UNIVERSITY OF TEXAS AT ARLINGTON

Stochastic Reliability Models for a General Server and Related Networks

Author:
Rachel Traylor

Supervisor:
Dr. Andrzej Korzeniowski

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Department of Mathematics

May 2016

Abstract

Stochastic Reliability Models for a General Server and Related Networks

by Rachel Traylor

There are many types of systems which can be dubbed servers, i.e. a retail checkout counter, a shipping company, a web server, or a customer service hotline. All of these systems have common general behavior: requests or customers arrive via a stochastic process, the service times vary randomly, and each request increases the stress on the server for some interval of time. A general stochastic model that describes the reliability of a server can provide the necessary information-for optimal resource allocation and efficient task scheduling, leading to significant cost savings and improved performance metrics. In this work, we consider several generalizations of existing stochastic reliability models that incorporate random workloads, load-balancing allocation, and clustered tasks. The efficiency of the described servers is studied extensively in order to facilitate the design and implementation of control policies for fast-paced environments such as IT applications. Finally, a method to determine the reliability of any network of general servers, both correlated and uncorrelated, is presented.

Acknowledgements

First, I would like to thank my advisor, Professor Andzrej Korzeniowski, for facilitating my development as an independent researcher and the strong support given as I pursued this project. I am forever grateful for his encouragement and positive attitude while I found my footing as a mathematician. I also give my sincerest thanks to my committee members, Drs. Gaik Ambartsoumian, Chien-Pai Han, D.L. Hawkins, and Shan Sun-Mitchell.

I would like to acknowledge my fellow graduate students, particularly Jason Hathcock, who became a dear friend as we bonded over mathematics, metal, and LaTeX. In addition, I offer my sincerest thanks to Dr. Barbara Shipman, who was never too busy to lend an ear to mathematics discussions spanning every branch. To my mother, Donna; stepfather,Barry; father, Ian; and stepmother, Kelly, I am indebted for their tireless support since my childhood. I'd like to acknowledge my siblings, Tess and Derek, for their encouragement as well.

Lastly, I'd like to acknowledge my husband, Nicholas Traylor. For years he has tolerated mathematics books and papers strewn about the house, mirrors used as whiteboards, and my inability to find car keys because my mind was preoccupied with Greek letters. I couldn't have made it this far and retained my sanity without him.

Contents

List of Figures

List of Tables

List of Abbreviations

NHPP	**N**on**h**omogenous **P**oisson **P**rocesses
RSBR	**R**andom **S**tress **B**reakdown **R**ate

List of Symbols

$\phi(x)$	structure function
$\lambda(x)$	intensity function for NHPP
$N(t)$	number of arrivals in $[0, t]$
$T_j/\mathfrak{T}$	random arrival time of request j/set of all arrival times
$W_j/\mathfrak{W}$	random time to completion of request j/set of all service times
$\mathcal{H}_j/\eta/\mathfrak{H}$	random stress to the server of request $j$fixed stress of request j/set of all service times
Y	server lifetime (length of renewal cycle)
$g_W(w)$	pdf of service time distribution
$\bar{G}_W(w)$	complement of CDF of service times
ψ	server efficiency
ν	mean time to server reboot
$M(t)/M$	jobs completed by the server in $[0, t]$/jobs completed by the server during a renewal cycle

To Andy and Debbie Kohler, who inspired me to become a mathematician

Chapter 1

Introduction

There are many types of systems which can be dubbed servers, such as a retail checkout counter, a shipping company, a web server, or a customer service hotline. All of these systems have common general behavior. Requests or customers arrive via a stochastic process, the service times vary randomly, and each request stresses the server if only temporarily. A general stochastic model that describes the reliability of such a server can provide the necessary information for optimal resource allocation and efficient task scheduling, leading to significant cost savings for businesses and improved performance metrics[6]. Such topics have been studied in literature for several decades [1, 2, 3, 21].

Much attention was devoted to reliability principles that model software failures and bug fixes, starting with Jelinski and Moranda in 1972 [11]. The hazard function under this model shows the time between the ith failure and the $i+1$st failure. Littlewood (1980) [17] extended this initial reliability model for software by assuming differences in error size. [12].

These models have been extended into software testing applications [4, 5, 20] and optimal software release times [7, 8, 18, 24]. The explosion of e-commerce and the resulting increase in internet traffic have led to the development of reliability models for Web applications. Heavy traffic can overload and crash a server; thus, various control policies for refreshing content and admission of page requests were created [10, 13, 16, 19, 26].

In particular, Cha and Lee (2011) [9] proposed a stochastic breakdown model for an unreliable web server whose requests arrive at random times according to a nonhomogenous Poisson process and bring a constant stress factor to the system that dissipates upon service completion. The authors provide a fairly general survival function under any service distribution $g_W(w)$, define server efficiency to measure performance, and illustrate a possible admission control policy due to an observed property of the server efficiency under a specific numerical example.

Thus far, no extensions of [9] have been proposed. This work generalizes the model put forth by Cha and Lee in a variety of ways. First, the assumption of constant job stress is relaxed and replaced by a random variable, and a new survival function and efficiency equation are derived.

Motivated by the numerical example in [9], the effect of the service distribution on the efficiency is studied for several key cases: the Erlang, uniform, and Exponential classes. Further extensions of the single-server model presented herein include multichannel servers under a load-balancing allocation scheme and under clustered task assignment, and the respective failure distributions are derived. Futhermore, systems of servers under a random stress environment are studied. The system survival function for both correlated systems and independent systems are derived, and an isomorphism with the reliability structure function is established, allowing for a straightforward procedure to determine the survival function of any system of servers under very few general assumptions. This work, while suitable for IT applications, is general enough for use in almost any industry, including logistics, retail, manufacturing, and engineering systems.

The remainder of this chapter provides some preliminary concepts and theorems relating to the work of the model proposed by Cha and Lee in [9].

Chapter 2 details the first generalization of [9] in which the constant stress assumption is relaxed in favor of a general random variable. The survival function and efficiency are derived. Chapter 3 presents a study on the effect of the service time distribution on the existence of a finite maximum efficiency for several choice distributions. In particular, it is shown that the exponential distribution is a poor choice for modeling service life distribution. Chapter 4 extends the single-server model presented in Chapter 2 to two different instances of a multichannel server. A server under a load-balancing allocation scheme common in many different scenarios is investigated in addition to a server whose requests "pick" tasks in clusters (both independently and dependently). Chapter 5 builds systems of the random stress servers developed in Chapter 2 and shows the isomorphism between the conditional survival function of the system and the reliability structure function, thus providing an immediate and straightforward methodology for obtaining any system survival function under the assumptions presented in Chapter 2. Auxiliary lemmas and other details are given in Appendices A and B.

1.1 Preliminaries

Nonhomogeneous Poisson Process

Definition 1.1 (Nonhomogenous Poisson Process). *A counting process* $\{N(t), t \geq 0\}$ *is a* ***nonhomogenous Poisson process*** *(NHPP) with intensity function* $\lambda(t) \geq 0$ *if the following hold:*

(1) $N(0) = 0$

(2) $\{N(t), t \geq 0\}$ *has independent increments*

(3) $P(N(t+h) - N(t) \geq 2) = o(h)$

(4) $P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$

Fact.

$$P(N(t+s) - N(t) = n) = \frac{e^{-[m(t+s)-m(t)]}\,(m(t+s) - m(t))^n}{n!} \tag{1.1}$$

A NHPP process has independent increments but is no longer stationary; events may be more likely to occur at certain times than others.

Renewal Theory: Renewal Reward Processes

Definition 1.2 (Renewal Process)**.** *Let* $\{X_n : n \in \mathbb{N}\}$ *be a sequence of nonnegative independent random variables with common distribution* F*, where* X_n *is the time between the* $(n-1)$*st and* n*th event. Let* $S_n = \sum_{i=1}^{n} X_i$*, where* $S_0 = 0$ *be the time of the* n*th event. Define* $N(t) = \sup\{n : S_n \leq t\}$*. Then* $\{N(t), t \geq 0\}$ *is called a* ***renewal process***

Suppose $\{N(t), t \geq 0\}$ is a renewal process, and upon each renewal a reward is received, denoted by R_n. $\{R_n : n \in \mathbb{N}\}$ are i.i.d. random variables but may depend on the length of the renewal interval X_n. Thus (X_n, R_n) are i.i.d. Let $R(t) = \sum_{i=1}^{N(t)} R_n$ be the cumulative reward by time t. We have the following:

Theorem 1.1. *Suppose* $E[R_n] < \infty$ *and* $E[X_n] < \infty$*. Then*

(i) $\frac{R(t)}{t} \xrightarrow{W.P.1} \frac{E[R_n]}{E[X_n]}$ *as* $t \to \infty$

(ii) $\frac{E[R(t)]}{t} \to \frac{E[R_n]}{E[X_n]}$ *as* $t \to \infty$

The above are standard facts from renewal theory and may be found, for example, in [23].

Reliability Theory and Survival Analysis

Structure Functions

A system is defined as a collection on n components, each with a binary assumption of failed or functional.

Definition 1.3 (Component State)**.** *The* ***state*** *of component* i *is denoted* x_i *and is defined by*

$$x_i := \begin{cases} 1, & \text{component } i \text{ is functional} \\ 0, & \text{component } i \text{ is failed} \end{cases}$$

Definition 1.4 (System State Vector)**.** *For* n *components in a system, each with state* x_i*, the system state is defined by* $\mathbf{x} = (x_1, ..., x_n)$*.*

There are 2^n possible state vectors $\mathbf{x}$ for any particular $n-$component system. The notion of order of state vectors is done component-wise. $\mathbf{x} < \mathbf{y}$ if and only if $x_i \leq y_i$ for all $i = 1, ..., n$ and $x_i < y_i$ for some i.

Definition 1.5 (Structure Function). *The **structure function** of a system $\phi : \mathbf{x} \to \{0, 1\}$ is defined as*

$$\phi(\mathbf{x}) := \begin{cases} 1, & \textit{the system is functioning when the state vector is } \mathbf{x} \\ 0, & \textit{the system has failed when the state vector is } \mathbf{x} \end{cases}$$

Basic Systems and Block Diagrams

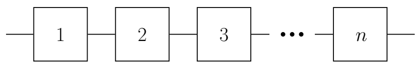


FIGURE 1.1: Block Diagram-Series System

A **series system** of n components is functioning if and only if all components are functioning. Thus, the structure function of a series system is given by

$$\phi_{\text{series}}(\mathbf{x}) = \prod_{i=1}^{n} x_i$$

The series system is represented in a *block diagram*, given in Figure 1.1. The block diagram is similar to a circuit diagram, in that if a path from left to right can be traced through functioning components, the system is operational. It should be noted that the block diagram is logical rather than physical.

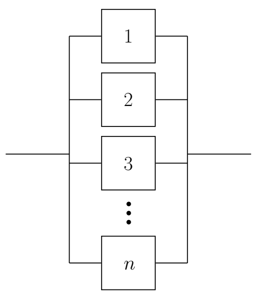


FIGURE 1.2: Block Diagram-Parallel System

A **parallel system** of n components functions if and only if at least one component is functioning. It can be defined equivalently by failure, in that a parallel system fails if and only if all components fail. Using the latter definition, the structure function is given by

$$\phi_{\text{parallel}}(\mathbf{x}) = 1 - \prod_{i=1}^{n}(1 - x_i)$$

Definition 1.6 (Minimal Path Set)**.** $\mathbf{x}$ *is a* ***path set*** *if* $\phi(\mathbf{x}) = 1$. $\mathbf{x}$ *is a* ***minimal path set*** *if* $\phi(\mathbf{y}) = 0$ *for any* $\mathbf{y} < \mathbf{x}$.

From Leemis (2.2) [15] ,

Theorem 1.2 (Decomposition of Systems into Series/Parallel Subsystems)**.** *Let* $P_1, ..., P_s$ *be the minimal path sets for a system. Then*

$$\phi(\mathbf{x}) = 1 - \prod_{i=1}^{s}\left(1 - \prod_{j \in P_i} x_j\right)$$

Proof. Let $\alpha_i = \prod_{j \in P_i} x_j$. The system fails if and only if no path exists through the system; i.e. all minimal path sets have failed. Thus

$$\phi(\mathbf{x}) = 1 - \prod_{i=1}^{s}(1 - \alpha_i) = 1 - \prod_{i=1}^{s}\left(1 - \prod_{j \in P_i} x_j\right)$$

□

Lifetime Distribution Representations

Definition 1.7 (Survivor Function)**.** *The* ***survivor function****, denoted* $S(t)$*, is the probability the system lifetime* Y *exceeds the time* t*.*

$$S(t) := P(Y \geq t), t \geq 0$$

The lifetime Y also has a pdf $f(t) := -\frac{dS}{dt}$.

Definition 1.8 (Hazard Function/Breakdown Rate)**.** *The* ***hazard function****, also called the breakdown rate function, measures the amount of risk associated with a system at time* t*. It may also be interpreted as the instantaneous failure rate.*

$$h(t) := \frac{f(t)}{S(t)}$$

Lemma 1.1. *The survivor function* $S(t)$ *may be expressed in terms of the hazard function* $h(t)$*:*

$$S(t) = e^{-\int_0^t h(s)ds}$$

Proof.

$$h(x) := \frac{f(x)}{S(x)} = \frac{-S'(x)}{S(x)}$$

Then $-h(x) = \frac{S'(x)}{S(x)}$. Integrating from 0 to t, $\ln(S(t)) = -\int_0^t h(x)dx$. □

Lemma 1.2 (Expected Lifetime)**.** *The expected lifetime Y is given by $E[Y] = \int_0^\infty S(t)dt$.*

Proof.

$$\begin{aligned} E[Y] &= \int_0^\infty xf(x)dx \\ &= -\int_0^\infty xS'(x)dx \end{aligned}$$

Integrating by parts,

$$E[Y] = -\left(tS(t) - \int_0^\infty S(t)dt \right)$$

Since it is assumed that $S(t) \xrightarrow{t\to\infty} 0$, $S(0) = 1$, and $tS(t) \xrightarrow{t\to\infty} 0$, the result is immediate. □

1.2 Original Model (Cha and Lee [9])

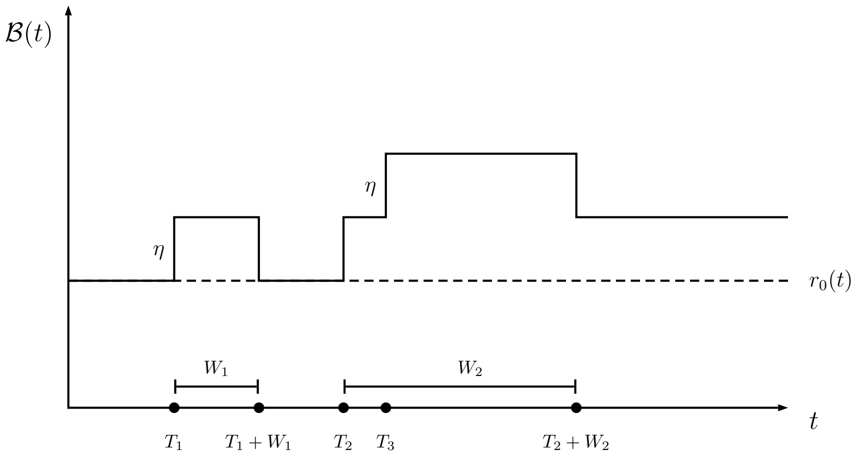


FIGURE 1.3: Sample Trajectory of Breakdown Rate Process Under Original Model

System Description and Survival Function

Cha and Lee considered a web server wherein each request arrives via a nonhomogenous Poisson process $\{N(t) : t \geq 0\}$ with intensity function $\lambda(t)$. Each request adds a constant stress η, increasing the breakdown rate for the duration of service. Suppose $r_0(t)$ is the breakdown rate of the idle server. Then the breakdown rate process $\mathcal{B}(t)$ is defined as

$$\mathcal{B}(t) := r_0(t) + \eta \sum_{j=1}^{N(t)} \mathbb{1}(T_j \leq t \leq T_j + W_j)$$

where $N(t)$, $\{T_j\}_{j=1}^{N(t)}$, $\{W_j\}_{j=1}^{N(t)}$ are the random variables that describe the number of arrivals, arrival times, and service times, respectively. It is assumed that $\{T_j\}_{j=1}^{N(t)}$ are independent of each other Furthermore, $\{W_j\}_{j=1}^{N(t)} \overset{\text{i.i.d.}}{\sim} g_W(w)$ and are mutually independent of all T_j's.

Under these conditions, Cha and Lee proved the following theorem:

Theorem. *Suppose that $\{N(t), t \geq 0\}$ is a nonhomogenous Poisson process with intensity function $\lambda(t)$, i.e. $m(t) \equiv \int_0^t \lambda(x)dx$. Assuming the conditional survival function is given by*

$$P\left(Y > t \,\middle|\, N(t), \{T_j\}_{j=1}^{N(t)}, \{W_j\}_{j=1}^{N(t)}\right) = \bar{F}_0(t)\exp\left(-\eta \sum_{j=1}^{N(t)} \min(W_j, t - T_j)\right)$$

and $m(t)$ has an inverse, the survival function of Y is given by

$$S_Y(t) = \bar{F}_0(t)\exp\left(-\eta \int_0^t \exp(-\eta w)\bar{G}_W(w)m(t-w)dw\right)$$

and the hazard function of Y, denoted $r(t)$, is given by

$$r(t) = r_0(t) + \eta \int_0^t e^{-\eta w}\bar{G}_W(w)\lambda(t-w)dw$$

Efficiency of the Server

It is natural to develop some measure of server performance. Cha and Lee measure such performance by defining the *efficiency*, ψ, of the web server as the long-run expected number of jobs completed per unit time. That is,

$$\psi := \lim_{t\to\infty} \frac{E[M(t)]}{t}$$

Upon breakdown and rebooting, the server is assumed to be 'as good as new', in that performance of the server does not degrade during subsequent reboots. In addition, the model assumes the arrival process after reboot, denoted $\{N^*(t), t \geq 0\}$, is a nonhomogenous Poisson process with the same intensity function $\lambda(t)$ as before, and that $\{N^*(t), t \geq 0\}$ is independent of the arrival process before reboot. In a practical setting, this model assumes no 'bottlenecking' of arrivals occurs in the queue during server downtime that would cause an initial flood to the rebooted server. In addition, the reboot time is assumed to follow a continuous distribution $H(t)$ with expected value ν. This process is a renewal reward process, with the renewal $\{R_n\} = \{M_n\}$, the number of jobs completed. The length of a renewal cycle is $Y_n + H_n$, where Y_n is the length of time the server was operational, and H_n is the time to reboot after a server crash. Then, by Theorem 1.1,

$$\psi = \frac{E[M]}{E[Y] + \nu} \tag{1.2}$$

where M is the number of jobs completed in a particular renewal cycle, ν is the mean time to reboot of the server, and Y is the length of a particular renewal cycle. Then, using (1.2), the following closed form of the efficiency of a server under all assumptions of Cha and Lee's model is derived.

Theorem. *Suppose* $\{N(t), t \geq 0\}$ *is a nonhomogenous Poisson process with intensity* $\lambda(t) \geq 0$. *Then the efficiency is given by*

$$\psi = \frac{1}{\int_0^\infty S_Y(t)dt + \nu} \left[\exp\left(-\int_0^t r_0(x)dx - \int_0^t \lambda(x)dx + a(t) + b(t) \right) \times (r_0(t)a(t) + \eta a(t)b(t)) \right]$$

where $a(t) = \int_0^t e^{-\eta v} g_W(v) m(t-v)dv$, $b(t) = \int_0^t e^{-\eta(t-r)} \bar{G}_W(t-r)\lambda(r)dr$, $\bar{G}_W(x) = 1 - \int_0^x g_W(s)ds$, *and* $m(x) = \int_0^x \lambda(s)ds$.

Numerical Example and Control Policies

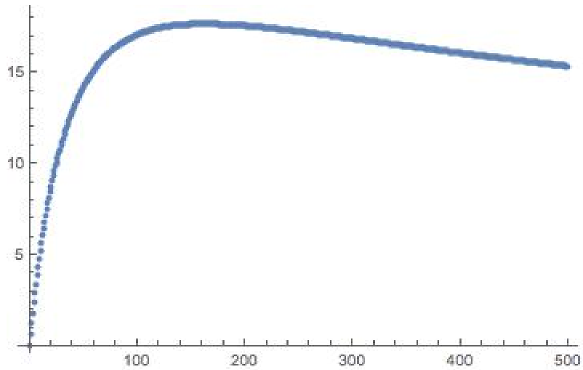


FIGURE 1.4: $\psi(\lambda)$ under Rayleigh Service Time Distribution

As an illustrative example, Cha and Lee considered the case when $\lambda(t) \equiv \lambda$, $r_0(t) \equiv r_0 = 0.2$, $\eta = 0.01$, $\nu = 1$, and $g_W(w) = we^{-w^2/2}$ (the PDF of the Rayleigh distribution). As shown in Figure 1.4, there exists a λ^* such that $\psi(\lambda)$ is maximized. Thus one may implement the obvious optimal control policy for server control to avoid server overload:

(1) If the real time arrival rate $\lambda < \lambda^*$, do not interfere with arrivals.

(2) If $\lambda \geq \lambda^*$, facilitate some appropriate measure of interference.

Examples of interference for a web server in particular include rejection of incoming requests or possible re-routing. Cha and Lee give an interference policy of rejection with probability $1 - \frac{\lambda^*}{\lambda}$. The next chapter presents a generalization of the above model.

Chapter 2

Server Under Random Stress Requests

The original model proposed by Cha and Lee (2011) assumed the workload brought by each random arrival T_j was a constant η. However, it is unrealistic to assume that each request brings the same workload to a server. For a web server, requests may vary from simple page views to many database queries during online commerce. A graphic designer may receive commissions for a simple web page to a dynamic and interactive site. A bridge has many different weights of foot or automobile traffic crossing. If we consider a fighter aircraft as a server, each mission may be viewed as a request having a different stress on the aircraft. An order placed for shipment can have varying stresses on warehouse inventory. In all these examples, these drastically different requirements of each arriving customer will strain the server resources in different and random ways.

In this chapter, the first extension of [9] is presented. The restrictive assumption of constant stress across all arrivals and all time is relaxed, and it is assumed that the individual stress is a random variable $\mathcal{H}_j$, independent of other job stresses and arrival times. The survival function of the server under a random stress environment is derived, and the closed form of the server efficiency is presented.

2.1 Survival Function of Server under Random Stress Environment

Model Assumptions and Random Breakdown Rate Process

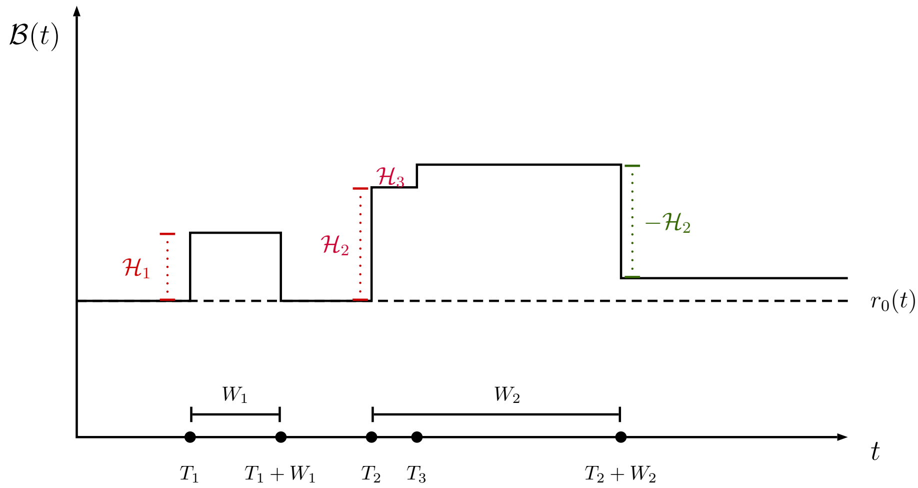


FIGURE 2.1: Sample Trajectory of Breakdown Rate Process under Random Stress Model

Assume that each job j coming into the server adds a random stress $\mathcal{H}_j$ to the server for the duration of its time in the system. Suppose $\{\mathcal{H}_j\}_{j=1}^{N(t)} \overset{i.i.d.}{\sim} \mathcal{H}$, where WLOG $\mathcal{H}$ is a discrete random variable with a finite sample space $S = \{\eta_i : \eta_i \in \mathbb{R}^+, i = 1, ..., m \text{ for } m \in \mathbb{N}\}$ and probability distribution given by

$$P(\mathcal{H} = \eta_i) = p_i, \quad i = 1, ..., m$$

The following assumptions from [9] are retained:

(CL1) Requests arrive via a nonhomogenous Poisson Process $\{N(t), t \geq 0\}$ with intensity $\lambda(t)$.

(CL2) Arrival times $\{T_j\}_{j=1}^{N(t)}$ are independent.

(CL3) Service times $\{W_j\}_{j=1}^{N(t)}$ are i.i.d. with pdf $g_W(w)$ and mutually independent of all arrival times.

Then, the random stress breakdown rate (RSBR) process $\mathcal{B}(t)$ is given by

$$\mathcal{B}(t) = r_0(t) + \sum_{j=1}^{N(t)} \mathcal{H}_j \mathbb{1}(T_j < t \leq T_j + W_j), \quad t \geq 0 \tag{2.1}$$

Compare the sample trajectory shown in Figure 2.1 to Figure 1.3. The random stress brought by each request still disappears upon job completion, but the effect on the server is no longer deterministic. Thus, for the same set of arrival times and respective completion times, the realization of the breakdown rate process under the RSBR model has one more element of variation.

Let Y be the random time to breakdown of the web server given the workload from client requests. Let $\mathfrak{T} = \{T_j\}_{j=1}^{N(t)}$, $\mathfrak{W} = \{W_j\}_{j=1}^{N(t)}$, and $\mathfrak{H} = \{\mathcal{H}_j\}_{j=1}^{N(t)}$, with observed values $\mathfrak{t} = \{t_j\}_{j=1}^{N(t)}$, $\mathfrak{w} = \{w_j\}_{j=1}^{N(t)}$, and $\mathfrak{h} = \{\eta_{i_j}\}_{i_j=1}^{N(t)}$. Then the conditional survival function of the server, given the arrival process of client requests ($N(t)$), job stresses ($\mathfrak{H}$), service times ($\mathfrak{W}$), and arrival times ($\mathfrak{T}$) is

$$\begin{aligned} S_{Y|N(t),\mathfrak{T},\mathfrak{W},\mathfrak{H}}(t|n,\mathfrak{t},\mathfrak{w},\mathfrak{h}) &= e^{-\int_0^t \mathcal{B}(x)dx} \\ &= \bar{F}_0(t)e^{-\sum_{j=1}^{N(t)} \mathcal{H}_i \min(W_j, t-T_j)} \end{aligned}$$

where $\overline{F}_0(t) = \exp\left(-\int_0^t r_0(x)dx\right)$.

Survival Function for the RSBR Web Server

Under the RSBR generalization, the survival function of the server is given in the following theorem.

Theorem 2.1 (Survival Function of RSBR Server)**.** *Suppose that jobs arrive to a server according to a nonhomogenous Poisson process $\{N(t), t \geq 0\}$ with intensity function $\lambda(t) \geq 0$ and $m(t) \equiv E[N(t)] = \int_0^t \lambda(x)dx$. Let the arrival times $\{T_j\}_{j=1}^{N(t)}$ be independent, and let the service times $\{W_j\}_{j=1}^{N(t)} \overset{i.i.d.}{\sim} g_W(w)$ be mutually independent of all arrival times. Assume the random job stresses $\mathcal{H}_j \overset{i.i.d.}{\sim} \mathcal{H}$. Then*

$$S_Y(t) = \bar{F}_0(t)\exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right]\right) \tag{2.2}$$

where $\bar{F}_0(t) = \exp\left(-\int_0^t r_0(s)ds\right)$.

Proof. Taking the expectation of the conditional survival function (2.2)

$$S_Y(t) = \bar{F}_0(t)E\left[\exp\left(-\sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j, t-T_j)\right)\right]$$

Using the law of total expectation:

$$E\left[e^{-\sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j,t-T_j)}\right] = E\left[E\left[e^{-\sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j,t-T_j)} \middle| N(t),\mathfrak{H}\right]\right]$$

Conditioned on $N(t) = n$ and $\mathcal{H}_j = \eta_{i_j}$ for some $i_j \in \{1,...,m\}$,

$$f_{T_1,\dots,T_n|N(t),\mathfrak{H}}(t_1 \dots t_n|n,\mathfrak{h}) = f_{T_1,\dots,T_n|N(t)}(t_1,\cdots,t_n|n)$$

since the sets $\mathfrak{H}$ and $\mathfrak{T}$ are mutually independent.

Let $T_1', ...T_n'$ be i.i.d. random variables with pdf $f(x) = \frac{\lambda(x)}{m(t)}$. By Lemma A.1,

$$f_{T_1,...,T_n|N(t)}(t_1, \cdots, t_n|n) = n! \prod_{j=1}^{n} \frac{\lambda(t_j)}{m(t)} \tag{2.3}$$

for $0 \leq t_1 \leq \cdots \leq t_n \leq t$. Then

$$E\left[e^{-\sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j, t-T_j)} \middle| N(t), \mathfrak{H}\right] = E\left[e^{-\sum_{j=1}^{n} \eta_{i_j} \min(W_j, t-T_j)}\right] \quad = E\left[e^{-\sum_{j=1}^{n} \eta_{i_{[j]}} \min(W_j, t-T_{[j]})}\right]$$

By Lemma A.1 (Appendix A),

$$\begin{aligned} E\left[e^{-\sum_{j=1}^{n} \eta_{i_{[j]}} \min(W_j, t-T_{[j]})}\right] &= E\left[e^{-\sum_{j=1}^{n} \eta_{i_{j'}} \min(W_j, t-T_{j'})}\right] \\ &= E\left[\prod_{j=1}^{n} e^{-\eta_{i_j} \min(W_j, t-T_{j'})}\right] \\ &= \prod_{j=1}^{n} E\left[e^{-\eta_{i_{j'}} \min(W_j, t-T_{j'})}\right] \end{aligned}$$

The equalities hold because the elements of $\mathfrak{W}$ and $\mathfrak{T}$ are i.i.d., respectively, and are mutually independent. Now, fix j'; then $\eta_{i_{j'}}$ is also fixed. By Lemma A.2 (Appendix A),

$$E\left[e^{-\eta_{i_{j'}} \min(W_j, t-T_{j'})}\right] = \frac{1}{m(t)} \left(m(t) - \eta_{i_{j'}} \int_0^t e^{-\eta_{i_{j'}} w} m(t-w) \bar{G}_W(w) dw\right) \tag{2.4}$$

Equation (2.4) is true $\forall\, j$, so

$$E\left[e^{-\sum_{j=1}^{N(t)} \mathcal{H}_j \min(W_j, t-T_j)} \middle| N(t), \mathfrak{H}\right] = \prod_{j=1}^{n} \frac{1}{m(t)} \left(m(t) - \eta_{i_j} \int_0^t e^{-\eta_{i_j} w} m(t-w) \bar{G}_W(w) dw\right) \tag{2.5}$$

Finally, the expectation of (2.5) over $N(t), \mathcal{H}_1, ..., \mathcal{H}_{N(t)}$ is taken. Denote

$$h_j(t) = m(t) - \eta_{i_j} \int_0^t e^{-\eta_{i_j} w} m(t-w) \bar{G}_W(w) dw$$

Denote $\vec{i} = (i_1, ..., i_n)$, $\vec{i}_{-j} = (i_1, ..., i_{j-1}, i_{j+1}, ..., i_n)$, $\eta_{\vec{i}} = (\eta_{i_1}, ..., \eta_{i_n})$, and

$$\sum_{i_n=1}^{m} \cdots \sum_{i_2=1}^{m} \sum_{i_1=1}^{m} (\cdot) P(\mathcal{H}_1 = \eta_{i_1}) P(\mathcal{H}_2 = \eta_{i_2}) \cdots P(\mathcal{H}_n = \eta_{i_n}) = \sum_{\vec{i}} (\cdot) P(\mathfrak{H} = \eta_{\vec{i}})$$

Then

$$
\begin{aligned}
E\left[\prod_{j=1}^{n} h_j(t)\right] &= \sum_{n=0}^{\infty}\sum_{\vec{i}}\left(\prod_{j=1}^{n}\frac{1}{m(t)}h_j\right)P(\mathcal{H}=\eta_{\vec{i}})P(N(t)=n)\\
&= \sum_{n=0}^{\infty}\frac{1}{m(t)^n}\left(\sum_{i_1=1}^{m} h_1 P(\mathcal{H}_1=\eta_{i_1})\right)\sum_{\vec{i}_{-1}}\prod_{j=2}^{n} h_j P(\mathfrak{H}_{-1}=\eta_{\vec{i}_{-1}})P(N(t)=n)\\
&= \sum_{n=0}^{\infty}\frac{1}{m(t)^n}\prod_{j=1}^{n}\left(\sum_{j=1}^{m} h_j P(\mathcal{H}_j=\eta_{i_j})\right)P(N(t)=n)\\
&= \sum_{n=0}^{\infty}\frac{1}{m(t)^n}\prod_{j=1}^{n}\left(\sum_{j=1}^{m} h_j P(\mathcal{H}_j=\eta_{i_j})\right)\frac{m(t)^n}{n!}e^{-m(t)}\\
&= \sum_{n=0}^{\infty}\frac{1}{m(t)^n}\left(\prod_{j=1}^{n} E_{\mathcal{H}_j}[h_j]\right)\frac{m(t)^n}{n!}e^{-m(t)}
\end{aligned}
$$

Let $h(t) = m(t) - \mathcal{H}\int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw$ Since $\mathcal{H}_j \overset{\text{i.i.d.}}{\sim} \mathcal{H}\ \forall\ j$, $E_{\mathcal{H}_j}[h_j] = E_{\mathcal{H}}[h(t)]$ for all j. Thus

$$
\begin{aligned}
&\sum_{n=0}^{\infty}\frac{1}{m(t)^n}\left(\prod_{j=1}^{n} E_{\mathcal{H}_j}[h_j]\right)\frac{m(t)^n}{n!}e^{-m(t)}\\
&\qquad= \sum_{n=0}^{\infty}\frac{1}{m(t)^n}\left(E_{\mathcal{H}}[h(t)]\right)^n\frac{m(t)^n}{n!}e^{-m(t)}\\
&\qquad= e^{-m(t)}\sum_{n=0}^{\infty}\frac{m(t)^n}{n!}\left(E_{\mathcal{H}}[h(t)]\right)^n\\
&\qquad= e^{-m(t)}\exp\left(m(t) - E_{\mathcal{H}}\left[\mathcal{H}\int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right]\right)\\
&\qquad= \exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right]\right)
\end{aligned}
$$

□

where the third equality uses the Taylor series representation of e^x. The compound failure rate function $r(t)$ is given by

$$
r(t) = -\frac{d}{dt}\ln(S_Y(t)) = r_0(t) + E_{\mathcal{H}}\left[\mathcal{H}\int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right]
$$

as defined in 1.1.)

2.2 Efficiency measure of the server under RSBR

Upon server crash, the server must be rebooted. This section gives the server efficiency as defined in [9]. The following assumptions from the original model are retained:

(E1) The arrival process after rebooting, $\{N^{rb}(t), t \geq 0\}$, remains a nonhomogenous Poisson process with the same intensity function $\lambda(t), t \geq 0$ as before.

(E2) $\{N^{rb}(t), t \geq 0\}$ is independent of the arrival process of client requests before rebooting. Hence, $\{N^{rb}(t), t \geq 0\} = \{N(t), t \geq 0\}$, since it retains all the same characteristics as before.

(E3) The time to reboot the server follows a continuous distribution $H(t)$ with mean ν.

Recall that $M(t)$ is defined as the total number of jobs completed by the server during the time $(0, t]$. Also, recall the definition of server efficiency from [9]:

$$\psi \equiv \lim_{t \to \infty} \frac{E[M(t)]}{t}$$

The efficiency of the server under a random stress environment is given in the following theorem.

Theorem 2.2 (Server Efficiency under Random Stress Environment)**.** *Suppose that $\{N(t) : t \geq 0\}$ is a nonhomogenous Poisson process with intensity function $\lambda(t), t \geq 0$. Suppose also the conditions of Theorem 2.1 and the conditions (E1)-(E3) are met. Then the efficiency of the server is given by*

$$\psi = \frac{1}{\int_0^\infty S_Y(t)dt + \nu} \left\{ \int_0^\infty e^{-\int_0^t r_0(x) - \int_0^t \lambda(x)dx + E_{\mathcal{H}}[a(t)+b(t)]} (r_0(t) E_{\mathcal{H}}[a(t)] + E_{\mathcal{H}}[\mathcal{H}a(t)b(t)])dt \right\} \tag{2.6}$$

where $a(t) = \int_0^t e^{-\mathcal{H}v} g_W(v) m(t-v) dv$ and $b(t) = \int_0^t e^{-\mathcal{H}(t-r)} \bar{G}_W(t-r)\lambda(r)dr$.

Proof. From Theorem 1.1 and Section 1.2, $\psi = \frac{E[M]}{E[Y]+\nu}$, where Y is the length of time the server is operational during a particular renewal cycle and ν is the mean time to reboot. By Lemma 1.2, $E[Y] = \int_0^\infty S_Y(t)dt$, where $S_Y(t)$ is the unconditional survival function from Theorem 2.1. Therefore, the completion of the proof relies on deriving $E[M]$.
$M = \sum_{j=1}^{N(Y)} \mathbb{1}(T_j + W_j \leq Y)$ which may be rewritten as

$$M = \sum_{j=1}^{N(Y)} \mathbb{1}(R_j + V_j \leq Y)$$

where $\{(R_j, V_j)\}_{j=1}^{N(Y)}$ may be regarded as a random permutation of $\{(T_j, W_j)\}_{j=1}^{N(Y)}$ due to the mutual independence of $\{T_j\}, \{W_j\}$ and the respective i.i.d nature of both. Therefore,

$$E[M] = E\left[\sum_{j=1}^{N(Y)} \mathbb{1}(R_j + V_j \leq Y)\right]$$

For convenience and clarity, the following notation is introduced:

$$\mathfrak{R} = \{R_1, ..., R_n\}, \qquad \mathfrak{V} = \{V_1, ..., V_n\},$$
$$\mathfrak{H} = \{\mathcal{H}_1, ..., \mathcal{H}_n\}$$

with observed values

$$\mathfrak{r} = \{r_1, ..., r_n\}, \qquad \mathfrak{v} = \{v_1, ..., v_n\},$$
$$\mathfrak{h} = \{\eta_{i_1}, ..., \eta_{i_n}\}$$

By Bayes's Theorem,

$$f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},Y,N}(\mathfrak{r},\mathfrak{v},\mathfrak{h},t,n) = f_{Y|\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(t|\mathfrak{r},\mathfrak{v},\mathfrak{h},n) f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(\mathfrak{r},\mathfrak{v},\mathfrak{h},n)$$

By Lemma A.3(Appendix A), the conditional distribution $f_{Y|\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(t|\mathfrak{r},\mathfrak{v},\mathfrak{h},n)$ is given by

$$f_{Y|\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(t|\mathfrak{r},\mathfrak{v},\mathfrak{h},n) = e^{-\int_0^t r_0(s)ds - \sum_{j=1}^n \eta_{i_j} \min(v_j, t-r_j)} \left(r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}(v_j > t - r_j) \right) \tag{2.7}$$

Since all $\mathcal{H}_j \in \mathfrak{H}$ are i.i.d. and mutually independent of $\mathfrak{R}, \mathfrak{V}$, and N,

$$f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(\mathfrak{r},\mathfrak{v},\mathfrak{h},n) = f_{\mathfrak{R},\mathfrak{V},N}(\mathfrak{r},\mathfrak{v},n) f_{\mathfrak{H}}(\mathfrak{h}) = f_{\mathfrak{R},\mathfrak{V},N}(\mathfrak{r},\mathfrak{v},n) \prod_{j=1}^n P(\mathcal{H}_j = \eta_{i_j})$$

By Lemma A.4(Appendix A)

$$f_{\mathfrak{R},\mathfrak{V},N}(\mathfrak{r},\mathfrak{v},n) = \frac{1}{n!} \prod_{j=1}^n e^{\int_0^t \lambda(x)dx} \lambda(r_j) g_W(v_j)$$

$$f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(\mathfrak{r},\mathfrak{v},\mathfrak{h},n) = \frac{1}{n!} \prod_{j=1}^n e^{\int_0^t \lambda(x)dx} \lambda(r_j) g_W(v_j) \prod_{j=1}^n P(\mathcal{H}_j = \eta_{i_j}) \tag{2.8}$$

Finally, by multiplying (2.7) and (2.8)

$$f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},Y,N(t)}(\mathfrak{r},\mathfrak{v},\mathfrak{h},t,n) = \frac{e^{-\int_0^t r_0(x)dx - \int_0^t \lambda(x)dx}}{n!} \left[\prod_{j=1}^n e^{-\eta_{i_j} \min(v_j, t-r_j)} \lambda(r_j) g_W(v_j) P(\mathcal{H}_j = \eta_{i_j}) \right] \times \left[r_0(t) + \sum_{j=1}^n \eta_{i_j} \mathbb{1}(v_j > t - r_j) \right]$$

Denote $\mathfrak{v}_{-j} = (v_1, ..., v_{j-1}, v_{j+1}, ..., v_n)$, $\mathfrak{r}_{-j} = (r_1, ..., r_{j-1}, r_{j+1}, ..., r_n)$.
Let $\int_{\vec{a}}^{\vec{b}} f(\mathbf{x})d\mathbf{x} = \int_a^b \cdots \int_a^b \int_a^b f(x_1, ..., x_n)dx_1 \cdots dx_n$ for $a, b \in \mathbb{R}$.

$$\begin{aligned}
E[M] &= E\left[\sum_{j=1}^{N(Y)} \mathbb{1}(R_j + V_j \leq Y)\right] \\
&= \sum_{n=1}^{\infty} \int_0^t \left[\sum_{j=1}^{n} \sum_{i_j=1}^{m} \int_0^t \int_0^{t-r_j} \int_{\vec{0}}^{\vec{t}} \int_{\vec{0}}^{\vec{\infty}} f_{\mathfrak{R},\mathfrak{V},\mathfrak{H},Y,N(t)}(\mathfrak{r}, \mathfrak{v}, \mathfrak{h}, t, n) d\mathfrak{v}_{-j} d\mathfrak{r}_{-j} dv_j dr_j\right] dt \\
&= \sum_{n=1}^{\infty} \int_0^{\infty} \frac{1}{n!} r_0(t) e^{-\int_0^t r_0(x)dx - \int_0^t \lambda(x)dx} n E_{\mathcal{H}}\left[\int_0^t \int_0^{t-r} e^{-\mathcal{H}v} g_W(v) dv \lambda(r) dr\right] \\
&\quad \times \left\{E_{\mathcal{H}}\left[\int_0^t \int_0^{t-r} e^{-\mathcal{H}v} g_W(v) dv \lambda(r) dr + \int_0^t e^{-\mathcal{H}(t-r)} \bar{G}_W(t-r) \lambda(r) dr\right]\right\}^{n-1} dt \\
&+ \sum_{n=1}^{\infty} \int_0^{\infty} \frac{1}{n!} e^{-\int_0^t r_0(x)dx - \int_0^t \lambda(x)dx} \\
&\quad \times n(n-1) E_{\mathcal{H}}\left[\mathcal{H} \int_0^t \int_0^{t-r} e^{-\mathcal{H}v} g_W(v) dv \lambda(r) dr \int_0^t e^{-\mathcal{H}(t-r)} \bar{G}_W(t-r) \lambda(r) dr\right] \\
&\quad \times \left\{E_{\mathcal{H}}\left[\int_0^t \int_0^{t-r} e^{-\mathcal{H}v} g_W(v) dv \lambda(r) dr + \int_0^t e^{-\mathcal{H}(t-r)} \bar{G}_W(t-r) \lambda(r) dr\right]\right\}^{n-2} dt
\end{aligned}$$

Let $a(t) = \int_0^t \int_0^{t-r} e^{-\mathcal{H}v} g_W(v) dv \lambda(r) dr$. Through a change of variables, $a(t) = \int_0^t e^{-\mathcal{H}v} g_W(v) m(t-v) dv$. Let $b(t) = \int_0^t e^{-\mathcal{H}(t-r)} \bar{G}_W(t-r) \lambda(r) dr$. Then

$$\begin{aligned}
E[M] &= \sum_{n=1}^{\infty} \int_0^{\infty} \frac{1}{(n-1)!} r_0(t) E_{\mathcal{H}}[a(t)] (E_{\mathcal{H}}[a(t) + b(t)])^{n-1} e^{-\int_0^t r_0(x)dx - \int_0^t \lambda(x)dx} dt \\
&\quad + \sum_{n=2}^{\infty} \int_0^{\infty} \frac{1}{(n-2)!} E_{\mathcal{H}}[\mathcal{H} a(t) b(t)] (E_{\mathcal{H}}[a(t) + b(t)])^{n-2} dt \\
&= \int_0^{\infty} r_0(t) e^{-\int_0^t r_0(x) - \int_0^t \lambda(x)dx} E_{\mathcal{H}}[a(t)] \left(\sum_{n=1}^{\infty} \frac{1}{(n-1)!} (E_{\mathcal{H}}[a(t) + b(t)])^{n-1}\right) dt \\
&\quad + \int_0^{\infty} e^{-\int_0^t r_0(x) - \int_0^t \lambda(x)dx} E_{\mathcal{H}}[\mathcal{H} a(t) b(t)] \left(\sum_{n=2}^{\infty} \frac{1}{(n-2)!} (E_{\mathcal{H}}[a(t) + b(t)])^{n-2}\right) dt \\
&= \int_0^{\infty} r_0(t) e^{-\int_0^t r_0(x) - \int_0^t \lambda(x)dx} E_{\mathcal{H}}[a(t)] e^{E_{\mathcal{H}}[a(t)+b(t)]} dt \\
&\quad + \int_0^{\infty} e^{-\int_0^t r_0(x) - \int_0^t \lambda(x)dx} E_{\mathcal{H}}[\mathcal{H} a(t) b(t)] e^{E_{\mathcal{H}}[a(t)+b(t)]} dt \\
&= \int_0^{\infty} e^{-\int_0^t r_0(x) - \int_0^t \lambda(x)dx + E_{\mathcal{H}}[a(t)+b(t)]} \left[r_0(t) E_{\mathcal{H}}[a(t)] + E_{\mathcal{H}}[\mathcal{H} a(t) b(t)]\right] dt \qquad (2.9)
\end{aligned}$$

□

2.3 Remarks and Implications

Note that $\mathcal{H}$ was assumed discrete, but the proofs of Theorema 2.1 and 2.2 are unaffected if $\mathcal{H}$ is continuous. Thus the generality of this model is significantly stronger than in [9].

For certain distributions of $\mathcal{H}$, $S_Y(t)$ has a fairly compact form. Section 4.2 examines the case where $\mathcal{H}$ has a binomial distribution, formed from both independent and dependent Bernoulli trials. Chapter 3 explores the properties of ψ further under various service distributions.

Chapter 3

On Server Efficiency

3.1 Motivation

In [9], the authors provide a numerical illustration of the efficiency under the Rayleigh service life distribution. The Rayleigh distribution has applications in physics, typically when the magnitude of a vector is related to its directional components [22]. The Rayleigh distribution is a special case of the Weibull distribution which is widely used in survival analysis, failure analysis, weather forecasting, and communications [1, 25].

These distributions are not typically used to model service times. The exponential distribution is the most common due to its memoryless properties, followed by the Erlang and uniform distributions [2].

The efficiency, ψ, under the Rayleigh distribution example given in [9] and reproduced in Figure 1.4, assuming a constant intensity $\lambda(t) \equiv \lambda$, shows the existence of a $0 < \lambda^* < \infty$ such that $\psi(\lambda)$ is maximized at λ^*.

This useful feature of $\psi(\lambda)$ in this case allows for the implementation of a simple control policy for arrivals to the server to prevent overload: (1) if $\lambda(t) \leq \lambda^*$, do nothing, and (2) if $\lambda(t) > \lambda^*$, intercept arrivals in some fashion. Cha and Lee ([9]) choose to interfere by rejecting each arrival thereafter with probability $1 - \frac{\lambda^*}{\lambda}$.

This binary nature of the control policy would require real-time data on arrival rates for implementation. Numerical simulations under a variety of possible distribution classes, including convex, concave, exponential, uniform, and Erlang suggest that the mathematical properties of ψ are heavily influenced by the choice and characteristics of service time distribution $g_W(w)$.

In particular, it is of interest to seek sufficient conditions of $g_W(w)$ that will guarantee the existence of a λ^* that maximizes ψ. This is done for the uniform, compact support, and Erlang classes. Furthermore, it is shown under certain conditions, not only does the server efficiency lack a maximum, but ψ increases without bound. This is not representative of real server behavior.

3.2 Efficiency under Uniform Service Life Distribution

Suppose $\lambda(x) \equiv \lambda$, and suppose $r_0(x) \equiv r_0 = \max_{x\in(0,\infty)} r_0(x)$. The efficiency ψ is given in Section 1.2 for constant η and λ and reproduced here:

$$\psi(\lambda) = \frac{1}{\int_0^\infty S_Y(t)dt + \nu}\left[\int_0^\infty \exp\left(-r_0 t - \lambda t + a(t) + b(t)\right)(r_0 + b(t))a(t)dt\right] \tag{3.1}$$

where $S_Y(t)$ is the survival function of the node, $a(t) = \int_0^t e^{-\eta v} g(v)(t-v)dv$,
$b(t) = \int_0^t e^{-\eta(t-r)}\bar{G}(t-r)dr$, $g(v)$ is the pdf of the service time distribution, and
$\bar{G}(x) = 1 - \int_0^x g(s)ds$.

The following theorem gives sufficient conditions for the uniform distribution and η that guarantee the existence of a finite maximum efficiency.

Theorem 3.1. *Suppose the service life distribution is given by Uniform(c,d) for some $0 < c < d$. Then if $\sigma > \frac{ce^{-c\eta}}{\sqrt{12}\phi(-\eta)(1+\eta(c+d))+c\eta-1}$ where $\phi(-\eta)$ the standard deviation of the service life W, $\psi(\lambda)$ has a maximum on $(0,\infty)$ is the moment generating function of a uniform distribution evaluated at $-\eta$.*

Sketch of proof: The proof will proceed in the following steps.

(i) Note that $\psi(0) = 0$, and $\psi(\lambda) \geq 0 \ \forall \ \lambda \in [0,\infty)$

(ii) Construct $h(\lambda) \geq \psi(\lambda)$ such that $h(\lambda) \xrightarrow{\lambda\to\infty} 0$

(iii) Then $\psi(\lambda) \xrightarrow{\lambda\to\infty} 0$. Since $\psi(\lambda)$ is continuous, and clearly $\psi(1) > \psi(0)$, the existence of the maximum is established.

Proof. (i) is clear, since all components are nonnegative for all λ. To accomplish (ii), note that $\int_0^\infty S_Y(t)dt + \nu \geq 0$. Thus the construction of $h(\lambda)$ will focus on dominating the numerator of $\psi(\lambda)$ by upper estimates of $a(t)$ and $b(t)$ such that the numerator of $h(\lambda)$, denoted henceforth as $h_N(\lambda) \xrightarrow{\lambda\to\infty} 0$.

The uniform distribution is given by $g(w) = \frac{1}{d-c}\mathbb{1}(w)$ for $c < d$, and

$$\bar{G}_W(w) = \begin{cases} 1, & w < c \\ \frac{d-w}{d-c}, & c \leq w < d \\ 0, & w > d \end{cases}$$

Recall $a(t) := \int_0^t e^{-\eta v} g(v)(t-v)dv$. Plugging in the uniform service life distribution,

$$\begin{aligned} a(t) &= \frac{1}{d-c}\int_0^t e^{-\eta v}\mathbb{1}(v)(t-v)dv \\ &= \begin{cases} 0, & t<c \\ \frac{1}{(d-c)\eta^2}\left[e^{-t\eta}+e^{-c\eta}(t\eta-c\eta-1)\right], & c\le t<d \\ \frac{1}{(d-c)\eta^2}\left(e^{-d\eta}(1+d\eta-t\eta)+e^{-c\eta}(t\eta-c\eta-1\right), & t\ge d \end{cases} \end{aligned}$$

and

$$\frac{da}{dt} = \begin{cases} 0 & t<c \\ \frac{1}{(d-c)\eta}\left(-e^{-\eta t}+e^{-c\eta}\right) & c\le t<d \\ \frac{1}{(d-c)\eta}\left(e^{-c\eta}-e^{-d\eta}\right) & t\ge d \end{cases}$$

Then

$$\frac{da}{dt} \le \begin{cases} 0 & t<c \\ \frac{\lambda e^{-c\eta}}{\eta(d-c)} & t\ge c \end{cases}$$

Thus $a(t)$ may be dominated on $(0,\infty)$ by the piecewise linear

$$\tilde{a}(t) = \begin{cases} 0 & t<c \\ \frac{\lambda e^{-c\eta}}{\eta(d-c)}t & t\ge c \end{cases}$$

Similarly, $b(t) := \int_0^t e^{-\eta(t-r)}\bar{G}(t-r)dr$. For the uniform distribution,

$$\bar{G}(t-r) = \begin{cases} 0, & r<t-d \\ \frac{d-(t-r)}{d-c}, & t-d\le r\le t-c \\ 1, & r>t-c \end{cases}$$

Then plugging into $b(t)$,

$$\begin{aligned} b(t) &= \left[0+\int_{t-d}^{t-c} e^{-\eta(t-r)}\frac{d-(t-r)}{d-c}dr+\int_{t-c}^{t} e^{-\eta(t-r)}dr\right] \\ &= \left[\frac{1}{\eta}+\frac{(e^{-d\eta}-e^{-c\eta})(1+c\eta+d\eta)}{(d-c)\eta^2}\right] \\ &:= B \end{aligned}$$

The numerator of $\psi(\lambda)$ is then dominated by the numerator of $h(\lambda)$, denoted by $h_N(\lambda)$ and given by

$$\begin{aligned} N(\lambda) &= \int_0^\infty e^{-r_0 t - \lambda t + \lambda \tilde{a}(t) + \lambda B} \left(r_0 + \lambda B\right) \lambda \tilde{a}(t) dt \\ &= \int_c^\infty e^{-r_0 t - \lambda t + \lambda \frac{e^{-c\eta}}{\eta(d-c)} t + \lambda B} \left(r_0 + \lambda B\right) \lambda \left(\frac{e^{-c\eta}}{\eta(d-c)}\right) t dt \end{aligned}$$

To guarantee $h_N(\lambda) \to 0$ in λ, it is sufficient for $B - c - \frac{e^{-c\eta}}{\eta(d-c)} c < 0$. The moment generating function for the uniform distribution is given by $\phi(x) = \frac{e^{xd} - e^{xc}}{x(d-c)}$, and the standard deviation of a uniform distribution is given by $\sigma = \frac{d-c}{\sqrt{12}}$

$$\begin{aligned} B - c - \frac{e^{-c\eta}}{\eta(d-c)} c &= \frac{1}{\eta} + \frac{(e^{-d\eta} - e^{-c\eta})(1 + \eta(c+d))}{\eta^2 (d-c)} - c + \frac{ce^{-c\eta}}{\eta(d-c)} \\ &= \frac{1}{\eta} - \phi(-\eta) \cdot \frac{1 + \eta(c+d)}{\eta} - c + \frac{ce^{-c\eta}}{\eta\sqrt{12}\sigma} \end{aligned}$$

Then

$$\begin{aligned} -\phi(-\eta)(1 + \eta(c+d)) + \frac{ce^{-c\eta}}{\sqrt{12}\sigma} &< \eta c - 1 \\ \frac{ce^{-c\eta}}{\sqrt{12}\sigma} &< \phi(-\eta)(1 + \eta(c+d)) + c\eta - 1 \end{aligned}$$

Solving for σ gives

$$\sigma > \frac{ce^{-c\eta}}{\sqrt{12}\phi(-\eta)(1 + \eta(c+d)) + c\eta - 1}$$

□

Numerical simulations suggest that ψ increases without bound for $c = 0, d > 1$. The following lemma proves this fact.

Lemma 3.1. *Suppose the service life distribution is given by Uniform(0,d), with $d > 1$. Then ψ increases without bound.*

Proof. The proof is similar to the proof of Theorem 3.1, but in this case, ψ will be bounded from below by a constructed function $h(\lambda)$ such that $h(\lambda) \to \infty$ in λ.

$$a(t) = \begin{cases} \frac{1}{d\eta^2}(t\eta + e^{-t\eta} - 1), & t \le d \\ \frac{1}{d\eta^2}(e^{-d\eta}(1 + d\eta - t\eta) + t\eta - 1), & t > d \end{cases}$$

Then $a(t) \ge \frac{t}{d\eta} - \frac{1}{d\eta}$, since $a(t)$ is nonnegative, and $\frac{da}{dt} \le \frac{1}{d\eta}$.

$$b(t) = \int_0^t e^{-\eta(t-r)} \bar{G}(t-r) dr = \int_{t-d}^t e^{-\eta(t-r)} \left(\frac{d - (t-r)}{d}\right) = \frac{1}{\eta} + \frac{e^{-d\eta} - 1}{d\eta^2}$$

Then the numerator of ψ is bounded below by $h_N(\lambda)$, where

$$h_N(\lambda) = \int_0^\infty e^{-r_0 t - \lambda t + \lambda\left(\frac{t}{d\eta} - \frac{1}{d\eta}\right)\lambda B}(r_0 + \lambda B)\lambda\left(\frac{t}{d\eta} - \frac{1}{d\eta}\right) dt$$

It suffices to seek conditions such that $B - \frac{1}{d\eta} > 0$, or equivalently, that

$$d\eta - \eta + e^{-d\eta} - 1 > 0$$

Thus $d > 1$ is sufficient to guarantee that $N(\lambda) \to \infty$. Thus for $c = 0, d > 1$, ψ will increase without bound. □

3.3 Extension of the Uniform Distribution: Compact Support

The ideas and techniques presented in Section 3.2 yield a powerful extension to any service life distribution with compact support away from 0. Supposing $g_W(w)$ has compact support $[a, b]$, it may be bounded above by a positively scaled uniform distribution. In practice, service times are finite and nonzero, thus this extension allows for very simple control policies to be implemented for a much larger class of distributions.

Theorem 3.2. *Let $g_W(w)$ be the pdf of the service times having compact support $[c, d], d > 0$. Let $m = \max_w g_W(w) < \infty$, and $R = (d - c)$ be the length of the support. Then $\psi(\lambda)$ has a maximum if $m < \frac{c}{R\eta + e^{-b\eta} - e^{-a\eta} + \eta e^{-a\eta}}$.*

Proof. Let $M = m(d - c)$. Then $g_W(w) \le m\mathbb{1}_{[c,d]}(w)$, and

$$\bar{G}_W(w) \le \begin{cases} M, & w \le c \\ M - m(w - c), & c \le w \le d \\ 0, & w \ge d \end{cases}$$

We construct a function $h_N(\lambda)$ that will bound the numerator of $\psi(\lambda)$ such that $h_N(\lambda) \to 0$ as $\lambda \to \infty$, just as in the proof of Theorem 3.1.

$$a(t) \le \begin{cases} \frac{m}{\eta^2}(e^{-t\eta} + e^{-c\eta}(t\eta - (c\eta + 1))), & c \le t \le d \\ \frac{m}{\eta^2}\left(t\eta\left(e^{-c\eta} - e^{-d\eta}\right) + e^{-d\eta}(1 + d\eta) - e^{-c\eta}(1 + c\eta)\right), & t \ge d \end{cases}$$

The above may be bounded one more time by $\tilde{a}(t) = \frac{me^{-c}t}{\eta}, t \geq c$. Now from the bounds on $\bar{G}_W(w)$,

$$\begin{aligned} b(t) &= \int_{t-d}^{t-c} e^{-\eta(t-r)}(M - m((t-r)-c))dr + \int_{t-c}^{t} Me^{-\eta(t-r)}dr \\ &= \frac{m}{\eta^2}\left(R\eta + e^{-d\eta} - e^{-c\eta}right)\right) \\ &:= B \end{aligned}$$

Then

$$N(\lambda) = \int_c^\infty e^{-r_0 t - \lambda t + \lambda \tilde{a}(t) + \lambda B}(r_0 + \eta\lambda B)\lambda\tilde{a}(t)dt$$

In order for the above to be integrable, $\frac{me^{-c}c}{\eta} < 1$, or $m < \eta e^{c\eta}$. Supposing this is true, $h_N(\lambda) \to 0$ if $B - c + \frac{me^{-c}c}{\eta} < 0$. Simplifying this in terms of m gives the desired result. □

Comments and Illustrations

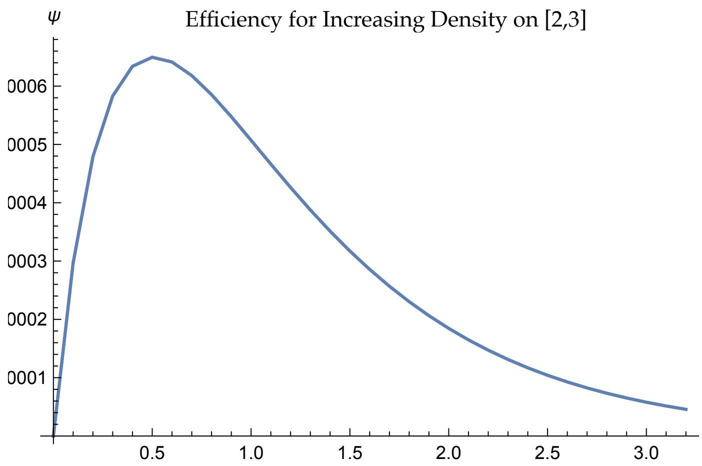


FIGURE 3.1: $\psi(\lambda)$ under $g_W(w) = \frac{2}{5}w\mathbb{1}_{[2,3]}(w)$

Example 3.1. *Let* $g_W(w) = \frac{2}{5}w\mathbb{1}_{[2,3]}(w)$, *and let* $r_0 = \nu = \eta = 1$. *Then* $m = \frac{6}{5}$. *By Theorem 3.2,* $m < \frac{2}{e^{-2}+e^{-3}+1} \approx 1.678$. *Thus, the existence of a maximum* ψ *is guaranteed. Figure 3.1 gives the numerical result with step size of 0.1. The maximum occurs around* $\lambda^* = 0.5$.

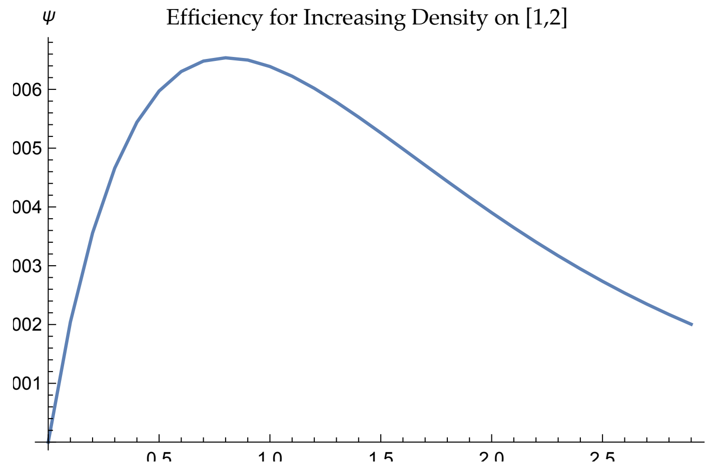


FIGURE 3.2: $\psi(\lambda)$ under $g_W(w) = \frac{2}{3}w\mathbb{1}_{[1,2]}(w)$

The condition is rather weak and is only a sufficient condition. As an illustration, consider the same increasing density shifted to a support of $[1, 2]$. Thus
$g_W(w) = \frac{2}{3}w\mathbb{1}_{[1,2]}(w)$. Retain all other assumptions from Example 3.1. Then

$$m = \frac{4}{3} > \frac{1}{e^{-2}+1} \approx 0.88$$

which violates the condition. However, consider Figure 3.2 above. Clearly ψ has a maximum at approximately $\lambda^* = 0.8$.

The condition given in Theorem 3.2 relies on overestimating $g_W(w)$ by a constant. If the variance σ_W^2 of $g_W(w)$ is large, the maximum of the pdf will decrease and be more comparable to the rest of the distribution. In these cases, bounding $g_W(w)$ by its max m over the support $[a, b]$ will give a reasonable approximation. However, in the case of a small support and high enough skew compared to the size and location of the support, much of the mass is concentrated at the right end of the support, and m will be higher. In these cases, bounding $g_W(w)$ by m will result in a large amount of overestimation, and thus the condition may fail, but ψ still has a maximum. The numerical example wherein $g_W(w) = \frac{2}{3}w\mathbb{1}_{[1,2]}(w)$ illustrates the conservative nature of this type of estimation.

3.4 Efficiency under Erlang Service Life Distribution

Now suppose $g(v)$ is of the Erlang class but shifted $\delta > 0$ to the right. For motivation, consider that service times can never be 0 in a practical setting. Then the PDF and the complement of the

CDF are given by

$$g(v;k,\gamma,\delta) = \begin{cases} 0, & 0 \le v \le \delta \\ \frac{\gamma^k (v-\delta)^{k-1} e^{-\gamma(v-\delta)}}{(k-1)!}, & v \ge \delta; \end{cases}$$
$$\bar{G}(v;k,\gamma,\delta) = \begin{cases} 1, & 0 \le v \le \delta \\ e^{\gamma(\delta-v)} \sum_{j=0}^{k-1} \frac{\gamma^j (v-\delta)^j}{j!}, & v \ge \delta \end{cases} \tag{3.2}$$

Theorem 3.3. *Let $\delta > 0, \eta > 0$. Let $\alpha(\delta) = \left(\frac{\gamma}{\gamma+\eta}\right)^k e^{-\eta\delta} + \frac{\gamma^k e^{\gamma\delta}(k-1)}{(\eta+\gamma)^{k-1}}$, and $0 < \beta(\delta,\eta) < 1$. If the service life distribution is of the $\delta-$shifted Erlang class, then $\psi(\lambda)$ has a maximum in λ on $(0,\infty)$ for δ, η such that $\alpha(\delta) + \beta(\delta,\eta) < 1$.*

Proof. The proof will proceed using the same steps (i)-(iii) from the proof of Theorem 3.1. By Lemma A.5,

$$a(t) = \left(\frac{\gamma}{\gamma+\eta}\right)^k e^{-\eta\delta} t + \frac{k\gamma^k \left(e^{-(\eta+\gamma)t+\gamma\delta} - e^{-\eta\delta}\right)}{(\gamma+\eta)^{k+1}} + \frac{\gamma^k e^{-(\eta+\gamma)t+\gamma\delta}}{(k-1)!} \sum_{j=1}^{k-1} \left[\frac{(t-\delta)^j}{(\gamma+\eta)^{k-j+1}} \left(\frac{k!}{j!} - \frac{(k-1)!}{(j-1)!}\right)\right] \tag{3.3}$$

Let $f_a(t) = \sum_{j=1}^{k-1} \left[\frac{(t-\delta)^j}{(\gamma+\eta)^{k-j+1}} \left(\frac{k!}{j!} - \frac{(k-1)!}{(j-1)!}\right)\right]$. Then

$$a(t) = \left(\frac{\gamma}{\gamma+\eta}\right)^k e^{-\eta\delta} t + \frac{k\gamma^k \left(e^{-(\eta+\gamma)t+\gamma\delta} - e^{-\eta\delta}\right)}{(\gamma+\eta)^{k+1}} + + \frac{\gamma^k e^{-(\eta+\gamma)t+\gamma\delta}}{(k-1)!} f_a(t)$$

and

$$\begin{aligned} \frac{da}{dt} &= \left(\frac{\gamma}{\gamma+\eta}\right)^k e^{-\eta\delta} - \frac{e^{-(\eta+\gamma)t+\gamma\delta} k\gamma^k}{(\gamma+\eta)^k} - \frac{\gamma^k(\gamma+\eta)e^{-(\eta+\gamma)t+\gamma\delta}}{(k-1)!} f_a(t) + \frac{\gamma^k e^{-(\eta+\gamma)t+\gamma\delta}}{(k-1)!} \dot{f}_a \\ &\le \left(\frac{\gamma}{\gamma+\eta}\right)^k e^{-\eta\delta} + \frac{\gamma^k e^{\gamma\delta}(k-1)}{(\eta+\gamma)^{k-1}} \end{aligned}$$

Thus $a(t) \le \alpha(\delta) t$, where $\alpha(\delta) = \left(\frac{\gamma}{\gamma+\eta}\right)^k e^{-\eta\delta} + \frac{\gamma^k e^{\gamma\delta}(k-1)}{(\eta+\gamma)^{k-1}} < 1$

By Lemma A.6,

$$b(t) = \frac{1-e^{-\delta\eta}}{\eta} + \sum_{j=0}^{k-1} \left[\frac{\gamma^j e^{-\eta\delta}}{(\gamma+\eta)^{j+1}} - \gamma^j e^{-(\eta+\gamma)t+\eta\delta} \left(\sum_{i=0}^{j} \frac{(\eta+\gamma)^i (t-\delta)^i}{i!}\right)\right] \tag{3.4}$$

Now, $b(t) \xrightarrow{t\to\infty} B(\delta,\eta,\gamma)$ where $B(\delta,\eta,\gamma) = \frac{1-e^{-\delta\eta}}{\eta} + \sum_{j=0}^{k-1} \frac{\gamma^j e^{-\gamma\delta}}{(\eta+\gamma)^{j+1}}$.

Let $f_b(t) := \sum_{j=0}^{k-1} \gamma^j \sum_{i=0}^{j} \frac{(\eta+\gamma)^i (t-\delta)^i}{i!}$. Then $b(t)$ may be rewritten as

$$b(t) = B(\delta,\gamma,\eta) - e^{-(\eta+\gamma)t+\eta\delta} f_b(t)$$

By Lemma A.7, $b(t)$ is concave. Through straightforward calculus, $b'(\delta) = 1$ and $b(t)$ is only defined for $t \geq \delta$.

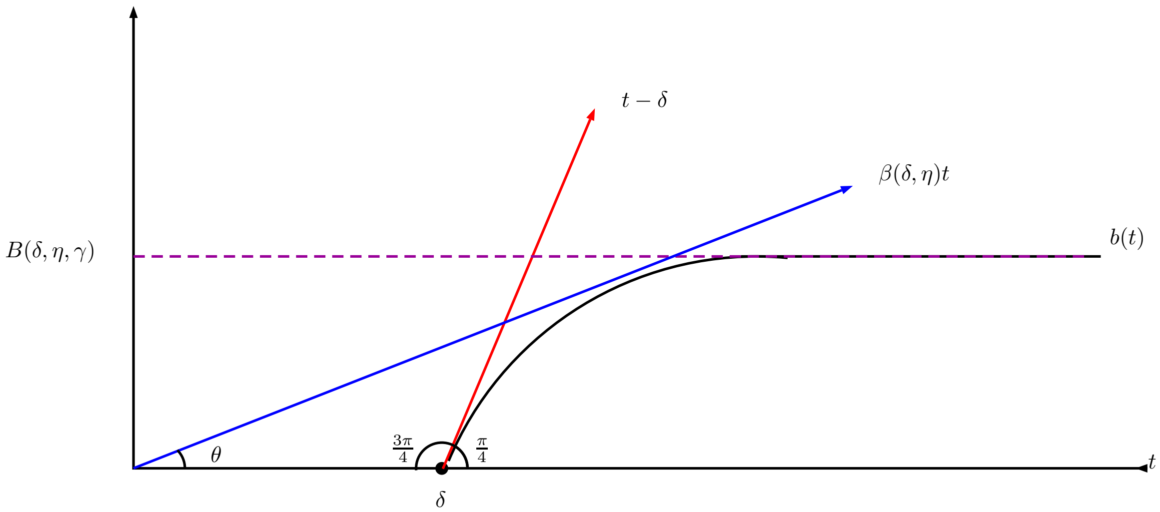


FIGURE 3.3: Construction of Linear Domination of $b(t)$

Since $b(t)$ is concave, $b(t) \leq t - \delta$. We seek a domination $\beta(\delta,\eta)t \geq b(t)$ such that $\beta(\delta,\eta) < 1$, $\beta(\delta,\eta)t \geq b(t)\ \forall\ t \geq 0$, and $\alpha(\delta) + \beta(\delta,\eta) < 1$. To see that at least one such $\beta(\delta,\eta)$ exists, refer to Figure 3.3. Let θ be the angle between $\beta(\delta,\eta)t$ and $y = 0$. Since $b'(\delta) = 1$, the angle between $t - \delta$ and $y = 0$ is $\frac{\pi}{4}$. Then the largest angle of the triangle formed by $\beta(\delta,\eta)t, t-\delta$, and $y = 0$ is $\frac{3\pi}{4}$. Thus $\theta < \frac{\pi}{4}$, and therefore there exists a $\beta(\delta,\eta) < 1$. To ensure $\beta(\delta,\eta) + \alpha(\delta) < 1$, choose δ and corresponding η such that this is true and θ is minimized, while $\beta(\delta,\eta)t \geq b(t)$[1].

In summary, we have that $a(t) \leq \alpha(\delta)t$ and $b(t) \leq \beta(\delta,\eta)t$, where $\beta(\delta,\eta) + \alpha(\delta) < 1$. Therefore,

$$h_N(\lambda) = \int_0^\infty e^{-r_0 t - \lambda t + \lambda(\alpha(\delta)t + \beta(\delta,\eta)t)}(r_0 + \eta\beta(\delta,\eta)t)\alpha(\delta)tdt$$

Denote $\xi = 1 - \alpha(\delta) - \beta(\delta,\eta)$. Then

$$\begin{aligned} h_N(\lambda) &= e^{-r_0 t - \xi\lambda t}(r_0 + \eta\lambda\beta(\delta,\eta)t)\lambda\alpha(\delta)tdt \\ &= \frac{\alpha(\delta)r_0^2\lambda + \lambda^2(\alpha(\delta)r_0\xi + 2\beta(\delta,\eta)\eta)}{(r_0+\xi\lambda)^3} \\ &\xrightarrow{\lambda\to\infty} 0 \end{aligned}$$

[1]Note: $\beta(\delta,\eta)$ will likely have to be determined numerically for specific cases.

Thus step (ii) is complete, and therefore $\psi(\lambda)$ has a maximum on $(0, \infty)$ for the η, δ that ensure $\beta(\delta, \eta) + \alpha(\delta) < 1$. □

3.5 Efficiency under Exponential Distribution

Lambert W-Function

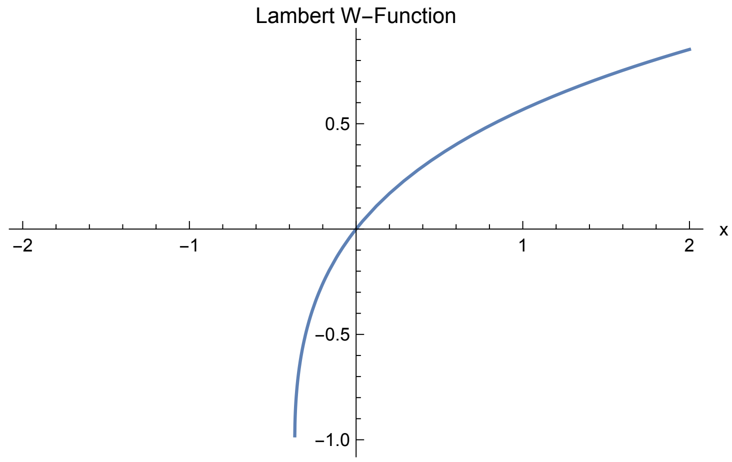


FIGURE 3.4: Lambert W-Function for Real Values

The proof of Theorem 3.4 requires the use of the Lambert $W-$Function, defined as the inverse function of $f(W) = We^W$. Figure 3.4 shows the function for $W \in \mathbb{R}$. Note that the function is positive for positive x and negative for negative x.

Suppose $g(v)$ is of the exponential class. That is, $g(v) = \gamma e^{-\gamma v}, \gamma > 0$. It will be proven that under certain conditions on η and γ, an exponential $g(v)$ causes ψ to increase without bound.

Theorem 3.4. *Suppose* $g(v) = \gamma e^{-\gamma v}$. *Then if* $\frac{2\gamma}{\gamma+\eta} > 1 + \frac{2}{\frac{2}{\gamma} + W\left(-\frac{\gamma}{\gamma+\eta} e^{-2-\frac{2\eta}{\gamma}}\right)}$, $\psi(\lambda) \xrightarrow{\lambda \to \infty} \infty$.

Proof. The proof will proceed via constructing a lower estimate for the numerator of $\psi(\lambda)$, denoted $N_L(\lambda)$ such that $N_L(\lambda) \to \infty$. Then the function $h(\lambda) = \frac{N_L(\lambda)}{\int_0^\infty S_Y(t)dt + v} \leq \psi(\lambda)$, and $h(\lambda) \to \infty$, thus $\psi(\lambda) \to \infty$.

To construct $N_L(\lambda)$, lower estimates for $a(t)$ and $b(t)$ will be obtained.

$$a(t) = \int_0^t e^{-(\eta+\gamma)v}(t - v)dv = \frac{\gamma t}{\gamma + \eta} + \frac{\gamma e^{-t(\eta+\gamma)}}{(\eta + \gamma)^2} - \frac{\gamma}{(\eta + \gamma)^2}$$

Now, $\frac{da}{dt} \leq \frac{\gamma}{\gamma+\eta}$, and $a(t) \geq 0$. Then $a(t) \geq \frac{\gamma t}{\gamma+\eta} - \frac{1}{(\eta+\gamma)}$.

Similarly,

$$b(t) = \int_0^t e^{-(\eta+\gamma)(t-r)}dr = \frac{1 - e^{-t(\eta+\gamma)}}{\eta + \gamma}$$

$b(t)$ is clearly concave, nonnegative, and $b(t) \to \frac{1}{\eta+\gamma}$; thus the lower bound for $b(t)$ will be piecewise linear.

$$b(t) \geq \begin{cases} \frac{\gamma t}{\gamma+\eta} - \frac{1}{(\eta+\gamma)} & ,0 \leq t \leq t_0 \\ b(t_0), & t > t_0 \end{cases}$$

where t_0 is the point of intersection between $b(t)$ and $y(t) = \frac{\gamma t}{\gamma+\eta} - \frac{1}{(\eta+\gamma)}$. Solving,

$$t_0 = \frac{2}{\gamma} + \frac{W\left(-\frac{\gamma}{\gamma+\eta}e^{-2-\frac{2\eta}{\gamma}}\right)}{\gamma+\eta}$$

where $W(\cdot)$ is the Lambert W-function described earlier. Thus

$$N_L(\lambda) = \int_0^{t_0} e^{-r_0 t - \lambda t + 2\lambda\left(\frac{\gamma t}{\gamma+\eta} - \frac{1}{(\eta+\gamma)}\right)} \left[r_0 + \eta\lambda\left(\frac{\gamma t}{\gamma+\eta} - \frac{1}{(\eta+\gamma)}\right)\right]\left[\frac{\lambda\gamma t}{\gamma+\eta} - \frac{1}{(\eta+\gamma)}\right] dt$$
$$+ \int_{t_0}^{\infty} e^{-r_0 t - \lambda t + \lambda\left(\frac{\gamma t}{\gamma+\eta} - \frac{1}{(\eta+\gamma)}\right) + \lambda b(t_0)} (r_0 + \eta\lambda b(t_0)) \left(\frac{\lambda\gamma t}{\gamma+\eta} - \frac{1}{(\eta+\gamma)}\right) dt$$

It is sufficient to ensure $-t_0 + \frac{2\gamma t_0}{\gamma+\eta} - \frac{2}{(\eta+\gamma)} > 0$. Then

$$\frac{2\gamma}{\gamma+\eta} > 1 + \frac{2}{t_0(\eta+\gamma)}$$

Plugging in t_0,

$$\frac{2\gamma}{\gamma+\eta} > 1 + \frac{2}{\frac{2}{\gamma} + W\left(-\frac{\gamma}{\gamma+\eta}e^{-2-\frac{2\eta}{\gamma}}\right)} \tag{3.5}$$

Thus, γ, η such that (3.5) is met ensure that $\psi(\lambda) \to \infty$. □

3.6 Extension to Random Stress and Nonconstant Intensity

Theorems 3.1 - 3.4 all assumed constant stress η and constant intensity λ. This section generalizes the analyses in Sections 3.2- 3.5 for nonconstant intensity and random stress.

The stochastic reliability models in both [9] and Chapter 2 both assumed a time-dependent intensity $\lambda(t)$. By setting $\lambda \equiv \max_t \lambda(t)$, Theorems 3.1- 3.3 provide a conservative set of conditions under which a maximum efficiency may be obtained. In these cases, ψ is actually a function of all possible $\lambda_{\max}$.

If the job stresses are random, as in Theorem 2.2, the above sections may still be utilized. Assume the sample space for $\mathcal{H}$ is compact and nonnegative. The efficiency is given by Theorem 2.2

and reproduced here:

$$\psi = \frac{1}{\int_0^\infty S_Y(t)dt + \nu} \left\{ \int_0^\infty e^{-\int_0^t r_0(x) - \int_0^t \lambda(x)dx + E_{\mathcal{H}}[a(t)+b(t)]} (r_0(t)E_{\mathcal{H}}[a(t)] + E_{\mathcal{H}}[\mathcal{H}a(t)b(t)])dt \right\} \tag{3.6}$$

WLOG, again suppose the sample space of $\mathcal{H}$ is discrete, given by $\{\eta_1, ..., \eta_m\}$ with respective probabilities $p_i, i = 1, ..., m$. Now suppose all mass is concentrated at $\eta_{[m]}$. Let $a_m(t) = \lambda \int_0^t e^{-\eta_m v} g(v)(t - v)dv$, and $b_m(t) = \lambda \int_0^t e^{-\eta_m(t-r)} \bar{G}_W(t-r)dr$. Then, the following are true:

(a) $E_{\mathcal{H}}[a(t) + b(t)] \leq a_m(t) + b_m(t)$

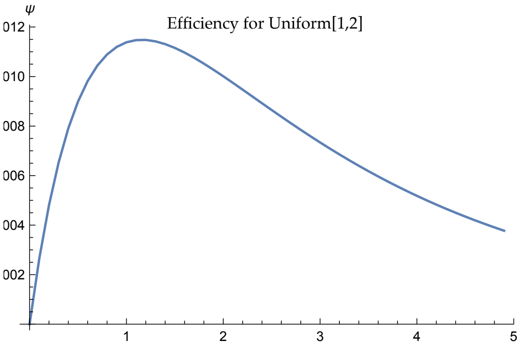
(b) $E_{\mathcal{H}}[a(t)] \leq a_m(t)$

(c) $E_{\mathcal{H}}[\mathcal{H}a(t)b(t)] \leq \eta_m a_m(t) b_m(t)$

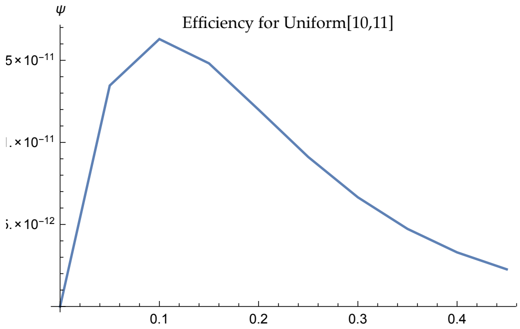
Thus, by replacing the expectations in (a) - (c) with their respective upper bounds in Theorems 3.1 - 3.3, analyses of the efficiency for the uniform, compact support, and Erlang classes may proceed as previously detailed. These estimates are conservative but sufficient.

For an exponential service life distribution and random stress, create a lower bound for the expectations in (a) - (c) by concentrating all mass at $\eta_{[1]}$. Then the conditions in Theorem 3.4 guarantee an explosion for the exponential distribution.

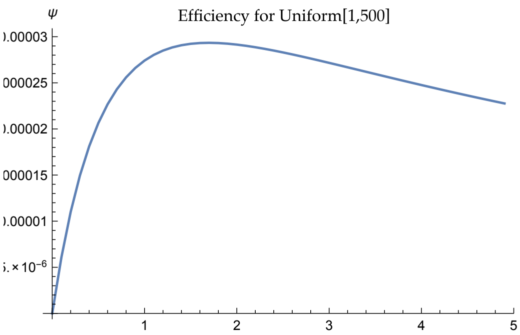
3.7 Implications and Numerical Illustrations



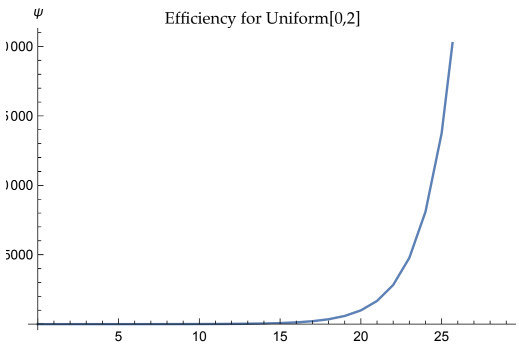
(A) $g_W(w) \sim$ Uniform[1,2]



(B) $g_W(w) \sim$ Uniform[10,11]



(C) $g_W(w) \sim$ Uniform[1,500]



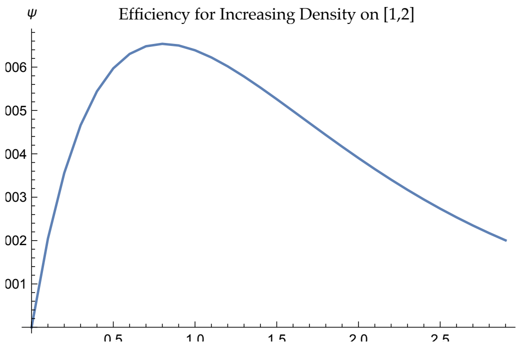
(D) $g_W(w) \sim$ Uniform[0,2]

FIGURE 3.5: $\psi(\lambda)$ under Various Uniform Service Distributions

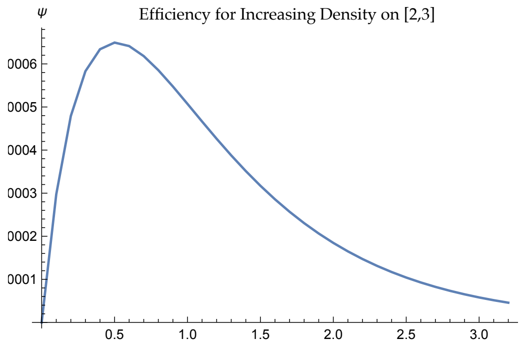
$g_W(w)$	σ^2	μ	Approximate Range of ψ	Approximate λ^*
Uniform[1,2]	1/12	1.5	$(0, 0.012)$	1.3
Unform[10,11]	1/12	10.5	$(0, 2 \times 10^{-11})$	0.1
Uniform[1,500]	499/12	250.5	$(0, 3 \times 10^{-5})$	1.6

TABLE 3.1: Comparison of Various Uniform Service Distributions and Resulting Effects on ψ

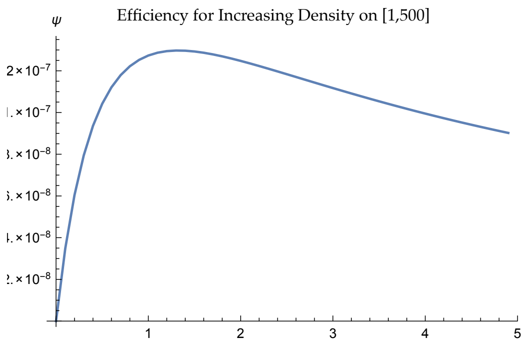
For the uniform service distribution, both the variance of $g_W(w)$ and the location of the support affect the efficiency itself. Figure 3.5 shows $\psi(\lambda)$ for various uniform distributions as an illustration. In all four cases, $r_0 = \nu = \eta = 1$. The variance of a uniform distribution is given by $\sigma^2 = \frac{d-c}{12}$. With the exception of Figure 3.5(d), which illustrates the explosion of ψ when the distribution has positive mass at 0, Table 3.1 compares possible values of ψ for different Uniform distributions. Notice that while the variance σ^2 does affect the range of ψ by several orders of magnitude, the location of the support has a much more powerful effect. Thus, if all service times are equally likely, a server is less efficient if it is consistently but mildly slow (Uniform[10,11]) compared to an inconsistent server (Uniform[1,500]).



(A) $g_W(w) = \frac{2x}{3}\mathbb{1}_{[1,2]}(x)$



(B) $g_W(w) = \frac{2x}{5}\mathbb{1}_{[2,3]}(x)$



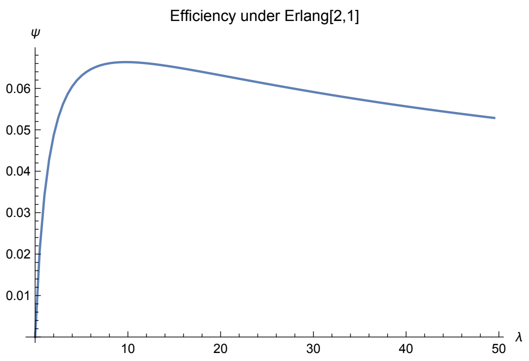
(C) $g_W(w) = \frac{2x}{500^2-1}\mathbb{1}_{[1,500]}(x)$

FIGURE 3.6: $\psi(\lambda)$ under Various Increasing Service Densities.

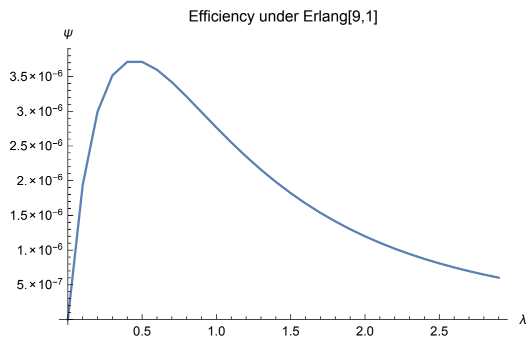
$g_W(w)$	σ^2	μ	Approximate Range of ψ	Approximate λ^*
$\frac{2x}{3}\mathbb{1}_{[1,2]}(x)$	≈ 0.0802	≈ 1.56	$(0, 7 \times 10^{-3})$	.75
$\frac{2x}{5}\mathbb{1}_{[2,3]}(x)$	≈ 0.0822	≈ 2.53	$(0, 7 \times 10^{-4})$	0.5
$\frac{2x}{500^2-1}\mathbb{1}_{[1,500]}(x)$	≈ 13888.5	333.35	$(0, 1.3 \times 10^{-7})$	1.4

TABLE 3.2: Comparison of Various Increasing Compact Service Densities and Resulting Effects on ψ

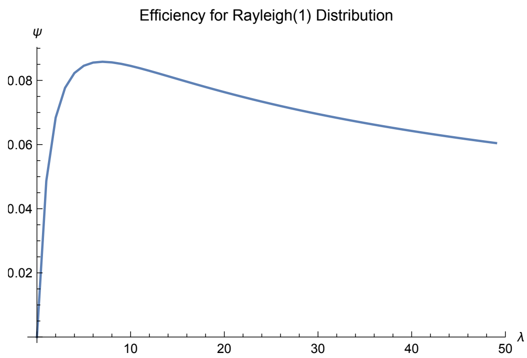
As an illustration of a distribution on compact support, consider the class of increasing densities $g_W(w) = cx\mathbb{1}_{[a,b]}(x)$. Several examples are given in Figure 3.6 and Table 3.2. For both compact supports of length 1, the variance is approximately the same, but the mean changes, producing an order of magnitude decrease in efficiency. Compared to the compact support of length 499, with a much larger mean, the efficiency decreases by 3 orders of magnitude. Notice, however, that the decline after the maximum is much less sharp.



(A) Erlang(2,1)



(B) Erlang(9,1)



(C) $g_W(w) \sim$ Rayleigh(1)

FIGURE 3.7: $\psi(\lambda)$ under Various Erlang and Rayleigh Service Distributions

$g_W(w)$	σ^2	μ	Approximate Range of ψ	Approximate λ^*
Erlang(2,1)	2	2	$(0, 0.7)$	9
Erlang(9,1)	9	9	$(0, 4 \times 10^{-6})$	0.5
Rayleigh(1)	$(4-\pi)/2$	$\sqrt{\pi/2}$	$(0, 0.9)$	8

TABLE 3.3: Comparison of Various Erlang and Rayleigh Service Distributions and Resulting Effects on ψ

Figure 3.7 gives two examples of ψ under an Erlang distribution. Notice the change in the efficiency as the mean increases. Here, since $\lambda = 1$, $\sigma^2 = \mu$, so the mean likely has the largest effect on ψ.

Comparing all examples in Figures 3.5- 3.7, the Rayleigh(1) service distribution imposes the highest maximum efficiency, followed closely by the Erlang(2,1) service distribution, with the Uniform[1,2] service distribution following. λ^* under the Erlang(2,1) service distribution is larger than for the Rayleigh(1) service distribution, indicating that a server whose service times follow the former distribution can handle a larger arrival intensity before its efficiency begins to decline than the latter.

The means for the Rayleigh(1), Erlang(2,1), and Uniform[1,2] distributions are similar, as shown in Tables 3.1 and 3.3, but the Uniform[1,2] distribution has equal probability of any service time in its support with large negative excess kurtosis. It is posulated that kurtosis, skew, and variance play large roles in the behavior and range of ψ. Compare the efficiency under the Erlang(2,1) service distribution with the efficiency under the Erlang(9,1) service distribution. Not only is the mean much lower for the Erlang(2,1) distribution, but the distribution is more strongly positive-skewed than the Erlang(9,1). Thus, more mass is concentrated at the left side of the distribution, indicating that the service times are more often shorter.

Finally, to note the effect of the typical stress level η on the range of ψ, compare Figure 1.4 with Figure 3.6(c). The service distribution and all other quantities remain the same, but Cha and Lee's numerical example set $\eta = 0.01$, whereas Figure 3.6(c) shows ψ under $\eta = 1$. The range of ψ decreases by two orders of magnitude with a 100 fold increase in η, with the shape remaining similar. In addition, the location of the maximum λ^* also inversely varies by the same magnitude.

Studying the efficiency under various service distributions aids not only in deciding when to implement a server intervention, but also aids in evaluating the performance of various servers given their service times.

Chapter 4

Extensions of the Single Server Model

4.1 Load Balancing Allocation for a Multichannel Server

Model Description

Previously, we had assumed that a web server functions as a single queue that attempts to process jobs as soon as they arrive. These jobs originally brought a constant stress η to the server, with the system stress reducing by η at the completion of each job.

Now, suppose we have a server partitioned into K channels. Denote each channel as $Q_k, \quad k = 1, ..., K$. Jobs arrive via a nonhomogenous Poisson process with rate $\lambda(t)$. Upon arrival, each job falls (or is routed) to the channel with the shortest queue length. If all queue lengths are equal or multiple channels have the shortest length, the job will enter one of the appropriate queues with equal probability.

We retain the previous notation for the baseline breakdown rate, or hazard function. This is denoted by $r_0(t)$ and is the hazard function under an idle system. We also retain the assumption that the arrival times $\mathfrak{T}$ are independent. In addition, the service times $\mathfrak{W}$ are i.i.d. with distribution $G_W(w)$. We assume that all channels are serving jobs at the same time, i.e. a job can be completed from any queue at any time. We do not require load balancing for service. In other words, any queue can empty with others still backlogged. We also retain the FIFO service policy for each queue.

Since we have now "balanced", or distributed, the load of jobs in the server, not all jobs will cause additional stress to the system. Suppose all jobs bring the same constant stress η upon arrival. Under load balancing, we will define the additional stress to the system as $\eta \max_k |Q_k|$. Figure 4.1 shows an example server with current stress of 4η.

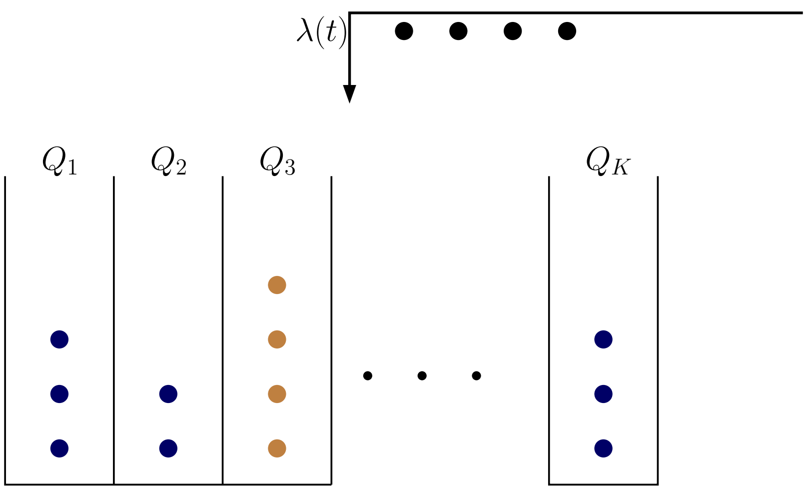
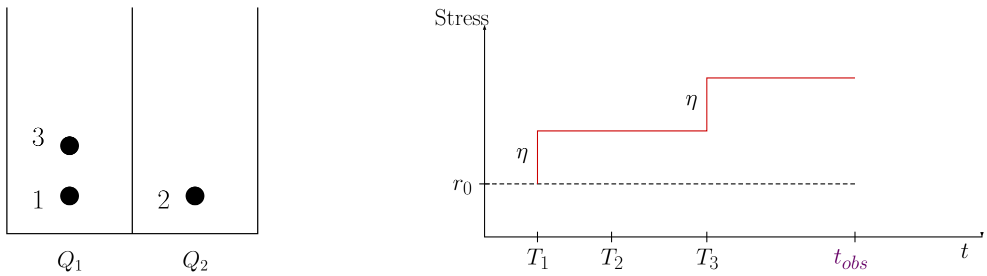


FIGURE 4.1: Partitioned Server with Load Balancing

Examples

Due to the dynamic nature of arrival times, allocation to queues, and service times, we have many possible configurations of jobs at any point in time. Therefore, the allocation scheme adds an additional layer of variation to the service times and order of service. The placement of jobs in the various queues (and thus the order of service and service times) is wholly dependent on all arrival times and service times of the prior arrivals. The following examples illustrate the effect on the workload stress added to the system in various scenarios.



(A) Queue Configuration

(B) Breakdown Rate Process Trajectory

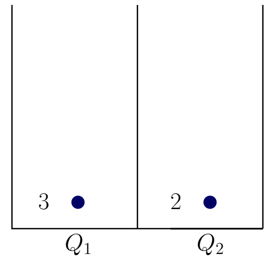
FIGURE 4.2: Example 4.1

Example 4.1. *Suppose for simplicity we have 2 channels. Suppose at the time of observation of the system, 3 jobs have arrived and none have finished. WLOG, suppose job 3 fell into* Q_1*. See Figure 4.2a. The stress to the system at* $t = t_{obs}$ *is* $r_0(t_{obs}) + 2\eta$*, as shown in Figure 4.2b.*

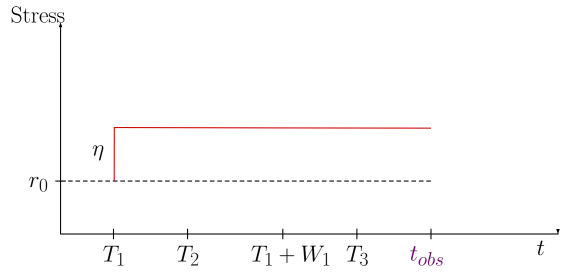
Note in example 4.1 that Job 2 does not add any additional stress to the system. Job 1 sees an empty queue upon arrival, and $\max_K |Q_K| = 1$ when it falls into any particular queue. Job 2 arrives as Job 1 is still being processed, and thus the placement of Job 1 forces Job 2 into the empty channel. Since $\max_K |Q_K|$ is still 1, the stress to the system doesn't change. Job 3 arrives as Jobs 1 and 2 are in service, and thus its choice of queue is irrelevant due to the configuration of the two queues at

T_3. Regardless of which queue Job 3 falls into, $\max_K |Q_K| = 2$. Thus the arrival of Job 3 increases the breakdown rate by η again.

The next example shows the change in system stress Job 1 from Example 4.1 when one job has finished processing before T_3.



(A) Queue Configuration



(B) Breakdown Rate Process Trajectory

FIGURE 4.3: Example 4.2

Example 4.2. *Consider the same two-channel system from Example 4.1. However, now suppose WLOG that* $T_3 < T_1 + W_1$*. In other words, service for Job 1 was completed before Job 3 arrived. Hence Job 3 will fall into the opposite queue as Job 2. The stress to the system at the time of observation would be* $r_0(t) + \eta$*. See Figures 4.3a and 4.3b.*

In this scenario, the workload due to Job 3 does not contribute any additional stress to the server. Also observe that upon completion of Job 1, the workload stress to the server does not decrease, as Job 2 still resides in the system and is being served.

Contrast this behavior with the breakdown rate process given in Chapter 2. In the single-channel, single-server model described in both [9] and Section 2.1, each job adds stress to the server upon arrival. Under the load balancing allocation scheme, the additional stress to the server depends on the arrival and service times of all prior jobs. From a stochastic perspective, this breakdown rate process has full memory.

The examples above illustrate that $\max_K |Q_K|$ depends on the intersection of the intervals $I_j = [T_j, T_j + W_j], j = 1, ..., N(t)$. The next section details the methodology to obtain the configuration of jobs in the server at time t by deccomposition of $\bigcup_{j=1}^{N(t)} I_j$ into disjoint atoms and derives the stochastic breakdown rate process under the load balancing allocation scheme.

Breakdown Rate Process and Conditional Survival Function

Let $\varepsilon = (\varepsilon_1, ..., \varepsilon_{N(t)})$ be a $N(t)-$tuple whose components $\varepsilon_j = \{\emptyset, c\}$, where $\emptyset$ denotes the empty set, and c denotes the complement of the set. Let $E = \{\varepsilon : \varepsilon_j = \{\emptyset, c\}\}$ denote the set of all possible

ε, excepting $\varepsilon = (c, ..., c)$. Then by Lemma A.8 (Appendix A),

$$\bigcup_{j=1}^{N(t)} I_j = \bigcup_{\varepsilon \in E} \bigcap_{j=1}^{N(t)} I_j^{\varepsilon_j} \tag{4.1}$$

Remark: $\cap_{j=1}^{N(t)} I_j^{\varepsilon_j}$ indicates which jobs are still in the server at time t. The union is disjoint; thus only one ε will describe the server configuration at any given time t. For example, if 3 jobs have arrived to the server at time t_{obs}, $|E| = 3 \times 2 - 1 = 5$. These may be enumerated:

- $I_1 \cap I_2 \cap I_3$
- $I_1^c \cap I_2 \cap I_3$
- $I_1^c \cap I_2^c \cap I_3$
- $I_1^c \cap I_2 \cap I_3^c$
- $I_1 \cap I_2^c \cap I_3^c$

As an illustration, refer to Example 4.1. All three jobs are in the system at $t = t_{\text{obs}}$ (that is, none have completed service), and thus $t_{\text{obs}} \in I_1 \cap I_2 \cap I_3$. Expanding, $t_{\text{obs}} \in [T_1, T_1 + W_1], [T_2, T_2 + W_2]$, and $[T_3, T_3 + W_3]$.

Compare the case with that of Example 4.2. In this case, three jobs have arrived at $t = t_{\text{obs}}$, but Job 1 has finished by t_{obs}. Thus $t_{\text{obs}} \notin I_1$, but since Jobs 2 and 3 are still in the system, $t_{\text{obs}} \in I_2 \cap I_3$. Thus $t_{\text{obs}} \in I_1^c \cap I_2 \cap I_3$.

Now, since the additional workload stress to the server is a multiple of $\eta \max_K |Q_K|$, it remains to derive the appropriate multiplier that accounts for the number of jobs that contribute additional stress to the system. Let $n = \sum_{j=1}^{N(t)} \mathbb{1}(\varepsilon_j = \emptyset | \varepsilon_j \in \varepsilon)$ for a particular ε, and let α_ε be the multiplier that indicates the number of jobs that contribute stress η to the system. Under [9] and the generalization in Section 2.1, every uncompleted job in the system contributes stress, thus $\alpha_\varepsilon = n$.

Under the load balancing scheme, $\alpha_\varepsilon = \lfloor \frac{n+1}{K} \rfloor$, where K is the number of channels in the server. This is due to the allocation scheme's attempts to evenly distribute jobs across channels. Thus, for Example 4.1, $n = 3$, and $K = 2$, meaning $\alpha_\varepsilon = 2$, as illustrated in Figure 4.2b and for Example 4.2, $\alpha_\varepsilon = \lfloor \frac{3+1}{2} \rfloor = 1$, as in Figure 4.3b.

Then, the stochastic breakdown rate process under the load balancing allocation scheme is given by

$$\mathcal{B}(t) = r_0(t) + \eta \sum_{\varepsilon \in E} \alpha_\varepsilon \mathbb{1}_{I_1^{\varepsilon_1} \cap I_2^{\varepsilon_2} \cap I_{N(t)}^{\varepsilon_{N(t)}}}(t)$$

Under this expression, only one indicator function will be nonzero at any given point in time, since all atoms are disjoint. Now, $I_1^{\varepsilon_1} \cap I_2^{\varepsilon_2} \cap ... \cap I_{N(t)}^{\varepsilon_{N(t)}}$ may be expressed as one interval $[L_\varepsilon, R_\varepsilon]$, where

$$L_\varepsilon = \max\left(\{T_j : \varepsilon_j = \emptyset\}_{j=1}^{N(t)}\right)$$
$$R_\varepsilon = \min\left(\{T_j + W_j : \varepsilon_j = \emptyset\}_{j=1}^{N(t)}, \{T_j : \varepsilon_j = c\}_{j=1}^{N(t)}\right)$$

Thus, for a server with K channels under a load balancing routing scheme with all jobs bringing constant stress η, the breakdown rate process $\mathcal{B}(t)$ may be expressed as

$$\mathcal{B}(t) = r_0(t) + \eta \sum_{\varepsilon \in E} \alpha_\varepsilon \mathbb{1}_{[L_\varepsilon, R_\varepsilon]}(t) \tag{4.2}$$

Thus, the conditional survival function under the load balancing scheme is given by

$$\begin{aligned} S_{Y|\mathfrak{T},\mathfrak{W},N(t)}(t|\mathfrak{t},\mathfrak{w},n) &= e^{-\int_0^t \mathcal{B}(s)ds} \\ &= \bar{F_0}(t) \exp\left(-\eta \int_0^t \sum_{\varepsilon \in E} \alpha_\varepsilon \mathbb{1}_{[L_\varepsilon, R_\varepsilon]}(s)ds\right) \\ &= \bar{F}_0(t) \exp\left(-\eta \sum_{\varepsilon \in E} \alpha_\varepsilon \min(t - L_\varepsilon, R_\varepsilon)\right) \end{aligned}$$

Remarks

Finding the survival function of the single-channel environment relied on the independence of the set of arrival times and service times. From (4.1), the independence is clearly lost. As noted before, the random breakdown process has full memory, and thus is completely dependent upon the entire trajectory up to $t = t_{\text{obs}}$.

4.2 Clustered Tasks in a Multichannel Server

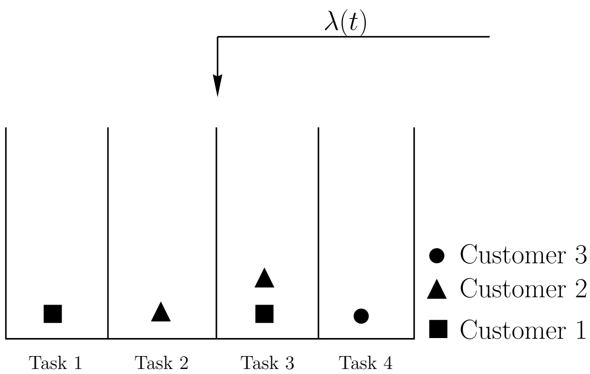


FIGURE 4.4: Illustration of Clustered Tasks in a Multichannel Server

The previous multichannel server model in Section 4.1 implicitly assumed each job comes with one task, and all channels are identical in their ability to serve any task brought by a job. A classic illustration is a block of registers at a retail establishment. Each customer will survey the length of the various queues at each register before choosing the shortest queue. Viewing each of these separate registers as a channel in a single server under these conditions gave rise to the load balancing allocation model detailed in the previous section. This section presents a different interpretation of a multichannel, single-server model.

Suppose a server has multiple channels $Q_1, ..., Q_K$, but each channel serves a different type of task. A customer arrives to the server and may select any number from 0 to K tasks for the server to perform. Said customer will select each possible task j with probability p_j. Figure 4.4 illustrates an example of such a situation in which three customers visit the server and each customer picks a different number and set of tasks at random. A customer is considered fully serviced (i.e. the job is complete) upon completion of the last task belonging to that particular customer.

Model Assumptions

The following mathematical assumptions are made for the multichannel server with clustered tasks:

(i) Customers arrive to the server with K channels via a nonhomogenous Poisson process (NHPP) with intensity $\lambda(t)$.

(ii) The breakdown rate of the idle server is given by $r_0(t)$.

(iii) Each channel corresponds to a different task the server can perform.

(iv) The selection of each task is a Bernoulli random variable with probability p_k. Thus the number of tasks selected by each customer is a binomial random variable.

(v) The workload stress to the server is a constant multiple η of the number of tasks requested by the customer, i.e. the additional stress is given by ηN, where N is the number of tasks requested.

(vi) The PDF of each channel's service time is given by $g_i(w), i = 1, ..., K$. Since the customer's service is not complete until all requested tasks have finished, the service life distribution for the customers is given by $\max_i G_i(w)$.

Under these assumptions, this model is a special interpretation of the random stress environment developed in Chapter 2. In this case, the random workload stress is ηN, where N is a binomial random variable, and the service life distribution $G_W(w) = \max_i G_i(w)$, which may be easily obtained through the mathematical properties of order statistics. Two variations are considered in this section: independent channels and correlated channels.

Independent Channels in a Clustered Task Server

Suppose the selection probabilities for each task in the server are identical, that is, $p_1 = p_2 = \ldots = p_K = p$. Then $N \sim \text{Bin}(K, p)$. Using Theorem 2.1, the survival function of the multichannel server is given in the following theorem:

Theorem 4.1 (Survival Function of Multichannel Server with Clustered Tasks and Independent Channels)**.** *Suppose conditions (i)-(vi) above are satisfied. In addition, assume $p_1 = p_2 = \ldots = p_K = p$. Then the survival function of the server is given by*

$$S_Y(t) = \bar{F}_0(t) \exp\left(-K\eta \left[e^{-\eta t}\left(1 - p + pe^{-\eta t}\right)^{K-1} - p(1-p)^{K-1}\right] \int_0^t m(t-w)\bar{G}_W(w)dw\right)$$

where $m(x) = \int_0^x \lambda(s)ds$, $\bar{F}_0(t) = e^{-\int_0^t r_0(s)ds}$, $\bar{G}_W(w) = 1 - G_W(w)$, and $G_W(w) = \max_i G_i(w)$.

Proof. Since $p_1 = \ldots = p_K = p$, the number of tasks selected by any particular customer $N \sim \text{Bin}(K, p)$. Then the $\mathcal{H}$ from Theorem 2.1 is given by $\mathcal{H} = \eta N$. Thus

$$S_Y(t) = \bar{F}_0(t) \exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right]\right)$$

In this case,

$$\begin{aligned}
&E\left[\mathcal{H}\int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right] \\
&\qquad = E\left[\eta N\int_0^t e^{-\eta N w} m(t-w)\bar{G}_W(w)dw\right] \\
&\qquad = \sum_{n=0}^{K}\left[\eta n\int_0^t e^{-\eta n w} m(t-w)\bar{G}_W(w)dw\right]\cdot P(N=n) \\
&\qquad = \sum_{n=0}^{K}\left[\eta n\int_0^t e^{-\eta n w} m(t-w)\bar{G}_W(w)dw\right]\binom{K}{n}p^n(1-p)^{K-n} \\
&\qquad = \eta\int_0^t m(t-w)\bar{G}_W(w)\left(\sum_{n=0}^{K} n e^{-\eta n w}\binom{K}{n}p^n(1-p)^{K-n}\right)dw
\end{aligned}$$

Now,

$$\begin{aligned}
\sum_{n=0}^{K} n e^{-\eta n w}\binom{K}{n}p^n(1-p)^{K-n} &= \sum_{n=0}^{K}\frac{K!}{(K-n)!n!} n e^{-\eta n w} p^n (1-p)^{K-n} \\
&= \sum_{n=0}^{K}\frac{K(K-1)!}{(n-1)!(K-1-(n-1))!} e^{-\eta n w} p^n (1-p)^{K-n} \\
&= \sum_{n=0}^{K} K\binom{K-1}{n-1} e^{-\eta n w} p^n (1-p)^{K-n}
\end{aligned}$$

Making a change of indices, let $j = n - 1$. Then

$$\sum_{n=0}^{K} K\binom{K-1}{n-1} e^{-\eta n w} p^n (1-p)^{K-n} = K \sum_{j=0}^{K-1} \binom{K-1}{j} p^{j+1}(1-p)^{K-(j+1)} e^{-\eta(j+1)w}$$

Note the above resembles a scaled and shifted moment generating function of a binomial random variable. Let $X \sim \text{Bin}(K-1, p)$. Then

$$\begin{aligned} K\sum_{j=0}^{K-1} \binom{K-1}{j} p^{j+1}(1-p)^{K-(j+1)} e^{-\eta(j+1)w} &= K\left(E\left[e^{-\eta(X+1)t}\right] - P(X=0)\right) \\ &= K\left(e^{-\eta t} E\left[e^{-\eta X t} - p(1-p)^{K-1}\right]\right) \\ &= K\left(e^{-\eta t}\left[1 - p + pe^{-\eta t}\right]^{K-1} - p(1-p)^{K-1}\right) \end{aligned}$$

Thus,

$$S_Y(t) = \bar{F}_0(t) \exp\left(-K\eta\left[e^{-\eta t}\left(1 - p + pe^{-\eta t}\right)^{K-1} - p(1-p)^{K-1}\right] \int_0^t m(t-w)\bar{G}_W(w)dw\right)$$

□

Correlated Channels in a Cluster Server

Now suppose the server tasks are correlated, in that the selection of one particular task may affect the selection of any or all of the other tasks. Thus the channels are a sequence of dependent Bernoulli random variables. The construction of dependent Bernoulli random variables is given in [14], and a summary is given.

Dependent Bernoulli Random Variables and the Generalized Binomial Distribution

Korzenwioski [14] constructs a sequence of dependent Bernoulli random variables using a binary tree that distributes probability mass over dyadic partitions of [0,1]. Let $0 \leq \delta \leq 1$, $0 < p < 1$, and $q = 1 - p$. Then define the following quantities:

$$\begin{aligned} q^+ &:= q + \delta p \qquad & p^+ &:= p + \delta q \\ q^- &:= q(1-\delta) \qquad & p^- &:= p(1-\delta) \end{aligned} \tag{4.3}$$

The quantities in (4.3) satisfy the following conditions:

$$\begin{aligned} q^+ + p^- = q^- + p^+ = q + p = 1 \\ qq^+ + pq^- = q, \quad qp^- + pp^+ = 1 \end{aligned} \tag{4.4}$$

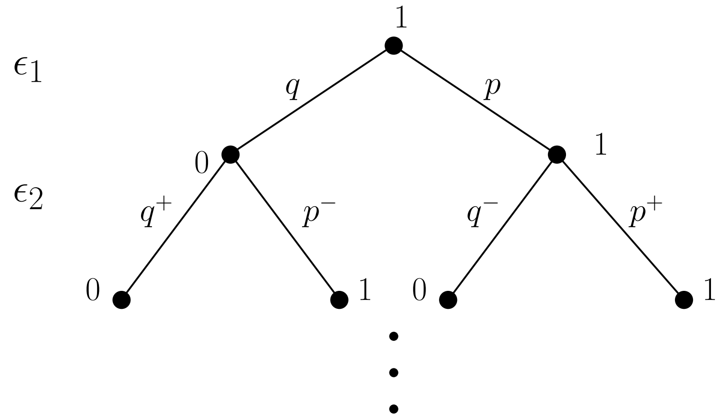


FIGURE 4.5: Construction of Dependent Bernoulli Random Variables

Figure 4.5 shows the construction shows the dependencies. The following examples using coin flips illustrate the effect of the dependency coefficient δ:

Example 4.3 **($\delta = 1$).** *For $\delta = 1$, $q^+ = q + p = 1, q^- = 0$, $p^+ = p + q = 1$, and $p^- = 0$. Supposing the first coin flip $\varepsilon_1 = 1$. Then every successive ε_i will also be 1. Similarly if $\varepsilon_1 = 0$. Thus the result of the first coin flip completely determines the outcomes of all the rest.*

Example 4.4 **($\delta = 0$).** *For $\delta = 0$, $q^+ = q^- = q$, and $p^+ = p^- = p$. Thus, the first coin flip (and all subsequent ones) have no effect on the ones that follow.*

Example 4.5 **($\delta = \frac{1}{4}$).** *Suppose $p = q = \frac{1}{2}$. Then $p^+ = q^+ = \frac{5}{8}$, and $p^- = q^- = \frac{3}{8}$. Then the subsequent outcomes $\varepsilon_i, i \geq 2$ are more likely to match the outcomes of ε_1 than not.*

Now suppose $p = \frac{1}{4}, q = \frac{3}{4}$. Then $p^+ = \frac{7}{16}, p^- = \frac{3}{16}, q^+ = \frac{13}{16}$, and $q^- = \frac{9}{16}$. In this example of an unfair coin, the dependency coefficient δ still attempts to skew the results following the first coin flip in favor of the outcome of ε_1. However, the dependency here heightens the effect of $\varepsilon_1 = 0$ on subsequent flips, and cannot overcome the discrepancy between the probability of success and failure to skew $\varepsilon_i, i \geq 2$ in favor of a 1 following the outcome of $\varepsilon_1 = 1$.

Using these dependent Bernoulli random variables, [14] presents a Generalized Binomial Distribution for identically distributed but dependent Bernoulli random variables.

Generalized Binomial Distribution

Let $X = \sum_{i=1}^{n} \varepsilon_i$, where $\varepsilon_i, i = 1, ..., n$ are identically distributed Bernoulli random variables with probability of success p and dependency coefficient δ. Then

$$P(X = k) = q\binom{n-1}{k}(p^-)^k(q^+)^{n-1-k} + p\binom{n-1}{k-1}(p^+)^{k-1}(q^-)^{n-1-(k-1)} \quad (4.5)$$

Survival Function of Correlated Channels in a Cluster Server

Suppose the selection of tasks may be modeled by the dependent Bernoulli random variables given in the previous section. That is, suppose the customer selects Tasks 1-K in sequence, and the selection or rejection of Task 1 affects all subsequent tasks by a dependency coefficient δ. From [14], the correlation between task selections $\varepsilon_i, \varepsilon_j$ is given by

$$\rho = \mathrm{Cor}(\varepsilon_i, \varepsilon_j) = \begin{cases} \delta, & i = 1; j = 2, ..., K \\ \delta^2, & i \neq j; i, j \geq 2 \end{cases} \tag{4.6}$$

This illustrates the dependency of Tasks 2-K on the outcome of Task 1, and notes that while Tasks 2-K are still correlated with each other, the dependency is much lower. In a similar fashion to the independent channel server, the survival function is derived.

Theorem 4.2 (Survival Function of Multichannel Server with Clustered Tasks and Dependent Channels)**.** *Suppose conditions (i)-(vi) above are satisfied. In addition, suppose the selection of channels* $1 - K$ *are determined by identically distributed Bernoulli random variables with dependency coefficient* δ *as defined in [14]. Then the survival function of the server is given by*

$$S_Y(t) = \bar{F}_0(t) \exp\left(-\eta \int_0^t m(t-w)\bar{G}_W(w)\mathcal{S}(w)dw\right) \tag{4.7}$$

where $m(x) = \int_0^x \lambda(s)ds$*, and*

$$\begin{aligned}\mathcal{S}(w) = &\sum_{n=0}^{K} e^{-\eta n w} \sum_{j=0}^{K-n-1} \binom{K-1}{n-1, j, K-1-n-j} p^{K-1-j}(1-p)^{j+1}\delta^{K-1-n-j}(1-\delta)^n \\ &+ \sum_{n=0}^{K} n e^{-\eta n w} \sum_{i=0}^{n-1} \binom{K-1}{K-1-n, i, n-1-i} p^{i+1}(1-p)^{K-n}\delta^{n-1-j}(1-\delta)^{K-n-j}\end{aligned}$$

Proof. By Theorem 2.1,

$$S_Y(t) = \bar{F}_0(t) \exp\left(-E\left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right]\right)$$

Similar to the proof of Theorem 4.1, $\mathcal{H} = \eta X$, where this time X has the generalized binomial distribution given in (4.5). Then

$$\begin{aligned}
E & \left[\mathcal{H}\int_0^t e^{-\mathcal{H}w} m(t-w)\bar{G}_W(w)dw\right] \\
&= \sum_{x=0}^{K}\left[\eta x\int_0^t e^{-\eta x w} m(t-w)\bar{G}_W(w)dw\right] P(X=x) \\
&= \sum_{x=0}^{K}\eta x\left[\int_0^t e^{-\eta x w} m(t-w)\bar{G}_W(w)dw\right]\left[q\binom{K-1}{x}(p^-)^x(q^+)^{K-1-x}\right] \\
&\quad + \sum_{x=0}^{K}\eta x\left[\int_0^t e^{-\eta x w} m(t-w)\bar{G}_W(w)dw\right]\left[p\binom{K-1}{x-1}(p^+)^{x-1}(q^-)^{K-x}\right] \\
&= \eta\int_0^t m(t-w)\bar{G}_W(w)(\mathcal{S}_1(w)+\mathcal{S}_2(w))dw
\end{aligned}$$

where $\mathcal{S}_1(w) = \sum_{x=0}^{K} xe^{-\eta x w} q\binom{K-1}{x}(p^-)^x(q^+)^{K-1-x}$
and $\mathcal{S}_2(w) = \sum_{x=0}^{K} xe^{-\eta x w} p\binom{K-1}{x-1}(p^+)^{x-1}(q^-)^{K-x}$. Using the definitions given in (4.3),

$$\begin{aligned}
\mathcal{S}_1(w) &= \sum_{x=0}^{K} xe^{-\eta x w}(1-p)\binom{K-1}{x}(p-\delta p)^x(1-p+\delta p)^{K-1-n} \\
&= \sum_{x=0}^{K} xe^{-\eta x w}(1-p)\binom{K-1}{x}p^x(1-\delta)^x\sum_{j=0}^{K-1-x}\binom{K-1-x}{j}(1-p)^j(\delta p)^{K-1-x-j}
\end{aligned}$$

Now, $x\binom{K-1}{x}\binom{K-1-x}{j} = \frac{(K-1)!}{(x-1)!j!(K-1-x-j)!} = \binom{K-1}{x-1,j,K-1-x-j}$. Then

$$\mathcal{S}_1(w) = \sum_{x=0}^{K} e^{-\eta x w}\sum_{j=0}^{K-x-1}\binom{K-1}{x-1,j,K-1-x-j}(1-p)^{j+1}(1-\delta)^x\delta^{K-1-x-j}p^{K-1-j}$$

Similarly,

$$\begin{aligned}
\mathcal{S}_2(w) &= \sum_{x=0}^{K} xe^{-\eta x w}p\binom{K-1}{x-1}(p+\delta(1-p))^{x-1}((1-p)(1-\delta))^{K-x} \\
&= \sum_{x=0}^{K} xe^{-\eta x w}p(1-\delta)^{K-x}(1-p)^{K-x}\sum_{i=0}^{x-1}\binom{x-1}{i}p^i(1-\delta)^i\delta^{x-1-i} \\
&= \sum_{x=0}^{K} xe^{-\eta x w}\sum_{i=0}^{x-1}x\binom{K-1}{K-1-x,i,x-1-i}p^{i+1}\delta^{x-1-i}(1-\delta)^{K-x+i}(1-p)^{K-x}
\end{aligned}$$

Clearly $\mathcal{S}(w) = \mathcal{S}_1(w) + \mathcal{S}_2(w)$ □

Chapter 5

Systems of Servers under a Random Stress Environment

5.1 Systems of Correlated Servers

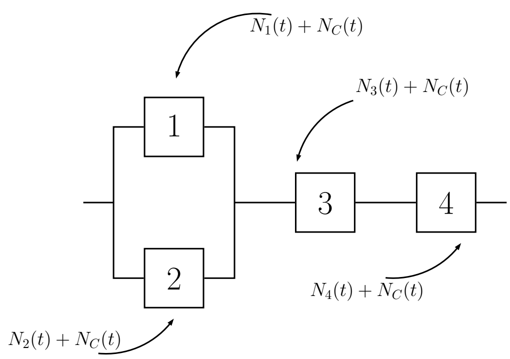


FIGURE 5.1: Logical Topology for a Hypothetical System with Correlated Traffic Streams

Chapter 2 presented a dynamic model for a single server under a random stress environment. In many applications, particularly retail and manufacturing, servers are organized into networks. Each server forms a node in the network, and the performance of each server affects the health of the network as a whole. In addition to the physical layout of the network, every system of components has a logical topology, or reliability topology.

It is common in the analysis of system reliability to employ *structure functions* which define the system state as a function of the component states. These structure functions give the system state (with a binary assumption of *working* or *failed*) as a function of component states [15]. Denote x_i as the state of component i. Then

$$x_i := \begin{cases} 0, & \text{if } i \text{ has failed} \\ 1, & \text{if } i \text{ is working} \end{cases}$$

Then the structure function of a system with n components is given by

$$\phi(\mathbf{x}) = \begin{cases} 0, & \text{if the system has failed when in state } \mathbf{x} \\ 1, & \text{if the system is working when in state } \mathbf{x} \end{cases}$$

where $\mathbf{x} = (x_1, ..., x_n)$ is known as the *state vector*.

The reliability topology of the network is given in a reliability block diagram (RBD), which shows the logical arrangement of the network. The existence of a "path" through the RBD equates to $\phi(\mathbf{x}) = 1$, where the path is traversed through the systems as nodes are operational; that is, $x_i = 1$. (See Section 1.1.)

As an example, refer to Figure 5.1. In this particular system, if either of Nodes 3 or 4 fails, then no path exists through the diagram. However, if only one of Nodes 1 or 2 fails, with both of Nodes 3 and 4 operational, then a path exists through the system. Thus the network is operational if any of the following sets of nodes are operational: $\{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}$. In this system, Nodes 1 and 2 are in *parallel*. The subsystem formed by Nodes 1 and 2, and Nodes 3 and 4 are all in *series*.

Consider a system with multiple nodes such that the external traffic goes to each node individually. Many systems in retail and manufacturing follow this model. For example, instead of viewing a series of checkout registers as a $G/G/K$ queueing system, the system may be interpreted as a parallel logical topology of K nodes with separate but correlated traffic streams to each node. A manufacturing system in which each manufacturing site is responsible for assembling a portion of a widget may have separate shipments of raw materials arriving to each location.

In each of these examples, the separate traffic streams are certainly correlated. The arrivals to each node will be modeled by a nonhomogenous Poisson processes (NHPP), as in Chapter 2, with the introduction of a *correlator process* that will ensure all nodes are correlated but conditionally independent.

Let $\{N_1(t) : t \geq 0\}, \{N_2(t) : t \geq 0\}, \ldots, \{N_K(t) : t \geq 0\}$ and a *correlator process* $\{N_c(t) : t \geq 0\}$ be mutually independent nonhomogenous Poisson processes (NHPPs) with intensities $\lambda_i(t), i = 1, ..., K$ and $\lambda_c(t)$, respectively. Now suppose there are K components (or queues) in this system, denoted $Q_{\ell,c}, \ell = 1, ..., K$ such that the arrival processes $\{\mathcal{N}_{\ell,c}(t) : t \geq 0\}, \ell = 1, ..., K$ are given by $\mathcal{N}_{\ell,c}(t) = N_\ell(t) + N_c(t)$.

By Lemma A.9, the sum of n independent NHPPs remains a NHPP. Thus, since $\{N_\ell\}_{\ell=1}^K$ and $\{N_c(t)\}$ are mutually independent, and $\mathcal{N}_\ell(t) = N_\ell(t) + N_c(t)$ is a NHPP, $E[\mathcal{N}_{\ell,c}(t)] = \lambda_\ell(t) + \lambda_c(t)$.

The covariance of $\mathcal{N}_i$ and $\mathcal{N}_j$ is given by

$$Cov(\mathcal{N}_{i,c}, \mathcal{N}_{j,c})(t) = E[N_{q_i} N_{q_j}] - E[N_{q_i}]E[N_{q_j}] = \lambda_c(t)$$

and the correlation between $\mathcal{N}_{i,c}$ and $\mathcal{N}_{j,c}$ is thus given by

$$\rho_{\mathcal{N}_{i,c},\mathcal{N}_{j,c}(t)} = \frac{Cov(\mathcal{N}_{i,c},\mathcal{N}_{j,c})}{\sigma_{\mathcal{N}_{i,c}}\sigma_{\mathcal{N}_{j,c}}} = \frac{\lambda_c(t)}{\sqrt{(\lambda_i(t)+\lambda_c(t))(\lambda_j(t)+\lambda_c(t))}}$$

Survival Function of System with Two Correlated Nodes

Series System

Suppose two nodes $Q_{\ell,1}$ and $Q_{\ell,2}$ are arranged in series with arrival processes $\mathcal{N}_1(t) = N_1(t) + N_c(t)$ and $\mathcal{N}_2(t) = N_2(t) + N_c(t)$. Let the NHPPs $\{N_\ell : t \geq 0\}, \ell = 1,2$ and c have arrival times $\{T_{j_\ell}\}_{j_\ell=1}^{N_\ell(t)}$, service times $\{W_{j_\ell}\}_{j_\ell=1}^{N_\ell(t)}$, and stresses $\{\mathcal{H}_{j_\ell}\}_{j_\ell=1}^{N_\ell(t)}$ as in Chapter 2. Assume that all stresses $\mathcal{H}_{j_\ell} \overset{i.i.d.}{\sim} \mathcal{H}$. In addition, all service times regardless of node are i.i.d. with distribution $G_W(w)$ and pdf $g_W(w)$. Let the baseline breakdown rate for $Q_{\ell,c}$ be $r_{0_{\ell,c}}(t)$. Jobs in both queues add stress to the server until completion. Then for $Q_{\ell,c}$, the breakdown rate process for each node is given by

$$\mathcal{B}_{\ell,c}(t) = r_{0_{\ell,c}}(t) + \sum_{j_\ell=1}^{N_\ell(t)} \mathcal{H}_{j_\ell}\mathbb{1}(T_{j_\ell} \leq t \leq T_{j_\ell} + W_{j_\ell}) + \sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c}\mathbb{1}(T_{j_c} \leq t \leq T_{j_c} + W_{j_c}) \tag{5.1}$$

The system survives past time t if and only if both $Q_{1,c}$ and $Q_{2,c}$ survive past time t. Let $Y_{i,c}, i = 1,2$ be the random length of the node lifetime under workload (or renewal cycle if the node can be rebooted) $Q_{i,c}, i = 1,2$, and Y_S the system life under workload.

$Q_{1,c}$ and $Q_{2,c}$ are conditionally independent under $\{N_c(t) : t \geq 0\}, \{T_{j_c} = t_{j_c}\}_{j_c=1}^{N_c(t)}$, $\{W_{j_c} = w_{j_c}\}_{j_c=1}^{N_c(t)}$, and $\{\mathcal{H}_{j_c} = \eta_{i_{j_c}}\}_{j_c=1}^{N_c(t)}$. Thus

$$\begin{aligned}
&P\left(Y_S > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right) \\
&\quad = P\left(Y_{1,c} > t \cap Y_{2,c} > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right) \\
&\quad = \prod_{i=1}^{2} P\left(Y_i > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right)
\end{aligned} \tag{5.2}$$

By Lemma A.10,

$$\begin{aligned}
&P\left(Y_{\ell,c} > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right) \\
&= \bar{F}_{0_{\ell,c}}(t) \exp\left(-\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c})\right) \\
&\qquad \times \exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t \exp(-\mathcal{H}w) m_\ell(t-w)\bar{G}_W(w)dw\right]\right)
\end{aligned} \tag{5.3}$$

where $\bar{F}_{0_{\ell,c}} = \exp\left(-\int_0^t r_0(x)dx\right)$. Thus from (5.2),

$$\begin{aligned}
&P\left(Y_s > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right) \\
&\quad = \bar{F}_{0_{1,c}}(t)\bar{F}_{0_{2,c}}(t)\exp\left(-2\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c}\min(W_{j_c}, t - T_{j_c})\right) \\
&\quad \times \exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t (m_1(t-w) + m_2(t-w))\exp(-\mathcal{H}w)\bar{G}_W(w)dw\right]\right) \qquad (5.4)
\end{aligned}$$

The unconditional survival function for the two-component correlated series system is given in the following theorem.

Theorem 5.1. *Let $\{N_1(t) : t \geq 0\}, \{N_2(t) : t \geq 0\}$, and $\{N_c(t) : t \geq 0\}$ be independent NHPPs with intensities $\lambda_1(t), \lambda_2(t)$, and $\lambda_c(t)$, respectively. Suppose all arrival times $\{T_{j_\alpha}\}_{j_\alpha=1}^{N_\alpha(t)}, \alpha = 1, 2; c$ are independent. Let all service times $\{W_{j_\alpha}\}_{j_\alpha=1}^{N_\alpha(t)}, \alpha = 1, 2; c$ be i.i.d. with pdf $g_W(w)$ and distribution $G_w(w)$ and mutually independent of all arrival times. Let all stresses $\{\mathcal{H}_{j_\alpha}\}_{j_\alpha=1}^{N_\alpha(t)}, \alpha = 1, 2; c$ be i.i.d. with distribution $\mathcal{H}$ as given in Theorem 2.1, and mutually independent of arrival times and service times. Suppose we have a system of two components ($Q_{1,c}$ and $Q_{2,c}$) arranged logically in series, where each component has a arrival process $\mathcal{N}_i(t) = N_i(t) + N_c(t), i = 1, 2$. Then the survival function of the system $S_{Y_s}(t)$ is given by*

$$\begin{aligned}
S_{Y_s}(t) = \bar{F}_{0_1}(t)\bar{F}_{0_2}(t)\exp&\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t (m_1(t-w) + m_2(t-w))\exp(-\mathcal{H}w)\bar{G}_W(w)dw\right]\right) \\
&\times \exp\left(-2E_{\mathcal{H}}\left[\mathcal{H}\int_0^t \exp(-2\mathcal{H}w)m_c(t-w)\bar{G}_w(w)dw\right]\right) \qquad (5.5)
\end{aligned}$$

Proof. The proof is analogous to the proof of Theorem 2.1. In particular, the conditional survival function is given by

$$\begin{aligned}
S_{Y_s}(t) &= E\left[E\left[P\left(Y_s > t \,\middle|\, N_z(t), \{T_{j_c}\}_{j_c=1}^{N_z(t)}, \{W_{j_c}\}_{j_c=1}^{N_z(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_z(t)}\right)\right] \middle| N_c(t), \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right] \\
&= \bar{F}_{0_1}(t)\bar{F}_{0_2}(t)\exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t (m_1(t-w) + m_2(t-w))\exp(-\mathcal{H}w)\bar{G}_W(w)dw\right]\right) \\
&\times E\left[E\left[\exp\left(-2\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c}\min(W_{j_c}, t - T_{j_c})\right) \middle| N_c(t), \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right]\right]
\end{aligned}$$

To obtain the survival function, the above expectation is obtained in a similar manner as in the proof of Theorem 2.2, replacing the $\mathcal{H}_{j_c}$ with $2\mathcal{H}_{j_c}$ to immediately see the given result. □

Parallel System

Now suppose the same conditions (1)-(4) are retained but the network is in parallel rather than series. In this case, the system fails only if both components fail. Thus, conditioning on the entire

$\{N_c\}$ process, both $Q_{1,c}$ and $Q_{2,c}$ are now independent, and

$$\begin{aligned} P\left(Y_s < t \,\middle|\, N_c(t), \{T_{z_j}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right) \\ = \prod_{\ell=1}^{2} P\left(Y_{\ell,c} < t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right) \end{aligned}$$

Hence, the conditional survival function of the parallel system is given by

$$\begin{aligned} P\left(Y_s > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right) \\ = 1 - P\left(Y_s < t \,\middle|\, N_c(t), \{T_{z_j}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right) \\ = \sum_{\ell=1}^{2} P\left(Y_{\ell,c} > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right) \\ - \prod_{\ell=1}^{2} P\left(Y_{\ell,c} > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right) \end{aligned}$$

Using Lemma A.10(Appendix A),

$$\begin{aligned} &P\left(Y_s > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right) \\ &= \bar{F}_{0_1}(t) \exp\left(-\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c} \right) \exp\left(-E_{\mathcal{H}}[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m_1(t-w) \bar{G}_W(w) dw] \right) \\ &+ \bar{F}_{0_2}(t) \exp\left(-\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c} \right) \exp\left(-E_{\mathcal{H}}[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m_2(t-w) \bar{G}_W(w) dw] \right) \\ &+ \bar{F}_{0_1}(t)\bar{F}_{0_2}(t) \exp\left(-2\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c} \right) \\ &\qquad \times \exp\left(-E_{\mathcal{H}}[\mathcal{H} \int_0^t (m_1(t-w) + m_2(t-w)) e^{-\mathcal{H}w} \bar{G}_W(w) dw] \right) \end{aligned} \tag{5.6}$$

Now we may find $S_{Y_S}(t)$ for the parallel system in the same manner as the series system. Then, denoting $f_{\mathcal{H}}^{r,q}(t) = q\mathcal{H} \int_0^t e^{-q\mathcal{H}w} m_r(t-w) \bar{G}_w(w) dw$, where $r = \{1, ..., K, c\}$ and $q \in \mathbb{N}$.

$$\begin{aligned} P(Y_s > t) = \bar{F}_{0_1}(t) \exp\left(-E_{\mathcal{H}}[f_{\mathcal{H}}^{1,1}(t) + f_{\mathcal{H}}^{c,1}(t)] \right) + \bar{F}_{0_2} \exp\left(-E_{\mathcal{H}}[f_{\mathcal{H}}^{2,1}(t) + f_{\mathcal{H}}^{c,1}(t)] \right) \\ - \bar{F}_{0_1}(t)\bar{F}_{0_2}(t) \exp\left(-E_{\mathcal{H}}[f_{\mathcal{H}}^{1,1}(t) + f_{\mathcal{H}}^{2,1}(t)] \right) \exp\left(-E_{\mathcal{H}}\left[f_{\mathcal{H}}^{c,2}(t) \right] \right) \end{aligned} \tag{5.7}$$

Generalization to Systems of K Components

The systems in the prevous subsections are generalized to K components, all with the same correlator process N_c. Thus we now have components $Q_{1,c}, ..., Q_{K,c}$ with arrival processes $\mathcal{N}_{\ell,c}(t) = N_\ell(t) + N_c(t), \ell = 1, ..., K$.

Survival Function of Series System with K Correlated Components

The conditional survival function for a system with K correlated nodes, all correlated by the same process $N_c(t)$ is a straightforward generalization of (5.4) and is given by

$$
\begin{aligned}
&P\left(Y_s > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right) \\
&\quad = \left(\prod_{\ell=1}^{K} \bar{F}_{\ell,c}\right) \exp\left(-K \sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c})\right) \\
&\quad \times exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t \left(\sum_{\ell=1}^{K} m_\ell(t-w)\right) \exp(-\mathcal{H}w)\bar{G}_W(w)dw\right]\right)
\end{aligned} \tag{5.8}
$$

and thus the unconditional survival function for a series system of K correlated components is given by

$$
\begin{aligned}
S_{Y_s}(t) = &\left(\prod_{\ell=1}^{K} \bar{F}_{\ell,c}\right) \exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t \left(\sum_{\ell=1}^{K} m_\ell(t-w)\right) \exp(-\mathcal{H}w)\bar{G}_W(w)dw\right]\right) \\
&\times \exp\left(-KE_{\mathcal{H}}\left[\mathcal{H}\int_0^t \exp(-K\mathcal{H}w) m_c(t-w)\bar{G}_w(w)dw\right]\right)
\end{aligned} \tag{5.9}
$$

Survival Function of Parallel System with K Correlated Components

For a system with K components in parallel, note again that the systems fails if and only if every component fails. Let $\xi_{\ell,c}(t) = P\left(Y_{\ell,c} > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right)$. Then the conditional survival function for the K-component parallel system is given by

$$
P\left(Y_S > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right) = 1 - \prod_{\ell=1}^{K}(1 - \xi_{\ell,c}(t))
$$

Denote $\xi = (\xi_{1,c}, ..., \xi_{K,c})$, $\mathbf{1}$ as the $K-$tuple of 1's. Then

$$
\begin{aligned}
1 - \prod_{\ell=1}^{K}(1 - \xi_{\ell,c}(t)) &= 1 - \sum_{\nu \le \mathbf{1}} \binom{\mathbf{1}}{\nu} (-1)^{\mathbf{1}-\nu} \xi^{\mathbf{1}-\nu} \\
&= 1 - \sum_{s_1=0}^{1} \cdots \sum_{s_K=0}^{1} (-1)^{1-s_1} \xi_{1,c}^{1-s_1} \cdots (-1)^{1-s_K} \xi_{K,c}^{1-s_K}
\end{aligned} \tag{5.10}
$$

Let $\mathcal{S} = \{s = (s_1, ..., s_K) : s_\ell = 0, 1\}$ denote all possible combinations of s in the terms of (5.10). Let $\mathcal{L}_\sigma = \{\ell : s_\ell = 0 \text{ in } s_\sigma, s_\sigma \in \mathcal{S}\}$. Then we may express (5.10) in the following way:

$$
1 - \prod_{\ell=1}^{K}(1 - \xi_{\ell,c}(t)) = 1 - \sum_{s \in \mathcal{S}} (-\mathbf{1})^s \xi^s \tag{5.11}
$$

Using Lemma A.10,

$$\xi^s = \left(\prod_{\ell\in\mathcal{L}_\sigma} \bar{F}_{0_{\ell,c}}(t)\right) \exp\left(-E_{\mathcal{H}}\left[\mathcal{H}\int_0^t \left(\sum_{\mathcal{L}_\sigma} m_{\ell(t-w)}\right) e^{-\mathcal{H}w}\bar{G}_W(w)dw\right]\right) \times \exp\left(-|\mathcal{L}_\sigma|\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c}\min(W_{j_c}, t-T_{j_c})\right) \tag{5.12}$$

Thus, in a similar fashion to Theorem 5.1, the survival function for a system of K correlated components in parallel is given by

$$\begin{aligned} S_{Y_s}(t) &= 1 - \sum_{s\in\mathcal{S}}(-1)^s E[\xi^s] \\ &= 1 - \sum_{s\in\mathcal{S}}(-1)^s \left(\prod_{\ell\in\mathcal{L}_\sigma} \bar{F}_{0_{\ell,c}}(t)\right) \exp\left(-E_{\mathcal{H}}\left[\int_0^t \bar{G}_W(w)\left(\mathcal{H}e^{-\mathcal{H}w}\sum_{\ell\in\mathcal{L}_\sigma}[m_\ell(t-w)]\right.\right.\right. \\ &\qquad \left.\left.\left. + \ |\mathcal{L}_\sigma|\mathcal{H}e^{-|\mathcal{L}_\sigma|\mathcal{H}w} m_c(t-w)\right)\right]\right) \end{aligned} \tag{5.13}$$

Selected Additional System Architectures and a Generalized Method for Obtaining System Survival Functions

This section extends the same principle of multiple nodes with one correlator process to other selected logical system architectures. As a brief example, the structure function of the series system with K components is given by $\phi_{\text{series}}(\mathbf{x}) = \prod_{\ell=1}^K x_\ell$. Replacing the binary x_ℓ by the conditional survival function

$$P\left(Y_{\ell,c} > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}\right)$$

for each component given by Lemma A.10, the conditional survival function for the series system given in (5.8) is completely analogous to the structure function ϕ.

The structure function for a parallel system is given by $\phi_{\text{parallel}}(\mathbf{x}) = 1 - \prod_{\ell=1}^K(1 - x_\ell)$ which may be expanded into the form of (5.10). Thus, using similar logic, the conditional survival function of a parallel system is analogous to the structure function of a parallel system. Therefore, for both a series and parallel system, the binary state variable x_ℓ and the conditional survival function for node ℓ may be viewed to be in a one-to-one correspondence of sorts. Thus, the system survival function is isomorphic to the system structure function for both series and parallel systems. Since every logical system architecture can be written either as a series system comprised of parallel subsystems or a parallel system comprised of series subsystems, we only need the structure function expressed as a linear combination of powers of x_ℓ, $\ell = 1, ..., K$ in order to obtain the system survival function.

In the subsequent subsections, a selection of other logical system architectures provides examples illustrating the above method.

Bridge System

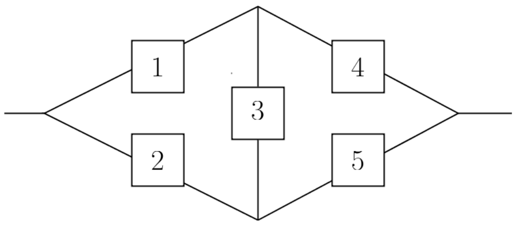


FIGURE 5.2: Block Diagram of a Bridge Structure

Figure 5.2 gives the logical block diagram for a system with a bridge-style reliability. Some communications networks may use a bridge system when there are alternative ways of connecting devices such as telephones or computers. The bridge system provides many possible ways to "complete the circuit" for relatively few components compared to a parallel system with higher reliability than a series system.

As before, each node still has its own arrival process; thus the diagram given above is logical and reveals the various combinations of working components required for the system to work. It can be easily seen that the system survives past time t if any one of the following sets of components all survive past t:

$$\{1,3,5\} \quad \{1,4\} \quad \{2,3,4\} \quad \{2,5\}$$

We may give an equivalent block diagram of the bridge system using repeated components in Figure 5.3.

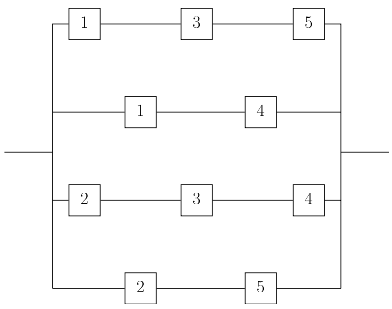


FIGURE 5.3: Alternative Representation of a Bridge Structure

Then the structure function may be easily derived using prior knowledge of series and parallel systems by breaking the above diagram into a parallel system of series subsystems. Therefore,

$$\phi(\mathbf{x}) = 1 - (1 - x_1x_3x_5)(1 - x_1x_4)(1 - x_2x_3x_4)(1 - x_2x_5)$$

Expanding the above and replacing x_ℓ by $P(Y_\ell > t|N_c(t), \{\mathcal{H}_{j_c}\}, \{T_{j_c}\}, \{W_{j_c}\})$ one may derive the conditional survival function for the bridge system. Let $\mathcal{S}_{j_c} = \sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c})$, and let $f_{\mathcal{H}}^{r,q}(t) = q\mathcal{H} \int_0^t e^{-q\mathcal{H}w} m_r(t-w)\bar{G}_w(w)dw$, where $r = \{1, ..., 5, c\}$ and $q \in \mathbb{N}$. Let $f_{\mathcal{H}}^{i_1;i_2;...;i_n}(t) = \sum_{j=1}^n f_{\mathcal{H}}^{i_j,1}(t)$, and let $\bar{F}^\alpha_{0_{m;n}}(t) = \bar{F}^\alpha_{0_m}(t)\bar{F}^\alpha_{0_n}(t)$. Then

$$\begin{aligned}
&P(Y_S > t|N_c(t), \{\mathcal{H}_{j_c}\}, \{T_{j_c}\}, \{W_{j_c}\}) \\
&= e^{-2\mathcal{S}_{j_c}} \left[\bar{F}_{0_{1;4}}(t) \exp\left(-E_{\mathcal{H}}\left[f_{\mathcal{H}}^{1;4}(t)\right]\right) + \bar{F}_{0_{2;5}}(t) \exp\left(-E_{\mathcal{H}}\left[f_{\mathcal{H}}^{2;5}(t)\right]\right)\right] \\
&\quad + e^{-3\mathcal{S}_{j_c}} \left[\bar{F}_{0_{1;3;5}}(t) \exp\left(-E_{\mathcal{H}}\left[f_{\mathcal{H}}^{1;3;5}(t)\right]\right) + \bar{F}_{0_{2;3;4}}(t) \exp\left(-E_{\mathcal{H}}\left[f_{\mathcal{H}}^{2;3;4}(t)\right]\right)\right] \\
&\quad - e^{-4\mathcal{S}_{j_c}} \left[\bar{F}_{0_{1;2;4;5}}(t) \exp\left(-E_{\mathcal{H}}\left[f_{\mathcal{H}}^{1;2;4;5}(t)\right]\right)\right] \\
&\quad - e^{-5\mathcal{S}_{j_c}} \left[\bar{F}_{0_{1;3;5}}(t) \exp\left(-E_{\mathcal{H}}\left[f_{\mathcal{H}}^{1;3;5}(t)\right]\right)\right. \\
&\qquad\qquad + \bar{F}^2_{0_1}(t)\bar{F}_{0_{3;4;5}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{1,1}(t) + f_{\mathcal{H}}^{3;4;5}(t)\right]\right) \\
&\qquad\qquad + \bar{F}^2_{0_2}(t)\bar{F}_{0_{3;4;5}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{2,1}(t) + f_{\mathcal{H}}^{3;4;5}(t)\right]\right) \\
&\qquad\qquad + \bar{F}^2_{0_4}(t)\bar{F}_{0_{1;2;3}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{4,1}(t) + f_{\mathcal{H}}^{1;2;3}(t)\right]\right) \\
&\qquad\qquad \left.+ \bar{F}^2_{0_5}(t)\bar{F}_{0_{1;2;3}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{5,1}(t) + f_{\mathcal{H}}^{1;2;3}(t)\right]\right)\right] \\
&\quad - e^{-6\mathcal{S}_{j_c}} \left[\bar{F}^2_{0_3}(t)\bar{F}_{0_{1;2;4;5}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{3,1}(t) + f_{\mathcal{H}}^{1;2;4;5}(t)\right]\right)\right] \\
&\quad + e^{-7\mathcal{S}_{j_c}} \left[\bar{F}^2_{0_{1;5}}(t)\bar{F}_{0_{2;3;4}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{1;5}(t) + f_{\mathcal{H}}^{2;3;4}(t)\right]\right)\right. \\
&\qquad\qquad + \bar{F}_{0_{1;3;5}}(t)\bar{F}^2_{0_{2;4}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{2;4;5}(t) + f_{\mathcal{H}}^{1;3}(t)\right]\right) \\
&\qquad\qquad \left.+ \bar{F}_{0_{1;4;5}}(t)\bar{F}^2_{0_{2;3}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{2;3;5}(t) + f_{\mathcal{H}}^{1;4}(t)\right]\right)\right] \\
&\quad + e^{-8\mathcal{S}_{j_c}} \left[\bar{F}^2_{0_{1;3;4}}(t)\bar{F}_{0_{2;5}}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{1;3;4;5}(t) + f_{\mathcal{H}}^{2,1}(t)\right]\right)\right] \\
&\quad - e^{-9\mathcal{S}_{j_c}} \left[\bar{F}^2_{0_{1;2;3;4}}(t)\bar{F}_{0_5}(t) \exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{1;2;3;4}(t) + f_{\mathcal{H}}^{5,1}(t)\right]\right)\right]
\end{aligned} \tag{5.14}$$

We may use the linearity of expectation to simply replace $e^{-q\mathcal{S}_{j_c}}$ by $\exp\left(E_{\mathcal{H}}\left[f_{\mathcal{H}}^{q,c}\right]\right)$ in (5.14) to obtain $S_{Y_S}(t)$ for the bridge system. While not completely tractable, the survival function is given in closed form and illustrates a general technique wherein we may use the expansion of a structure function in order to derive the conditional survival function for any system where all components are correlated by one process $\{N_c(t)\}$. This conditional survival function will always be linear in $e^{-q\mathcal{S}_{j_c}}$, and thus the unconditional survival function may be easily obtained using the already well-developed method. We illustrate this with another example.

k-of-n System

The k-of-n system has n nodes in parallel of which k must be functioning in order for the system to remain functioning. This is a generalization of the series system (n-of-n) and the parallel system (1-of-n). The structure function of a k-of-n system is given by

$$\phi(\mathbf{x}) = 1 - \prod_{\substack{\ell_j=1;\\ j=1,\dots,k;\\ \ell_1<\dots<\ell_k}}^{n} \left(1 - \prod_{j=1}^{k} x_{\ell_j}\right) \tag{5.15}$$

By replacing ℓ_j with the appropriate conditional survival function for component ℓ and expanding (5.15), we may again arrive at the system survival function conditioned upon $\{N_c(t) : t \geq 0\}$. Because this system may also be expressed as a parallel system of series subsystems, the conditional survival function will again be linear in $e^{-qS_{jc}}$ and thus the linearity of expectation allows for straightfoward computation of the system survival function.

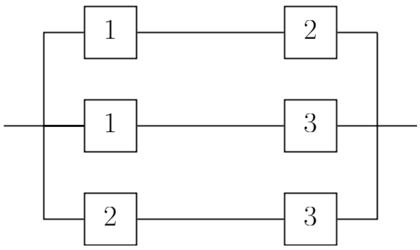


FIGURE 5.4: Block Diagram of 2-of-3 System

Example 5.1 (2-of-3 system)**.** *Figure 5.4 gives the logical block diagram for a 2-of-3 system. Thus, using* (5.15)*, the structure function for the 2-of-3 system is given by*

$$\begin{aligned}\phi(\mathbf{x}) &= 1 - (1 - x_1x_2)(1 - x_1x_3)(1 - x_2x_3)\\ &= \sum_{i=1}^{2}\sum_{j=i+1}^{3} x_ix_j - \sum_{\substack{i=1\\ j,k\neq i}}^{3} x_i^2x_jx_k + \prod_{i=1}^{3} x_i^2\end{aligned} \tag{5.16}$$

Using the technique described in Section 5.1, one may arrive at the unconditional survival function of the 2-of-3 system:

$$
\begin{aligned}
S_{Y_S}(t) = \exp\left(-E_{\mathcal{H}}[f_{\mathcal{H}}^{c,2}(t)]\right) &\left[\sum_{i=1}^{2}\sum_{j=i+1}^{3} \bar{F}_{0_i}(t)\bar{F}_{0_j}(t)\exp\left(-E_{\mathcal{H}}[f_{\mathcal{H}}^{i,1}(t)+f_{\mathcal{H}}^{j,1}(t)]\right)\right] \\
&\times \exp\left(-E_{\mathcal{H}}[f_{\mathcal{H}}^{c,4}(t)]\right)\left[\sum_{\substack{i=1\\ j,k\neq i}}^{3} \bar{F}_{0_i}^{2}(t)\bar{F}_{0_j}(t)\bar{F}_{0_k}(t)\exp\left(-E_{\mathcal{H}}\left[2f_{\mathcal{H}}^{i,1}(t)+f_{\mathcal{H}}^{j,1}(t)+f_{\mathcal{H}}^{k,1}(t)\right]\right)\right] \\
&+ \exp\left(-E_{\mathcal{H}}[f_{\mathcal{H}}^{c,6}(t)]\right)\left(\prod_{\ell=1}^{3}\bar{F}_{0_\ell}^{2}(t)\right)\exp\left(-E_{\mathcal{H}}\left[2\sum_{\ell=1}^{3} f_{\mathcal{H}}^{\ell,1}(t)\right]\right) \qquad (5.17)
\end{aligned}
$$

Chapter 6

Conclusion and Future Research

The growth of dynamic and complex servers and the associated networks necessitates general and adaptable reliability models for a variety of business concerns, including failure prediction, resource allocation, network architecture, and control policy design and implementation. Cha and Lee [9] created a fairly general dynamic model that accounted for nonhomogenous arrival times and any generic service time distribution. This work generalized [9] in several ways. In Chapter 2, the constant stress assumption was relaxed, and a model was created in which customers to a server can stress the server according to a distribution $\mathcal{H}$ with finite expectation. Cha and Lee also defined a performance measure of the server, ψ, called the server efficiency. ψ measured the long-term average number of jobs completed during a server renewal cycle, where a renewal cycle is defined as a server lifetime plus the reboot time after a crash. They provided a numerical example of ψ under constant stress η, constant intensity λ, and Rayleigh service time distribution. In this scenario, ψ had a global maximum at λ^*. Chapter 3 studied the efficiency ψ further, and explored the effect of service time distribution on the efficiency for the uniform, compact support, Erlang, and exponential classes. Sufficient conditions were derived in all but the exponential case that guaranteed the existence of a maximum ψ. The conditions derived are quite restrictive, in that failing to meet the condition does not imply ψ lacks a maximum. Numerical examples were given that illustrate this.

In addition, it was proven that for $g_W(w) = \text{Uniform}[0, b], b > 0$, and under certain conditions on the exponential class, ψ increases without bound. This is antithetical to actual server behavior; the contradiction likely stems from $g_W(w)$ having positive mass at 0. This would imply a job can be serviced in exactly 0 time with positive probability, which is impossible, as is a server with monotonically increasing efficiency for ever-increasing arrival rates. This contradictory behavior lends legitimacy to the model in that it behaves as it should under appropriate service time distributions, and gives nonsensical results for inappropriate distributions. Since the exponential distribution in particular is commonly used to model service times to due favorable mathematical properties such as memorylessness, the analysis in Chapter 3 provides a powerful justification for the abandonment of its use as a service distribution. Future research in this area would include the strengthening of sufficient conditions for the existence of a maximum in ψ, in addition to proving the postulation that ψ will never have a maximum if $g_W(w)$ has any positive mass at 0.

Chapter 4 considered additional extensions to the RSBR model from Chapter 2 to two different multi-channel cases. First, since many applications (particularly retail and logistics), experience load-balancing allocation to queues or channels, whether forced or natural, we presented the conditional survival function of a server under this allocation scheme. This allocation scheme causes the set of service intervals $\{I_j = [T_j, T_j + W_j]\}$ and the total server stress to become completely dependent. Thus, the breakdown rate process has full memory and lacks any level of independence. Future research would include deriving upper and lower estimates for the survival function of the server under a load-balancing allocation scheme.

Another extension considered in Chapter 4 was a clustered-task multichannel server in which customers arrive to a server and select 0 to N possible tasks. The selection of each task j is a Bernoulli random variable, and the set of Bernoulli random variables $\{\varepsilon j : j = 1, ..., N\}$ may or may not be correlated. Both cases were considered, and the survival function given for both. In the case of uncorrelated Bernoulli random variables, the number of possible tasks selected by a customer is a standard $Bin(N, p)$ distribution, and the corresponding server survival function has a pleasantly compact form. The correlated case was also considered wherein the Bernoulli random variables were constructed via the method in [14], and the corresponding survival function was presented.

Chapter 5 described the construction of networks of RSBR servers when the arrival processes to each server (or node) are correlated by a correlator process $\{N_c(t) : t \geq 0\}$. The conditional survival function for a series system and parallel system of K components was derived; since every system may be written as a parallel system of series subsystem with repeated components, and the conditional survival function of the system is isomorphic to the structure function of the system, the conditional survival function of any system may be determined in a straightforward manner. In addition, due to the conditionally independent nature of the nodes, the conditional survival function is linear in the correlator process and thus the survival function of any system may be obtained using the linearity of expectation and the strategy in the proof of Theorem 2.1.

Future research in the direction of Chapter 5 will include the addition of multiple correlator processes, creating a “correlation topology" in addition to the logical topology. The interaction between such topologies may prove not only mathematically interesting but may also aid in the design and modeling of highly sophisticated dynamic networks.

In addition, one can look at a hybrid stochastic server in which the stress distribution $\mathcal{H}$ is nonconstant in time. This approach would encompass many situations in which the workload itself, regardless of any change in arrival traffic, can vary with time. One possible scenario would be a warehouse with retail inventory. Orders for shipment would be the analogue to server requests, and the stress to the server (the warehouse in this case) would be the items and quantities ordered. Holiday seasons may cause a temporary shift in the order size, and thus the entire distribution of $\mathcal{H}$ may change seasonally.

A possible extension to the clustered-task mulichannel server with correlated tasks detailed in Chapter 4 would be to consider a nonconstant δ. The dependency coefficient in [14] is constant as the binary tree is built, but can be generalized to be a function of the level in the tree. As a contrived example, consider a group of people debating a joint restaurant location. The 6th person's response to a particular choice will be much more influenced by the previous answers than the 2nd. The reliability of a clustered-task multichannel server will likely change significantly under this even more generalized binomial distribution.

Appendix A

Auxiliary Lemmas

A.1 Chapter 2

Lemma A.1 (Conditional Joint Distribution of Arrival Times). *Let $\{N(t)\}$ be a nonhomgenous Poisson process describing the arrivals of client requests to the web server, and let $T_1, ...T_n$ be the arrival times of the client requests. Then, given $N(t) = n$, the conditional joint distribution of $T_1, ...T_n$, denoted $f_{T_1,...T_n|N(t)=n}(t_1, ...t_n)$ has distribution equal to the joint distribution of the order statistics $T'_{[1]}, ...T'_{[n]}$, where $T'_1, ...T'_n$ are i.i.d. with pdf $f(x) = \frac{\lambda(x)}{m(t)}$. The pdf is given by*

$$f_{T_1,...T_n|N(t)=n}(t_1, ...t_n|n) = \frac{n!}{m(t)} \prod_{i=1}^{n} \lambda(t_i), \qquad 0 \le t_1 \le ... \le t_n \le t$$

Proof. Let $N(t) = n$, and let $0 < t_1 < ... < t_n < t$. Let $h_i, i = 1, .., n$ be small enough such that $t_i + h_i < t_{i+1} \ \forall \ i = 1, .., n-1$. Denote A_i as the event that the server sees exactly 1 arrival in $[t_i, t_i + h_i]$, and B be the event that no events arrive outside the set
$U := [0, t_1] \cup [t_1, t_1 + h_1] \cup [t_2, t_2 + h_2] \cup ... \cup [t_n, t_n + h_n] \cup [t_n + h_n, t]$. Then

$$P(t_i \le T_i \le t_i + h_i, i = 1, ..., n|N(t) = n) = \frac{P(A_1 \cap A_2 \cap ... \cap A_n \cap B)}{P(N(t) = n)}$$

From (1.1) in the preliminaries, $P(N(t) = n) = \frac{e^{-m(t)} m(t)^n}{n!}$. For each $i = 1, .., n$

$$\begin{aligned} P(t_i \le T_i \le t_i + h_i) &= P(N(t_i + h_i) - N(t_i) = 1) \\ &= e^{-(m(t_i+h_i)-m(t_i))}(m(t_i + h_i) - m(t_i)) \end{aligned}$$

The next step is the calculation of $P(B)$. The complement of the set U can be broken into the disjoint intervals $[0, t_1], [t_1 + h_1, t_2], ...[t_i + h_i, t_{i+1}], ...[t_n + h_n, t]$ By items (1) - (4) in the definition of a NHPP,

$$P(B) = P[N(t) - N(t_n + h_n) = 0]P[N(t_1) - N(0) = 0] \prod_{i=1}^{n-1} P[N(t_{i+1}) - N(t_i + h_i) = 0]$$
$$= e^{-m(t_1)} e^{-(m(t) - m(t_n + h_n))} \prod_{i=1}^{n-1} e^{-(m(t_{i+1}) - m(t_i + h_i))}$$
$$= \exp\left(-\left[m(t) + \sum_{i=1}^{n} m(t_i) - \sum_{i=1}^{n} m(t_i + h_i)\right]\right)$$

Now, again using (2) in the definition, and simplifying,

$$P(t_i \le T_i \le t_i + h_i, i = 1, ..., n | N(t) = n) = \prod_{i=1}^{n} P(t_i \le T_i \le t_i + h_i)$$
$$= n! \prod_{i=1}^{n} \frac{m(t_i + h_i) - m(t_i)}{m(t)}$$

Letting $h_i \to 0 \; \forall \; i$,

$$f_{T_1, ... T_n | N(t) = n}(t_1, ... t_n | n) = n! \prod_{i=1}^{n} \frac{\lambda(t_i)}{m(t)}$$

□

Lemma A.2 (Expectation of $e^{-\eta_{i_{j'}} \min(W_j, t - T_{j'})}$**).**

$$E\left[E\left[e^{-\eta_{i_{j'}} \min(W_j, t - T'_j)} \,\middle|\, W_j\right]\right] = \frac{1}{m(t)}\left(m(t) - \eta_{i_{j'}} \int_0^t e^{-\eta_{i_{j'}} w} m(t - w) \bar{G}_W(w) dw\right)$$

Proof. Two cases must be considered: (1) $w \le t$ and (2) $w > t$. For $w \le t$,

$$E\left[e^{-\eta_{i_j} \min(W_j, t - T'_j)} | W_j = w\right] = \int_0^{t-w} e^{-\eta_{i_j} w} \frac{\lambda(x)}{m(t)} dx + \int_{t-w}^{t} e^{-\eta_{i_j}(t - x)} \frac{\lambda(x)}{m(t)} dx$$
$$= e^{-\eta_{i_j} w} \frac{m(t - w)}{m(t)} + e^{-\eta_{i_j} t} \int_0^t e^{\eta_{i_j} x} \frac{\lambda(x)}{m(t)} dx$$

For $w > t$,

$$E\left[e^{-\eta_{i_j} \min(W_j, t - T'_j)} | W_j = w\right] = e^{-\eta_{i_j} t} \int_0^t e^{\eta_{i_j} x} \frac{\lambda(x)}{m(t)} dx$$

Therefore,

$$\begin{aligned}
E_W\left[e^{-\eta_{i_j}\min(W_j,t-T'_j)}\right] &= E_W\left[E\left[e^{-\eta_{i_j}\min(W_j,t-T'_j)}\middle| W_j = w\right]\right] \\
&= \frac{1}{m(t)}\left(\int_0^t e^{-\eta_{i_j}w} m(t-w) g_W(w)dw\right. \\
&+ e^{-\eta_{i_j}t}\int_0^t\int_{t-w}^t e^{\eta_{i_j}x}\lambda(x)dx g_W(w)dw \qquad (*) \\
&\left.+ \bar{G}_W(t)e^{-\eta_{i_j}t}\int_0^t e^{\eta_{i_j}x}\lambda(x)dx\right)
\end{aligned}$$

Focusing on $(*)$, we make the change of variables $w = t - x$ and change the order of intergration, yielding a new second term.

$$\begin{aligned}
&= \frac{1}{m(t)}(\int_0^t e^{-\eta_{i_j}w} m(t-w) g_W(w)dw \\
&+ e^{-\eta_{i_j}t}\int_0^t e^{\eta_{i_j}x}\lambda(x)\int_{t-x}^t g_W(w)dwdx \\
&+ \bar{G}_W(t)e^{-\eta_{i_j}t}\int_0^t e^{\eta_{i_j}x}\lambda(x)dx)
\end{aligned}$$

Combining the second and third terms:

$$\begin{aligned}
&= \frac{1}{m(t)}\left(\int_0^t e^{-\eta_{i_j}w} m(t-w) g_W(w)dw\right. \\
&\left.+ e^{-\eta_{i_j}t}\int_0^t e^{\eta_{i_j}x}\lambda(x)\bar{G}_w(t-x)dx\right)
\end{aligned}$$

Changing variables again in the second term, using $w = t - x$, we get

$$\begin{aligned}
&= \frac{1}{m(t)}\left(\int_0^t e^{-\eta_{i_j}w} m(t-w) g_W(w)dw\right. \\
&\left.+ \int_0^t e^{-\eta_{i_j}w}\lambda(t-w)\bar{G}_w(w)dw\right)
\end{aligned}$$

Integrating the first term by parts, we get

$$\begin{aligned}
&= \frac{1}{m(t)}\left(\left[-e^{-\eta_{i_j}w} m(t-w)\bar{G}_w(w)\right]\Big|_0^t\right. \\
&- \int_0^t (\eta_{i_j}e^{-\eta_{i_j}w} m(t-w) + e^{-\eta w}\lambda(t-w))\bar{G}_W(w)dw \\
&\left.+ \int_0^t e^{-\eta_{i_j}w}\lambda(t-w)\bar{G}_w(w)dw\right) \\
&= \frac{1}{m(t)}\left(m(t) - \eta_{i_j}\int_0^t e^{-\eta_{i_j}w} m(t-w)\bar{G}_W(w)dw\right)
\end{aligned}$$

□

Lemma A.3 (Conditional distribution of Renewal Cycle Length)**.** *The conditional distribution $f_{Y|\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(t|\mathfrak{r},\mathfrak{v},\mathfrak{h},n)$ is given by*

$$f_{Y|\mathfrak{R},\mathfrak{V},\mathfrak{H},N}(t|\mathfrak{r},\mathfrak{v},\mathfrak{h},n) = \exp\left(-\int_0^t r_0(s)ds - \sum_{j=1}^{n} \eta_{i_j} \min(v_j, t-r_j)\right) \times \left(r_0(t) + \sum_{j=1}^{n} \eta_{i_j} \mathbb{1}(v_j > t - r_j)\right)$$

Proof. Denote the condition $C = \{\mathfrak{R} = \mathfrak{r}, \mathfrak{V} = \mathfrak{v}, \mathfrak{H} = \mathfrak{h}, N(t) = n\}$. Then

$$f_{Y|C}(t|c) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left(P(Y > t|C = c) - P(Y > t + \Delta t|C = c)\right) \tag{A.1}$$

From (2.2),

$$P(Y > t|C = c) = \exp\left(-\int_0^t r_0(s)ds - \sum_{j=1}^{n} \eta_j^{i_j} \min(v_j, t-r_j)\right)$$

Recall that $\mathcal{B}(s) = r_0(s) + \sum_{j=1}^{N(t)} \mathbb{1}(R_j \leq s \leq R_j + V_j)$. We now derive $P(Y > t + \Delta t|C = c)$.

Let $A_k = \{N(t + \Delta t) - N(t) = k\}$ be the event such that k requests arrived between t and $t + \Delta t$. From the definition of a nonhomogenous Poisson process [23],

$$\begin{aligned} P(A_0) &= 1 - \lambda(t)\Delta t + o(\Delta t) = e^{-(m(t+\Delta t)-m(t))} \\ P(A_1) &= \lambda(t)\Delta t + o(\Delta t) = (m(t+\Delta t) - m(t))e^{-(m(t+\Delta t)-m(t))} \\ &\vdots \\ P(A_k) &= \frac{(m(t+\Delta t) - m(t))^k}{k!} e^{-(m(t+\Delta t)-m(t))} \end{aligned}$$

for $k \geq 2$. By the Law of Total Probability,

$$P(Y > t + \Delta t|C = c) = \sum_{k=0}^{\infty} P(Y > t + \Delta t|C \cap A_k)P(A_k) \tag{A.2}$$

We look at each fixed k and show that the terms for $k \geq 2$ are of $o(\Delta k)$. For $k = 0$,

$$\begin{aligned} P(Y > t + \Delta t|C \cap A_0)P(A_0) &= e^{-\int_0^{t+\Delta t} \mathcal{B}(s)ds}(1 - \lambda(t)\Delta t + o(\Delta t)) \\ &= e^{-\int_0^t \mathcal{B}(s)ds} e^{-\int_t^{t+\Delta t} \mathcal{B}(s)ds}(1 - \lambda(t)\Delta t + o(\Delta t)) \end{aligned} \tag{A.3}$$

$$\int_t^{t+\Delta t} \mathcal{B}(s)ds = \int_t^{t+\Delta t} r_0(s)ds + \left(\sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}(v_j > t - r_j) \right) \Delta t$$
$$= r_0(t) + o(\Delta t) + \left(\sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}(v_j > t - r_j) \right) \Delta t$$

Using the fact that $e^{-x+o(x)} = 1 - x + o(x)$ for small x,

$$e^{-\int_t^{t+\Delta t} \mathcal{B}(s)ds} = 1 - \left(r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}(v_j > t - r_j) \right) \Delta t + o(\Delta t)$$

Substituting into (A.3),

$$P(Y > t + \Delta t | C \cap A_0) P(A_0) = \left[e^{-\int_0^t \mathcal{B}(s)ds} 1 - \left(r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}(v_j > t - r_j) \right) \Delta t + o(\Delta t) \right]$$
$$[1 - \lambda(t)\Delta t + o(\Delta t)]$$
$$= e^{-\int_0^t \mathcal{B}(s)ds} \left(1 - [r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}_{v_j > t - r_j}]\Delta t - \lambda(t)\Delta t + o(\Delta t) \right) \tag{A.4}$$

Now, for $k = 1$, we must contend with the arrival of one request in $[t, t + \Delta t]$ in addition to the n fixed arrivals given by C. Call this arrival t_1, with corresponding time to completion w_1. We then have two cases:

(1) $t_1 + w_1 < t + \Delta t$, that is, the request arrives and is serviced before $t + \Delta t$, or

(2) $t_1 + w_1 > t + \Delta t$. In this case, the service time is greater than Δt.

With both of these cases, the failure rate increases by an $\eta_{t_1}^{n_{i_{t_1}}}$ for a time smaller than Δt, which we will denote as $\Delta^1(t)$. Then

$$\Delta^1(t) = \begin{cases} w_1, & t_1 + w_1 < t + \Delta t \\ t + \Delta t - t_1, & t_1 + w_1 > t + \Delta t \end{cases}$$

In what follows we apply previous arguments to calculate $P(Y > t + \Delta t | C \cap A_1) P(A_1)$. Now,

$$\exp\left(-\int_t^{t+\Delta t} \mathcal{B}(s)ds \right) = \exp\left(-r_0(t) + \left[\sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}(v_j > t - r_j) \right] \Delta t + \eta_{t_1}^{i_{t_1}} \Delta^1 t + o(\Delta t) \right)$$

The above simplifies to

$$\exp\left(-\int_t^{t+\Delta t} \mathcal{B}(s)ds \right) = 1 - [r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}(v_j > t - r_j)]\Delta t + \eta_{t_1}^{i_{t_1}} \Delta^1 t + o(\Delta t)$$

Now,

$$
\begin{aligned}
P(Y > t + \Delta t | C \cap A_1) P(A_1) &= e^{-\int_0^t \mathcal{B}(s) ds} \left(\left[1 - \left[r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}(v_j > t - r_j) \right] \Delta t \right. \right. \\
&\left. \left. + \eta_{t_1}^{i_{t_1}} \Delta^1 t + o(\Delta t) \right] [\lambda(t) \Delta t + o(\Delta t)] \right) \\
&= e^{-\int_0^t \mathcal{B}(s) ds} (\lambda(t) \Delta t + o(\Delta t)) \qquad \text{(A.5)}
\end{aligned}
$$

Combining (A.5) with (A.4),

$$
\sum_{k=0}^{1} P(Y > t + \Delta t | C \cap A_k) P(A_k) = e^{-\int_0^t \mathcal{B}(s) ds} \left(1 - (r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}_{v_j > t - r_j}) \Delta t \right) + o(\Delta t) \qquad \text{(A.6)}
$$

For $k \geq 2$, we will now show that the contribution to (A.2) is negligible. Using similar notation established in the case of $k = 1$,

$$
P(Y > t + \Delta t | C \cap A_k) = e^{-\int_0^t \mathcal{B}(s) ds} \left[1 - [(r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}_{v_j > t - r_j}) \Delta t + \sum_{l=1}^{k} \eta_{t_l}^{i_{t_l}} \Delta^l t] + o(\Delta t) \right]
$$

Now, we see that $\forall\, k \geq 2$,

$$
P(Y > t + \Delta t | C \cap A_k) \leq e^{-\int_0^t \mathcal{B}(s) ds}
$$

and hence

$$
\begin{aligned}
\sum_{k=2}^{\infty} P(Y > t + \Delta t | C \cap A_k) P(A_k) &\leq e^{-\int_0^t \mathcal{B}(s) ds} \sum_{k=2}^{\infty} P(A_k) \\
&= e^{-\int_0^t \mathcal{B}(s) ds} (1 - P(A_0) - P(A_1)) \\
&= e^{-\int_0^t \mathcal{B}(s) ds} o(\Delta t)
\end{aligned}
$$

Therefore,

$$
P(Y > t + \Delta t | C) = e^{-\int_0^t \mathcal{B}(s) ds} \left(1 - (r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}_{v_j > t - r_j}) \Delta t \right) + o(\Delta t)
$$

Now, we see that

$$
\frac{1}{\Delta t} (P(Y > t | C = c) - P(Y > t + \Delta t | C = c)) = e^{-\int_0^1 \mathcal{B}(s) ds} \left(r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}_{v_j > t - r_j} + \frac{o(\Delta t)}{\Delta t} \right)
$$

Then, letting $\Delta t \to 0$,

$$
f_{Y|\mathfrak{R}, \mathfrak{V}, \mathfrak{H}, N}(t | \mathfrak{r}, \mathfrak{v}, \mathfrak{h}, n) = \left\{ \exp\left(-\int_0^t r_0(s) ds - \sum_{j=1}^{n} \eta_j^{i_j} \min(v_j, t - r_j) \right) \right\} \left(r_0(t) + \sum_{j=1}^{n} \eta_j^{i_j} \mathbb{1}_{v_j > t - r_j} \right)
$$

□

Lemma A.4 (Joint PDF of $\mathfrak{R}, \mathfrak{V}, N$).

$$f_{\mathfrak{R},\mathfrak{V},N(t)}(\mathfrak{r}, \mathfrak{v}, n) = \frac{1}{n!} \exp\left(-\int_0^t \lambda(s)ds\right) \prod_{j=1}^{n} \lambda(r_j) g_W(v_j)$$

Proof. Let $\mathcal{W} = (W_1, ..., W_n)$ be the service times under the condition that $N(t) = n$, and let $\mathcal{W} = \sqsupseteq = (w_1, ..., w_n)$. Let $(r_{[1]}, ..., r_{[n]})$ be the ordered vector of $\mathfrak{r}$. There are $n!$ possible orderings of $\mathfrak{r}$. Now, $P(r_{[i]} \le R_{[i]} \le r_{[i]} + h_i)$ for some $h_i > 0, i = 1, ..., n$ is given by

$$P(r_{[i]} \le R_{[i]} \le r_{[i]} + h_i) = e^{-(m(r_{[i]}+h_i) - m(r_{[i]})}(m(r_{[i]} + h_i) - m(r_{[i]}))$$

We may see that the joint distribution of $\mathfrak{R}$ is identical to the distribution of the order statistics of $\mathfrak{R}$:

$$f_{\mathfrak{R},N(t)}(\mathfrak{r}, n) = \frac{1}{n!} \prod_{j=1}^{n} \lambda(r_j) \exp\left(-\int_0^t \lambda(s)ds\right)$$

$\mathfrak{R}, \mathfrak{V}$ are mutually independent. Therefore,

$$f_{\mathfrak{R},\mathfrak{V},N(t)}(\mathfrak{r}, \mathfrak{v}, n) = \frac{1}{n!} \exp\left(-\int)_0^t \lambda(s)ds\right) \prod_{j=1}^{n} \lambda(r_j) g_W(v_j)$$

□

A.2 Chapter 3

Lemma A.5 ($a(t)$ under $\delta-$shifted Erlang distribution).

$$a(t) = \left(\frac{\gamma}{\gamma+\eta}\right)^k e^{-\eta\delta}(t-\delta) + \frac{k\gamma^k\left(e^{-(\eta+\gamma)t+\gamma\delta} - e^{-\eta\delta}\right)}{(\gamma+\eta)^{k+1}} + \frac{\gamma^k e^{-(\eta+\gamma)t+\gamma\delta}}{(k-1)!} \sum_{j=1}^{k-1} \left[\frac{(t-\delta)^j}{(\gamma+\eta)^{k-j+1}} \left(\frac{k!}{j!} - \frac{(k-1)!}{(j-1)!}\right)\right]$$

Proof.

$$\begin{aligned} a(t) &= \int_0^t e^{-\eta v} g(v)(t-v)dv \\ &= \int_\delta^t \frac{1}{(k-1)!} e^{-(\eta+\gamma)v+\delta\gamma} \gamma^k (v-\delta)^{k-1}(t-v)dv \\ &= \frac{\gamma^k e^{\delta\gamma}}{(k-1)!} \int_\delta^t e^{-(\eta+\gamma)v}(v-\delta)^{k-1}(t-v)dv \end{aligned}$$

Let $C = (\eta + \gamma)$. Splitting the integral,

$$a(t) = \frac{\gamma^k e^{\delta\gamma}}{(k-1)!}\left[t\int_\delta^t e^{-Cv}(v-\delta)^{k-1}dv - \int_\delta^t e^{-Cv}v(v-\delta)^{k-1}dv\right]$$

Making a change of variables, let $x = v - \delta$, then $v = x + \delta$, $dx = dv$, and $0 < x < t - \delta$. Thus

$$\begin{aligned}
a(t) &= \frac{\gamma^k e^{\delta\gamma}}{(k-1)!}\left[t\int_0^{t-\delta} e^{-C(x+\delta)}x^{k-1}dx - \int_0^{t-\delta} e^{-C(x+\delta)}(x+\delta)x^{k-1}dx\right]\\
&= \frac{\gamma^k e^{\delta\gamma}}{(k-1)!}\left[[te^{-C\delta}\int_0^{t-\delta} e^{-Cx}x^{k-1}dx - e^{-C\delta}\int_0^{t-\delta} e^{-Cx}x^k dx - \delta e^{-C\delta}\int_0^{t-\delta} e^{-Cx}x^{k-1}dx\right]\\
&= \frac{\gamma^k e^{\delta\gamma}}{(k-1)!}\left[(t-\delta)e^{-C\delta}\int_0^{t-\delta} e^{-Cx}x^{k-1}dx - e^{-C\delta}\int_0^{t-\delta} e^{-Cx}x^k dx\right]
\end{aligned}$$

One may see via induction that $\int_0^w e^{-Cx}x^p dx = \frac{-e^{-Cx}}{C^{p+1}}\sum_{j=0}^p \frac{p!C^j x^j}{j!}\Big|_{x=0}^w$. Thus,

$$\begin{aligned}
(t-\delta)e^{-C\delta}\int_0^{t-\delta} e^{-Cx}x^{k-1}dx &= (t-\delta)e^{-C\delta}\left[\frac{-e^{-Cx}}{C^k}\sum_{j=0}^{k-1}\frac{(k-1)!C^j x^j}{j!}\Bigg|_{x=0}^{t-\delta}\right]\\
&= (t-\delta)e^{-C\delta}\left[\frac{(k-1)!}{C^k} - \frac{e^{-C(t-\delta)}}{C^k}\sum_{j=0}^{k-1}\frac{(k-1)!C^j(t-\delta)^j}{j!}\right]\\
&= \frac{(k-1)!e^{-C\delta}}{C^k}(t-\delta) - \frac{e^{-Ct}}{C^k}\sum_{j=0}^{k-1}\frac{(k-1)!C^j(t-\delta)^{j+1}}{j!}
\end{aligned}$$

Similarly,

$$-e^{-C\delta}\int_0^{t-\delta} e^{-Cx}x^k dx = \frac{e^{-Ct}}{C^{k+1}}\sum_{j=0}^{k}\left(\frac{k!c^j(t-\delta)^j}{j!}\right) - \frac{e^{-C\delta}k!}{C^{k+1}}$$

Thus

$$\begin{aligned}
a(t) &= \frac{\gamma^k e^{\delta\gamma}}{(k-1)!}\left[\frac{(k-1)!e^{-C\delta}}{C^k}(t-\delta) - \frac{e^{-C\delta}k!}{C^{k+1}} + e^{-Ct}\left(\frac{k!}{C^{k+1}} + \sum_{j=1}^{k-1}\left[\frac{(t-\delta)^j}{C^{k-j+1}}\left(\frac{k!}{j!} - \frac{(k-1)!}{(j-1)!}\right)\right]\right)\right]\\
&= \left(\frac{\gamma}{\gamma+\eta}\right)^k e^{-\eta\delta}(t-\delta) - \frac{ke^{-\eta\delta}\gamma^k}{(\gamma+\eta)^{k+1}} + \frac{\gamma^k e^{-(\eta+\gamma)t+\gamma\delta}}{(k-1)!}\sum_{j=1}^{k-1}\left[\frac{(t-\delta)^j}{(\gamma+\eta)^{k-j+1}}\left(\frac{k!}{j!} - \frac{(k-1)!}{(j-1)!}\right)\right]
\end{aligned}$$

□

Lemma A.6 ($b(t)$ under shifted Erlang distribution)**.**

$$b(t) = \frac{1-e^{-\delta\eta}}{\eta} + \sum_{j=0}^{k-1}\left[\frac{\gamma^j e^{-\eta\delta}}{(\gamma+\eta)^{j+1}} - \gamma^j e^{-(\eta+\gamma)t+\eta\delta}\left(\sum_{i=0}^{j}\frac{(\eta+\gamma)^i(t-\delta)^i}{i!}\right)\right]$$

Proof.

$$b(t) := \int_0^t e^{-\eta(t-r)}\bar{G}(t-r)dr$$

From (3.2),

$$\begin{aligned}\bar{G}(t-r) &= \begin{cases}1, & 0 \le t-r < \delta\\ e^{\gamma(\delta-(t-r))}\sum_{j=0}^{k-1}\frac{\gamma^j(t-r-\delta)^j}{j!}, & t-r \ge \delta\end{cases}\\ &= \begin{cases}1, & 0 \le t-\delta \le r \le t\\ e^{\gamma(\delta-(t-r))}\sum_{j=0}^{k-1}\frac{\gamma^j(t-r-\delta)^j}{j!}, & 0 \le r \le t-\delta\end{cases}\end{aligned} \tag{A.7}$$

Then

$$\begin{aligned}b(t) &= \int_0^{t-\delta} e^{\delta\gamma-(\eta+\gamma)(t-r)}\sum_{j=0}^{k-1}\frac{\gamma^j(t-r-\delta)^j}{j!}dr + \int_{t-\delta}^{t} e^{-\eta(t-r)}dr\\ &= \sum_{j=0}^{k-1}\left(\int_0^{t-\delta} e^{\delta\gamma-(\eta+\gamma)(t-r)}\frac{\gamma^j(t-r-\delta)^j}{j!}dr\right)\end{aligned} \tag{A.8}$$

Fix j. Then

$$\int_0^{t-\delta} e^{\delta\gamma-(\eta+\gamma)(t-r)}\frac{\gamma^j(t-r-\delta)^j}{j!}dr = \frac{\gamma^j e^{\delta\gamma}}{j!}\int_0^{t-\delta} e^{-(\eta+\gamma)(t-r)}(t-r-\delta)^j dr$$

Make a change of variables. Let $x = t-r$. Then $-dx = dr$, $r = t-x$, and $\delta < x < t$. Then

$$\frac{\gamma^j e^{\delta\gamma}}{j!}\int_0^{t-\delta} e^{-(\eta+\gamma)(t-r)}(t-r-\delta)^j dr = -\frac{\gamma^j e^{\delta\gamma}}{j!}\int_\delta^t e^{-(\eta+\gamma)x}(x-\delta)^j dx$$

Make one more change of variables. Let $y = x-\delta$. Then $dy = dx$, $x = y+\delta$, and $0 < y < t-\delta$. Now,

$$\begin{aligned}-\frac{\gamma^j e^{\delta\gamma}}{j!}\int_\delta^t e^{-(\eta+\gamma)x}(x-\delta)^j dx &= -\frac{\gamma^j e^{\delta\gamma-(\eta+\gamma)\delta}}{j!}\int_0^{t-\delta} e^{-(\eta+\gamma)y}y^j dy\\ &= \frac{\gamma^j e^{-\eta\gamma}}{j!}\left[\frac{e^{-(\eta+\gamma)(t-\delta)}}{(\eta+\gamma)^{j+1}}\sum_{i=0}^{j}\left(\frac{j!}{i!}(\eta+\gamma)^i(t-\delta)^i\right) - \frac{j!}{(\eta+\gamma)^{j+1}}\right]\end{aligned}$$

Then

$$b(t) = \sum_{j=0}^{k-1} \left[\frac{\gamma^j e^{-\gamma\delta}}{(\eta+\gamma)^{j+1}} - \gamma^j e^{-(\eta+\gamma)t+\gamma\delta} \sum_{i=0}^{j} \frac{(\eta+\gamma)^i (t-\delta)^i}{i!} \right] + \frac{1-e^{-\delta\eta}}{\eta}$$

□

Lemma A.7. *$b(t)$ is concave.*

Proof. This will be shown via the standard second derivative test, and is straightforward computation.

$$\ddot{b} = e^{-(\eta+\gamma)t+\gamma+\delta} \left[-(\eta+\gamma)^2 f_b(t) + 2(\eta+\gamma)\dot{f}_b - \ddot{f}_b \right] \tag{A.9}$$

where $f_b(t) = \sum_{j=0}^{k-1} \gamma^j \sum_{i=0}^{j} \frac{(\eta+\gamma)^i (t-\delta)^i}{i!}$.

$$\begin{aligned} f_b(t) &= 1 + \sum_{j=1}^{k-1} \gamma^j \sum_{i=0}^{j} \frac{(\eta+\gamma)^i (t-\delta)^i}{i!} \\ &= 1 + \gamma \sum_{i=0}^{1} \frac{(\eta+\gamma)^i (t-\delta)^i}{i!} + \gamma^2 \sum_{i=0}^{2} \frac{(\eta+\gamma)^i (t-\delta)^i}{i!} \\ &\quad + \ldots + \gamma^{k-1} \sum_{i=0}^{k-1} \frac{(\eta+\gamma)^i (t-\delta)^i}{i!} \end{aligned} \tag{A.10}$$

Next,

$$\dot{f}_b(t) = \sum_{j=1}^{k-1} \gamma^j \sum_{i=1}^{j} i(\eta+\gamma)^i (t-\delta)^{i-1} \tag{A.11}$$

and

$$\ddot{f}_b(t) = \sum_{j=2}^{k-1} \gamma^j \sum_{i=2}^{j} i(i-1)(\eta+\gamma)^i (t-\delta)^{i-2} \tag{A.12}$$

It is sufficient to show that $2(\eta+\gamma)\dot{f}_b < (\eta+\gamma)^2 f_b(t) + \ddot{f}_b(t) \ \forall\ t \in (0, \infty)$

$$\begin{aligned} &2(\eta+\gamma)\dot{f}_b(t) - (\eta+\gamma)^2 f_b(t) - \ddot{f}_b(t) \\ &\quad = \sum_{j=0}^{k-1} \gamma^j \left[\sum_{i=0}^{j} \left(2 \cdot (i+1) - \left((i+2)(i+1) + \frac{1}{i!} \right) \right) (\eta+\gamma)^{i+2} (t-\delta)^i \right] \\ &\quad \leq 0 \text{ for } t \geq \delta \end{aligned}$$

□

A.3 Chapter 4

Lemma A.8 (Decomposition of a Union of Events into Disjoint Atoms)**.** *Let $A_1, ..., A_n$ be events. Let $F = \{\emptyset, c\}, C_n = \{\underbrace{c, \ldots, c}_{n}\}$, and $E_n = (F \times^n F) \setminus C_n$, where $\times^n$ denotes the $n-$fold cross product. Let*

$\varepsilon = (\varepsilon_1, ..., \varepsilon_n) \in E_n$ denote the $n-$tuple where $\varepsilon_i \in F$.

$$\bigcup_{i=1}^{n} A_i = \bigcup_{\varepsilon \in E_n} \bigcap_{i=1}^{n} A_i^{\varepsilon_i}$$

Proof. First, it will be shown that $E_{n+1} = [(E_n \cup C_n) \times F] \setminus C_{n+1}$.

$$\begin{aligned} E_{n+1} &= (F \times^n F) \setminus C_{n+1} \\ &= ([F \times^n F] \times F) \setminus C_{n+1} \\ &= [(E_n \cup C_n) \times \{\emptyset, c\}] \setminus C_{n+1} \end{aligned}$$

Now, it remains to be shown that the tuples $\varepsilon \in E_n$ account for all atoms of $\cup_{i=1}^{n} A_i$. Consider $n = 2$. Then $E_2 = \{(\emptyset, c), (c, \emptyset), (\emptyset, \emptyset)\}$. Then

$$\begin{aligned} &(A_1 \cap A_2^c) \cup (A_1^c \cap A_2) \cup (A_1 \cap A_2) \\ &\qquad = (A_1 \setminus A_2) \cup (A_2 \setminus A_1) \cup (A_1 \cap A_2) \\ &\qquad = (A_1 \cup (A_1 \cap A_2)) \setminus (A_2 \setminus (A_1 \cap A_2)) \cup (A_2 \setminus A_1) \\ &\qquad = [A_1 \cup (A_1 \cap A_2) \cup (A_2 \setminus A_1)] \setminus [A_2 \setminus (A_1 \cap A_2) \setminus (A_2 \setminus A_1)] \\ &\qquad = (A_1 \cup A_2) \setminus \emptyset \\ &\qquad = A_1 \cup A_2 \end{aligned}$$

Now assume for $k \leq n$, $\bigcup_{i=1}^{n} A_i = \bigcup_{\varepsilon \in E_n} \bigcap_{i=1}^{n} A_i^{\varepsilon_i}$. Letting $Y = [(E_{n-1} \cup C_{n-1}) \times \{\emptyset, c\}] \setminus C_n$

$$\bigcup_{i=1}^{n} A_i = \left[\bigcup_{i=1}^{n-1} A_i \right] \bigcup A_n$$

Now, $\varepsilon_n \in \{\emptyset, c\}$. Thus,

$$\begin{aligned} \bigcup_{i=1}^{n} A_i &= \bigcup_{\varepsilon \in Y} \bigcap_{i=1}^{n} A_i^{\varepsilon_i} \\ &= \bigcup_{\varepsilon \in E_n} \bigcap_{i=1}^{n} A_i^{\varepsilon_i} \end{aligned}$$

□

A.4 Chapter 5

Lemma A.9. *Let* $\{N_i(t) : t \geq 0\}_{i=1}^n$ *be independent nonhomogeneous Poisson processes with intensities* $\lambda_i(t), i = 1, ..., n$. *Let* $N(t) = \sum_{i=1}^n N_i(t)$ *is a nonhomogeneous Poisson process with intensity* $\lambda(t) = \sum_{i=1}^n \lambda_i(t)$

Proof. The proof will proceed via induction. As a base case, let $n = 2$. It suffices to show that $N(0) = 0$, and

$$P(N(t+s) - N(t) = n) = \frac{\exp\left(-(m(t+s) - m(t))\right)[m(t+s) - m(t)]^n}{n!}$$

where $m(t) = m_1(t) + m_2(t)$ and $m_i(t) = \int_0^t \lambda_i(s)ds$. Clearly, $N(0) = N_1(0) + N_2(0) = 0 + 0 = 0$. Now, since $\{N_1(t)\}, \{N_2(t)\}$ are independent, we may find the distribution of $\{N(t)\}$ via convolution. Thus,

$$\begin{aligned}
&P(N(t+s) - N(t) = n) \\
&\quad = P((N_1 + N_2)(t+s) - (N_1 + N_2)(t)) = n) \\
&\quad = P(N_1(t+s) + N_2(t+s) - N_1(t) - N_2(t) = n) \\
&\quad = \sum_{x=0}^{n} P([N_1(t+s) - N_1(t) = x] \cap [N_2(t+s) - N_2(t) = n - x]) \\
&\quad = \sum_{x=0}^{n} P(N_1(t+s) - N_1(t) = x) P(N_2(t+s) - N_2(t) = n - x) \\
&\quad = \frac{1}{n!} e^{-(m_1+m_2)(t+s)-(m_1+m_2)(t)} (m_1(t+s) - m_1(t) + m_2(t+s) - m_2(t))^n \\
&\quad = \frac{e^{-(m_1+m_2)(t+s)-(m_1+m_2)(t)}}{n!} ((m_1 + m_2)(t+s) - (m_1 + m_2)(t))^n \\
&\quad = \frac{e^{-(\lambda_1+\lambda_2)(s)} (\lambda_1 + \lambda_2)(s))^n}{n!}
\end{aligned} \tag{A.13}$$

Now, assume that $N_\kappa(t) = \sum_{i=1}^k N_i(t)$ is a NHPP with intensity $\lambda(t) = \sum_{i=1}^k \lambda_i(t)$. Then let $N(t) = N_\kappa(t) + N_{k+1}(t)$, where $\{N_{k+1}(t)\}$ is a NHPP with intensity $\lambda_{k+1}(t)$. Then using the same procedure as above, we see that

$$P(N(t+s) - N(t) = n) = \frac{e^{-\left(\sum_{i=1}^{k+1} \lambda_i\right)(s)} \left[\left(\sum_{i=1}^{k+1} \lambda_i\right)(s)\right]^n}{n!}$$

and thus the sum of nonhomogeneous Poisson processes remains a NHPP. □

Lemma A.10. *Under the condition that* $N_c(t) = n_c, \mathcal{H}_{j_c} = \eta_{i_{j_c}}, i \in \{1, ..., m\}, j_c = 1, ..., n_c$, *we have that*

$$P\left(Y_{\ell,c} > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right)$$
$$= \bar{F}_{0_{\ell,c}} \exp\left(-\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c}) \right) \exp\left(-E_{\mathcal{H}} \left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m_\ell(t-w) \bar{G}_W(w) dw \right] \right)$$

where $m_\ell(x) = \int_0^x \lambda_\ell(s) ds$.

Proof. As in the proof of Theorem 2.2, we have that

$$P\left(Y_\ell > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)}, N_\ell(t), \{T_{j_\ell}\}_{j_\ell=1}^{N_\ell(t)}, \{W_{j_\ell}\}_{j_\ell=1}^{N_\ell(t)}, \{\mathcal{H}_{j_\ell}\}_{j_\ell=1}^{N_\ell(t)} \right)$$
$$= \exp\left(-\int_0^t \mathcal{B}_\ell(t) \right)$$
$$= \bar{F}_{0_1}(t) \exp\left(-\sum_{j_\ell=1}^{N_\ell(t)} \mathcal{H}_{j_\ell} \min(W_{j_\ell}, t - T_{j_\ell}) - \sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c}) \right)$$
$$= \bar{F}_{0_1}(t) \exp\left(-\sum_{j_\ell=1}^{N_\ell(t)} \mathcal{H}_{j_\ell} \min(W_{j_\ell}, t - T_{j_\ell}) \right) \exp\left(-\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c}) \right) \tag{A.14}$$

Now,

$$P\left(Y_{\ell,c} > t \,\middle|\, N_c(t), \{T_{j_c}\}_{j_c=1}^{N_c(t)}, \{W_{j_c}\}_{j_c=1}^{N_c(t)}, \{\mathcal{H}_{j_c}\}_{j_c=1}^{N_c(t)} \right)$$
$$= E_{N_\ell, \{\mathcal{H}_{j_\ell}\}} \left[\bar{F}_{0_{\ell,c}}(t) \exp\left(-\sum_{j_\ell=1}^{N_\ell(t)} \mathcal{H}_{j_\ell} \min(W_{j_\ell}, t - T_{j_\ell}) \right) \exp\left(-\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c}) \right) \right]$$
$$= \bar{F}_{0_{\ell,c}} \exp\left(-\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c}) \right) E_{N_\ell, \{\mathcal{H}_{j_\ell}\}} \left[\exp\left(-\sum_{j_\ell=1}^{N_\ell(t)} \mathcal{H}_{j_\ell} \min(W_{j_\ell}, t - T_{j_\ell}) \right) \right]$$

But this case reduces to the previous RSBR case, and hence we have

$$= \bar{F}_{0_{\ell,c}} \exp\left(-\sum_{j_c=1}^{N_c(t)} \mathcal{H}_{j_c} \min(W_{j_c}, t - T_{j_c}) \right) \exp\left(-E_{\mathcal{H}} \left[\mathcal{H} \int_0^t e^{-\mathcal{H}w} m(t-w) \bar{G}_W(w) dw \right] \right)$$

□

Appendix B

Relevant Code

B.1 Generate Numerical Approximation for $\psi(\lambda)$

Uniform Distribution

The following is a sample of Mathematica code which generates a numerical approximation for $\psi(\lambda)$ under the service density of the uniform distribution on the interval [10,11].

```
(*          Numerical Investigation of efficiency for Uniform Distribution       *)



(* Declare the necessary constants \[Eta], r0, and \[Nu] *)
\[Eta] = 1;
r0 = 1;
\[Nu] = 1;



(*  First, we take care of the denominator. Denote the integral by S. Then
S is broken into the three parts defined by Sa[t], Sab[t,a,b], and Sb[t,a,b]       *)

(* For t < a *)

Sa[t_] := Integrate[Exp[-\[Eta]*w]*(t - w), {w, 0, t}]

(* For a <= t <= b *)

Sab[t_, a_, b_] := Integrate[Exp[-\[Eta]*w]*(t - w), {w, 0, a}] +
                        Integrate[Exp[-\[Eta]*w]*((b - w)/(b - a))*(t - w), {w, 0, t}]

(*For t > b *)

Sb[t_, a_, b_] :=  Integrate[Exp[-\[Eta]*w]*(t - w), {w, 0, a}] +
                       Integrate[Exp[-\[Eta]*w]*((b - w)/(b - a))*(t - w), {w, a, b}]
```

```
(*  Now we handle the numerator.   *)


(*—————————————————————————— a(t)————————————————————————————————*)

              (*For t < a, a(t) = 0. For a <= t < b *)

Aab[t_, a_, b_] := (1/(b − a))*Integrate[Exp[−\[Eta]*v]*(t − v), {v, a, t}]

              (*For t > b*)

Ab[t_, a_, b_] := (1/(b − a))*Integrate[Exp[−\[Eta]*v]*(t − v), {v, a, b}]

(*————————————————————————END a(t)————————————————————————————————*)

(*———————————————————————————b(t) ————————————————————————————————*)

b[t_, a_, b_] :=
              Integrate[Exp[−\[Eta]*(t − r)]*((b − (t − r))/(b − a)), {r, t − b, t − a}]
            +  Integrate[Exp[−\[Eta]*(t − r)], {r, t − a, t}]

(*———————————————————————— END b(t) ————————————————————————————————*)




(*    Create the loop for various values of \[Lambda]    *)


Clear[psi];
psi = {};

For[\[Lambda] = 0, \[Lambda] < 0.5, \[Lambda] += 0.05,
      Num =  NIntegrate[Exp[−r0*t − \[Lambda]*t + \[Lambda]*Aab[t, 10, 11]
                     + \[Lambda]*b[t, 10, 11]]*(r0 + \[Eta]*\[Lambda]*b[t, 10, 11])
                              *\[Lambda]*Aab[t, 10, 11], {t, 10, 11}]
                     + NIntegrate[Exp[−r0*t − \[Lambda]*t + \[Lambda]*Ab[t, 10, 11]
                     + \[Lambda]*b[t, 10, 11]]*(r0 + \[Eta]*\[Lambda]*b[t, 10, 11])
                          *\[Lambda]* Ab[t, 10, 11], {t, 11, Infinity}];
     Den = NIntegrate[Exp[−r0*t − \[Eta]*\[Lambda]*Sa[t]], {t, 0, 10}]
                 + NIntegrate[Exp[−r0*t − \[Eta]*\[Lambda]*Sab[t, 10, 11]], {t, 10, 11}]
                     + NIntegrate[Exp[−r0*t − \[Eta]*\[Lambda]*Sb[t, 10, 11]], {t, 11, Infinity}];
```

```
    AppendTo[psi, Num/(Den + \[Nu])]
  ]

  lamb = Range[0, .45, .05]

  ListLinePlot[Transpose[{lamb, psi}],
   AxesLabel -> {"\[Lambda]", "\[Psi]"}]
```

Increasing Density on Compact Support

```
g[w_] := (2/3)*w


(*Define parameters *)

\[Eta] = 1

\[Nu] = 1

r0 = 1

(*Calculate Denominator. *)

SP0[t_] := Integrate[Exp[-\[Eta]*w]*(t - w), {w, 0, t}]

SP1[t_] := Integrate[Exp[-\[Eta]*w]*((4 - w^2)/3)*(t - w), {w, 1, t}]

SP2[t_] :=  Integrate[Exp[-\[Eta]*w]*((4 - w^2)/3)*(t - w), {w, 1, 2}]

(* Numerator *)

Na1[t_] := Integrate[Exp[-\[Eta]*v]*(g[v] )*(t - v), {v, 1, t}]

Na2[t_] := Integrate[Exp[-\[Eta]*v]*(g[v] )*(t - v), {v, 1, 2}]

Nb[t_] := Integrate[Exp[-\[Eta]*(t - r)], {r, t - 1, t}] +
  Integrate[
   Exp[-\[Eta]*(t - r)]*((4 - (t - r)^2)/3), {r, t - 2, t - 1}]

(* Loop *)

Clear[psi]
```

```
psi = {}

lamb = Range[0, 2.9, .1]

For[\[Lambda] = 0, \[Lambda] < 3, \[Lambda] += .1,
 Num = NIntegrate[
    Exp[-r0*t - \[Lambda]*t + \[Lambda]*Na1[t] + \[Lambda]*
        Nb[t]]*(r0 + \[Eta]*\[Lambda]*Nb[t])*\[Lambda]*Na1[t], {t, 2,
     3}] + NIntegrate[
    Exp[-r0*t - \[Lambda]*t + \[Lambda]*Na2[t] + \[Lambda]*
        Nb[t]]*(r0 + \[Eta]*\[Lambda]*Nb[t])*\[Lambda]*Na2[t], {t, 3,
     Infinity}];
 Den = NIntegrate[
    Exp[-r0*t]*Exp[-\[Eta]*\[Lambda]*SP0[t]], {t, 0, 2}] +
   NIntegrate[Exp[-r0*t]*Exp[-\[Eta]*\[Lambda]*SP1[t]], {t, 2, 3}] +
   NIntegrate[
    Exp[-r0*t]*Exp[-\[Eta]*\[Lambda]*SP2[t]], {t, 3, Infinity}];
  AppendTo[
  psi, Num/(Den + \[Nu])]
 ]

points = Transpose[{lamb, psi}]

ListLinePlot[points, AxesLabel -> {"\[Lambda]", "\[Psi]"},
 PlotLabel -> "Increasing Distribution on [1,2]"]
```

Rayleigh Service Distribution

```
(* Efficiency Calculation for Rayleigh Distribution *)

In[5]:= $Assumptions = t > 0

 r0 = 1
 \[Eta] = 1

(*Define parameters *)

\[Eta] = 1
\[Nu] = 1
r0 = 1

(*Calculate Denominator. *)

In[9]:= SP0[t_] := Integrate[Exp[-\[Eta]*w]*(t - w), {w, 0, t}]
```

```
 SP1[t_] :=
 Integrate[
  Exp[−\[Eta]∗w]∗
   SurvivalFunction[RayleighDistribution[1], w]∗(t − w), {w, 0, t}]

(∗ Numerator ∗)

 Na1[t_] :=
 Integrate[
  Exp[−\[Eta]∗v]∗PDF[RayleighDistribution[1], v]∗(t − v), {v, 0, t}]

 Nb[t_] :=
 Integrate[
  Exp[−\[Eta]∗(t − r)]∗
   SurvivalFunction[RayleighDistribution[1], t − r], {r, 0, t}]

 \[Nu] = 1

Clear[psi]

 psi = {}

 lamb = Range[0, 49, 1]

For[\[Lambda] = 0, \[Lambda] < 50, \[Lambda] += 1,
 Num = NIntegrate[
   Exp[−r0∗t − \[Lambda]∗t + \[Lambda]∗Na1[t] + \[Lambda]∗
       Nb[t]]∗(r0 + \[Eta]∗\[Lambda]∗Nb[t])∗\[Lambda]∗Na1[t], {t, 0,
    Infinity}] ;
 Den = NIntegrate[
    Exp[−r0∗t]∗Exp[−\[Eta]∗\[Lambda]∗SP0[t]], {t, 0, 0}] +
   NIntegrate[
    Exp[−r0∗t]∗Exp[−\[Eta]∗\[Lambda]∗SP1[t]], {t, 0, Infinity}] ;
  AppendTo[
  psi, Num/(Den + \[Nu])]
 ]

points = Transpose[{lamb, psi}]

 ListLinePlot[points, AxesLabel −> {"\[Lambda]", "\[Psi]"},
 PlotLabel −> "Efficiency for Rayleigh(1) Distribution"]
```

Bibliography

[1] Robert R. Abernethy. *The New Weibull Handbook: Reliability and Statistical Analysis for Predicting Life, Safety, Supportability, Risk, Cost and Warranty Claims*. Dr. Robert. Abernethy, 2006.

[2] Arnold O. Allen. *Probability, Statistics, and Queuing Theory with Computer Science Appliations, 2nd edition*. Academic Press, Inc., 1990.

[3] Fred Douglis Mark Chamness Guanlin Lu Darren Sawyer Surendar Chandra Windsor Hu Ao Ma Rachel Traylor. "RAIDShield: Characterizing, Monitoring, and Proactively Protecting Against Disk Failures". In: *ACM Transactions on Storage* 11 (2015).

[4] M.A. Augustin and E.A. Pea. "A Dynamic Competing Risks Model". In: *Probability in the Engineering and Informational Sciences* 13 (1999), pp. 333–358.

[5] M. Bargout. "Predicting Software Reliability Using an Imperfect Debugging Jelinski Moranda Nonhomogenous Poisson Process model". In: *Model Assisted Statistical Applications* 5 (2010), pp. 31–41.

[6] N. Bhatti and R. Freidrich. "Web Server Support for Tiered Services". In: *IEEE Network* 13 (1999), pp. 64–71.

[7] P.J. Boland and Ni Chuiv. "Optimal Times for Software Release when Repar is Imperfect". In: *Statistical and Probability Letters* 77 (2007), pp. 1176–1184.

[8] P.J. Boland and H. Singh. "Determining the Optimal Time for Software in the Geometric Poisson Reliability Model". In: *International Journal of Reliability and Quality Safety Engineering* 9 (2002), pp. 201–213.

[9] Ki Hwan Cha and Eui Yong Lee. "A Stochastic Breakdown Model for an Unreliable Web Server System and an Optimal Admission Control Policy". In: *Journal of Applied Probability* 48.2 (2011), pp. 453–466.

[10] David Culler and Matt Welsh. "Adaptive Overload Control for Busy Internet Servers". In: *In Proceedings of the 4th conference on USENIX Symposium on Internet Technologies and Systems*. Vol. 4. 2003.

[11] Z. Jelinski and P. Moranda. "Software Reliability Research". In: *Statistical Computer Performance Evaluation* (1972), pp. 465–484.

[12] Stephen H. Kan. *Metrics and Models in Software Quality Engineering*. Addison-Wesley, 2003.

[13] Nigel Thomas Katja Gilly Carlos Juiz and Ramon Puigjaner. "Adaptive Admission Control Algorithm in a QoS-aware Web system". In: *Journal of Information Sciences* 199 (2012), pp. 58–77.

[14] Andrzej Korzeniowski. "On Correlated Random Graphs". In: *Journal of Probability and Statistical Science* 11 (2013), pp. 43–58.

[15] Lawrence M. Leemis. *Reliability: Probabilistic Models and Statistical Methods, 2nd edition*. Prentice-Hall, 2009.

[16] Y. Ling and J. Mi. "An optimal trade-off between content freshness and refresh cost". In: *Journal of Applied Probability* 41 (2004), pp. 721–734.

[17] B. Littlewood. "The Littlewood-Verrall model for software reliability compared with some rivals". In: *Journal of Systems and Software* 1 (1980), pp. 251–258.

[18] J. Mi. "Age Replacement Policy and Optimal Work Size". In: *Journal of Applied Probability* 39 (2002), pp. 296–311.

[19] Alexandru Iosup Nezih Yigitbasi Ozan Sonmez and Dick Epema. "Performance Evaluation of Overload Control in Multi-Cluster Grids". In: *Proceedings of the 2011 IEEE/ACM 12th annual Conference on Grid Computing*. 2011, pp. 173–180.

[20] H. Singh P.J. Boland and B. Cukic. "Stochastic Orders in Partition and Random Testing of Software". In: *Journal of Applied Probability* 39 (2002), pp. 555–565.

[21] Marvin Rausand and Arnljot Høyland. *System Reliability Theory: Models, Statistical Methods, and Applications, 2nd edition*. John Wiley and Sons, 2004.

[22] *Rayleigh Distribution*. URL: `https://en.wikipedia.org/wiki/Rayleigh_distribution`.

[23] Sheldon Ross. *Stochastic Processes, 2nd edition*. 1996.

[24] N.D. Singpurwalla. "Determining an optimal time interval for testing and debugging software". In: *IEEE Transactions on Software Engineering* 17 (1991), pp. 313–319.

[25] *Weibull Distribution*. URL: `https://en.wikipedia.org/wiki/Weibull_distribution`.

[26] Chuan Ye and Haining Wang. "Profit-Aware Overload Protection in E-Commerce Websites". In: *Journal of Network and Computer Applications* 32 (2 2009), pp. 347–356.