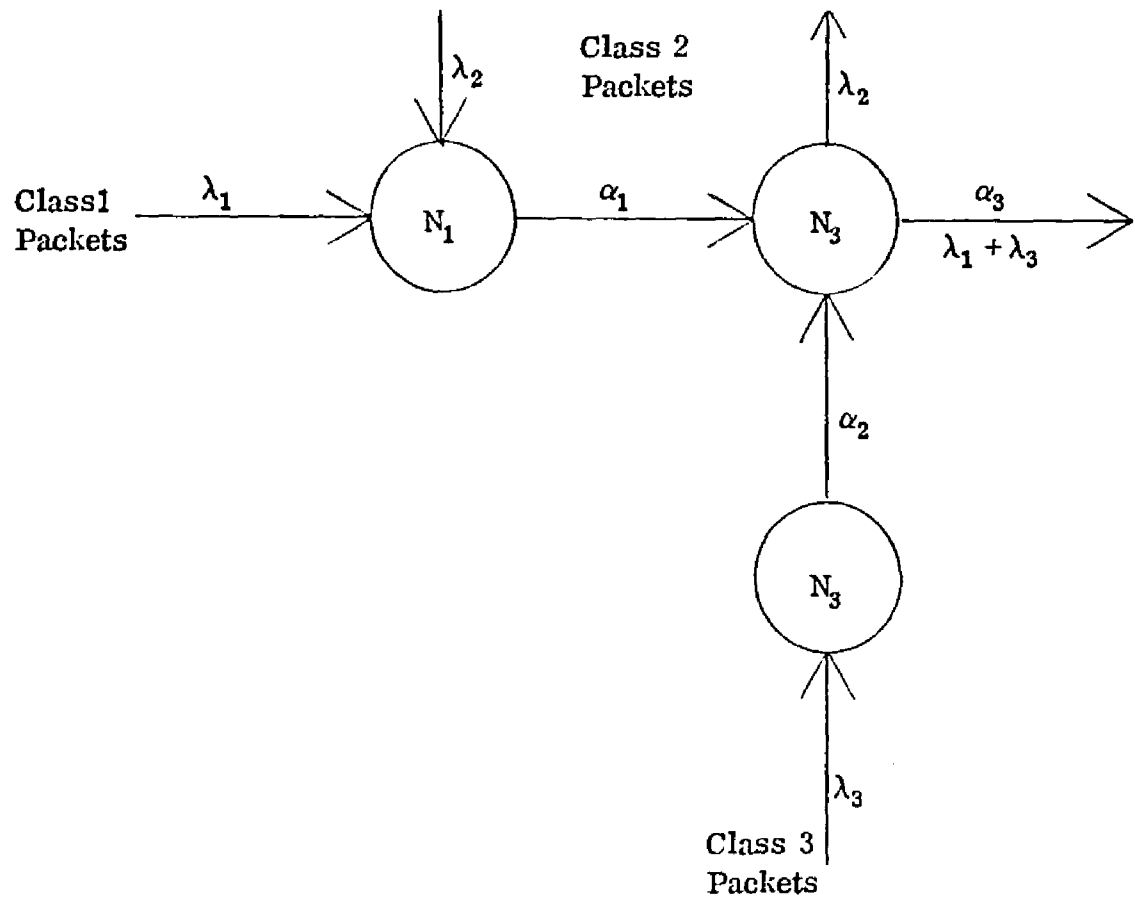


Fig. 1.4-4 : A Queuing Network in which Node N_3 is a Combined Merger and Separation Node

2.1 Current Methodology

In order to design more efficient computer communications networks, it would be advantageous to have an analytical technique for the modeling and analysis of various performance measures; such as, the steady-state mean message waiting and response times, the mean queue length at each processing station and the throughput of the network considered. The major difficulty in obtaining accurate analytical methods has been the strong interdependence that exists among the various queues of a given network. To overcome this difficulty most of the past work in this area has made use of an "independence assumption" first introduced by Kleinrock [11]. To understand this independence assumption and its application, let us look at Figure 2.1-1.

In Figure 2.1-1 we see a model of a two stage tandem queueing network. In this network, we assume that messages initially enter the network at node N_1 , arriving, from the external environment, randomly in time according to a Poisson process with intensity λ (messages/sec.); that is,

$$P\{k \text{ arrivals in a time interval } t\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots \quad (2.1-1)$$

At N_1 these messages are processed, according to a first-come first-served queueing discipline, by a channel whose capacity

is equal to C_1 (bits/sec.). Once processed at N_1 , the messages are transmitted to node N_2 where they receive additional processing. If we now further assume that N_1 has an exponential server; that is, the length of each message is exponentially distributed with a mean of β (bits/message); such that,

$$P\{\text{message length} \leq x\} = 1 - e^{-\beta x} \quad x \geq 0 \quad (2.1-2)$$

and that N_1 has unlimited storage available, then node N_1 is said to constitute an M/M/1 queueing facility, the statistics of which are well known and documented [12, pp.94-99].

As an example, if we define

- $\bar{\alpha}_i \triangleq$ the mean service time of a message at node N_i
- $\bar{W}^{(i)} \triangleq$ the steady-state mean message waiting time (queueing time) at node N_i
- $\bar{N}_q^{(i)} \triangleq$ the steady-state mean queue length at node N_i

one could easily calculate the steady-state mean waiting time experienced by a message at node N_1 as

$$\bar{W}^{(1)} = \frac{\rho_1 \bar{\alpha}_1}{(1-\rho_1)} \quad (\text{sec/message}) \quad (2.1-3)$$

where $\bar{\alpha}_1 = \frac{\beta}{C_1}$ (sec/message) and $\rho_1 = \lambda_1 \bar{\alpha}_1 \triangleq$ the utilization factor of the queue.

Furthermore, using Little's Theorem [14], which relates the average time spent by a message on queue to the number of

of messages on the queue we find that the steady-state mean number of messages on queue at node N_1 is

$$\bar{N}_q^{(1)} = \lambda \bar{W}^{(1)} = \frac{\rho_1^2}{(1-\rho_1)} \text{ (messages)}. \quad (2.1-4)$$

Other performance measures for an M/M/1 queue are easily derived and can be found in the literature.

As for the output process from an M/M/1 queue, Burke [2] has shown that this process is Poisson; thus, the arrival process at node N_2 of Figure 2.1-1 is also Poisson. However, it is clear that the processing time required by a message at node N_2 , is dependent on the processing time that the message had required at node N_1 . This is so because the length of a message does not change as it travels from stage to stage in the network. It is here that the independence assumption is applied. This assumption amounts to rechoosing the length of a given message independently at each node of a given network. Thus, in Figure 2.1-1, node N_2 is now assumed to constitute an M/M/1 queueing system and calculations for steady-state mean queueing time and queue length, $\bar{W}^{(2)}$ and $\bar{N}_q^{(2)}$ respectively, would yield approximations of

$$\bar{W}^{(2)} \approx \frac{\rho_2 \bar{\alpha}_2}{(1-\rho_2)} \text{ (sec./message)} \quad (2.1-5)$$

$$\bar{N}_q^{(2)} \approx \frac{\rho_2^2}{(1-\rho_2)} \text{ (messages)} \quad (2.1-6)$$

For packet-switching networks, where each message is assumed to consist of a single packet and each packet is of

fixed length, the independence assumption must be redefined in the following manner: Since each packet is of fixed length, rather than being distributed exponentially, the amount of time required to process each packet at a given channel is constant. Thus, in this case, each stage of a given network can be represented by a constant server queue whose service time is a function of that node's channel capacity. For Figure 2.1-1, assuming a fixed packet length of β (bits/packet), this would mean that node N_1 now represents what is called an M/D/1 queueing system with a constant packet service time of $\alpha_1 = \beta/C_1$ (sec/packet). Steady-state results for this type of queueing system are well known [12, ch.5]. As an example, the steady-state mean waiting time at node N_1 would now be

$$\bar{w}^{(1)} = \frac{\rho_1 \alpha_1}{2(1-\rho_1)} . \quad (2.1-7)$$

Furthermore, since we now have a constant server queue, Burke's theorem no longer applies to the output process of node N_1 and it is here that the independence assumption is made: we assume that the arrival process at node N_2 remains Poisson. Thus, node N_2 can now also be treated as an M/D/1 queueing system. Consequently, the approximate steady-state mean waiting time at N_2 would be

$$\bar{w}^{(2)} \approx \frac{\rho_2 \alpha_2}{2(1-\rho_2)} . \quad (2.1-8)$$

It has been shown by Rubin [20] that steady-state results,

calculated using this independence assumption, generally constitute a loose upper bound to the actual values as obtained by use of computer simulation programs. More recently, Friedman [9] and Rubin [17-19] have been able to remove this independence assumption and obtain exact steady-state waiting time distributions for packet-switching communication paths in isolation; that is, for networks in which messages originating from different external sources never interfere with each other. This was done by realizing that the overall network waiting time experienced by a packet in an n-channel communications path, as shown by Figure 2.1-2, is determined by the slowest channel in that path. Thus, with respect to overall path waiting time, the entire n-channel communications path can be viewed as a single M/D/1 queue whose service time is equal to that of the slowest channel in the path. Applying this principle to the network of Figure 2.1-1, we would have the following two cases:

$$\text{Case 1: } C_2 \geq C_1 \quad (\alpha_2 \leq \alpha_1)$$

In this case, the overall steady-state mean packet waiting time, \bar{S} , will be determined by the channel capacity of the server at node N_1 . Furthermore, it is clear, that the waiting time experienced by a packet at node N_2 will be zero. Thus, we have

$$\begin{aligned} \bar{S} &= \bar{W}^{(1)} = \frac{\rho_1 \alpha_1}{2(1-\rho_1)} & (2.1-9) \\ \bar{W}^{(2)} &= 0 \end{aligned}$$

Case 2: $C_2 \leq C_1$ ($\alpha_2 \geq \alpha_1$)

In this case, the overall steady-state mean waiting time is determined by C_2 (or α_2) and is given by

$$\bar{S} = \frac{\rho_2 \alpha_2}{2(1-\rho_2)} \quad (2.1-10)$$

Consequently, the mean waiting time experienced by a packet at node N_2 will be

$$\begin{aligned} \bar{W}^{(2)} &= \bar{S} - \bar{W}^{(1)} \\ &= \frac{\rho_2 \alpha_2}{2(1-\rho_2)} - \frac{\rho_1 \alpha_1}{2(1-\rho_1)} \end{aligned} \quad (2.1-11)$$

However, as pointed out earlier, this method applies only to communication paths. When Rubin [20] applied his ideas to networks in which packet interference is allowed to occur, such as the network shown in Figure 2.1-3, he too was able to obtain only approximate results.

In the section that follows, we shall develop a principle which we call "delay decomposition." We shall show that, using this principle, it is possible to decompose the waiting time experienced by a packet at a constant server channel into two basic components which we shall call self-delay and interference-delay. In the next section we shall develop the theory and discuss more fully the meaning and application of our principle. It will be shown that for networks in which

the merger node (that is, the node at which the packet streams are multiplexed) has the slowest service time of any node up to that point, the principle of the delay decomposition will yield exact results. In succeeding sections we shall look at networks in which the merger node is not the slowest and find that application of our principle will lead to exact results in some instances and to a tight upper bound approximation in others.

2.2 Principle of Delay Decomposition: Theory

Consider the two queueing networks depicted by Figs. 2.2-1 and 2.2-2. In Figure 2.2-1, node N represents a queueing facility which is being fed by two independent external packet streams. We assume that the interarrival time between packets, for each of the two input streams, is governed by some general probability distribution with mean interarrival time of $1/\lambda_1$ and $1/\lambda_2$ (sec/packet) for the class 1 and class 2 streams respectively, $\lambda_1, \lambda_2 \geq 0$. Furthermore, we assume that all packets are of a constant length and that node N has a single server with a constant service time of α (sec/packet). Figure 2.2-2, on the other hand, represents a network of three single server queues, each of which is assumed to have a constant service time of α (sec/packet). In addition, the class 1 and class 2 input streams entering the network of Figure 2.2-2 are assumed to be the exact same streams entering the network of Figure 2.2-1. Finally, for both networks, we assume that each queue has infinite storage capability and that each queue operates according to a first-come first-served queueing discipline.

As time passes, each queue shown will alternate between periods of time where its server is busy and periods of time where its server is idle [12, p. 206]. As such, we define the following quantities:

$t_i \triangleq$ the arrival time of the i^{th} arbitrary class packet into the network

$M_j \triangleq$ the number of packets serviced at node N of Figure 2.2-1 during its j^{th} busy period.

$M_j^{(k)} \triangleq$ the number of packets serviced at node N_k of Figure 2.2-2 during its j^{th} busy period,
 $k = 1, 2, 3$

$I_j \triangleq$ the duration, in time, of the j^{th} idle period at node N of Figure 2.2-1

$I_j^{(k)} \triangleq$ the duration of the j^{th} idle period at node N_k of Figure 2.2-2

We will now proceed to derive some interesting equivalencies that exist between the two networks considered.

Theorem 2.2-1:

The number of packets serviced at node N of Figure 2.2-1 during its j^{th} busy period equals the number of packets serviced at node N_3 of Figure 2.2-2 during its j^{th} busy period; that is,

$$M_j = M_j^{(3)}, \quad j = 1, 2, 3, \dots \quad (2.2-1)$$

Proof: Let us begin by looking at the first busy period at node N of Figure 2.2-1. We observe that in order to have a busy period of length M_1 , the packet arrival times must be

$$\begin{aligned}
 t_1 &= t_1 && (2.2-2) \\
 t_1 &\leq t_2 \leq t_1 + \alpha \\
 t_2 &\leq t_3 \leq t_1 + 2\alpha \\
 &\vdots \\
 t_{M_1-1} &\leq t_{M_1} \leq t_1 + (M_1-1)\alpha
 \end{aligned}$$

Furthermore, since it will take $M_1\alpha$ seconds to service these M_1 packets, the arrival time of the $(M_1+1)^{st}$ packet, which will be the first packet of the second busy period at node N, must be

$$t_{M_1+1} > t_1 + M_1\alpha . \quad (2.2-3)$$

This arrival time will insure a busy period of exactly M_1 packets.

We now turn to the network of Figure 2.2-2. Let us define

- $t'_i \triangleq$ the arrival time at node N_3 of Figure 2.2-2 of the packet that initially entered the network at time t_i ,
- $T'_i \triangleq$ the arrival time at node N_3 of the i^{th} packet to arrive at node N_3 ,
- $D'_i \triangleq$ the departure time from node N_3 of the i^{th} packet to have arrived at node N_3 .

With the external arrival times as before (2.2-2,2.2-3), we find that

$$\begin{aligned}
 t_1' &= t_1 + \alpha & (2.2-4) \\
 t_1 + \alpha &\leq t_2' \leq t_1 + 2\alpha \\
 t_2 + \alpha &\leq t_3' \leq t_1 + 3\alpha \\
 &\vdots \\
 t_{M_1-1} + \alpha &\leq t_{M_1}' \leq t_1 + M_1\alpha \\
 t_{M_1+1}' &> t_1 + (M_1+1)\alpha
 \end{aligned}$$

Furthermore, from (2.2-4), one can calculate that

$$\begin{aligned}
 T_1' &= t_1 + \alpha & (2.2-5) \\
 T_2' &\leq t_1 + 2\alpha \\
 T_3' &\leq t_1 + 3\alpha \\
 &\vdots \\
 T_{M_1}' &\leq t_1 + M_1\alpha \\
 T_{M_1+1}' &> t_1 + (M_1+1)\alpha
 \end{aligned}$$

In addition, using (2.2-5) and the fact that the service time at N_3 is a constant, we find

$$\begin{aligned}
 D_1' &= t_1 + 2\alpha & (2.2-6) \\
 D_2' &= t_1 + 3\alpha \\
 &\vdots \\
 D_{M_1}' &= t_1 + (M_1+1)\alpha
 \end{aligned}$$

Thus, from (2.2-5) and (2.2-6), we can conclude that

$$\begin{aligned}
 T_{i+1}' &\leq D_i' , \quad i = 1, 2, \dots, M_1-1 & (2.2-7) \\
 T_{M_1+1}' &> D_{M_1}'
 \end{aligned}$$

Equation (2.2-7) implies that we have had a busy period, at node N_3 , whose length was exactly M_1 packets; thus, $M_1^{(3)} = M_1$. Furthermore, by simple induction, this result extends to all subsequent busy periods. Hence,

$$M_j^{(3)} = M_j, \quad j = 1, 2, \dots \quad \text{Q.E.D.}$$

Thus, we have shown that busy period by busy period the two networks considered will service the exact same number of packets. Using this fact, we can immediately derive the following corollaries.

Corollary 2.2-1: The probability distributions of the number of packets serviced during a busy period at nodes N of Figure 2.2-1 and N_3 of Figure 2.2-2 are identical; that is,

$$P\{M_j = m\} = P\{M_j^{(3)} = m\}, \quad m = 1, 2, \dots \quad (2.2-8)$$

Corollary 2.2-2: The duration of the j^{th} idle period at node N of Figure 2.2-1 is equal to the duration of the j^{th} idle period at node N_3 of Figure 2.2-2; that is,

$$I_j = I_j^{(3)}, \quad j = 1, 2, \dots \quad (2.2-9)$$

Corollary 2.2-3: The probability distributions of the duration of the j^{th} idle period at nodes N and N_3 are identical; that is,

$$P\{I_j \leq x\} = P\{I_j^{(3)} \leq x\}, \quad x > 0. \quad (2.2-10)$$

Corollaries 2.2-2 and 2.2-3 can be proven by considering the time between the departure of the last packet of a given busy period and the arrival of the first packet of the succeeding busy period both at node N of Figure 2.2-1 and node N_3 of Figure 2.2-2.

We now proceed to derive a second theorem for the two networks considered.

Theorem 2.2-2:

For the j^{th} busy period of nodes N and N_3 , the total waiting time (queueing time) experienced by the M_j packets in the network of Figure 2.2-1 equals the total waiting time experienced by the $M_j^{(3)}$ packets over the entire network of Figure 2.2-2.

Proof: With all arrival times (e.g. t_i , t_i' , T_i') now referring to the j^{th} busy period, let us define

$W_i \triangleq$ the waiting time, at node N of Figure 2.2-1 on the i^{th} packet to arrive in the j^{th} busy period

$W_i^{(k)} \triangleq$ the waiting time, at node N_k of Figure 2.2-2 on the i^{th} packet, belonging to the j^{th} busy period of node N_3 , to arrive at node N_k , $k = 1, 2, 3$.

We begin by calculating the total waiting time experienced by the M_j packets of the j^{th} busy period at node N of Figure 2.2-1.

From general queueing theory [12, p. 277], we can, at node N, write that

$$W_1 = 0 \quad (2.2-11)$$

$$W_{i+1} = [W_i + \alpha - (t_{i+1} - t_i)]^+, \quad i=1,2,\dots$$

$$\text{where } [x]^+ = \max(0, x).$$

Furthermore, during a busy period the bracketted expression in (2.2-11) must by definition be nonnegative.

Thus, for a busy period of length M_j , (2.2-11) yields

$$W_1 = 0 \quad (2.2-12)$$

$$W_2 = t_1 - t_2 + \alpha$$

$$W_3 = t_1 - t_3 + 2\alpha$$

$$\vdots$$

$$W_{M_j} = t_1 - t_{M_j} + (M_j - 1)\alpha$$

Consequently, the total waiting time in this busy period is

$$\begin{aligned} W_{\text{total}} &= \sum_{i=1}^{M_j} W_i = \sum_{i=1}^{M_j} t_1 - t_i + (i-1)\alpha = M_j t_1 - \sum_{i=1}^{M_j} t_i \\ &\quad + [M_j(M_j-1)/2]\alpha. \end{aligned} \quad (2.2-13)$$

We will now prove our theorem by mathematical induction.

In the case $M_j = M_j^{(3)} = 1$, it is clear that the total waiting time, in both networks, is identically zero. Let us now assume that (2.2-13) is true for both networks for $M_j = M_j^{(3)} \leq m$. We will proceed to show that this assumption implies that for $M_j = M_j^{(3)} = m+1$, the total waiting

time experienced in each of the two networks by these $m+1$ packets will still be the same.

For Figure 2.2-1, the extra waiting time introduced by having M_j equal to $m+1$ instead of m , is seen from (2.2-12) to be

$$(W_{\text{extra}})_{\text{Fig. 2.2-1}} = W_{m+1} = t_1 - t_{m+1} + m\alpha \quad (2.2-14)$$

We now turn to Figure 2.2-2 where we assume (arbitrarily) that this $(m+1)^{\text{st}}$ packet, arriving at time t_{m+1} , enters the network at node N_1 . This packet will then arrive at node N_3 at time

$$t'_{m+1} = t_{m+1} + W_k^{(1)} + \alpha \quad (2.2-15)$$

where we have assumed this packet to be the k^{th} , $1 \leq k \leq m+1$, to arrive at N_1 . At node N_3 we can write (similar to (2.2-11)), that

$$W_1^{(3)} = 0 \quad (2.2-16)$$

$$W_{i+1}^{(3)} = [W_i^{(3)} + \alpha - (T'_{i+1} - T'_i)]^+, \quad i=1,2,\dots$$

Thus, for a busy period of length $(m+1)$ packets we find, from (2.2-16), that

$$W_1^{(3)} = 0 \quad (2.2-17)$$

$$W_2^{(3)} = T'_1 - T'_2 + \alpha$$

$$W_3^{(3)} = T'_1 - T'_3 + 2\alpha$$

$$\vdots$$

$$W_m^{(3)} = T'_1 - T'_m + (m-1)\alpha$$

$$W_{m+1}^{(3)} = T'_1 - T'_{m+1} + m\alpha$$

where, from (2.2-5), $T'_1 = t_1 + \alpha$.

In general, this $(m+1)^{\text{st}}$ packet will be the n^{th} packet of the j^{th} busy period to arrive at node N_3 , $1 \leq n \leq m+1$; that is, $t'_{m+1} = T'_n$. Furthermore, from (2.2-17), we see that the extra waiting time, at node N_3 , caused by having $M_j^{(3)}$ equal to $m+1$ instead of m , is given by

$$W_{\text{extra}}^{(3)} = T'_1 - T'_n + m\alpha \quad (2.2-18)$$

We now substitute for T'_1 and T'_n by using relations derived from (2.2-5) and (2.2-15) respectively and find

$$W_{\text{extra}}^{(3)} = t_1 - t_{m+1} - W_k^{(1)} + m\alpha \quad (2.2-19)$$

Consequently,

$$\begin{aligned} (W_{\text{extra}})_{\text{Fig. 2.2-2}} &= W_{\text{extra}}^{(1)} + W_{\text{extra}}^{(3)} & (2.2-20) \\ &= W_k^{(1)} + W_{\text{extra}}^{(3)} \\ &= t_1 - t_{m+1} + m\alpha. \end{aligned}$$

On comparing (2.2-14) and (2.2-20), we find that

$$(W_{\text{extra}})_{\text{Fig.2.2-1}} = (W_{\text{extra}})_{\text{Fig.2.2-2}} \quad (2.2-21)$$

Thus, the total waiting time in the j^{th} busy period, for $M_j = M_j^{(3)} = m+1$, is the same in both networks. This completes our induction procedure and proves our theorem.

Q.E.D.

In summary, Theorems 2.2-1 and 2.2-2 have shown that for any given busy period both the number of packets serviced and the total waiting time experienced by these

packets will be equivalent for the two networks considered. This leads directly to the following two corollaries.

Corollary 2.2-4: For the networks of Figures 2.2-1 and 2.2-2, the probability distributions of the total waiting time experienced by the $M_j = M_j^{(3)}$ packets of the j^{th} busy period are identical; that is,

$$P\left\{\sum_{i=1}^{M_j} W_i < x\right\} = P\left\{\left[\sum_{i=1}^{K_1} W_i^{(1)} + \sum_{i=1}^{K_2} W_i^{(2)} + \sum_{i=1}^{M_j^{(3)}} W_i^{(3)}\right] < x\right\}, \quad x \geq 0, \quad (2.2-22)$$

where K_1 and K_2 are equal to the number of class 1 and class 2 packets, respectively, of the j^{th} busy period of node N_3 , and $K_1 + K_2 = M_j^{(3)}$.

Before stating our final corollary, we must introduce the following quantities:

- $\bar{w}_{(j)}^{(k)} \triangleq$ the steady-state mean waiting time on a class j packet at node N_k ,
- $\bar{w}^{(k)} \triangleq$ the steady-state mean waiting time on an arbitrary class packet at node N_k ; that is,
 $\bar{w}^{(k)} = \sum_j \bar{w}_{(j)}^{(k)} P\{\text{a packet is class } j\},$
- $\bar{s} \triangleq$ the steady-state overall mean waiting time experienced by an arbitrary class packet; that is, $\bar{s} = \sum_k \bar{w}^{(k)}$.
- $\bar{s}_{(j)} \triangleq$ the steady-state overall mean waiting time experienced by a class j packet; that is,
 $\bar{s}_{(j)} = \sum_k \bar{w}_{(j)}^{(k)}.$

We now state our corollary which we believe to be one of the more important results of this dissertation.

Corollary 2.2-5: For the networks of Figures 2.2-1 and 2.2-2, the steady-state overall mean waiting time experienced by an arbitrary class packet is the same in both networks; that is,

$$(\bar{S})_{\text{Fig.2.2-1}} = (\bar{S})_{\text{Fig.2.2-2}} \quad (2.2-23)$$

Corollary 2.2-5 implies an important concept in delay decomposition which can best be explained in the following manner.

Let us consider the waiting time experienced by a class 1 packet (arbitrary) as it travels through the network of Figure 2.2-2. We note that, if $\lambda_2 = 0$, any waiting time experienced by a class 1 packet would occur totally at node N_1 . This is so because the network consists entirely of queues with equivalent constant service times. We shall refer to this waiting time as self delay since it is due solely to the arrival process of the class 1 packets. In addition, we further note that, when $\lambda_2 > 0$ any additional waiting time that a class 1 packet will experience must occur at node N_3 and that this additional delay is due solely to the merger or interference of the two packet streams at node N_3 . We shall call this type of waiting time interference delay. Thus, we have the important concept that the total overall waiting time

experienced by a packet consists of two basic components; namely, self delay and interference delay. Consequently the steady-state mean waiting time on an arbitrary packet at node N of Figure 2.2-1 must consist of the sum of its steady-state mean self delay and its steady-state mean interference delay. Furthermore, the mean self delay experienced by a packet must be the same in both networks. Consequently, it is implied that, with respect to the steady-state mean waiting time of an arbitrary class packet, a node that consists of the form of Figure 2.2-1 may be replaced by or decomposed into the network of Figure 2.2-2.

As an application of this decomposition principle and (2.2-23), one could proceed to calculate the steady-state mean waiting time on an arbitrary class packet at node N_3 of Figure 2.2-2 and find

$$\begin{aligned} \bar{W}^{(3)} &= (\bar{S})_{\text{Fig. 2.2-2}} - \bar{W}^{(1)} - \bar{W}^{(2)} && (2.2-24) \\ &= (\bar{S})_{\text{Fig. 2.2-1}} - \bar{W}_{(1)}^{(1)} P\{\text{a packet is class 1}\} \\ &\quad - \bar{W}_{(2)}^{(2)} P\{\text{a packet is class 2}\}. \end{aligned}$$

As a second application, one can demonstrate that the five queueing networks depicted in Figure 2.2-3 have equivalent overall steady-state mean packet waiting time under the condition that the service time at the merger node (node N) is the slowest in the network; that is,

$$\alpha_n = \max(\alpha_{11}, \alpha_{12}, \dots, \alpha_{21}, \alpha_{22}, \dots, \alpha_n). \quad \text{This equivalency}$$

can be shown in a step-by-step manner beginning with the network of Figure 2.2-3e. In this network, the class 1 packets pass through an arbitrarily long path of tandem queues after which they enter node N. Along the tandem path, each node has a constant packet service time given by $\alpha_{11}, \alpha_{12}, \dots$ respectively, while at node N the packet service time is α_n . Similarly, the class 2 packets traverse a path of queues with service times of $\alpha_{21}, \alpha_{22}, \dots$ and then enter node N where they are merged with the class 1 stream. We make note of the fact that the various service times in the network (i.e. $\alpha_{11}, \alpha_{12}, \dots, \alpha_{21}, \alpha_{22}, \dots, \alpha_n$) need not all be the same. Furthermore, the number of nodes traversed by the class 1 packets (i.e. the path length) must neither be the same as the number that are traversed by the class 2 packets. On analysis, one notes that node N of Figure 2.2-3e is of the form of the network of Figure 2.2-1. Thus, node N may be replaced by (or decomposed into) a network of the form of Figure 2.2-2. Consequently, we have transformed the network of Figure 2.2-3e into the network of Figure 2.2-3d. We now turn our attention to Figure 2.2-3d where we apply a principle introduced by Friedman [9] and Rubin [17-19] for paths of tandem queues. Simply stated, this principle says that, over a path of constant server tandem queues, the overall waiting time is determined solely by the slowest server in the path. Therefore, if $\alpha_n = \max(\alpha_{11}, \alpha_{12}, \dots, \alpha_{21}, \alpha_{22}, \dots, \alpha_n)$

the network of Figure 2.2-3d may be reduced to the network of Figure 2.2-3c. Furthermore, Figures 2.2-3c and 2.2-3b are seen to be exactly Figures 2.2-2 and 2.2-1 respectively. Finally, if the joint arrival process consisting of the superposition of the two independent external packet streams is found, Figure 2.2-3b may be transformed into Figure 2.2-3a where $\lambda = \lambda_1 + \lambda_2$. Thus, we have proven that, under the aforementioned conditions, the five networks of Figure 2.2-3 yield equivalent overall steady-state mean waiting time on an arbitrary class packet.

We shall now illustrate some of our results by a specific example.

2.3 Example: Exponentially Distributed Interarrival Times

As an example, we consider the case where the interarrival time, for each of the two external input streams, is exponentially distributed. We immediately note that the distribution of the combined arrival process is also exponential with mean $1/(\lambda_1 + \lambda_2)$ (sec/packet). Thus, Figure 2.2-1 can be considered an M/D/1 queue with input parameter $\lambda_1 + \lambda_2$ (packets/sec). Furthermore, nodes N_1 and N_2 of Figure 2.2-2 are also M/D/1 queues with parameters λ_1 and λ_2 respectively. Consequently, we will make use of the well known results for an M/D/1 queue [12, Ch. 5], [18], [16, pp. 32-38].

Using (2.2-8), we find that the probability distribution of the number of packets serviced at node N_3 of Figure 2.2-2 during a busy period is

$$P\{M_j^{(3)}=m\} = P\{M_j=m\} = (m!)^{-1} [(\lambda_1 + \lambda_2)m\alpha]^{m-1} \exp[-(\lambda_1 + \lambda_2)m\alpha] \quad (2.3-1)$$

Furthermore, from (2.2-10), the distribution of the duration of an idle period is

$$P\{I_j^{(3)} \leq x\} = P\{I_j \leq x\} = 1 - \exp[-(\lambda_1 + \lambda_2)x] \quad (2.3-2)$$

In addition, using (2.2-23) and (2.2-24) we find that

$$(\bar{S})_{\text{Fig. 2.2-2}} = (\bar{S})_{\text{Fig. 2-2-1}} = \frac{(\rho_1 + \rho_2)\alpha}{2(1 - \rho_1 - \rho_2)} \quad (2.3-3)$$

and

$$\begin{aligned} \bar{W}^{(3)} = & \frac{(\rho_1 + \rho_2)\alpha}{2(1 - \rho_1 - \rho_2)} - \frac{\rho_1\alpha}{2(1 - \rho_1)} (\lambda_1 / (\lambda_1 + \lambda_2)) \\ & - \frac{\rho_2\alpha}{2(1 - \rho_2)} (\lambda_2 / (\lambda_1 + \lambda_2)) \end{aligned} \quad (2.3-4)$$

where $\rho_i = \lambda_i\alpha$.

Before leaving this example, let us look at the waiting time experienced by the various packet streams with respect to self delay and interference delay. In the network of Figure 2.2-1 we find that the self delay on a specific class packet is given by

$$\bar{S}_{(1) \text{ (self) Fig. 2.2-1}} = \frac{\rho_1\alpha}{2(1 - \rho_1)} \quad (2.3-5a)$$

and

$$\bar{S}_{(2) \text{ (self) Fig. 2.2-1}} = \frac{\rho_2\alpha}{2(1 - \rho_2)} \quad (2.3-5b)$$

where we have calculated the self delay on a class i packet by assuming the class j , $j \neq i$, arrival stream to be zero. From (2.3-5), we can now proceed to calculate the mean interference delay on an arbitrary class packet. Doing this we find

$$\begin{aligned} \bar{S}_{\text{ (inter) Fig. 2.2-1}} &= \bar{S}_{\text{ Fig. 2.2-1}} - \bar{S}_{(1) \text{ (self) Fig. 2.2-1}} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) \\ &\quad - \bar{S}_{(2) \text{ (self) Fig. 2.2-1}} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \quad (2.3-6) \\ &= \frac{(\rho_1 + \rho_2)\alpha}{2(1 - \rho_1 - \rho_2)} - \frac{\rho_1\alpha}{2(1 - \rho_1)} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) - \frac{\rho_2\alpha}{2(1 - \rho_2)} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \end{aligned}$$

On comparison with (2.3-4), we find that $\bar{S}_{\text{ (inter) Fig. 2.2-1}}$ is exactly equal to $\bar{W}^{(3)}$. Turning to Figure 2.2-2 we find

that the total self delay experienced by a class 1 packet would occur at node N_1 . Similarly the self delay on a class 2 packet occurs solely at node N_2 . Thus, we conclude that

$$\bar{S}_{(1)(\text{self})\text{Fig.2.2-2}} = \bar{W}_{(1)}^{(1)} = \frac{\rho_1 \alpha}{2(1-\rho_1)} \quad (2.3-7a)$$

and

$$\bar{S}_{(2)(\text{self})\text{Fig.2.2-2}} = \bar{W}_{(2)}^{(2)} = \frac{\rho_2 \alpha}{2(1-\rho_1)} \quad (2.3-7b)$$

and with respect to interference delay on an arbitrary class we find

$$\begin{aligned} \bar{S}_{(\text{inter})\text{Fig.2.2-2}} &= \bar{W}^{(3)} \quad (2.3-8) \\ &= \frac{(\rho_1 + \rho_2) \alpha}{2(1-\rho_1-\rho_2)} - \frac{\rho_1 \alpha}{2(1-\rho_1)} \left[\frac{\lambda_1}{\lambda_1 + \lambda_2} \right] - \frac{\rho_2 \alpha}{2(1-\rho_2)} \left[\frac{\lambda_2}{\lambda_1 + \lambda_2} \right] \end{aligned}$$

Thus, on comparing the steady-state waiting times in the two networks, we see that the amount of self delay experienced by a class i , $i = 1, 2$, packet is equivalent in both networks. This is seen by comparing (2.3-5) and (2.3-7). Furthermore, the interference delay on an arbitrary class packet is also equivalent in both networks. As for the interference delay on a specific class packet, we have as yet been unable to calculate these quantities in either of the two networks. However, we believe that these quantities will not be equivalent for the two networks considered.

2.4 Generalization to Networks with an Arbitrary Number of External Inputs

In the previous sections of this chapter, we have been limiting our study of networks to those that contained at most two external input streams of packets. In this section we show that all the theorems and corollaries presented earlier can easily be shown to apply to networks with an arbitrary number of independent external input streams.

In Figure 2.4-1 we have a queueing network consisting of a single queue with a constant service time of α sec/packet which is being fed by k external independent packet streams, each of which follow some general probability distribution. In Figure 2.4-2 we have these same external streams being merged at node N_{k+1} ; however, in this case each of the packets has received some processing at an earlier node before being merged at node N_{k+1} . In this network, all queues are assumed to have a constant service time of α sec/packet.

Let us now turn to the proofs of Theorems 2.2-1 and 2.2-2 and note that nowhere have we used the fact that we were dealing with networks having merely two external inputs. Thus, it is clear that the theorems and corollaries presented previously can be generalized to networks with an arbitrary number of external inputs. Therefore, one can state that:

1. The number of packets serviced at node N of Figure 2.4-1 during its j^{th} busy period equals the number

of packets serviced at node N_{k+1} of Figure 2.4-2 during its j^{th} busy period.

2. For the j^{th} busy period of nodes N and N_{k+1} , the total waiting time (queueing time) experienced by the M_j packets in the network of Figure 2.4-1 equals the total waiting experienced by the $M_j^{(k+1)}$ packets of the entire network of Figure 2.4-2.
3. For the networks of Figures 2.4-1 and 2.4-2, the steady-state overall mean waiting time experienced by an arbitrary class packet is the same in both networks; that is,

$$(\bar{S})_{\text{Fig.2.4-1}} = (\bar{S})_{\text{Fig.2.4-2}} \quad (2.4-1)$$

We could at this point draw networks similar to those of Figure 2.2-3 and then proceed to make claims, with regard to overall network steady-state waiting times, similar to those presented above while discussing Figure 2.2-3. Instead, we shall illustrate these claims by analyzing the network shown in Figure 2.4-3.

In Figure 2.4-3 we see a network consisting of four single server queues being fed by four independent external inputs each of which enter the network at a different point. As before, each external input is assumed to follow a general probability distribution with a mean arrival rate of λ_i packets/sec., $i = 1, 2, 3, 4$. Furthermore, the constant service time associated with node N_k is given by α_k where we have assumed $\alpha_4 \geq \alpha_3$ and $\alpha_4 \geq \alpha_2 \geq \alpha_1$ (thus assuring that

N_4 is the slowest queue). Finally, as was done previously, we assume each queue to have infinite storage capabilities and to operate on a first-come first-served basis.

The first result we mention is that the overall mean steady-state waiting time on an arbitrary packet would be determined solely by the value of α_4 and not the other service times present in the network; that is,

$$\bar{S} = f(\alpha_4) \quad (2.4-2a)$$

As an example, if all the external inputs were to be Poisson in nature we would obtain

$$\bar{S} = \frac{\rho\alpha_4}{2(1-\rho)} \quad (2.4-2b)$$

where $\rho = \sum_1 \lambda_i' \alpha_4 < 1$.

As a second result we could then proceed to calculate the overall mean steady-state waiting experienced by each of the individual packet streams and find:

$$\bar{S}_{(1)} = \bar{W}_{(1)}^{(1)} + \bar{W}_{(1)}^{(2)} + \bar{W}_{(1)}^{(4)} \quad (2.4-3a)$$

$$\bar{S}_{(2)} = \bar{W}_{(2)}^{(2)} + \bar{W}_{(2)}^{(4)} \quad (2.4-3b)$$

$$\bar{S}_{(3)} = \bar{W}_{(3)}^{(3)} + \bar{W}_{(3)}^{(4)} \quad (2.4-3c)$$

$$\bar{S}_{(4)} = \bar{W}_{(4)}^{(4)} \quad (2.4-3d)$$

where the $\bar{W}_{(j)}^{(k)}$ depend on the nature of the external inputs.

Once again as a specific example, we choose to let the external inputs be Poisson in nature and find

$$\bar{w}_{(1)}^{(1)} = \frac{\rho_{11}\alpha_1}{2(1-\rho_{11})} \quad (2.4-4a)$$

$$\begin{aligned} \bar{w}_{(1)}^{(2)} \left(\frac{\lambda_1}{\lambda_1+\lambda_2} \right) + \bar{w}_{(2)}^{(2)} \left(\frac{\lambda_2}{\lambda_1+\lambda_2} \right) &= \\ &= \frac{(\rho_{12}+\rho_{22})\alpha_2}{2(1-\rho_{12}-\rho_{22})} - \frac{\rho_{11}\alpha_1}{2(1-\rho_{11})} \frac{\lambda_1}{\lambda_1+\lambda_2} \end{aligned} \quad (2.4-4b)$$

$$\bar{w}_{(3)}^{(3)} = \frac{\rho_{33}\alpha_3}{2(1-\rho_{33})} \quad (2.4-4c)$$

$$\begin{aligned} \bar{w}_{(4)}^{(4)} &= \sum_{j=1}^4 w_{(j)}^{(4)} \left(\frac{\lambda_j}{\lambda_1+\lambda_2+\lambda_3+\lambda_4} \right) \quad (2.4-4d) \\ &= \bar{s} - \bar{w}_{(1)}^{(1)} - \bar{w}_{(2)}^{(2)} - \bar{w}_{(3)}^{(3)} \\ &= \frac{\rho\alpha_4}{2(1-\rho)} - \frac{(\rho_{12}+\rho_{22})\alpha_2}{2(1-\rho_{12}-\rho_{22})} \left(\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3+\lambda_4} \right) \\ &\quad - \frac{\rho_{33}\alpha_3}{2(1-\rho_{33})} \left(\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3+\lambda_4} \right) \end{aligned}$$

where $\bar{w}_{(1)}^{(1)}$ is found by considering the subnetwork consisting of the single node N_1 , $\bar{w}_{(1)}^{(2)} \left(\frac{\lambda_1}{\lambda_1+\lambda_2} \right) + \bar{w}_{(2)}^{(2)} \left(\frac{\lambda_2}{\lambda_1+\lambda_2} \right)$ is determined by the subnetwork of nodes N_1 and N_2 , $\bar{w}_{(3)}^{(3)}$ by the subnetwork consisting of node N_3 , and $\bar{w}_{(4)}^{(4)}$ by considering the total network of all four nodes.

2.5 Networks with Fast Merger Nodes

Let us look at the network of Figure 2.5-1. Here, we see a network consisting of three nodes and two independent external inputs. In the previous sections, we have limited our discussion of the network to the case where the service time of the merger node was the slowest; that is,

$\alpha_3 = \max(\alpha_1, \alpha_2, \alpha_3)$. We saw that when this condition was true, node N_3 became the dominant node of the network with respect to the overall steady-state waiting time on an arbitrary packet. In this section, we shall study the network of Figure 2.5-1 for cases when the service time of the merger node is not the slowest in the system. To do this, we shall divide our discussion into two cases, the first being where α_3 is greater than the parallel combination of α_1 and α_2 and the second where α_3 is less than or equal to the parallel combination of α_1 and α_2 ; that is, $\alpha_3 > \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$ and $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$ respectively. The reason for this division is that for the case $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, it is easily shown that the queue length at node N_3 must remain finite for all values of λ_1 and λ_2 . Furthermore, it is easily proven that the number of packets on queue at node N_3 can never exceed a total of one.

Case 1: $\alpha_3 > \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, i.e. $\alpha_3 > (\alpha_1 // \alpha_2)$

For the case $\alpha_3 > \alpha_1 // \alpha_2$, we have, as yet, been unable to obtain any exact results. However, we have been

able to derive a tight upper bound to the mean steady-state waiting time at node N_3 for an arbitrary class packet. We shall derive the upper bound and then proceed to compare it to results obtained using approximation methods of Kleinrock [11,13] and Rubin [20]. We shall derive the upper bound by considering three separate subcases.

Subcase 1.1. $\alpha_1 \geq \alpha_3 \geq \alpha_2$

Let us consider the subcase in the specific situation where the external inputs are Poisson in nature. We note that, were we to consider the case $\alpha_3 = \max(\alpha_1, \alpha_2, \alpha_3)$, the mean steady-state waiting time on an arbitrary class packet at node N_3 could be found exactly using the methods presented earlier (Sections 2.2-2.3) yielding

$$\bar{w}^{(3)} = \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} - \frac{\rho_{11}\alpha_1}{2(1 - \rho_{11})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) - \frac{\rho_{22}\alpha_2}{2(1 - \rho_{22})} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \quad (2.5-1)$$

Furthermore,

$$(\bar{w}_{(1)}^{(3)})_{\text{self}} = \frac{\rho_{13}\alpha_3}{2(1 - \rho_{13})} - \frac{\rho_{11}\alpha_1}{2(1 - \rho_{11})} \quad (2.5-2)$$

Similarly,

$$(\bar{w}_{(2)}^{(3)})_{\text{self}} = \frac{\rho_{23}\alpha_3}{2(1-\rho_{23})} - \frac{\rho_{22}\alpha_2}{2(1-\rho_{22})} \quad (2.5-3)$$

Thus, the mean interference delay on an arbitrary packet at node N_3 is found to be

$$\begin{aligned} \bar{w}_{\text{inter}}^{(3)} &= \bar{w}^{(3)} - \left[\bar{w}_{(1)}^{(3)} \right]_{\text{self}} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) - \left[\bar{w}_{(2)}^{(3)} \right]_{\text{self}} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \\ &= \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1-\rho_{13}-\rho_{23})} - \frac{\rho_{13}\alpha_3}{2(1-\rho_{13})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) \\ &\quad - \frac{\rho_{23}\alpha_3}{2(1-\rho_{23})} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \end{aligned} \quad (2.5-4)$$

We make note of two important points. First, we see from (2.5-4), that the interference delay at node N_3 on an arbitrary class packet is independent of α_1 and α_2 . Second, and of more importance, we see from (2.5-1), that as α_1 approaches α_3 , the steady-state mean waiting time, $\bar{w}^{(3)}$, is reduced in value. A similar statement could be made as α_2 approaches α_3 .

Let us now return to the present case of interest; that is, $\alpha_1 \geq \alpha_3 \geq \alpha_2$. It should be clear from the discussion of the previous paragraph that a tight upper bound to the mean steady-state waiting time at node N_3 can be obtained from (2.5-1) by substituting α_3 for α_1 . Thus, we obtain

$$\bar{w}(3) \leq \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} - \frac{\rho_{13}\alpha_3}{2(1 - \rho_{13})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) - \frac{\rho_{22}\alpha_2}{2(1 - \rho_{22})} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \quad (2.5-5)$$

Subcase 1.2. $\alpha_2 \geq \alpha_3 \geq \alpha_1$

It is clear that this subcase is symmetric to subcase 1.1; thus, we obtain

$$\bar{w}(3) \leq \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} - \frac{\rho_{11}\alpha_3}{2(1 - \rho_{11})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) - \frac{\rho_{23}\alpha_3}{2(1 - \rho_{23})} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \quad (2.5-6)$$

Subcase 1.3. $\alpha_3 \leq \alpha_1$ and $\alpha_3 \leq \alpha_2$

For this subcase, we substitute α_3 for both α_1 and α_2 in (2.5-1) yielding

$$\bar{w}(3) \leq \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} - \frac{\rho_{13}\alpha_3}{2(1 - \rho_{13})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) - \frac{\rho_{23}\alpha_3}{2(1 - \rho_{23})} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \quad (2.5-7)$$

In summary, we see that an upper bound to the waiting time at node N_3 may be obtained by replacing each $\alpha_i \leq \alpha_3$ by α_3 and proceeding with the techniques developed previously. Furthermore, though we have proven this fact only in the case of Poisson external inputs, it should be clear that this would be true for any general external input. Thus, we state the following theorem.

THEOREM 2.5-1:

For external inputs following any general probability distribution, $\alpha_3 > \frac{\alpha_1\alpha_2}{\alpha_1 + \alpha_2}$ and $\alpha_3 < \alpha_i$ for at least one value of i , $i = 1, 2$, the mean steady-state waiting time can

an arbitrary class packet at node N_3 of Figure 2.5-1 can be upper bounded by considering a network with $\alpha_i = \alpha_3$ for each $\alpha_i > \alpha_3$.

We will now compare results obtained using our approximate method to those obtained using the independence assumption and to the approximate results one would obtain using the methods of Rubin [20]. In this comparison, we shall assume that the external inputs are Poisson in nature.

If one had applied the independence assumption at node N_3 , thus assuming that the arrival streams into N_3 are Poisson, one would obtain

$$\bar{w}_P^{(3)} = \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} \quad (2.5-8)$$

where $\bar{w}_P^{(3)}$ is the mean steady-state waiting time on an arbitrary class packet at node N_3 found by using the independence assumption. On comparison, one notes that results obtained using the independence assumption, (2.5-8), are always greater than those obtained using our approximation method, (2.5-5), (2.5-6), (2.5-7). Thus, since we have proven our method to be an upper bound to the actual value of the waiting time at N_3 , it is clear that our approximation method yields a much tighter upper bound to that of the independence assumption.

On the other hand, using the approximation method of Rubin [20], one would proceed by writing

$$\bar{w}_R^{(3)} = \bar{w}_R^{(1)} + \bar{w}_R^{(2)} \quad (2.5-9)$$

where $\bar{W}_{R(1)}^{(3)}$ is an approximation to the part of the waiting time at N_3 on an arbitrary class packet determined by the stochastic character of the class 1 stream and deterministic like interference by the class 2 stream. Similarly, $\bar{W}_{R(2)}^{(3)}$ can be regarded as that part of the waiting time on an arbitrary packet caused by the stochastic nature of the class 2 stream and a deterministic like interference by the class 1 stream. Subsequently, the sum of these two parts, $\bar{W}_{R(1)}^{(3)}$ and $\bar{W}_{R(2)}^{(3)}$, will constitute an approximation to the waiting time on an arbitrary class packet at node N_3 , which we denote by $\bar{W}_R^{(3)}$. Furthermore, quoting from Rubin's [20] paper, one would obtain,

$$\bar{W}_{R(1)}^{(3)} = \begin{cases} \frac{\rho_{13}\alpha_3}{2(1-\rho_{13}-\rho_{23})} - \frac{\rho_{11}\alpha_1(1-\rho_{23})}{2(1-\rho_{11})} & \text{if } \alpha_1 < \frac{\alpha_3}{1-\rho_{23}} \\ 0 & \text{if } \alpha_1 \geq \frac{\alpha_3}{1-\rho_{23}} \end{cases} \quad (2.5-10)$$

and

$$\bar{W}_{R(2)}^{(3)} = \begin{cases} \frac{\rho_{23}\alpha_3}{2(1-\rho_{13}-\rho_{23})} - \frac{\rho_{22}\alpha_2(1-\rho_{13})}{2(1-\rho_{22})} & \text{if } \alpha_2 < \frac{\alpha_3}{1-\rho_{13}} \\ 0 & \text{if } \alpha_2 \geq \frac{\alpha_3}{1-\rho_{13}} \end{cases} \quad (2.5-11)$$

To compare results obtained using our approximation method to those obtained using Rubin's method, let us look at one of the limiting points for the case under consideration. We shall choose the case of $\alpha_3 = \alpha_1 = \alpha_2$, a point

at which our method has been proven to yield an exact result. At this point, using Rubin's method, one would obtain from (2.5-9), (2.5-10), and (2.5-11) that

$$\bar{W}_R^{(3)} = \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} - \frac{\rho_{11}\alpha_1(1 - \rho_{23})}{2(1 - \rho_{11})} - \frac{\rho_{22}\alpha_2(1 - \rho_{13})}{2(1 - \rho_{22})} \quad (2.5-12)$$

Using our method, one would obtain an exact result given by (2.5-1). To compare the two results, we can rewrite (2.5-1) to yield

$$\bar{W}^{(3)} = \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} - \frac{\rho_{11}\alpha_1}{2(1 - \rho_{11})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right) - \frac{\rho_{22}\alpha_2}{2(1 - \rho_{22})} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right) \quad (2.5-13a)$$

$$= \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} - \frac{\rho_{11}\alpha_1}{2(1 - \rho_{11})} \left(1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \right) - \frac{\rho_{22}\alpha_2}{2(1 - \rho_{22})} \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) \quad (2.5-13b)$$

$$= \frac{(\rho_{13} + \rho_{23})\alpha_3}{2(1 - \rho_{13} - \rho_{23})} - \frac{\rho_{11}\alpha_1}{2(1 - \rho_{11})} \left(1 - \frac{\rho_{23}}{\rho_{13} + \rho_{23}} \right) - \frac{\rho_{22}\alpha_2}{2(1 - \rho_{22})} \left(1 - \frac{\rho_{13}}{\rho_{13} + \rho_{23}} \right) \quad (2.5-13c)$$

On comparing (2.5-12) to (2.5-13c), one notes that since

$(\rho_{13} + \rho_{23}) < 1$ (from stability considerations) we must have $\bar{W}_{(R)}^{(3)} < \bar{W}^{(3)}$. * Furthermore, since we have proven (2.5-13)

* We note that $\bar{W}_R^{(3)} = \bar{W}^{(3)}$ if and only if $\lambda_1 = 0$, or $\lambda_2 = 0$, or $\rho_{13} + \rho_{23} = 1$, all of which are uninteresting cases.

to be exact for $\alpha_3 = \alpha_1 = \alpha_2$, we see that (2.5-12) constitutes a lower bound to the waiting time at node N_3 .

In a similar manner, one could now proceed and show that for any point in the region $\alpha_3 > \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_3}$, Rubin's method will always yield a lower value of waiting time at N_3 than would be obtained using Theorem 2.5-1. Finally, though proven only for cases where the method of Theorem 2.5-1 yields exact results, one can state that, whereas Theorem 2.5-1 has been proven to yield an upper bound approximation to the waiting time at node N_3 , Rubin's method will yield a lower bound approximation.

Having established the relationship among the approximation methods of Theorem 2.5-1, the independence assumption, and that of Rubin, we shall proceed to compare results obtained from these various methods to results obtained via a computer simulation of the network of Figure 2.5-1. In each simulation run, we assumed that the two external inputs are Poisson in nature with a fixed mean packet arrival rate of λ_1 and λ_2 respectively, and that the service time at nodes N_1 and N_2 were fixed at α_1 and α_2 respectively. We then had the simulation program obtain a plot of the mean steady-state waiting time on an arbitrary packet at node N_3 as a function of the service time at node N_3 ; namely, α_3 . The results of simulation runs for various values of λ_1 , λ_2 , α_1 , and α_2 are shown in the graphs of Figures 2.5-2 through 2.5-10. On each of these graphs we have plotted the simulation results along with the results that would be

obtained by each of the three approximation methods mentioned previously. From these graphs, it is clear that results obtained using the independence assumption are a loose upper bound to actual values. Furthermore, Rubin's method seems to yield a rather loose lower bound for the case considered. Finally, results obtained using the method introduced by Theorem 2.5-1 tend to be a tight upper bound approximation to actual values. We mention that all of these results were expected as described above.

Finally, before concluding the case under consideration, we mention that, once again, although we have proven Theorem 2.5-1 for networks with only two external inputs, it may easily be extended to networks with an arbitrary number of external inputs; such as, the network of Figure 2.5-11. In doing so, Theorem 2.5-1 would be rewritten to state that for external inputs following any general probability distribution, $\alpha_{k+1} > (\alpha_1 // \alpha_2 // \dots // \alpha_k)$ and $\alpha_{k+1} < \alpha_i$ for at least one value of i , $i = 1, 2, \dots, k$, the mean steady-state waiting time on an arbitrary class packet at node N_{k+1} of Figure 2.5-11 can be upper bounded by considering a network with $\alpha_i = \alpha_{k+1}$ for each $\alpha_i > \alpha_{k+1}$.

Case 2. $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, i.e. $\alpha_3 \leq (\alpha_1 // \alpha_2)$

We now return to the network of Figure 2.5-1 and analyze the situation when $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$. As mentioned earlier, the case of $\alpha_3 \leq (\alpha_1 // \alpha_2)$ has the interesting

characteristic that the total number of packets on queue at node N_3 may never exceed one; thus, the absolute total waiting time on any packet may never exceed α_3 sec. We shall now state this result as a theorem and proceed to prove it.

Theorem 2.5-2:

For the case $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, the total number of packets on queue at node N_3 may never exceed one.

Proof: We shall prove our theorem by an analysis of the worst possible case which is seen to occur when one sets $\alpha_3 = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$. That this is the worst case is clear because $\alpha_3 = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$ will allow for the shortest possible interarrival time between packets of a single class at node N_3 , thus, causing the greatest possibility of the queue length growing in size.

When $\alpha_3 = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, we may write that

$$\alpha_1 = k\alpha_3, \quad 1 < k \leq 2 \quad (2.5-14a)$$

This implies that

$$\alpha_2 = \frac{k\alpha_3}{k-1} \quad (2.5-14b)$$

(We note that we could have selected $k \geq 2$; however, this would imply $1 < \frac{k}{k-1} \leq 2$.)

We now proceed to construct an arrival pattern at node N_3 where we shall assume that the arrival pattern was begun

by the arrival of both a class 1 and a class 2 packet at approximately the exact same time. Furthermore, for the worst case analysis, we shall assume that class 1 packets are arriving regularly at intervals of α_1 secs. and class 2 packets are doing similarly at intervals of α_2 secs. Thus, with

$t_{ij}^{(m)} \triangleq$ the arrival time at node m of the j^{th} class i packet,

we may write

$$\begin{aligned} t_{11}^{(3)} &= 0 \\ t_{12}^{(3)} &= k\alpha_3 \\ &\vdots \\ t_{1j}^{(3)} &= (j-1)k\alpha_3 \end{aligned} \tag{2.5-15a}$$

and

$$\begin{aligned} t_{21}^{(3)} &= 0 \\ t_{22}^{(3)} &= \frac{k\alpha_3}{k-1} \\ t_{23}^{(3)} &= 2 \frac{k\alpha_3}{k-1} \\ &\vdots \\ t_{2n}^{(3)} &= (n-1) \frac{k\alpha_3}{k-1} \end{aligned} \tag{2.5-15b}$$

where we have assumed that the arrivals at node N_3 commence at time $t = 0$.

Using the limits of k , we may rewrite (2.5-15) and obtain

$$\begin{aligned}
 t_{11}^{(3)} &= 0 \\
 \alpha_3 &< t_{12}^{(3)} \leq 2\alpha_3 \\
 2\alpha_3 &< t_{13}^{(3)} \leq 4\alpha_3 \\
 &\vdots \\
 (j-1)\alpha_3 &< t_{1j}^{(3)} \leq 2(j-1)\alpha_3
 \end{aligned}
 \tag{2.5-16a}$$

and

$$\begin{aligned}
 t_{21}^{(3)} &= 0 \\
 2\alpha_3 &\leq t_{22}^{(3)} < \infty \\
 4\alpha_3 &\leq t_{23}^{(3)} < \infty \\
 &\vdots \\
 2(n-1)\alpha_3 &\leq t_{2n}^{(3)} < \infty
 \end{aligned}
 \tag{2.5-16b}$$

Using the notation of Section 2.2 one may now proceed to use (2.5-15) and (2.5-16) and find (see Appendix A) the arrival time at N_3 of the i^{th} arbitrary class packet which is given by

$$\begin{aligned}
 T_1' &= 0 \\
 T_2' &= 0 \\
 \alpha_3 &< T_3' \leq 2\alpha_3 \\
 2\alpha_3 &\leq T_4' \leq 3\alpha_3 \\
 3\alpha_3 &\leq T_5' \leq 4\alpha_3 \\
 &\vdots \\
 (n-2)\alpha_3 &\leq T_n' \leq (n-1)\alpha_3
 \end{aligned}
 \tag{2.5-17}$$

Furthermore, with respect to departure times from node N_3 , one may use (2.5-17) and calculate

$$\begin{aligned} D'_1 &= \alpha_3 \\ D'_2 &= 2\alpha_3 \\ D'_3 &= 3\alpha_3 \\ D'_4 &= 4\alpha_3 \\ D'_5 &= 5\alpha_3 \\ &\vdots \\ D'_n &= n\alpha_3 \end{aligned} \tag{2.5-18}$$

Thus, from (2.5-17) and (2.5-18) we can conclude that

$$T'_{n+2} \geq D'_n \tag{2.5-19}$$

which proves our theorem. Q.E.D.

Having proven Theorem 2.5-2, we can immediately state the following corollary.

Corollary 2.5-1

For the case $\alpha_3 \leq \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}$, the waiting time on any packet at node N_3 of Figure 2.5-1 may never exceed α_3 sec.; that is,

$$W^{(3)} \leq \alpha_3 . \tag{2.5-20}$$

Thus (2.5-20) may serve as an upper bound on any results we obtain in the region $\alpha_3 \leq \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}$.

We now make the observation that if $\alpha_3 \leq \frac{1}{2} \min(\alpha_1, \alpha_2)$, then the maximum length of a busy period at node N_3 is clearly two. This is true since it is clear that node N_3 will be simultaneously occupied by at most one class 1 and one class 2 packet. Thus, we shall subdivide our discussion of case 2 into two subcases.

Subcase 2.1: $\alpha_3 \leq \frac{1}{2} \min(\alpha_1, \alpha_2)$

For the subcase $\alpha_3 \leq \frac{1}{2} \min(\alpha_1, \alpha_2)$, we have been successful in obtaining exact results for the mean steady-state waiting time on a packet at N_3 .

Theorem 2.5-3:

For the subcase $\alpha_3 \leq \frac{1}{2} \min(\alpha_1, \alpha_2)$, the mean steady-state waiting time on an arbitrary class packet at node N_3 of Figure 2.5-1 is given by

$$\bar{w}(3) = \frac{\rho_{13}\rho_{23}}{\lambda_1 + \lambda_2} \quad (2.5-21)$$

and is valid for external inputs following any general probability distribution.

Proof: Since the distributions of the class 1 and class 2 external inputs are assumed independent from each other, class 1 packets will arrive at node N_3 in an independent manner from the arrival pattern of the class 2 packets and vice versa. Furthermore, since, as we have stated above, a busy period at N_3 can never exceed a length of two, we may write that

$$\bar{W}_{(1)}^{(3)} = E\left\{W_{(1)}^{(3)} \mid \text{a class 1 arrival finds } N_3 \text{ busy with a class 2 packet}\right\} \\ \cdot P\left\{\text{a class 1 arrival finds } N_3 \text{ busy with a class 2 packet}\right\} \quad (2.5-22a)$$

and

$$\bar{W}_{(2)}^{(3)} = E\left\{W_{(2)}^{(3)} \mid \text{a class 2 arrival finds } N_3 \text{ busy with a class 1 packet}\right\} \\ \cdot P\left\{\text{a class 2 arrival finds } N_3 \text{ busy with a class 1 packet}\right\} \quad (2.5-22b)$$

Furthermore, since the class 1 and class 2 packets arrive at N_3 independently of each other, it is clear that a packet that arrives at N_3 and finds it busy has a waiting time that is uniformly distributed between zero and α_3 .

Thus,

$$E\left\{W_{(i)}^{(3)} \mid \text{a class } i \text{ arrival finds } N_3 \text{ busy}\right\} = \frac{\alpha_3}{2} \quad (2.5-23)$$

Also, from general queueing theory and the fact that N_3 is occupied with no more than a single packet from a given class we have

$$P\left\{\text{a class } i \text{ arrival finds } N_3 \text{ busy with a class } j \text{ packet}\right\} \\ = \rho_{j3}, \quad i=1,2; \quad j=1,2; \quad i \neq j \quad (2.5-24)$$

Consequently, if we substitute (2.5-23) and (2.5-24) into (2.5-22), we find

$$\bar{w}_{(1)}^{(3)} = \frac{\rho_{23}^{\alpha_3}}{2} \quad (2.5-25a)$$

and

$$\bar{w}_{(2)}^{(3)} = \frac{\rho_{13}^{\alpha_3}}{2} \quad (2.5-25b)$$

Finally, we can find $\bar{w}^{(3)}$ to be

$$\begin{aligned} \bar{w}^{(3)} &= \bar{w}_{(1)}^{(3)} P\left\{\begin{array}{l} \text{a packet is class 1} \\ \text{is class 2} \end{array}\right\} + \bar{w}_{(2)}^{(3)} P\left\{\begin{array}{l} \text{a packet} \\ \text{is class 2} \end{array}\right\} \\ &= \frac{\rho_{23}^{\alpha_3}}{2} \left[\frac{\lambda_1}{\lambda_1 + \lambda_2} \right] + \frac{\rho_{13}^{\alpha_3}}{2} \left[\frac{\lambda_2}{\lambda_1 + \lambda_2} \right] \\ &= \frac{\rho_{13} \rho_{23}}{\lambda_1 + \lambda_2} \end{aligned} \quad (2.5-26)$$

thus proving our theorem. Q.E.D.

We now introduce the following quantities.

$P_k^{(j)} \triangleq$ the probability that node j contains k packets (both on queue and in the server) , $k=0,1,\dots$

$\bar{N}^{(j)} \triangleq$ the mean steady-state number of arbitrary class packets at node j ; that is, $\bar{N}^{(j)} = \sum_{k=0}^{\infty} k P_k^{(j)}$.

$\bar{N}_q^{(j)} \triangleq$ the mean steady-state number of arbitrary class packets waiting on queue at node j ; that is, $\bar{N}_q^{(j)} = \sum_{k=1}^{\infty} k P_{k+1}^{(j)} = \bar{N}^{(j)} - \rho_j$, where ρ_j is the utilization factor of node j .

$\bar{N}_{q(i)}^{(j)} \triangleq$ the mean steady-state number of class i packets waiting on queue at node j .

$\bar{T}^{(j)} \triangleq$ the total mean steady-state system time (queueing plus service) on an arbitrary class packet at node j ;

that is, $\bar{T}^{(j)} = \bar{W}^{(j)} + \alpha_j$.

From (2.5-21) we see that

$$\bar{T}^{(3)} = \frac{\rho_{13}\rho_{23}}{\lambda_1 + \lambda_2} + \alpha_3 \quad (2.5-27)$$

If we now apply Little's formula [14] to (2.5-21)

and (2.5-27) we find

$$\bar{N}_q^{(3)} = (\lambda_1 + \lambda_2) \bar{W}^{(3)} = \rho_{13}\rho_{23} \quad (2.5-28)$$

and

$$\bar{N}^{(3)} = (\lambda_1 + \lambda_2) \bar{T}^{(3)} = \rho_{13}\rho_{23} + (\rho_{13} + \rho_{23}) \quad (2.5-29)$$

Furthermore, from Theorem 2.5-2, it is clear that at any random point in time, node N_3 has either zero, one, or two packets in its facility, thus, we need only find $P_0^{(3)}$, $P_1^{(3)}$ and $P_2^{(3)}$. To calculate these quantities we need only apply the definitions of $\bar{N}_q^{(3)}$, $\bar{N}^{(3)}$ and the fact that $\sum_{k=0}^{\infty} P_k = 1$. Doing this we find

$$P_2^{(3)} = \bar{N}_q^{(3)} = \rho_{13}\rho_{23} \quad (2.5-30a)$$

$$P_1^{(3)} = \bar{N}^{(3)} - 2P_2^{(3)} = \rho_{13} + \rho_{23} - \rho_{13}\rho_{23} \quad (2.5-30b)$$

$$P_0^{(3)} = 1 - P_1^{(3)} - P_2^{(3)} = 1 - \rho_{13} - \rho_{23} \quad (2.5-30c)$$

Thus, we have obtained the exact steady-state distribution of the number of packets at node N_3 .

As a final point of interest, using (2.5-25) and Little's formula, we find

$$\bar{N}_{q(1)}^{(3)} = \lambda_1 \bar{W}_{(1)}^{(3)} = \frac{\rho_{13} \rho_{23}}{2} \quad (2.5-31a)$$

and

$$\bar{N}_{q(2)}^{(3)} = \lambda_2 \bar{W}_{(2)}^{(3)} = \frac{\rho_{13} \rho_{23}}{2} = \bar{N}_{q(1)}^{(3)} \quad (2.5-31b)$$

This is a rather surprising, however, intuitive result.

At this point, it is of interest to see how the results just derived for a network of two external inputs can be applied to yield an exact result in a network which has three external input streams. Such a network is depicted by Figure 2.5-12.

In Figure 2.5-12 we have three independent external inputs whose packet streams are merged at node N_4 after each packet has received some processing at an earlier node. If we assume $\alpha_4 \leq \frac{1}{3} \min(\alpha_1, \alpha_2, \alpha_3)$, it is clear and easily proven, as was done above for two external inputs, that node N_4 will at no time be occupied by more than a single packet from each input class, thus, limiting the maximum number of packets that may be present at N_4 to three. Furthermore, since a given class of packets arrive at N_4 independently of the arrival of packets from the other two packet streams, it is clear that a packet arriving at N_4 sees a network consisting of the other two packet streams whose state probabilities at node N_4 are of the form given by (2.5-30).

In addition, since a packet from a given class arrives at a random point in time with respect to the other two classes, a packet arriving to find N_4 occupied by a single packet will have a waiting time uniformly distributed between zero and α_4 sec., while a packet arriving and finding N_4 occupied with two packets will have its waiting time uniformly distributed between α_4 and $2\alpha_4$ sec. Thus, using these facts one may write

$$\begin{aligned} \bar{W}_{(1)}^{(4)} = & E\left\{W_{(1)}^{(4)} \mid \text{a class 1 packet arrives at } N_4 \text{ and} \right. & (2.5-32) \\ & \left. \text{finds } N_4 \text{ busy with one packet} \right\} \\ & \cdot P\left\{\text{a class 1 packet arrives at } N_4 \text{ and finds} \right. \\ & \left. N_4 \text{ busy with one packet} \right\} \\ & + E\left\{W_{(1)}^{(4)} \mid \text{a class 1 packet arrives at } N_4 \text{ and} \right. \\ & \left. \text{finds } N_4 \text{ busy with two packets} \right\} \\ & \cdot P\left\{\text{a class 1 packet arrives at } N_4 \text{ and finds} \right. \\ & \left. N_4 \text{ busy with two packets} \right\} . \end{aligned}$$

Then, using the results of (2.5-30) and the fact of uniform distributions on the waiting time of an arriving packet, we find

$$\begin{aligned} \bar{W}_{(1)}^{(4)} = & \left(\frac{\alpha_4}{2}\right) (\rho_{24} + \rho_{34} - \rho_{24}\rho_{34}) + \left(\frac{3\alpha_4}{2}\right) (\rho_{24}\rho_{34}) \\ & = (\rho_{24} + \rho_{34}) \left(\frac{\alpha_4}{2}\right) + \rho_{24}\rho_{34}\alpha_4 . \end{aligned} \tag{2.5-33a}$$

In a similar manner one calculates that

$$\begin{aligned}\bar{w}_{(2)}^{(4)} &= \left(\frac{\alpha_4}{2}\right) (\rho_{14} + \rho_{34} - \rho_{14}\rho_{34}) + \left(\frac{3\alpha_4}{2}\right) (\rho_{14}\rho_{34}) & (2.5-33b) \\ &= (\rho_{14} + \rho_{34}) \left(\frac{\alpha_4}{2}\right) + \rho_{14}\rho_{34}\alpha_4\end{aligned}$$

$$\begin{aligned}\bar{w}_{(3)}^{(4)} &= \left(\frac{\alpha_4}{2}\right) (\rho_{14} + \rho_{24} - \rho_{14}\rho_{24}) - \left(\frac{3\alpha_4}{2}\right) (\rho_{14}\rho_{24}) & (2.5-33c) \\ &= (\rho_{14} + \rho_{24}) \left(\frac{\alpha_4}{2}\right) + \rho_{14}\rho_{24}\alpha_4.\end{aligned}$$

Then, using (2.5-33), one can determine the mean steady-state waiting time on an arbitrary class packet at node N_4 and find

$$\begin{aligned}\bar{w}^{(4)} &= \sum_{i=1}^3 \bar{w}_{(i)}^{(4)} P\{\text{a packet is class } i\} & (2.5-34) \\ &= (\rho_{14}\rho_{24} + \rho_{14}\rho_{34} + \rho_{24}\rho_{34} + 3\rho_{14}\rho_{24}\rho_{34}) \left(\frac{1}{\lambda_1 + \lambda_2 + \lambda_3}\right).\end{aligned}$$

Finally, applying Little's theorem to (2.5-34) yields

$$\bar{N}_q^{(4)} = \rho_{14}\rho_{24} + \rho_{14}\rho_{34} + \rho_{24}\rho_{34} + 3\rho_{14}\rho_{24}\rho_{34} . \quad (2.5-35)$$

At this point, we mention that we as yet have been unable to mathematically extend the results just derived to networks with more than three external inputs. We do, however, have some conjectures as to what these extensions may be. These conjectures have been included in Appendix B for the interested reader.

Subcase 2.2: $\frac{1}{2} \min(\alpha_1, \alpha_2) \leq \alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$

We now return to Figure 2.5-1 and undertake the analysis of our final subcase; that is, when $\frac{1}{2} \min(\alpha_1, \alpha_2) \leq \alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$. The difficulty in analyzing this subcase occurs because, though Theorem 2.5-2 still applies, it is possible for node N_3 to be concurrently occupied by two packets from the same class. Thus, we must use a technique different from the one used for subcase 2.1.

As yet, we have been unable to obtain either exact results or even a good upper bound approximation method that is valid for the entire region under discussion. The reason for this will become apparent in the ensuing paragraphs. We have, however, been successful in deriving exact results for certain special cases that exist in the region of interest.

We begin by stating the following theorem.

Theorem 2.5-4

For the network of Figure 2.5-1, the mean steady-state waiting time on an arbitrary class packet at node N_3 , when $\frac{1}{2} \min(\alpha_1, \alpha_2) \leq \alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$; $\alpha_2 = k\alpha_1$, $k = 1, 2, 3, \dots$; $\lambda_1 = \frac{1}{\alpha_1}$; and $\lambda_2 \leq \frac{1}{\alpha_2}$, is given by

$$\bar{w}^{(3)} = \frac{\rho_{13} \rho_{23}}{\rho_{13} + \rho_{23}} \left(\sum_{j=1}^{\infty} [j\alpha_3 - (j-1)\alpha_1]^+ \right). \quad (2.5-36)$$

In addition, the mean steady-state waiting time at N_3 on a packet from a specific class is

$$\bar{w}_{(1)}^{(3)} = \frac{\rho_{23}\alpha_3}{2} + \rho_{23} \left\{ \sum_{j=2}^{\infty} [j\alpha_3 - (j-1)\alpha_1]^+ \right\} \quad (2.5-37a)$$

and

$$\bar{w}_{(2)}^{(3)} = \frac{\rho_{13}\alpha_3}{2} . \quad (2.5-37b)$$

Proof: Before beginning our derivation of Theorem 2.5-4, we note that our choice of $\alpha_2 = k\alpha_1$ instead of $\alpha_1 = k\alpha_2$, is of course totally arbitrary and that a similar theorem could be stated for $\alpha_1 = k\alpha_2$ by simply interchanging subscripts 1 and 2 each time they appear in the above theorem.

We now begin our derivation of Theorem 2.5-4 by analyzing the special case when $\alpha_2 = 2\alpha_1$, $\alpha_3 = \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}$, $\lambda_1 = \frac{1}{\alpha_1}$, and $\lambda_2 = \frac{1}{\alpha_2}$. We note that, for the values just chosen, we have $\rho_{11} = \rho_{22} = 1$ which would indeed cause an infinite queue length to occur in steady-state at nodes N_1 and N_2 . However, as proven above, the total number of packets at N_3 will never exceed two as long as $\alpha_3 \leq \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}$. Consequently, as far as node N_3 is concerned, in the steady-state, it sees a continuous stream of class 1 and class 2 packets arriving where the class 1 arrivals are uniformly spaced by α_1 sec. and the class 2 packets are uniformly spaced by α_2 sec..

Let us suppose that the class 1 stream arrives at node N_3 , in the steady-state, at times

$$t_0 + j\alpha_1, \quad j=0,1,2,\dots \quad (2.5-38)$$

The class 2 arrivals, on the other hand, are assumed to

arrive at times

$$t' + k\alpha_2, \quad k=0,1,2,\dots \quad (2.5-39)$$

It should be obvious that we lose no generality by limiting t' to be in the region

$$t_0 < t' < t_0 + \alpha_1 \quad (2.5-40)$$

Furthermore, by using the relations $\alpha_2 = 2\alpha_1$ and $\alpha_3 = \frac{\alpha_1\alpha_2}{\alpha_1 + \alpha_2}$, we would find

$$\alpha_1 = \frac{3}{2} \alpha_3 \quad (2.5-41a)$$

and

$$\alpha_2 = 3 \alpha_3 \quad (2.5-41b)$$

which leads directly to the conclusion that the busy period at node N_3 will be of length three, consisting of two class 1 packets and a single class 2 packet.

For $t_0 < t' \leq t_0 + \alpha_3$, it should be obvious that, a busy period at node N_3 will consist of the packets arriving at times t_0 , t' , and $t_0 + \alpha_1$. The waiting time on the class 1 packet arriving at time t_0 , W_{t_0} , is clearly zero. On the class 2 packet arriving at t' , the waiting time will be

$$W_{t'} = t_0 + \alpha_3 - t' \quad (2.5-42a)$$

Substituting the limits on t' , we find

$$0 \leq W_{t'} \leq \alpha_3 \quad (2.5-42b)$$

Furthermore, since the class 2 packets arrive at N_3

independently of the class 1 packets, $W_{t'}$, must be uniformly distributed, thus,

$$\bar{W}_{t'} = \frac{\alpha_3}{2} \quad (2.5-43)$$

Finally, for the class 1 packet arriving at $t_0 + \alpha_1$ the waiting time is

$$W_{t_0 + \alpha_1} = (t_0 + 2\alpha_3) - (t_0 + \alpha_1) = 2\alpha_3 - \alpha_1 \quad (2.5-44)$$

Thus, the mean waiting time on a class 1 packet is

$$\bar{W}_{\text{class 1}} = \frac{1}{2} \left(W_{t_0} + W_{t_0 + \alpha_1} \right) = \frac{2\alpha_3 - \alpha_1}{2} \quad (2.5-45)$$

Suppose now that $t_0 + \alpha_3 \leq t' < t_0 + \alpha_1$. In this case the busy period will consist of packets arriving at t' , $t_0 + \alpha_1$, and $t_0 + 2\alpha_1$. Clearly, the class 2 packet arriving at t' has no waiting time, thus,

$$W_{t'} = \bar{W}_{t'} = 0 \quad (2.5-46)$$

For the class 1 packets arriving at $t_0 + \alpha_1$ and $t_0 + 2\alpha_1$, we may write

$$W_{t_0 + \alpha_1} = (t' + \alpha_3) - (t_0 + \alpha_1) \quad (2.5-47a)$$

$$W_{t_0 + 2\alpha_1} = (t' + 2\alpha_3) - (t_0 + 2\alpha_1) \quad (2.5-47b)$$

Substituting the limits on t' we find

$$2\alpha_3 - \alpha_1 \leq W_{t_0 + \alpha_1} \leq \alpha_3 \quad (2.5-48a)$$

$$0 = 3\alpha_3 - 2\alpha_1 \leq W_{t_0 + 2\alpha_1} \leq 2\alpha_3 - \alpha_1 \quad (2.5-48b)$$

Once again, this waiting time must be uniformly distributed; thus,

$$\bar{w}_{t_0+\alpha_1} = \frac{3\alpha_3 - \alpha_1}{2} \quad (2.5-49a)$$

$$\bar{w}_{t_0+2\alpha_1} = \frac{5\alpha_3 - 3\alpha_1}{2} \quad (2.5-49b)$$

and

$$\begin{aligned} \bar{w}_{\text{class 1}} &= \frac{1}{2} \left(\bar{w}_{t_0+\alpha_1} + \bar{w}_{t_0+2\alpha_1} \right) \\ &= 2\alpha_3 - \alpha_1. \end{aligned} \quad (2.5-50)$$

At this point, we are ready to calculate the mean steady-state waiting time on a packet for each of the two specific packet classes. For the class 2 stream, the mean waiting time is given by

$$\begin{aligned} \bar{w}_{(2)}^{(3)} &= E\left\{w_t \mid t_0 < t' \leq t_0 + \alpha_3\right\} P\left\{t_0 < t' \leq t_0 + \alpha_3\right\} \\ &\quad + E\left\{w_t' \mid t_0 + \alpha_3 \leq t' < t_0 + \alpha_1\right\} P\left\{t_0 + \alpha_3 \leq t' < t_0 + \alpha_1\right\} \\ &= \left(\frac{\alpha_3}{2}\right) \left(\frac{\alpha_3}{\alpha_1}\right) + (0) \left(\frac{\alpha_1 - \alpha_3}{\alpha_1}\right) \\ &= \left(\frac{\alpha_3}{2}\right) \left(\frac{\rho_{13}}{\rho_{11}}\right) \\ &= \frac{\rho_{13}\alpha_3}{2} \end{aligned} \quad (2.5-51)$$

For the class 1 packets, we have

$$\begin{aligned} \bar{w}_{(1)}^{(3)} &= E\left\{w_{\text{class 1}} \mid t_0 < t' \leq t_0 + \alpha_3\right\} P\left\{t_0 < t' \leq t_0 + \alpha_3\right\} \\ &\quad + E\left\{w_{\text{class 1}} \mid t_0 + \alpha_3 \leq t' < t_0 + \alpha_1\right\} P\left\{t_0 + \alpha_3 \leq t' < t_0 + \alpha_1\right\} \end{aligned} \quad (2.5-52)$$

$$\begin{aligned}
 &= \left(\frac{2\alpha_3 - \alpha_1}{2} \right) \left(\frac{\alpha_3}{\alpha_1} \right) + (2\alpha_3 - \alpha_1) \left(\frac{\alpha_1 - \alpha_3}{\alpha_1} \right) \\
 &= \left(\frac{2\alpha_3 - \alpha_1}{2} \right) \left(\frac{\rho_{13}}{\rho_{11}} \right) + (2\alpha_3 - \alpha_1) \left(\frac{\rho_{11} - \rho_{13}}{\rho_{11}} \right) \\
 &= (2\alpha_3 - \alpha_1) \rho_{23} + \left(2\alpha_3 - \frac{3}{2} \alpha_3 \right) \rho_{23} \\
 &= \frac{\rho_{23} \alpha_3}{2} + \rho_{23} (2\alpha_3 - \alpha_1) .
 \end{aligned}$$

Thus, on comparing (2.5-51) and (2.5-52) with (2.5-37), we have proven the validity of (2.5-37) for $\alpha_2 = 2\alpha_1$, $\alpha_3 = \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$ and $\lambda_2 = 1/\alpha_2$. Clearly, (2.5-36) follows using the definition of $\bar{w}^{(3)}$.

Let us now relax the condition on α_3 and allow $\frac{1}{2} \min(\alpha_1, \alpha_2) \leq \alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$. Once again, the packets will arrive at node N_3 according to (2.5-38) and (2.5-39). Clearly, on the class 2 packets, the analysis that led to (2.5-43) and (2.5-46) is still valid, thus, yielding a mean waiting time, $\bar{w}_{(2)}^{(3)}$, given by (2.5-51). Turning to the class 1 packets, we note that, for $t_0 < t' \leq t_0 + \alpha_3$, (2.5-44) and (2.5-45) are still valid. However, the region $t_0 + \alpha_3 \leq t' < t_0 + \alpha_1$ must now be subdivided into two separate regions.

For the region $t_0 + \alpha_3 \leq t' \leq t_0 + (2\alpha_1 - 2\alpha_3)$, we find a busy period consisting of two packets arriving at t' and $t_0 + \alpha_1$ which, on the class 1 packets, yield waiting times of

$$W_{t_0} = 0 \quad (2.5-53a)$$

$$2\alpha_3 - \alpha_1 \leq W_{t_0 + \alpha_1} \leq \alpha_1 - \alpha_3 \quad (2.5-53b)$$

The mean waiting times are then found to be

$$\bar{W}_{t_0} = 0 \quad (2.5-54a)$$

$$\bar{W}_{t_0 + \alpha_1} = \frac{\alpha_3}{2} \quad (2.5-54b)$$

and

$$\bar{W}_{\text{class 1}} = \frac{\alpha_3}{4} \quad (2.5-54c)$$

On the other hand, for the region $t_0 + (2\alpha_1 - 2\alpha_3) \leq t' \leq t_0 + \alpha_1$, the busy period consists of packets arriving at t' , $t_0 + \alpha_1$ and $t_0 + 2\alpha_1$. In this case, the waiting times on the class 1 packets are found to be

$$\alpha_1 - \alpha_3 \leq W_{t_0 + \alpha_1} \leq \alpha_3 \quad (2.5-55a)$$

$$0 \leq W_{t_0 + 2\alpha_1} \leq 2\alpha_3 - \alpha_1 \quad (2.5-55b)$$

which will yield mean values of

$$\bar{W}_{t_0 + \alpha_1} = \frac{\alpha_1}{2} \quad (2.5-56a)$$

$$\bar{W}_{t_0 + 2\alpha_1} = \frac{2\alpha_3 - \alpha_1}{2} \quad (2.5-56b)$$

and

$$\bar{W}_{\text{class 1}} = \frac{\alpha_3}{2} \quad (2.5-56c)$$

Consequently, using (2.5-45), (2.5-54c) and (2.5-56c) we calculate the mean steady-state waiting time on a class 1 packet to be

$$\begin{aligned}
 \bar{w}_{(1)}^{(3)} &= E\left\{W_{\text{class } 1} \mid t_0 < t' \leq t_0 + \alpha_3\right\} P\left\{t_0 < t' \leq t_0 + \alpha_3\right\} & (2.5-57) \\
 &+ E\left\{W_{\text{class } 1} \mid t_0 + \alpha_3 \leq t' \leq t_0 + (2\alpha_1 - 2\alpha_3)\right\} \\
 &\quad P\left\{t_0 + \alpha_3 \leq t' \leq t_0 + (2\alpha_1 - 2\alpha_3)\right\} \\
 &+ E\left\{W_{\text{class } 1} \mid t_0 + (2\alpha_1 - 2\alpha_3) \leq t' < t_0 + \alpha_1\right\} \\
 &\quad P\left\{t_0 + (2\alpha_1 - 2\alpha_3) \leq t' < t_0 + \alpha_1\right\} \\
 &= \left(\frac{2\alpha_3 - \alpha_1}{2}\right) \left(\frac{\alpha_3}{\alpha_1}\right) + \left(\frac{\alpha_3}{4}\right) \left(\frac{2\alpha_1 - 3\alpha_3}{\alpha_1}\right) + \left(\frac{\alpha_3}{2}\right) \left(\frac{2\alpha_3 - \alpha_1}{\alpha_1}\right) \\
 &= \left(\frac{2\alpha_3 - \alpha_1}{2}\right) \left(\frac{\rho_{13}}{\rho_{11}}\right) + \left(\frac{\alpha_3}{4}\right) \left(\frac{2\rho_{11} - 3\rho_{13}}{\rho_{11}}\right) + \left(\frac{\alpha_3}{2}\right) \left(\frac{2\rho_{13} - \rho_{11}}{\rho_{11}}\right) \\
 &= \frac{\rho_{23}\alpha_3}{2} + \rho_{23}(2\alpha_3 - \alpha_1) ,
 \end{aligned}$$

where we have used the substitutions $\rho_{11} = 1$ and $\rho_{13} = 2\rho_{23}$.

Hence, we have extended our proof to allow for $\frac{1}{2} \min(\alpha_1, \alpha_2) \leq \alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$.

We are now in a position to relax the condition $\lambda_2 = \frac{1}{\alpha_2}$ and allow for $\lambda_2 \leq \frac{1}{\alpha_2}$. To show that this extension is valid we begin by noting that, since $\lambda_1 = \frac{1}{\alpha_1}$, equation (2.5-38) is still valid for class 1 arrivals at node N_3 . The class 2 packets, however, no longer arrive at N_3 according to (2.5-39). Nonetheless, it is critical to note that each arriving class 2 packet sees node N_3 in the same random manner as when $\lambda_2 = \frac{1}{\alpha_2}$. This is true since no busy period can have more than one class 2 packet. Consequently, it is clear that the mean waiting time on a class 2 packet is given by (2.5-37a) or (2.5-51). As for the class 1 packets,

they too see a similar network; as when $\lambda_2 = \frac{1}{\alpha_2}$, each time the busy period has a length greater than one. The only difference is that when $\lambda_2 \leq \frac{1}{\alpha_2}$, the number of such busy period decreases in a linear manner with λ_2 . Thus, we conclude that the mean waiting time on a class 1 packet will decrease linearly with decreasing λ_2 . On analysis of (2.5-37a), (or (2.5-52) or (2.5-57)), we see that this is indeed the case for the equation of $\bar{W}_{(1)}^{(3)}$ given.

Hence we have proven the validity of Theorem 2.5-4 for the case $\alpha_2 = 2\alpha_1$.

We now turn to the case $\alpha_2 = 3\alpha_1$ and will prove the validity of (2.5-36) and (2.5-37) for the special case $\alpha_3 = \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}$ and $\lambda_2 = \frac{1}{\alpha_2}$. The steps involved in the proof are essentially the same as for $\alpha_2 = 2\alpha_1$; thus, we shall proceed through the derivation rather quickly.

Clearly, equation (2.5-38), (2.5-39) and (2.5-40) still apply for the arrival times at N_3 . For $\alpha_2 = 3\alpha_1$, (2.5-41) becomes

$$\alpha_1 = \frac{4}{3} \alpha_3 \quad (2.5-58a)$$

$$\alpha_3 = 4\alpha_3 \quad (2.5-58b)$$

On simple analysis, it is seen that we now have busy periods of length four, consisting of three class 1 packets and a single class 2 packet. Once again, we must consider two subregions for t' . For $t_0 < t' \leq t_0 + \alpha_3$, we have a busy period consisting of packets arriving at t_0 , t' , $t_0 + \alpha_1$,

and $t_0+2\alpha_1$. The waiting time on these packets are found to be limited to

$$W_{t_0} = 0 \quad (2.5-59a)$$

$$0 \leq W_{t'} \leq \alpha_3 \quad (2.5-59b)$$

$$W_{t_0+\alpha_1} = 2\alpha_3 - \alpha_1 \quad (2.5-59c)$$

$$W_{t_0+2\alpha_1} = 3\alpha_3 - 2\alpha_1 \quad (2.5-59d)$$

The mean waiting times are then found to be

$$\bar{W}_{t'} = \frac{\alpha_3}{2} \quad (2.5-60a)$$

$$\bar{W}_{\text{class 1}} = \frac{2\alpha_3 - \alpha_1}{3} + \frac{3\alpha_3 - 2\alpha_1}{3} \quad (2.5-60b)$$

For $t_0+\alpha_3 \leq t' < t_0+\alpha_1$, we have a busy period consisting of packets arriving at t' , $t_0+\alpha_1$, $t_0+2\alpha_1$, and $t_0+3\alpha_1$. Thus, the waiting times are now calculated as

$$W_{t'} = 0 \quad (2.5-61a)$$

$$2\alpha_3 - \alpha_1 \leq W_{t_0+\alpha_1} \leq \alpha_3 \quad (2.5-61b)$$

$$3\alpha_3 - 2\alpha_1 \leq W_{t_0+2\alpha_1} \leq 2\alpha_3 - \alpha_1 \quad (2.5-61c)$$

$$0 = 4\alpha_3 - 3\alpha_1 \leq W_{t_0+3\alpha_1} \leq 3\alpha_3 - 2\alpha_1 \quad (2.5-61d)$$

As for the mean values, we obtain

$$\bar{W}_{t_0+\alpha_1} = \frac{3\alpha_3 - \alpha_1}{2} \quad (2.5-62a)$$

$$\bar{W}_{t_0+2\alpha_1} = \frac{5\alpha_3 - 3\alpha_1}{2} \quad (2.5-62b)$$

$$\bar{W}_{t_0+3\alpha_1} = \frac{7\alpha_3 - 5\alpha_1}{2} \quad (2.5-62c)$$

which leads to

$$\bar{w}_{\text{class 1}} = \frac{15\alpha_3 - 9\alpha_1}{6} = \frac{5\alpha_3 - 3\alpha_1}{2} \quad (2.5-63)$$

Finally, proceeding as in (2.5-51) and (2.5-52), we find

$$\begin{aligned} \bar{w}_{(2)}^{(3)} &= \left(\frac{\alpha_3}{2}\right) \left(\frac{\alpha_3}{\alpha_1}\right) + (0) \left(\frac{\alpha_1 - \alpha_3}{\alpha_1}\right) \\ &= \left(\frac{\alpha_3}{2}\right) \left(\frac{\rho_{13}}{\rho_{11}}\right) \\ &= \frac{\rho_{13}\alpha_3}{2} \end{aligned} \quad (2.5-64a)$$

and

$$\begin{aligned} \bar{w}_{(1)}^{(3)} &= \left[\frac{2\alpha_3 - \alpha_1}{3} + \frac{3\alpha_3 - 2\alpha_1}{3} \right] \left(\frac{\alpha_3}{\alpha_1}\right) + \left(\frac{5\alpha_3 - 3\alpha_1}{2}\right) \left(\frac{\alpha_1 - \alpha_3}{\alpha_1}\right) \\ &= \left[\frac{2\alpha_3 - \alpha_1}{3} + \frac{3\alpha_3 - 2\alpha_1}{3} \right] (\rho_{13}) + \left(\frac{5\alpha_3 - 3\alpha_1}{2}\right) (1 - \rho_{13}) \\ &= [(2\alpha_3 - \alpha_1) + (3\alpha_3 - 2\alpha_1)] \rho_{23} + \left(\frac{5\alpha_3 - 4\alpha_3}{2}\right) (\rho_{23}) \\ &= \frac{\rho_{23}\alpha_3}{2} + \rho_{23} [(2\alpha_3 - \alpha_1) + (3\alpha_3 - 2\alpha_1)] \end{aligned} \quad (2.5-64b)$$

which are the relations we were seeking.

At this point, we could relax the restriction on α_3 and λ_2 and allow $\frac{1}{2} \min(\alpha_1, \alpha_2) \leq \alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$ and $\lambda_2 \leq \frac{1}{\alpha_2}$. The proofs involved parallel those above for $\alpha_2 = 2\alpha_1$ and are, therefore, omitted at this time. With these relaxations, we have proven Theorem 2.5-4 for $\alpha_2 = 3\alpha_1$.

To complete the proof of the theorem, we assume that Theorem 2.5-4 is valid for $\alpha_2 = k\alpha_1$, k an integer and show that this implies validity for $\alpha_2 = (k+1)\alpha_1$. Proceeding

as was done for $k = 2$ and 3 , we would find in a direct manner that $\bar{W}_{(2)}^{(3)}$ remains equal to $\frac{\rho_{13}\alpha_3}{2}$ and that the extra mean delay experienced by a class 1 packet by having $\alpha_2 = (k+1)\alpha_1$ would be

$$(\bar{W}_{(1)}^{(3)})_{\text{extra}} = \rho_{23} [(k+1)\alpha_3 - k\alpha_1]^+ \quad (2.5-65)$$

which would prove our theorem by induction. Q.E.D.

As yet, we have been unable to remove the restriction in Theorem 2.5-4 that λ_1 be equal to $1/\alpha_1$. We believe that unlike Theorem 2.5-4 which has been shown true for external inputs following any general probability distribution, if we allow $\lambda_1 \leq 1/\alpha_1$, exact results for the mean packet waiting time at node N_3 of Figure 2.5-1 would depend on the probability distribution of the input streams.

Before leaving this section, we would like to make a conjecture on the mean packet waiting time at node N_3 for a situation where α_2 is not an integer multiple of α_1 .

Conjecture 2.5-1. For the network of Figure 2.5-1, the steady-state mean waiting time on a specific class packet, when $\alpha_3 = \frac{\alpha_1\alpha_2}{\alpha_1 + \alpha_2}$, $\lambda_1 = 1/\alpha_1$, $\lambda_2 = 1/\alpha_2$ and $\alpha_2 = (k_1/k_2)\alpha_1$ where k_1 and k_2 are integers is given by

$$\bar{W}_{(1)}^{(3)} = \frac{\rho_{23}\alpha_3}{2} + \frac{\rho_{23}}{k_2} \sum_{j=1}^{k_1-1} \left(\left\lceil \frac{j\alpha_1}{\alpha_3} \right\rceil \alpha_3 - j\alpha_1 \right) \quad (2.5-66a)$$

$$\bar{W}_{(2)}^{(3)} = \frac{\rho_{13}\alpha_3}{2} + \frac{\rho_{13}}{k_1} \sum_{j=1}^{k_2-1} \left(\left\lceil \frac{j\alpha_2}{\alpha_3} \right\rceil \alpha_3 - j\alpha_2 \right) \quad (2.5-66b)$$

where $\lceil x \rceil \triangleq$ the least integer greater than or equal to x . Furthermore, the mean steady-state waiting time on an arbitrary class packet is given by

$$\begin{aligned} \bar{w}^{(3)} = & \rho_{13}\rho_{23} \left[\alpha_3 + \frac{1}{k_2} \sum_{j=1}^{k_1-1} \left(\left\lceil \frac{j\alpha_1}{\alpha_3} \right\rceil \alpha_3 - j\alpha_1 \right) \right. \\ & \left. + \frac{1}{k_1} \sum_{n=1}^{k_2-1} \left(\left\lceil \frac{n\alpha_2}{\alpha_3} \right\rceil \alpha_3 - n\alpha_2 \right) \right] \end{aligned} \quad (2.5-67)$$

The conjecture was obtained by deriving the exact results for various specified values of k_1 and k_2 ; for example, $\alpha_2 = \frac{3}{2} \alpha_1$, $\alpha_2 = \frac{5}{2} \alpha_1$, and $\alpha_2 = \frac{4}{3} \alpha_1$. The procedure used is similar to that used in deriving Theorem 2.5-4 and does yield exact results for any specific values of k_1 and k_2 tried. To briefly outline the procedure one would use, we begin by noting that (2.5-38) and (2.5-39) are still applicable for the arrival times at N_3 . Equation (2.5-41) can be generalized to $\alpha_i = \frac{k_1+k_2}{k_i} \alpha_3$, $i = 1, 2$. On careful analysis, one finds that each busy period at N_3 will consist of k_1 class 1 and k_2 class 2 packets. Finally, the two regions that must be considered are $t_0 < t' \leq t_0 + \alpha_3/k_2$ and $t_0 + \alpha_3/k_2 \leq t' < t_0 + \alpha_1/k_2$. To complete the procedure, one would then derive equations of the form (2.5-42)-(2.5-57).

We believe (2.5-66) and (2.5-67) to be exact results. We further believe that we may drop the restriction on α_3 and allow for $\alpha_3 \leq \frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}$. However, we have as yet not proven these results true for random k_1 and k_2 and, therefore, have the results as a conjecture.

Finally, in Figure 2.5-13, we have plotted $\bar{W}^{(3)}/\alpha_3$ vs. α_2/α_1 using (2.5-67). The graph clearly indicates some very nonintuitive results. The waiting time $\bar{W}^{(3)}$ is seen to decrease between the points where α_2 is an integer multiple of α_1 . However, as α_2/α_1 goes to infinity, we see that $\bar{W}^{(3)}$ approaches $\alpha_3/2$ asymptotically. Thus, Figure 2.5-13 clearly exhibits the inherent difficulties in obtaining exact results in the region considered.

2.6 Message Switching Networks

Until this point we have limited our discussion to packet switching networks where the length of all messages entering the network was fixed at a single packet. In this section, we relax this restriction and allow variable length messages; that is, we allow each message entering the network to consist of a random number of fixed length packets where this random number is said to follow any general probability distribution. We shall refer to a message as having arrived at a given node as soon as the first packet of the message has arrived at that node. On the other hand, we shall refer to a message as having departed from a given node of a network only after the last packet of the message has completely left the given node. We shall assume that the server, located at a given node, may begin processing a message as soon as its first packet has arrived at that node. We refer to this type of network as a message switching network in the sense that once a server has begun processing the first packet of a message, it shall remain dedicated to that message and shall process all packets belonging to that message before being allowed to begin service on another message that may have already arrived and is waiting in the nodes buffer. Finally, as done previously, we assume that each node of a network has infinite storage capability and will service messages according to a first-come first-served queueing discipline.

We now return to the networks of Figures 2.2-1 and 2.2-2 and prove that all the theorems and corollaries derived in Section 2.2 for the packet switching networks may be extended and applied to message switching networks.

Before beginning our derivations, we redefine all the notation presented in earlier sections to refer to messages instead of packets; for example,

$t_i \triangleq$ the arrival time of the i^{th} arbitrary class message into the network,

$M_j \triangleq$ the number of messages serviced at node N of Figure 2.2-1 during its j^{th} busy period,

and so forth. The only exception is the service time at a node, α or α_i , which is still in terms of time to service a packet. In addition, we introduce the following two quantities:

$L_i \triangleq$ the length, in packets, of the message that initially entered the network at time t_i

$L'_i \triangleq$ the length, in packets, of the i^{th} message to arrive at node N_3 of Figure 2.2-2

$\bar{L} \triangleq E\{L_i\} = E\{L'_i\}.$

Theorem 2.6-1.

The number of messages serviced at node N of Figure 2.2-1 during its j^{th} busy period equals the number of messages serviced at node N_3 of Figure 2.2-2 during its j^{th} busy period; that is,

$$M_j = M_j^{(3)}, \quad j = 1, 2, 3, \dots \quad (2.6-1)$$

Proof: Once again, we begin by looking at the first busy period at node N of Figure 2.2-1. We observe that in order to have a busy period of length M_1 , the message arrival times must be

$$\begin{aligned} t_1 &= t_1 & (2.6-2) \\ t_1 &\leq t_2 \leq t_1 + L_1 \alpha \\ t_2 &\leq t_3 \leq t_1 + (L_1 + L_2) \alpha \\ &\vdots \\ t_{M_1-1} &\leq t_{M_1} \leq t_1 + \sum_{i=1}^{M_1-1} L_i \alpha \end{aligned}$$

Then, since it will take $\sum_{i=1}^{M_1} L_i \alpha$ seconds to service these M_1 messages, the busy period will be broken by the arrival time of the $(M_1+1)^{st}$ message at

$$t_{M_1+1} > t_1 + \sum_{i=1}^{M_1} L_i \alpha \quad (2.6-3)$$

Continuing as before, we find that for Figure 2.2-2

$$\begin{aligned} t_1' &= t_1 + \alpha & (2.6-4) \\ t_1 + \alpha &\leq t_2' \leq t_1 + (L_1 + 1) \alpha \\ t_2 + \alpha &\leq t_3' \leq t_1 + (L_1 + L_2 + 1) \alpha \\ &\vdots \\ t_{M_1-1} + \alpha &\leq t_{M_1}' \leq t_1 + \left(\sum_{i=1}^{M_1-1} L_i \alpha \right) + \alpha \\ t_{M_1+1}' &> t_1 + \left(\sum_{i=1}^{M_1} L_i \alpha \right) + \alpha. \end{aligned}$$

Furthermore, from (2.6-4), one can calculate that

$$T_1' = t_1 + \alpha \quad (2.6-5)$$

$$T_2' \leq t_1 + (L_1 + 1)\alpha$$

$$T_3' \leq t_1 + (L_1 + L_2 + 1)\alpha$$

⋮

$$T_{M_1}' \leq t_1 + \left(\sum_{i=1}^{M_1-1} L_i \alpha \right) + \alpha$$

$$T_{M_1+1}' > t_1 + \left(\sum_{i=1}^{M_1} L_i \alpha \right) + \alpha$$

We now examine the departure times of messages from node N_3 of Figure 2.2-2. To begin,

$$D_1' = T_1' + L_1' \alpha = t_1 + L_1' \alpha + \alpha \quad (2.6-6)$$

We now have two possibilities, $L_1' = L_1$ or $L_1' \neq L_1$.

If $L_1' = L_1$ then

$$D_1' = t_1 + (L_1 + 1)\alpha \geq T_2' \quad (2.6-7)$$

On the other hand, if $L_1' \neq L_1$ then we have that

$$D_1' \geq t_1 + 2\alpha \quad (2.6-8)$$

from which we can state that

$$D_1' > t_1' = T_2' \quad (2.6-9)$$

Thus, in either case, $L_1' = L_1$ or $L_1' \neq L_1$, we have insured $T_2' \leq D_1'$ which insures that the busy period will last at least for a second message.

Having established that the busy period will last for at least two messages, we may write

$$D_2' = t_1 + (L_1' + L_2')\alpha + \alpha \quad (2.6-10)$$

We now test all possibilities on L_1' and L_2' and show that the busy period extends for at least three messages.

If $L_1' = L_1$ and $L_2' = L_2$, then

$$D_2' = t_1 + (L_1 + L_2 + 1)\alpha \geq T_3' \quad (2.6-11)$$

If $L_1' = L_1$ and $L_2' \neq L_2$, then

$$D_2' \geq t_1 + L_1\alpha + 2\alpha \quad (2.6-12)$$

and

$$T_3' \leq t_2' \leq t_1 + (L_1 + 1)\alpha < D_2' \quad (2.6-13)$$

If $L_1' \neq L_1$ and $L_2' = L_2$, this implies $L_2' = L_1$ and (2.6-11) is applicable. Finally, if $L_1' \neq L_1$ and $L_2' \neq L_2$, this implies $L_2' = L_1$ and this time (2.6-12) and (2.6-13) are seen to apply. Thus, for all cases, we have $T_3' \leq D_2'$, implying that the busy period at N_3 will continue for at least another message.

We now assume that the busy period has lasted for $M_1 - 1$ messages; that is,

$$T_{i+1}' \leq D_i' \quad , \quad i=1,2,\dots,M_1-2 \quad (2.6-14)$$

and show that this implies that the busy period must last for another message; that is,

$$T_{M_1}' \leq D_{M_1-1}' \quad (2.6-15)$$

Furthermore, since $t'_{M_1+1} > t'_i$, $i = 1, 2, \dots, M_1$, we can write that

$$D'_{M_1-1} = t_1 + \sum_{i=1}^{M_1} L_i \alpha + \alpha - L'_{M_1} \alpha \quad (2.6-16)$$

where L'_{M_1} is equal to the length of that message that has yet to be processed at N_3 . In addition, we can say

$$T'_{M_1} = t'_j \quad (2.6-17a)$$

and

$$L'_{M_1} = L_j, \quad j=2 \text{ or } 3 \text{ or } 4 \text{ or } \dots \text{ or } M_1 \quad (2.6-17b)$$

Finally, using (2.6-4) and (2.6-16) we see that for any value of j , we must have $T'_{M_1} \leq D'_{M_1-1}$ which is what we were seeking. Thus, we conclude that the busy period must last for at least M_1 messages. Furthermore, from (2.6-16) and (2.6-15) we have

$$D'_{M_1} = t_1 + \sum_{i=1}^{M_1} L_i \alpha + \alpha \quad (2.6-18)$$

Thus, we have

$$T'_{M_1+1} > D'_{M_1} \quad (2.6-19)$$

causing the busy period to end after exactly M_1 messages. Consequently, we have shown $M_1 = M_1^{(3)}$. Furthermore, by simple induction, this result extends to all subsequent busy periods, thus, proving our theorem. Q.E.D.

As was the case with Theorem 2.2-1, Theorem 2.6-1 can be used to derive the following three corollaries.

Corollary 2.6-1. The probability distributions of the number of messages serviced during a busy period at nodes N of Figure 2.2-1 and N_3 of Figure 2.2-2 are identical; that is,

$$P\{M_j=m\} = P\{M_j^{(3)}=m\} , \quad m=1,2,\dots \quad (2.6-20)$$

Corollary 2.6-2. The duration of the j^{th} idle period at node N of Figure 2.2-6 equals the duration of the j^{th} idle period at node N_3 of Figure 2.2-2; that is,

$$I_j = I_j^{(3)} , \quad j=1,2,\dots \quad (2.6-21)$$

Corollary 2.6-3. The probability distributions of the duration of the j^{th} idle period at nodes N and N_3 are identical; that is,

$$P\{I_j \leq x\} = P\{I_j^{(3)} \leq x\} , \quad x > 0 . \quad (2.6-22)$$

The derivation of these corollaries parallel those of Corollaries 2.2-1, 2.2-2, and 2.2-3 and are, therefore, omitted at this time.

We now proceed, as was done in the packet switching case, to derive a theorem concerning the message waiting time in the networks considered. For a message, the waiting time it experiences will be considered to be the time that the first packet of a given message must wait on queue. The delay on succeeding packets of a given message are, in a message switching system, easily calculated from this quantity.

Theorem 2.6-2

For the j^{th} busy periods of nodes N and N_3 , the mean total waiting time experienced by the M_j messages in the network of Figure 2.2-1 equals the mean total waiting time experienced by the $M_j^{(3)}$ messages over the entire network of Figure 2.2-2.

Proof: The derivation parallels that of Theorem 2.2-2. Once again, we begin by calculating the total waiting experienced by the M_j messages of the j^{th} busy period at node N of Figure 2.2-1.

For message switching networks, equation (2.2-11) may be rewritten as

$$W_1 = 0 \tag{2.6-23}$$

$$W_{i+1} = [W_i + L_i \alpha - (t_{i+1} - t_i)]^+, \quad i=1, 2, \dots$$

For a busy period of length M_j , (2.6-23) yields

$$W_1 = 0 \tag{2.6-24}$$

$$W_2 = t_1 - t_2 + L_1 \alpha$$

$$W_3 = t_1 - t_3 + (L_1 + L_2) \alpha$$

$$\vdots$$

$$W_{M_j} = t_1 - t_{M_j} + \sum_{i=1}^{M_j-1} L_i \alpha$$

Hence, the mean total waiting time in the busy period is

$$\bar{W}_{\text{total}} = E \left\{ \sum_{i=1}^{M_j} W_i \right\} = M_j t_1 - \sum_{i=1}^{M_j} t_i + \frac{M_j (M_j - 1)}{2} \bar{L} \alpha \tag{2.6-25}$$

We will now prove our theorem using mathematical induction.

For the case $M_j = M_j^{(3)} = 1$, it is clear that the total waiting time, in both networks, is identically zero. We now assume that (2.6-25) is true for both networks for $M_j = M_j^{(3)} \leq m$. We will show that this assumption implies that, for $M_j = M_j^{(3)} = m+1$, the mean total waiting time experienced in each of the two networks by these $m+1$ messages will still be the same.

For Figure 2.2-1, the extra waiting time introduced by having M_j equal to $m+1$ instead of m , is seen, from (2.6-24), to be

$$(W_{\text{extra}})_{\text{Fig.2.2-1}} = W_{m+1} = t_1 - t_{m+1} + \sum_{i=1}^m L_i \alpha \quad (2.6-26a)$$

and

$$(\bar{W}_{\text{extra}})_{\text{Fig.2.2-1}} = E\{W_{m+1}\} = t_1 - t_{m+1} + m \bar{L} \alpha \quad (2.6-26b)$$

We now turn to Figure 2.2-2 where we again assume that this $(m+1)^{\text{st}}$ message, arriving at time t_{m+1} , enters the network at node N_1 . This message will then arrive at node N_3 at time

$$t'_{m+1} = t_{m+1} + W_k^{(1)} + \alpha \quad (2.6-27)$$

where we have assumed the message to be the k^{th} , $1 \leq k \leq m+1$, to arrive at N_1 . At node N_3 , we can write (similar to (2.6-23), that

$$\begin{aligned} W_1^{(3)} &= 0 \\ W_{i+1}^{(3)} &= [W_i^{(3)} + L_i' - (T'_{i+1} - T'_i)]^+, \quad i=1,2,\dots \end{aligned} \quad (2.6-28a)$$

For a busy period of length $m+1$, (2.6-28a) yields

$$\begin{aligned}
 W_1^{(3)} &= 0 & (2.6-28b) \\
 W_2^{(3)} &= T_1' - T_2' + L_1' \alpha \\
 W_3^{(3)} &= T_1' - T_3' + (L_1' + L_2') \alpha \\
 &\vdots \\
 W_m^{(3)} &= T_1' - T_m' + \sum_{i=1}^{m-1} L_i' \alpha \\
 W_{m+1}^{(3)} &= T_1' - T_{m+1}' + \sum_{i=1}^m L_i' \alpha
 \end{aligned}$$

We now take the expected values in (2.6-28b), and we obtain

$$\begin{aligned}
 \bar{W}_1^{(3)} &= 0 & (2.6-29) \\
 \bar{W}_2^{(3)} &= T_1' + T_2' + \bar{L} \alpha \\
 \bar{W}_3^{(3)} &= T_1' - T_3' + 2\bar{L} \alpha \\
 &\vdots \\
 \bar{W}_m^{(3)} &= T_1' - T_m' + (m-1) \bar{L} \alpha \\
 \bar{W}_{m+1}^{(3)} &= T_1' - T_{m+1}' + m \bar{L} \alpha
 \end{aligned}$$

In general, this $(m+1)^{st}$ message will be the n^{th} message of the j^{th} busy period to arrive at node N_3 , $1 \leq n \leq m+1$; that is, $t_{m+1}' = T_n'$. Furthermore, from (2.6-29), we see that the mean extra waiting time, at node N_3 , caused by having $M_j^{(3)}$ equal to $m+1$ instead of m , is given by

$$\bar{W}_{extra}^{(3)} = T_1' - T_n' + m\bar{L}\alpha \quad (2.6-30)$$

If we now substitute for T_1' and T_n' by using relations derived from (2.6-5) and (2.6-27) respectively, we find

that

$$\bar{w}_{\text{extra}}^{(3)} = t_1 - t_{m+1} - W_k^{(1)} + m\bar{L}\alpha \quad (2.6-31)$$

Hence

$$\begin{aligned} (\bar{w}_{\text{extra}}^{(3)})_{\text{Fig.2.2-2}} &= \bar{w}_{\text{extra}}^{(1)} + \bar{w}_{\text{extra}}^{(3)} & (2.6-32) \\ &= W_k^{(1)} + \bar{w}_{\text{extra}}^{(3)} \\ &= t_1 - t_{m+1} + m\bar{L}\alpha \\ &= (\bar{w}_{\text{extra}}^{(3)})_{\text{Fig.2.2-1}} \end{aligned}$$

Thus the mean total waiting time in the j^{th} busy period, for $M_j = M_j^{(3)} = m+1$, is the same in both networks. This completes our induction procedure and proves our theorem.

Q.E.D.

In summary Theorems 1 and 2 have shown that for any given busy period both the number of messages serviced and the mean total waiting time experienced by these messages will be equivalent for the two networks considered. This leads directly to the following corollary.

Corollary 2.6-4 For the networks of Figures 2.2-1 and 2.2-2, the steady-state overall mean waiting time experienced by an arbitrary class message is the same in both networks; that is,

$$(\bar{S})_{\text{Fig.2.2-1}} = (\bar{S})_{\text{Fig.2.2-2}} \quad (2.6-33)$$

Thus, we have shown the equivalency with respect to overall mean steady-state message waiting time that exists

between the two networks. As before, we may utilize this equivalency to calculate the mean steady-state message waiting time at node N_3 of Figure 2.2-2 and find

$$\begin{aligned} \bar{W}^{(3)} &= (\bar{S})_{\text{Fig.2.2-2}} - \bar{W}^{(1)} - \bar{W}^{(2)} & (2.6-34) \\ &= (\bar{S})_{\text{Fig.2-2-1}} - \bar{W}_{(1)}^{(1)} P\{\text{a message is class 1}\} \\ &\quad - \bar{W}_{(2)}^{(2)} P\{\text{a message is class 2}\}. \end{aligned}$$

Having completed our analysis of Figures 2.2-1 and 2.2-2, we may, as was done in the packet switching case, clearly extend our results to networks as depicted by Figure 2.2-3; that is, to networks where the service time at the merger node is the slowest in the network. In addition, the generalizations, to networks with an arbitrary number of external inputs, discussed in Section 2.4, for packet switched networks, can easily be seen to directly apply in the message switched case being considered. Finally, we note that the discussion of self delay and interference delay is also applicable to message switching networks.

Turning to networks with fast merger nodes, we remark that for message switching networks, we can no longer differentiate between cases where the service time at the merger node is greater than and the case where the service time is less than the parallel combination of the service times of the channels feeding the merger node.

The reason for this is, of course, because we now allow multi-packet messages, thus allowing the possibility of having an infinite queue length at the merger node. Thus, we need only consider networks where the service time at the merger node is less than or equal to the service time of at least one of the incoming channels.

For the case of fast merger nodes, we can at this time only extend the results of Section 2.5 which yield an upper bound on the actual waiting time. Using a similar procedure to that of Section 2.5, we could consider a network as depicted in Figure 2.5-11 and obtain the following theorem.

Theorem 2.6-3

For external inputs following any general probability distribution and $\alpha_{k+1} < \alpha_i$ for at least one value of i , $i = 1, 2, \dots, k$, the mean steady-state waiting time on an arbitrary class message at node N_{k+1} of Figure 2.5-11 can be approximated by considering a network with $\alpha_i = \alpha_{k+1}$ for each $\alpha_i > \alpha_{k+1}$.

This concludes our discussion of message switched networks. This section also concludes our discussion of networks with single output channels.

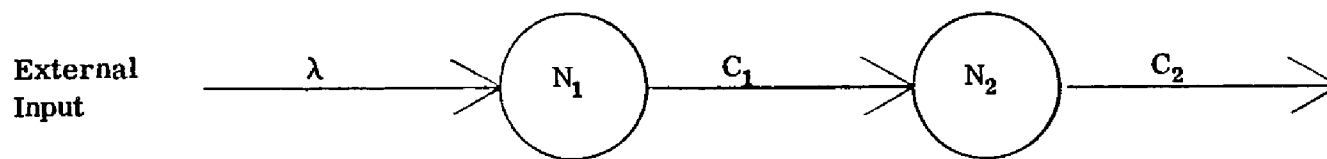


Fig. 2.1-1: A Two Stage Tandem Queueing Network

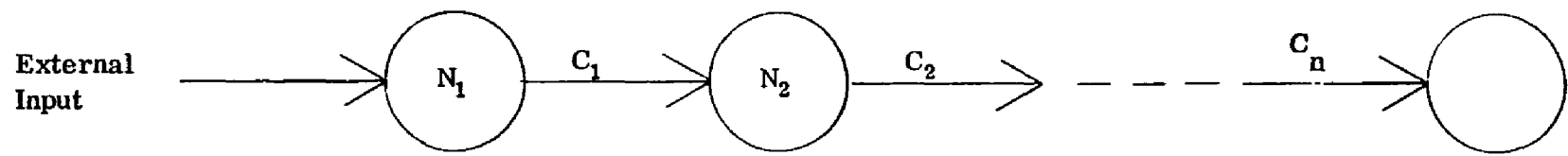


Fig. 2.1-2: An n-Channel Communications Path

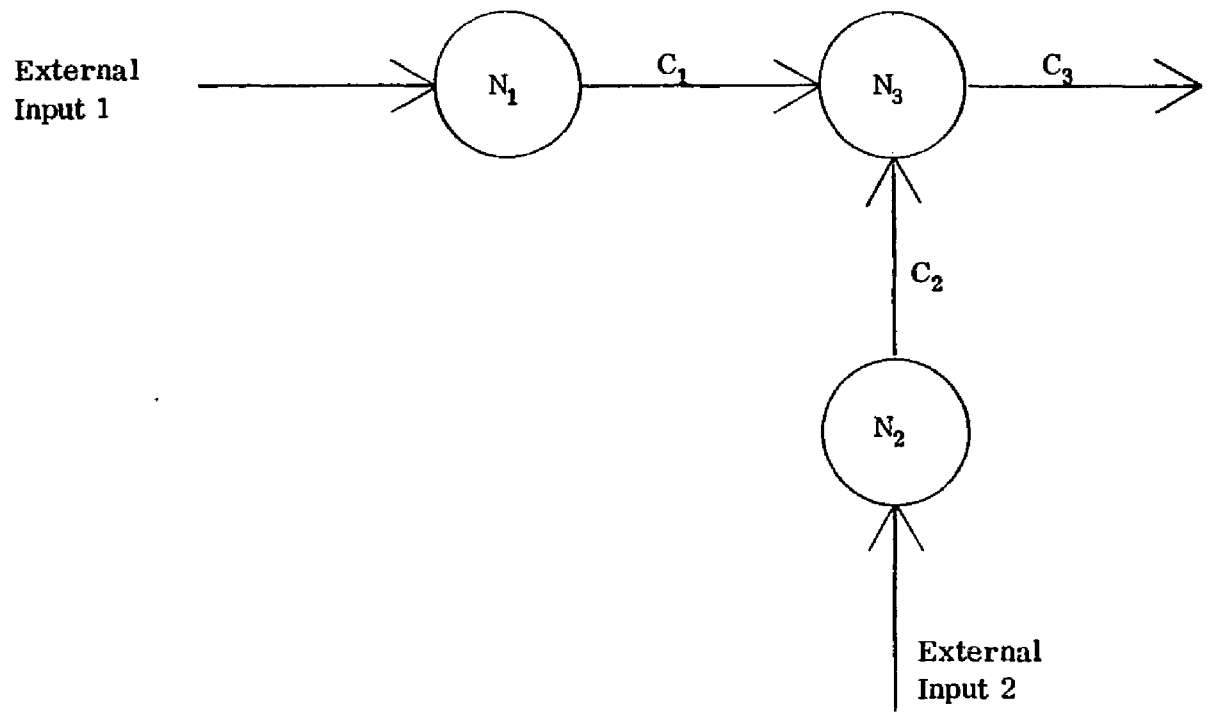


Fig. 2.1-3: Network in which Packet Interference Occurs at Node N_3

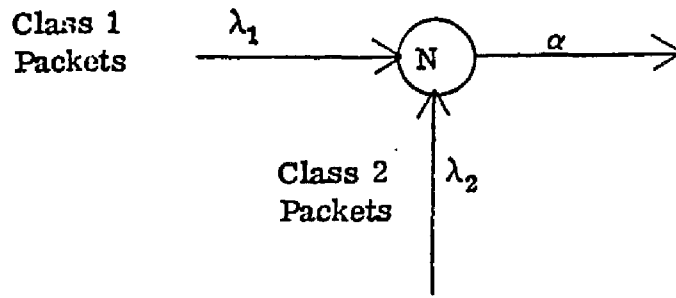


Fig.2. 2-1: A Queueing Network Consisting of a Single Server Queue and Two External Input Streams

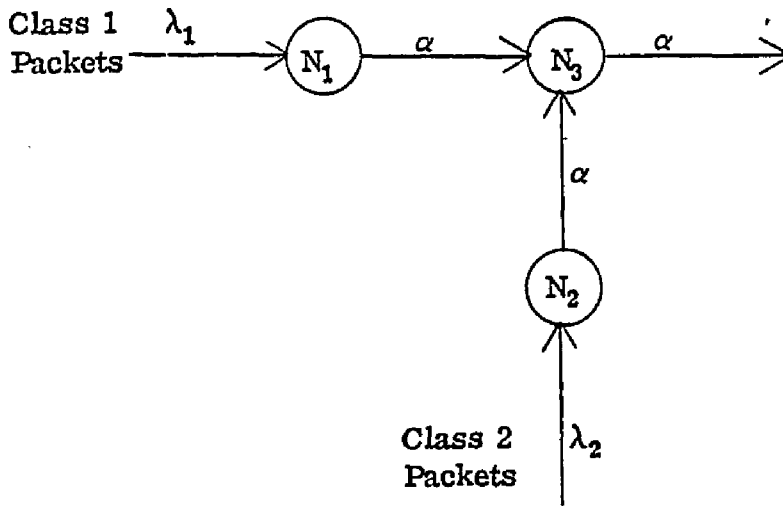
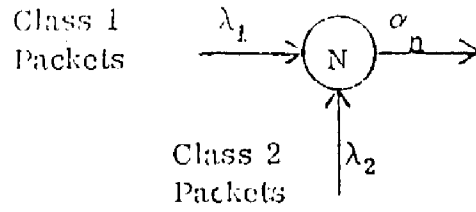


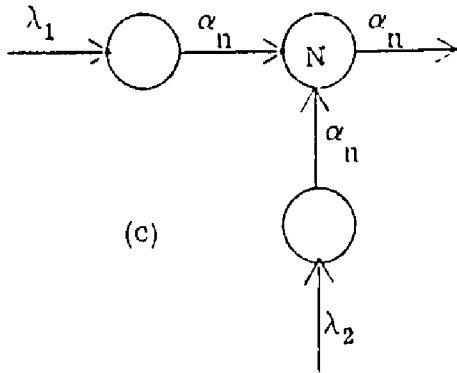
Fig. 2.2-2: A Queueing Network Consisting of Three Single Server Queues and Two External Input Streams



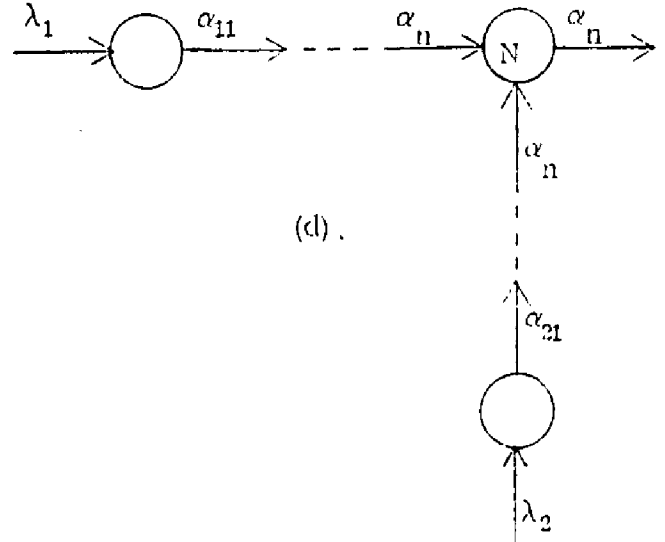
(a)



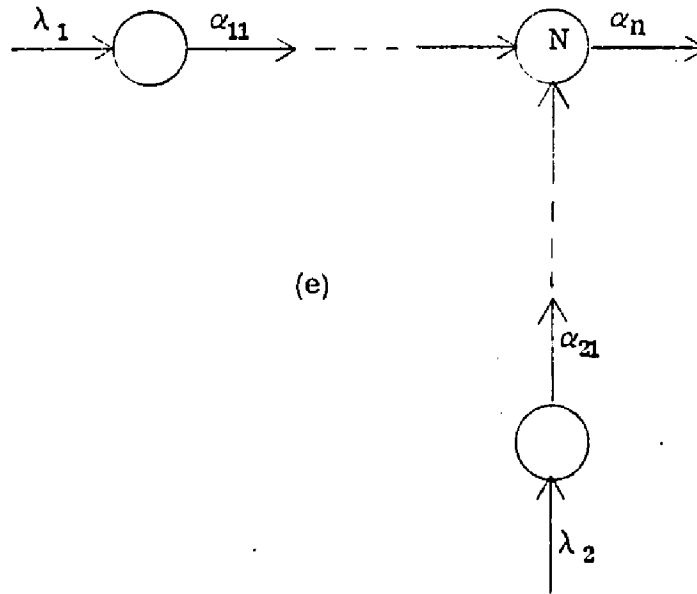
(b)



(c)



(d)



(e)

Fig. 2.2-3: Networks With Equivalent Overall Steady-State Mean Waiting Time when $\alpha_n = \max(\alpha_{11}, \alpha_{12}, \dots, \alpha_{21}, \alpha_{22}, \dots, \alpha_n)$ and $\lambda = \lambda_1 + \lambda_2$

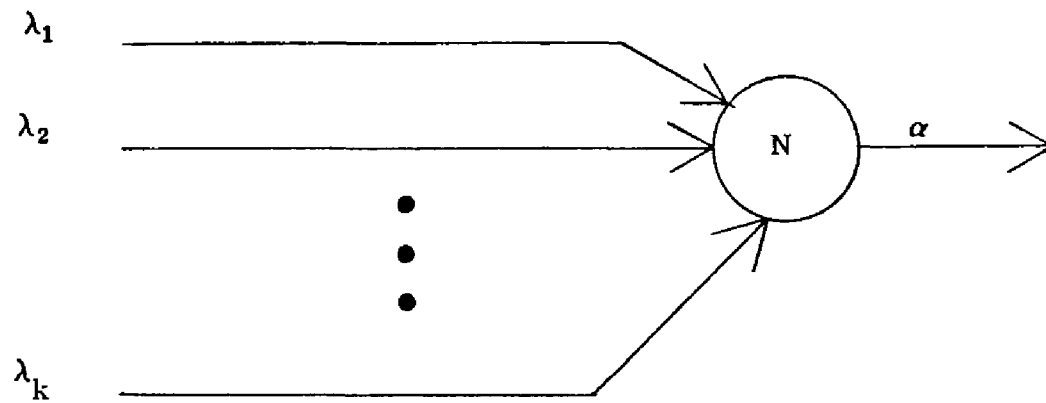


Fig. 2.4-1: A Queueing Network Consisting of a Single Server Queue and
" k " External Inputs

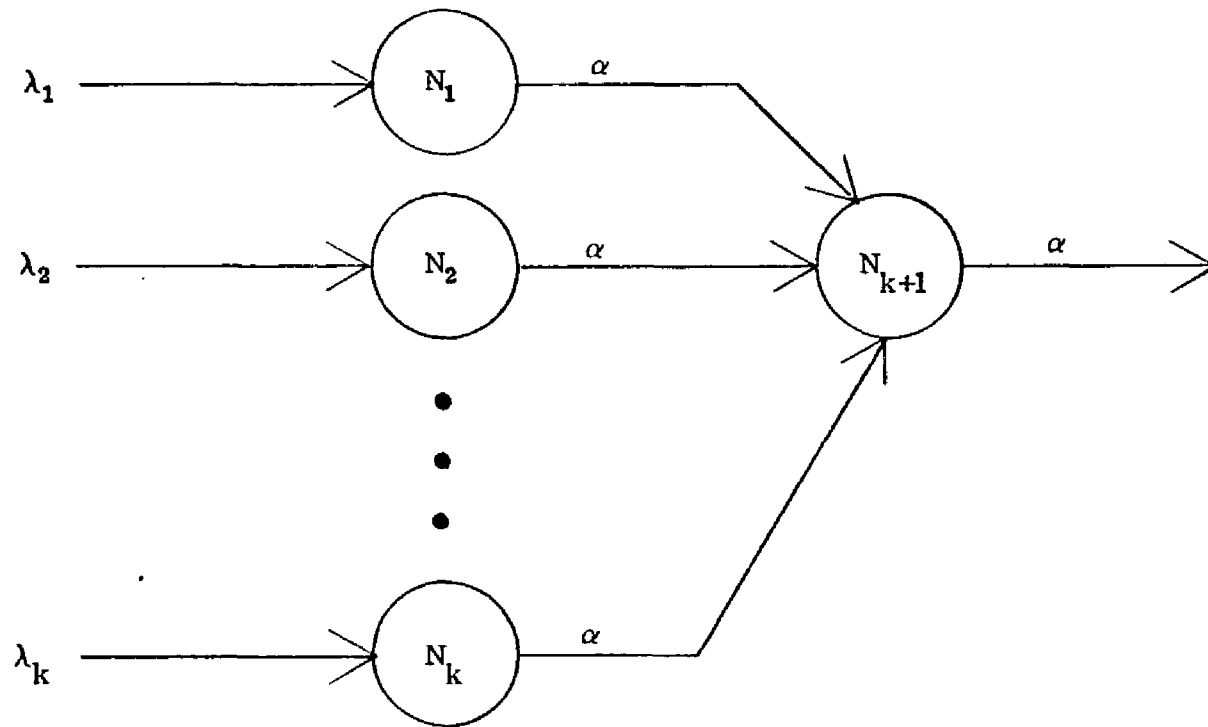


Fig. 2.4-2: A Queueing Network Consisting of " $k+1$ " Single Server Queues and
 " k " External Inputs

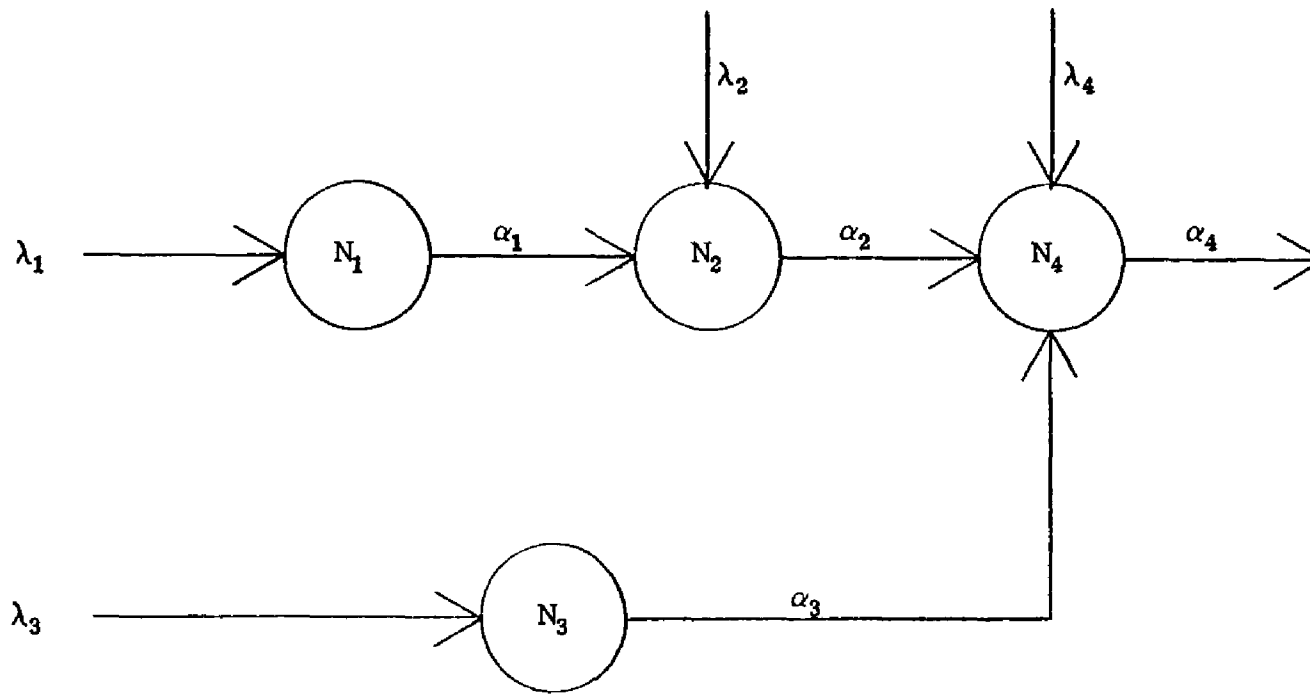


Fig. 2.4-3: A Queueing Network Consisting of Four Single Server Queues where $\alpha_4 \geq \alpha_3$ and $\alpha_4 \geq \alpha_2 \geq \alpha_1$

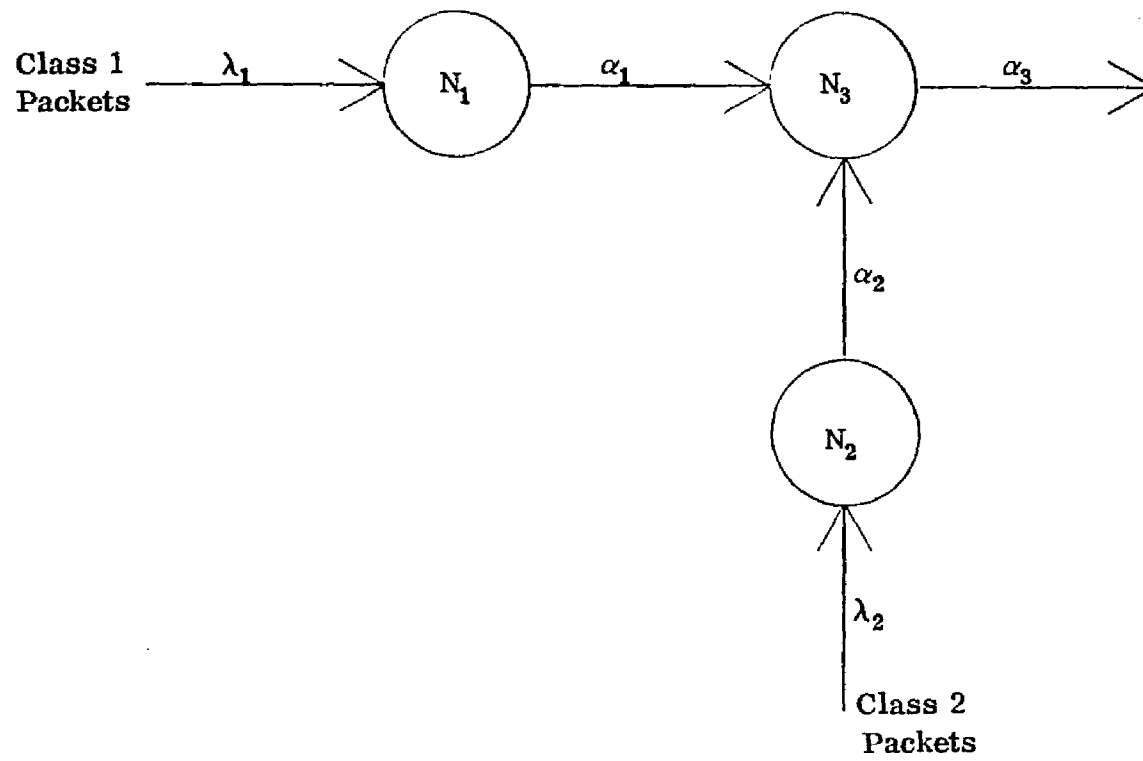


Fig. 2.5-1: A Three Node Network with Two External Inputs

Waiting Time at N_2 as a Function of α_3

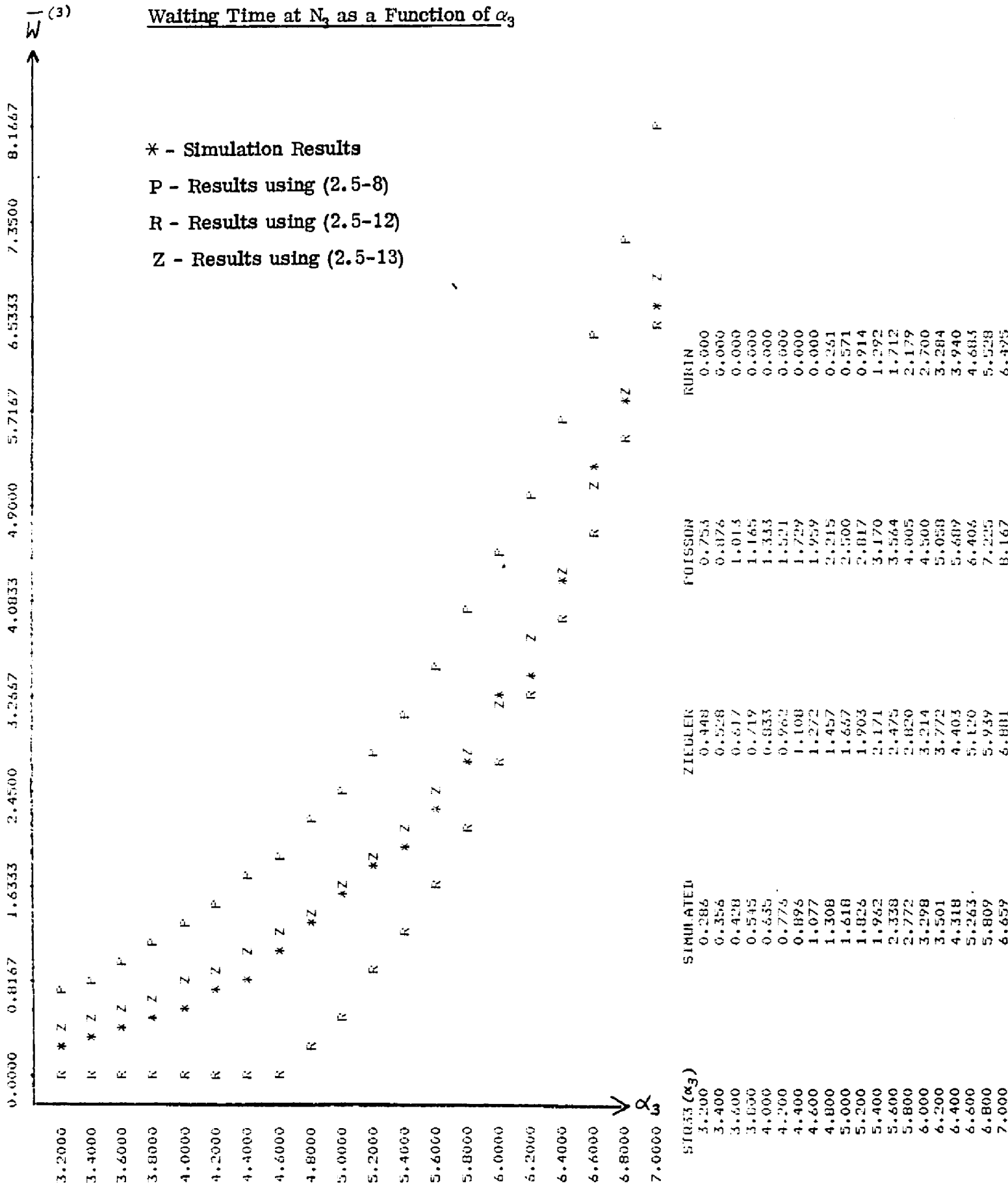


Fig. 2.5-2: Comparison of Results Obtained Using Eqs. (2.5-8), (2.5-12)

and (2.5-13) for the case $\lambda_1 = \lambda_2 = 0.05$ and $\alpha_1 = \alpha_2 = 5$

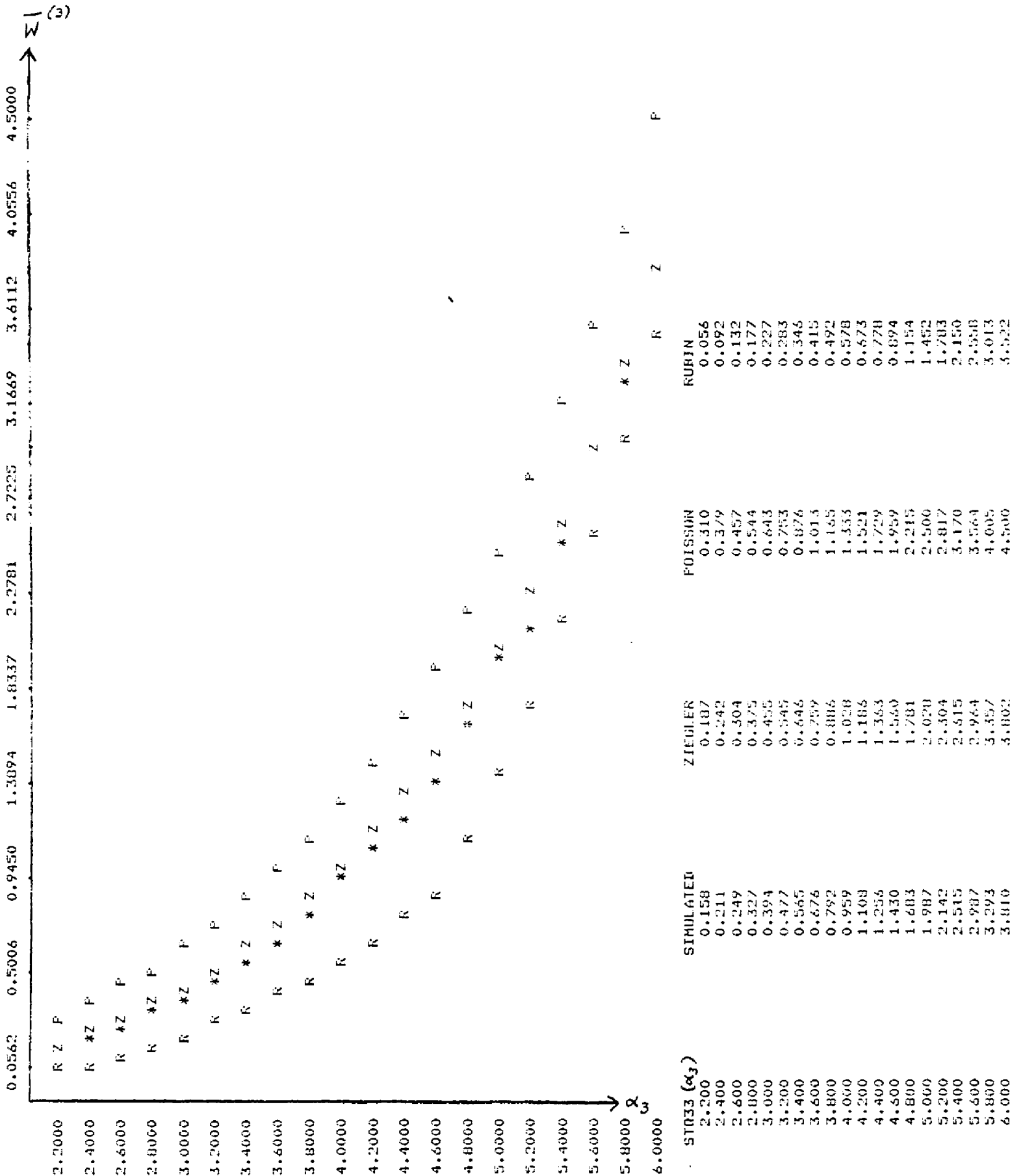
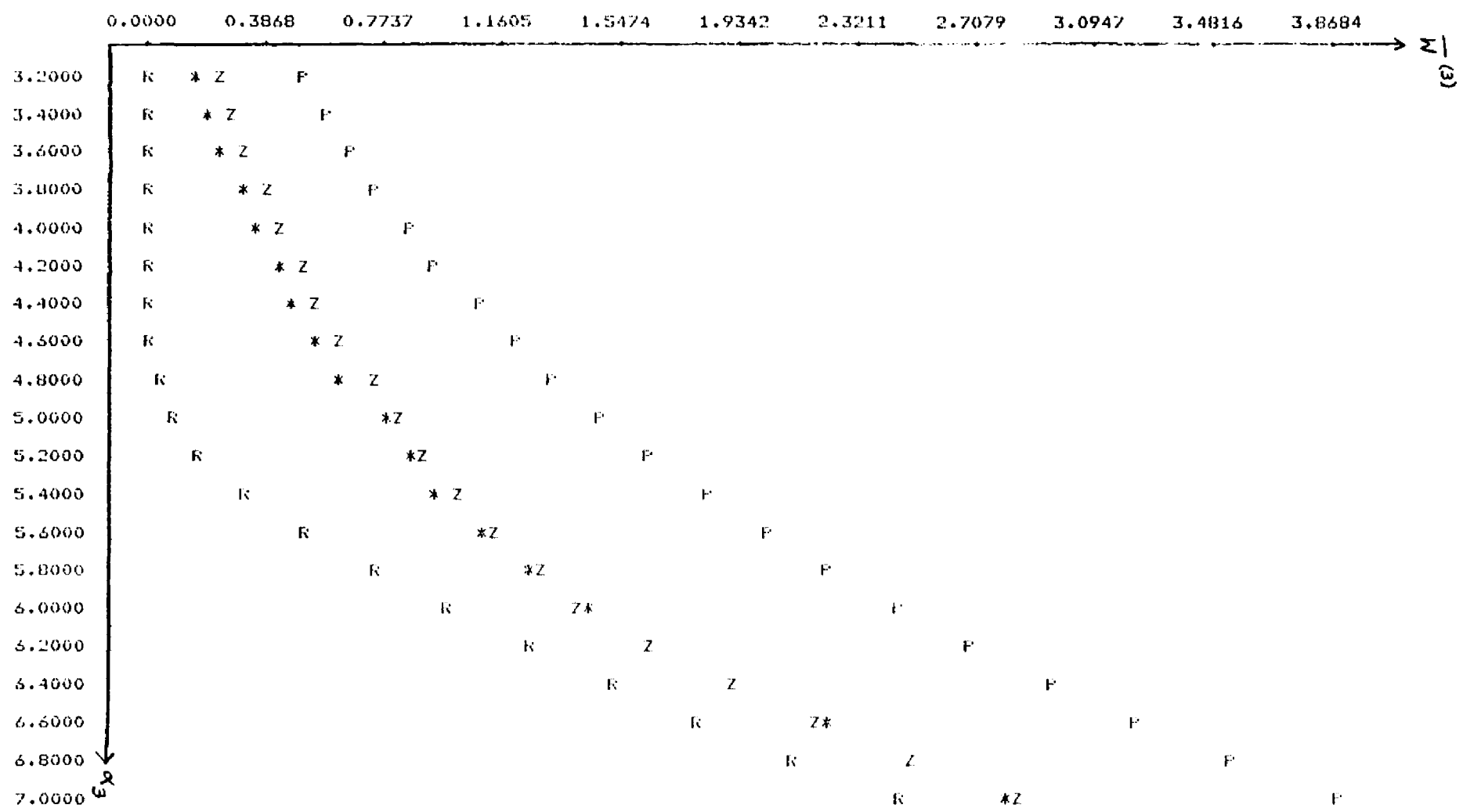


Fig. 2.5-3: Comparison of Results for $\lambda_1 = \lambda_2 = 0.05$, $\alpha_1 = 6$ and $\alpha_2 = 2$

Fig. 2.5-4: Results for $\lambda_1 = 0.05$, $\lambda_2 = 0.025$, and $\alpha_1 = \alpha_2 = 6$



α_3	STR33 (α_3)	SIMULATED	ZIEGLER	POTSSON	RIBIN
3.200	0.190	0.190	0.256	0.505	0.000
3.400	0.227	0.227	0.297	0.582	0.000
3.600	0.255	0.255	0.343	0.666	0.000
3.800	0.329	0.329	0.394	0.757	0.000
4.000	0.374	0.374	0.450	0.857	0.000
4.200	0.427	0.427	0.511	0.966	0.000
4.400	0.465	0.465	0.579	1.084	0.000
4.600	0.562	0.562	0.654	1.211	0.000
4.800	0.625	0.625	0.736	1.350	0.048
5.000	0.777	0.777	0.825	1.500	0.103
5.200	0.888	0.888	0.924	1.662	0.162
5.400	0.966	0.966	1.032	1.838	0.339
5.600	1.114	1.114	1.150	2.028	0.541
5.800	1.247	1.247	1.279	2.233	0.758
6.000	1.450	1.450	1.421	2.455	0.991
6.200	1.630	1.630	1.661	2.694	1.243
6.400	1.898	1.898	1.920	2.954	1.514
6.600	2.240	2.240	2.261	3.235	1.806
6.800	2.493	2.493	2.505	3.539	2.122
7.000	2.812	2.812	2.835	3.868	2.454

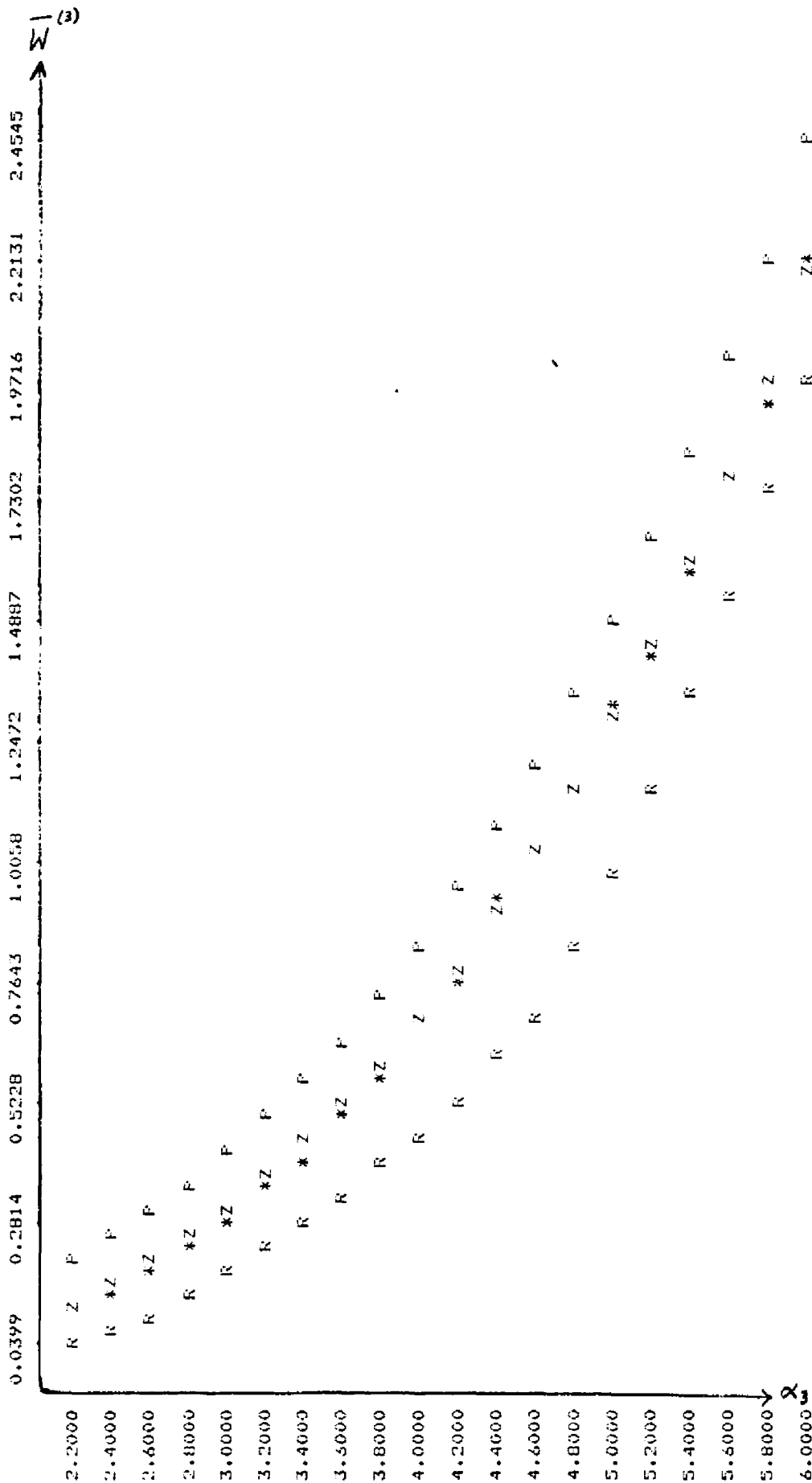


Fig. 2.5-5: Results for $\lambda_1 = 0.025$, $\lambda_2 = 0.05$, $\alpha_1 = 6$, and $\alpha_2 = 2$

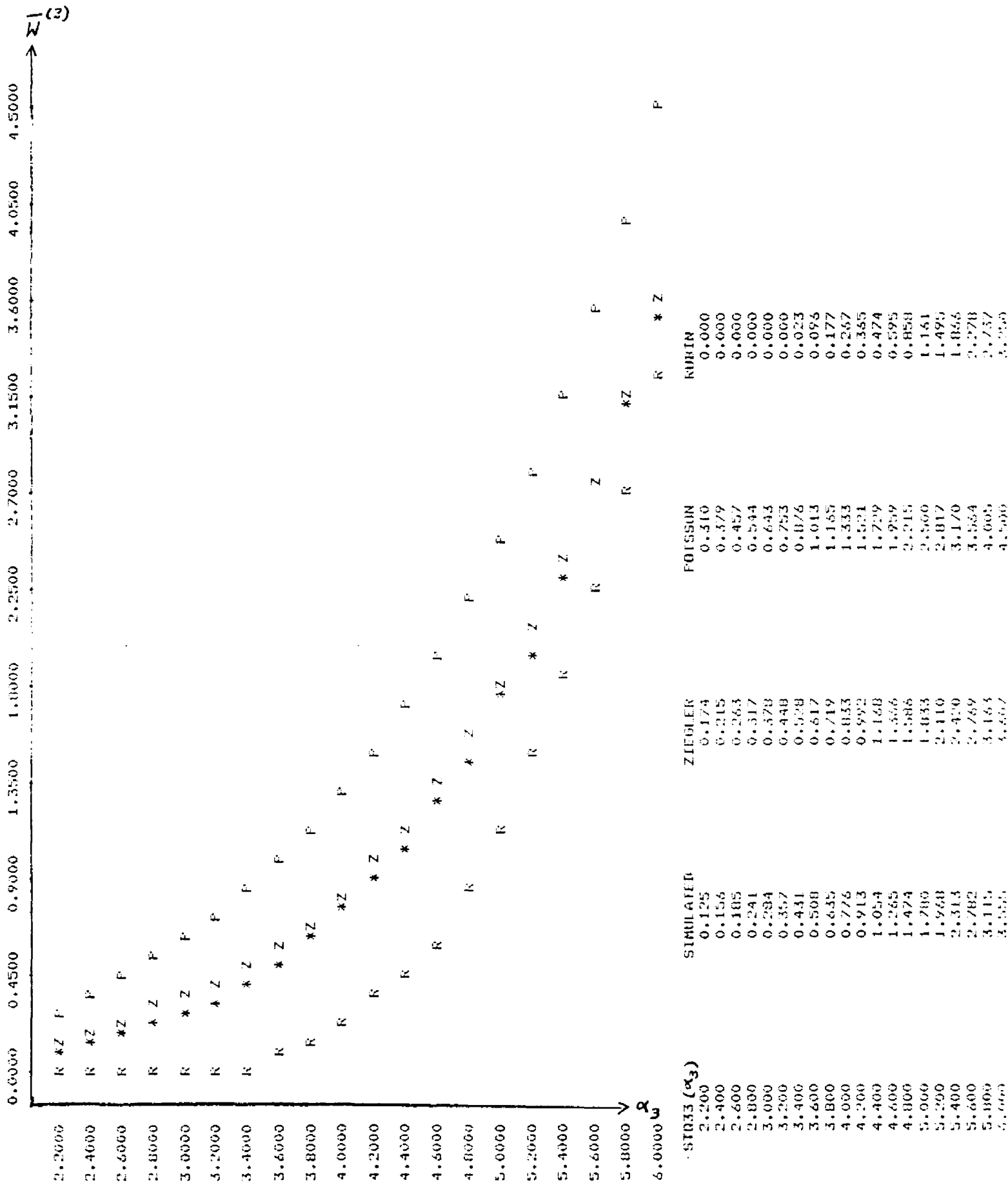


Fig. 2.5-7: Results for $\lambda_1 = \lambda_2 = 0.05$, $\alpha_1 = 6$, and $\alpha_2 = 4$

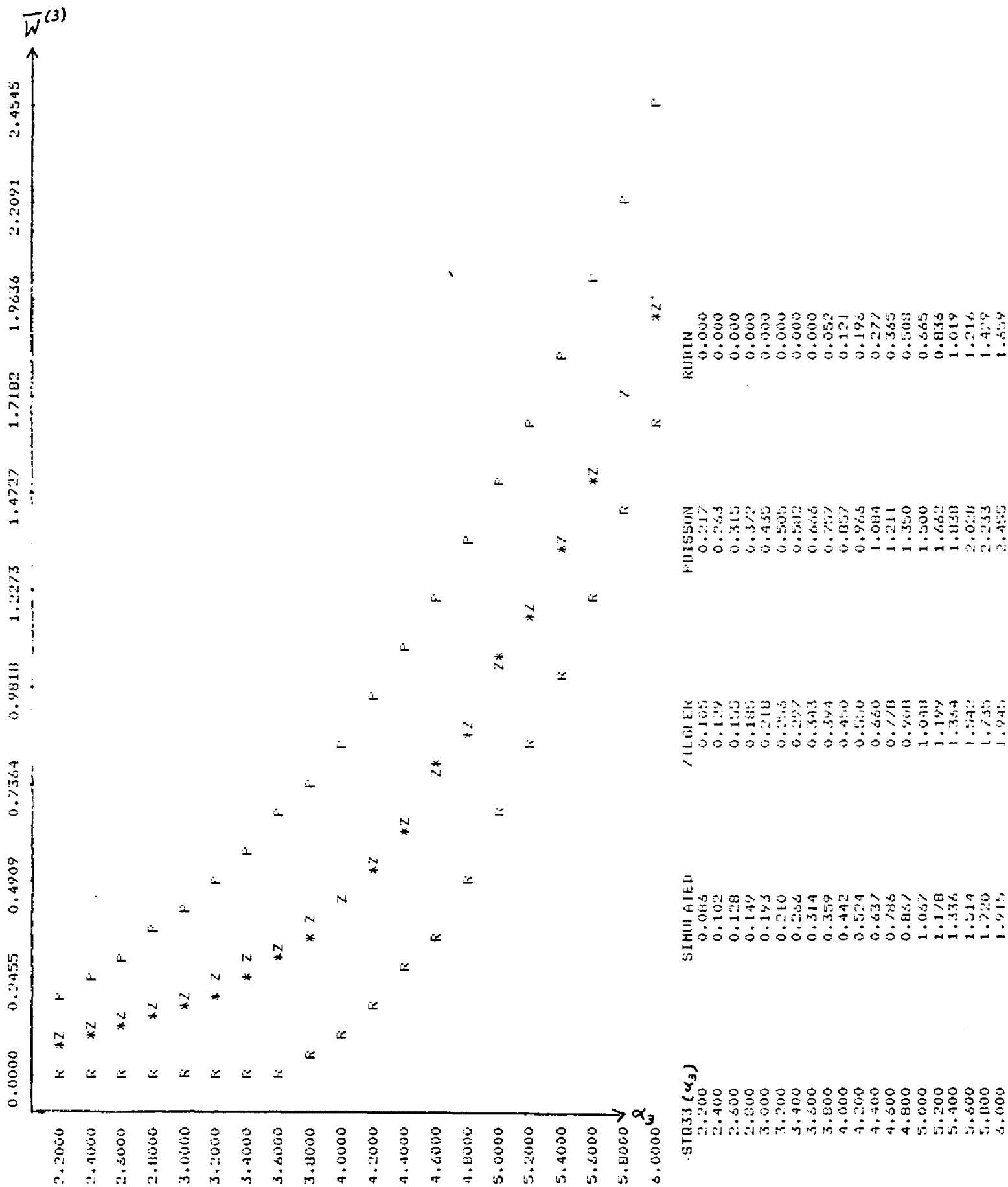


Fig. 2.5-8: Results for $\lambda_1 = 0.025$, $\lambda_2 = 0.05$, $\alpha_1 = 6$, and $\alpha_2 = 4$

\bar{M} (e)

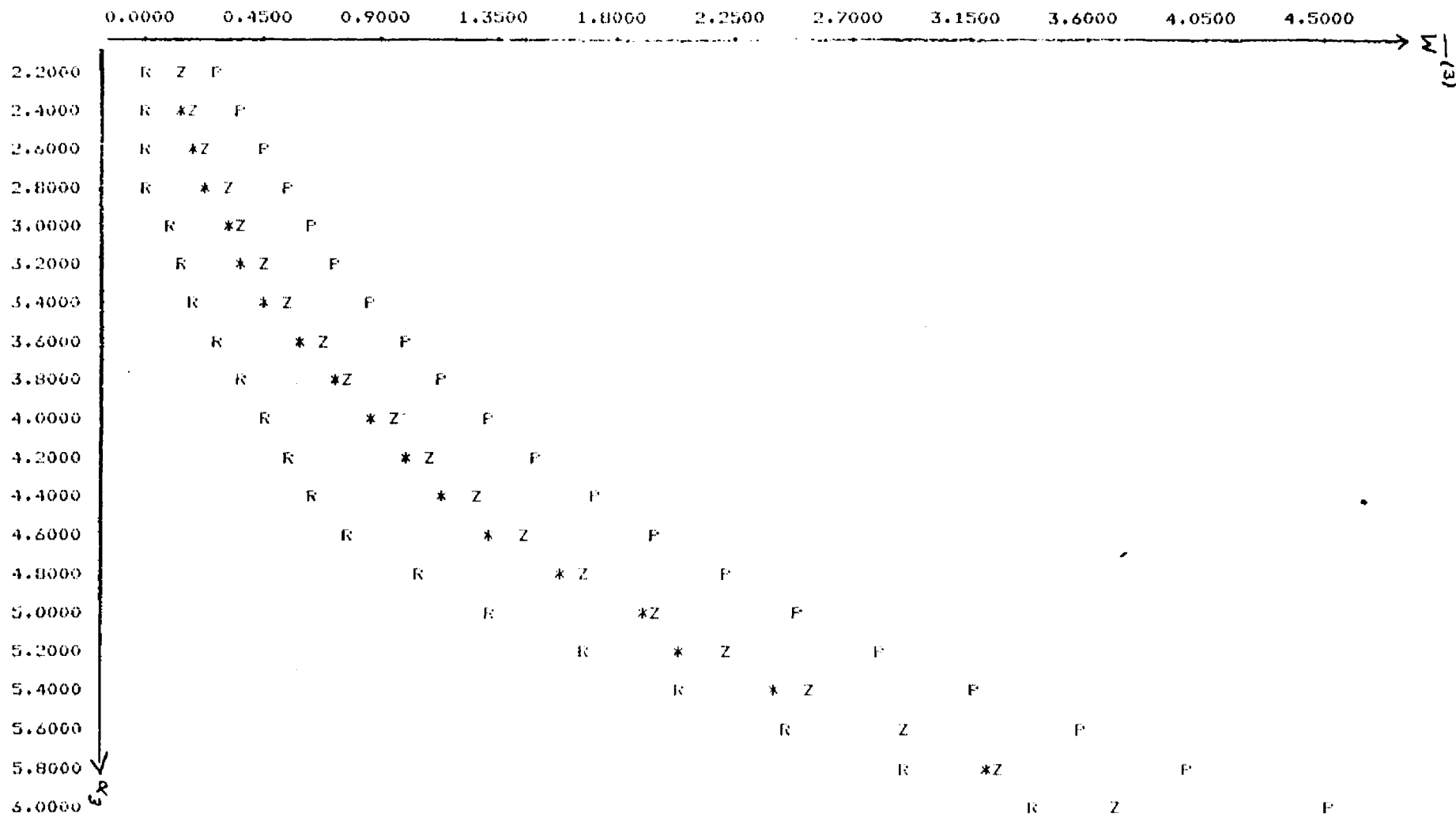
0.0000 0.2455 0.4909 0.7354 0.9818 1.2273 1.4727 1.7182 1.9636 2.2091 2.4545



σ_p (k ₃)	SIMULATED	ZIEGLER	FOLSIEN	ROBIN
2,200	0.083	0.105	0.217	0.000
2,400	0.103	0.129	0.263	0.000
2,600	0.122	0.155	0.315	0.000
2,800	0.145	0.185	0.372	0.000
3,000	0.173	0.218	0.435	0.000
3,200	0.194	0.255	0.505	0.000
3,400	0.237	0.297	0.582	0.010
3,600	0.278	0.343	0.665	0.040
3,800	0.315	0.394	0.757	0.072
4,000	0.390	0.450	0.857	0.108
4,200	0.465	0.519	0.966	0.145
4,400	0.554	0.595	1.084	0.188
4,600	0.641	0.679	1.211	0.233
4,800	0.705	0.771	1.350	0.281
5,000	0.816	0.870	1.500	0.333
5,200	0.906	0.979	1.662	0.390
5,400	1.045	1.098	1.838	0.563
5,600	1.249	1.298	2.028	0.762
5,800	1.388	1.369	2.133	0.976
6,000	1.507	1.503	2.185	1.206

Fig. 2.5-9: Results for $\lambda_1 = 0.05$, $\lambda_2 = 0.025$, $\alpha_1 = 6$, and $\alpha_2 = 4$

Fig. 2.5-10: Results for $\lambda_1 = \lambda_2 = 0.05$, $\alpha_1 = 6$, and $\alpha_2 = 3$



α_3	STUSS (α_3)	SIMULATED	ZIEGLER	FOUSSON	RUBIN
2.200	0.139	0.139	0.174	0.310	0.000
2.400	0.172	0.172	0.215	0.379	0.000
2.600	0.206	0.206	0.263	0.457	0.000
2.800	0.259	0.259	0.317	0.544	0.045
3.000	0.316	0.316	0.378	0.643	0.096
3.200	0.397	0.397	0.468	0.753	0.154
3.400	0.491	0.491	0.569	0.876	0.218
3.600	0.590	0.590	0.683	1.013	0.289
3.800	0.720	0.720	0.809	1.165	0.368
4.000	0.885	0.885	0.951	1.333	0.455
4.200	1.031	1.031	1.109	1.521	0.551
4.400	1.156	1.156	1.286	1.729	0.658
4.600	1.350	1.350	1.483	1.959	0.776
4.800	1.618	1.618	1.704	2.215	1.037
5.000	1.907	1.907	1.951	2.500	1.337
5.200	2.044	2.044	2.228	2.817	1.659
5.400	2.420	2.420	2.538	3.170	2.038
5.600	2.890	2.890	2.887	3.564	2.447
5.800	3.210	3.210	3.280	4.005	2.904
6.000	3.708	3.708	3.725	4.500	3.415

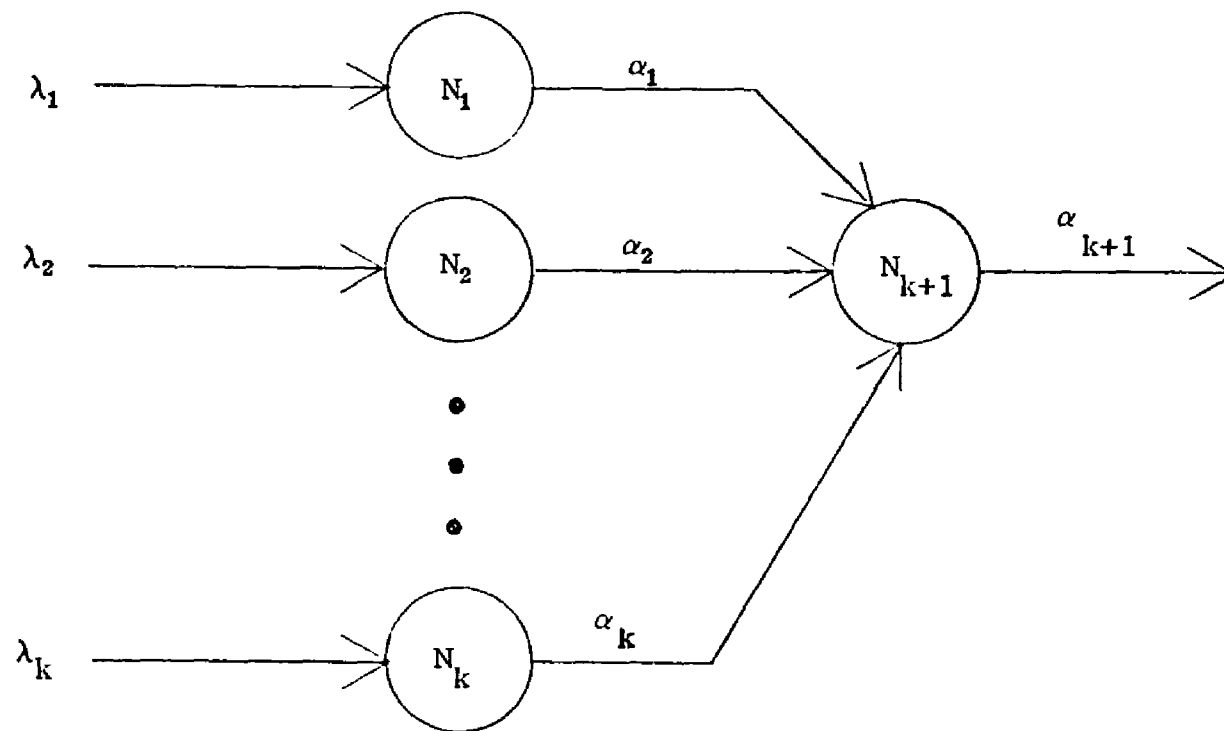


Fig. 2.5-11: A Queueing Network with " k " External Inputs

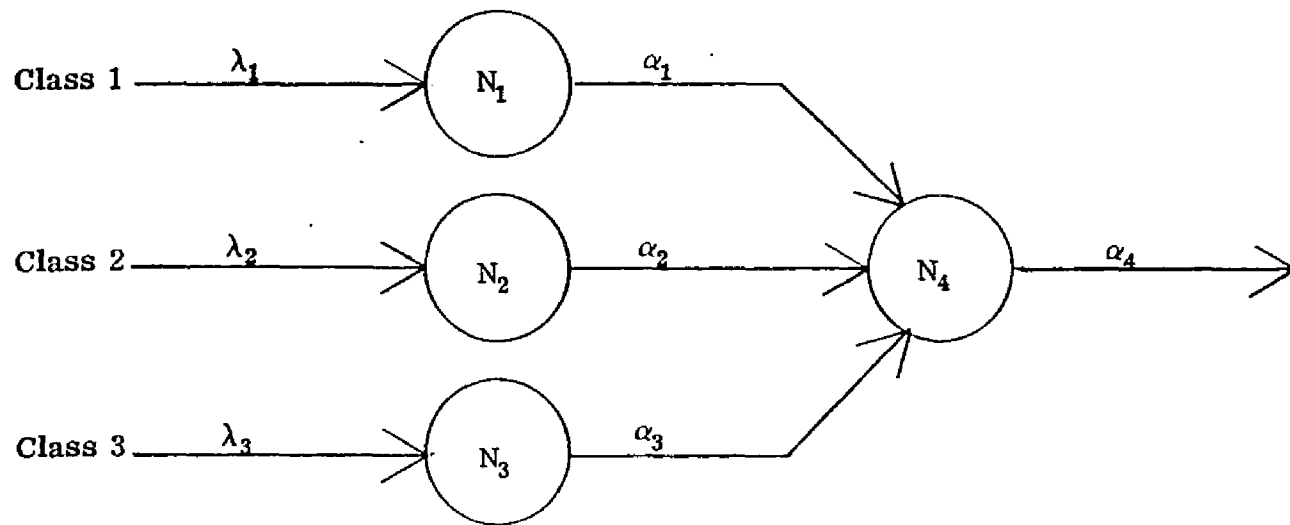


Fig. 2.5-12: A Three Input , Four Node Queueing Network with $\alpha_4 \leq 1/3 \min(\alpha_1, \alpha_2, \alpha_3)$

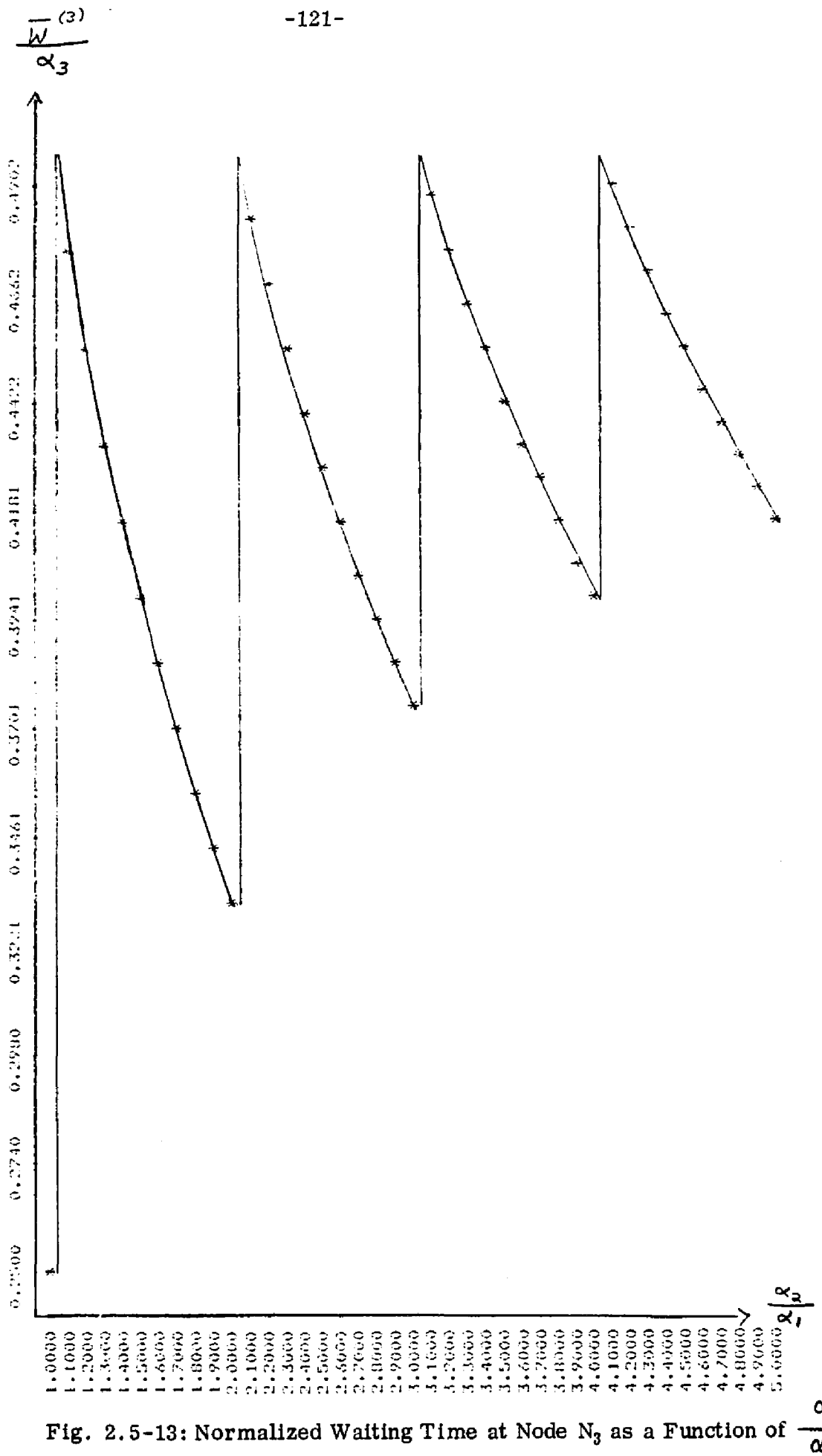


Fig. 2.5-13: Normalized Waiting Time at Node N_3 as a Function of $\frac{\alpha_2}{\alpha_1}$

3.1. NETWORKS WITH PARALLEL SERVERS

In this section we turn our attention to networks where each node of the network is assumed to have a given number of parallel channels over which it can service arriving customers. As examples of networks with parallel channels we have, the take-off and landing operation at an airport with several runways, the multiple channels that may connect various points in the telephone network, and the parallel channels that exist between nodes in some computer networks for reliability purposes. We shall commence our discussion with a review of results obtained by Friedman [9] for tandem networks with parallel servers and after this, we shall proceed to demonstrate how many of the results derived in the previous chapter, for networks of single server queues, may be generalized and extended to networks of multiple server queues.

We begin our review of Friedman's results by considering the network of Fig. 3.1-1. The network shown is a tandem queueing system consisting of n stages, N_1, N_2, \dots, N_n , in series where stage N_i is assumed to have m_i parallel servers each having the same constant service time α_i sec. Once again, we shall assume that each customer arriving into the networks consists of a single fixed length packet. The packets are assumed to arrive at the first stage or node

according to any general probability distribution and proceed through the stages, in order, on a first-come first-served basis. At each node, the channel over which each packet is processed is assumed to be selected in a cyclic manner. Finally, we assume that each stage has unlimited storage available.

At this point we introduce some notation as used by Friedman. We write $N_i = (\alpha_i : m_i)$ to indicate that node N_i consists of m_i parallel channels each with the same constant service time of α_i . Thus, the network of Fig. 3.1-1 may be written as

$$\rightarrow (\alpha_1 : m_1) \rightarrow (\alpha_2 : m_2) \rightarrow (\alpha_3 : m_3) \rightarrow \cdots \rightarrow (\alpha_n : m_n) \rightarrow . \quad (3.1-1)$$

We now state Friedman's first result.

Result 1:

Given any input sequence of packets to any system given by (3.1-1), the output sequence from the last stage is independent of the order of the stages.

The implication inherent in Result 1 is that a packet entering the network at a given time, will depart from the final stage of the network at the exact same time regardless of the permutation of the various stages. Thus, the amount of waiting time experienced by a packet passing through the tandem network is independent of the order of the stages.

We now introduce Friedman's principle of dominance. We say that stage N_i dominates stage N_j if, in the system $\rightarrow N_i \rightarrow N_j \rightarrow$, no packet ever waits at N_j regardless of the input sequence to N_i . The statement N_i dominates N_j is a property of nodes N_i and N_j and is independent of system context; for example, if N_i dominates N_j , then in the system $\rightarrow N_k \rightarrow N_j \rightarrow N_l \rightarrow N_i \rightarrow \dots$ we may still say N_i dominates N_j . We now state the remaining three results of Friedman's paper.

Result 2:

$N_i = (\alpha_i : m_i)$ dominates $N_j = (\alpha_j : m_j)$ if and only if $\alpha_j \leq M\alpha_i$ where $M = \lceil \frac{m_j}{m_i} \rceil$ and $\lceil x \rceil$ is the greatest integer less than or equal to x .

Result 3:

Dominance is transitive; that is, if N_i dominates N_j and N_j dominates N_k , then N_i dominates N_k .

Result 4:

Dominance is persistent; that is, if N_i dominates N_j and N_i and N_j are nonadjacent stages in a system $\dots \rightarrow N_i \rightarrow \dots \rightarrow N_r \rightarrow N_j \rightarrow \dots$, then no waiting time ever occurs at N_j , regardless of the input sequence to the system.

Result 4 implies that for a tandem network of nodes with parallel servers, the waiting time experienced by a packet is completely determined by the dominant nodes of the network

and that the waiting time experienced by a packet may be found by considering a reduced network formed from the original system by removing all nodes that are dominated by at least one other node of the network. As an example, the system

$$\rightarrow (5:2) \rightarrow (4:2) \rightarrow (2:1) \rightarrow (3:1) \rightarrow (7:3) \rightarrow \quad (3.1-2a)$$

can be reduced to a network consisting of the single stage

$$\rightarrow (3:1) \rightarrow \quad (3.1-2b)$$

without changing the packet waiting time since (3:1) dominates all the other stages as seen by application of Result 2. As a second example, the network

$$\rightarrow (3:1) \rightarrow (2:1) \rightarrow (7:2) \rightarrow (13:4) \rightarrow (5:4) \quad (3.1-3a)$$

may be reduced to

$$\rightarrow (3:1) \rightarrow (7:2) \rightarrow , \quad (3.1-3b)$$

but no further since (3:1) dominates (2:1), and (7:2) dominates both (13:4) and (5:4); however, neither (3:1) nor (7:2) dominate each other.

Clearly, the results of Friedman may be interpreted as a generalization of the results of Rubin [17-19], as presented above in Chapter 2, for tandem networks of single server queues, to tandem networks of multiple server queues. However, as in the case of single server systems, no exact results have been obtained by Friedman for non-tandem networks. It is for these non-tandem networks that we shall show that many of the theorems, derived in the previous chapter for single server systems, can be generalized and extended to multiple server systems under certain conditions.

We now turn our attention to Figs. 3.1-2 and 3.1-3. These networks are assumed to have all the properties as the networks of Figs. 2.2-1 and 2.2-2 respectively, which were described at the beginning of Section 2.2, except, that in the present case, each node of each network is assumed to have two parallel servers each with a constant packet service time of α sec.

As time passes, each server of each queue shown will alternate between busy and idle periods. As such, we shall derive the following theorem:

Theorem 3.1-1:

The number of packets serviced by the i^{th} server, $i = 1, 2$, at node N of Fig. 3.1-2 during its j^{th} busy period equals the number of packets serviced by the i^{th} server at node N_3 of Fig. 3.1-3 during its j^{th} busy

period, $j = 1, 2, 3, \dots$, where the i^{th} server of a given queue is defined as that server that processes the i^{th} packet, $i = 1, 2$, to arrive at the particular queue under consideration.

Proof:

With all quantities defined in Section 2.2, for Figs. 2.2-1 and 2.2-2, now referring to the networks of Figs. 3.1-2 and 3.1-3 respectively, we once again begin our proof by looking at the very first busy period of server 1 at node N of Fig. 3.1-2. We observe that server 1 will process all odd numbered packets to arrive at node N and server 2 all the even numbered ones. This is clearly true because of the cyclic nature of choosing a server at a given node. Thus, for server 1 to have a busy period of M_1 packets, the arrival time of the odd numbered packets must be

$$\begin{aligned} t_1 &= t_1 && (3.1-4a) \\ t_2 &\leq t_3 \leq t_1 + \alpha \\ t_4 &\leq t_5 \leq t_1 + 2\alpha \\ &\vdots \\ t_{2M_1-2} &\leq t_{2M_1-1} \leq t_1 + (M_1-1)\alpha \\ t_{2M_1+1} &> t_1 + M_1\alpha \end{aligned}$$

As for the even numbered packets, we have

$$\begin{aligned}
 t_1 &\leq t_2 \leq t_1 + \alpha && (3.1-4b) \\
 t_3 &\leq t_4 \leq t_1 + 2\alpha \\
 &\vdots \\
 t_{2M_1-3} &\leq t_{2M_1-2} \leq t_1 + (M_1-1)\alpha
 \end{aligned}$$

Using the external arrival times as defined by (3.1-4), we now turn to Fig. 3.1-3 and find

$$\begin{aligned}
 t'_1 &= t_1 + \alpha && (3.1-5) \\
 t_1 + \alpha &\leq t'_2 \leq t_1 + 2\alpha \\
 t_2 + \alpha &\leq t'_3 \leq t_1 + 2\alpha \\
 t_3 + \alpha &\leq t'_4 \leq t_1 + 3\alpha \\
 t_4 + \alpha &\leq t'_5 \leq t_1 + 3\alpha \\
 &\vdots \\
 t_{2M_1-3} + \alpha &\leq t'_{2M_1-2} \leq t_1 + M_1\alpha \\
 t_{2M_1-2} + \alpha &\leq t'_{2M_1-1} \leq t_1 + M_1\alpha
 \end{aligned}$$

and

$$t'_{2M_1+1} > t_1 + (M_1+1)\alpha$$

At node N_3 , once again the odd numbered packets to arrive at N_3 will be processed by server 1 at node N_3 . Thus, using (3.1-5), we can calculate that

$$\begin{aligned}
 T'_1 &= t_1 + \alpha & (3.1-6) \\
 T'_3 &\leq t_1 + 2\alpha \\
 T'_5 &\leq t_1 + 3\alpha \\
 &\vdots \\
 T'_{2M_1-1} &\leq t_1 + M_1\alpha \\
 T'_{2M_1+1} &> t_1 + (M_1+1)\alpha
 \end{aligned}$$

Furthermore, using (3.1-6), we find that packets departing from server 1 at N_3 do so at times

$$\begin{aligned}
 D'_1 &= t_1 + 2\alpha & (3.1-7) \\
 D'_3 &= t_1 + 3\alpha \\
 &\vdots \\
 D'_{2M_1-1} &= t_1 + (M_1+1)\alpha
 \end{aligned}$$

Thus, from (3.1-6) and (3.1-7), we can conclude that

$$\begin{aligned}
 T'_{2i+1} &\leq D'_{2i-1} & i = 1, 2, \dots, M_1-1 & (3.1-8) \\
 T'_{2M_1+1} &> D'_{2M_1-1}
 \end{aligned}$$

which implies that we have had a busy period of exactly M_1 packets at node N_3 . Furthermore, though specifically proven for server 1 only, it is clear that a simple derivation can be constructed for server 2 of nodes N and N_3 . In addition,

by simple induction, these results extend to all subsequent busy periods at nodes N and N_3 ; thus, this proves our theorem. Q.E.D.

Consequently, we have shown that, busy period by busy period, the i^{th} server of nodes N and N_3 will process exactly the same number of packets. Using this fact, we can, as was done in Section 2.2, derive the following three corollaries:

Corollary 3.1-1:

The probability distributions of the number of packets serviced during a busy period of the i^{th} server, $i = 1, 2$, at nodes N and N_3 are identical.

Corollary 3.1-2:

The duration of the j^{th} idle period of the i^{th} server of node N equals the duration of the j^{th} idle period of the i^{th} server of node N_3 , $i = 1, 2, j = 1, 2, \dots$.

Corollary 3.1-3:

The probability distributions of the duration of the j^{th} idle period of the i^{th} server at nodes N and N_3 are identical.

At this point, we would desire to proceed and state a theorem concerning the steady-state mean waiting time experienced by a packet in the two networks considered. However, before doing this, we must first define the concept of

a combined busy period over all servers attached to a given node.

To describe our definition of a combined busy period, let us assume that all servers of a given node are initially busy. As such, the combined busy period is said to continue until such time as one of the servers becomes idle for a finite amount of time greater than zero. This signals the conclusion of the present busy period and all packets currently in the other servers are considered belonging to the old busy period. We note that, because of the cyclic nature of server assignment, there can never be any packets on queue when a busy period is said to terminate. In addition, we observe that we say the combined busy period has ended though the other servers may still be processing packets. Having become idle in the combined sense, the node in question awaits the arrival of a new packet. On arrival, this new packet will, of course, be assigned to that server whose becoming idle terminated the previous busy period. This assignment will signal the beginning of a new combined busy period. Having commenced a new busy period, we now turn to the next server in the assignment cycle. If this server is presently busy, we shall say that for a new packet to arrive and belong to the newly started combined busy period, it must do so before the server becomes idle; that is, before the server finishes processing the packet that belonged to a previous busy period. On the

other hand, if we turn to the next server in the cycle and find it idle, then, for a packet to belong to the newly started busy period, it may arrive so long as the packet which initiated the newly started combined busy period is still in service.

Thus, having defined a combined busy period, we state the following theorem:

Theorem 3.1-2:

The number of packets serviced at node N of Fig. 3.1-2 during its j^{th} combined busy period equals the number of packets serviced at node N_3 of Fig. 3.1-3 during its j^{th} combined busy period, $j = 1, 2, \dots$.

Proof:

The proof of this theorem will, once again, be similar to that of theorem 2.2-1; therefore, we shall go through the steps of the proof rather quickly.

Once again, we begin by looking at the very first combined busy period at node N of Fig. 3.1-2. For a combined busy period of length M_1 packets, the packet arrival times must be

$$\begin{aligned} t_1 &= t_1 && (3.1-9) \\ t_1 &\leq t_2 \leq t_1 + \alpha \\ t_2 &\leq t_3 \leq t_1 + \alpha \\ t_3 &\leq t_4 \leq t_2 + \alpha \\ t_4 &\leq t_5 \leq t_1 + 2\alpha \end{aligned}$$

$$\begin{aligned}
 t_5 &\leq t_6 \leq t_2 + 2\alpha \\
 &\vdots \\
 t_{M_1-3} &\leq t_{M_1-2} \leq t_i + \left(\frac{M_1-2-i}{2}\right)\alpha \\
 t_{M_1-2} &\leq t_{M_1-1} \leq t_{3-i} + \left(\frac{M_1-1-(3-i)}{2}\right)\alpha = t_{3-i} + \left(\frac{M_1-4+i}{2}\right)\alpha \\
 t_{M_1-1} &\leq t_{M_1} \leq t_i + \left(\frac{M_1-i}{2}\right)\alpha \\
 & \\
 t_{M_1+1} &> t_{3-i} + \left(\frac{M_1-2+i}{2}\right)\alpha
 \end{aligned}$$

where

$$i = \begin{cases} 1 & \text{if } M_1 \text{ is odd} \\ 2 & \text{if } M_1 \text{ is even.} \end{cases}$$

With external arrival times as given by (3.1-9), we now turn to Fig. 3.1-3 and calculate

$$\begin{aligned}
 t'_1 &= t_1 + \alpha && (3.1-10) \\
 t_1 + \alpha &\leq t'_2 = t_2 + \alpha \leq t_1 + 2\alpha \\
 t_2 + \alpha &\leq t'_3 \leq t_1 + 2\alpha \\
 t_3 + \alpha &\leq t'_4 \leq t_2 + 2\alpha \leq t_1 + 3\alpha \\
 &\vdots \\
 t_{M_1-3} + \alpha &\leq t'_{M_1-2} \leq t_i + \left(\frac{M_1-i}{2}\right)\alpha \\
 t_{M_1-2} + \alpha &\leq t'_{M_1-1} \leq t_{3-i} + \left(\frac{M_1-2+i}{2}\right)\alpha \\
 t_{M_1-1} + \alpha &\leq t'_{M_1} \leq t_i + \left(\frac{M_1-i+2}{2}\right)\alpha \\
 & \\
 t'_{M_1+1} &> t_{3-i} + \left(\frac{M_1+i}{2}\right)
 \end{aligned}$$

Using (3.1-10), we then find the ordinal packet arrival times at N_3 to be

$$\begin{aligned}
 T'_1 &= t_1 + \alpha & (3.1-11) \\
 T'_2 &= t_2 + \alpha \leq t_1 + 2\alpha \\
 T'_3 &\leq t_1 + 2\alpha \\
 T'_4 &\leq t_2 + 2\alpha \leq t_1 + 3\alpha \\
 &\vdots \\
 T'_{M_1-2} &\leq t_i + \left(\frac{M_1-i}{2}\right)\alpha \\
 T'_{M_1-1} &\leq t_{3-i} + \left(\frac{M_1-2+i}{2}\right)\alpha \\
 T'_{M_1} &\leq t_i + \left(\frac{M_1-i+2}{2}\right)\alpha \\
 T'_{M_1+1} &> t_{3-i} + \left(\frac{M_1+i}{2}\right)\alpha
 \end{aligned}$$

As for the departure times from N_3 , we have

$$\begin{aligned}
 D'_1 &= t_1 + 2\alpha & (3.1-12) \\
 D'_2 &= t_2 + 2\alpha \leq t_1 + 3\alpha \\
 D'_3 &= t_1 + 3\alpha \\
 D'_4 &= t_2 + 3\alpha \leq t_1 + 4\alpha \\
 &\vdots \\
 D'_{M_1-2} &= t_i + \left(\frac{M_1-i+2}{2}\right)\alpha
 \end{aligned}$$

$$D'_{M_1-1} = t_{3-i} + \left(\frac{M_1+i}{2}\right)\alpha$$

$$D'_{M_1} = t_i + \left(\frac{M_1-i+4}{2}\right)\alpha$$

Thus, from (3.1-11) and (3.1-12) we see that

$$T'_{i+2} \leq D'_i \quad i = 1, 2, \dots, M_1-2 \quad (3.1-13)$$

$$T'_{M_1+1} > D'_{M_1-1}$$

Consequently, we have shown that the very first busy period, at both nodes N and N_3 , will have a length of exactly M_1 packets. Furthermore, using the departure times from N and (3.1-12) for N_3 , we can calculate that when the busy period ends in server i , the remaining service time on the packet in server $3-i$ is

$$t_i - t_{3-i} + (2-i)\alpha \leq \alpha \quad (3.1-14)$$

(see, for example, $D'_{M_1} - D'_{M_1-1}$).

To complete the proof, let us look at the second busy period at node N . This, second busy period, which will have a length of M_2 packets, begins with the arrival of a packet at node N at time t_{M_1+1} . In a straightforward manner, we may show, using (3.1-9), that the departure time of packet M_1 from node N occurs at time

$$D_{M_1} = t_i + \left(\frac{M_1 - i + 2}{2} \right) \alpha$$

Thus, we have the following two cases:

Case 1: $t_{M_1+1} > D_{M_1}$

For this case, we must have at N_3

$$t'_{M_1+1} = T'_{M_1+1} > D_{M_1} + \alpha = t_i + \left(\frac{M_1 - i + 4}{2} \right) \alpha > D'_{M_1} \quad (3.1-16)$$

Consequently, it is obvious that the second busy period for both nodes N and N_3 have the same general form as the first and the proof just completed for the first busy period is clearly applicable to the second.

Case 2: $t_{M_1+1} \leq D_{M_1}$

In this case, the arrival of packet M_1+1 at time t_{M_1+1} finds packet M_1 still in service. Thus, for the busy period to continue, packet M_1+2 must arrive before packet M_1 departs. Consequently, the arrival times at node N of the packets of the second busy period must occur at

$$t_{M_1+1} = t_{M_1+1} \quad (3.1-17)$$

$$t_{M_1+1} \leq t_{M_1+2} \leq D_{M_1} = t_i + \left(\frac{M_1 - i + 2}{2} \right) \alpha \leq t_{M_1+1} + \alpha$$

$$t_{M_1+2} \leq t_{M_1+3} \leq t_{M_1+1} + \alpha$$

$$\begin{aligned}
 t_{M_1+3} &\leq t_{M_1+4} \leq D_{M_1} + \alpha \leq t_{M_1+1} + 2\alpha \\
 &\vdots \\
 t_{M_1+M_2-3} &\leq t_{M_1+M_2-2} \leq t_{M_1+k} + \left(\frac{M_2-2-k}{2}\right)\alpha \\
 t_{M_1+M_2-2} &\leq t_{M_1+M_2-1} \leq t_{M_1+3-k} + \left(\frac{M_2-4+k}{2}\right)\alpha \\
 t_{M_1+M_2-1} &\leq t_{M_1+M_2} \leq t_{M_1+k} + \left(\frac{M_2-2+k}{2}\right)\alpha \\
 t_{M_1+M_2+1} &> t_{M_1+3-k} + \left(\frac{M_2-2+k}{2}\right)\alpha
 \end{aligned}$$

where

$$k = \begin{cases} 1 & \text{if } M_2 \text{ is odd} \\ 2 & \text{if } M_2 \text{ is even.} \end{cases}$$

At this point we can turn to node N_3 and construct equations similar to those of (3.1-10)-(3.1-13), thus, proving the equality between the length of the second combined busy periods at node N and at node N_3 . The actual process is left as a simple exercise.

Thus, we have proven that, in all cases, the lengths of the second combined busy periods at nodes N and N_3 are equal.

We are now in a position to state that by simple induction, this result extends to all subsequent busy periods, thus, proving our theorem. Q.E.D.

At this point, we can once again state three corollaries similar to those following theorem 3.1-1. Clearly, the duration of the j^{th} combined idle periods at node N and N_3 are equal. Furthermore, the probability distribution associated with the length of j^{th} combined busy period and j^{th} combined idle period at node N are identical to those associated with node N_3 respectively. However, instead of formally deriving these corollaries, we leave them to the interested reader and, instead, proceed to another theorem of more importance.

Theorem 3.1-3:

For the j^{th} combined busy period of nodes N and N_3 , the total waiting time experienced by the M_j packets in the network of Fig. 3.1-2 equals the total waiting time experienced by the $M_j^{(3)} = M_j$ packets over the entire network of Fig. 3.1-3.

Proof:

Once again, we adopt the notation of Chapter 2. With all arrival times now referring to the j^{th} combined busy period, we obtain relations for the waiting time on a packet of the j^{th} combined busy period of node N as

$$\begin{aligned}
 W_1 &= 0 && (3.1-18) \\
 W_2 &= [D_L - t_2]^+ \\
 W_{k+2} &= [W_k + \alpha - (t_{k+2} - t_k)]^+ && k = 1, 2, \dots
 \end{aligned}$$

where D_L is the departure time from node N of the last packet of the $(j-1)^{st}$ combined busy period. Expanding (3.1-18), we have

$$\begin{aligned}
 W_1 &= 0 & (3.1-19) \\
 W_2 &= [D_L - t_2]^+ \\
 W_3 &= t_1 - t_3 + \alpha \\
 W_4 &= W_2 + t_2 - t_4 + \alpha \\
 W_5 &= t_1 - t_5 + 2\alpha \\
 W_6 &= W_2 + t_2 - t_6 + 2\alpha \\
 &\vdots \\
 W_{M_{j-1}} &= (2-i)W_2 + t_{3-i} - t_{M_{j-1}} + \left(\frac{M_j - 4 + i}{2}\right)\alpha \\
 W_{M_j} &= (i-1)W_2 + t_i - t_{M_j} + \left(\frac{M_j - i}{2}\right)\alpha
 \end{aligned}$$

where

$$i = \begin{cases} 1 & \text{if } M_j \text{ is odd} \\ 2 & \text{if } M_j \text{ is even.} \end{cases}$$

Summing over all M_j packets, we find

$$\begin{aligned}
 W_{TOTAL} &= \sum_{k=1}^{M_j} W_k = \sum_{k=1}^{\frac{M_j+2-i}{2}} [t_1 - t_{2k-1} + (k-1)\alpha] + & (3.1-20) \\
 &+ \sum_{k=1}^{\frac{M_j+i-2}{2}} [W_2 + t_2 - t_{2k} + (k-1)\alpha] =
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{M_j+2-i}{2} t_1 + \frac{M_j+i-2}{2} t_2 - \sum_{k=1}^{M_j} t_k + \\
 &\quad + \left[\frac{(M_j-i)(M_j-i-2)}{4} + \frac{i(M_j-i)}{2} \right] \alpha = \\
 &= \frac{M_j+2-i}{2} t_1 + \frac{M_j+i-2}{2} t_2 - \sum_{k=1}^{M_j} t_k + \left[\frac{(M_j-i)(M_j+i-2)}{4} \right] \alpha.
 \end{aligned}$$

We will now prove our theorem by mathematical induction.

For the case $M_j = M_j^{(3)} = 1$, it is clear that the total waiting time, in both networks, is identically zero. We now assume that (3.1-20) is true for both networks for $M_j = M_j^{(3)} \leq m$. We proceed to show that, for $M_j = M_j^{(3)} = m + 1$, the total waiting in each of the two networks will still be equal.

At node N, the extra waiting time introduced by having M_j equal $m+1$ instead of m , is seen, from (3.1-15), to be

$$\begin{aligned}
 (W_{\text{EXTRA}})^{\text{Fig 3.1-2}} &= W_{m+1} = & (3.1-21) \\
 &= (2-i)W_2 + t_{3-i} - t_{m+1} + \left(\frac{m-2+i}{2}\right)\alpha.
 \end{aligned}$$

Turning to Fig. 3.1-3, we, once again, arbitrarily assume that this $(m+1)^{\text{st}}$ packet, arriving at t_{m+1} , enters the network at node N_1 . This packet will arrive at node N_3 at time

$$t'_{m+1} = t_{m+1} + W_k^{(1)} + \alpha \quad (3.1-22)$$

where we have assumed this packet to be the k^{th} , $1 \leq k \leq m+1$, to arrive at N_1 . At node N_3 , we can write, similar to (3.1-18), that

$$\begin{aligned} W_1^{(3)} &= 0 & (3.1-23) \\ W_2^{(3)} &= [D'_L - T'_2]^+ \\ W_{k+2}^{(3)} &= [W_k^{(3)} + \alpha - (T'_{k+2} - T'_k)]^+ \quad k = 1, 2, \dots \end{aligned}$$

where D'_L is the departure time from N_3 of the last packet of the previous busy period. Thus, for a busy period of $(m+1)$ packets, we have

$$\begin{aligned} W_1^{(3)} &= 0 & (3.1-24) \\ W_2^{(3)} &= [D'_L - T'_2]^+ \\ W_3^{(3)} &= T'_1 + T'_3 + \alpha \\ W_4^{(3)} &= W_2^{(3)} + T'_2 - T'_3 + \alpha \\ &\vdots \\ W_m^{(3)} &= (i-1)W_2^{(3)} + T'_i - T'_m + \left(\frac{m-i}{2}\right)\alpha \\ W_{m+1}^{(3)} &= (2-i)W_2^{(3)} + T'_{3-i} - T'_{m+1} + \left(\frac{m-2+i}{2}\right)\alpha \end{aligned}$$

where

$$i = \begin{cases} 1 & \text{if } m \text{ is odd} \\ 2 & \text{if } m \text{ is even.} \end{cases}$$

In general, this $(m+1)^{st}$ packet will be the n^{th} packet, of the j^{th} combined busy period, to arrive at node N_3 , $1 \leq n \leq m+1$; that is, $t'_{m+1} = T'_n$. Furthermore, from (3.1-24), we see that the extra waiting time, at N_3 , caused by having $M_j^{(3)}$ equal to $m+1$ instead of m , is given by

$$W_{EXTRA}^{(3)} = (2-i)W_2^{(3)} + T'_{3-i} - T'_n + \left(\frac{m-2+i}{2}\right)\alpha \quad (3.1-25)$$

We now substitute for T'_n , using (3.1-22), and T'_{3-i} , using the relation $T'_{3-i} = t_{3-i} + \alpha$, and find

$$W_{EXTRA}^{(3)} = (2-i)W_2^{(3)} + t_{3-i} - t_{m+1} - W_k^{(1)} + \left(\frac{m-2+i}{2}\right)\alpha \quad (3.1-26)$$

Consequently,

$$\begin{aligned} (W_{EXTRA})_{Fig\ 3.1-3} &= W_{EXTRA}^{(1)} + W_{EXTRA}^{(3)} & (3.1-27) \\ &= W_k^{(1)} + W_{EXTRA}^{(2)} \\ &= (2-i)W_2^{(3)} + t_{3-i} - t_{m+1} + \left(\frac{m-2+i}{2}\right)\alpha \\ &= (W_{EXTRA})_{Fig.\ 3.1-2} \end{aligned}$$

where we have used the fact that $W_2 = W_2^{(3)}$ which, in itself, can be proven from the discussion leading to equation (3.1-14).

This completes our induction procedure and proves our theorem. Q.E.D.

Having proven theorem 3.1-3, it is clear that the probability distributions of the total waiting time experienced by the $M_j = M_j^{(3)}$ packets of the j^{th} combined busy period is identical for the networks of Figs. 3.1-2 and 3.1-3. Of more importance, theorems 3.1-2 and 3.1-3 directly imply the following corollary:

Corollary 3.1-4:

For the networks of Figs. 3.1-2 and 3.1-3, the steady-state mean overall waiting time experienced by an arbitrary class packet is the same in both networks; that is,

$$(\bar{S})_{\text{Fig. 3.1-2}} = (\bar{S})_{\text{Fig. 3.1-3}} \quad (3.1-28)$$

Clearly, corollary 3.1-4 implies the same principle in delay decomposition, with respect to self delay and interference delay, as discussed previously for networks of single server queues. Furthermore, we once again obtain the result that the steady-state mean waiting time on an arbitrary class packet at node N_3 of Fig. 3.1-3 is given by

$$\begin{aligned} \bar{W}^{(3)} &= (\bar{S})_{\text{Fig. 2.1-3}} - \bar{W}^{(1)} - \bar{W}^{(2)} & (3.1-29) \\ &= (\bar{S})_{\text{Fig. 3.1-2}} - \bar{W}_{(1)}^{(1)} P\{\text{a packet is class 1}\} - \\ &\quad - \bar{W}_{(2)}^{(2)} P\{\text{a packet is class 2}\}. \end{aligned}$$

Thus, we have shown that, the theorems and corollaries developed previously in section 2.2 are applicable and extendable to networks where each node has two parallel servers. In fact, in a manner directly parallel to that just used, we can show that the theorems and corollaries just developed for the networks of Figs. 3.1-2 and 3.1-3, networks where each node has only two parallel servers, are extendable and applicable to networks where each node of the network has k parallel servers, as depicted by Figs. 3.1-4 and 3.1-5, where each server has a packet service time of α sec. The details of the actual proof are left for the interested reader. At this time, we merely mention that, with the cyclic assignment policy used, each server i , of a given node, will process packets numbered $i + nk$, $i = 1, 2, \dots, k$, $n = 1, 2, \dots$, where we refer to the ordinal arrival of packets to a given node. As a result, if one desired to construct an initial combined busy period of M_1 packets at node N of Fig. 3.1-4, (see eq. (3.1-9)), the arrival time of packets to node N must follow

$$\begin{aligned} t_1 &= t_1 && (3.1-30) \\ t_1 &\leq t_2 \leq t_1 + \alpha \\ t_2 &\leq t_3 \leq t_1 + \alpha \\ &\vdots \\ t_{k-1} &\leq t_k \leq t_1 + \alpha \end{aligned}$$

$$\begin{aligned}
 t_k &\leq t_{k+1} \leq t_1 + \alpha \\
 t_{k+1} &\leq t_{k+2} \leq t_2 + \alpha \\
 &\vdots \\
 t_{2k-1} &\leq t_{2k} \leq t_k + \alpha \\
 t_{2k} &\leq t_{2k+1} \leq t_1 + 2\alpha \\
 t_{2k+1} &\leq t_{2k+2} \leq t_2 + 2\alpha \\
 &\vdots \\
 t_{M_1-3} &\leq t_{M_1-2} \leq t_{i-2+k(\delta(i-1)+\delta(i-2))} + \\
 &\quad + \frac{M_1-2-(i-2+k(\delta(i-1)+\delta(i-2)))}{k} \alpha \\
 t_{M_1-2} &\leq t_{M_1-1} \leq t_{i-1+k(\delta(i-1))} + \frac{M_1-1-(i-1+k\delta(i-1))}{k} \alpha \\
 t_{M_1-1} &\leq t_{M_1} \leq t_i + \frac{M_1-i}{k} \alpha \\
 t_{M_1+1} &> t_{i+1-k\delta(i-k)} + \frac{M_1+1-(i+1-k\delta(i-k))}{k} \alpha
 \end{aligned}$$

where $M_1 = nk+i$, $1 \leq i \leq k$, n an integer and $\delta(x-x_0)$ is the dirac delta function; i.e.,

$$\delta(x-x_0) = \begin{cases} 1, & x = x_0 \\ 0, & x \neq x_0 \end{cases} \quad (3.1-31)$$

As stated before, the remaining steps in the required derivations are left to the interested reader.

Having shown the equivalency, with respect to steady-state overall mean waiting time, that exists between the

networks of Figs. 3.1-4 and 3.1-5, we can proceed to the five networks of Fig. 3.1-6, and, using this equivalency together with the results of Friedman, show that, if stage $N = \alpha_n:k_n$, the merger node in each network, dominates all other stages in each of the networks, all five networks will have the same steady-state overall mean waiting time on a packet. Once again, as done for the networks of Fig. 2.2-3, we would begin at Fig. 3.1-6e and use our decomposition principle at N to transform the network into the network of Fig. 3.1-6d. Then using the results of Friedman, Fig. 3.1-6d can be transformed into the Fig. 3.1-6c. Having attained Fig. 3.1-6c, we note that Figs. 3.1-6b and 3.1-6c are exactly the networks of Figs. 3.1-4 and 3.1-5 respectively. Finally, Fig. 3.1-6a is obtained from Fig. 3.1-6b by finding the joint arrival process of the two independent packet streams of Fig. 3.1-6b. Thus, we have proven the desired equivalency in steady-state waiting time that exists among these five networks.

In summary, we state that we have formulated a technique which, given a network consisting of stages $N_i = (\alpha_i:k_i)$, such as, the network of Fig. 3.1-7, the steady-state overall mean waiting time experienced by an arbitrary class packet passing through a merger node, may be determined exactly, provided that the merger node dominates all stages of the

network that are feeding packets into the merger node. Whether the waiting time experienced by an arbitrary class packet at the merger node itself may be calculated would depend on the existence of a dominant stage in each path entering the merger node; that is, if for each path entering the merger node there exists a stage that dominates all other stages in the given path and, thus, solely determines the overall packet waiting time for the given path. If such stages exist for each path, an exact determination of the steady-state mean waiting time at the merger node is possible, otherwise, if such dominant stages do not exist, an exact determination is not yet possible. Thus, in Fig. 3.1-7, the exact waiting on a packet at node N_3 may be found exactly provided N_3 dominates both N_1 and N_2 . Furthermore, though shown as simple nodes, N_1 and N_2 may each represent entire networks of nodes in which each, N_1 and N_2 , were dominant.

Clearly, many problems remain to be answered. In the case of a merger node that is not dominant over all other stages feeding the merger node, we have as yet been unable to obtain any satisfactory results. In the previous chapter, sec. 2.5, for networks in which each node had only a single server, we were able to obtain some exact and some tight approximations in the case of fast, non-dominating, merger nodes. However, we have as yet been unsuccessful in

extending the techniques developed there, or finding new methods to apply to networks with parallel servers and non-dominant merger nodes such as Fig. 3.1-7 when N_3 does not dominate both N_1 and N_2 . The problem of non-dominant merger nodes is important and remains to be solved.

3.2. DELAY ANALYSIS AT A SEPARATION NODE

Up to this point, we have been limiting our discussion of networks to those whose topological complexity did not go beyond merger nodes. In other words, in the networks considered previously, if two independent packet streams entered a common node for processing, the packets from these streams would be merged in the sense that all packets from both streams would have an identical route of nodes to follow from this point onward. From the point of practicality these feed-forward types of networks have their applications in such areas as centralized computer networks in which all users of the network are attempting to access some centralized processing facility. Indeed, in this case we have a feed-forward type of network in which streams of packets from different users, once merged, continue along the same route of nodes to the centralized processing facility. However, if we investigate the return route; that is, the route that a packet must follow from the central node back to an individual user, we would find that instead of passing through a series of merger nodes as in the forward trip, the packet would instead pass through a series of separation nodes, nodes at which the common stream of packets leaving the central node is decomposed or separated into a number of different streams each of which will then proceed to follow an individual route so as to reach its

final destination. Thus, we shall now turn our attention to delay analysis at separation nodes.

In Fig. 3.2-1 we have depicted a two node queueing network. At node N_1 , two external independent packet streams enter the network and are merged and processed by a single server channel, at rate α_1 sec./packet, over which the merged packet stream travels to node N_2 . Upon arrival at node N_2 , the merged packet stream is decomposed such that all the class 1 packets will be processed by a channel with a constant service time of α_2 sec./packet while the class 2 packets will all be processed by some other channel. Thus, node N_2 is a separation node since at it the merged packet stream is separated into its class 1 and class 2 components. Our interest, then, lies in calculating the delay experienced by a class 1 packet at node N_2 .

Clearly, when $\alpha_2 \leq \alpha_1$, the waiting time experienced by a class 1 packet at node N_2 is zero. Thus, our interest lies in the analysis of the case when $\alpha_2 > \alpha_1$.

For $\alpha_2 > \alpha_1$, we have as yet been unable to obtain an exact method for analysis. For the case of Poisson inputs, a fairly tight upper bound approximation has been obtained by Rubin [20] by simply assuming $\lambda_2 = 0$. In doing this, the delay experienced by a class 1 packet at N_2 may then be calculated by considering the tandem network of nodes

N_1 and N_2 being fed by a single external input stream of class 1 packets. In doing this we find that for $\alpha_2 > \alpha_1$,

$$\bar{w}_{(1)}^{(2)} \leq \frac{\rho_{12}\alpha_2}{2(1-\rho_{12})} - \frac{\rho_{11}\alpha_1}{2(1-\rho_{11})} \quad (3.2-1)$$

An interesting lower bound approximation may be obtained by considering the network characteristics as λ_2 approaches $(1-\rho_{11})/\alpha_1$ (that is, $\rho_{11} + \rho_{21} \rightarrow 1$ with constant λ_1 and increasing λ_2). As λ_2 approaches $(1-\rho_{11})/\alpha_1$, node N_1 , of course, becomes unstable; that is, the steady-state queue length and waiting time become infinite. However, of importance to us is the fact that the departure stream of packets from N_1 tends to become deterministic in nature; that is, with the queue length tending to infinity, there will be a departure of an arbitrary class packet from N_1 every α_1 sec. If we further assume that the class 1 and class 2 input streams are Poisson in character, then the probability that a given departure from N_1 is of class i is clearly given by

$$P\{\text{a departure is class } i\} = \frac{\lambda_i}{\lambda_1 + \lambda_2} \quad i = 1, 2 \quad (3.2-2)$$

This being true because of the memoryless property of the Poisson process. Turning to node N_2 , we note that node N_2

is being fed by a discrete time arrival process where there is an arrival of either a class 1 or a class 2 packet in each time slot, where a time slot consists of α_1 sec.. If we further view the arrival of a class 1 packet in a given slot as a success and the arrival of a class 2 packet as a failure, we note the arrivals of packets at N_2 form a Bernouli sequence with

$$\begin{aligned} P\{\text{success}\} &= P\{\text{a class 1 packet arrives}\} = & (3.2-3) \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \end{aligned}$$

If we now concern ourselves with the number of class 1 packets that arrive at N_2 in a given number of slots, we find that this process will have a binomial probability

distribution with parameter $\frac{\lambda_1}{\lambda_1 + \lambda_2}$. Thus,

$$\begin{aligned} P\{k \text{ class 1 packets arriving in } m \text{ slots}\} &= & (3.2-4) \\ &= \binom{m}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{m-k} \end{aligned}$$

where $\binom{m}{k} \triangleq$ binomial coefficients $= \frac{m!}{k!(m-k)!}$. The binomial is, of course, the discrete time counterpart of the continuous time Poisson process, enjoying the same memoryless properties. To complete the approximation

procedure, we note that if α_2 is an integer multiple of α_1 ; that is $\alpha_2 = k\alpha_1$, $k = 1, 2, \dots$, then node N_2 forms a discrete time queueing system whose input follows a binomial distribution. Discrete time networks, as this, were studied by Meisling [15] and he derived exact results for the steady-state queue length and mean waiting time at such a discrete time queue. Using the results of Meisling, we have for $\lambda_2 = \frac{1-\rho_{11}}{\alpha_1}$, the steady-state mean waiting time experienced by a class 1 packet at N_2 is given by

$$\bar{w}_{(1)}^{(2)} = \frac{\rho_{12}(\alpha_2 - \alpha_1)}{2(1-\rho_{12})} = \frac{\rho_{12}\alpha_2}{2(1-\rho_{12})} - \frac{\rho_{12}\alpha_1}{2(1-\rho_{12})} \quad (3.2-5)$$

Equation (3.2-5) may then be used as a lower bound approximation to the actual waiting time experienced by a class 1 packet at N_2 .

A second, interesting approach to the problem of finding the waiting time at a separation node, is through the use of priority queueing. In Fig. 3.2-2 we have a two node tandem queueing network being fed by two external inputs. Node N_1 is assumed to operate according to a first-come first-served queueing discipline. On the other hand, node N_2 will operate using a priority discipline where we shall assume that class 1 packets have priority over class 2 packets. Indeed, we note that were the class 1 packets

to have a preemptive priority over the class 2 packets, then the waiting time experienced by a class 1 packet at node N_2 of Fig. 3.2-2 would be exactly the same as it would experience at node N_2 of Fig. 3.2-1. Thus, we have equated the problem of finding the delay on a packet at a separation node to that of finding the delay on a packet in a tandem network where the first stage operates on a first-come first-served basis and the second stage uses a preemptive priority discipline. However, we have as yet been unable to obtain exact results using the priority approach. Nevertheless, we have been successful in obtaining good approximations by assuming that node N_2 operates using a non-preemptive priority discipline.

For a non-preemptive priority discipline, one may write that the mean steady-state waiting time experienced by a class 2 packet at node N_2 of Fig. 3.2-2 is given by (see Kleinrock [13] and Saaty [22]),

$$\bar{w}_{(2)}^{(2)} = \bar{U} + \bar{N}_W \alpha_2 \quad (3.2-6)$$

where

$\bar{U} \stackrel{\Delta}{=} \text{the mean unfinished work remaining at node } N_2 \text{ as seen by an arriving packet;}$
 $\bar{N}_W \stackrel{\Delta}{=} \text{the mean number of class 1 packets to arrive at } N_2 \text{ during the waiting time of a class 2 packet.}$

Clearly, \bar{U} is equal to the mean steady-state waiting time that an arbitrary class packet would experience at N_2 were node N_2 to operate on a first-come first-served basis. Indeed, this must be true because the unfinished work remaining at a node will not change regardless of the order of service; that is, the unfinished work must be conserved (see Kleinrock [13]). This must be true even for G/G/1. Thus, we have that, for a non-preemptive priority at N_2 ,

$$\bar{U} = \bar{W}^{(2)}. \quad (3.2-7)$$

Turning to \bar{N}_W , it is here that we must make a simplifying assumption. With the arrival rate of class 1 packets to N_2 being λ_1 packets/sec., we shall assume that during the waiting time of a class 2 packet at N_2 , the number of class 1 packets to arrive will be

$$\bar{N}_W \approx \lambda_1 \bar{W}^{(2)}. \quad (3.2.8)$$

Consequently, using (3.2-7) and (3.2-8), we may rewrite (3.2-6) to read

$$\begin{aligned} \bar{W}^{(2)} &= \bar{W}^{(2)} + \lambda_1 \bar{W}^{(2)} \alpha_2 \\ &= \bar{W}^{(2)} + \rho_{12} \bar{W}^{(2)} \end{aligned} \quad (3.2-9)$$

which, upon solving for $\bar{w}_{(2)}^{(2)}$, yields

$$\bar{w}_{(2)}^{(2)} = \frac{\bar{w}^{(2)}}{1-\rho_{12}} . \quad (3.2-10)$$

Once $\bar{w}_{(2)}^{(2)}$ has been found, $\bar{w}_{(1)}^{(2)}$ may be calculated by noting that the mean steady-state waiting time experienced by an arbitrary class packet at a node with non-preemptive priorities, must be the same as at a node using a first-come first-served discipline. Therefore,

$$\begin{aligned} \bar{w}^{(2)} &= \bar{w}_{(1)}^{(2)} P\{\text{a packet is class 1}\} + \\ &\quad + \bar{w}_{(2)}^{(2)} P\{\text{a packet is class 2}\} = \\ &= \bar{w}_{(1)}^{(2)} \left[\frac{\lambda_1}{\lambda_1 + \lambda_2} \right] + \bar{w}_{(2)}^{(2)} \left[\frac{\lambda_2}{\lambda_1 + \lambda_2} \right] . \end{aligned} \quad (3.2-11)$$

Solving (3.2-11) for $\bar{w}_{(1)}^{(2)}$, we find

$$\bar{w}_{(1)}^{(2)} = \bar{w}^{(2)} \left[\frac{\lambda_1 + \lambda_2}{\lambda_1} \right] - \bar{w}_{(2)}^{(2)} \left[\frac{\lambda_2}{\lambda_1} \right] . \quad (3.2-12)$$

Substituting from (3.2-10) for $\bar{w}_{(2)}^{(2)}$, yields

$$\bar{w}_{(1)}^{(2)} = (1-\rho_{12}-\rho_{22})\bar{w}^{(2)} . \quad (3.2-13)$$

Hence, we have derived the result we were seeking.

$\bar{w}_{(1)}^{(2)}$, as given by (3.2-13), or equivalently by (3.2-12),

may now be used as an approximation to the steady-state waiting time experienced by a class 1 packet at node N_2 , the separation node, of Fig. 3.2-1.

As an example of results yielded by this procedure, we once again assume Poisson external input streams. In this case

$$\begin{aligned} \bar{W}^{(2)} &= \frac{(\rho_{12} + \rho_{22})\alpha_2}{2(1 - \rho_{12} + \rho_{22})} - \frac{(\rho_{11} + \rho_{21})\alpha_1}{2(1 - \rho_{11} - \rho_{21})} & (3.2-14) \\ &= \frac{[(\rho_{11} + \rho_{21}) + (\rho_{12} + \rho_{22}) - (\rho_{11} + \rho_{21})(\rho_{12} + \rho_{22})](\alpha_2 - \alpha_1)}{2(1 - \rho_{11} - \rho_{21})(1 - \rho_{12} - \rho_{22})} \end{aligned}$$

Then, using (3.2-10) followed by (3.2-13) we would obtain that

$$\bar{W}_{(1)}^{(2)} \approx \frac{[(\rho_{11} + \rho_{21}) + (\rho_{12} + \rho_{22}) - (\rho_{11} + \rho_{21})(\rho_{12} + \rho_{22})](\alpha_2 - \alpha_1)}{2(1 - \rho_{11} - \rho_{21})(1 - \rho_{12})} \quad (3.2-15)$$

This, then, is our approximate result for the separation node, N_2 , in Fig. 3.2-1. Shortly, we shall compare the approximation to the others derived earlier by means of computer simulation. However, before doing this, we shall take a closer look at (3.2-15) and derive two other approximate answers.

On observation of (3.2-15), we note the presence of the term ρ_{22} in the numerator. Clearly, the exact results for $\bar{W}_{(1)}^{(2)}$ cannot contain this factor. This is true since

the class 2 packets do not pass over the channel with service time α_2 in the network of Fig. 3.2-1. Thus, we suggest a new approximate result obtained by eliminating the factor ρ_{22} each time it appears in (3.2-15). Doing this, we find

$$\bar{W}_{(1)}^{(2)} \approx \frac{[\rho_{11} + \rho_{21} + \rho_{12} - (\rho_{11} + \rho_{21})\rho_{12}] (\alpha_2 - \alpha_1)}{2(1 - \rho_{11} - \rho_{21})(1 - \rho_{12})} \quad (3.2-16)$$

As a final approximation, we shall state a result which we have found through a trial and error procedure using (3.2-16). We have found that the elimination of the first ρ_{21} that appears in the numerator and the ρ_{21} of the denominator, will yield a very good approximation to simulation results. With this process, we have

$$\bar{W}_{(1)}^{(2)} \approx \frac{[\rho_{11} + \rho_{12} - (\rho_{11} + \rho_{21})\rho_{12}] (\alpha_2 - \alpha_1)}{2(1 - \rho_{11})(1 - \rho_{12})} \quad (3.2-17)$$

Of interest, in (3.2-17), is to note that all that remains of the interference of the class 2 packet stream on the class 1 stream, is the lone ρ_{21} term present in the numerator.

We shall now compare results obtained, at a separation node, using the approximations we have derived; namely, equations (3.2-1), (3.2-5), (3.2-15), (3.2-16), and (3.2-17), with those obtained from a simulation program.

In Figs. 3.2-3 - 3.2-6, we have plotted results of various simulation runs. On these graphs, we have plotted the mean steady-state waiting time on a class 1 packet at node N_2 of Fig. 3.2-1 for varying values of α_2 . Graphs A,B,C,D, and E represent results obtained using equations (3.2-1), (3.2-5), (3.2-15), (3.2-16), and (3.2-17) respectively. Simulation results are represented by an asterisk.

From these graphs, we conclude that results obtained using (3.2-15) and (3.2-16) are rather loose upper bounds. Eq. (3.2-5), as expected, yields a lower bound that is approached as $\rho_{11} + \rho_{21} \rightarrow 1$. On the other hand, (3.2-1) yields a fairly tight upper bound while (3.2-17) produces a tight lower bound approximation. In conclusion, we state that, at the moment, the techniques used to obtain (3.2-1) are the most useful for delay analysis at separation nodes. However, additional work using priority queueing seems promising to yield better methods.

3.3 DELAY ANALYSIS AT A COMBINED SEPARATION AND MERGER NODE

Having devised an approximate analysis procedure for a separation node, we now face the task of extending these methods such as to allow our newly separated stream of packets to be interfered by and merged with a new stream of packets. This leads us to the study of nodes which are both separation and merger points.

In Fig. 3.3-1, we have depicted a three node queueing network. At node N_1 , we have a merger node where class 1 and class 2 packets are combined and serviced by a single channel, with constant service time α_1 , over which the merged stream of packets travels to node N_3 . Class 3 packets enter the network at N_2 and are serviced at a constant rate of α_2 sec./packet by a channel over which they too travel to node N_3 . At node N_3 , the class 1 and class 2 packet streams are separated. However, on being separated from the class 2 stream, the class 1 packets are then merged with the class 3 packets and are processed over a single channel with a constant service time of α_3 sec./packet. Thus, node N_3 is both a separation and a merger node. Thus, our interest lies in devising a technique to obtain an accurate measurement of the mean steady-state waiting time that would be experienced by a class 1 or class 3 packet at node N_3 .

In the discussion that follows, we shall, once again, limit ourselves to the case of Poisson input streams though, as we shall point out later, for the case where we obtain exact results, they will be seen to apply for inputs following any general probability distribution.

For Poisson inputs, it was seen in the last section that a practical approximation at a separation node could be obtained by assuming the interference to be nonexistent when calculating the waiting time at the separation node. If we apply this principle to our present case, we find that for calculating the waiting time on a class 1 or class 3 packet at N_3 , the network of Fig. 3.3-1 transposes into the network of Fig. 3.3-2. On careful study, one immediately notes that the network of Fig. 3.3-2 is exactly the same as the network of Fig. 2.5-1. Consequently, all the procedures and results obtained in the analysis of network characteristics at node N_3 of Fig. 2.5-1 can now be applied to node N_3 of Fig. 3.3-2 and thus serve as our procedure for the analysis of the waiting time at node N_3 of Fig. 3.3-1.

In the analysis of the waiting time at node N_3 of Fig. 3.3-2, we must account, as was done in the analysis of Fig. 2.5-1, for the various possible relationships that may exist among the magnitudes of the various service times present in the network. This leads us to consider several different cases.

Case 1: $\alpha_3 \geq \alpha_1$ and $\alpha_3 \geq \alpha_2$.

For this case, node N_3 is the dominant node of Fig. 3.3-2. Hence, for Poisson inputs, the mean steady-state waiting time that would be experienced at N_3 by an arbitrary packet belonging either to class 1 or class 3, is of the form of equation (2.5-1) and is given by

$$\begin{aligned} \bar{w}_3^{(3)} = & \frac{(\rho_{13} + \rho_{33})\alpha_3}{2(1 - \rho_{13} - \rho_{33})} - \frac{\rho_{11}\alpha_1}{2(1 - \rho_{11})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_3} \right) - & (3.3-1) \\ & - \frac{\rho_{32}\alpha_2}{2(1 - \rho_{32})} \left(\frac{\lambda_3}{\lambda_1 + \lambda_3} \right) \end{aligned}$$

Case 2: $\alpha_3 > \frac{\alpha_1\alpha_2}{\alpha_1 + \alpha_2}$ and $\alpha_3 < \alpha_i$ for at least one value of i , $i = 1, 2$.

In this case, we once again apply results derived earlier; namely, theorem 2.5-1, which stated that a good approximation of the actual waiting time could be obtained by assuming $\alpha_i = \alpha_3$ for each $\alpha_i > \alpha_3$. As an example, for $\alpha_2 > \alpha_3 \geq \alpha_1$, we would obtain, similar to (2.5-6), that

$$\begin{aligned} \bar{w}_3^{(3)} \leq & \frac{(\rho_{13} + \rho_{33})\alpha_3}{2(1 - \rho_{13} - \rho_{33})} - \frac{\rho_{11}\alpha_1}{2(1 - \rho_{11})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_3} \right) - & (3.3-2) \\ & - \frac{\rho_{33}\alpha_3}{2(1 - \rho_{33})} \left(\frac{\lambda_3}{\lambda_1 + \lambda_3} \right) \end{aligned}$$

In a similar manner, for the cases $\alpha_1 > \alpha_3 \geq \alpha_2$ and $\alpha_3 < \alpha_1$ and $\alpha_3 < \alpha_2$, one would obtain equations similar to those of (2.5-5) and (2.5-7) respectively.

Case 3: $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$

As in the previous two cases, we draw on the results derived previously in section 2.5. On careful study, one notes that all of the theorems and corollaries derived in section 2.5 for the case under consideration, (case 2 in section 2.5), are directly applicable to our present network. Indeed, we may at this point drop the restriction of Poisson input streams, for a careful examination of the results and proofs of section 2.5 will show their applicability to our present network for inputs that follow any general probability distribution.

To begin, it is clear that theorem 2.5-2 applies in our network; thus, the total number of packets on queue at node N_3 of Fig. 3.3-2 may never exceed one. Of consequence, as stated by corollary 2.5-1, is the fact that the waiting time on any packet at node N_3 of Fig. 3.3-2 can never exceed α_3 sec.; that is,

$$w^{(3)} \leq \alpha_3 \tag{3.3-3}$$

Having arrived at this point, we must, as was done in section 2.5, consider two separate subcases; namely, $\alpha_3 \leq \frac{1}{2} \min (\alpha_1, \alpha_2)$ and $\frac{1}{2} \min (\alpha_1, \alpha_2) \leq \alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$.

Subcase 3.1: $\alpha_3 \leq \frac{1}{2} \min (\alpha_1, \alpha_2)$

For this case, the maximum length of a busy period at node N_3 of Fig. 3.3-2 is two, where a busy period of length two would consist of a single class 1 packet and a single class 3 packet. At this point, theorem 2.5-3 is applicable and following the derivation of that theorem, we find, similar to (2.5-25) and (2.5-26), that, at node N_3 of Fig. 3.3-2, we have

$$\bar{w}_{(1)}^{(3)} = \frac{\rho_{33} \alpha_3}{2} \quad (3.3-4a)$$

$$\bar{w}_{(3)}^{(3)} = \frac{\rho_{13} \alpha_3}{2} \quad (3.3-4b)$$

and

$$\bar{w}^{(3)} = \frac{\rho_{13} \rho_{33}}{\lambda_1 + \lambda_3} \quad (3.3-4c)$$

Of special note is that these results are exact, not only at node N_3 of Fig. 3.3-2, but also at node N_3 of Fig. 3.3-1, the combination separation and merger node.

Before moving on to the next subcase, we mention that, similar to (2.5-30), the distribution of the number of packets at node N_3 of Fig. 3.3-2 (and Fig. 3.3-1) is given by

$$P_0^{(3)} = 1 - \rho_{13} - \rho_{33} \quad (3.3-5a)$$

$$P_1^{(3)} = \rho_{13} + \rho_{33} - \rho_{13}\rho_{33} \quad (3.3-5b)$$

$$P_2^{(3)} = \rho_{13}\rho_{33} \quad (3.3-5c)$$

Subcase 3.2: $\frac{1}{2} \min(\alpha_1, \alpha_2) \leq \alpha_3 \leq \frac{\alpha_1\alpha_2}{\alpha_1 + \alpha_2}$

For this subcase, we run into the same difficulties as were discussed in section 2.5. The difficulties result from the fact that node N_3 may be concurrently occupied by two packets which belong to the same class. However, applying theorem 2.5-4 for the case $\alpha_1 = k\alpha_2$, $k = 1, 2, 3, \dots$; $\lambda_1 \leq \frac{1}{\alpha_1}$; and $\lambda_3 = \frac{1}{\alpha_2}$ we have, at node N_3 of Fig. 3.3-2 (and Fig. 3.3-1), that

$$\bar{W}_{(1)}^{(3)} = \frac{\rho_{33}\alpha_3}{2} \quad (3.3-6a)$$

$$\bar{W}_{(3)}^{(3)} = \frac{\rho_{13}\alpha_3}{2} + \rho_{13} \left(\sum_{j=2}^{\infty} [j\alpha_3 - (j-1)\alpha_2]^+ \right) \quad (3.3-6b)$$

and

$$\bar{W}^{(3)} = \frac{\rho_{13}\rho_{33}}{\rho_{13} + \rho_{33}} \left(\sum_{j=1}^{\infty} [j\alpha_3 - (j-1)\alpha_2]^+ \right) \quad (3.3-6c)$$

In addition, it should be pointed out that (3.3-6a) is an exact result even when we relax the restriction on λ_3 and allow for $\lambda_3 \leq \frac{1}{\alpha_2}$. On the other hand, (3.3-6b) and

(3.3-6c) are no longer exact when we allow $\lambda_3 \leq \frac{1}{\alpha_2}$; however, they may indeed serve as useful approximations to actual results.

Thus, we have been able to obtain a few results when α_1 is an integer multiple of α_2 . If we now reverse this condition such that we now have $\alpha_2 = k\alpha_1$, $k = 1, 2, 3, \dots$, and apply the conditions $\lambda_3 \leq \frac{1}{\alpha_2}$ and $\lambda_1 + \lambda_2 = \frac{1}{\alpha_1}$, we find that the mean waiting time on a class 3 packet at node N_3 of Fig. 3.3-2 (and Fig. 3.3-1) is given by

$$\bar{W}_{(3)} = \frac{\rho_{13}\alpha_3}{2} \quad (3.3-7)$$

similar to (3.3-6a). However, with respect to the mean steady-state waiting time on a class 1 packet, we have as yet been unable to formulate a useful expression. The difficulty in this comes from the fact that $\lambda_1 = \frac{1}{\alpha_1} - \lambda_2$ such that we do not have class 1 packets arriving at N_3 every α_1 seconds. Instead, either a class 1 or a class 2 packet arrives from N_1 to N_3 every α_1 seconds. However, it is noteworthy that as $\lambda_2 \rightarrow 0$, we find $\lambda_1 \rightarrow \frac{1}{\alpha_1}$, thus, yielding

$$\bar{W}_{(1)} \rightarrow \frac{\rho_{33}\alpha_3}{2} + \rho_{33} \left[\sum_{j=2}^{\infty} [j\alpha_3 - (j-1)\alpha_1]^+ \right] \quad (3.3-8a)$$

and

$$\bar{w}^{(3)} \rightarrow \frac{\rho_{13}\rho_{33}}{\rho_{13}+\rho_{33}} \left(\sum_{j=1}^{\infty} [j\alpha_3 - (j-1)\alpha_1]^+ \right) . \quad (3.3-8b)$$

As a second example of a combined separation and merger node, we consider the network of Fig. 3.3-3. In this network, node N_2 is a combination node in that at N_2 the class 1 packets are separated from the merged stream of class 1 and class 2 packets arriving from node N_1 and are merged with the class 3 packets which are entering N_3 from the external environment.

The main difference in analyzing this network, as compared to the network of Fig. 3.3-1, is that the class 3 packets entering the combination node, N_2 , do so from the external environment; thus, node N_3 may become unstable for all possible values of α_2 . Consequently, we need only consider two cases; namely, $\alpha_2 \geq \alpha_1$ and $\alpha_2 \leq \alpha_1$. In addition, we shall once again limit our discussion to Poisson input streams.

Case 1: $\alpha_2 \geq \alpha_1$

Once again, our technique is to assume $\lambda_2 = 0$ for purposes of finding an upper bound approximation as to the mean steady-state waiting time that will be experienced by an arbitrary class 1 or class 3 packet at node N_2 . Doing this, we would find node N_2 is a dominant node; therefore, using the methods developed earlier, we find, as an approximation, that

$$\bar{w}(2) \leq \frac{(\rho_{12} + \rho_{32})\alpha_2}{2(1 - \rho_{12} - \rho_{32})} - \frac{\rho_{11}\alpha_1}{2(1 - \rho_{11})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_3} \right) \quad (3.3-9)$$

Case 2: $\alpha_2 \leq \alpha_1$

For this case, we once again use the additional approximation method of theorem 2.5-1; that is, we set $\alpha_1 = \alpha_2$ for purposes of finding a useful approximate result. Consequently, on application, we find

$$\bar{w}(2) \leq \frac{(\rho_{12} + \rho_{32})\alpha_2}{2(1 - \rho_{12} - \rho_{32})} - \frac{\rho_{12}\alpha_2}{2(1 - \rho_{12})} \left(\frac{\lambda_1}{\lambda_1 + \lambda_3} \right) . \quad (3.3-10)$$

This concludes our analytical results for combination separation and merger nodes.

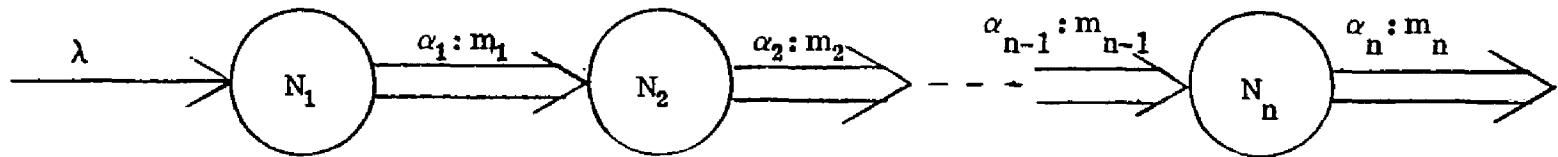


Fig. 3.1-1: A Tandem Queueing Network where each Stage has m_i Parallel Servers each having a Constant Service Time of α_i seconds, $i=1, 2, \dots, n$

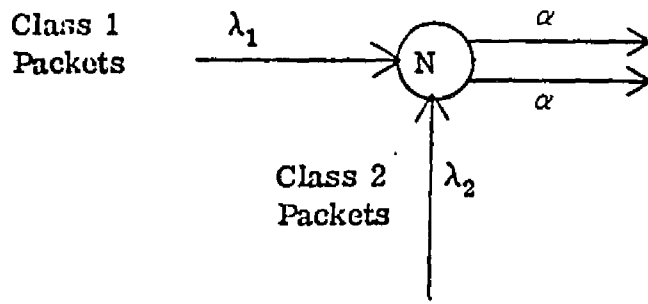


Fig. 3.1-2: A Single Node Network with Two Parallel Servers

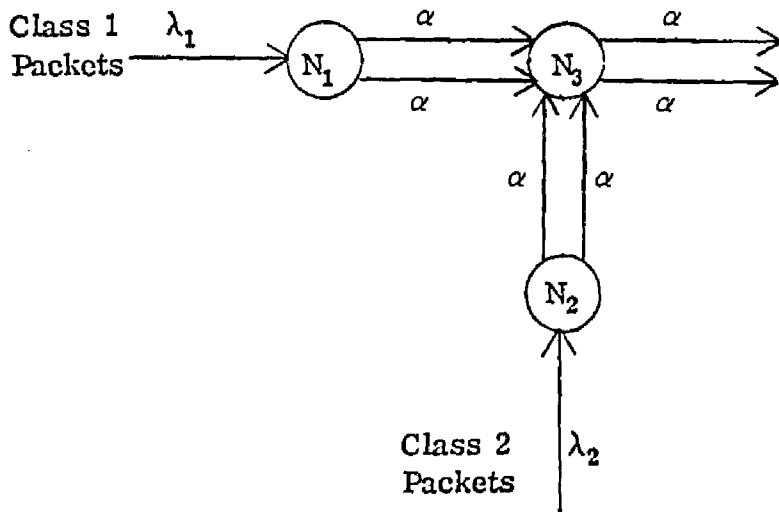


Fig. 3.1-3: A Three Node Network where Each Node has Two Parallel Servers

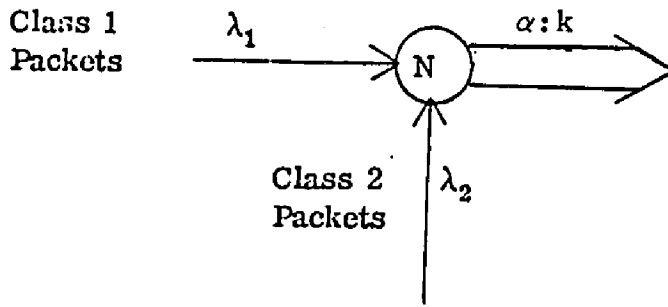


Fig. 3.1-4: A Single Node Network with " k " Parallel Servers

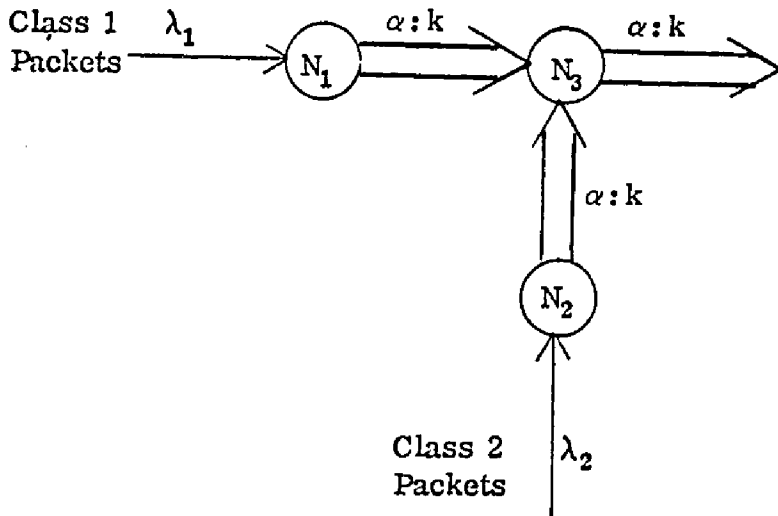
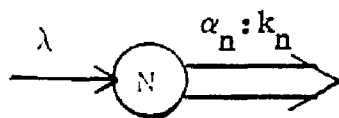
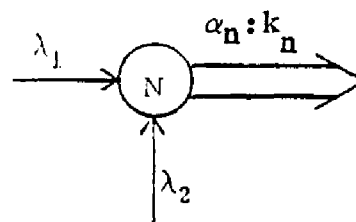


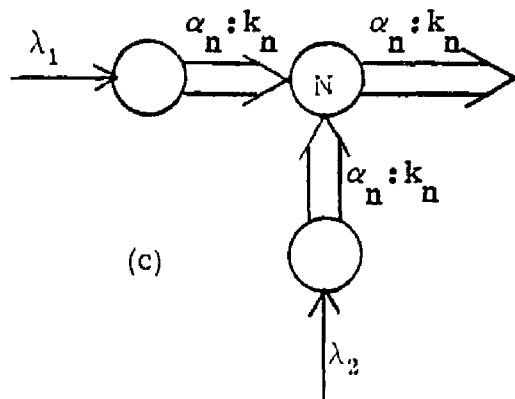
Fig. 3.1-5: A Three Node Network where Each Node has " k " Parallel Servers



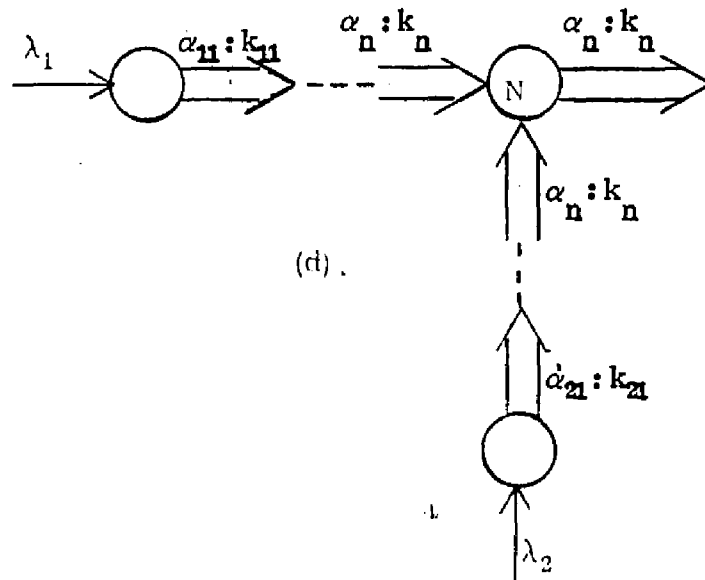
(a)



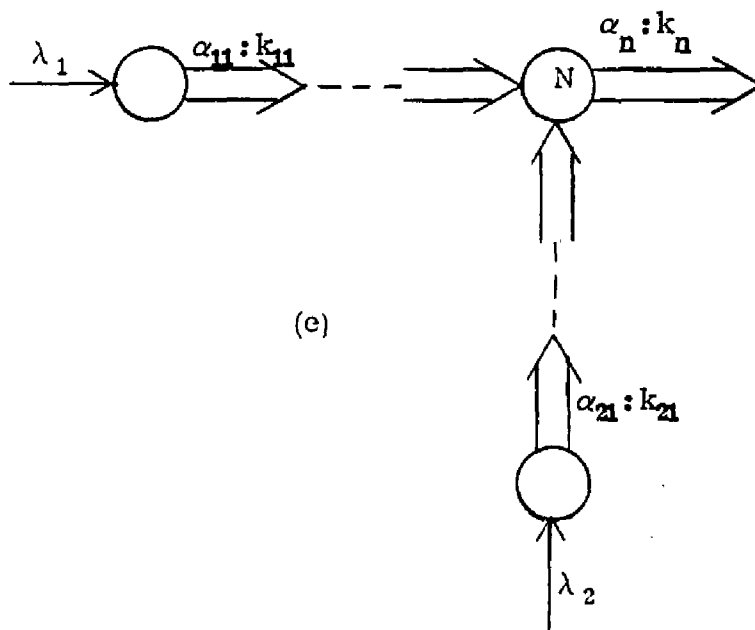
(b)



(c)



(d)



(e)

Fig. 3.1-6: Networks with Equivalent Overall Mean Steady-State Waiting Time when Stage $N = (\alpha_n : k_n)$ Dominates All Other Stages of the Network

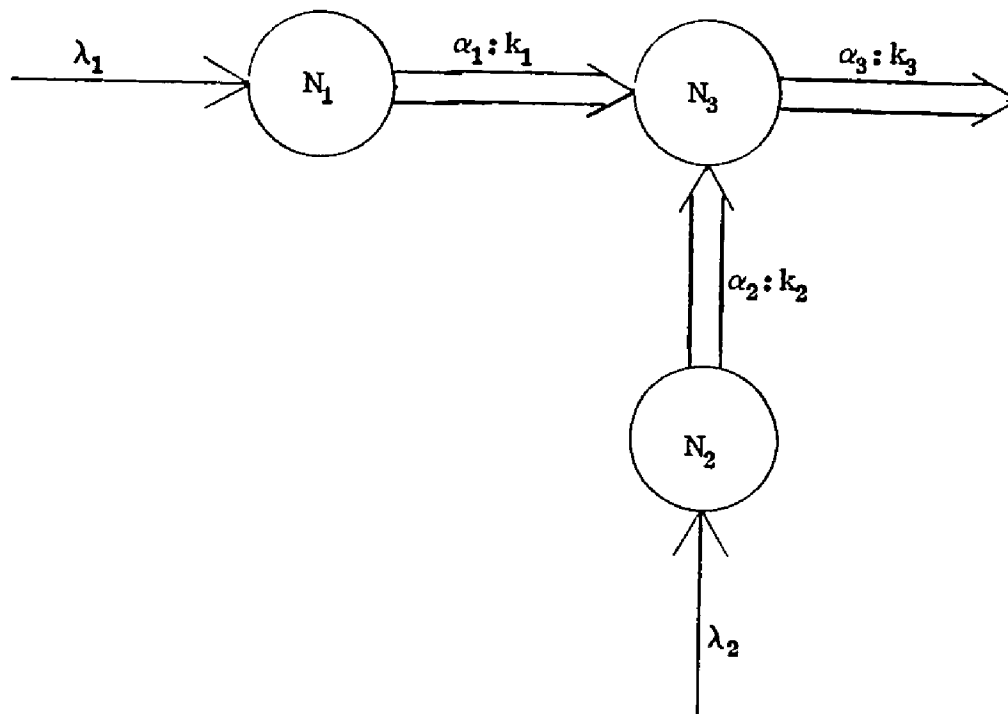


Fig. 3.1-7: A Three Node Network with Two External Input Streams

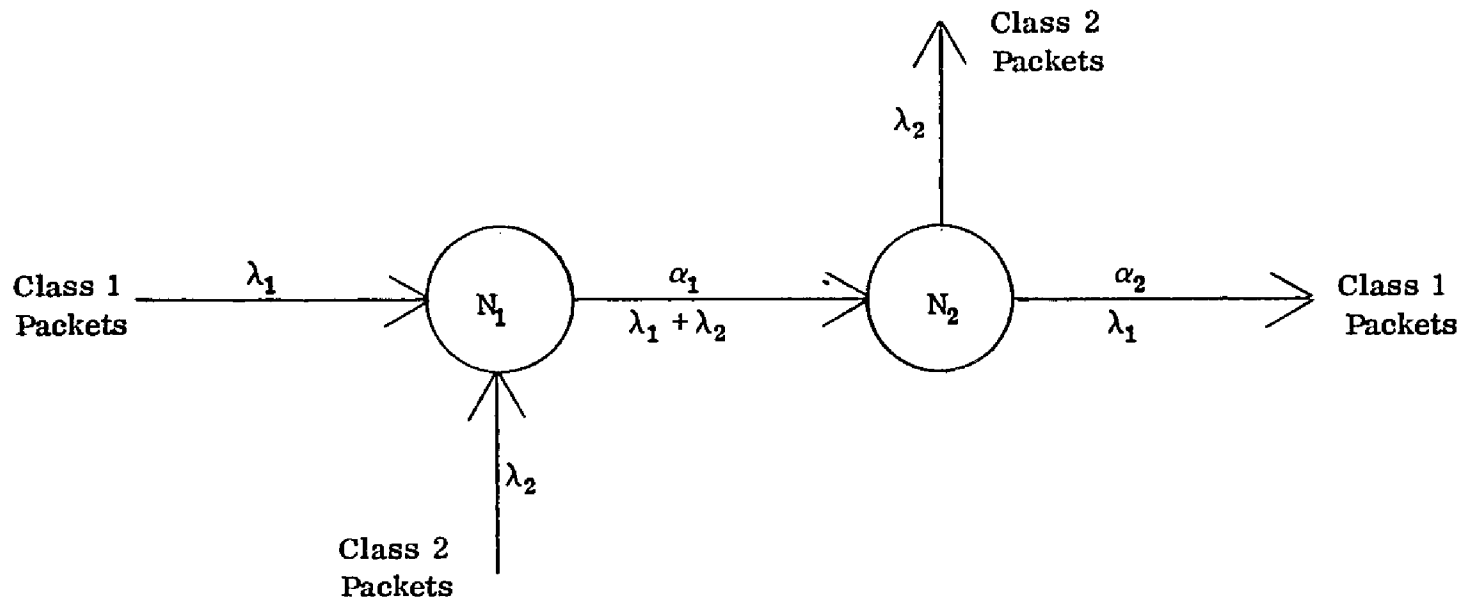


Fig. 3.2-1: A Queuing Network Consisting of a Merger Node Followed by a Separation Node

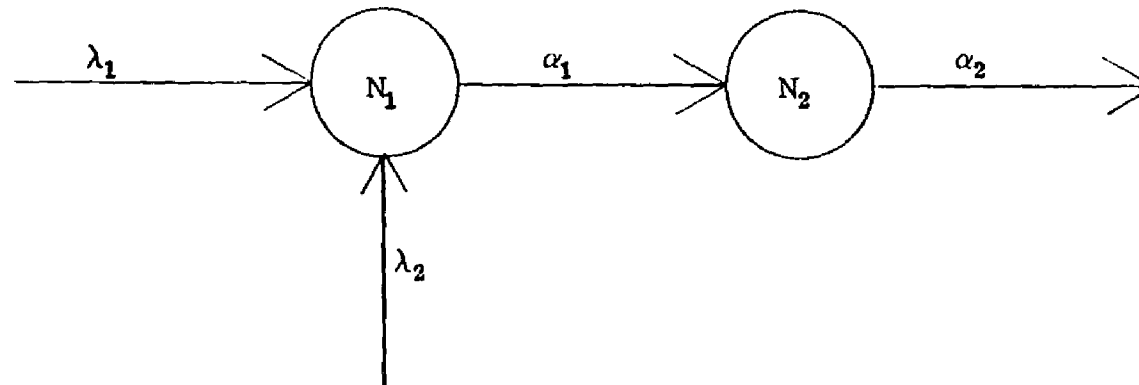


Fig. 3.2-2: A Tandem Queueing Network where Node N_1 has a First-Come First-Served Discipline while Node N_2 has a Priority Discipline

Waiting Time on a class 1 Packet at N₂ as a Function of α_2

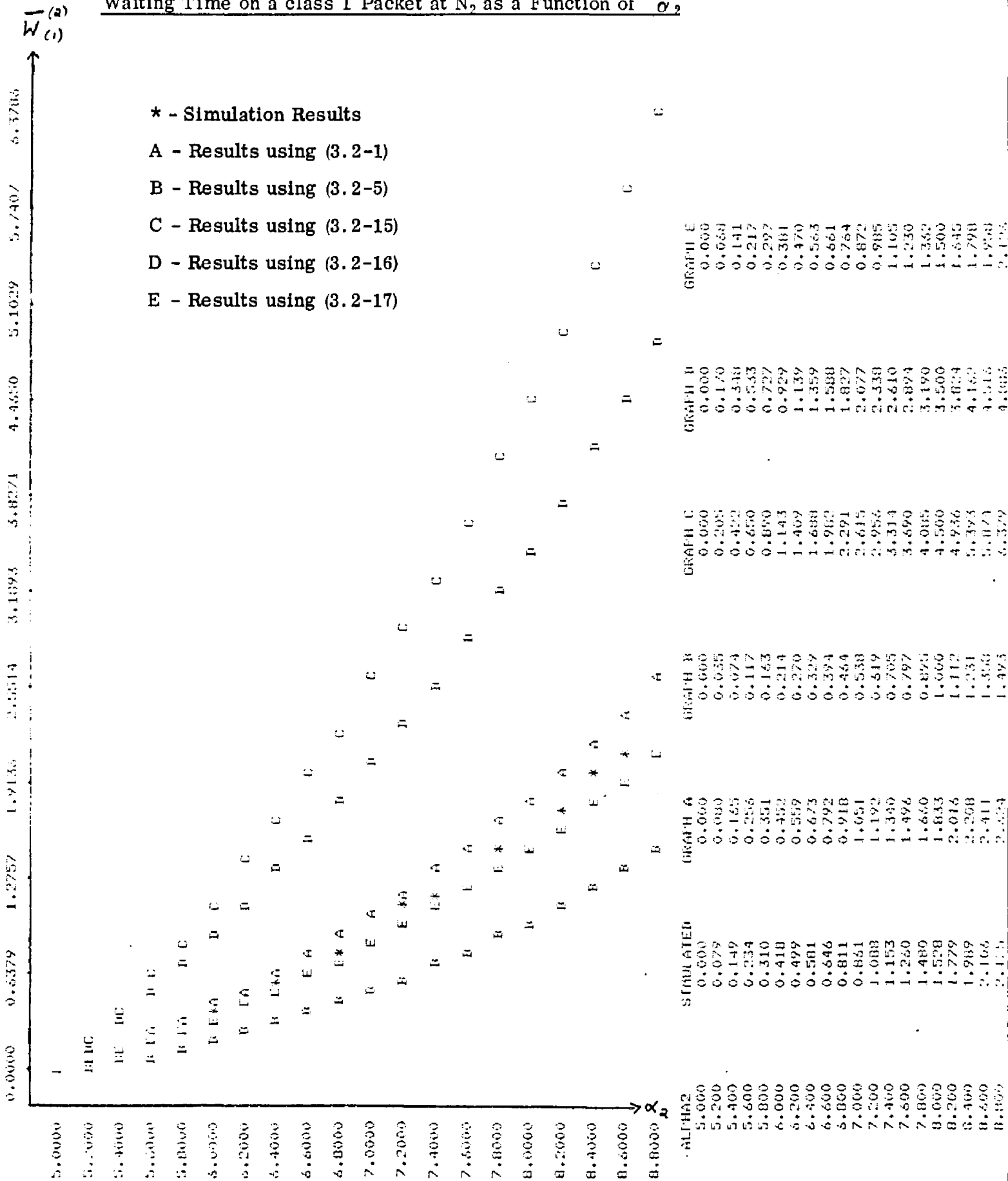


Fig. 3.2-3: Comparison of Results Obtained Using Eqs. (3.2-1), (3.2-5), (3.2-15), (3.2-16), and (3.2-17) for the case $\lambda_1 = \lambda_2 = 0.05$ and $\alpha_1 = 5$

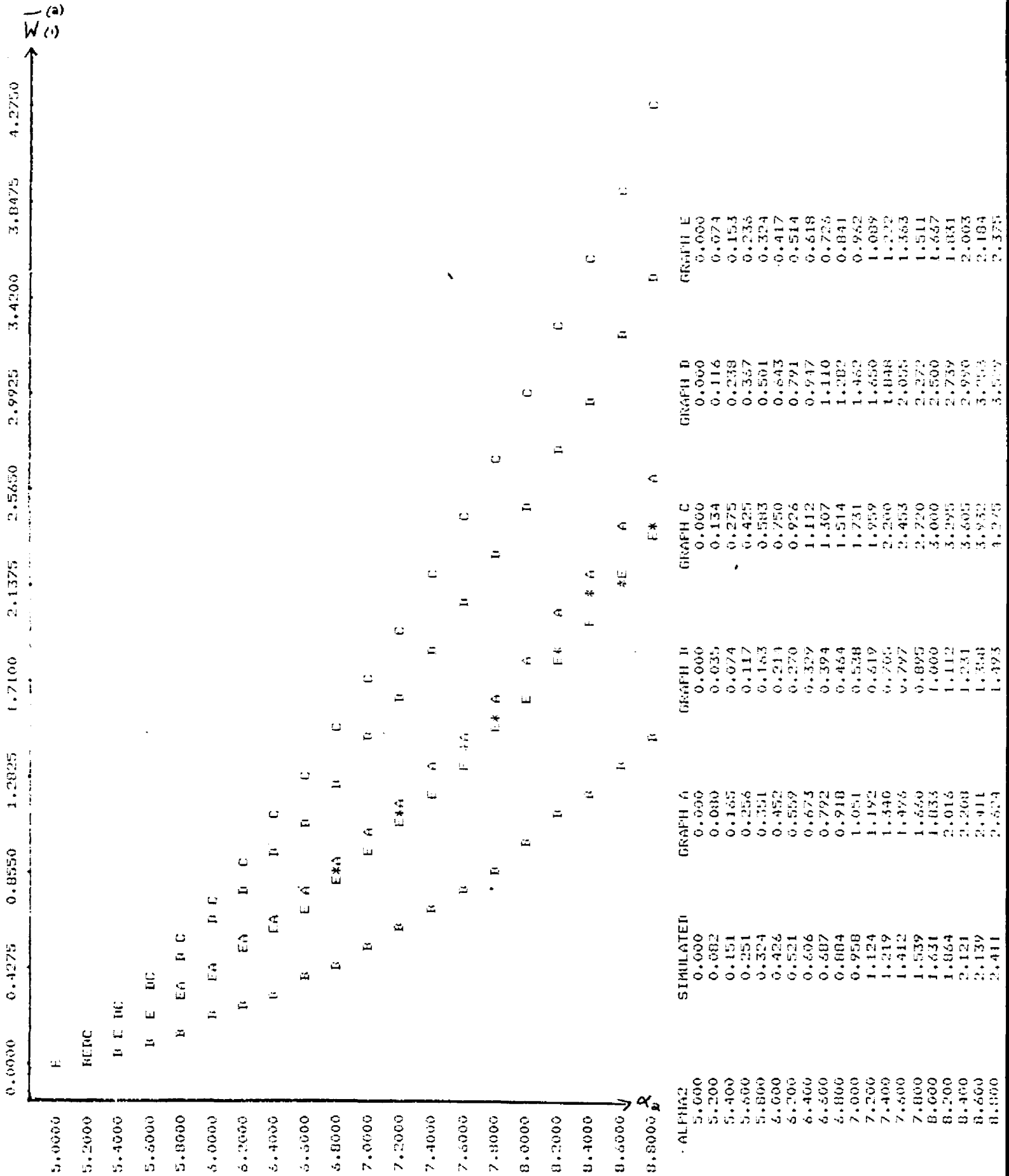


Fig. 3.2-4: Results for $\lambda_1 = 0.05$, $\lambda_2 = 0.025$, and $\alpha_1 = 5$

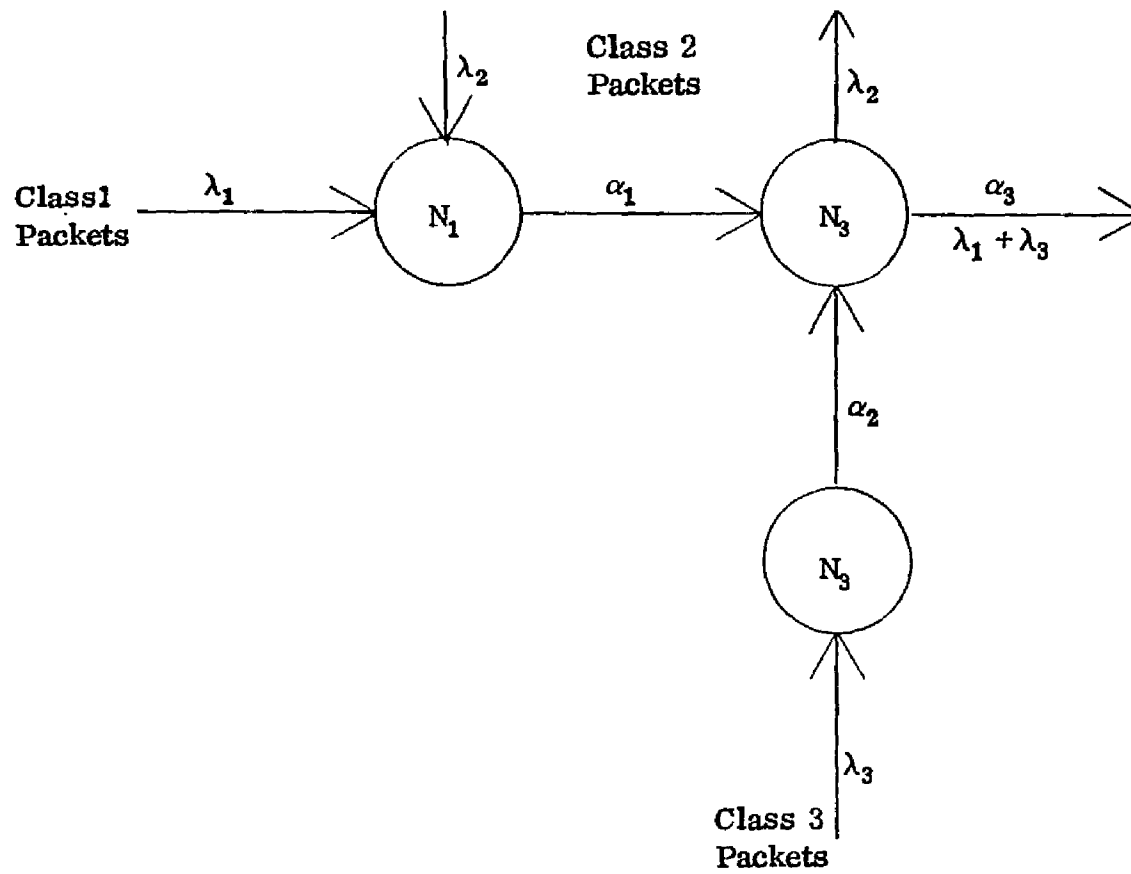


Fig. 3.3-1: A Queuing Network in which Node N_3 is a Combined Merger and Separation Node

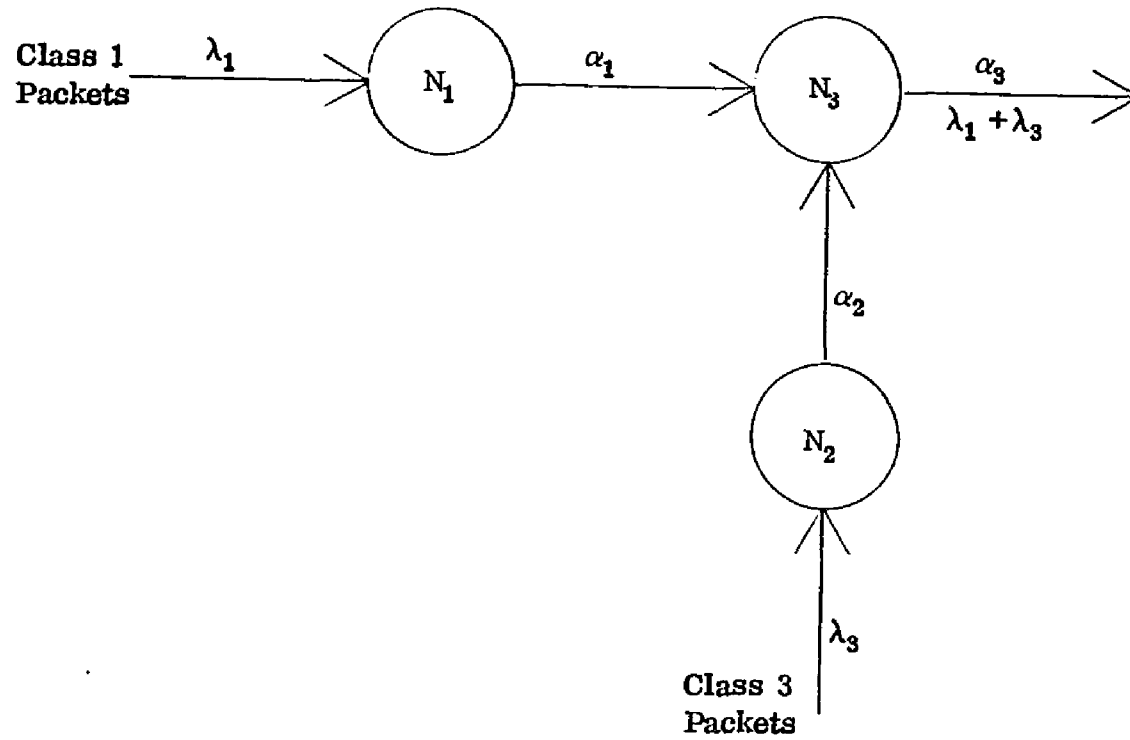


Fig. 3.3-2: Network Derived from Fig. 3.3-1 by Assuming $\lambda_2 = 0$

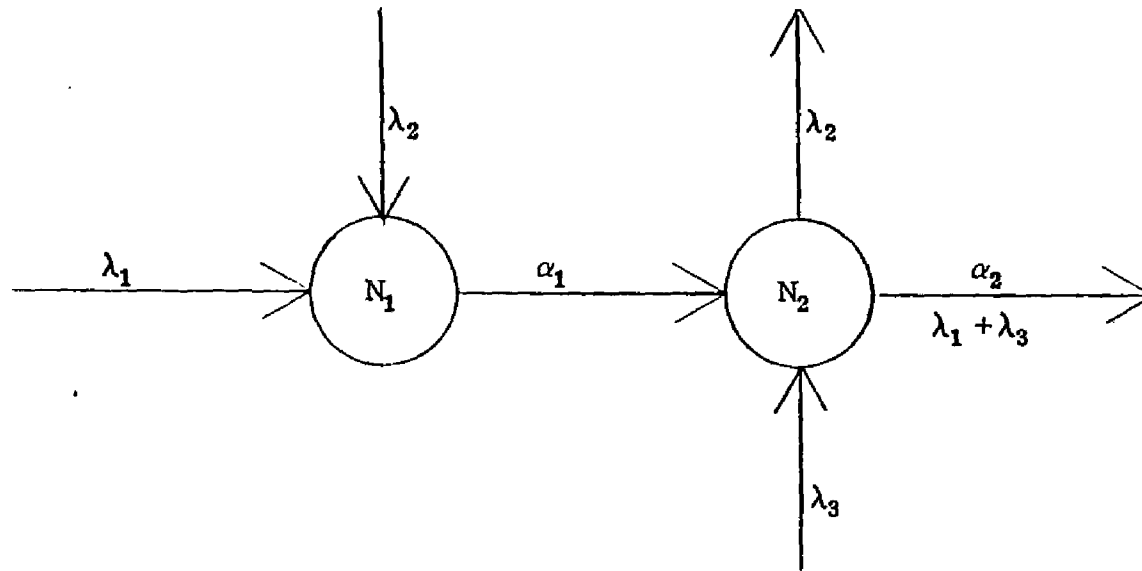


Fig. 3.3-3: A Network in which Node N_2 Merges One Internal and One External Packet Stream

4.1. CONCLUSIONS

In this dissertation, we have developed some useful analytic methods for use in the analysis and modeling of packet switched computer communications networks. Though not as general as one would hope, the techniques developed in the earlier chapters are seen as tools which may lead to the more efficient design of computer networks, especially centralized networks which have the general structure and topology of the networks that were considered in this dissertation.

For networks of single server queues; that is, for networks in which each node of the network has at most one outgoing channel over which to process packets, we have, in Chapter 2, developed a systematic method for their analysis. This was done by the derivation of techniques for the analysis of the mean steady-state waiting time on a packet at a merger node. For dominant merger nodes; that is, for merger nodes whose service time is slower than any of the preceding nodes that are feeding the merger node, we have shown that the overall steady-state mean waiting experienced by an arbitrary class packet as it passed through the network up to the merger node, would be determined solely by the service time of this dominant node. On the other hand, at fast merger nodes, we developed techniques which at times yielded exact results and at other times yielded tight upper bound approximations to actual values.

In Chapter 3, we turned our attention to networks which allowed for multiple servers at any given node of a network. Initially, we showed how some of the results obtained for feed-forward networks of single server queues may be extended to feed-forward networks where a number of parallel servers were allowed to exist between adjacent nodes. Indeed, the concept of dominance at a merger node, mentioned in the previous paragraph, was shown to be applicable in networks with parallel servers when dominance was properly defined for such a network. We then proceeded to study the waiting time characteristics at a separation node. Though we were unable to obtain any exact results, we were successful in developing some useful approximation techniques through the use of priority queueing and other procedures. Finally, we considered the case of a combination separation and merger node. Here, we were able to show the direct applicability of methods and results developed in Chapter 2, for feed-forward networks, to yield useful approximations in most situations, and in some specific cases were shown to yield exact results.

In total, we may conclude that the techniques and results developed in this dissertation, allow for the systematic and efficient analysis of centralized computer networks. The path that a packet travels from the user to the central station is seen to consist entirely of either tandem nodes or merger nodes, all of which are acting in a feed-forward manner. Thus, the

methods developed in Chapter 2, for networks of single server queues, are seen to be directly applicable. As for the return trip, that is, the route a packet follows from the central facility to the user terminal, it is seen to consist entirely of either tandem nodes or separation nodes. Hence, the methods used for tandem networks and the procedures developed in Chapter 3 for separation nodes would be used for analysis. Indeed, for a centralized network in which all channels, or servers, have the exact same capacity, or service time, the techniques that we have developed would yield an exact result at each node of the entire network, both for the trip from the user to the central station and for the return trip from the central station to the user terminal.

As a final point, we note that we have introduced the principle of delay decomposition, in which the waiting time experienced by a packet may be decomposed into two parts called self delay and interference delay. We believe this principle to be of importance in the future development of analysis procedures for more complex network structures than those considered in the context of this dissertation.

4.2. SUGGESTIONS FOR FUTURE WORK

In this dissertation, we have developed some useful analysis techniques for use in the analysis and modeling of packet switching computer networks. However, in the process, we have left many interesting questions unanswered. Indeed, the development of our methods has, as will be discussed shortly, raised many new, intriguing problems for the interested researcher.

Let us return to the network depicted by Fig. 2.5-1. Throughout Chapter 2, we had developed techniques for the analysis of the mean steady-state waiting time that would be experienced by a packet at node N_3 , the merger node of the network. For the case where the merger node was the dominant node in the network; that is, the node with the slowest packet service time, we developed in Sec. 2.2, a method for finding the exact waiting time experienced by a packet at the merger node. However, this exact result applied only to an arbitrary class packet. As for the waiting time experienced by a packet of a specific class, we have as yet been unable to obtain exact results. Thus, this is the first problem that remains to be answered. In Sec. 2.5, we turned our attention to the case where $\alpha_3 < \alpha_i$ for at least one value of i , $i = 1, 2$. For the case

$\alpha_3 > \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, we were successful in deriving a tight upper

bound approximation to actual results for the waiting time on an arbitrary class packet. Thus, in addition to the previously mentioned problem of finding the waiting time on a specific class packet, there is still the problem of finding a method that would yield exact results. Finally, for the case $\alpha_3 \leq \frac{\alpha_1 \alpha_2}{\alpha_1 + \alpha_2}$, we find that for $\alpha_3 \leq \frac{1}{2} \min(\alpha_1, \alpha_2)$, we were successful in obtaining methods that yielded exact waiting time results, not only for an arbitrary class packet, but also for packets from a specific class. However, for the subcase $\alpha_3 > \frac{1}{2} \min(\alpha_1, \alpha_2)$, we encountered many difficulties in attempting to derive any useful results. Our best effort resulted in theorem 2.5-4, which serves as our approximate method for this region. On study of theorem 2.5-4, one notes the added restrictions; namely, the two channels feeding the merger node must have service times, α_1 and α_2 , that are integer multiples of each other, and that arrival of packets along the faster of these two channels must be deterministic in the steady-state. Hence, the removal of these two restrictions, especially the latter, is of great interest.

In attempting a solution for the above mentioned problems, we believe that the principle of delay decomposition, introduced in Sec. 2.2, may be the tool with which to proceed. The principle of delay decomposition identified the components

of the waiting time experienced by a packet as self delay and interference delay. Indeed, it was shown that the delay experienced by a packet at a fast merger node consists entirely of interference delay. We believe that further development of this principle may lead to the derivation of more accurate techniques for the analysis of a fast merger node. Furthermore, it may also produce a solution to the problem of the waiting experienced by a specific class packet for both slow and fast merger nodes.

In Chapter 3, we ventured into the realm of networks in which a node was allowed to have multiple servers. This permitted the possibility for merged packet streams to separate at a separation node.

In Sec. 3.1, we showed how our techniques developed, in Sec. 2.2, for slow, dominant merger nodes, could be extended to yield exact results at dominant merger nodes in networks that allowed parallel servers to exist between adjacent nodes. However, for non-dominant merger nodes, we have as yet been unable to obtain satisfactory results for even a useful approximate method. We do believe, however, that a more careful study of the problem is necessary and would lead to some useful results.

In Sec. 3.2, we entered into a discussion of separation nodes. Through the discussion, we were able to derive at some approximation methods to be utilized at a separation node. However, much work is still needed in refining our

techniques at a separation node. We introduced the idea that one may calculate the waiting time on a packet at a separation node by analyzing a non-separation node which employed a preemptive priority discipline in use of its server. We were, however, unable to develop the concept to the extent we would have wished and thus this too remains as an area for possible future research.

In the final section of Chapter 3, Sec. 3.3, we concluded our study with a look at a combination node; that is, a node that is both a separation and a merger node. For this case, we saw how the methods developed in Chapter 2 for pure merger nodes could be used to obtain useful approximations for combination nodes and indeed some exact results under certain conditions. However, once again much more research could be used in developing better techniques at these combination nodes. We believe that the principle of delay decomposition, mentioned previously may possibly serve as an important tool in deriving better analysis methods for both combination and pure separation nodes.

In addition to the problems just mentioned with respect to specific questions that remain to be answered for the specific topics covered within the context of this dissertation, there exist many other interesting problems to which the results derived in this dissertation may have application.

The techniques developed in this dissertation were shown to have application in the analysis of centralized computer

networks. As for distributed networks, networks in which loops and meshes are allowed, no direct extension of the methods developed has been, as yet, found. Thus, the problem of extending our results so as to obtain useful analysis techniques for distributed networks remains to be attempted. The complexity of this task is much greater than for centralized networks due to the many additional factors that must be considered when analyzing distributed networks; for example, the routing strategy employed is of critical importance when dealing with distributed networks having loops and meshes within their topologies. As for current methods used in analyzing distributed networks, we refer the interested reader to the work of Kellinrock [12, pp. 320-9].

Yet another area of interest is in the study of networks containing queues with finite buffer lengths. The techniques developed above, were done by considering infinite storage available at every queue within the network. No extensions to networks with finite length queues have, as yet, been developed. The problem of finite buffers is indeed of interest for few results, especially for tandem networks of finite length queues, are existent at this time. For present methods and results available, the interested reader is referred to the series of papers by Chu [3-7] and Rudin [21] with regard to statistical multiplexors. The reader is also referred to

Abramson and Kuo [1, pp. 237-68] and Schwartz [23, pp. 154-70] for additional background material.

A final possible area, for which the techniques developed in this dissertation may have application, is in the capacity and flow assignment problems in computer networks. These problems are discussed in detail by Kleinrock [13, pp. 314-60] and Schwartz [23, pp. 58-102] and the references mentioned therein. Much of the work in the area has relied heavily on the "independence assumption" and it is our belief that it may be possible to apply our methods, which do not use the independence assumption, so as to obtain more efficient solutions to both the capacity and flow assignment problems, especially when considering these problems with respect to centralized networks.

Appendix A. Derivation of Equation (2.5-17)

In this appendix, we shall derive equation (2.5-17) in a more detailed manner.

To begin with, it was assumed that a busy period at node N_3 of Figure 2.5-1 would begin with the simultaneous arrivals of both a class 1 and class 2 packet at time $t = 0$. Thus, we have

$$T_1' = 0 \quad (A-1)$$

$$T_2' = 0 \quad (A-2)$$

since, from (2.5-15), $t_{11}^{(3)} = t_{21}^{(3)} = 0$.

If the busy period at node N_3 is to last for at least a third packet, then this third packet must be a class 1 packet since $t_{12}^{(3)} < t_{22}^{(3)}$ for all allowable values of k , $1 < k < 2$. (The case $k = 2$ is not allowable since for $k = 2$ the busy period may never exceed a length of two.) Thus, using the limits of $t_{12}^{(3)}$ from (2.5-16a) we find

$$\alpha_3 < T_3' \leq 2\alpha_3 \quad (A-3)$$

If we now allow the busy period at N_3 to be at least of length four, we note that, depending on the value of k , this fourth packet may be either class 1 or class 2. To find the region where the packet will be class 1, we must solve the inequality

$$t_{13}^{(3)} < t_{22}^{(3)} \quad (A-4)$$

which from (2.5-15) yields

$$2k\alpha_3 < \frac{k\alpha_3}{k-1} \quad (\text{A-5a})$$

$$(2k^2-3k)\alpha_3 < 0 \quad (\text{A-5b})$$

$$k < 3/2 \quad (\text{A-5c})$$

Therefore, in the region $1 < k < 3/2$, the fourth packet of the busy period must be a class 1 packet whose arrival time at N_3 is found from (2.5-15a) to be bounded as

$$2\alpha_3 < t_{13}^{(3)} < 3\alpha_3, \quad (\text{A-6a})$$

and in the region $3/2 < k < 2$, the fourth packet of the busy period must be a class 2 packet whose arrival time at N_3 is found from (2.5-15b) to be bounded as

$$2\alpha_3 < t_{22}^{(3)} < 3\alpha_3 \quad (\text{A-6b})$$

Consequently, from (A-6) we see that

$$2\alpha_3 \leq T_4' \leq 3\alpha_3 \quad (\text{A-7})$$

We shall now conclude this appendix by deriving the limits on T_5' .

For a busy period of length five at N_3 , this fifth packet of the busy period may once again be either class 1 or class 2. However, if it is class 1, it will be either the third or fourth class 1 packet of the busy period, and if it is class 2, it must be only the second class 2 packet of the busy period. All of this will, of course, depend on k . We shall now derive the regions of k for each of the three possible events just mentioned.

From (A-5), we see that for $3/2 < k < 2$, we must have $t_{22}^{(3)} < t_{13}^{(2)}$; then, in this region, the fifth packet of the busy period will be the third class 1 packet of the busy period whose arrival time at N_3 is limited to be

$$3\alpha_3 < t_{13}^{(3)} < 4\alpha_3 \quad (\text{A-8})$$

Let us now solve the inequality

$$t_{14}^{(3)} < t_{22}^{(3)} \quad (\text{A-9})$$

which will yield the region of k for which the fifth packet will be the fourth class 1 packet of the busy period. Using (2.5-15) we find

$$3k\alpha_3 < \frac{k\alpha_3}{k-1} \quad (\text{A-10a})$$

$$(3k^2 - 4k)\alpha_3 < 0 \quad (\text{A-10b})$$

$$k < \frac{4}{3} \quad (\text{A-10c})$$

Thus, in the region $1 < k < 4/3$, the fifth packet of a busy period at N_3 will be class 1, arriving at $t_{14}^{(3)}$ where

$$3\alpha_3 < t_{14}^{(3)} < 4\alpha_3 \quad (\text{A-11})$$

Finally, for the region $4/3 < k < 3/2$, this fifth packet must be a class 2 whose arrival time at N_3 , $t_{22}^{(3)}$, is bound as

$$3\alpha_3 < t_{22}^{(3)} < 4\alpha_3 \quad (\text{A-12})$$

Consequently, from (A-8), (A-11), and (A-12), we conclude that

$$3\alpha_3 \leq T_5' \leq 4\alpha_3 \quad (\text{A-13})$$

In a straightforward manner, this result may be extended to show that

$$(n-2)\alpha_3 \leq T_n' \leq (n-1)\alpha_3, \quad (\text{A-14})$$

thus proving (2.5-17).

Appendix B. Conjecture on the Extension of Results

Obtained in Section 2.5, Case 2, Subcase 2.1

In this appendix, we state a conjecture on the extension of results derived in Section 2.5, Case 2, Subcase 2.1 for networks with at most three external inputs, to networks with an arbitrary number of external packet streams. Though we have no formal proofs for what we are about to state, the conjectures that follow have been shown to be extremely accurate by use of simulation programs.

Our point of departure is equation (2.5-35), where we have derived the exact expression for the mean steady-state queue length at node N_4 of Figure 2.5-12, a network with three external inputs, for the case $\alpha_4 < \frac{1}{3} \min(\alpha_1, \alpha_2, \alpha_3)$. We will now utilize the same approach, that was used in going from a two input network to a three input network (see (2.5-30)-(2.5-35)), to extend our results from a three input network to one with four external inputs; such as, the network depicted in Figure B-1. Thus, the first step in our extension process is to find the queue length distribution at node N_4 of Figure 2.5-12. It is here that we make our first conjecture.

We hypothesize that the queue length distribution at node N_4 of Figure 2.5-12 is the following.

$$P_0^{(4)} = 1 - \rho_{14} - \rho_{24} - \rho_{34} \quad (B-1a)$$

$$P_1^{(4)} = \rho_{14} + \rho_{24} + \rho_{34} - \rho_{14}\rho_{24} - \rho_{14}\rho_{34} - \rho_{24}\rho_{34} \quad (B-1b)$$

$$P_2^{(4)} = \rho_{14}\rho_{24} + \rho_{14}\rho_{34} + \rho_{24}\rho_{34} - 3\rho_{14}\rho_{24}\rho_{34} \quad (B-1c)$$

$$P_3^{(4)} = 3\rho_{14}\rho_{24}\rho_{34} \quad (B-1d)$$

The reason we have come to this hypothesis is a knowledge of (2.5-30) and a need to satisfy (2.5-35), and a similar equation of $\bar{N}^{(4)}$, which we have done in the simplest way possible.

If we accept the conjecture of (B-1), we may now proceed to the network of Figure B-1, a network having four external packet streams merged at node N_5 with $\alpha_5 \leq \frac{1}{4} \min(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, and calculate the waiting time statistics at node N_5 using the approach of equations (2.5-32)-(2.5-34). Doing this, we obtain the following:

$$\begin{aligned} \bar{w}_{(1)}^{(5)} &= \left(\frac{\alpha_5}{2}\right) (\rho_{25} + \rho_{35} + \rho_{45} - \rho_{25}\rho_{35} - \rho_{25}\rho_{45} - \rho_{35}\rho_{45}) \quad (\text{B-2a}) \\ &\quad + \left(\frac{3\alpha_5}{2}\right) (\rho_{25}\rho_{35} + \rho_{25}\rho_{45} + \rho_{35}\rho_{45} - 3\rho_{25}\rho_{35}\rho_{45}) \\ &\quad + \left(\frac{5\alpha_5}{2}\right) (3\rho_{25}\rho_{35}\rho_{45}) \\ &= (\rho_{25} + \rho_{35} + \rho_{45}) \left(\frac{\alpha_5}{2}\right) + (\rho_{25}\rho_{35} + \rho_{25}\rho_{45} + \rho_{35}\rho_{45})\alpha_5 \\ &\quad + 3\rho_{25}\rho_{35}\rho_{45}\alpha_5 \end{aligned}$$

$$\begin{aligned} \bar{w}_{(2)}^{(5)} &= (\rho_{15} + \rho_{35} + \rho_{45}) \frac{\alpha_5}{2} + (\rho_{15}\rho_{35} + \rho_{15}\rho_{45} + \rho_{35}\rho_{45})\alpha_5 \\ &\quad + 3\rho_{15}\rho_{35}\rho_{45}\alpha_5 \quad (\text{B-2b}) \end{aligned}$$

$$\begin{aligned} \bar{w}_{(3)}^{(5)} &= (\rho_{15} + \rho_{25} + \rho_{45}) \frac{\alpha_5}{2} + (\rho_{15}\rho_{25} + \rho_{15}\rho_{45} + \rho_{25}\rho_{45})\alpha_5 \\ &\quad + 3\rho_{15}\rho_{25}\rho_{45}\alpha_5 \quad (\text{B-2c}) \end{aligned}$$

$$\begin{aligned} \bar{w}_{(4)}^{(5)} &= (\rho_{15} + \rho_{25} + \rho_{35}) \frac{\alpha_5}{2} + (\rho_{15}\rho_{25} + \rho_{15}\rho_{35} + \rho_{25}\rho_{35})\alpha_5 \\ &+ 3\rho_{15}\rho_{25}\rho_{35}\alpha_5 \end{aligned} \quad (\text{B-2d})$$

Consequently, we find that

$$\begin{aligned} \bar{w}^{(5)} &= \sum_{i=1}^4 \bar{w}_{(i)}^{(5)} P\{\text{a packet is class } i\} \quad (\text{B-3}) \\ &= \left(\frac{1}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} \right) (\rho_{15}\rho_{25} + \rho_{15}\rho_{35} + \rho_{15}\rho_{45} + \rho_{25}\rho_{35} \\ &+ \rho_{25}\rho_{45} + \rho_{35}\rho_{45} + 3\rho_{15}\rho_{25}\rho_{35} + 3\rho_{15}\rho_{25}\rho_{45} \\ &+ 3\rho_{15}\rho_{35}\rho_{45} + 3\rho_{25}\rho_{35}\rho_{45} + 12\rho_{15}\rho_{25}\rho_{35}\rho_{45}) \\ &= \left(\frac{1}{\sum_{i=1}^4 \lambda_i} \right) \left(\sum_{i=1}^3 \sum_{j=i+1}^4 \rho_{i5}\rho_{j5} + \right. \\ &\left. + \sum_{i=1}^2 \sum_{j=i+1}^3 \sum_{k=j+1}^4 3\rho_{i5}\rho_{j5}\rho_{k5} + 12\rho_{15}\rho_{25}\rho_{35}\rho_{45} \right) \end{aligned}$$

Finally,

$$\begin{aligned} \bar{N}_q^{(5)} &= (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4) \bar{w}^{(5)} \quad (\text{B-4}) \\ &= \sum_{i=1}^3 \sum_{j=i+1}^4 \rho_{i5}\rho_{j5} + \sum_{i=1}^2 \sum_{j=i+1}^3 \sum_{k=j+1}^4 3\rho_{i5}\rho_{j5}\rho_{k5} \\ &+ 12\rho_{15}\rho_{25}\rho_{35}\rho_{45} \end{aligned}$$

Having arrived at equation (B-4), we are now in a position where we can once again hypothesize on the queue length distribution at node N_5 of Figure B-1 and in this manner conjecture on results for a network with five

external input streams. Instead, we believe that we recognize an emerging pattern and we shall therefore conjecture on the extension to networks with an arbitrary number of external inputs, such as the network of Figure B-2.

In Figure B-2 we have a network with n external input streams being merged at node N_{n+1} with $\alpha_{n+1} \leq \frac{1}{n} \min(\alpha_1, \alpha_2, \dots, \alpha_n)$. It is our conjecture that the queue length distribution will be

$$\begin{aligned}
 P_0^{(n+1)} &= 1 = \sum_{i=1}^n \rho_{i,n+1} & (B-5) \\
 P_1^{(n+1)} &= \sum_{i=1}^n \rho_{i,n+1} - \sum_{i=1}^{n-1} \sum_{j=i+1}^n \rho_{i,n+1} \rho_{j,n+1} \\
 P_2^{(n+1)} &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \rho_{i,n+1} \rho_{j,n+1} \\
 &\quad - \left(\frac{3!}{2}\right) \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n \rho_{i,n+1} \rho_{j,n+1} \rho_{k,n+1} \\
 P_3^{(n+1)} &= \left(\frac{3!}{2}\right) \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n \rho_{i,n+1} \rho_{j,n+1} \rho_{k,n+1} \\
 &\quad - \left(\frac{4!}{2}\right) \sum_{i=1}^{n-3} \sum_{j=i+1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{\ell=k+1}^n \rho_{i,n+1} \rho_{j,n+1} \rho_{k,n+1} \rho_{\ell,n+1} \\
 \dots & \\
 P_{n-2}^{(n+1)} &= \frac{(n-2)!}{2} \sum_{i=1}^3 \sum_{j=i+1}^4 \dots \sum_{k=n-2}^n \rho_{i,n+1} \rho_{j,n+1} \dots \rho_{k,n+1} \\
 &\quad - \frac{(n-1)!}{2} \sum_{i=1}^2 \sum_{j=i+1}^3 \dots \sum_{k=n-1}^n \rho_{i,n+1} \rho_{j,n+1} \dots \rho_{k,n+1}
 \end{aligned}$$

$$P_{n-1}^{(n+1)} = \frac{(n-1)!}{2} \sum_{i=1}^2 \sum_{j=i+1}^3 \cdots \sum_{k=n-1}^n \rho_{i,n+1} \rho_{j,n+1} \cdots \rho_{k,n+1}$$

$$- \frac{n!}{2} \prod_{i=1}^n \rho_{i,n+1}$$

$$P_n^{(n+1)} = \frac{n!}{2} \prod_{i=1}^n \rho_{i,n+1}$$

where it is clear that there can never be more than n packets, one from each class, at node N_{n+1} at any one instant of time.

From (B-5), one can proceed in a direct manner and find our conjecture for the mean steady-state queue length at node N_{n+1} to be

$$\bar{N}_q^{(n+1)} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \rho_{i,n+1} \rho_{j,n+1} \tag{B-6}$$

$$+ \left(\frac{3!}{2}\right) \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n \rho_{i,n+1} \rho_{j,n+1} \rho_{k,n+1}$$

$$+ \dots + \frac{(n-1)!}{2} \sum_{i=1}^2 \sum_{j=i+1}^3 \cdots \sum_{k=n-1}^n \rho_{i,n+1} \rho_{j,n+1}$$

$$\cdots \rho_{k,n+1}$$

$$+ \frac{n!}{2} \prod_{i=1}^n \rho_{i,n+1}$$

Clearly, the waiting time on an arbitrary packet at node N_{n+1} may be found using Little's formula; that is,

$$\bar{W}^{(n+1)} = \left(\sum_{i=1}^n \lambda_i \right)^{-1} \bar{N}_q^{(n+1)} \tag{B-7}$$

Finally, equations for $\bar{W}_{(j)}^{(n+1)}$, $j = 1, 2, \dots, n$, may be found from (B-5) and application of the method introduced earlier in the derivation of (B-2).

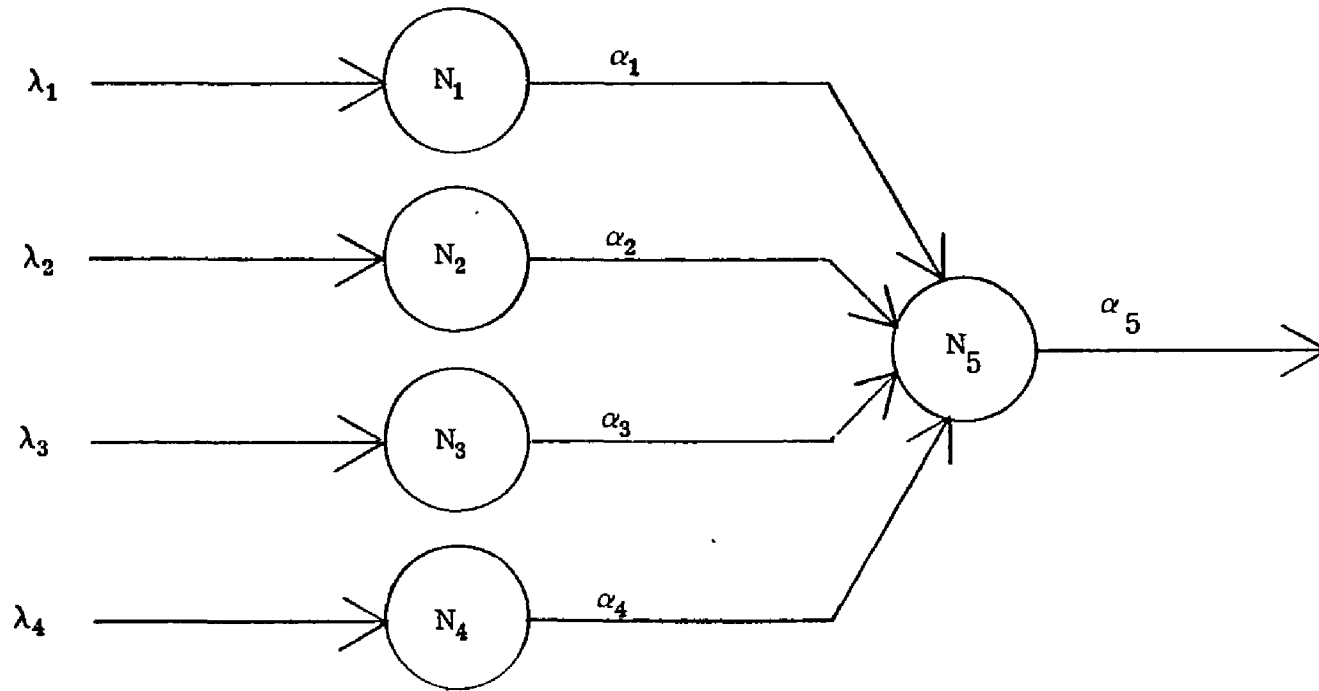


Fig. B-1: A Five Node Network with Four External Inputs and $\alpha_5 \leq \frac{1}{2} \min (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$

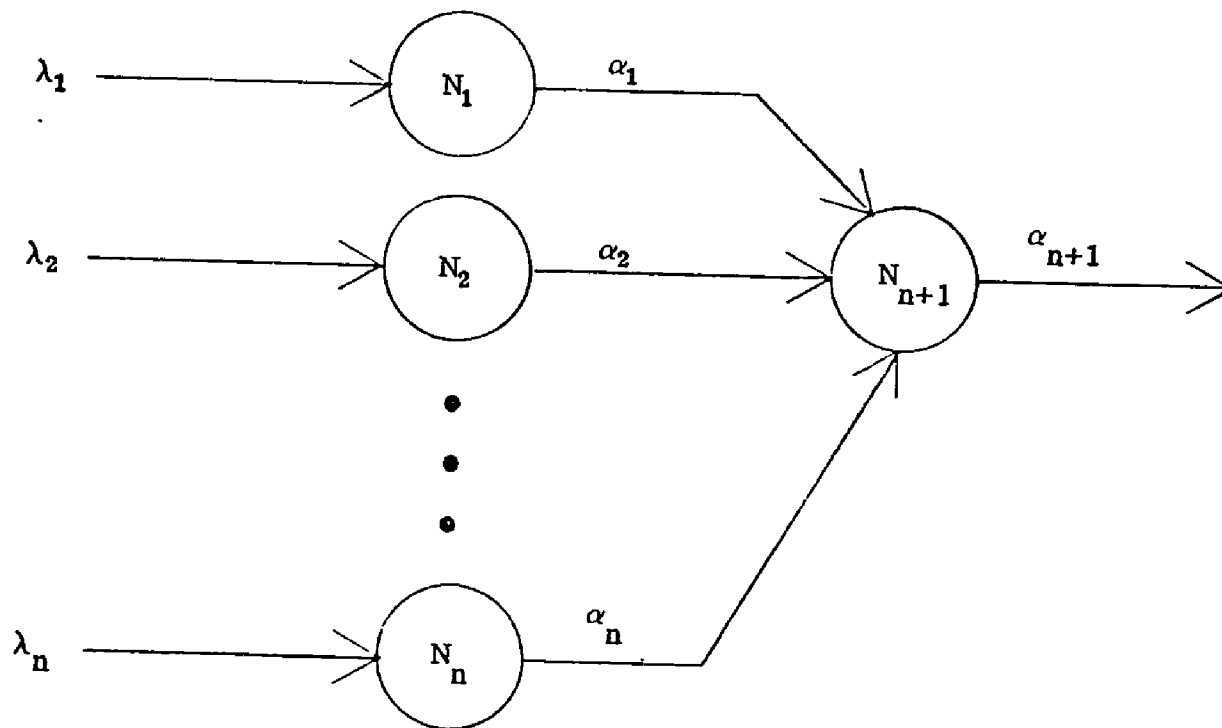


Fig. B-2: A Network with " n " External Inputs and $\alpha_{n+1} \leq 1/n \min(\alpha_1, \alpha_2, \dots, \alpha_n)$

REFERENCES

- [1] N. Abramson and F.F. Kuo, eds., Computer-Communications Networks, Prentice-Hall, Englewood Cliffs, N.J., 1973.
- [2] P.J. Burke, "The Output of a Queueing System," *Oper. Res.*, Vol. 4, pp. 699-704, Dec. 1956.
- [3] W.W. Chu, ed., Advances in Computer Communications, Artech House, Dedham, MA., 1974.
- [4] W.W. Chu, "Buffer Behavior for Poisson Arrivals and Multiple Constant Outputs," *IEEE Trans. Computer*, Vol. C-19, pp. 530-34, June 1970.
- [5] _____, "Buffer Behavior for Batched Poisson Arrivals and Single Constant Output," *IEEE Trans. Commun.*, Vol. COM-18, pp. 613-18, Oct. 1970.
- [6] _____, "Demultiplexing Considerations for Statistical Multiplexors," *IEEE Trans. Commun.* Vol. COM-20, pp. 603-609, June 1972.
- [7] W.W. Chu and L.C. Liang, "Buffer Behavior for Mixed Input Traffic and Single Constant Output Rate," *IEEE Trans. Commun.*, Vol. COM-20, pp. 230-235, April 1972.
- [8] D.W. Davies and D.L.A. Barber, Communication Networks for Computers, John Wiley & Sons, New York, 1973.
- [9] H.D. Friedman, "Reduction Methods for Tandem Queueing Systems," *Oper. Res.*, Vol. 13, pp. 121-31, 1965.
- [10] P.E. Green and R.W. Lucky, eds., Computer Communications, IEEE Press, New York, 1974.
- [11] L. Kleinrock, Communication Nets, McGraw-Hill, New York, 1964, Dover, New York, 1972.
- [12] _____, Queueing Systems Volume I: Theory, John Wiley & Sons, New York, 1975.
- [13] _____, Queueing Systems Volume II: Computer Applications, John Wiley & Sons, New York, 1976.
- [14] J.D.C. Little, "A Proof for the Queueing Formula $L = \lambda W$," *Oper. Res.*, pp. 383-87, May 1961.

- [15] T. Meisling, "Discrete-time Queueing Theory," *Oper. Res.*, pp. 96-105, 1958.
- [16] N.U. Prabhu, Queues and Inventories, John Wiley & Sons, New York, 1965.
- [17] I. Rubin, "Tandem Queues with Constant Channel Service Times and Group Arrivals," *UCLA, Tech. Rep.*, UCLA-Eng-7417, March 1974.
- [18] _____, "Communication Networks: Message Path Delays," *IEEE Trans. Inform. Theory*, Vol. IT-20, pp. 38-45, Nov. 1974.
- [19] _____, "Message Path Delays in Packet-Switching Communication Networks," *IEEE Trans. Commun.*, Vol. COM-23, pp. 186-92, Feb. 1975.
- [20] _____, "An Approximate Time-Delay Analysis for Packet Switching Communications Networks," *IEEE Trans. Commun.*, Vol. COM-24, pp. 210-221, Feb. 1976.
- [21] H. Rudin, "Performance of Simple Multiplexor-Concentrators for Data Communications," *IEEE Trans. Commun.*, Vol. COM-19, pp. 178-87, April 1971.
- [22] T. Saaty, Elements of Queueing Theory, McGraw-Hill, New York, 1961.
- [23] M. Schwartz, Computer-Communication Network Design and Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1977.
- [24] Special Issue on Computer Communications, *IEEE Trans. Commun.*, Vol. COM-25, Jan. 1977.