THE SIGNIFICANCE OF TRANSIENTS FOLLOWING FAILURES AND REPAIRS IN PACKET-SWITCHED NETWORKS

by

John E. Kobza

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Industrial and Systems Engineering

APPROVED:

Joel A. Nachlas, Chairman

Jeffrey D. Tew

Ira Jacobs

Jeffrey C. Woldstad

Scott F. Midkiff

February, 1993

Blacksburg, Virginia
THE SIGNIFICANCE OF TRANSIENTS FOLLOWING FAILURES AND REPAIRS IN PACKET-SWITCHED NETWORKS

by

John E. Kobza

Joel A. Nachlas, Chairman

Industrial and Systems Engineering

(ABSTRACT)

A system composed of unreliable components can experience different levels of performance as its configuration changes due to failures and repairs. One approach used to measure overall system performance is to weight the level of performance measured for each system state by the probability that the system is in that state and then sum across all system states. Many performance measures have a transient behavior following a change in the state of the system. Because of the difficulty associated with transient analysis, the system is often assumed to be in steady state when measuring the performance for each system state.

When this approach is used to analyze packet-switched communication networks, which consist of highly reliable high-speed links and switching nodes, it is argued that the steady-state assumption is justified on the basis of the large difference in rates of traffic-related events, such as call completions and packet transmissions, compared to component-related events, such as failures and repairs.

To investigate the validity of this assumption, we define lower bounds for the length of the transient phase following link failures and repairs. For both cases, we obtain a distribution for the length of the lower bound. The transient phase is significant when its length exceeds a given fraction of the time until the next change in network state. Using the distributions for these lengths,
we derive an expression for the probability that the transient phase is significant in terms of the amount of traffic on the link and the ratio of the rates for traffic-related events and network state changes.

These results show that the difference in rates between traffic-related events and component-related events is not enough by itself to justify the steady-state assumption. The amount of traffic carried on the link and the size of the network must also be considered. These results indicate some situations where the steady-state assumption is inappropriate. We also obtain sufficient conditions for transient-phase significance following link failures. Although these results do not indicate when it is safe to use the steady-state assumption, they provide a measure of the risk associated with using it.
ACKNOWLEDGEMENTS

During any work which spans a number of years, there are many people who, in one way or another, contribute to its completion and its final form. I would like to thank a few of these people at this time.

First, the two men largely responsible for my undertaking and completing this work: Dr. John Daigle and Dr. Joel Nachlas. Dr. John Daigle spent many years teaching me how to formulate and solve problems. He also taught me many life lessons for which I will always be grateful. Dr. Joel Nachlas gave me the vision to see that this work could be done and guided me to its completion. I have enjoyed working with him and hope to have the opportunity to do so again.

The members of my committee, both past and present, for their helpful suggestions during the course of this work.

Dr. Sara Anderson for her many enlightening discussions about the English language.

Leslie Graham who typed major portions of this document.

Finally, my family. Autumn, the study cat, for her many hours of late-night companionship, especially during Real Analysis. My daughter Kate, who teaches me about the really important things in life. My wife Katherine, who supported and encouraged me during these years and made greater sacrifices than I did to see this work completed.
TABLE OF CONTENTS

1. INTRODUCTION
   1.1 Relationship Between Structure and Performance  ..... 1
   1.2 A Model of Network Structure  ..... 2
   1.3 The Steady-State Assumption  ..... 4
   1.4 Illustration of Transients  ..... 5
   1.5 Statement of Problem  ..... 6

2. NETWORK EFFECTIVENESS MEASURES  ..... 9
   2.3 Level One - Structural Measures  ..... 10
   2.4 Level Two - Capability Measures  ..... 11
   2.5 Level Three - Load-Dependent Measures  ..... 12

3. TIME-DEPENDENT AND TRANSIENT RESULTS
   FOR NETWORKS OF QUEUES  ..... 15
   3.1 Numerical Techniques for Markov Systems  ..... 16
   3.2 Approximations for Markovian Networks  ..... 17
   3.3 Numerical Techniques for General Queueing Networks  ..... 18
   3.4 Networks with Time-Varying Rates  ..... 18

4. THE NETWORK:
   ASSUMPTIONS, PROTOCOLS, AND TRANSIENT COMPONENTS  ..... 20
   4.1 Network Traffic Assumptions  ..... 20
   4.2 Network Protocols  ..... 23
   4.3 The Physics of Failure - Network Behavior Following Failures  ..... 24
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Network Structure</td>
<td>3</td>
</tr>
<tr>
<td>4.1</td>
<td>ISO OSI Reference Model</td>
<td>22</td>
</tr>
<tr>
<td>5.1</td>
<td>Death Process for Secondary Calls</td>
<td>32</td>
</tr>
<tr>
<td>5.2</td>
<td>Probability Repair Transient is Significant, $\delta = 0.001$</td>
<td>45</td>
</tr>
<tr>
<td>5.3</td>
<td>Probability Repair Transient is Significant, $\delta = 0.01$</td>
<td>46</td>
</tr>
<tr>
<td>5.4</td>
<td>Probability Repair Transient is Significant, $\delta = 0.1$</td>
<td>47</td>
</tr>
<tr>
<td>5.4</td>
<td>Example Network</td>
<td>50</td>
</tr>
<tr>
<td>6.1</td>
<td>Mixed Queueing Model</td>
<td>55</td>
</tr>
<tr>
<td>6.2</td>
<td>Single-Queue System with One Closed Customer and Immediate Feedback</td>
<td>60</td>
</tr>
<tr>
<td>6.3</td>
<td>Single-Queue System with Multiple Closed Customers and Immediate Feedback</td>
<td>61</td>
</tr>
<tr>
<td>6.4</td>
<td>Output Process for Single-Customer System</td>
<td>63</td>
</tr>
<tr>
<td>6.5</td>
<td>Converging Distributions</td>
<td>72</td>
</tr>
<tr>
<td>6.6</td>
<td>Cycle Structure for Multiple-Customer System</td>
<td>80</td>
</tr>
<tr>
<td>6.7</td>
<td>Probability of Failure Transient Significance $\rho = 0.6$, $\epsilon = 0.1$, $\delta = 0.01$</td>
<td>93</td>
</tr>
<tr>
<td>6.8</td>
<td>Probability of Failure Transient Significance $\rho = 0.8$, $\epsilon = 0.1$, $\delta = 0.01$</td>
<td>94</td>
</tr>
<tr>
<td>6.9</td>
<td>Probability of Failure Transient Significance $\rho = 0.9$, $\epsilon = 0.1$, $\delta = 0.01$</td>
<td>95</td>
</tr>
<tr>
<td>6.10</td>
<td>Probability of Failure Transient Significance $\rho = 0.8$, $\epsilon = 0.1$, $\delta = 0.001$</td>
<td>95</td>
</tr>
<tr>
<td>6.11</td>
<td>Probability of Failure Transient Significance $\rho = 0.8$, $\epsilon = 0.1$, $\delta = 0.1$</td>
<td>97</td>
</tr>
<tr>
<td>6.12</td>
<td>Probability of Failure Transient Significance $\rho = 0.8$, $\epsilon = 0.01$, $\delta = 0.01$</td>
<td>98</td>
</tr>
<tr>
<td>6.13</td>
<td>Probability of Failure Transient Significance $\rho = 0.8$, $\epsilon = 0.006$, $\delta = 0.01$</td>
<td>99</td>
</tr>
<tr>
<td>7.1</td>
<td>Single-Queue System with One Closed Customer and Delayed Feedback</td>
<td>111</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 5.1</td>
<td>Repair Transient Significance, $\delta = 0.001$</td>
<td>44</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Repair Transient Significance, $\delta = 0.01$</td>
<td>44</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Repair Transient Significance, $\delta = 0.1$</td>
<td>44</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>Failure Transient Significance, $\rho = 0.6, \varepsilon = 0.1, \delta = 0.01$</td>
<td>89</td>
</tr>
<tr>
<td>Table 6.2</td>
<td>Failure Transient Significance, $\rho = 0.8, \varepsilon = 0.1, \delta = 0.01$</td>
<td>89</td>
</tr>
<tr>
<td>Table 6.3</td>
<td>Failure Transient Significance, $\rho = 0.9, \varepsilon = 0.1, \delta = 0.01$</td>
<td>90</td>
</tr>
<tr>
<td>Table 6.4</td>
<td>Failure Transient Significance, $\rho = 0.8, \varepsilon = 0.1, \delta = 0.001$</td>
<td>90</td>
</tr>
<tr>
<td>Table 6.5</td>
<td>Failure Transient Significance, $\rho = 0.8, \varepsilon = 0.1, \delta = 0.1$</td>
<td>91</td>
</tr>
<tr>
<td>Table 6.6</td>
<td>Failure Transient Significance, $\rho = 0.8, \varepsilon = 0.01, \delta = 0.01$</td>
<td>91</td>
</tr>
<tr>
<td>Table 6.7</td>
<td>Failure Transient Significance, $\rho = 0.8, \varepsilon = 0.006, \delta = 0.01$</td>
<td>92</td>
</tr>
<tr>
<td>Table B.1</td>
<td>Failure Transient, $\rho = 0.6, \varepsilon = 0.1, \delta = 0.01$</td>
<td>124</td>
</tr>
<tr>
<td>Table B.2</td>
<td>Failure Transient, $\rho = 0.8, \varepsilon = 0.1, \delta = 0.01$</td>
<td>125</td>
</tr>
<tr>
<td>Table B.3</td>
<td>Failure Transient, $\rho = 0.9, \varepsilon = 0.1, \delta = 0.01$</td>
<td>126</td>
</tr>
<tr>
<td>Table B.4</td>
<td>Failure Transient, $\rho = 0.8, \varepsilon = 0.1, \delta = 0.001$</td>
<td>127</td>
</tr>
<tr>
<td>Table B.5</td>
<td>Failure Transient, $\rho = 0.8, \varepsilon = 0.1, \delta = 0.1$</td>
<td>128</td>
</tr>
<tr>
<td>Table B.6</td>
<td>Failure Transient, $\rho = 0.8, \varepsilon = 0.01, \delta = 0.01$</td>
<td>129</td>
</tr>
<tr>
<td>Table B.7</td>
<td>Failure Transient, $\rho = 0.8, \varepsilon = 0.006, \delta = 0.01$</td>
<td>130</td>
</tr>
</tbody>
</table>
CHAPTER ONE
INTRODUCTION

1.1 Relationship Between Structure and Performance

A system's performance depends on its structure. As system resources fail and are repaired, the structure and performance of the system change. Much work has been done to quantify this relationship between structure and performance [32, 33, 59, 73, 91]. Dorrough [32, 33] develops the Performance Capability Measurement (PCM) as a system effectiveness measure for use in designing computer systems with redundancy. His primary concern is developing a measure that considers "not only the length of performance, but also the quality of performance." For each noncatastrophic system state, he determines the time-dependent probability of the system operating within allowable limits. The PCM is the expected value of this probability at time \( t \), taken with respect to the probabilities of state occurrence. He obtains this by weighting the time-dependent probability found for each noncatastrophic system state by the probability of the system being in that state at time \( t \) and summing across all states. This notion of averaging measures across sets of system states is the fundamental approach used in measuring the effectiveness of unreliable systems.

Meyer [59] rigorously formalizes this approach using a hierarchical system view to characterize the relationship between a system's performance and its structure over an interval of time. At the highest level, the system accumulates reward over time based on its performance. Since the performance, and thus the accumulated reward, may vary over time as the system structure changes, the system's effectiveness is measured by the expected worth of the system over the time interval. The next lower level quantifies system behavior in terms of "performability": the probability that the system performs within a given set of values over the time interval. At the lowest level, the system moves through a sequence of states defined in terms of the system and its environment.
Each level in the hierarchy defines random variables to map lower-level states into higher-level measures. This approach provides a powerful technique for relating structure and performance in a single measure of system effectiveness. See [73] for a brief review of some of its applications.

1.2 A Model of Network Structure

These types of approaches are sometimes used in analysis of communication networks. The structure of these networks is typically represented by a graph consisting of unreliable nodes and links (see Figure 1.1). The network structure changes as these resources fail and are repaired. A common model for representing this behavior uses the following random variable to indicate the state of the \(i\)th resource:

\[
r_i(t) = \begin{cases} 
0 & \text{if resource } i \text{ is down (failed) at time } t \\
1 & \text{if resource } i \text{ is up (operational) at time } t.
\end{cases}
\]

Over time each resource alternates back and forth between these two states. The resources are typically assumed to fail and be repaired independently of each other with constant rates for both these events, i.e., times are exponentially distributed. Under these assumptions, \(\{r_i(t), t \geq 0\}\) is a continuous-time Markov chain (CTMC), and it is possible to obtain the probability of a resource being in a particular state at time \(t\). To represent the equilibrium values of these probabilities, let

\[
p_i = \lim_{t \to \infty} P\{r_i(t) = 1\}. \tag{1.1}
\]

The \(n\)-tuple \(S(t)\) defines the state of the overall network in terms of the individual resource states as follows

\[
S(t) = (r_1(t), r_2(t), \ldots, r_n(t)).
\]

2
The network moves from state to state as resources fail and are repaired. Assuming constant failure and repair rates also makes \( \{S(t), t \geq 0\} \) a CTMC. The independence of failures and repairs implies the probability of the network occupying a given state is the product of the probabilities of the individual resources being in their respective states. All these probabilities may be time-dependent or steady state. The equilibrium probabilities are given by

\[
P(\lim \limits_{t \to \infty} S(t) = (s_1, s_2, \ldots, s_n)) = \prod_{l=1}^{n} \left[ s_l p_l + (1 - s_l)(1 - p_l) \right],
\]

where

\[ s_l = 0 \text{ or } 1. \]

1.3 The Steady-State Assumption

To measure the effectiveness of the network using either the time-dependent or equilibrium network state probabilities, first determine the performance in each state, and then average these state-dependent values over all the possible network states. Chapter Two presents a survey of network effectiveness measures obtained in this manner. They range from structurally-oriented measures, such as connectivity and availability, to capability-based measures related to availability of capacity, and (finally) to load-dependent measures, such as throughput and delay.

These load-dependent measures refer to networks with traffic on them. Following a change in the state of the network, there is a transient phase as the traffic converges to an equilibrium pattern for the new network configuration. In order to obtain a performance measure for a particular configuration, it is common to ignore this transient phase and to assume that the traffic is in equilibrium following a change in system state [21, 57, 86]. An early example of this appears in [60] as part of the analysis of a computer system with a finite input buffer and several parallel processors modeled as a finite M/M/n queue. The state of the system changes as the components fail. Failure
of the input buffer is catastrophic, whereas processor failures degrade system performance. To avoid a time-dependent solution for the state probabilities, the system is assumed to be in steady state immediately following processor failures. The large relative difference between component failure rates and the rates of job arrival and service is used to justify this steady-state assumption.

The same argument is made to justify the use of the assumption in the analysis of communication networks. However, in the computer system, the number of components was small, and system transients were due only to the instant change in traffic intensity following a processor failure. In communication networks, the change in traffic rates is not instantaneous following failures or repairs. The traffic rates change over a period of time following a network state change.

1.4 Illustration of Transients

One of these state changes occurs when a network link fails and calls using it can no longer transfer information. The network responds by rerouting these interrupted calls, but the process is not instantaneous. Time is needed to release the old connections, update routing tables to reflect the new state of the network, and establish new connections. As a result, network traffic rates change as rerouted calls begin appearing on the network. In addition, traffic on each of these calls is not in steady state. After rerouting, each call starts with a backlog of packets to transmit. Some of these are unacknowledged packets lost on the old connection at the time of failure, which the source must retransmit. There are also new packets that arrived during the rerouting process. Clearing this backlog for each call is necessary before the traffic process on the network can reach steady state. The many newly rerouted calls in a short amount of time, each starting with a backlog of packets, creates a surge of traffic that ripples through the network. As the backlogs clear and traffic on the rerouted calls settles down, the network has a new steady-state traffic pattern and the queues in the network converge to new equilibrium values.
Another type of network state change occurs when network links are repaired. Following a link repair, it may be possible to improve the performance of some of the calls existing on the network by rerouting them over the repaired link in order to shorten their connection or use lower utilized resources. However, this would interrupt service and require a complicated network communication protocol with global knowledge of call information. As a result, existing calls continue to use their initial route through the network. There are now two types of calls on the network: primary and secondary. Secondary calls would experience improved performance if allowed to reroute over the repaired link, whereas primary calls would experience no improvement. Since secondary calls are not rerouted over the repaired resource, each continues to use its current routing until it completes and releases its resources. Meanwhile, new calls may use the repaired link. The traffic pattern on the network changes as the number of calls using the repaired resource starts from zero and converges to its equilibrium distribution and the number of secondary calls decreases to zero as they terminate and disappear from the network.

1.5 Statement of Problem

As a result of these effects, there is a period of time when traffic rates change and traffic patterns settle to their equilibrium distributions following network state changes. The difference in rates between component failure rates and packet arrival and service rates may justify ignoring the transient phase following network state changes. However, there appears to be no work addressing how large the difference in rates must be to ensure that the transient phase is small relative to the time between changes in network state. In addition, the large number of network components increases the overall rate of failures and, thus, tends to reduce the difference between rates of network state change and rates of packet arrival and service. These issues suggest the need for work examining the validity of the steady-state assumption used in state-based measures of network effectiveness. To do this, we test the hypothesis that the length of the transient phase is significant relative to the
time between changes in network state. That is, the transient phase cannot be ignored when analyzing network performance.

Since we are investigating whether it is possible to ignore the transient phase following network state changes, we begin by surveying the existing work on communication network transients. Chapter Three shows that these results tend to focus on exact response, such as time-dependent occupancy distributions, and to assume an instantaneous change in rates. Bounds on the time necessary for the network to reach equilibrium exist only for small Jackson networks, and these also assume instantaneous rate changes.

In Chapter Four we examine a communication network in order to understand how it behaves following resource failures and repairs. We first study the operation of an appropriate set of communication protocols used in the network, laying out the steps, events, and actions taken following a failure or repair to identify components of the transient phase. We also make a number of assumptions about the traffic carried by the network. Many of these are taken from the network effectiveness literature reviewed in Chapter Two. Additional assumptions, such as those concerning call arrival processes and length distributions, add a new layer of complexity to the traffic models.

In order to test our hypothesis, we model network behavior following a failure or repair. Rather than obtaining exact network response, we are primarily interested in the time until the traffic patterns reach equilibrium following a change in the state of the network. Since this cannot happen until traffic rates stabilize, we use characteristics of the network, including network traffic processes and protocol operation, to bound the transient phase by determining the length of time during which these rates change following a failure or repair. For equilibrium network state behavior, we use the Markov model presented earlier, determining the dwell time for each network state from the failure and repair rates of the individual resources. Finally, we compare the lengths of these dwell times with the lower bound of the transient phase length to determine the significance of the transient phase relative to the time between changes in network state.
In Chapter Five we apply this approach to the case of link repairs. By using a death process to model the number of secondary calls, we obtain the distribution of a lower bound for the length of the transient phase. To determine whether the transient phase is significant, we derive the probability that the lower bound of the transient phase exceeds a given fraction of the time to the next change in network state. We find that a large relative difference between the rates of call arrivals and departures and the rates of individual component failure or repair is not sufficient for justifying use of the steady-state assumption for communication networks. It is also necessary to consider the average number of calls carried on a link and the overall number of components in the network.

Chapter Six covers the case of link failures. In order to obtain a lower bound for the length of the transient phase, we ignore the time necessary to detect the failure and reroute the affected calls. We then approximate the time necessary for traffic of a single call to reach equilibrium after it is rerouted. Next, we derive the probability that this time exceeds a given fraction of the time to the next change in network state. Assuming a Poisson-distributed number of rerouted calls, we obtain the probability that the transient for any call exceeds a given fraction of the time to the next change in network state. This is our measure of transient phase significance following link failures. We find that the probability of significance for the transient phase depends on the traffic intensity of packets on the network, the expected number of rerouted calls, how narrowly we define convergence for the approximate single-call model, the given fraction of time between network state changes, and the ratio of the rate of packet transmission to the rate of network state changes. Again, a large relative difference between the rates of packet arrival and transmission and the rates of individual component failure or repair is not sufficient for justifying use of the steady-state assumption for communication networks. Finally, Chapter Seven summarizes our findings and suggests areas for future work.
CHAPTER TWO
NETWORK EFFECTIVENESS MEASURES

Effectiveness measures provide a single overall performance measure for systems that can experience different levels of performance as the system configuration changes due to component failure and repair. These types of measures have found a number of useful applications in the area of communication networks. In this chapter, we briefly survey the development of several classes of effectiveness measures for networks and describe how they are used for network design.

We define three levels of effectiveness. Each generalizes the level below it by including another aspect of system behavior. The following questions about the system and its performance present these levels.

- Level One: Does the system perform at all?
- Level Two: Can the system meet a desired level of performance?
- Level Three: How well does the system perform?

To illustrate the levels of effectiveness, we ask these questions while restoring an old car. Initially, the car’s basic condition is the restorer’s primary concern. Will it start? Can it move under its own power? This is the first level of effectiveness. After the restorer has done enough work, the car almost always starts and moves under its own power. She can then focus on the car’s ability to perform at a certain level, and may ask, “Can it get to town and back without breaking down?” Eventually, perhaps after significant work, confidence in this level of performance increases and the next level of performance becomes relevant. How well does the car run? What will it do? Here, the focus is on quantitative aspects, such as the miles per gallon, the car’s ability to accelerate or
carry a load, and the distance it can travel between failures. Having restored the car, we now apply these levels of effectiveness to communication networks.

2.3 Level One - Structural Measures

Reliability of the network, expressed as its operational status, is the main concern at this level. The user is happy if the network is up and running. The design goal is to improve reliability, usually for an unspecified level of traffic performance. This may involve adding redundancy, swapping components with different reliabilities, or changing the configuration of the network. This level is most important for small networks or those with relatively unreliable components.

The quantities of interest at the structural level are typically graph-oriented measures. Examples of these include the percentage of nodes connected [12], the length of the shortest surviving path between each pair of nodes [36, 37], network connectivity, and cohesion. Connectivity is the minimum number of edges or nodes that must be removed to break all paths between any given pair of nodes. Network cohesion is the minimum number of links or nodes that must be removed to isolate any subgraph of nodes from the rest of the graph [87, 98]. Another measure is all-terminal reliability, which is the probability that all nodes are connected [98]. A more frequently used measure is two-terminal reliability: the probability that a connection exists between a given pair of nodes [3, 5, 11, 69, 70, 97]. Since structural measures are graph-oriented, determining them typically involves graph theoretic concepts such as cutsets, spanning trees, or shortest paths. As a result, combinatoric issues can arise while taking combinations of resources to enumerate all paths or cutsets.

The network design process often involves assigning links to connect a given set of nodes, usually under a constraint. For example, in laying out the network, there may be a constraint regarding the connectivity of the network [61, 78]. Sometimes the structure or topology of the network is
given and designers must maximize reliability by constructing it from a given set of links with varying reliabilities [44]. Another issue is determining where to make changes in the network to improve reliability [46, 62].

Structural measures view network components as being up or down, functioning or failed. If components have capacities associated with them, the total amount of capacity available varies as the components fail and are repaired. The next level of effectiveness addresses this added dimension.

2.4 Level Two - Capability Measures

This level of effectiveness focuses on system capability: the ability of the system to meet a specific performance objective. System components now have capacities or service rates. Although reliability is no longer the primary issue, it must still be considered. For example, network components may be more reliable or the network large enough so that it is almost always connected, but failures may affect the capability of the network so that it varies for a given level of reliability. Another feature of this level is that significant improvements in the old measure of reliability are no longer possible. For example, the network may have so many links that adding another one may not noticeably improve connectivity.

Capability measures show how the system capacity or its ability to do work varies with the state of the network. They are user independent and consider the system without user demands. The network model is a graph with capacities or service rates for the components. There is no traffic on the network and the system capacity varies as the components fail and become operational.
A performance threshold is the prime indicator of capability. The basic measure of capability is the probability that the system performs at or above this threshold, i.e., the probability that the system is capable. A more complete picture is given by the distribution of capability. A capability measure in the area of computer systems is computation capacity: the amount of useful computation per unit time available on the system [17].

The primary measure of capability for networks is the capacity existing between two given nodes. This could be in bits per second or in the number of channels available, depending on the type of network. One area of work involves the probability that capacity between two nodes exceeds a given threshold [4, 13, 56, 80]. This is sometimes used as a definition of network availability with the network being available if capacity exceeds the threshold. The mean capacity [2, 81, 93] and the distribution of capacity [22, 85] are also topics of interest.

Design at this level involves structuring the network and assigning capacities to the links or nodes. In this process, the concept of capacity availability is useful for determining circuit routing [18, 65], assigning capacities across network states [63, 101], and comparing system alternatives [71]. Approaches exist for assigning component probabilities to maximize network availability, defined in terms of capacity as above, subject to cost constraints or to minimize costs subject to availability constraints [13]. It is also possible to minimize costs subject to constraints on computation reliability, system reliability, and computation availability for computer systems [66].

2.5 Level Three - Load-Dependent Measures

Traffic now appears on the network, and effectiveness depends on how network changes affect it. Network components are highly reliable and there is very little probability of failures disconnecting the network.
The type of network determines the type of traffic. Circuit-switched networks establish circuits or channels between users. The primary issue is whether or not channels are available for arriving calls. Important measures for these networks include blocking probability [9, 54, 55], average lost call traffic [82, 83], and the effect of traffic on network availability [15].

Packet-switched networks receive traffic in the form of packets of information. The packets move through the network and may queue for transmission or processing. An arrival process and a packet-length distribution define the traffic. Usually a Poisson process generates packet arrivals, and their lengths are exponential and independently generated at each hop in the network. Arrival rates for traffic between all nodes in a network are given in the form of a matrix. Measures for these networks include the distribution of packet delay in the network, i.e., sojourn time [18, 21], average delay in the network [82], equilibrium throughput of the network (packets delivered per second) [48], and congestion [52, 53].

Burst-switching networks share characteristics of both circuit-switched and packet-switched networks. One measure for these networks is traffic loss rate (in bits per second or bursts per second not delivered to destination by network due to failure) [47].

Determining these measures often requires averaging over all the network states. The number of states can be quite large for some systems. A common approach for overcoming this problem is to examine only the most probable states [28, 50, 57, 84, 100]. Simulation is another technique for generating states to consider [49].

Design again involves structuring the network and assigning resource capacities, but under different constraints. Techniques exist to minimize lost traffic [27, 82, 83], and to minimize cost subject to constraints on delay performability [58]. Call-blocking probabilities as a constraint are considered in [38].
One of the primary difficulties associated with evaluating these load-dependent measures is that traffic on the network experiences transient behavior following a change in the state of the network. Rather than performing a time-dependent analysis, the common approach for overcoming this difficulty is to assume the network traffic is in steady state when determining the value of the performance measure for each network state. In order to justify this assumption, a claim is made that the rates of traffic-related events, such as call arrivals or packet transmissions, are large enough relative to component-related events, such as link failures, that the traffic transients quickly disappear. However, no work is cited to justify this claim or to indicate how large the rates must be for it to hold. Hoping to find such results, we searched the literature related to transient analysis of networks. The next chapter surveys our findings.
CHAPTER THREE
TIME-DEPENDENT AND TRANSIENT RESULTS FOR NETWORKS OF QUEUES

We searched the literature for results indicating when it is appropriate to ignore the transient response of a queuing network following failures. Although we did not find results directly addressing our problem, we found a sizable body of work related to time-dependent or transient behavior of networks of queues. This chapter presents these results.

The Kolmogorov forward differential equations governing the time-dependent behavior of a continuous-time Markov chain (CTMC) are

$$\frac{d}{dt} P(t) = P(t)Q$$

(3.1)

where $P(t)$ is a vector of which the $i$th component is the time-dependent probability of the network being in state $i$ at time $t$ and $Q$ is the rate matrix for the CTMC. When the system is in a steady-state condition, the distribution does not change over time. For the transient or time-dependent behavior of the system, the solution to the set of equations in (3.1) is

$$P(t) = P(0) e^{Qt}$$

(3.2)

where the definition of the matrix exponential is

$$e^{Qt} = \sum_{l=0}^{\infty} \frac{(Qt)^l}{l!}.$$
3.1 Numerical Techniques for Markov Systems

One approach for computing the matrix exponential uses eigenvalue techniques. Using a spectral representation of $Q$, the matrix exponential is

$$e^{Qt} = \sum_{i=0}^{\infty} \frac{(Qt)^i}{i!} = e^{t\lambda_1}B_1 + \ldots + e^{t\lambda_n}B_n$$  (3.4)

where the $\lambda_i$ are the eigenvalues of $Q$ [29]. Stern [89] uses this spectral decomposition approach to obtain a sum of exponentials, which he then truncates to approximate the time-dependent behavior of the occupancy distribution for the M/M/1 queue. For a network of queues, the state is usually defined as an $n$-tuple with the $i$th element representing the number of customers in the $i$th queue. Poisson arrivals and exponential service times make this network model a CTMC. For a network of finite exponential queues, Protonotarios and Aravantinos [68] describe how to number the states and construct the rate matrix. They then find the time-dependent behavior of the network by exponentiating the matrix $Q$ directly or using a numerical technique. For large networks, they recommend using sparse matrix techniques. However, some eigenvalue techniques tend to destroy the sparsity of the matrix [39, 72] making large problems too complex to solve.

An alternative to decomposition approaches is to directly solve for $P(t)$ using a variety of numerical techniques. Runge-Kutta integration [39, 72, 95] is a common approach. Implicit techniques [72] are also available, but they require solving a set of equations at every step, so that they are most efficient for stiff Markov chains, where the real parts of the eigenvalues of the rate matrix $Q$ have large differences in magnitude. One of the most flexible techniques for obtaining $P(t)$ is Jensen's method, also known as randomization or uniformization [39, 40, 72]. This approach embeds a discrete-time Markov chain into the continuous-time chain by adding self transitions to the states so that each has the same dwell time. The transient problem is then easily solved in discrete time. Harrison [43] presents a technique for closed networks that uses an iterative functional approximation that converges to the solution for the Kolmogorov differential equations.
3.2 Approximations for Markovian Networks

If the network queues have infinite capacity, the state space for the CTMC network model is infinite, making it difficult to solve directly for the state probabilities. Although these probabilities may be of interest themselves, they are often used to obtain values for system parameters such as the mean and variance of the number of customers in the system. It is sometimes possible to approximate these parameters directly by using the state probabilities to form a set of differential equations describing the behavior of the desired parameters. A closure approximation then reduces the infinite set of equations to a finite set that can be solved numerically or may even have a closed-form solution. Closed-form results obtained using this approach include time-dependent expressions for the average queue size in the M/M/1 queue [76], and the mean and variance of the number in the M/M/s queue [77]. For networks of exponential queues, results include finite sets of differential equations for the time-dependent mean and variance of the number at each station, which can be integrated numerically [8, 26].

Another approach is to bound the time a system parameter takes to reach steady state using a concept called relaxation time. This idea is based on observations that many Markov systems tend to converge exponentially to steady state [30, Section III.7.3; 64; 89]. A definition of the relaxation time is

\[ T(f) = \inf \{ T | f(t) - f(\infty) = O(e^{-T/t}), \quad t \to \infty \} \]

where \( f \) is a system measure such as the mean number in the system or the probability the system is empty. The relaxation time is also called the time constant for the system.

Blanc and van Doorn [20] define relaxation time using the probability that the system is empty. They present exact results for networks of M/M/\( \infty \) queues and Jackson networks with two single server queues. In addition, they make a conjecture about the form of relaxation times for arbitrary Jackson networks. In examining networks of M/M/\( \infty \) queues, Blanc [19] obtains time-dependent
results for the state probabilities and the mean number of jobs in the network. Using these, he defines the relaxation time in terms of the mean number of jobs in the network and derives time constants for arbitrary networks of M/M/∞ queues. He also presents upper bounds on the time until steady state for M/M/∞ networks when initially empty. In addition, he states results for relaxation times of tandem M/M/1 queues defined in terms of the probability that the system is empty. Finally, he conjectures a bound on the time for a Jackson network to reach steady state.

3.3 Numerical Techniques for General Queueing Networks

The diffusion approximation [34, 92] is a technique for general queues. It approximates a discrete-time queueing process with a continuous-time one, resulting in a partial differential diffusion equation subject to a set of boundary conditions. The solution to this equation approximates the time-dependent occupancy distribution for the queue. Duda [34] applies this technique to networks of general queues by looking at each queue in isolation. Given the flows from upstream queues, he applies the single-queue diffusion approximation model to find the queue length distribution, and then approximates the departure process. By applying the technique to each of the queues in the network, he obtains a new set of input flows. By repeating the procedure, he can iterate to a network solution.

3.4 Networks with Time-Varying Rates

Most of the work discussed above assumes that arrival and service rates are constant or at least constant over intervals between changes. The review in [92] contains results for single-queue models with time-varying rates. For networks of queues, Fan [35] simulates a network with time-varying rates by approximating the Kolmogorov equations with a set of difference equations,
whereas van As [95] integrated the equations using a fourth order Runge-Kutta technique. The approximations of [8, 26] also hold for time-varying rates since they integrate numerically.

For the network model we wish to examine, the traffic rates are discrete and non-stationary following network state changes. System rates are constant over intervals, but these intervals may be short immediately following repairs and failures of network resources. Although it may be possible to obtain their distribution as a function of time, there is no closed form for the rates as a function of time. Most importantly, the primary measure of interest is the time that it takes the network parameters to reach steady state following a change in network state, rather than the time-dependent network state probabilities following changes. Relaxation time results may be useful for bounding the transient phase, but results for networks are limited. In order to determine if we can apply them to our problem, we now take a closer look at the network and how it behaves following component failures and repairs.
CHAPTER FOUR

THE NETWORK:

ASSUMPTIONS, PROTOCOLS, AND TRANSIENT COMPONENTS

This chapter presents a packet-switched network providing virtual-circuit service. We begin by making assumptions about the generation of user calls and traffic and the general behavior of the network. We briefly discuss network protocols, which are the rules used to transfer information across the network. We then present a detailed description of how a particular set of protocols respond to failures in the network. General behavior following the repair of network resources is also examined. Finally, we define a set of time intervals that make up the transient phase following failure and repair of network resources.

4.1 Network Traffic Assumptions

We make a number of assumptions regarding traffic on the network. Calls arrive to the network according to a Poisson process that is independent of the state of the network and is always in steady state. This assumption is not always appropriate. For example, failure resulting from some catastrophe would probably be accompanied by additional traffic.

Arriving calls are randomly assigned to an origin/destination pair. If the origin or destination node is down, the call is lost. Otherwise a connection is established. Thus, calls are only lost or blocked when nodes are down. There is no blocking due to over-utilized resources. The call is up (remains active) for an exponentially distributed amount of time; then it completes and releases the connection. The number of calls carried by the network varies and is a random variable.
Connections are in one direction only, i.e., traffic flows in only one direction. Network connectivity is assumed to be high so alternate routes through the network always exist following failures.

It is necessary to assume that there is enough capacity available so link utilizations are less than one following rerouting of traffic. Without this, there will be no steady state. Also, all links have a fixed, though not necessarily identical, transmission rate. Following a link failure, there is no change in the number of calls on the network; all calls using the failed link are rerouted. There is also no blocking of new calls since other routes, which do not use the failed link, are available. As a result, the call process representing the number of active calls on the network remains in steady state following a link failure. On the other hand, node failures and repairs affect the number of calls on the network. Following a node failure, calls originating or terminating at the failed node are lost, whereas transit calls being switched at the failed node are rerouted. Arriving calls assigned to the failed node as an origin or destination are also lost. They do not wait to connect at the time of node repair.

Traffic on each connection consists of packets arriving according to a Poisson process. Packet lengths are exponentially distributed and independently redrawn at each hop in the network. Together with the fixed transmission rates, these latter assumptions lead to independent exponentially distributed transmission times. There is only a single type of traffic.

Papers in the performability literature using the steady-state assumption typically assume Poisson traffic between node pairs. The rates are assumed to be constant and are given in the form of a matrix. This is equivalent to having a fixed number of calls that are always on. Recent research concerned with correlated arrival processes has found that the Poisson assumption tends to give optimistic steady-state results [88]. It is unclear whether the use of Poisson arrivals will give optimistic results for the length of the transient phase relative to correlated arrivals. However, since we are investigating the applicability of the steady-state assumption for papers that assume the traffic is independent, we shall make the same assumption.
Figure 4.1 ISO OSI Reference Model
4.2 Network Protocols

When users call each other, the network establishes a connection between them called a virtual circuit. The path that this connection takes through the network from call origin to call destination is a route. At the time of call setup, a routing algorithm or routing tables at each node specify which network resources the call uses. Thus, the state of the network when the call is established determines the route used.

The network uses rules called protocols to transfer information between users [41, 90]. The International Standards Organization (ISO) has developed the Open Systems Interconnect (OSI) Reference Model in an attempt to provide common rules so equipment from different vendors can connect and work with each other. The Reference Model groups communication rules into seven layers based on their functionality. These layers sit on top of each other with protocols in one layer using protocols in the layer below it to carry out their functions. In establishing a circuit, the network creates software processes or entities within each layer. These communicate with each other to transfer information. One goal of the Reference Model is to hide the implementation details of each layer so that changes in implementing a protocol in one layer do not affect the operation of protocols in other layers. Some of the layers operate end-to-end and do not appear at intermediate nodes along the way. Others operate point-to-point and may appear at several locations along a connection. Figure 4.1 illustrates this structure.

What are the seven layers and what do they do? The application layer is closest to the user and governs how applications use the network. It addresses issues such as distributing work on different machines and making the network transparent to the user. The presentation layer focuses on representing data and contains functions for encryption, code conversion, and file conversion. The session layer manages the session between users. Its functions include synchronizing and initializing the user connection. The transport layer governs end-to-end transport of information through the network and focuses on reliable transportation across the network. Network layer issues concern
routing and congestion control. This layer contains functions for forwarding packets to adjacent nodes. Beneath it, the data link layer focuses on reliable transfer of information across transmission links. Finally, the physical layer defines bit transmission and concerns mechanical and electrical issues of the physical medium.

4.3 The Physics of Failure - Network Behavior Following Failures

Protocols control network behavior by specifying the actions that protocol entities take under different network conditions. Failures affect the communication of some of the protocol entities on the network. When an entity realizes that there is a problem, it tries to correct it. If it does not succeed, the entity notifies the next higher layer of the problem. We choose a representative set of protocols for the network and examine their behavior by identifying different failures that can occur within each layer and how the protocols respond to them. These protocols include X.21, X.25/LAPB, X.25, and X.224/Class 4 [23; 24, Section 2; 24; 25]. For clarity, we refer to local and opposite protocol entities relative to the location of the failure or the reacting entity. Local entities are at the same location, whereas opposite entities are generally on the opposite end of the link or connection.

Both immediate and delayed failures can occur at the physical layer. An immediate failure occurs when the interface to the data link layer enters an out-of-order state [24, Section 1.1.2 and 1.2.2]. The link layer entity immediately notifies the packet layer entity [24, Section 3.5, 4.6], which then begins clearing the call. A delayed failure is due to the loss of an incoming signal or a loss of alignment [24, Section 1.1.3, 1.2.3, and 1.3]. In this case, the link entity may delay for several seconds before considering the interface out of order and notifying the packet layer.
At the data link layer, operating under X.25/LAPB, there are total and partial failures. A total failure occurs if a link entity stops communicating. The opposite link entity then detects an idle link and waits T3 seconds before notifying the packet entity of an excessive idle channel [24, Section 2.3.5.5]. The value of variables such as T3 may be given by the protocol specification or fixed when the protocol is implemented. We assume that the local packet entity immediately detects the disappearance of the link entity and begins clearing the call. During a partial failure, the opposite link entity sees an active link, perhaps an interframe time fill, but receives no usable or expected frames. If the opposite link entity is waiting for an acknowledgement [24, Section 2.4.5.9], it waits T1 seconds before requesting an acknowledgement. It receives none, repeats this procedure N2 times, then tries to reset the link. After N2 attempts to reset the link, it notifies its local packet entity and enters the disconnect phase [24, Section 2.4.7.2]. The total time to notify the packet entity of the failure is 2(T2*N2). If the opposite link entity tries to reset the link immediately, the time to notify the packet entity will be T1*N2. Values for T1 and N2 are defined in X.25, Section 2.4.8.1 and 2.4.8.4. Typical values are three seconds for T1, ten times for N2, and 187 seconds for T3 [94].

There is apparently no mechanism in X.25 to detect the disappearance of a packet entity. Other packet-layer protocols exchange routing and congestion information between packet-layer entities [90]. For example, ARPANET updates every ten seconds, whereas DECNET updates every fifteen seconds or whenever there is a critical event. Failure to receive one of these updates could be interpreted as a failure and appropriate actions could be taken. Sometimes an entity generates inappropriate or undefined packets causing a partial failure. The opposite packet entity ignores some of these [24, Table C-1]. For others, the entity requests a restart or a clear and waits time T10 [24, Section 3.3.2, Annex D]. When time T10 expires, the entity resets and again waits T10. After T10 expires a second time, the entity clears the call. Annex D suggests a value of sixty seconds for time T10.
At the transport layer, protocol X.224 operating under class TP4 can take a variety of actions. When the transport layer entity receives a disconnect indication from the adjacent packet entity, it sets a time out and attempts to set up a new connection [25, Section 6.12]. Actions taken following a retransmission time out are similar to LAPB, but are carried out end-to-end [25, Section 6.19, Section 12.2.1.21]. A time out $T_1$ is set. (Note, this is a different timer than the $T_1$ used by the packet-layer entity.) After $N$ attempts, the transport entity waits an additional amount of time to allow for the packet to travel to the other end, be processed, and return before it releases the packet layer and notifies the session layer. Another action at the transport layer is inactivity control [25, Section 6.21, Section 12.2.3.1.1, Section 12.2.3.3]. The transport entities periodically send acknowledgements to each other to ensure the connection is operational. If a time out expires without receipt of an acknowledgement, the entity starts a release procedure. This will fail if the connection is down since the packet must reach the other end.

4.4 The Physics of Repair - Network Behavior Following Repairs

Consider the repair of a network resource. For purposes of illustration, assume no further failures or additional repairs occur. Existing calls are not using the newly repaired resource so it is initially idle. Each of these calls is primary or secondary relative to the repaired resource. Primary calls use their preferred route through the network; rerouting them to include the repaired resource would not improve their performance. Rerouting a secondary call over the repaired resource, however, would improve its performance by shortening the connection or using lower utilized resources.

Although it may improve performance, there is no rerouting of existing calls following resource repair. Primary calls already use their preferred route. Rerouting secondary calls would improve their performance, but would interrupt service and require a complicated routing protocol with global knowledge of the routes of all existing calls and the ability to determine which calls would
be improved by rerouting. Thus, secondary calls continue to use their original routing until the call completes and releases its resources.

After the time of repair, all new calls are primary. They choose the best route available in the network. Because there are no new secondary calls, the number of them decreases as they conclude and disconnect. In contrast, the number of primary calls using the repaired resource varies as calls arrive and complete, but tends to increase for a time since it starts at zero.

Before the network can reach equilibrium, the number of calls using each resource must be in steady state. The number using the repaired resource must reach steady state starting from zero, whereas the number using other resources reaches steady state when all the secondary calls disappear. Once the number of calls settles, traffic on these calls must reach steady state and the network must reach equilibrium for the new steady-state traffic flows.

4.5 Transient Phase Components

A transient condition exists whenever event rates, such as packet arrivals, or system-variable distributions, such as occupancy or departure-time distributions, change over time. Based on the behavior of the network following failures and repairs, we identify a number of transient phase components. Following a failure, these components are the time to

- F1. Detect the failure
- F2. Release and reroute a call using the failed resource
- F2A. Release a call using the failed resource
- F2B. Establish a connection over an alternate route
and the time for the

- F3. Traffic on a rerouted call to reach steady state.

The components following network repairs include the time for the

- R1. Number of primary calls to reach steady state

- R2. Traffic on a new primary connection to reach steady state.

How do these components relate to each other? First, they are not sequential for the network as a whole. For example, a call having several hops through the network may still be rerouting while a shorter call waits for traffic along its new route to reach equilibrium. Next, they may describe behavior of the network or an individual call. Network components include component R1 and component F1, which has the same value for all calls on the network. Components F2 and F3 are call components. They depend on each call's characteristics, such as the number of links in both the original and new routes. They may also depend on the window size used for flow control over the connection.

The network begins responding at the moment of failure. Traffic rates keep changing until traffic on the last call reaches steady state, i.e., completes F3. However, the rates tend to converge as more and more calls reach equilibrium. As a result, the network probably reaches equilibrium shortly after traffic on the last call reaches steady state. Also, as the number of rerouted calls with transient behavior dwindles, the remaining transient changes may be lost in the randomness of the call process arrivals and completions so the network may reach equilibrium before the last rerouted call does.
The transient phase following repairs is probably longer, but not as costly or disruptive as that following failures. The time-scale of component R1 depends on call lengths rather than the traffic that the calls carry. Since the call holding time is long compared to the packet interarrival and service times, R1 should last longer than the failure transients, which are primarily driven by packet parameters. Repairs do not interrupt calls so they are less disruptive.

Link and node failures affect the network differently. During link failures, the number of calls remains constant, but their routing may change. Node repairs and failures affect both the number and the routing of calls on the network. The number of calls changes as calls using that node as an origin or a destination appear or disappear.

The following chapters present models of network transient behavior. To simplify our problem, we only consider link failures and repairs. The models also assume that no further failures or repairs occur. In addition, we focus our effort on the one transient component that we consider most significant for each event. For link failures, this is component F3, the time for rerouted traffic to reach steady state on the new connection. Following link repairs, this is component R1, the time for the call process to reach steady state. Examining a single component simplifies the problem and leads to a lower bound for the network transient phase. Chapter Five considers transients following link repairs and Chapter Six considers transients following link failures.
CHAPTER FIVE
REPAIR TRANSIENT MODEL

Component R2 is the time for traffic on a new primary connection to reach steady state. A transient model with an instantaneous change in rates, in this case starting from zero, is appropriate here. However, assuming that the call reaches equilibrium before it completes, the length of this transient is probably small compared to the time necessary for the call process for component R1, the time for the number of primary calls to reach steady state. We therefore ignore component R2 and focus on component R1 as a model of transients following link repairs.

Recall that the call process is independent of the state of the network and is assumed to be in steady state always. We examine component R1's behavior following a link repair. Before the repair time, the link is unavailable, and calls preferring to use it must choose a secondary route. After the repair time, the link is available, and all new calls can use their primary routes. However, the existing secondary calls are not rerouted. The time until the number of primary calls using the link reaches steady state is then equivalent to the time until all secondary calls complete and release their resources, i.e., until there are zero secondary calls. We show this using the following circus analogy.

5.1 Elephants on parade

Elephants arrive at the center ring of a circus according to a Poisson process with rate \( \lambda \) and parade around the ring an exponential amount of time with rate \( \mu \) before leaving. The parade is quite exciting, especially to those noticing that prior to time \( T \), all the arrivals are African elephants, whereas after time \( T \), they are Indian elephants. If the number of elephants in the ring is in equi-
librium prior to time $T$, how long after time $T$ is it before the number of Indian elephants reaches its steady-state distribution?

To answer this question, let $N(t)$ be the number of elephants in the center ring at time $t$, and assume the birth and death process $\{ N(t), t \geq 0 \}$ is always in steady state. Define $N_d(t)$ as the number of African elephants in the center ring at time $t$ and $N_i(t)$ as the number of Indian elephants, then $N(t) = N_d(t) + N_i(t)$. Prior to time $T$, $N_d(t) = 0$ and $N(t) = N_i(t)$. After time $T$, there are no new arrivals of African elephants. The number of African elephants in the ring decreases as they leave until they are all gone so $N_d(\infty) = 0$ and $\{ N_d(t), t \geq T \}$ is a pure death process with 0 as an absorption state. Similarly, after time $T$, Indian elephants both arrive and depart, making $\{ N_i(t), t \geq T \}$ a birth and death process and, eventually, $N(t) = N_d(t)$.

Since $N(t)$ is assumed to be in steady state, $N_d(t)$ will be in steady state when $N(t) = N_i(t)$, i.e., when $N_d(t) = 0$. Therefore, the time for $\{ N_d(t), t \geq T \}$ to reach steady state is the absorption time of $\{ N_d(t), t \geq T \}$ starting at time $T$, or, equivalently, the first passage time of $\{ N_d(t), t \geq T \}$ to 0 after time $T$.

In relating this to the transient phase following link repairs, the elephants correspond to calls using or desiring to use the repaired link. African elephants represent secondary calls, and Indian elephants represent primary calls on the repaired link. The link becomes available at time $T$. The primary calls on the network before time $T$ can be ignored since they are not affected by the link repair.
Figure 5.1 Death Process for Secondary Calls
5.2 Absorption-Time Model

We now determine the absorption time for the transient continuous-time Markov chain (CTMC) used to model the time to steady state for component R1 following link repairs. Calls arrive according to a Poisson process with rate \( \lambda \) and have an exponential holding time with rate \( \mu \). Secondary calls use a secondary route through the network since the failed resource is unavailable. Because of the random assignment of arriving calls to origin/destination pairs, secondary calls arriving while the link is down can be thought of as arriving according to a Poisson process with rate \( \alpha \lambda \) where \( \alpha \) is the probability of an arriving call being assigned to an (O/D) pair requiring the use of a secondary route. Secondary calls stop arriving after the link becomes available and the time until the last one completes is the time until the number of primary calls on the link (and the network) reaches steady state.

The pure death process in Figure 5.1 models the number of secondary calls active after the link becomes available at time \( T \). If there are \( i \) secondary calls active when the link is repaired, the first completes after an exponentially distributed length of time with rate \( i\mu \) since the call lengths are exponential with rate \( \mu \) and the minimum of a set of exponential random variables is exponential with a rate equal to the sum of the individual rates. The remaining \( i - 1 \) calls then start over because of the memoryless property of the exponential distribution, and the next call completes after an exponential time with rate \( (i - 1)\mu \). This continues until there are no remaining calls. The absorption time (\( AT \)) of the process is then the sum of an exponential length of time with rate \( i\mu \), an exponential length of time with rate \( (i - 1)\mu \), and so on until the final length of time, which is exponential with rate \( \mu \). The distribution of the length of this sum is the convolution of the individual distributions and its Laplace-Stieltjes Transform (LST) is the product of the individual LSTs.
Given a random variable $Y$ with distribution $F_Y(y)$, we define its LST as

$$F_Y^*(s) = \int_0^\infty e^{-sy}dF_Y(y).$$

The LST of the distribution of the absorption time, given that there are initially $i$ calls, is then

$$F_i^*(s) = \left( \frac{i\mu}{s + i\mu} \right) \left( \frac{(i-1)\mu}{s + (i-1)\mu} \right) \ldots \left( \frac{2\mu}{s + 2\mu} \right) \left( \frac{\mu}{s + \mu} \right)$$

$$= \prod_{k=1}^i \frac{k\mu}{s + k\mu}, \quad i = 1, 2, \ldots$$

where $k\mu/(s + k\mu)$ is the LST of an exponential distribution with rate $k\mu$.

If $i = 0$, there are initially no calls and the time until absorption is 0, so $F_0(s) = 1$ and

$$F_i^*(s) = \begin{cases} 1, & i = 0 \\ \prod_{k=1}^i \frac{k\mu}{s + k\mu}, & i = 1, 2, \ldots \end{cases}$$

We now define

$$p_i = P\{i \text{ calls are initially active}\}.$$

By unconditioning on the number of initial calls, the LST of the distribution of the absorption time is

$$F_{AT}^*(s) = \sum_{i=0}^{\infty} p_i F_i^*(s). \quad (5.1)$$

To obtain the distribution, we must invert the transform. Although each of the $F_i^*(s)$ is given as a product of "easier" LSTs, the linearity property generally makes it easier to invert a summation than a product. This property states that the transform of the sum is the sum of the transforms,
so each term in the summation can be inverted individually. With this in mind, we first convert each $F_i^*(s)$ from a product into a summation using a partial fraction expansion, then invert $F_{i\tau}(s)$ term by term.

After expansion,

$$F_i^*(s) = \prod_{k=1}^{i} \frac{k\mu}{s + k\mu} = \sum_{k=1}^{i} \frac{B_k}{s + k\mu}, \quad i = 1, 2, \ldots \quad (5.2)$$

where

$$B_k = \left[ (s + k\mu) \prod_{j=1}^{l} \frac{j\mu}{s + j\mu} \right]_{s = -k\mu} = \frac{il\mu}{\prod_{j=1}^{l} (j - k), \quad j \neq k}.$$ 

By splitting the product in the denominator into two parts, the first from 1 to $k - 1$ and the second from $k + 1$ to $i$, then substituting a variable for $j - k$, we obtain

$$\prod_{j=1}^{l} (j - k) = \prod_{j=k+1}^{l} (j - k) \prod_{j=1}^{k-1} (j - k) = \prod_{\ell=1}^{l-k} \prod_{\ell=1}^{k-1} - \ell = (i - k)/(k - 1)!(-1)^{k-1}.$$ 

Now,

$$B_k = \frac{il\mu}{(i - k)/(k - 1)!(-1)^{k-1}} = \left( \frac{il}{(i - k)!k!} \right) \left( \frac{k\mu}{(-1)^{k-1}} \right) = \left( \frac{i}{k} \right)(-1)^{k-1}k\mu.$$ 

35
Using this and substituting Equation (5.2) into Equation (5.1) yields

\[ F^*_A(s) = p_0 + \sum_{l=1}^{\infty} p_l \sum_{k=1}^{l} \binom{i}{k} (-1)^{k-1} \frac{k \mu}{s + k \mu}. \]

After interchanging the order of summation and grouping terms, we obtain

\[ F^*_A(s) = p_0 + \sum_{k=1}^{\infty} A_k \left( \frac{k \mu}{s + k \mu} \right) \]  \hspace{1cm} (5.3)

where

\[ A_k = (-1)^{k-1} \sum_{i=k}^{\infty} \binom{i}{k} p_i, \quad k = 1, 2, \ldots \]

We can now easily invert the transform term by term since \( k \mu/(s + k \mu) \) is the LST of an exponential distribution with rate \( k \mu \). The resulting distribution of the absorption time is

\[ F_A(t) = p_0 + \sum_{k=1}^{\infty} A_k (1 - e^{-k \mu t}), \quad t \geq 0 \]  \hspace{1cm} (5.4)

where

\[ A_k = (-1)^{k-1} \sum_{i=k}^{\infty} \binom{i}{k} p_i. \]
Since secondary calls arrive according to a Poisson process with rate $\alpha \lambda$ and have exponential holding times with rate $\mu$, the distribution of the number of active calls in steady state (and thus when the link becomes available at time $T$) is identical to the occupancy distribution of an $M/M/\infty$ queue. This distribution is [45]

$$p_i = \frac{(\alpha \lambda / \mu)^i}{i!} e^{-\alpha \lambda / \mu}, \quad i = 0, 1, 2, \ldots$$

with the expected number being

$$N = \alpha \lambda / \mu.$$

Using this in Equation (5.4), we have

$$A_k = (-1)^{k-1} \sum_{i=k}^{\infty} \frac{N^i}{i!} \frac{N^{i-k}}{(i-k)!} e^{-N}$$

$$= (-1)^{k-1} e^{-N} \frac{N^k}{k!} \sum_{i=k}^{\infty} \frac{N^{i-k}}{(i-k)!}$$

$$= \frac{(-N)^k}{k!}, \quad k = 1, 2, \ldots$$

Then

$$F_{AT}(t) = e^{-N} + \sum_{k=1}^{\infty} \frac{(-N)^k}{k!} (1 - e^{-k \mu t})$$

$$= e^{-N} - \sum_{k=0}^{\infty} \frac{(-N)^k}{k!} + \sum_{k=0}^{\infty} \frac{(-N e^{-\mu t})^k}{k!}$$

(5.5)

$$= e^{-N e^{-\mu t}}, \quad t \geq 0$$

where $N$ is the average number of secondary calls active at time $T$ when the link is repaired. This result has the form of an extreme-value distribution and is intuitively satisfying. The time until absorption is the length of the longest call since the number of secondary calls goes to zero when the last one completes.
Using an extreme-value approach, a simpler derivation of Equation (5.5) is possible. Begin by conditioning on the number of secondary calls at time $T$,

$$F_{AT}(t) = \sum_{i=0}^{\infty} p_i F_i(t)$$

where $F_i(t)$ is the probability that the longest of $i$ initially active calls terminates by time $t$. This happens only when all $i$ calls terminate by time $t$. For $i$ exponential calls with rate $\mu$

$$F_i(t) = [1 - e^{-\mu t}]^i, \quad t \geq 0.$$  

Substituting for $F_i(t)$ and $p_i$,

$$F_{AT}(t) = \sum_{i=0}^{\infty} \frac{N^i}{i!} e^{-N} [1 - e^{-\mu t}]^i, \quad t \geq 0$$

$$= e^{-N} \sum_{i=0}^{\infty} \frac{[N(1 - e^{-\mu t})]^i}{i!}$$

$$= e^{-N} e^{N(1 - e^{-\mu t})}$$

so

$$F_{AT}(t) = e^{-Ne^{-\mu t}}, \quad t \geq 0.$$  

This is the distribution of the length of a lower bound for the length of the transient phase following link repairs.
5.3 Significance of Transient Phase Following Link Repairs

We now use this distribution to compare the length of the transient phase with the time to the next change in network state. From Chapter One, the CTMC model of the network, \( \{ S(t), t \geq 0 \} \), implies this time is exponentially distributed. (We assume with rate \( \beta \).) In order to ignore the transient phase \( (TP) \), the network state dwell time \( (S) \) must be large relative to the transient phase. How large? Say at least \( \kappa \) times as large, or

\[
S \geq \kappa \cdot TP.
\]

The probability that the steady-state assumption is valid is then

\[
P(S \geq \kappa \cdot TP) = P(TP \leq \delta S)
\]

where

\[
\delta = \frac{1}{\kappa}.
\]

The probability that the transient phase is significant is then

\[
P(TP > \delta S) = 1 - P(TP \leq \delta S).
\]

We next determine

\[
P(TP \leq \delta S)
\]

where

\( TP = \) length of transient phase following link repair
\( S = \) time until the next change in network state

and \( \delta \in [0, 1] \).
Conditioning on the time until the next state change,

\[ P(TP \leq \delta S) = \int_0^\infty P(TP \leq \delta S \mid S = t) dF_S(t) \]

\[ = \int_0^\infty P(TP \leq \delta t) dF_S(t) \]

\[ = \int_0^\infty F_{TP}(\delta t) dF_S(t) \]

where

\[ F_{TP}(t) = e^{-Ne^{-\mu t}}, \quad t \geq 0 \]

and

\[ dF_S(t) = \beta e^{-\beta t} dt, \quad t \geq 0. \]

After substituting,

\[ P(TP \leq \delta S) = \int_0^\infty \beta e^{-\beta t} e^{-Ne^{-\mu t}} dt. \]

Expanding \( e^{-Ne^{-\mu t}} \) in a power series, then interchanging the order of integration and summation yields

\[ P(TP \leq \delta S) = \sum_{i=0}^\infty \frac{(-N)^i \beta}{i!} \int_0^\infty e^{-(\beta + i\delta \mu) t} dt. \]
After integrating and rearranging terms,

\[ P(TP \leq \delta S) = \sum_{l=0}^{\infty} \frac{(-N)^l}{l!} \left( \frac{\frac{1}{\delta \mu}}{1 + i \delta \frac{\mu}{\beta}} \right) . \]

A closed form for this summation is given in [42] as

\[ \sum_{k=0}^{\infty} \frac{1}{k!(ka + b)} (-Z)^k = \frac{1}{a} \left( -\frac{b}{a} \right)^{-\frac{b}{a}} \gamma\left( \frac{b}{a}, Z \right) . \]

Substituting yields

\[ P(TP \leq \delta S) = uN^{-u} \gamma(u, N) \quad (5.6) \]

where

\[ u = \frac{\beta}{\delta \mu} = \frac{1}{\delta R} \]

and

\[ R = \frac{1/\beta}{1/\mu} = \frac{\mu}{\beta} \]

is the ratio of the mean time between network state changes to the mean call holding time. It is also the ratio of the rate of call completions to the rate of network state changes. This value is our measure of the relative difference in rates commonly used to justify the steady-state assumption. Its effect on the probability of transient-phase significance is of primary interest to us. The definition of the incomplete Gamma function is
\[ \gamma(u, x) = \int_0^x e^{-t} t^{u-1} \, dt. \]

To evaluate Equation (5.6), we follow the approach of [51]. Beginning with the following recursion from [1]

\[ \gamma(u+1, N) = u \gamma(u, N) - N^u e^{-N}. \]

Rearrange to obtain

\[ \gamma(u, N) = \frac{N^u e^{-N}}{u} + \frac{1}{u} \gamma(u+1, N). \quad (5.7) \]

Successively substituting (5.7) into its own right-hand side yields

\[
\begin{align*}
\gamma(u, N) &= \frac{N^u e^{-N}}{u} + \frac{1}{u} \left[ \frac{N^u + e^{-N}}{u + 1} + \frac{1}{u + 1} \left[ \frac{N^u + 2e^{-N}}{u + 2} + \ldots \right] \right] \\
&= \frac{N^u}{u} e^{-N} \left[ 1 + \frac{N}{u + 1} + \frac{N^2}{(u + 1)(u + 2)} + \ldots \right] \\
&= \frac{1}{u} N^u e^{-N} \sum_{i=0}^{\infty} C_i(u, N)
\end{align*}
\]

where

\[ C_0(u, N) = 1 \quad \text{and} \quad C_i(u, N) = \frac{N}{u + i} C_{i-1}(u, N). \]

Substituting this result into (5.6) yields

\[
P(TP \leq \delta S) = u N^{-u} \left[ \frac{1}{u} N^u e^{-N} \sum_{i=0}^{\infty} C_i(u, N) \right] \]

\[ = e^{-N} \sum_{i=0}^{\infty} C_i(u, N) \]

\[
(5.8)
\]
where

\[ u = \frac{1}{\delta R}, \quad R = \frac{\mu}{\beta}, \quad C_0(u, N) = 1 \quad \text{and} \quad C_i(u, N) = \frac{N}{u + i} C_{i-1}(u, N). \]

The Pascal computer program listed in Appendix A uses Equation (5.8) to determine the probability that the transient phase following a link repair is significant relative to the time until the next change in network state, i.e.,

\[ P\{TP > \delta S\} = 1 - P\{TP \leq \delta S\}. \quad (5.9) \]

Tables 5.1, 5.2 and 5.3 present the probability of transient-phase significance over a range of values for \( N \) and \( R \) obtained using this program for given values of \( \delta \). Figures 5.2, 5.3 and 5.4 chart these probabilities for their respective tables. Recall that \( N \) is the expected number of secondary calls and \( R \) is the ratio of the rate of call completions to the rate of network state changes.

To begin, we must consider what level of probability indicates that the transient cannot be ignored. For purposes of illustration, we choose 10%. That is, we consider the steady-state assumption inappropriate if the probability that the length of the transient phase exceeds a fraction \( \delta \) of the time to the next network state change is greater than 10%. There is nothing special about choosing 10% other than that it seems neither too big nor too small. The choice would also depend on the value of \( \delta \) and the user’s desired level of certainty.

For our baseline case, we choose \( \delta = 0.01 \). This value of \( \delta \) means that the length of the transient phase must exceed 1% of the length of time until the next change in network state in order to be significant. This value seems reasonable without being too restrictive. The probabilities for this case appear in Table 5.2. (The program experienced underflow problems for \( N > 10000 \). In order to get a clearer picture of how the curves behave for the baseline case, two more rows of data were obtained for Table 5.2 using a short, but long-running, Mathematica program.)
Table 5.1 Repair Transient Significance, delta=0.001

<table>
<thead>
<tr>
<th>N</th>
<th>R = 1</th>
<th>R = 10</th>
<th>R = 100</th>
<th>R = 1K</th>
<th>R = 10K</th>
<th>R = 100K</th>
<th>R = 1M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63175268</td>
<td>0.62844213</td>
<td>0.59565922</td>
<td>0.36787944</td>
<td>0.07160280</td>
<td>0.00787781</td>
<td>0.000079571</td>
</tr>
<tr>
<td>10</td>
<td>0.99995414</td>
<td>0.99980396</td>
<td>0.99980329</td>
<td>0.90000454</td>
<td>0.24431564</td>
<td>0.02830781</td>
<td>0.00287484</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99000000</td>
<td>0.39973824</td>
<td>0.05042618</td>
<td>0.00516816</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99900000</td>
<td>0.52319514</td>
<td>0.07204112</td>
<td>0.00745621</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99990000</td>
<td>0.62126044</td>
<td>0.09316405</td>
<td>0.00973900</td>
</tr>
</tbody>
</table>

Table 5.2 Repair Transient Significance, delta=0.01

<table>
<thead>
<tr>
<th>N</th>
<th>R = 1</th>
<th>R = 10</th>
<th>R = 100</th>
<th>R = 1K</th>
<th>R = 10K</th>
<th>R = 100K</th>
<th>R = 1M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.62844200</td>
<td>0.59565900</td>
<td>0.36788000</td>
<td>0.07160290</td>
<td>0.00787792</td>
<td>0.00079582</td>
<td>0.00007976</td>
</tr>
<tr>
<td>10</td>
<td>0.99995000</td>
<td>0.99980300</td>
<td>0.90000500</td>
<td>0.24431530</td>
<td>0.02830780</td>
<td>0.00287484</td>
<td>0.00028793</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99900000</td>
<td>0.39973800</td>
<td>0.05042620</td>
<td>0.00516816</td>
<td>0.00051810</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99900000</td>
<td>0.52319500</td>
<td>0.07204110</td>
<td>0.00745621</td>
<td>0.00074821</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.62126000</td>
<td>0.09316400</td>
<td>0.00973900</td>
<td>0.00097827</td>
</tr>
<tr>
<td>100000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.69915600</td>
<td>0.11380600</td>
<td>0.01201650</td>
<td>0.00120828</td>
</tr>
</tbody>
</table>

Table 5.3 Repair Transient Significance, delta=0.1

<table>
<thead>
<tr>
<th>N</th>
<th>R = 1</th>
<th>R = 10</th>
<th>R = 100</th>
<th>R = 1K</th>
<th>R = 10K</th>
<th>R = 100K</th>
<th>R = 1M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.59565922</td>
<td>0.36787944</td>
<td>0.07160280</td>
<td>0.00787781</td>
<td>0.00079571</td>
<td>0.00007965</td>
<td>0.0000797</td>
</tr>
<tr>
<td>10</td>
<td>0.99980329</td>
<td>0.90000454</td>
<td>0.24431564</td>
<td>0.02830781</td>
<td>0.00287484</td>
<td>0.00028793</td>
<td>0.0002880</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>0.99900000</td>
<td>0.39973824</td>
<td>0.05042618</td>
<td>0.00516816</td>
<td>0.00051810</td>
<td>0.0005182</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>0.99900000</td>
<td>0.52319514</td>
<td>0.07204112</td>
<td>0.00745621</td>
<td>0.00074821</td>
<td>0.0007485</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>0.99950000</td>
<td>0.62126044</td>
<td>0.09316405</td>
<td>0.00973900</td>
<td>0.00097827</td>
<td>0.0009787</td>
</tr>
</tbody>
</table>
$N = \text{Average Number of Calls}$

Figure 5.2 Probability Repair
Transient is Significant,
delta=0.001
Figure 5.3  Probability Repair
Transient is Significant,
\[ \text{delta}=0.01 \]
$N = \text{Average Number of Calls}$

Figure 5.4  Probability Repair  
Transient is Significant,  
delta=0.1
Figure 5.3 plots the probabilities of Table 5.2 against $N$ for each value of $R$. Increasing $R$ lowers the probability of significance. The higher ratio increases the mean time between network state changes as measured in mean call holding times. As a result, the length of the transient phase shrinks with respect to the time between network state changes and the probability of significance decreases. As $N$ increases for a given value of $R$, there are more secondary calls, and it takes longer for them to disappear from the network. This increases the length of the transient phase relative to the time between network state changes, increasing the probability that the transient phase cannot be ignored.

Note that a large value of $R$ does not prevent the transient phase from being significant. The figure shows that the probability of significance can exceed 10% even when $R = 10,000$. The number of affected calls, i.e., the amount of network traffic, is also a factor. In addition, note that we have defined $R$ in terms of the rate of network state changes rather than failure and repair rates for individual network components. It reflects the number of components in the network, i.e., the size of the network. The point is that a large relative difference between the rates of call arrival or completion and the rate of failure for a single link is not a sufficient condition for ignoring the transient phase.

The figure also shows that when $R \geq 10,000$, an order of magnitude increase in $R$ reduces the probability of significance of the transient phase by an order of magnitude, shifting the curve down. Also, for $R \geq 10,000$, Table 5.2 shows that increasing $N$ by an order of magnitude increases the probability by approximately $200/R$. This implies that the probability of significance increases almost linearly with $N$. It is not clear how far this relationship would hold, although for $R = 10,000$ it seems to hold for probabilities as high as 0.13. These patterns appear in all three sets of data with the pattern values changing for each case. Interestingly, the respective values appear to be $R \geq 100/\delta$ with a proportionality constant of $2/(\delta R)$. Expressing these “rules of thumb” in terms of $u$: whenever $0.01 \geq u$, the probability of significance decreases by an order of magnitude with an order of magnitude decrease in $u$, and increases approximately linearly in $N$ with pro-
portionality constant $2u$. The source of these effects does not appear to be obvious from the forms of Equations (5.6), (5.8) and (5.9).

Tables 5.1, 5.2 and 5.3 illustrate the effect of increasing $\delta$. Notice that increasing $\delta$ by an order of magnitude shifts the data in the table one column to the left so that it now corresponds to a value of $R$ that is an order of magnitude less than before. Since $\delta$ and $R$ always appear together in Equations (5.6) and (5.8), the effect of increasing $\delta$ by an order of magnitude should be the same as increasing $R$ by an order of magnitude. However, here there is no threshold value for $R$. All the curves shift. Actually, the curves do not shift: their relative spacing remains fixed, but their labels shift. Another way to look at this is that Figure 5.3 represents the set of $R$-curves for all values of $\delta$. Changing $\delta$ merely changes the scale of the probability axis.

Finally, we note that these results are not as useful to the network designer as may at first appear. They represent the probability that the transient phase is significant and indicate when the transient phase cannot be ignored. This is not the same as indicating when it is safe to use the steady-state assumption. It is possible that the transient phase is significant, even though the probability of the lower bound being significant is low. However, the results are still important because they indicate when it is not possible to ignore network transients following a link repair. They also show that a large difference in the relative rates of call events and individual component events is not sufficient for justifying use of the steady-state assumption. The amount of network traffic and the size of the network must also be considered. Finally, although they may not provide sufficient conditions for using the steady-state assumption, they can provide some measure of when it may be safe to use the assumption. For example, if the probability of significance is $10^{-4}$, there may be less risk than when it is $10^{-3}$. 
Figure 5.5 Example Network
5.4 Example

As an example of how network designers can use Equations (5.8) and (5.9), consider the network shown in Figure 5.5 from [62]. Since each undirected link represents two directed links, the network consists of 10 nodes and 28 links. Using data from [99], assume that the links are optical fibers with a mean time to failure of 2 million hours and the nodes are packet switches with a mean time to failure of 200,000 hours. Mean repair time is 7 hours. All failures and repairs are independent and occur with fixed rates; their times are exponentially distributed. The limiting probabilities defined by Equation (1.1) are then:

\[ P_N = P(\text{Node is up}) = \frac{200,000}{200,007} = 1 - 3.5 \times 10^{-5}. \]

\[ P_L = P(\text{Link is up}) = \frac{2,000,000}{2,000,007} = 1 - 3.5 \times 10^{-6}. \]

Using Equation (1.2),

\[ P(\text{Network fully operational}) = (p_N^{10}) (p_L^{28}) = 0.999552. \]

This is the most probable state of the network. We now investigate the transient following the repair of a failed link, which returns the network to this state. In the most probable state, the rate at which the next change in network state occurs is

\[ \beta = \frac{10}{200,000} + \frac{28}{2,000,000} = 6.4 \times 10^{-5} \text{ failures per hour}. \]

If call lengths are exponentially distributed with a mean of 3 minutes,

\[ \mu = 20 \text{ call terminations per hour}. \]
Then

\[ R = \frac{\mu}{\beta} = 312500 \text{ call terminations per failure}. \]

Notice that the ratio of the rate of call terminations to the rate of failures for a single link is \(4 \times 10^7\). Taking the size of the network into account in \( R \) reduces the ratio of event rates by two orders of magnitude. From Equation (5.9) with \( N = 50,000 \) and \( \delta = 0.01 \), the probability that the transient phase is significant following a link repair, which returns the network to its most probable state, is 0.00364. The probability of having a single link failed is less than the probability of having a single component failed, which is \( 4.48 \times 10^{-4} \).

For transients following repairs when there is still another component failed,

\[ \beta = \frac{1}{7} + \frac{9}{200,000} + \frac{28}{2,000,000} = 0.142916 \text{ events per hour} \]

if the remaining failed component is a node and

\[ \beta = \frac{1}{7} + \frac{10}{200,000} + \frac{27}{2,000,000} = 0.14292 \text{ events per hour} \]

if the remaining failed component is a link. We use \( \beta = 0.142916 \) to consider both cases, making

\[ R = \frac{\mu}{\beta} = 140 \text{ call terminations per event}. \]

Using \( R = 140 \) and \( N = 50,000 \), Equation (5.9) gives the probability that the transient phase is significant following repairs with a remaining component failed as 0.999599. We note that this situation only occurs following a link repair with another component failed. The probability of having two failed components, one of which is a link, is less than the probability of having two components failed which is \( 9.4 \times 10^{-9} \). Although the transient phase is significant following this repair, the probability of this situation occurring may not make it worth considering.
CHAPTER SIX
FAILURE TRANSIENT MODEL

This chapter focuses on the development of a model for transient component F3, the time for traffic on a newly rerouted call to reach steady state. Once the new connection is established, users have a backlog of packets to transmit. Until the user clears this backlog, traffic on the call does not reach steady state. Some of these packets are unacknowledged packets that were on the network at the time of failure. Others arrived while the network was rerouting the interrupted call. The user cannot dump all these packets on the network at once because of the network congestion control mechanism.

6.1 Network Congestion Control

Congestion can cause delays or even deadlock in the network. To control congestion, the network uses a mechanism to prevent users from overloading the network. One approach is sliding-window flow control [90]. This mechanism limits the number of packets on the network by giving each source a number of containers used to send packets. Once these are gone, the user must wait until they return before sending additional packets. The number of containers, i.e., the window size, may be fixed or allowed to vary with network traffic conditions. Another approach is a rate-based mechanism, such as the leaky bucket [10]. This mechanism generates containers at a fixed rate at the source. If no containers are available, packets must wait. However, if no packets are waiting, the user may save only a few containers. Regardless of the mechanism used, the backlog must be cleared.
We assume that the network uses a sliding-window flow-control mechanism. Models of this mechanism typically use tandem queues to represent a path through the network for both steady-state [67, 74, 75] and transient analysis [7, 96]. These models must consider both traffic for the modeled path and traffic from other paths in the network. Some models assume path arrivals are Poisson, but differ in their treatment. For example, loss models [67, 74] discard arriving packets if a container is not available, whereas wait models [6, 7] queue packets until containers are available. Another approach assumes a packet is always available at the source waiting for the next container to arrive [96]. We refer to such models as packet-available models. When considering interfering traffic from other paths in the network, some approaches model it explicitly [67] or account for it by altering the service rate [6, 75] whereas others ignore it [96].

There has been little work on transient analysis of window flow-control mechanisms. In [7], simulation is used to compare the average total delay for fixed and dynamic window sizes for time-varying traffic. In [96], a deterministic analysis, which ignores interfering traffic, is used to study the dynamic behavior of queues along the path. We model the transient behavior of the window flow-control mechanism using a model similar to the one used for steady-state analysis in [67].

This queueing model is shown in Figure 6.1. It has mixed traffic. Open traffic arrives at each node according to independent Poisson processes and departs the network after receiving service at a single node. This represents traffic from other paths in the network. Penoti [67] labels these external customers. Closed traffic represents traffic on the modeled path. It circulates along a closed loop connecting the queues along the modeled path. In [67] these closed customers are called link customers.
Figure 6.1 Mixed Queueing Model
We use several of the assumptions made in [67] to analyze this model. All packet lengths are drawn from the same exponential distribution and are chosen independently at each queue. Transmission rates are constant, implying that packets have exponential transmission times. Queueing occurs only for transmission; there is no nodal processing delay. Also, there are no transmission errors or packet losses.

Because we are interested in the transient behavior of the system during the backlog, we make a number of additional assumptions. First, since the system is in backlog, there is always a packet waiting for transmission. We have a packet-available model so we eliminate the wait queue used in [67]. As a result, containers reenter the network as soon as they become available at the source. Also, containers are available back at the source immediately after arriving at the destination.

The closed customers initially arrive at the first queue as a batch at time 0 to represent a full window. These customers then spread out as they circulate along the path and settle into an equilibrium pattern, which can be found using the technique presented in [14]. Penot and Schwartz [67] used this technique in their steady-state analysis. They define the network state as $(n, m)$ where the $i$th component of vector $n$ represents the number of link or closed customers in the $i$th queue and the $i$th component of the $m$ vector represents the number of external or open customers in the $i$th queue. Assuming that the network is in equilibrium and having Poisson arrivals for both types of customers implies that the customers are randomly mixed in each queue and allows the authors to obtain a probability for the type of customer in service.

For a transient analysis, this probability is not constant so the state-variable definition must include information about the type of customer in service and the ordering of customers within each queue. Consider a system with one closed customer and two open customers. It behaves differently when the customer ordering is closed, open, open than when it is open, closed, open. To include this information, we define the following general system state

$$S(t) = (n, r).$$
The $i$th component of the $n$ vector represents the number of customers, both open and closed, in the $i$th queue. The $i$th component of the $r$ vector is an ordered pair. The first component indicates the queue containing the $i$th closed customer and the second component is the number of customers, both closed and open, ahead of it in that queue.

It is still possible to use a continuous-time Markov chain to represent the behavior of the system based on this state-variable definition. Numerical techniques, such as Jensen's method, could then be used to analyze its transient behavior. However, the size of the state space rapidly increases. For example, two queues, each with a capacity of five customers, including the customer in service, and three circulating closed customers requires 36,000 states. Fortunately, the state occupancy probabilities are not needed since the measure of interest is the time it takes the system to settle down.

To determine this time, we use an output-process approach to analyze the mixed packet-available queueing system of Figure 6.1. The server in each queue alternates between periods of serving open customers and closed customers. As will be shown, the distribution of the lengths of these periods converges as the system reaches equilibrium. By characterizing each period and the number of them needed for convergence, we can obtain the distribution of the time until the system reaches equilibrium. Applying this model to the flow-control mechanism gives the distribution of the time until the window mechanism reaches equilibrium, given that there is always a backlog of packets awaiting transmission. This is not the same as component F3, which includes the time to clear the backlog followed by the time for the containers to settle into their equilibrium distribution along the path for the average call-packet arrival rate. However, for reasons we discuss below, we use it as our model of component F3.

While the backlog exists, the flow-control window is always full and there is always a packet waiting for a container. Thus, the packet-available queueing system models the behavior of component F3 until the backlog clears at which point a wait model is more appropriate. In the wait
model, call-packet arrivals are Poisson and in steady state. If no container is available, the packets wait. When the container arrives and finds packets waiting, it immediately reenters the system as in the packet-available model. If no packets are available, the container must wait for one before reentering the network. This is equivalent to entering a queue for closed customers only with a service rate equal to the packet arrival rate as is done in the loss models [67].

The initial distribution of closed customers along the path is the same for both models, that of the packet-available model at the moment the backlog clears. This distribution must then settle to its equilibrium distribution. Comparing the time required for this convergence in both models determines the appropriateness of using the packet-available model to represent transient component F3.

The relationship between convergence times for the packet-available model and the wait model for component F3 after the backlog clears is driven by the average packet arrival rate for the call. The probability the container does not have to wait is directly proportional to the call-packet arrival rate. For high arrival rates, it becomes more likely that packets are waiting and the wait model reduces to the packet-available model. If it is large enough, the flow-control mechanism could reach steady state before the backlog clears. In this case, the packet-available model would underestimate the length of component F3. If the backlog clears before the system reaches equilibrium, the behavior of the wait model after the backlog clears will be comparable with that of the packet-available model since it is unlikely that the container must wait for a packet. Of course, this probability is time-dependent so modeling it explicitly increases the complexity of the wait model.

For moderate arrival rates, the wait model looks like the packet-available model with an additional intermittent queue for closed customers. It is intermittent in the sense that sometimes it is there and sometimes it is not. If call packets are waiting when the containers arrive at the source, the queue does not appear. If no call packets are waiting, the containers must wait in the queue. Since the closed customers have another queue to wait in along their closed loop, the additional
queue would tend to increase the settling time of the system over the remaining settling time of the packet-available model by itself.

For low arrival rates, it becomes less likely that a packet is waiting when the container arrives. Containers almost always have to wait so the system looks like a loss system with its additional queue. The low arrival rate implies that it will not take long for convergence to take place after the backlog clears. Also, the time to clear the backlog will tend to be shorter so the packet-available model will be farther from its steady-state distribution when the backlog clears. Thus, the packet-available model may not be appropriate in this case.

As a suggestion for determining the time to clear the backlog, note that analysis of the mixed-traffic queueing model gives container-dependent arrival times back at the source. Using this as a non-stationary service process, it may be possible to model the backlog as an $M/G/1$ queue in isolation where arriving call packets queue to wait for a container and container arrivals correspond to the start of a service time. The length of a busy period for this system is the time necessary to clear the backlog. The probability of container $i$ in cycle $k$ arriving to find no waiting packets is the probability that the backlog is cleared. Given that the time container $i$ arrives during cycle $k$, we have an approximate distribution for the time until the backlog clears.

Because of the difficulty involved in analyzing the multiple queues in the mixed packet-available system of Figure 6.1, we analyze the following single-queue systems:

- Single-queue system having one closed customer with immediate loopback (Figure 6.2)

- Single-queue system having multiple closed customers with immediate loopback (Figure 6.3).
Figure 6.2 Single-Queue System with One Closed Customer and Immediate Feedback
Figure 6.3 Single-Queue System with Multiple Closed Customers and Immediate Feedback
For these models, we obtain the distribution of the time until the number of open customers between each closed customer reaches its equilibrium distribution (i.e., until the interdeparture time distribution for closed customers reaches its equilibrium distribution). We expect convergence for the single-queue systems will tend to be faster than for multiple-queue systems under similar traffic intensities so that these models will tend to underestimate the time necessary for the multiple-queue systems to reach equilibrium.

6.2 Single-Queue System with One Closed Customer and Immediate Loopback

The simplest version of the mixed queueing system of Figure 6.1 is a single open queueing system with one closed customer and an immediate loopback as shown in Figure 6.2.

The server is never idle since the closed customer loops back immediately after it completes service. A cycle is the time from one arrival of the closed customer to the next. During each cycle, the server works on open customers for a period of time and then serves the closed customer. By definition, a closed period consists of only one customer service time so an open period may have zero length. The output process then has the form of an alternating renewal process as shown in Figure 6.4.

Open customers arrive according to a Poisson process with rate $\lambda$ and both types of customers draw their service times from the same exponential distribution with rate $\mu$. As a result, the closed periods are independent and identically distributed (i.i.d.).
Figure 6.4 Output Process for Single-Customer System
Definitions used in the analysis follow.

\[ O_i = \text{length of the } i\text{th open period} \]
\[ C_i = \text{length of the } i\text{th closed period} \]
\[ W_i = O_i + C_i = \text{length of the } i\text{th cycle (or window)} \]
\[ N_i = \text{number of open customers arriving during } i\text{th cycle} \]

\[ F_{O_i}(x) = \text{distribution of } O_i \]
\[ F_{C_i}(x) = \text{distribution of } C_i \]
\[ F_{W_i}(x) = \text{distribution of } W_i \]
\[ F_{N_i}(n) = \text{distribution of } N_i \]

\[ F_{O_i}^*(s) = \text{Laplace-Stieltjes Transform of } F_{O_i}(x) \]
\[ F_{C_i}^*(s) = \text{Laplace-Stieltjes Transform of } F_{C_i}(x) \]
\[ \Phi_{N_i}(x) = \text{probability generating function (pgf) of } F_{N_i}(n) \]

Since the \( C_i \) are i.i.d. we drop the subscript, so

\[ F_C(x) = F_{C_i}(x) = 1 - e^{-\mu x}, \quad t \geq 0 \quad \text{for all } i \]

and

\[ F_C^*(s) = F_{C_i}^*(s) = \frac{\mu}{s + \mu} \quad \text{for all } i. \]

Cycle \( i \) begins as the closed customer enters the system for the \( i \)th time. The closed customer then waits while the open customers it finds in the system, if any, receive service. The length of this wait is \( O_i \), which can be zero. After the open customers are gone, the closed customer receives service and then reenters the system starting cycle \( i + 1 \). The number of open customers left by the closed customer at the end of cycle \( i \) is the number it finds at the start of cycle \( i + 1 \). If it looks back just before leaving the system, the closed customer sees all the open customers arriving since
it entered the system. Because the last arrival of the closed customer started cycle $i$, this is $N_i$. As the closed customer then reenters the system, starting cycle $i + 1$, it finds the $N_i$ open customers there. The time to serve these customers is $O_{i+1}$. Thus, the length of cycle $i + 1$ depends on the number of open arrivals during cycle $i$, which depends on the length of cycle $i$.

More formally, begin with the distribution of the number of customers found by the arriving closed customer at the start of cycle $i$ or, equivalently, the number of open customers arriving in cycle $i - 1$. Using this distribution and the Law of Total Probability,

$$O_i^*(s) = \sum_{j=0}^{\infty} \left[ O_i^*(s) \mid j \text{ arrivals in cycle } (i - 1) \right] P\{j \text{ arrivals in cycle } (i - 1)\}.$$

The time to serve $j$ customers is the sum of their individual service times and its distribution is the $j$-fold convolution of their individual service-time distributions. Since the service times are all drawn from the same exponential distribution with rate $\mu$, the LST of the convolution is given by the LST of this distribution raised to the $j$th power so

$$O_i^*(s) = \sum_{j=0}^{\infty} \left( \frac{\mu}{\lambda + \mu} \right)^j P\{j \text{ arrivals in cycle } (i - 1)\}. \quad (6.1)$$

If there were no arrivals in cycle $i - 1$, $O_i$ will have zero length, and $O_i^*(s) = 1 = \left( \frac{-\mu}{\lambda + \mu} \right)^0$.

The length of cycle $i$ is

$$W_i = O_i + C_i.$$

The service time of the closed customer is independent of everything, and the open customers making up $O_i$ clear the system before the closed customer begins service so $O_i$ is independent of $C_i$. Thus,
\[ W_i^*(s) = O_i^*(s)C_i^*(s) = O_i^*(s)C^*(s). \] (6.2)

We now make use of [3!, Theorem 4.2], which states that "... the probability generating function for the number of events from a Poisson process that occur during a random period of time is given by the LST for the distribution of the length of the period of time with the transform variable, \( s \), evaluated at the point \( \lambda[1-z] \)." Applying this gives the pgf of the distribution of the number of arrivals in cycle \( i \) as

\[ \mathcal{H}_N(z) = W_i^*(\lambda[1-z]). \]

Inverting this then gives the distribution to use as a starting point for analyzing cycle \( i + 1 \).

6.2.1 System Behavior Starting in Equilibrium

Assume that the system is in steady state with respect to open customers before the first arrival of the closed customer. The closed customer then finds an M/M/1 queue in equilibrium when it first arrives. The occupancy distribution for this system is [45]

\[ P(j \text{ customers in system}) = (1 - \rho)\rho^j, \quad j = 0, 1, \ldots. \] (6.3)

This is then the distribution of the number of open customers that the closed customer finds in the system on its first arrival. Substituting this into Equation (6.1) yields
\[ O'_1(s) = \sum_{j=0}^{\infty} \left( \frac{\mu}{s + \mu} \right)^j (1 - \rho) \rho^j \]

\[ = (1 - \rho) \sum_{j=0}^{\infty} \left( \frac{\lambda}{s + \mu} \right)^j \]

\[ = (1 - \rho) \frac{1}{1 - \frac{\lambda}{s + \mu}} \]

\[ = \frac{(1 - \rho)(s + \mu)}{s + \mu - \lambda} \quad (6.4) \]

The occupancy distribution of Equation (6.3) is geometric. Thus, if the closed customer finds at least one open customer, it must wait a geometric number of exponential periods, which is itself an exponential length of time. If the closed customer finds the system empty, there is no delay. The LST of the mixture of these two distributions is given by Equation (6.4). Using Equation (6.2), the LST of the length of the first cycle is then

\[ W'_1(s) = O'_1(s) C'_1(s) \]

\[ = \frac{(1 - \rho)(s + \mu)}{s + \mu - \lambda} \frac{\mu}{s + \mu} \]

\[ = \frac{\mu - \lambda}{s + \mu - \lambda} \quad (6.5) \]

This is the LST of an exponential distribution with rate \( \mu - \lambda \). To explain, the total number of services in the cycle is one more than the number of open-customer services and is therefore geometric. The additional service eliminates the possibility of a zero-length cycle, so there are a geometric number of exponential service times, which is exponential. This result implies that the sojourn time of the closed customer in the mixed system is equivalent to a service time in a system with the open customers removed and the server slowed by the arrival rate for open customers. This is the approach used to account for interfering traffic in the flow-control models of [6, 75].
Applying the theorem about Poisson arrivals,

\[ \mathcal{H}_{N_1}(z) = \frac{\mu - \lambda}{\mu - \lambda z}. \]

Inverting this yields

\[ P(\text{j arrivals in cycle } i) = (1 - \rho)^j \rho^i, \quad j = 0, 1, \ldots. \tag{6.6} \]

The distribution of the number of customers left in the system by the closed customer is the same as the distribution of the number found when it arrived. The closed customer leaves the system as it finds it. The system is in equilibrium and future cycles all have the same behavior. Although the future cycles are identically distributed, they are not independent because the number of arrivals in a cycle affects the length of the next cycle.

### 6.2.2 System Behavior Starting with Empty System

Assume that open customers do not begin arriving until after the first arrival of the closed customer. Since it arrives to find an empty system at the start of cycle 1, \( O_1 \) has length zero and

\[
\begin{align*}
O_1^*(t) &= 1, \\
W_1^*(t) &= \frac{\mu}{s + \mu}, \\
\mathcal{H}_{N_1}(z) &= \frac{\lambda[1 - z] + \mu}{s + \mu} = \frac{1}{1 + \rho} \frac{1}{1 - \rho z} \\
&= \frac{1}{1 + \rho} \frac{1}{1 - \rho z} \\
&= \frac{1}{1 + \rho} \sum_{j=0}^{\infty} \left( \frac{\rho}{1 + \rho} \right)^j.
\end{align*}
\tag{6.7}
\]
Inverting $\mathcal{H}_{N_1}(z)$ yields

$$\quad P(j \text{ arrivals in cycle } i) = \left( \frac{1}{1 + \rho} \right) \left( \frac{\rho}{1 + \rho} \right)^j \quad j = 0, 1, \ldots \ .$$

We now state the following results for cycle $i$

$$\quad O^*_i(s) = \frac{(1 - \rho)(s + \mu)}{(1 - \rho^i)s + \mu - \lambda}, \quad i = 1, 2, \ldots \ .$$

$$\quad W^*_i(s) = \frac{\mu - \lambda}{(1 - \rho^i)s + \mu - \lambda}, \quad i = 1, 2, \ldots \ .$$

(6.8)

and

$$\quad P(j \text{ arrivals in cycle } i) = \left( \frac{1 - \rho}{1 - \rho^{i+1}} \right) \left( \frac{\rho(1 - \rho^i)}{1 - \rho^{i+1}} \right)^j, \quad i = 1, 2, \ldots, j = 0, 1, \ldots \ .$$

As the first step in an inductive proof, note that the results hold for $i = 1$. This can easily be seen by using

$$\quad \mu - \lambda = \mu(1 - \rho) \ .$$

and

$$\quad \frac{1 - \rho}{1 - \rho^2} = \frac{1}{1 + \rho} .$$

For the inductive step, assume that

$$\quad P(j \text{ arrivals in cycle } i - 1) = \left( \frac{1 - \rho}{1 - \rho^i} \right) \left( \frac{\rho(1 - \rho^{i-1})}{1 - \rho^i} \right)^j, \quad i = 1, 2, \ldots, j = 0, 1, \ldots .$$
Then using Equations (6.1) and (6.2) as before,

\[ O_i^*(s) = \sum_{j=0}^{\infty} \left( \frac{-\rho}{s+\mu} \right)^j \left( \frac{1-\rho}{1-\rho^i} \right)^j \left( \frac{\rho(1-\rho^i-1)}{1-\rho^i} \right)^j \]

\[ = \left( \frac{1-\rho}{1-\rho^i} \right) \sum_{j=0}^{\infty} \left[ \frac{\mu}{s+\mu} \left( \frac{\rho - \rho^i}{1-\rho^i} \right)^j \right] \]

\[ = \left( \frac{1-\rho}{1-\rho^i} \right) \left( \frac{1}{1 - \left( \frac{\mu}{s+\mu} \right) \frac{\rho - \rho^i}{1-\rho^i}} \right) \]

\[ = \frac{(1-\rho)(s+\mu)}{(1-\rho^i)(s+\mu)-\mu(\rho - \rho^i)} \]

\[ = \frac{(1-\rho)(s+\mu)}{(1-\rho^i)s+\mu-\lambda} \]

and

\[ W_i^*(s) = \frac{\mu-\lambda}{(1-\rho^i)s+\mu-\lambda} \]

Applying the Poisson arrival theorem as before

\[ \mathcal{H}_N(z) = \frac{\mu-\lambda}{(1-\rho^i)\lambda [1-z] + \mu - \lambda} \]

\[ = \frac{\mu-\lambda}{\mu - \lambda \rho^i - \lambda (1-\rho^i)z} \]

\[ = \frac{1-\rho}{1-\rho^{i+1} - \rho(1-\rho^i)z} \]

\[ = \left( \frac{1-\rho}{1-\rho^{i+1}} \right) \left( \frac{1}{1 - \frac{\rho(1-\rho^i)}{1-\rho^{i+1}}} \right) \]
from which

\[ P\{j \text{ arrivals in cycle } i\} = \left( \frac{1 - \rho}{1 - \rho^{i+1}} \right) \left( \frac{\rho(1 - \rho)^i}{1 - \rho^{i+1}} \right)^j, \quad i = 1, 2, \ldots, j = 0, 1, \ldots. \]

This concludes the proof. The results of Equation (6.8) are true. As \( i \to \infty \), these reduce to the previous results for the steady-state system as we would expect.

6.2.3 Output Process Convergence Time

To determine the convergence time for the output process, Equation (6.8) implies the distribution of the length of the \( \text{th} \) cycle is given by

\[ F_i(x) = 1 - \exp\left( -\left( \frac{\mu - \lambda}{1 - \rho^i} \right) x \right), \quad x \geq 0. \]

(For convenience, we drop the \( W \) in the subscript.) The equilibrium cycle length has distribution

\[ F(x) = 1 - \exp( - (\mu - \lambda)x), \quad x \geq 0. \]

Note that

\[
\mu - \lambda < \frac{\mu - \lambda}{1 - \rho^{i+1}} < \frac{\mu - \lambda}{1 - \rho^i},
\]

which implies

\[ F(x) \leq F_{i+1}(x) \leq F_i(x), \quad x \geq 0. \]

The sequence of functions \( \{F_i(x)\}, i = 1, 2, \ldots \) is monotonically decreasing and bounded from below as illustrated in Figure 6.5.
Figure 6.5 Converging Distributions
We now prove that it converges uniformly to \( F(x) \), that is, for every \( \varepsilon > 0 \) there is an integer \( l \) such that \( i \geq l \) implies

\[
| F_i(x) - F(x) | \leq \varepsilon, \quad x \geq 0.
\]

To prove this result, we make use of the following theorem [79, Theorem 7.9]

Suppose

\[
\lim_{i \to \infty} F_i(x) = F(x), \quad x \geq 0.
\]

Put

\[
M_i = \sup_{x \geq 0} | F_i(x) - F(x) |.
\]

Then \( F_i(x) \to F(x) \) uniformly for \( x \geq 0 \), if and only if \( M_i \to 0 \) as \( i \to \infty \).

We have already observed that

\[
\lim_{i \to \infty} F_i(x) = F(x), \quad x \geq 0.
\]

Since \( F_i(x) \) and \( F(x) \) are cumulative probability distributions

\[
F_i(0) = F(0) = 0,
F_i(\infty) = F(\infty) = 1
\]

and

\[
0 \leq | F_i(x) - F(x) | \leq 1.
\]

They are also continuous and monotonically increasing. Thus, their difference has a maximum value occurring at least at a single point in \( x \in (0, \infty) \) which we label \( x_{\text{max}} \). Let

\[
M_i = \max_{x \geq 0} ( F_i(x) - F(x) ) = F_i(x_{\text{max}}) - F(x_{\text{max}}).
\]
To find $M_i$, we first find the location of $x_{\text{max}}$ and substitute its value into the difference.

To find $x_{\text{max}}$, differentiate $F_i(x) - F(x)$ and set it equal to zero.

\[
\frac{d}{dx} [F_i(x) - F(x)] = \frac{d}{dx} \left[ 1 - \exp\left( - \left( \frac{\mu - \lambda}{1 - \rho^i} \right) x \right) - (1 - \exp(- (\mu - \lambda)x)) \right] \\
= \frac{d}{dx} \left[ \exp(-(\mu - \lambda)x) - \exp\left( - \left( \frac{\mu - \lambda}{1 - \rho^i} \right) x \right) \right] \\
= \left( \frac{\mu - \lambda}{1 - \rho^i} \right) \exp\left( - \left( \frac{\mu - \lambda}{1 - \rho^i} \right) x \right) - (\mu - \lambda) \exp(- (\mu - \lambda)x) \\
= (\mu - \lambda) \exp(- (\mu - \lambda)x) \left[ \frac{1}{1 - \rho^i} \exp\left( - \frac{\rho^i (\mu - \lambda)}{1 - \rho^i} x \right) - 1 \right].
\]

Setting the derivative equal to zero,

\[
\frac{1}{1 - \rho^i} \exp\left( - \frac{\rho^i (\mu - \lambda)}{1 - \rho^i} x_{\text{max}} \right) - 1 = 0 \\
\exp\left( - \frac{\rho^i (\mu - \lambda)}{1 - \rho^i} x_{\text{max}} \right) = 1 - \rho^i.
\]

Take the log of both sides and solve for $x_{\text{max}}$ to obtain:

\[
x_{\text{max}} = \frac{-(1 - \rho^i) \ln(1 - \rho^i)}{\rho^i (\mu - \lambda)}.
\]
The maximum difference is then

\[
F(x_{\text{max}}) - F(x_{\text{max}}) = \exp(-\mu) - \exp\left(-\left(\frac{\mu}{1 - \rho^i}x_{\text{max}}\right)\right)
\]

\[
= \exp\left(-\frac{\rho^i}{1 - \rho^i}\ln(1 - \rho^i)\right) - \exp\left(-\frac{1}{\rho^i} \ln(1 - \rho^i)\right)
\]

\[
= (1 - \rho^i)(1 - \rho^i)\rho^i - (1 - \rho^i)^{1/\rho^i}
\]

\[
= \rho^i(1 - \rho^i)(1 - \rho^i)^{1/\rho^i}.
\]

Thus

\[
M_i = \rho^i(1 - \rho^i)(1 - \rho^i)^{1/\rho^i} = \rho^i(1 - \rho^i)^{1/\rho^i}(1 - \rho^i)^{-1}.
\]

To show \(M_i \rightarrow 0\) as \(i \rightarrow \infty\), let

\[
a_i = \rho^i,
\]

\[
b_i = (1 - \rho^i)^{1/\rho^i},
\]

and

\[
c_i = (1 - \rho^i)^{-1}.
\]

It is easily seen that for \(0 < \rho < 1\)

\[
a = \lim_{i \rightarrow \infty} a_i = \lim_{i \rightarrow \infty} \rho^i = 0,
\]

and

\[
c = \lim_{i \rightarrow \infty} c_i = \lim_{i \rightarrow \infty} (1 - \rho^i)^{-1} = 1.
\]

To evaluate \(b = \lim_{i \rightarrow \infty} b_i\), note [16, Theorem 25.6] that if \(G\) is a continuous function, then

\[
G(\lim_{x \rightarrow \infty} F(x)) = \lim_{x \rightarrow \infty} G(F(x)).
\]
Then
\[
\ln(b) = \ln \left( \lim_{i \to \infty} (1 - \rho^i)^{1/\rho^i} \right) = \lim_{i \to \infty} \ln((1 - \rho^i)^{1/\rho^i}) = \lim_{i \to \infty} \frac{1}{\rho^i} \ln((1 - \rho^i)) = \frac{0}{0}.
\]

Using L'Hopital's rule
\[
\ln(b) = \lim_{i \to \infty} \frac{-i \rho^{i-1} / (1 - \rho^i)}{i \rho^{i-1} / (1 - \rho^i)} = \lim_{i \to \infty} \frac{-1}{1 - \rho^i} = -1.
\]

Then
\[
b = \lim_{i \to \infty} (1 - \rho^i)^{1/\rho^i} = e^{-1}.
\]

Now
\[
M_i = a_i b_i c_i.
\]

Since the individual limits exist, the limit of the product is the product of the limits [16, Theorem 15.6] and
\[
\lim_{i \to \infty} M_i = \lim_{i \to \infty} a_i b_i c_i = a b c = 0,
\]

which completes the proof.

To find the value of \( I \) given above in the definition of uniform convergence, it is sufficient to choose \( I \) such that \( M_i \leq \varepsilon \) or
\[
\rho^i (1 - \rho^i)^{1 - \rho^i/\rho^i} \leq \varepsilon.
\]

Taking the log of both sides
\[
I \ln \rho + \left[ \frac{1 - \rho^i}{\rho^i} \right] \ln(1 - \rho^i) \leq \ln \varepsilon.
\]
Rearranging and assuming $0 < \rho < 1$ yields

$$\ln \varepsilon = \left[ \frac{1 - \rho^I}{\rho^I} \right] \ln(1 - \rho^I)$$

$$I \geq \frac{\ln \varepsilon}{\ln \rho}.$$ 

Define

$$k(x) = \left[ \frac{1 - \rho^x}{\rho^x} \right] \ln(1 - \rho^x)$$

$$h(u) = \frac{\ln \varepsilon}{\ln \rho}. \quad (6.9)$$

We have been unable to prove that Equation (6.9) is a contraction. However, in practice, it acts like one. Successive substitution results in a converging sequence of values leading to a fixed point of the form

$$h(u) = u.$$ 

Choosing $I = \text{ceiling}(u)$ gives the smallest $I$ satisfying the uniform convergence criterion where $\text{ceiling}(u)$ is the smallest integer greater than or equal to $u$.

Using the result of Equation (6.9), the time to equilibrium of the single-queue system with one closed customer is the sum of the lengths of the first $I$ cycles so that its distribution is the convolution of the distributions of the lengths of these cycles. The LST of its distribution is then the product of the cycle LSTs.

$$TE_1^*(s) = \prod_{k=1}^I W_K(s)$$

$$= \prod_{k=1}^I \frac{\mu - \lambda}{(1 - \rho^k)s + \mu - \lambda}$$

$$= \prod_{k=1}^I \frac{(\mu - \lambda)/(1 - \rho^k)}{s + (\mu - \lambda)/(1 - \rho^k)}$$

77
where $TE_i$ denotes the length of the time to equilibrium for a single-queue system with one closed customer. To invert this, use a partial fraction expansion to obtain

$$TE_i^*(s) = \sum_{k=1}^{l} \frac{A_k (\mu - \lambda)(1 - \rho^k)}{s + (\mu - \lambda)(1 - \rho^k)}.$$ 

To find $A_k$,

$$
\left(\frac{\mu - \lambda}{1 - \rho^k}\right) A_k = \left[\left( S + \left( \frac{\mu - \lambda}{1 - \rho^k} \right) \right) \prod_{j=1}^{l} \frac{(\mu - \lambda)(1 - \rho^j)}{s + (\mu - \lambda)(1 - \rho^j)} \right] \bigg|_{s = \frac{k - \lambda}{1 - \rho^k}} = \frac{\prod_{j=1}^{l} \left( \frac{\mu - \lambda}{1 - \rho^j} \right)}{\prod_{j=1}^{l} \left( s + \left( \frac{\mu - \lambda}{1 - \rho^j} \right) \right)} \bigg|_{s = \frac{k - \lambda}{1 - \rho^k}}.
$$

Dividing through by $(\mu - \lambda)(1 - \rho^k)$ and evaluating for $s$ yields

$$A_k = \frac{\prod_{j=1}^{l} \left( \frac{\mu - \lambda}{1 - \rho^j} \right)}{\prod_{j=1}^{l} \left( \frac{\mu - \lambda}{1 - \rho^j} - \left( \frac{\mu - \lambda}{1 - \rho^k} \right) \right)} = \frac{\prod_{j=1}^{l} \left( \frac{\mu - \lambda}{1 - \rho^j} \right)}{\prod_{j=1}^{l} \left( \left( \frac{\mu - \lambda}{1 - \rho^j} \right) - \left( \frac{\mu - \lambda}{1 - \rho^k} \right) \right)}.$$
Dividing the numerator and denominator by \((\mu - \lambda)/(1 - \rho)\) results in

\[
A_k = \prod_{\substack{j = 1 \atop j \neq k}}^i \frac{1}{i - \rho_j} = \prod_{\substack{j = 1 \atop j \neq k}}^i \frac{1 - \rho_j}{\rho_j - \rho_k}.
\]

Inverting

\[
TE_1^*(s) = \sum_{k = 1}^i \frac{A_k(\mu - \lambda)(1 - \rho_k)}{s + (\mu - \lambda)(1 - \rho_k)}
\]

term by term yields

\[
f_{TE1}(t) = \sum_{k = 1}^i A_k \left( \frac{\mu - \lambda}{1 - \rho_k} \right) e^{-(\mu - \lambda)t(1 - \rho_k)} , \quad t \geq 0
\]

and

\[
F_{TE1}(t) = \sum_{k = 1}^i A_k \left[ 1 - e^{-\mu(1 - \rho_k)t} \right] , \quad t \geq 0
\]

where

\[
A_k = \prod_{\substack{j = 1 \atop j \neq k}}^i \frac{1 - \rho_j}{\rho_j - \rho_k}.
\]
Figure 6.6 Cycle Structure for Multiple-Customer System
6.3 Single-Queue System With Multiple Closed Customers and Immediate Loopback

We now increase the number of customers circulating along the closed path. Let the number of closed customers be $m$ and label them $C_1, C_2, \ldots, C_m$ such that customer $C_1$ precedes $C_2$ who, in turn, precedes $C_3$, and so on.

A cycle is the time between successive arrivals of customer $C_1$ and contains $m$ subcycles, each consisting of an open period followed by a closed-customer service time. The structure of cycle $i$ is illustrated in Figure 6.6.

For ease of illustration, we will discuss the behavior of subcycle $k$ in cycle $i$ for $i > 1$ and $1 < k < m$. Subcycle $k$ begins as closed customer $C_{i-1}$ completes service and a period of serving open customers begins. The length of this period is $O_k$, which can be zero. After the open customers are gone, closed customer $C_i$ receives service, then departs and reenters the system starting subcycle $k + 1$. As it reenters the system, $C_i$ finds a number of open customers between itself and $C_{i-1}$. These are the open customers arriving to the system after $C_{i-1}$ and before $C_i$, i.e., the open customers arriving during subcycle $k$. Given the LST of the distribution of the length of the $k$th subcycle, we can once again use the theorem about Poisson arrivals to determine the distribution of the number of open arrivals. These are the customers who will be served in the open customer service period of subcycle $k$ during the next cycle, $O_k(i + 1)$. Thus, we see that the behavior of subcycle $k$ is independent of the other subcycles and depends only on the number of open customer arrivals during subcycle $k$ of the previous cycle. As a result, the behavior of the $k$th subcycle over all cycles looks exactly like the behavior of a single-closed-customer system and the cycles of the system with $m$ closed customers looks like an interleaving of cycles from $m$ single-customer systems.

Using the single-customer results and the independence of the subcycles, we can describe the behavior of the cycle in the multiple-customer system. The number of cycles until equilibrium can
be determined by examining the behavior of the "slowest" subcycle. "Slowest" here refers to the subcycle taking the longest time for convergence based on its initial conditions. The number of cycles for convergence is equal to the number of cycles required for the single-customer system representing the "slowest" subcycle to converge. Given the LST for the distribution of each cycle's length and the number of cycles required for equilibrium, it is possible to obtain the distribution of the time until equilibrium for the multiple-customer system.

6.3.1 Single-Queue System Containing $m$ Closed Customers Starting in Equilibrium

Let the system be in equilibrium with respect to the open customers before the closed customers enter the system. Let the closed customers arrive together, as a batch, so that there are no open customers between them. On its initial arrival, the first closed customer, $C_1$, finds an M/M/1 queue in equilibrium. The distribution of customers found is then given by Equation (6.3). As with the single-closed-customer system, the LST of the distribution of the time to serve these open customers is given by Equation (6.4) and the LST of the distribution of the length of the first subcycle will be given by Equation (6.5). In addition, the distribution of the number of open customer arrivals before $C_1$ reenters the system is given by Equation (6.6). These arrivals, and only these arrivals, along with the service time of customer $C_1$ will determine the length of subcycle 1 in cycle 2. Thus, the same process repeats itself and the first subcycle in each cycle behaves like a single-queue system with one closed customer starting in equilibrium as described by Equations (6.4), (6.5) and (6.6) so

$$O^*_1(t) = \frac{(1 - \rho)(s + \mu j)}{s + \mu - \lambda}$$

$$W^*_1(t) = \frac{\mu - \lambda}{s + \mu - \lambda}$$

$P(j \text{ arrivals in subcycle 1}) = (1 - \rho)\rho^j, \quad j = 0, 1, \ldots$. 
Similarly, consider the remaining closed customers \( C_k, k = 2, 3, ..., m \). Since they arrive as a batch, there are no open customers between them to be served so \( O_{k1} = 0, k = 2, 3, ..., m \) and the results of Equation (6.7) hold

\[
O_{k1}^*(s) = l, \ k = 2, 3, ..., m, \\
W_{k1}^*(s) = \frac{\mu}{s + \mu}, \ k = 2, 3, ..., m,
\]

\[
P(j \text{ arrivals in subcycle } k1) = \left( \frac{1}{1 + \rho} \right) \left( \frac{\rho}{1 + \rho} \right)^j \ j = 0, 1, ..., k = 2, 3, ..., m.
\]

Because the subcycles are independent, the future behavior of each is then described by Equation (6.8). For \( k = 2, 3, ..., m \) we have

\[
O_{ik}^*(s) = \frac{(1 - \rho)(s + \mu)}{(1 - \rho)^i s + \mu - \lambda}, \ i = 1, 2, ...
\]

\[
W_{ik}^*(s) = \frac{\mu - \lambda}{(1 - \rho)^i s + \mu - \lambda}, \ i = 1, 2, ...
\]

and

\[
P(j \text{ arrivals in subcycle } ki) = \left( \frac{1 - \rho}{1 - \rho^{i + 1}} \right) \left( \frac{\rho(1 - \rho)^i}{1 - \rho^{i + 1}} \right)^j, \ i = 1, 2, ..., j = 0, 1, ...
\]

For each cycle we now have

\[
W_i^*(s) = \prod_{k=1}^{m} W_{ik}^*(s), \ i = 1, 2, ...
\]

\[
= \left( \frac{\mu - \lambda}{s + \mu - \lambda} \right) \left( \frac{\mu - \lambda}{(1 - \rho)^{i} s + \mu - \lambda} \right)^{m-1}, \ i = 1, 2, ...
\]

The number of cycles needed for convergence will be the same as for the single-queue system with one closed customer starting empty, say \( l \). Then the LST of the distribution of the time to equilibrium for the single-queue system with \( m \) closed customers starting in equilibrium is
\[ TE_m(s) = \prod_{i=1}^{l} W_i(s) \]
\[ = \prod_{i=1}^{l} \left[ \frac{(\frac{\mu - \lambda}{s + \mu - \lambda})(\frac{\mu - \lambda}{(1 - \rho) s + \mu - \lambda})^{m-1}}{1 - \rho} \right] \]
\[ = \left( \frac{\mu - \lambda}{s + \mu - \lambda} \right)^l \prod_{i=1}^{l} \left( \frac{\mu - \lambda}{(1 - \rho) s + \mu - \lambda} \right)^{m-1}, \quad m = 2, 3, \ldots \]

Since \( m - 1 \) of the subcycles behave like a single-queue system with one closed customer, the results given in Equation (6.10) provide a lower bound for the time to equilibrium of the system with multiple closed customers.
6.4 Significance of Transient Phase Following Link Failures

The time to equilibrium for the single-queue system with one closed customer starting empty provides a lower bound for the time to equilibrium of the multiple-queue system with multiple closed customers, which we are using as a proxy for the length of component F3. Thus, our model for the length of the transient phase for a single call following a link failure, $TP_1$, is the time to equilibrium for the single-queue system with one closed customer when starting empty, $TE_1$.

To investigate the significance of the transient phase for a single call relative to the time between network state changes, we follow the approach used for the repair transient.

\[
P(TP_1 \leq \delta S) = \int_0^\infty P(TP_1 \leq \delta S | S = t)dF_3(t)
\]
\[
= \int_0^\infty P(TP_1 \leq \delta t)dF_3(t)
\]
\[
= \int_0^\infty F_{TP_1}(\delta t)dF_3(t).
\]

Assuming that the time to the next network failure or repair is again exponentially distributed with parameter $\beta$ and using Equation (6.10),

\[
P(TP_1 \leq \delta S) = \int_0^\infty \sum_{k=1}^l A_k \left[ 1 - \exp \left( - \left( \frac{\mu - \lambda}{1 - \rho^k} \right) \delta t \right) \right] \beta \exp(-\beta t)dt.
\]

Interchanging the order of integration and summation, then distributing $\beta \exp(-\beta t)dt$,

\[
P(TP_1 \leq \delta S) = \sum_{k=1}^l A_k \left[ \int_0^\infty \beta \exp(-\beta t)dt - \int_0^\infty \beta \exp \left( - \left[ \beta + \delta \left( \frac{\mu - \lambda}{1 - \rho^k} \right) \right] t \right) dt \right].
\]

85
The value of the first integral is 1 since we are integrating across a probability distribution so

\[ P(\mathit{TP}_1 \leq \delta S) = \sum_{k=1}^{l} A_k \left[ 1 - \frac{\beta}{\beta + \delta \left( \frac{\mu - \lambda}{1 - \rho^k} \right)} \int_0^\infty \left[ \beta + \delta \left( \frac{\mu - \lambda}{1 - \rho^k} \right) \right] e^{-(\beta + \delta \left( \frac{\mu - \lambda}{1 - \rho^k} \right)) t} dt \right] \]

The integral here is also equal to one. Thus

\[ P(\mathit{TP}_1 \leq \delta S) = \sum_{k=1}^{l} A_k \left[ 1 - \frac{\beta}{\beta + \delta \left( \frac{\mu - \lambda}{1 - \rho^k} \right)} \right] \]

\[ = \sum_{k=1}^{l} A_k \left[ 1 - \frac{1}{\beta + \delta \left( \frac{1 - \rho^k}{1 - \rho^k} \right) \frac{\delta \mu}{\beta}} \right] \]

where

\[ A_k = \prod_{j=1}^{l} \frac{1 - \rho^k}{1 - \rho^j}, \quad k = 1, 2, \ldots \]

Given there are \( n \) calls disrupted at the time of failure, the length of the overall transient phase will equal the longest transient phase of the individual calls. Note that the longest transient phase call is not significant if each of the individual calls is not significant. That is,

\[ P(\mathit{TP}_n \leq \delta S) = \left[ P(\mathit{TP}_1 \leq \delta S) \right]^n \]

Assume the distribution of the number of interrupted calls is the same as used in the repair case. Then

\[ P(\text{n calls interrupted}) = p_n = \frac{(N)^n}{n!} e^{-N}, \quad n = 0, 1, 2, \ldots \]
where

\[
N = \frac{\lambda_{\text{call}}}{\nu_{\text{call}}} = \text{average number of active calls interrupted.}
\]

Now, unconditioning on the number of calls

\[
P(TP \leq \delta S) = \sum_{n=0}^{\infty} p_n P(TP_n \leq \delta S)
\]

\[
= \sum_{n=0}^{\infty} \frac{(N)^n}{n!} e^{-N} [P(TP_1 \leq \delta S)]^n
\]

\[
= e^{-N} \sum_{n=0}^{\infty} \frac{(N^* P(TP_1 \leq \delta S))^n}{n!}
\]

\[
= e^{-N} e^{N^* P(TP_1 \leq \delta S)}
\]

\[
= e^{-N^* P(TP_1 > \delta S)}
\]

(6.11)

Thus, the probability that the length of the transient phase following a failure is significant, based on the single-queue model with one closed customer is

\[
P(TP > \delta S) = 1 - e^{-N^* P(TP_1 > \delta S)}
\]

(6.12)

where

\[
P(TP_1 > \delta S) = 1 - \sum_{k=1}^{l} A_k \left[ 1 - \frac{1}{i + \left( \frac{1 - \rho}{1 - \rho^k} \right) \delta R} \right],
\]

\[
A_k = \prod_{j=1, j \neq k}^{l} \frac{1 - \rho^k}{\rho^j - \rho^k}, \quad k = 1, 2, \ldots,
\]

\[
R = \frac{\mu}{\bar{\rho}}
\]
and $l$ is the minimum number of cycles needed for convergence given by the contraction of Equation (6.9).

The Pascal computer program listed in Appendix C implements these results given a set of values for $\rho$, $\varepsilon$ and $\delta$. Recall that $\rho$ is the traffic intensity of open customers, $\varepsilon$ is the bound for convergence of the output process, and $\delta$ is the fraction of time to the next network state transition that defines significance of the transient phase. Tables B.1 through B.7 present program output for several cases. In each table, the program first echoes the given values of $\rho$, $\varepsilon$ and $\delta$. It then uses the contraction of Equation (6.9) to determine $l$, the minimum number of cycles needed to guarantee convergence of the cycle-length distribution for the given value of $\varepsilon$. Using this value, it then computes and lists the $A_s$ coefficients used in Equation (6.10) for the distribution of the length of time to equilibrium for the single-queue system with one closed customer, and again in Equation (6.12) for the probability of the network transient phase following failures being significant. Finally, the program determines this probability over a range of values for $N$ and $R$ where $N$ is the expected number of rerouted calls and $R$ is the ratio of the rate of packet transmission to the rate of network state changes. These probabilities are presented again in Tables 6.1 through 6.7 and plotted in Figures 6.7 through 6.13 respectively.

Table 6.2 represents our baseline case: $\rho = 0.8$, $\varepsilon = 0.1$ and $\delta = 0.01$. We choose $\rho = 0.8$ to indicate a moderately heavy traffic intensity for the open customers in the single-queue system with one closed customer. Using $\varepsilon = 0.1$ in the convergence criterion gives a loose definition of convergence. The value is small enough to get close, but still leaves a large margin for error. For the definition of significance, we use $\delta = 0.01$. The transient phase is significant if it is larger than 1% of the time to the next change in network state. As in the repair case, this value is small enough to be reasonable without being too restrictive.
Table 6.1 Failure Transient Significance \( \rho=0.6, \ \epsilon=0.1, \ \delta=0.01 \)

\[
P( \text{Transient Phase is Significant} )
\]

<table>
<thead>
<tr>
<th>N</th>
<th>R=1</th>
<th>R=10</th>
<th>R=100</th>
<th>R=1K</th>
<th>R=10K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212044</td>
<td>0.63202505</td>
<td>0.60742638</td>
<td>0.29159069</td>
<td>0.04326254</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995448</td>
<td>0.99991306</td>
<td>0.96816956</td>
<td>0.35741916</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.98799733</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>R=100K</th>
<th>R=1M</th>
<th>R=10M</th>
<th>R=100M</th>
<th>R=1000M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00453562</td>
<td>0.00045576</td>
<td>0.00004560</td>
<td>0.00000456</td>
<td>0.00000046</td>
</tr>
<tr>
<td>10</td>
<td>0.04444159</td>
<td>0.00454822</td>
<td>0.00045588</td>
<td>0.00004560</td>
<td>0.00000456</td>
</tr>
<tr>
<td>100</td>
<td>0.36529434</td>
<td>0.04456248</td>
<td>0.00454948</td>
<td>0.00045589</td>
<td>0.00004560</td>
</tr>
<tr>
<td>1000</td>
<td>0.98938980</td>
<td>0.36609863</td>
<td>0.04457460</td>
<td>0.00454960</td>
<td>0.00045590</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>0.98952319</td>
<td>0.36617723</td>
<td>0.04457581</td>
<td>0.00454962</td>
</tr>
</tbody>
</table>

Table 6.2 Failure Transient Significance \( \rho=0.8, \ \epsilon=0.1, \ \delta=0.01 \)

\[
P( \text{Transient Phase is Significant} )
\]

<table>
<thead>
<tr>
<th>N</th>
<th>R=1</th>
<th>R=10</th>
<th>R=100</th>
<th>R=1K</th>
<th>R=10K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.63207284</td>
<td>0.55530838</td>
<td>0.15822524</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.99995454</td>
<td>0.99970433</td>
<td>0.82136810</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99999979</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>R=100K</th>
<th>R=1M</th>
<th>R=10M</th>
<th>R=100M</th>
<th>R=1000M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01880315</td>
<td>0.00191546</td>
<td>0.00019190</td>
<td>0.00001919</td>
<td>0.00000192</td>
</tr>
<tr>
<td>10</td>
<td>0.17289348</td>
<td>0.01899030</td>
<td>0.00191738</td>
<td>0.00019192</td>
<td>0.00001919</td>
</tr>
<tr>
<td>100</td>
<td>0.85016459</td>
<td>0.17446969</td>
<td>0.01900917</td>
<td>0.00191757</td>
<td>0.00019192</td>
</tr>
<tr>
<td>1000</td>
<td>0.99999999</td>
<td>0.85299563</td>
<td>0.17462847</td>
<td>0.01901106</td>
<td>0.00191759</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.85327812</td>
<td>0.17464436</td>
<td>0.01901125</td>
</tr>
</tbody>
</table>
### Table 6.3 Failure Transient Significance \( \rho=0.9, \epsilon=0.1, \delta=0.01 \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=1 )</th>
<th>( R=10 )</th>
<th>( R=100 )</th>
<th>( R=1K )</th>
<th>( R=10K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.6307235</td>
<td>0.39127180</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.9995284</td>
<td>0.99301391</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.0000000</td>
<td>1.0000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.0000000</td>
<td>1.0000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=100K )</th>
<th>( R=1M )</th>
<th>( R=10M )</th>
<th>( R=100M )</th>
<th>( R=1000M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.06570437</td>
<td>0.00700733</td>
<td>0.00070537</td>
<td>0.00007058</td>
<td>0.00000706</td>
</tr>
<tr>
<td>10</td>
<td>0.49319236</td>
<td>0.06790443</td>
<td>0.00703138</td>
<td>0.00070562</td>
<td>0.00007059</td>
</tr>
<tr>
<td>100</td>
<td>0.99888203</td>
<td>0.50500885</td>
<td>0.06813015</td>
<td>0.00703379</td>
<td>0.00070564</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>0.99911683</td>
<td>0.50619927</td>
<td>0.06815282</td>
<td>0.00703399</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99913796</td>
<td>0.50631840</td>
<td>0.06815467</td>
</tr>
</tbody>
</table>

### Table 6.4 Failure Transient Significance \( \rho=0.8, \epsilon=0.1, \delta=0.001 \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=1 )</th>
<th>( R=10 )</th>
<th>( R=100 )</th>
<th>( R=1K )</th>
<th>( R=10K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.6307284</td>
<td>0.55630838</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.9995454</td>
<td>0.99970433</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.0000000</td>
<td>1.0000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.0000000</td>
<td>1.0000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.0000000</td>
<td>1.0000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=100K )</th>
<th>( R=1M )</th>
<th>( R=10M )</th>
<th>( R=100M )</th>
<th>( R=1000M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15822524</td>
<td>0.01880315</td>
<td>0.00191546</td>
<td>0.00019190</td>
<td>0.00001919</td>
</tr>
<tr>
<td>10</td>
<td>0.82136810</td>
<td>0.17289348</td>
<td>0.01899030</td>
<td>0.00191738</td>
<td>0.00019192</td>
</tr>
<tr>
<td>100</td>
<td>0.99999997</td>
<td>0.85016459</td>
<td>0.17464969</td>
<td>0.0190917</td>
<td>0.00191757</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>0.99999999</td>
<td>0.85299563</td>
<td>0.17462847</td>
<td>0.01901106</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.85327812</td>
<td>0.17464436</td>
</tr>
</tbody>
</table>
Table 6.5 Failure Transient Significance \( \rho=0.8, \varepsilon=0.1, \delta=0.1 \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=1 )</th>
<th>( R=10 )</th>
<th>( R=100 )</th>
<th>( R=1K )</th>
<th>( R=10K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212056</td>
<td>0.63207284</td>
<td>0.55630838</td>
<td>0.15822524</td>
<td>0.01880315</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995454</td>
<td>0.99970433</td>
<td>0.82138610</td>
<td>0.17289348</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99999997</td>
<td>0.85016459</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99999999</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=100K )</th>
<th>( R=1M )</th>
<th>( R=10M )</th>
<th>( R=100M )</th>
<th>( R=1000M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00191546</td>
<td>0.00019190</td>
<td>0.00001919</td>
<td>0.00000192</td>
<td>0.00000019</td>
</tr>
<tr>
<td>10</td>
<td>0.01899030</td>
<td>0.00191738</td>
<td>0.00019192</td>
<td>0.00001919</td>
<td>0.00000192</td>
</tr>
<tr>
<td>100</td>
<td>0.17446969</td>
<td>0.01900917</td>
<td>0.00191757</td>
<td>0.00019192</td>
<td>0.00001919</td>
</tr>
<tr>
<td>1000</td>
<td>0.85299563</td>
<td>0.17462847</td>
<td>0.01901106</td>
<td>0.00191759</td>
<td>0.00019192</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>0.85327812</td>
<td>0.17464436</td>
<td>0.01901125</td>
<td>0.00191759</td>
</tr>
</tbody>
</table>

Table 6.6 Failure Transient Significance \( \rho=0.8, \varepsilon=0.01, \delta=0.01 \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=1 )</th>
<th>( R=10 )</th>
<th>( R=100 )</th>
<th>( R=1K )</th>
<th>( R=10K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63248805</td>
<td>0.63252450</td>
<td>0.63247964</td>
<td>0.63355478</td>
<td>0.38391787</td>
</tr>
<tr>
<td>10</td>
<td>0.99995505</td>
<td>0.99995510</td>
<td>0.99995504</td>
<td>0.99995634</td>
<td>0.99212254</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=100K )</th>
<th>( R=1M )</th>
<th>( R=10M )</th>
<th>( R=100M )</th>
<th>( R=1000M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.06789751</td>
<td>0.01934175</td>
<td>0.01550356</td>
<td>0.00778206</td>
<td>0.01165035</td>
</tr>
<tr>
<td>10</td>
<td>0.50496410</td>
<td>0.17742244</td>
<td>0.14465467</td>
<td>0.07515119</td>
<td>0.11058159</td>
</tr>
<tr>
<td>100</td>
<td>0.99911617</td>
<td>0.85816984</td>
<td>0.79038861</td>
<td>0.54216664</td>
<td>0.69021445</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99999984</td>
<td>0.99999535</td>
<td>0.99999186</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>
Table 6.7 Failure Transient Significance $\rho = 0.8$, $\epsilon = 0.006$, $\delta = 0.01$

$P\{\text{Transient Phase is Significant}\}$

<table>
<thead>
<tr>
<th>N</th>
<th>$R=1$</th>
<th>$R=10$</th>
<th>$R=100$</th>
<th>$R=1K$</th>
<th>$R=10K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.22119922</td>
<td>0.93607214</td>
<td>0.99999969</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10</td>
<td>0.91791500</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>$R=100K$</th>
<th>$R=1M$</th>
<th>$R=10M$</th>
<th>$R=100M$</th>
<th>$R=1000M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>
Figure 6.7 Probability of Failure Transient
Significance $\rho = 0.6$, $\varepsilon = 0.1$, $\delta = 0.01$
Figure 6.8 Probability of Failure Transient
Significance rho=0.8, epsilon=0.1, delta=0.01
Figure 6.9 Probability of Failure Transient
Significance $\rho=0.9$, $\epsilon=0.1$, $\delta=0.01$
Figure 6.10 Probability of Failure Transient
Significance rho=0.8, epsilon=0.1, delta=0.001
Figure 6.11 Probability of Failure Transient
Significance rho=0.8, epsilon=0.1, delta=0.1
Figure 6.12 Probability of Failure Transient
Significance rho=0.8, epsilon=0.01, delta=0.01
Figure 6.13 Probability of Failure Transient
Significance rho=0.8, epsilon=0.006, delta=0.01
The data in Table 6.2 for the probability of the network transient phase following link failures being significant relative to the time between changes in network state appears in Figure 6.8. It shows this probability plotted against \( N \), for a range of \( R \) values. Notice that the probability of significance can exceed 15% even when \( R \) is 100 million. As in the repair case, \( R \) by itself is not enough. The size of the network and the amount of traffic must also be considered.

Figure 6.8 shows that the failure case also has some interesting patterns. When \( R \geq 10,000 \), increasing \( R \) by an order of magnitude decreases the probability of significance by an order of magnitude and shifts the curve to the right without changing its shape. Also, when \( \frac{R}{N} > 10,000 \), increasing \( N \) by an order of magnitude increases the probability of significance by an order of magnitude. Both patterns appear in almost all the other figures and seem to use the same threshold, although its value changes as the parameters change. The exceptions are for the cases with low \( \epsilon \). Tightening the criterion for convergence significantly increases the probability of significance. For the cases of varying \( \delta \), the threshold values appears to be \( 100/\delta \) as in the repair case. For increasing \( \rho \), the threshold value increases.

As a result of these patterns, it would appear that it may be possible to specify sufficient conditions for significance in terms of the ratio of \( R \) to \( N \). For example, in Table 6.2, if \( \frac{R}{N} < 10,000 \) then the probability that the transient phase is significant is greater than 0.18. We tried to find a sufficient condition for significance in terms of \( \frac{R}{N} \), but were unsuccessful. However, we were able to obtain a sufficient condition in terms of \( N \) and \( P(TP_1 > \delta S) \).

Let the transient phase be significant when

\[
P(TP > \delta S) > \alpha.
\]

Substituting from Equation (6.12) and rearranging gives

\[e^{-N*P(TP_1 > \delta S)} < 1 - \alpha.
\]
Taking the log of both sides yields

\[-N^*P\{TP_1 > \delta S\} < \ln(1 - \alpha).\]

Sufficient conditions for significance of the transient phase are then

\[N > \frac{-\ln(1 - \alpha)}{P\{TP_1 > \delta S\}}\]

or

\[P\{TP_1 > \delta S\} > \frac{-\ln(1 - \alpha)}{N}.\]

Parameters here are \(N\), \(\rho\), and \(\frac{\mu}{\beta}\), for a given \(\varepsilon\), \(\delta\), and \(\alpha\). We see that the transient phase is significant wherever \(N\), the average number of rerouted calls, is too large or the probability of any one call being significant becomes too great.

Tables 6.1, 6.2 and 6.3 show the effect of increasing \(\rho\) for fixed \(\varepsilon\) and \(\delta\). Figures 6.7, 6.8 and 6.9 plot these results for \(\rho = 0.6\), \(\rho = 0.8\) and \(\rho = 0.9\) respectively, with \(\varepsilon = 0.1\) and \(\delta = 0.01\). They show that as \(\rho\) increases, \(\alpha\) does the probability of significance. This is because increasing \(\rho\) increases the number of cycles required for convergence. The increase in the right-hand side of Equation (6.9) is driven by the denominator, which decreases as \(\rho\) increases. The large number of cycles increases the length of time for a single call to reach equilibrium and, thus, the length of the network transient phase.

Tables 6.2, 6.6 and 6.7 show the effect of tightening the convergence criterion, i.e., decreasing \(\varepsilon\) for fixed \(\rho\) and \(\delta\). Figures 6.8, 6.12 and 6.13 plot these results for \(\varepsilon = 0.1\), \(\varepsilon = 0.01\) and \(\varepsilon = 0.006\) respectively, with \(\rho = 0.8\) and \(\delta = 0.01\). (Smaller values of \(\varepsilon\) resulted in numerical problems.) Not surprisingly, tightening the convergence criterion increases the probability of significance. Again, the number of cycles required by Equation (6.9) increases. In this case, the increase is driven by the \(\ln \varepsilon\) term in the numerator.
Finally, the effect of varying the significance criterion appears in Tables 6.2, 6.4 and 6.5. Figures 6.10, 6.8 and 6.11 plot results for $\delta = 0.001$, $\delta = 0.01$ and $\delta = 0.1$ respectively, with $\rho = 0.8$ and $\epsilon = 0.1$. Since $\delta$ and $R$ appear together in Equation (6.12), increasing $\delta$ by an order of magnitude has the same effect as increasing $R$ by an order of magnitude. The effect is the same as in the repair case. As $\delta$ increases by an order of magnitude, the data in the table shifts one column to the left and now represents a value of $R$ that is an order of magnitude less than before.
6.5 Example

Recall the example network of Chapter Five. For failure transients, assume the speed of the link is 2 Gbps and the average packet size is 1000 bits then

\[ \mu = 7.2 \times 10^9 \text{ packets per hour}. \]

Let \( \rho = 0.8, \varepsilon = 0.1, \) and \( \delta = 0.01. \) Following the failure of a link, we again have

\[ \beta = \frac{1}{7} + \frac{10}{200,000} + \frac{27}{2,000,000} = 0.14292 \text{ events per hour}. \]

Now

\[ R = \frac{\mu}{\beta} = 5.04 \times 10^{10} \text{ packet transmissions per event}. \]

From Equation (6.12), for \( N = 50,000, \) the probability that the transient phase is significant following failure is 0.0019.

For the transient following a link failure with another component failed,

\[ \beta = \frac{2}{7} + \frac{9}{200,000} + \frac{27}{2,000,000} = 0.28577 \text{ events per hour}. \]

Now

\[ R = \frac{\mu}{\beta} = 2.52 \times 10^{10} \text{ packet transmissions per event} \]

and the significance of the transient phase is 0.0038.
CHAPTER SEVEN
CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

7.1 Conclusions

Chapter Two presents a brief review of network effectiveness measures. A fundamental approach for these types of measures is to weight the value of the performance measure for each network state by the probability the network is in that state, then sum across all states. This approach is used for both steady-state and time-dependent analysis of network effectiveness. For load-dependent measures, it is common to assume the system is in equilibrium when determining the value of the performance measure for a given state. This leads to the problem discussed in Chapter One: Is this assumption appropriate? This question does not appear to have been previously investigated.

We begin our study of the transient phase following failures and repairs of network resources with a survey of the existing transient and time-dependent results for networks of queues, which is presented in Chapter Three. With the exception of relaxation times, these tools generally focus on obtaining or approximating time-dependent system measures such as state probabilities or moments based on them. In addition, relaxation-time results, which describe the length of the transient phase, are limited for networks of queues.

In order to model transients in the network, it is necessary to understand how the network behaves following repairs and failures. In Chapter Four, we review the operation of protocols in a communication network, make a number of assumptions about network traffic and, finally, identify a number of transient-phase components. On the basis of this understanding of the network's behavior, we then develop models for use in analyzing the significance of network transients following
repairs and failures. In order to develop conservative models which would underestimate the length of the transient phase, we choose to ignore several of the components and focus on a single component for transients following repairs and a single component for transients following failures. Also, primarily for ease of analysis, we choose to model only transients following repair and failure of communication links.

In Chapter Five, we develop a model of the transient phase following repairs of communication links. Following these repairs, some calls are no longer routed over the best available path in the network. All these calls must terminate before traffic on the network can reach steady state. The time necessary for them to disappear from the network is then a lower bound for the transient following a link repair. The distribution of the length of the time necessary for all these calls to terminate is given by Equation (5.5) and has the form of an extreme-value distribution. Using this as the distribution of the length of the transient phase following a link repair, we determine the probability that the lower bound for the transient phase is not significant following the repair of communication links in the network, as given by Equations (5.6) and (5.8). The steady-state assumption is not appropriate if the length of the lower bound for the transient phase is significant relative to the time until the next change in network state. However, lack of significance of the lower bound is not a sufficient condition for use of the steady-state assumption. The transient phase could still be significant.

A large relative difference in the rates of system events is typically used to justify the use of the steady-state assumption. Our result for the probability of significance is expressed in terms of \( R \), the ratio of the mean time between network state changes to the mean call holding time. This is also the ratio of the rate of call completions to the rate of network state changes. Results indicate that this ratio, by itself, is not enough. The probability of transient-phase significance is also influenced by the average number of calls affected by the repair. Since \( R \) also reflects the number of components in the network, the result is that a large difference in the relative rates of call events and individual component events is not sufficient for justifying use of the steady-state assumption.
in analysis of communication networks. The amount of network traffic and the size of the network must also be considered.

Another result is that, under certain conditions, the probability of significance decreases by an order of magnitude with an order of magnitude increase in \( R \) and increases approximately linearly in \( N \), the number of secondary calls.

We next examine transients following failures of communication links in the network. We ignore several of the previously identified components and choose to focus on the time for packets along the new path of a rerouted call to reach their equilibrium distribution. The transient behavior of the network flow-control mechanism directly influences the length of this transient phase. At this point, in order to develop a tractable model, we assume that a backlog always exists at the source; that is, there are always packets waiting to be transmitted. This differs from the actual behavior of the system where the backlog eventually clears and the pattern of packets along the path then settles into its equilibrium distribution based on the average arrival rate for cell packets.

Our model of the behavior of the network flow-control mechanism during the backlog is a mixed tandem queueing model, which we refer to as the packet-available model. Transient analysis of this system using existing techniques is difficult because of the rapidly increasing size of the state space. In order to overcome this problem, we use an output-process approach to determine the length of time necessary for the system to reach equilibrium. To reduce the complexity of the system, we use a single queue with one closed customer as our initial model. This single-queue system should converge faster than the multiple-queue system and, thus, should provide a lower bound.

The output-process approach views the server as alternating between periods of serving open customers and a service time for a closed customer. We determine the distribution of the length of these cycles based on different initial conditions for the system. Each cycle has an exponential
length, even when the system is not in equilibrium, although they are not identically distributed. These cycle-length distributions converge uniformly to the equilibrium cycle-length distribution. This leads us to develop a contraction map for determining the number of cycles necessary for convergence. Using this, we are able to determine the distribution of the time until equilibrium for the single-queue system with one closed customer (Equation (6.10)). After increasing the number of closed customers in the system, we find that the single-queue system with multiple closed customers looks like the interleaving of a number of single-closed-customer systems. We then determine the LST of the distribution of the time until equilibrium for this system. Our transient analysis of the mixed-traffic single-queue system appears to be new.

Because of the "interleaved" appearance of the multiple-customer system, the single-closed-customer system gives a lower bound for its time to equilibrium. For this reason, we use it as our model of the transient behavior of the network flow-control mechanism. Using the single-queue system with one closed customer as the model for a single call, we determine the probability that the transient phase of a single call is significant relative to the time between network state changes. By assuming the number of affected calls is Poisson, this result can be used to determine the probability that the transient phase for the entire network is significant relative to the time between network state changes, which is given by Equation (6.12).

As in the repair case, we find that a large difference in the relative rates of packet transmissions and individual component failures and repairs is not sufficient to justify use of the steady-state assumption in analysis of communication networks. The amount of network traffic and the size of the network must also be considered. Also, under certain conditions, the probability of significance decreases by an order of magnitude with an order of magnitude increase in \( R \), and increases by an order of magnitude with an order of magnitude increase in \( N \), the number of rerouted calls. Finally, we are able to determine sufficient conditions for significance of the transient phase.
It is interesting to note that in both the repair and failure models of network-transient significance, $R$ is not enough to guarantee the transient phase will not be significant. They also have similar behavior for the probabilities when varying $R$ and $N$. These similarities exist despite their different approaches. The repair-transient model is based on the number of non-optimal calls. The failure-transient model is based on the transient behavior of the flow-control mechanism for a single call, viewing network traffic at the packet level. Perhaps one reason for the similarities is the extreme-value approach used in both when defining significance and determining its probability. Following repairs, the transient phase is not significant if the longest secondary call is less than the significance threshold defined by $\delta$. Following failures, the transient phase is not significant if the longest call transient is less than the threshold. However, this common extreme-value approach leads to different forms of the results as given by Equations (5.6), (5.8) and (6.11). Finally, although they may not provide sufficient conditions for using the steady-state assumption, these models can provide a useful measure of the riskiness of using the assumption. For example, there may be less risk associated with using the steady-state assumption if the probability of significance is $10^{-3}$ than when it is $10^{-6}$.

7.2 Extensions

One possible extension of this work is to model the transient phase following node repairs. This is more complicated than the link-repair case because the number of calls is no longer in steady state. There are no calls waiting for reconnect following a repair; however, there will be newly arriving calls having an origin or a destination at the repaired node. The number of these calls starts at zero and eventually reaches a steady-state distribution. Therefore, we have the same problem as the link-repair case with the addition of new origin and destination traffic at the repaired node. One possible approach to analyzing this problem is to keep the death process of secondary calls previously defined for the link-repair case and add a birth-death process for the origin and destination traffic at the repaired node. This birth-death process starts at zero and converges to its
steady-state distribution. Its behavior is equivalent to the birth-death process modeling an
M/M/∞ queueing system starting empty. The total time for the network to reach steady state is
then the maximum of the settling times for these two components.

Following the failure of a network node, calls that originated and terminated at the failed node
disappear from the network. The calls routed through the failed node will be rerouted. Given a
distribution for the number of these rerouted calls, some of the results from the link failure case
hold. In fact, if the number of calls is Poisson distributed, all the results hold. However, the loss
of a node from the network significantly alters its structure compared to loss of a link. As a result,
transient component F4, the time for the network to reach equilibrium with a new traffic pattern,
may become the dominant factor in determining the length of the transient phase and other existing
results could be used [34].

Our work ignores several of the identified transient-phase components. Models of these com-
ponents could be used with our models for components R1 and F3 to develop more sophisticated
models for both the repair and failure transients.

Component F1 is the time needed for the network to detect the occurrence of a failure. This
is the time between when the failure occurs and call release begins at the packet entity. Where the
failure occurs affects how long the network takes to begin clearing the call. If the failure is in the
physical layer, the link-layer entity detects it within zero to a few seconds. A data-link-layer failure
requires zero to sixty seconds for detection. For packet-layer failures, there is apparently no
mechanism in X.25 for detecting such a failure. However, for ARPANET and DECNET, times
are on the order of zero to fifteen seconds. Finally, transport-layer failures result in a released
connection after an inactivity timer expires without the transport entity receiving an acknowledge-
ment from the opposite transport entity [25, Section 12.2.3.3]. Obtaining a distribution for the time
to detect a failure requires information about the frequency of the different types of failures. Net-
work measurements or protocol simulations could be used to obtain this information.
The time to release and reroute calls using the failed resource, component F2, has two separate subcomponents. These are the time to release calls using the failed resource and the time to establish a new connection over an alternate route. The first involves sending a Call_Clear packet from the packet entities adjacent to the failure back to the source and destination of the connection. To establish a new connection, a Call_Request packet must travel from the source to the destination over the new route and a Call_Connected packet is returned. An upper bound at the transport layer for reassigning a call can be determined from timeout values [25, Section 6.12.3].

A possible model for component F2 is the sojourn time through a set of independent tandem M/M/1 queues. Given the length or number of links of the old call and the new call, we make an assumption about where along the connection the failure occurs in order to obtain the total number of queues that the Call_Clear, Call_Request, and Call_Connected packets must travel through, say N. Under a set of appropriate assumptions, we could then model component F2 as the sojourn time through N independent M/M/1 queues.

Our approach has been to underestimate the length of the transient phase. Our results are only useful for identifying situations where the transient phase is clearly significant and the steady-state assumption is inappropriate. They do not indicate all situations where the assumption may be inappropriate. A more useful result for network designers would be models that could provide sufficient conditions under which the transient phase is not significant and the steady-state assumption is appropriate. An approach based on upper bounds for the transient phase would lead to such results.

Finally, rather than comparing the lengths of the transient phase and the time between network state changes to determine if the steady-state assumption is appropriate, it would be interesting to compare the relative magnitudes of the transient and steady-state performance measures.
Figure 7.1 Single-Queue System with One Closed Customer and Delayed Feedback
7.3 Transient Analysis of Mixed Queueing Systems

Although not directly related to the validity of the steady-state assumption, a whole body of future work remains for applying the output-process approach to transient analysis of mixed queueing systems.

Several extensions of the single-queue results involve the service-time distribution. The first keeps the exponential service times, but to uses different rates for the open and closed customers. The second extension is to use non-exponential service times. The usefulness of the theorem about Poisson arrivals does not depend on exponential service times. However, applicability of the approach could be limited by difficulty in inverting the resulting pgf for the number of arrivals during the previous cycle.

Another set of extensions adds a delayed feedback to the single-queue system with one closed customer, as shown in Figure 7.1. These models isolate the fundamental difficulty appearing in multiple-queue systems. With immediate feedback, the distribution of the number of open customers found by the arriving closed customer depends only on the number of arrivals since its last entry into the system. With delayed feedback, open customers may arrive and depart while the closed customer is delayed. This significantly increases the difficulty of analyzing the system. The form of the delay distribution is another issue. An exponential distribution would seem to be a good place to start, although another possibility is a degenerate delay distribution, i.e., a deterministic delay.

Once results are available for the delayed-feedback system with one closed customer, we believe the independence of subcycles for the closed customers will quickly lead to “interleaved” results for a single-queue system with multiple closed customers and delayed feedback. Because of the property of exponential sojourn times for the M/M/1 queue, results for exponential feedback imme-
diately solve part of the two-queue system with one closed customer, opening the door for analysis of multiple-queue systems with multiple closed customers.
REFERENCES


[34] Duda, A. (1986), "Diffusion Approximations for Time-Dependent Queueing Systems", *IEEE Journal on Selected Areas in Communications*, v. 4, n. 6, 905-918.


APPENDIX A
PASCAL PROGRAM FOR COMPUTING REPAIR TRANSIENT SIGNIFICANCE

PROGRAM Repair_Transient (input, output):
USES
SANE;
VAR
  count, count2 : integer;
  Tab : char;
  Allowed_error, N, delta, R, a : extended;
  wind : rect;

FUNCTION Measure8 (u, N : extended) : extended;
{This function computes P(TP26) based on the approach used in}
{Algorithm AS147 for the Incomplete Gamma Function.}
{There is also a correction in AS, vol34, pg. 326, 1985.}
{By Computing the Incomplete Gamma function for positive parameters u,N}
{using an infinite series.}
VAR
  i : integer;
  F, C, sum, denom : extended;
BEGIN
{Check admissibility of arguments and value of F}
  IF (u <= 0.0) OR (N <= 0.0) THEN
    BEGIN
      writeln('ERROR in Method8 -- parameter less than or equal zero');
      writeln(' passed value for u is ', u : 10 : 5);
      writeln(' passed value for N is ', N : 10);
      Measure8 := 0.0;
    END
  ELSE
    BEGIN
      F := exp(-N);
{write('Exponential term is ', F : 10);}
      IF F = 0.0 THEN
        BEGIN
          writeln('ERROR in Method8 -- Zero value for F');
          Measure8 := 0.0;
        END
        ELSE
          BEGIN
{Series Begins}
            C := 1.0;
            sum := 1.0;
            denom := u;
            i := 0;
{write(' i = ', i : 3, ' C = ', C : 8 : 5, ' sum = ', sum : 8 : 5);}
            WHILE (C / denom) > Allowed_Error DO
              BEGIN
                denom := denom + 1.0;
              END
\[
C := C \times N / \text{denom}; \\
\text{sum} := \text{sum} + C; \\
i := i + 1; \\
\{ \text{writeln('i = ', i : 3, 'C = ', C : 8 : 5, 'sum = ', sum : 8 : 5);} \}
\]

```
BEGIN;
{ writeln('Last term ', i : 5, ' is', C : 8);} 
Measure8 := sum * F; 
END;
END;
END; {Measure8}
```

```
BEGIN{Repair_Transient}
wind.top := 50; 
wind.left := 10; 
wind.bottom := 300; 
wind.right := 450; 
settextrect(wind); 
showtext; 
Tab := chr(9); 
Allowed_Error := 1e-8; 
delta := 0.1;
```

```
writeln('Probability that Transient Phase following Repairs is significant = P{ TP > 6s }'); 
writeln('parameters are'); 
writeln(' Allowed_Error = ', Allowed_Error : 5); 
writeln(' delta = ', delta : 8 : 5); 
writeln; 
write('N', Tab : 1, ' R = 1 ', Tab : 1, ' R = 10 ', Tab : 1, ' R = 100 ', Tab : 1); 
writeLN(' R = 1K ', Tab : 1, ' R = 10K ', Tab : 1, ' R = 100K ', Tab : 1, ' R = 1M '); 
```

```
FOR count := 0 TO 4 DO 
BEGIN 
N := XpwrY(10, count); 
write(N : 5); 
END; 
FOR count2 := 0 TO 6 DO 
BEGIN 
R := XpwrY(10, count2); 
a := 1 / (delta * R); 
write(Tab : 1, 1.0 - Measure8(a, N) : 10 : 8); 
END; 
writeln; 
END; 
```

END.(Repair_Transient)
APPENDIX B
COMPUTER OUTPUT FOR FAILURE TRANSIENT
Table B.1 Failure Transient \( \rho=0.6, \epsilon=0.1, \delta=0.01 \)

**SINGLE QUEUE SYSTEM WITH ONE CLOSED CUSTOMER**

**PROBABILITY OF FAILURE TRANSIENT SIGNIFICANCE**

\[
\begin{align*}
\rho &= 0.6 \\
\epsilon &= 0.1 \\
\delta &= 0.01
\end{align*}
\]

**TIME TO EQUILIBRIUM DISTRIBUTION**

The definition of steady state for the output process is that the distribution of the \( i \)th cycle time for the closed customer is within \( \epsilon \) of the equilibrium cycle time distribution for any \( x \), for the given parameters, the contraction gives the minimum number of cycles to convergence as 3.

The probability that the output process reaches equilibrium by time \( t \) is computed as the sum of exponentials. The coefficients for each term are:

<table>
<thead>
<tr>
<th>( j )</th>
<th>Coefficient</th>
<th>( j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.74E+00</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-1.19E+01</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.11E+01</td>
<td></td>
</tr>
</tbody>
</table>

\( P\{ \text{Transient Phase is Significant} \} \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=1 )</th>
<th>( R=10 )</th>
<th>( R=100 )</th>
<th>( R=1K )</th>
<th>( R=10K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212044</td>
<td>0.63202505</td>
<td>0.60742638</td>
<td>0.29159069</td>
<td>0.04326254</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995448</td>
<td>0.99991306</td>
<td>0.96816956</td>
<td>0.35741916</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.98799733</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=100K )</th>
<th>( R=1M )</th>
<th>( R=10M )</th>
<th>( R=100M )</th>
<th>( R=1000M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00453562</td>
<td>0.00045576</td>
<td>0.00004560</td>
<td>0.00000456</td>
<td>0.00000046</td>
</tr>
<tr>
<td>10</td>
<td>0.04444159</td>
<td>0.00454822</td>
<td>0.00045588</td>
<td>0.00004560</td>
<td>0.00000456</td>
</tr>
<tr>
<td>100</td>
<td>0.36529434</td>
<td>0.04456248</td>
<td>0.00045948</td>
<td>0.00045589</td>
<td>0.00004560</td>
</tr>
<tr>
<td>1000</td>
<td>0.98938980</td>
<td>0.36596683</td>
<td>0.04457460</td>
<td>0.00454960</td>
<td>0.00045590</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>0.98952319</td>
<td>0.36617723</td>
<td>0.04457581</td>
<td>0.00454962</td>
</tr>
</tbody>
</table>

Page 124
Table B.2 Failure Transient rho=0.8, epsilon=0.1, delta=0.01

SINGLE QUEUE SYSTEM WITH ONE CLOSED CUSTOMER
PROBABILITY OF FAILURE TRANSIENT SIGNIFICANCE

\[
\begin{align*}
\text{rho} & = \quad 0.8 \\
\text{epsilon} & = \quad 0.1 \\
\text{delta} & = \quad 0.01
\end{align*}
\]

TIME TO EQUILIBRIUM DISTRIBUTION

The definition of steady state for the output process is that
the distribution of the ith cycle time for the closed customer is within
epsilon of the equilibrium cycle time distribution for any x,
For the given parameters,
the contraction gives the minimum number of cycles to convergence as

the probability that the output process reaches equilibrium by time t is
computed as the sum of exponentials. The coefficients for each term are

<table>
<thead>
<tr>
<th>j</th>
<th>Coefficient j</th>
<th>j</th>
<th>Coefficient j</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.37E-02</td>
<td>6</td>
<td>-6.27E+03</td>
</tr>
<tr>
<td>2</td>
<td>-9.08E+00</td>
<td>7</td>
<td>2.57E+03</td>
</tr>
<tr>
<td>3</td>
<td>2.57E+02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-1.90E+03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.36E+03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \Pr \text{ Transient Phase is Significant} \]

<table>
<thead>
<tr>
<th>N</th>
<th>R=1</th>
<th>R=10</th>
<th>R=100</th>
<th>R=1K</th>
<th>R=10K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.53207284</td>
<td>0.55630838</td>
<td>0.15822524</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.99995454</td>
<td>0.99970433</td>
<td>0.82136810</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99999997</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>R=100K</th>
<th>R=1M</th>
<th>R=10M</th>
<th>R=100M</th>
<th>R=1000M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01880315</td>
<td>0.00191546</td>
<td>0.00019190</td>
<td>0.00001919</td>
<td>0.00000192</td>
</tr>
<tr>
<td>10</td>
<td>0.17289348</td>
<td>0.01899030</td>
<td>0.00191738</td>
<td>0.00019192</td>
<td>0.00001919</td>
</tr>
<tr>
<td>100</td>
<td>0.85076459</td>
<td>0.17446969</td>
<td>0.01900917</td>
<td>0.00191757</td>
<td>0.00019192</td>
</tr>
<tr>
<td>1000</td>
<td>0.99999999</td>
<td>0.85299563</td>
<td>0.17462847</td>
<td>0.01901106</td>
<td>0.00191759</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.85327812</td>
<td>0.17464436</td>
<td>0.01901125</td>
</tr>
</tbody>
</table>
Table B.3 Failure Transient \( \rho_{\theta}=0.9 \), \( \epsilon=0.1 \), \( \Delta=0.01 \)

**SINGLE QUEUE SYSTEM WITH ONE CLOSED CUSTOMER**

**PROBABILITY OF FAILURE TRANSIENT SIGNIFICANCE**

\[
\begin{align*}
\rho &= 0.9 \\
\epsilon &= 0.1 \\
\Delta &= 0.01
\end{align*}
\]

**TIME TO EQUILIBRIUM DISTRIBUTION**

The definition of steady state for the output process is that the distribution of the \( \nu \)th cycle time for the closed customer is within \( \epsilon \) of the equilibrium cycle time distribution for any \( x \). For the given parameters, the contraction gives the minimum number of cycles to convergence as

the probability that the output process reaches equilibrium by time \( t \) is computed as the sum of exponentials. The coefficients for each term are

<table>
<thead>
<tr>
<th>( j )</th>
<th>Coefficient</th>
<th>( j )</th>
<th>Coefficient</th>
<th>( j )</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.66E-08</td>
<td>6</td>
<td>5.47E+05</td>
<td>11</td>
<td>-3.83E+08</td>
</tr>
<tr>
<td>2</td>
<td>2.95E-03</td>
<td>7</td>
<td>-5.62E+08</td>
<td>12</td>
<td>3.34E+08</td>
</tr>
<tr>
<td>3</td>
<td>-3.59E+00</td>
<td>8</td>
<td>3.31E+07</td>
<td>13</td>
<td>-1.62E+08</td>
</tr>
<tr>
<td>4</td>
<td>5.78E+02</td>
<td>9</td>
<td>-1.19E+08</td>
<td>14</td>
<td>3.37E+07</td>
</tr>
<tr>
<td>5</td>
<td>-2.73E+04</td>
<td>10</td>
<td>2.69E+08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Transient Phase is Significant)

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=1 )</th>
<th>( R=10 )</th>
<th>( R=100 )</th>
<th>( R=1K )</th>
<th>( R=10K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.63072235</td>
<td>0.39127180</td>
</tr>
<tr>
<td>10</td>
<td>0.999995460</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.99995284</td>
<td>0.99301391</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>( N )</td>
<td>( R=100K )</td>
<td>( R=1M )</td>
<td>( R=10M )</td>
<td>( R=100M )</td>
<td>( R=1000M )</td>
</tr>
<tr>
<td>1</td>
<td>0.06570437</td>
<td>0.00700733</td>
<td>0.00070537</td>
<td>0.00007058</td>
<td>0.00000706</td>
</tr>
<tr>
<td>10</td>
<td>0.49319236</td>
<td>0.06790414</td>
<td>0.00070313</td>
<td>0.00007062</td>
<td>0.000007059</td>
</tr>
<tr>
<td>100</td>
<td>0.99888203</td>
<td>0.50500085</td>
<td>0.06813015</td>
<td>0.00703379</td>
<td>0.00070564</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>0.99911683</td>
<td>0.50619827</td>
<td>0.06815282</td>
<td>0.00703399</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99913796</td>
<td>0.50631840</td>
<td>0.06815467</td>
</tr>
</tbody>
</table>
Table B.4 Failure Transient \( \rho=0.8, \varepsilon=0.1, \delta=0.001 \)

SINGLE QUEUE SYSTEM WITH ONE CLOSED CUSTOMER
PROBABILITY OF FAILURE TRANSIENT SIGNIFICANCE

\[
\begin{align*}
\rho_0 &= 0.8 \\
\varepsilon &= 0.1 \\
\delta &= 0.001 \\
\end{align*}
\]

TIME TO EQUILIBRIUM DISTRIBUTION

the definition of steady state for the output process is that the distribution of the \( i \)th cycle time for the closed customer is within \( \varepsilon \) of the equilibrium cycle time distribution for any \( x \), for the given parameters, the contraction gives the minimum number of cycles to convergence as

the probability that the output process reaches equilibrium by time \( t \) is computed as the sum of exponentials. The coefficients for each term are

<table>
<thead>
<tr>
<th>( j )</th>
<th>Coefficient ( j )</th>
<th>( j )</th>
<th>Coefficient ( j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.37E-02</td>
<td>6</td>
<td>-6.27E+03</td>
</tr>
<tr>
<td>2</td>
<td>-9.08E+00</td>
<td>7</td>
<td>2.57E+03</td>
</tr>
<tr>
<td>3</td>
<td>2.57E+02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-1.90E+03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.36E+03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( P(\text{Transient Phase is Significant}) \)

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=1 )</th>
<th>( R=10 )</th>
<th>( R=100 )</th>
<th>( R=1K )</th>
<th>( R=10K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.63212056</td>
<td>0.63207284</td>
<td>0.55630838</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.99995460</td>
<td>0.99995454</td>
<td>0.99970433</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.000000000</td>
<td>1.000000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.000000000</td>
<td>1.000000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.000000000</td>
<td>1.000000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( N )</th>
<th>( R=100K )</th>
<th>( R=1M )</th>
<th>( R=10M )</th>
<th>( R=100M )</th>
<th>( R=1000M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15822524</td>
<td>0.01880315</td>
<td>0.00191546</td>
<td>0.00019190</td>
<td>0.00001919</td>
</tr>
<tr>
<td>10</td>
<td>0.82136810</td>
<td>0.17289348</td>
<td>0.01899030</td>
<td>0.00191738</td>
<td>0.00019192</td>
</tr>
<tr>
<td>100</td>
<td>0.99999997</td>
<td>0.85016459</td>
<td>0.17446969</td>
<td>0.01900917</td>
<td>0.00191757</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>0.99999999</td>
<td>0.85295663</td>
<td>0.17462847</td>
<td>0.01901106</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.85327812</td>
<td>0.17464436</td>
</tr>
</tbody>
</table>

127
Table B.5 Failure Transient $\rho=0.8$, $\epsilon=0.1$, $\delta=0.1$

SINGLE QUEUE SYSTEM WITH ONE CLOSED CUSTOMER

PROBABILITY OF FAILURE TRANSIENT SIGNIFICANCE

$$\begin{align*}
\rho &= 0.8 \\
\epsilon &= 0.1 \\
\delta &= 0.1
\end{align*}$$

TIME TO EQUILIBRIUM DISTRIBUTION

The definition of steady state for the output process is that the distribution of the $i$th cycle time for the closed customer is within $\epsilon$ of the equilibrium cycle time distribution for any $x$. For the given parameters, the contraction gives the minimum number of cycles to convergence as 7.

The probability that the output process reaches equilibrium by time $t$ is computed as the sum of exponentials. The coefficients for each term are

<table>
<thead>
<tr>
<th>$j$</th>
<th>Coefficient $j$</th>
<th>$j$</th>
<th>Coefficient $j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.37E-02</td>
<td>6</td>
<td>-6.27E+03</td>
</tr>
<tr>
<td>2</td>
<td>-9.08E+00</td>
<td>7</td>
<td>2.57E+03</td>
</tr>
<tr>
<td>3</td>
<td>2.57E+02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-1.90E+03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.36E+03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$P(\text{Transient Phase is Significant})$

<table>
<thead>
<tr>
<th>$N$</th>
<th>$R=1$</th>
<th>$R=10$</th>
<th>$R=100$</th>
<th>$R=1K$</th>
<th>$R=10K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63212056</td>
<td>0.63207284</td>
<td>0.55630838</td>
<td>0.15822524</td>
<td>0.01880315</td>
</tr>
<tr>
<td>10</td>
<td>0.99995460</td>
<td>0.99995454</td>
<td>0.99970433</td>
<td>0.82136810</td>
<td>0.17289348</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99999997</td>
<td>0.85016459</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99999999</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$N$</th>
<th>$R=10K$</th>
<th>$R=1M$</th>
<th>$R=10M$</th>
<th>$R=100M$</th>
<th>$R=1000M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00191546</td>
<td>0.00019190</td>
<td>0.00001919</td>
<td>0.00000192</td>
<td>0.00000019</td>
</tr>
<tr>
<td>10</td>
<td>0.01899030</td>
<td>0.00191738</td>
<td>0.00019192</td>
<td>0.00001919</td>
<td>0.00000192</td>
</tr>
<tr>
<td>100</td>
<td>0.17446969</td>
<td>0.01900917</td>
<td>0.00191757</td>
<td>0.00019192</td>
<td>0.00001919</td>
</tr>
<tr>
<td>1000</td>
<td>0.85299563</td>
<td>0.17462847</td>
<td>0.01901106</td>
<td>0.00191759</td>
<td>0.00019192</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>0.85327812</td>
<td>0.17464436</td>
<td>0.01901125</td>
<td>0.00191759</td>
</tr>
</tbody>
</table>

128
Table B.6 Failure Transient \( \rho=0.8, \epsilon=0.01, \delta=0.01 \)

\[
\begin{align*}
\rho &= 0.8 \\
\epsilon &= 0.01 \\
\delta &= 0.01 \\
\end{align*}
\]

**TIME TO EQUILIBRIUM DISTRIBUTION**

The definition of steady state for the output process is that the distribution of the ith cycle time for the closed customer is within \( \epsilon \) of the equilibrium cycle time distribution for any \( x \). For the given parameters, the contraction gives the minimum number of cycles to convergence as 17.

The probability that the output process reaches equilibrium by time \( t \) is computed as the sum of exponentials. The coefficients for each term are

<table>
<thead>
<tr>
<th>j</th>
<th>Coefficient</th>
<th>j</th>
<th>Coefficient</th>
<th>j</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.17E-08</td>
<td>6</td>
<td>-8.19E+09</td>
<td>11</td>
<td>2.35E+15</td>
</tr>
<tr>
<td>2</td>
<td>-1.04E-01</td>
<td>7</td>
<td>2.83E+11</td>
<td>12</td>
<td>-7.80E+15</td>
</tr>
<tr>
<td>3</td>
<td>8.20E+02</td>
<td>8</td>
<td>-5.45E+12</td>
<td>13</td>
<td>1.74E+16</td>
</tr>
<tr>
<td>4</td>
<td>-6.16E+05</td>
<td>9</td>
<td>6.37E+13</td>
<td>14</td>
<td>-2.55E+16</td>
</tr>
<tr>
<td>5</td>
<td>1.15E+08</td>
<td>10</td>
<td>-4.76E+14</td>
<td>15</td>
<td>2.36E+16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>j</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>-1.23E+16</td>
</tr>
<tr>
<td>17</td>
<td>2.79E+15</td>
</tr>
</tbody>
</table>

\( P(\text{Transient Phase is Significant}) \)

<table>
<thead>
<tr>
<th>N</th>
<th>( R=1 )</th>
<th>( R=10 )</th>
<th>( R=100 )</th>
<th>( R=1K )</th>
<th>( R=10K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63248805</td>
<td>0.63252450</td>
<td>0.63247964</td>
<td>0.63355478</td>
<td>0.38391787</td>
</tr>
<tr>
<td>10</td>
<td>0.99995505</td>
<td>0.99995510</td>
<td>0.99995504</td>
<td>0.99995634</td>
<td>0.99212254</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>( R=100K )</th>
<th>( R=1M )</th>
<th>( R=10M )</th>
<th>( R=100M )</th>
<th>( R=1000M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.06789751</td>
<td>0.01934175</td>
<td>0.01650356</td>
<td>0.00778206</td>
<td>0.01165035</td>
</tr>
<tr>
<td>10</td>
<td>0.50496410</td>
<td>0.17742244</td>
<td>0.14465467</td>
<td>0.07515119</td>
<td>0.11058159</td>
</tr>
<tr>
<td>100</td>
<td>0.99911617</td>
<td>0.85816984</td>
<td>0.79038861</td>
<td>0.54216664</td>
<td>0.69021445</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>0.99999984</td>
<td>0.9999535</td>
<td>0.99999186</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>j</td>
<td>Coefficient</td>
<td>j</td>
<td>Coefficient</td>
<td>j</td>
<td>Coefficient</td>
</tr>
<tr>
<td>----</td>
<td>-------------</td>
<td>----</td>
<td>-------------</td>
<td>----</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>4.02E-09</td>
<td>8</td>
<td>-1.64E+14</td>
<td>15</td>
<td>6.15E+19</td>
</tr>
<tr>
<td>2</td>
<td>-3.45E-02</td>
<td>9</td>
<td>3.43E+15</td>
<td>16</td>
<td>-8.37E+19</td>
</tr>
<tr>
<td>3</td>
<td>7.94E+02</td>
<td>10</td>
<td>-4.56E+16</td>
<td>17</td>
<td>7.29E+19</td>
</tr>
<tr>
<td>4</td>
<td>-1.39E+06</td>
<td>11</td>
<td>4.05E+17</td>
<td>18</td>
<td>-3.65E+19</td>
</tr>
<tr>
<td>5</td>
<td>5.34E+08</td>
<td>12</td>
<td>-2.46E+18</td>
<td>19</td>
<td>7.93E+18</td>
</tr>
<tr>
<td>6</td>
<td>-7.37E+10</td>
<td>13</td>
<td>1.03E+19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.73E+12</td>
<td>14</td>
<td>-3.04E+19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P{ Transient Phase is Significant}

<table>
<thead>
<tr>
<th>N</th>
<th>R=1</th>
<th>R=10</th>
<th>R=100</th>
<th>R=1K</th>
<th>R=10K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.22119922</td>
<td>0.93607214</td>
<td>0.99999969</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10</td>
<td>0.91791500</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>R=100K</th>
<th>R=1M</th>
<th>R=10M</th>
<th>R=100M</th>
<th>R=1000M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>100</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>1000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
<tr>
<td>10000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
<td>1.00000000</td>
</tr>
</tbody>
</table>
APPENDIX C
PASCAL PROGRAM FOR COMPUTING FAILURE TRANSIENT SIGNIFICANCE

UNIT Cycle_Time_Distribution_Convergence;

INTERFACE
USES
  SANE;
CONST
  max_cycles = 25;
TYPE
  Coeff_vector = ARRAY[1..max_cycles] OF extended;
  Time_vector = ARRAY[0..100] OF extended;
FUNCTION Contraction (r, r : extended) : integer;

PROCEDURE A_Coeff (VAR A : Coeff_vector;
  i : integer;
  rho : extended);

PROCEDURE Print_Coeffs (A : Coeff_vector;
  i : integer);
FUNCTION Fss (A : Coeff_vector;
  i : integer;
  m, r, t : extended) : extended;

PROCEDURE OP_Convrg_Dist_Fcn (epsilon, rho : extended;
  VAR A : Coeff_vector;
  VAR min_cyc : integer);
PROCEDURE Plot_Convrg_Dist (A : Coeff_vector;
  min_cyc : integer;
  mu, rho : extended;
  step_size : extended;
  num_step : integer;
  VAR I's : Time_vector);
PROCEDURE Print_Time (Time_Fss, Time_Dif : Time_vector;
  Step_Size : extended;
  num_Steps : integer); IMPLEMENTATION

FUNCTION Contraction;
VAR
  Old_x, New_x, junk : extended;
BEGIN (Contraction)
  Old_x := 0.0;
  New_x := ln(e) / ln(r);
  WHILE abs(Old_x - New_x) > 0.001 DO
    BEGIN
      Old_x := New_x;
      junk := 1.0 - XpwrY(r, Old_x);
      New_x := (ln(e) - ln(junk) * junk) / XpwrY(r, Old_x) / ln(r);
    END;
  Contraction := trunc(new_x + 1.0);
END; {Contraction}

PROCEDURE A_Coeff;
VAR
  j, k: integer;
  num, denom: extended;
BEGIN{A_Coeff}
FOR j := 1 TO max_cycles DO
BEGIN
  IF j <= i THEN
  BEGIN
    num := exp((i - 1) * ln(1.0 - XpwrY(rho, j)));
    denom := 1.0;
    FOR k := 1 TO i DO
      IF k <> j THEN
        denom := denom * (XpwrY(rho, k) - XpwrY(rho, j));
      Al[j] := num / denom;
  END
  ELSE
    Al[j] := 0.0;
  END
END;
END{A_Coeff}

PROCEDURE Print_Coeffs;
VAR
  j: integer;
  Tab: char;
BEGIN{Print_Coeffs}
  Tab := chr(9);
  writeln(Tab, 'j', Tab, 'Coefficient j');
  FOR j := 1 TO i DO
    writeln(Tab, j : 3, Tab, Al[j] : 15);
  writeln;
END{Print_Coeffs}

FUNCTION Fss;
VAR
  k: integer;
  temp, junk: extended;
BEGIN{Fss}
  temp := 0.0;
  FOR k := 1 TO i DO
    BEGIN
      junk := m * (1 - r) / (1.0 - XpwrY(r, k));
      temp := temp + Al[k] * (1.0 - exp(-junk * t));
    END;
  Fss := temp;
END{Fss}

PROCEDURE OP_Converg_Dist_Fcn;
VAR
  Tab: char;
BEGIN{OP_Converg_Dist_Fcn}
  Tab := chr(9);
  writeln;
  writeln('TIME TO EQUILIBRIUM DISTRIBUTION');
  writeln;

  {Use a contraction to Compute the minimum number of cycles needed to meet steady state requirements}
min_cyc := Contraction(epsilon, rho);
writeln;
writeln('the definition of steady state for the output process is that');
writeln('the distribution of the ith cycle time for the closed customer is within ');
writeln('epsilon of the equilibrium cycle time distribution for any x, ');
writeln('for the given parameters, ');
writeln('the contraction gives the minimum number of cycles to convergence as');

{compute the coefficients of the exponentials in Fss(t)}
writeln;
writeln('the probability that the output process reaches equilibrium by time t is');
writeln('computed as the sum of exponentials. The coefficients for each term are');
writeln;
A_Coeff(A, min_cyc, rho);
Print_Coeff(A, min_cyc);
END;{OP_Converg_Dist_Fcn}
PROCEDURE Plot_Converg_Dist;
VAR
  t, value : extended;
  step_num : integer;
  Tab : char;
BEGIN(Plot_Converg_Dist)
  Tab := chr(9);
  {Compute Fss(t)}
  writeln;
  writeln('the probability that the output process reaches equilibrium by time t is Fss(t)');
  writeln;
  writeln(Tab, ' t = ', Tab, ' Fss(t) = ');
  t := 0.0;
  value := Fss(A, min_cyc, mu, rho, t);
  writeln(Tab, t : 8 : 5, Tab, value : 8 : 5);
  Fss[0] := value;
  FOR step_num := 1 TO num_step DO
    BEGIN{step_num}
      t := t + step_size;
      value := Fss(A, min_cyc, mu, rho, t);
      writeln(Tab, t : 8 : 5, Tab, value : 8 : 5);
      Fss[step_num] := value;
    END;{step_num}
  END;{Plot_Converg_Dist}
PROCEDURE Print_Time;
VAR
  i : integer;
  Tab : char;
BEGIN(Print_Time)
  Tab := chr(9);
  writeln;
  writeln(' Time ', Tab, ' Fss(t) ', Tab, 'Occ. Diff');
  FOR i := 0 TO num_Steps DO
BEGIN
writeln (Tab, (i * Step_Size) : 8 : 5, Tab, Time_Fss[i] : 8 : 5, Tab, Time_Diff[i] : 8 : 5);
END;[i]
writeln:
END;{Print_Time}

END.{Cycle_Time_Distribution_Convergence}

PROGRAM Failure_Transient_Signif (input, output);
USES
SANE, Cycle_Time_Distribution_Convergence;
TYPE
results_array = ARRAY[1..10, 1..10] OF extended;
VAR
Results : results_array;
C_k : Coeff_Vector;
op_cyc, k, odd, N_loop, R_loop : integer;
epsilon, delta : extended;
R, N, even_coeff, Odd_coeff : extended;
mu, rho : extended;
P_T_Call_ns, P_TP_Signif, junk1, junk2 : extended;
Tab : char;
wind : rect;

PROCEDURE Input_Parameter_report;
BEGIN{Input_Parameter_report}
writeln;
writeln(Tab, 'rho = ', Tab, rho : 7 : 3);
writeln(Tab, 'epsilon = ', Tab, epsilon : 7 : 3);
writeln(Tab, 'delta = ', Tab, delta : 7 : 3);
writeln;
END;{Input_Parameter_report}

BEGIN{Failure_Transient_Signif}
wind.top := 50;
wind.left := 10;
wind.bottom := 300;
wind.right := 450;
settextext(wind);
showtext;
Tab := chr(9);

writeln(' SINGLE QUEUE SYSTEM WITH ONE CLOSED CUSTOMER ');
writeln(' PROBABILITY OF FAILURE TRANSIENT SIGNIFICANCE');
rho := 0.8;
epsilon := 0.006;
delta := 0.01;
Input_Parameter_report;
OP_Covg_Dist_Fcn(epsilon, rho, C_k, op_cyc);
R := 1;
FOR R_loop := i TO 10 DO
BEGIN
junk1 := delta * R * (1 - rho);
Odd_coeff := 0;

```
Even_coff := 0;
odd := 1;
FOR k := 1 TO op_cyc DO
BEGIN
junk2 := 1 + junk1 / (1 - XpwrY(rho, k));
IF odd = 1 THEN
BEGIN
Odd_coff := Odd_coff + C_k[k] * (1 - 1 / junk2);
odd := 0;
END
ELSE
BEGIN
Even_coff := Even_coff + C_k[k] * (1 - 1 / junk2);
odd := 1;
END;
END;
END{k}
P_TP_1_Call_ns := Odd_coff + Even_coff;
{writeln(' R = ', R : 10 : 2, Tab, (1 - P_TP_1_Call_ns) : 8 : 5);}
N := 1;
FOR N_loop := 1 TO 5 DO
BEGIN
P_TP_Signif := 1.0 - exp(-N * (1.0 - P_TP_1_Call_ns));
Result[N_loop, R_loop] := P_TP_Signif;
writeln(Tab, ' N = ', N : 10 : 2, Tab, P_TP_Signif : 8 : 5);
N := N * 10;
END;
writeln(Tab, 'P (Transient Phase is Significant)');
writeln(Tab, ' R = 1 ', Tab, ' R = 10 ', Tab, ' R = 100 ');
writeln(Tab, ' R = 1K ', Tab, ' R = 10K ');
VITA

John Kobza was born in Anchorage, Alaska on 30 March 1959. He received the B.S. degree in Electrical Engineering from Washington State University in 1982, and the M.S. degree in Electrical Engineering from Clemson University in 1984. From 1984 to 1987 he was a Member of the Technical Staff at GTE Laboratories where he worked on traffic scheduling for integrated communications networks. He was an IBM Fellow at the William E. Simon Graduate School of Business, University of Rochester during the 1987/88 academic year. Since 1988 he has been a graduate student in the Industrial and Systems Engineering Department at Virginia Tech where he recently completed the requirements for the Ph.D. degree. His research interests include stochastic processes, queueing theory, and performance analysis of communications networks.

John E. Kobza