Heavy Tails and Instabilities in Large-Scale Systems with Failures

Evangelia Skiani

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

Heavy Tails and Instabilities in Large-Scale Systems with Failures

Evangelia Skiani

Modern engineering systems, e.g., wireless communication networks, distributed computing systems, etc., are characterized by high variability and susceptibility to failures. Failure recovery is required to guarantee the successful operation of these systems. One straightforward and widely used mechanism is to restart the interrupted jobs from the beginning after a failure occurs. In network design, retransmissions are the primary building blocks of the network architecture that guarantee data delivery in the presence of channel failures. Retransmissions have recently been identified as a new origin of power laws in modern information networks. In particular, it was discovered that retransmissions give rise to long tails (delays) and possibly zero throughput. To this end, we investigate the impact of the 'retransmission phenomenon' on the performance of failure prone systems and propose adaptive solutions to address emerging instabilities.

The preceding finding of power law phenomena due to retransmissions holds under the assumption that data sizes have infinite support. In practice, however, data sizes are upper *bounded* $0 \leq L \leq b$, e.g., WaveLAN's maximum transfer unit is 1500 bytes, YouTube videos are of limited duration, e-mail attachments cannot exceed 10MB, etc. To this end, we first provide a uniform characterization of the entire body of the distribution of the number of retransmissions, which can be represented as a product of a power law and the Gamma distribution. This rigorous approximation clearly demonstrates the transition from power law distributions in the main body to exponential tails. Furthermore, the results highlight the importance of wisely determining the size of data fragments in order to accommodate

the performance needs in these systems as well as provide the appropriate tools for this fragmentation.

Second, we extend the analysis to the practically important case of *correlated* channels using modulated processes, e.g., Markov modulated, to capture the underlying dependencies. Our study shows that the tails of the retransmission and delay distributions are asymptotically insensitive to the channel correlations and are determined by the state that generates the lightest tail in the independent channel case. This insight is beneficial both for capacity planning and channel modeling since the independent model is sufficient and the correlation details do not matter. However, the preceding finding may be overly optimistic when the best state is atypical, since the effects of 'bad' states may still downgrade the performance.

Third, we examine the effects of *scheduling* policies in queueing systems with failures and restarts. Fair sharing, e.g., processor sharing (PS), is a widely accepted approach to resource allocation among multiple users. We revisit the well-studied M/G/1 PS queue with a new focus on server failures and restarts. Interestingly, we discover a new phenomenon showing that PS-based scheduling induces complete instability in the presence of retransmissions, regardless of how low the traffic load may be. This novel phenomenon occurs even when the job sizes are bounded/fragmented, e.g., deterministic. This work demonstrates that scheduling one job at a time, such as first-come-first-serve, achieves a larger stability region and should be preferred in these systems.

Last, we delve into the area of *distributed* computing and study the effects of commonly used mechanisms, i.e., restarts, fragmentation, replication, especially in cloud computing services. We evaluate the efficiency of these techniques under different assumptions on the data streams and discuss the corresponding optimization problem. These findings are useful for optimal resource allocation and fault tolerance in rapidly developing computing networks.

In addition to networking and distributed computing systems, the aforementioned results improve our understanding of failure recovery management in large manufacturing and

service systems, e.g., call centers. Scalable solutions to this problem increase in significance as these systems continuously grow in scale and complexity. The new phenomena and the techniques developed herein provide new insights in the areas of parallel computing, probability and statistics, as well as financial engineering.

Table of Contents

List of Figures **iii**

List of Tables **v**

1 Introduction **1**

1.1 Related Work & Main Contribution 3

1.2 Notation 6

1.3 Thesis Outline 7

2 Distribution of the Number of Retransmissions of Bounded Documents **8**

2.1 Introduction 9

2.1.1 Description of the Channel 13

2.2 Main Results 15

2.3 Exact Asymptotics 20

2.4 Simulation Experiments 39

2.5 Concluding Remarks 44

3 Retransmissions over Correlated Channels **45**

3.1 Introduction 46

3.1.1 Description of the Channel 48

3.2 Main Results 50

3.3 Simulations 55

3.4 Concluding Remarks 57
3.4.1 A word of caution 58
3.5 Proofs 61

4 Instability of Sharing Systems in the Presence of Retransmissions 73

4.1 Introduction 73
4.2 Definitions and Notation 77
4.3 M/G/1 Queue with Restarts 79
4.3.1 Instability of Processor Sharing Queue 80
4.3.2 Stability of One Job at a Time Non-Preemptive Policy 89
4.4 GI/G/1 PS Queue with Restarts 91
4.5 Transient Behavior - Scheduling a Finite Number of Jobs 96
4.5.1 One Job at a Time Non-Preemptive Policy 98
4.5.2 Processor Sharing Discipline 100
4.6 Simulation 105
4.7 Concluding Remarks 113

5 Future Work & Conclusion 114

5.1 Towards Stabilizing Sharing Systems 114
5.1.1 Throughput 118
5.2 Reliability Tradeoffs in Cloud Computing 123
5.2.1 Fragmentation 123
5.2.2 Replication 125
5.2.3 Simulation Examples 126
5.2.4 Replication Or Fragmentation? 130
5.3 Concluding Remarks 130

List of Figures

1.1 Jobs executed in a system with failures. 6

2.1 Documents sent over a channel with failures. 14
2.2 *Example 1(a). Exact asymptotics for* $\alpha > 1$. 40
2.3 *Example 1(b). Exact asymptotics for* $\alpha < 1$. 40
2.4 *Example 2. Power law vs. exponential tail asymptotics.* 41
2.5 *Example 3. Power law region increases for lighter tails of* L, A. 41
2.6 *Example 4(a). Exact asymptotics for the case where L follows the Gamma distribution.* 42
2.7 *Example 4(b). The asymptotes from Theorems 2.3 & 2.4 for Gamma distributed L.* 42

3.1 Correlated channel dynamics. 49
3.2 Packets sent over a channel with failures. 49
3.3 Example 1. Asymptotics of $\mathbb{P}[N > n]$ and transmission delay $\mathbb{P}[T > t]$ for a two-state channel. 56
3.4 Example 2. Asymptotics of $\mathbb{P}[N > n]$ for a three-state channel. 56
3.5 Example 3. Logarithmic asymptotics for a two-state channel where data sizes and channel statistics are normally distributed. 57
3.6 Example (a). Exact asymptotes from (3.6) and (3.7) for a two state channel where $\alpha_1 = 2$ and $\alpha_2 = 1/2$. 59
3.7 Example (b). Asymptotics of $\mathbb{P}[N > n]$ for a three-state channel. 60

4.1 System with failures. 78
4.2 Jobs executed in a system with failures. 79
4.3 Example 1. Jobs completed over time. 105
4.4 Example 1. Queue size evolution. Subfigure (b) zooms in the time range $[0, 10^6]$ of Fig. 4.4; Q_t (y-axis) is shown on the logarithmic scale. 106
4.5 (a) Example 1. Queue size over time parameterized by fragment length; $\beta = 2, \lambda = 0.1$. (b) Example 2. Queue size over time parameterized by job size; $\beta = 4$. 107
4.6 Example 3. Non-preemptive policy: Logarithmic asymptotics when $\alpha = 2$ for exponential, superexponential ($\gamma > 1$) and subexponential ($\gamma < 1$). distributions. 108
4.7 (a) Example 4. Logarithmic asymptotics for different number of superexponential jobs when $\alpha = 4$ under PS and FCFS discipline. (b) Example 5. Logarithmic asymptotics under FCFS, PS with subexponential and superexponential jobs. 109
4.8 Example 6. Throughput vs. utilization tradeoff. 111

5.1 Example 1(a). Throughput $\gamma(k)$; the dotted line is the approximation from (5.4) 120
5.2 Example 1(b). Throughput $\gamma(k)$ for different failure rates μ. 121
5.3 Example 2. Drift for different initial queue sizes Q_0. 122
5.4 Example 1. $\mathbb{P}(\mathcal{T}_k > t)$ and $\mathbb{P}(\mathcal{N}_k > n)$ for different number of fragments/replicas k. 127
5.5 Example 2. $\mathbb{P}(\mathcal{N}_5 > n)$ for different distribution types. 128
5.6 Example 3. $\mathbb{E}[\mathcal{T}_k]$ for different types of distributions. 129

List of Tables

5.1 Transition matrix $\mathcal{P}$ 116

Acknowledgments[1]

"It is not Ithaca, but the journey that matters". Along my PhD journey, there were many exciting moments but also times when everything seemed impossible. Nothing would be possible without the help of many bright and talented people I met during my stay at Columbia. Most important among all I am indebted to my advisor, Predrag R. Jelenkovíc, who with his passion, support, and continuous encouragement, helped me tackle one of the hardest endeavors in my life. His strong personality and the ease with which he solved seemingly intractable problems will always be an inspiration for me.

My sincere gratitude goes to my great professors, Ioannis Karatzas and Augustin Chaintreau, for being my role models throughout these years as well as for their kindness and support. I would like to thank my thesis committee members: Aurel Lazar and Javad Ghaderi for the valuable feedback on my thesis. I am also grateful to Ranveer Chandra for being an excellent mentor at Microsoft Research, as well as Yigal Bejerano and Matthew Andrews for our collaboration at Bell Laboratories.

I am extremely thankful to all my friends that glorified my stay at Columbia: Orestis, Katerina, Aya and Sabrina. My beloved parents Dimitris and Eleni for their love, support and endless faith in me. My sister Natasa for listening, cheering me up, and most importantly being my best friend. My dearest grandma, Evangelia, the person I was attached the most, whose strength, passion and love will forever be remembered. To everyone above and all my family and friends I did not mention, I will always be thankful for making this journey worth taking.

[1]This work is supported by NSF Grant number 0915784.

Chapter 1

Introduction

This thesis focuses on addressing the challenges imposed by network design and large-scale systems, within a multi-disciplinary framework, involving theoretical components that range from the area of operations research to electrical engineering. The objective lies in improving systems design with respect to multiple parameters, such as throughput, robustness, reliability, scalability, etc. To this end, it is crucial to analyze the asymptotic behavior of large-scale systems and, most importantly, understand the underlying laws that govern their performance. Special emphasis is placed on reexamining the existing design principles inherent to all networking layers.

We study the retransmission phenomenon in failure prone channels, where the so-called "restart mechanism" is deployed. This is the case for most wireless networks, where frequent channel failures occur due to signal fading, multi-path effects, interference, node contention, and other environmental changes. One of the most straightforward and widely used failure recovery mechanism is to simply restart the system and all of the interrupted jobs from the beginning after a failure occurs. Retransmissions represent one of the most fundamental approaches in communication networks that guarantee data delivery in the presence of channel failures. These types of mechanisms have an impact on the entire protocol stack:

- *Physical layer.* Wireless links, especially for low-powered sensor networks, exhibit high error rates, thus resulting in long delays on the data link layer due to the packet

variability and channel failures. We stress the need for novel fragmentation techniques and efficient coding schemes, since the results suggest that when codewords are much smaller than the maximum size of the packets, the number of retransmissions could be distributed as power laws, instead of geometrically, as the traditional models assume.

- *Medium Access Control (MAC) layer.* ALOHA is a widely used protocol that provides a contention management scheme for multiple users that share the same medium. Failures result from collisions due to the simultaneous attempts of multiple users to access the common channel. Once a collision is detected, the users retransmit with random (exponential) back-offs. Due to its simplicity and distributed nature, ALOHA is the basis of many more sophisticated collision avoidance/resolution protocols, such as CSMA/CD.

- *Transport layer.* Network protocols, like TCP, use end-to-end acknowledgements for packets as an error control strategy. Namely, once the packet sent from the sender to the receiver is lost due to, e.g., finite buffers or link failures, this packet will be retransmitted by the sender. Furthermore, the number of hops that a packet traverses on its path to the destination is random, e.g., an end-user that is surfing the Web might download documents from diverse web sites. The delay for transversing paths with random number of hops can be power law even if the system variables are exponential.

- *Application layer.* Application-layer protocols implement specific user applications and other high-level functions. Many application protocols, e.g., HTTP, employ retransmissions of Internet files, such as webpages, as a primary failure recovery mechanism. For example, when an HTTP request fails, e.g., a webpage is downloaded with errors or the web server does not respond, the file is downloaded from scratch via a new HTTP query.

1.1 Related Work & Main Contribution

The preceding discussion reveals the significant impact of retransmission-based failure recovery mechanisms on existing networks, and clearly set the basis for more discoveries in this domain along the vertical (protocol stack), temporal and spatial network dimensions. Along these lines, we study the retransmission phenomenon in the presence of correlated channels and data streams, bounded packets, as well as spatial and temporal interactions of many channels.

Traditionally, retransmissions were thought to follow light-tailed, e.g., geometric, distributions. This is only true under the assumption that data sizes and transmission error probability are independent. Nevertheless, these two are often highly correlated. It was first recognized in [24, 26] that such mechanisms may result in long-tailed (power law) delays. In modern communication systems, it has been shown that several well-known retransmission based protocols in different layers of networking architecture can lead to power law delays, e.g., ALOHA type protocols in MAC layer [12, 14] and end-to-end acknowledgments in transport layer [11, 9] as well as in other layers [9, 10]. The preceding studies considered distributions with infinite support.

In reality, data sizes are upper bounded by the maximum transmission unit. This situation results in "truncated" power laws for the number of retransmissions: such distributions are characterized by a power law main body and an exponentially bounded tail; see Example 3 in [9] and Example 2 in [12]. The retransmissions of bounded documents were further studied in [39], where partial approximations of their distribution were derived. We establish a uniform approximation for the number of retransmissions [2, 1] when the data sizes are bounded. In fact, this distribution can be represented as the product of a power law, which dominates the main body, and the Gamma distribution, determining the exponential tail.

This uniform approximation allows for a characterization of the entire body of the distribution, so that we can explicitly estimate the region where the power law phenomenon

arises and predict the tradeoff points between packet sizes and delays in these systems. From an engineering perspective, these results further suggest that careful re-examination and redesign of retransmission based protocols might be necessary. In infrastructure-less, error-prone wireless systems, traditional approaches, e.g., blind data fragmentation, may be insufficient for achieving a good balance between throughput and resource utilization. One example of bad resource allocation is fragmenting a large data unit into many smaller ones of the same header (needed to reassemble them). Smaller data sizes fail less frequently but they result in unnecessary overhead and increased network traffic.

The preceding studies have considered an independent channel model. In practice, communication channels are highly correlated in the sense that they switch between states with different characteristics. We extend the previously studied independent model to the dependent case where the availability periods depend on the channel state, i.e., the channel is correlated [3, 6]. We introduce an underlying modulating process to capture the channel dependencies. In this setting, we show that the tails of the retransmission and delay distributions are asymptotically insensitive to the channel correlations and are determined by the 'best' channel state, i.e., the one that generates the lightest asymptotics under the independent channel model. Intuitively, as the channel switches between states, a large data unit is more likely to be transmitted when the channel is in a 'good' state.

This optimistic best case scenario prediction and the apparent insensitivity to the structure of the channel correlations can be very promising in system analysis and design. The result implies that the initial independent model might be sufficient for modeling, and can also be extended to even more complex failure-prone networks. However, this is partially true as there are certain circumstances under which this claim underestimates the intricacies of the system. In particular, using the tail as a primary performance measure might result in an overly optimistic design if the 'best' state is atypical, i.e., it occurs very rarely. In the latter case, the main body of the distribution may be (much) heavier than the tail. The results may also be applied in designing new protocols, or developing new fragmentation schemes, specifically for correlated channels. In addition, combining this study with our

results on bounded data units could provide an accurate estimate for the optimal sizes of the packet fragments.

Apart from studying system performance under the aforementioned realistic assumptions for modern information networks, we are also interested in scheduling policies under heavy-tailed service times induced by restart mechanisms for failure recovery. Sharing is a primary approach to fair scheduling and efficient management of the available resources, e.g., CDMA is a multiple access method used in communication networks, where several users can transmit information simultaneously over a single channel via sharing the available bandwidth, Processor Sharing (PS) [36]/ *Generalized* PS (GPS) [32] scheduling, where the capacity is equally shared between multiple classes of customers, *Discriminatory* PS (DPS) [17, 23, 31] which is used to model the Weighted Round Robin (WRR) scheduling and/or TCP connections, etc.

In general, PS-based scheduling disciplines have been widely used in computer and communication networks. Early investigations of PS queues were motivated by applications in multiuser computer systems [22]. The M/G/1 PS queue has been studied extensively in the literature [35]. The importance of scheduling in the presence of heavy tails was first recognized in [18], and later, in [29], the M/G/1 PS queue was studied assuming subexponential job sizes; see also [29] for additional references. Herein, we evaluate the effect of fair sharing on queueing systems with a new focus on failures and restarts. We discover a new phenomenon: processor sharing with restarts is always unstable. The intuition is the following. In a queueing system, if many packets of similar sizes arrive within a short interval of time, sharing the total capacity of the channel will lead to longer individual delays; these delays may be too long to allow for the queue to empty.

The above paradigms call for the thorough study of the resulting tradeoffs and the design of new dynamic algorithms possibly exploiting the channel's statistical characteristics. We discuss failure recovery management in large distributed systems, such as massively parallel computing, where scalable solutions to this problem increase in significance as networking systems continuously grow in scale and complexity. Our theoretical work includes model

formulation and system analysis through the utilization of probabilistic, statistical and analytical tools. In addition, a large number of scenarios are evaluated via simulation. Performance evaluation lends credence to our theoretical arguments and provides important feedback to system design. This feedback is crucial since existing protocols are in dire need of novel network algorithms that can demonstrate easy implementation, robustness, adaptability and near-optimality under some analytical conditions.

1.2 Notation

Throughout this thesis, we use the following generic model to analyze a system with failures. The most basic description of our model can be stated as follows. The system dynamics is described as a process $\{A, A_i\}_{i\geq 1}$ of i.i.d. available periods, where the channel is continuously available during periods A_i and fails between such periods. In each period of time that the channel becomes available, we attempt to execute a job of random size B. If $B < A_i$, the job is successfully completed; otherwise, we wait for the next period A_{i+1} when the channel is available and attempt to retransmit the data from the beginning. This model was first introduced in [24] in the computing context, and was later applied in the context of networking [9, 10]. Throughout the thesis, the assumptions of this model may change and the specifics will be provided in the corresponding chapters. A sketch of the model depicting the system is drawn in Figure 1.1.

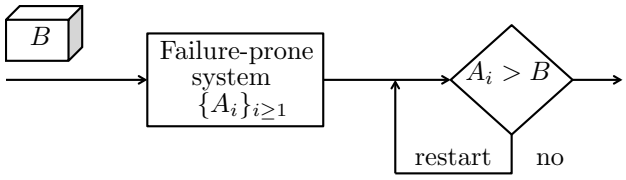


Figure 1.1: Jobs executed in a system with failures.

We also define the complementary cumulative distribution functions for A and B, respectively, as

$$\bar{G}(x) := \Pr(A > x) \quad \text{and} \quad \bar{F}(x) := \Pr(B > x).$$

Last, when jobs B refer to data units (packets) we use the variable L instead, which is explicitly stated in the corresponding chapters.

1.3 Thesis Outline

This thesis is organized as follows. In Chapter 2, we derive the distribution of the number of retransmissions when data sizes are bounded [1, 2]. Chapter 3 extends the previously studied i.i.d. channel model to the correlated case [3, 6], i.e., when the channel switches between different states. Next, in Chapter 4 we evaluate the impact of scheduling policies and, in particular, we discover a new phenomenon that fair sharing in queueing systems with failures and restarts always leads to instability [4, 5, 8, 7]. Last, Chapter 5 concludes the thesis and includes further extensions of this work in the area of cloud computing and distributed systems.

Chapter 2

Distribution of the Number of Retransmissions of Bounded Documents

Retransmission-based failure recovery represents a primary approach in existing communication networks that guarantees data delivery in the presence of channel failures. Recent work has shown that [24, 25, 9, 10], when data sizes have infinite support, retransmissions can cause long-tailed delays even if all traffic and network characteristics are light tailed. In this chapter, we investigate the practically important case of bounded data units $0 \leq L_b \leq b$ under the condition that the hazard functions of the distributions of data sizes and channel statistics are proportional. To this end, we provide an explicit and uniform characterization of the entire body of the retransmission distribution $\mathbb{P}[N_b > n]$ in both n and b. Our main discovery is that this distribution can be represented as the product of a power law and Gamma distribution. This rigorous approximation clearly demonstrates the coupling of a power law distribution, dominating the main body, and the Gamma distribution, determining the exponential tail. Our results are validated via simulation experiments and can be useful for designing retransmission-based systems with the required performance character-

istics. From a broader perspective, this study applies to any other system, e.g., computing, where restart mechanisms are employed after a job processing failure.

2.1 Introduction

Failure recovery mechanisms are employed in almost all engineering networks since complex systems of any kind are often prone to failures. One of the most straightforward and widely used failure recovery mechanism is to simply restart the system and all of the interrupted jobs from the beginning after a failure occurs. It was first recognized in [24, 25] that such mechanisms may result in long-tailed (power law) delays even if the job sizes and failure rates are exponential. In [9], it was noted that the same mechanism is at the core of modern communication networks where retransmissions are used on all protocol layers to guarantee data delivery in the presence of channel failures. Furthermore, [9] shows that the power law number of retransmissions and delay occur whenever the hazard functions of the data and failure distributions are proportional. Hence, power laws may arise even if the data and channel failure distributions are both Gaussian. In particular, retransmission phenomena can lead to zero throughput and system instabilities, and therefore need to be carefully considered for the design of fault tolerant systems.

More specifically, in communication networks, retransmissions represent the basic building blocks for failure recovery in all network protocols that guarantee data delivery in the presence of channel failures. These types of mechanisms have been employed on all networking layers, including, for example, Automatic Repeat reQuest (ARQ) protocol (e.g., see Section 2.4 of [19]) in the data link layer where a packet is resent automatically in case of an error; contention based ALOHA type protocols in the medium access control (MAC) layer that use random backoff and retransmission mechanism to recover data from collisions; end-to-end acknowledgment for multi-hop transmissions in the transport layer; HTTP downloading scheme in the application layer, etc. It has been shown that several well-known retransmission based protocols in different layers of networking architecture can

lead to power law delays, e.g., ALOHA type protocols in MAC layer [12, 14] and end-to-end acknowledgments in transport layer [11, 9] as well as in other layers [9]. For other (non-retransmission) mechanisms that can give rise to heavy tails see [15] and the references therein. In particular, the proportional growth/multiplicative models can result in heavy tails [15, 16].

Traditionally, retransmissions were thought to follow light-tailed distributions (with rapidly decaying tails), namely geometric, which requires the further assumption of independence between data sizes and transmission error probability. However, these two are often highly correlated in most communication systems, meaning that longer data units have higher probability of error, thus violating the independence assumption. Recent work [9, 12, 11, 10] has shown that, when the data size distribution has infinite support, all retransmission-based protocols could cause heavy-tailed behavior and possibly result in zero throughput, regardless of how light-tailed the distributions of data sizes and channel failures are. Nevertheless, in reality, data sizes are usually upper bounded. For example, WaveLAN's maximum transfer unit is 1500 bytes, YouTube videos are of limited duration, e-mail attachments cannot exceed an upper limit, say 25MB, etc. This fact motivates us to investigate the transmission of bounded data and approximate uniformly the entire body of the resulting retransmission distribution as it transits from the power law to the exponential tail.

We use the following generic channel with failures [9] to model the preceding situations. This model was first introduced in [24] in a different application context. The channel dynamics is described by the i.i.d. channel availability process $\{A,\ A_i\}_{i\geq 1}$, where the channel is continuously available during periods $\{A_i\}$ and fails between these periods. In each period of time that the channel becomes available, say A_i, we attempt to transmit the data unit of random size L_b. We focus on the situation when the data size has finite support on interval $[0, b]$. If $L_b < A_i$, we say that the transmission is successful; otherwise, we wait for the next period A_{i+1} when the channel is available and attempt to retransmit the data from the beginning. It was first recognized in [24] that this model results in

power law distributions when the distributions of $L \equiv L_\infty$ and A have a matrix exponential representation, and this result was rigorously proved and further generalized in [9, 10, 26]. A related study when $L = \ell$ is a constant and failure/arrival rates are time-dependent Poisson can be found in [27].

It was discovered in [9] that bounded data units result in truncated power law distributions for the number of retransmissions, see Example 3 in [9]; see also Example 2 in [12]. Such distributions are characterized by a power law main body and an exponentially bounded tail. However, the exponential behavior appears only for very small probabilities, often meaning that the number of retransmissions of interest may fall inside the region where the distribution behaves as a power law. It was argued in Example 3 of [9] that the power law region will grow faster than exponential if the distributions of A and L_b are lighter than exponential. The retransmissions of bounded documents were further studied in [39], where partial approximations of the distribution of the number of retransmissions on the logarithmic and exact scales were provided in Theorems 1 and 3 of [39], respectively. Herein, we present a uniform characterization of the entire body of such a distribution, both on the logarithmic as well as the exact scale.

Specifically, let N_b represent the number of retransmissions (until successful transmission) of a bounded random data unit of size $L_b \in [0, b]$ on the previously described channel. In order to study the uniform approximation in both n and b we construct a family of variables L_b, such that $\mathbb{P}[L_b \leq x] = \mathbb{P}[L \leq x]/\mathbb{P}[L \leq b]$, for $0 \leq x \leq b$ when $L = L_\infty$ is fixed. This scaling of L_b was also used in [39]. For the logarithmic scale, our result, stated in Theorem 2.2, provides a uniform characterization of the entire body of $\log \mathbb{P}[N_b > n]$, i.e., informally

$$\log \mathbb{P}[N_b > n] \approx -\alpha \log n + n \log \mathbb{P}[A \leq b]$$

for all n and b sufficiently large when the hazard functions of L and A are linearly related as $\log \mathbb{P}[L > x] \approx \alpha \log \mathbb{P}[A > x]$; see Theorem 2.2 for the precise assumptions. Note that the first term in the preceding approximation corresponds to the power law part $n^{-\alpha}$ of

the distribution, while the second part describes the exponential (geometric $\mathbb{P}[A \leq b]^n$) tail. Hence, it may be natural to define the transition point n_b from the power law to the exponential tail as a solution to $n_b \log \mathbb{P}[A \leq b] \approx \alpha \log n_b$.

In addition, under more restrictive assumptions, we discover a new exact asymptotic formula for the retransmission distribution that works uniformly for all large n, b. Surprisingly, the approximation admits an explicit form (see Theorems 2.3 and 2.4)

$$\mathbb{P}[N_b > n] \approx \frac{\alpha}{n^\alpha \ell(n \wedge \mathbb{P}[A > b]^{-1})} \int_{-n \log \mathbb{P}[A \leq b]}^{\infty} e^{-z} z^{\alpha-1} dz, \tag{2.1}$$

where $x \wedge y = \min(x, y)$ and $\ell(\cdot)$ is a slowly varying function; note that the preceding integral is the incomplete Gamma function $\Gamma(x, \alpha)$.

Clearly, when $-n \log \mathbb{P}[A \leq b] \downarrow 0$, the preceding approximation converges to a true power law $\Gamma(\alpha + 1)/(\ell(n) n^\alpha)$. And, when $-n \log(\mathbb{P}[A < b]) \uparrow \infty$, approximation (2.1), by the property $\Gamma(x, \alpha) \sim e^{-x} x^{\alpha-1}$ as $x \to \infty$, has a geometric leading term $\mathbb{P}[A \leq b]^n$. Interestingly, for the special case when α is an integer and $\ell(x) \equiv 1$, one can compute the exact expression for $\mathbb{P}[N_b > n]$, see Proposition 2.2. Furthermore, our results show that the length of the power law region increases as the corresponding distributions of L and A assume lighter tails. All of the preceding results are validated via simulation experiments in Section 2.4. It is worth noting that our asymptotic approximations are in excellent agreement with the simulations.

This uniform approximation allows for a characterization of the entire body of the distribution $\mathbb{P}[N_b > n]$, so that one can explicitly estimate the region where the power law phenomenon arises. Introducing the relationship between n and $\mathbb{P}[A > b]$ also provides an assessment method of efficiency and is important for diminishing the power law effects in order to achieve high throughput. Basically, when the power law region is significant, it could lead to nearly zero throughput ($\alpha < 1$), implying that the system parameters should be more carefully adjusted in order to meet the new requirements. On the contrary, if the exponential tail dominates, the system performance is more desirable. Our analytical

work could be applicable in network protocol design, possibly including data fragmentation techniques [13, 38, 37] and failure-recovery mechanisms.

Also, from an engineering perspective, our results further suggest that careful re-examination and possible redesign of retransmission based protocols in communication networks might be necessary. Specifically, current engineering trends towards infrastructure-less, error-prone wireless technology encourage the study of highly variable systems with frequent failures. In these types of systems, traditional approaches, e.g., blind data fragmentation, may be insufficient for achieving a good balance between throughput and resource utilization. For example, IP packets are lower bounded by the packet header of 20 bytes and cannot be more than 1500 bytes. Thus, it is not efficient to create very small packets since the 20-byte packet header carries no useful information. In fact, one may consider merging smaller packets to reduce the overhead and, hence, increase the efficiency. Overall, we consider a generic model when the maximum size of data units is limited, which, in general, can be used towards improving the design of future complex and failure-prone systems in many different applications.

The rest of the chapter is organized as follows. After a detailed description of the channel model in the next Section 2.1.1, we present our main results in Section 2.2. Finally, Section 2.4 contains simulation examples that verify our theoretical work, while Section 2.5 concludes the chapter.

2.1.1 Description of the Channel

In this section, we formally describe our model and provide necessary definitions and notation. Consider transmitting a generic data unit of random size L_b over a channel with failures. Without loss of generality, we assume that the channel is of unit capacity. As stated in the introduction, the channel dynamics is modeled by the channel availability process $\{A, A_i\}_{i\geq 1}$, where the channel is continuously available during time periods $\{A_i\}$ whereas it fails between such periods. In each period of time that the channel becomes available, say A_i, we attempt to transmit the data unit and, if $A_i > L_b$, we say that the

transmission was successful; otherwise, we wait for the next period A_{i+1} when the channel is available and attempt to retransmit the data from the beginning. A sketch of the model depicting the system is drawn in Figure 4.1.

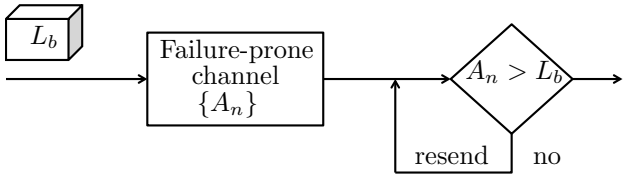


Figure 2.1: Documents sent over a channel with failures.

We are interested in computing the number of attempts N_b (retransmissions) that is required until L_b is successfully transmitted, which is formally defined as follows.

Definition 2.1.1. *The total number of retransmissions for a generic data unit of length L_b is defined as*

$$N_b \triangleq \inf\{n : A_n > L_b\}.$$

We denote the complementary cumulative distribution functions for A and L, respectively, as

$$\bar{G}(x) \triangleq \mathbb{P}[A > x] \quad \text{and} \quad \bar{F}(x) \triangleq \mathbb{P}[L > x],$$

where L is a generic random variable that is used to define the distribution of L_b.

Throughout this section we assume that L and A are continuous (equivalently, $\bar{F}(x)$ and $\bar{G}(x)$ are absolutely continuous) and have infinite support, i.e., $\bar{G}(x) > 0$ and $\bar{F}(x) > 0$ for all $x \geq 0$. Then, the distribution of L_b is defined as

$$\mathbb{P}[L_b \leq x] = \frac{\mathbb{P}[L \leq x]}{\mathbb{P}[L \leq b]}, \quad 0 \leq x \leq b. \tag{2.2}$$

To avoid trivialities, we assume that b is large enough such that $\mathbb{P}[L \leq b] > 0$.

We use the following standard notations. For any two real functions $a(t)$ and $b(t)$ and fixed $t_0 \in \mathbb{R} \bigcup \{\infty\}$, we use $a(t) \sim b(t)$ as $t \to t_0$ to denote $\lim_{t\to t_0} a(t)/b(t) = 1$. Similarly, we say that $a(t) \gtrsim b(t)$ as $t \to t_0$ if $\liminf_{t\to t_0} a(t)/b(t) \geq 1$; $a(t) \lesssim b(t)$ has a complementary

definition.

2.2 Main Results

In this section, we present our main results. Under mild conditions, we first prove a general upper bound for the distribution of N_b on the logarithmic scale in Proposition 2.1. In Theorem 2.2, we present our first main result, which under more stringent assumptions, characterizes the entire body of the distribution on the logarithmic scale uniformly for all large n and b, i.e., informally we show that

$$\log \mathbb{P}[N_b > n] \approx -\alpha \log n + n \log \mathbb{P}[A \leq b],$$

as previously mentioned in the introduction. Roughly speaking, when $-\log \mathbb{P}[A \leq b] = o(\log n/n)$, $\mathbb{P}[N_b > n]$ is a power law of index α. Our results on the exact asymptotics are given in the next Subsection 2.3 in Theorems 2.3 and 2.4; the results are stated in two different theorems since Theorem 2.4 requires slightly stronger assumptions. The uniform approximation implied by these two theorems is presented in (2.7), or previously in (2.1).

Recall that the distribution of L_b has finite support on $[0, b]$, given by (2.2). First, we prove the following general upper bound.

Proposition 2.1. *Assume that*

$$\liminf_{x\to\infty} \frac{\log \mathbb{P}[L > x]}{\log \mathbb{P}[A > x]} \geq \alpha$$

and let b_0 be such that $\mathbb{P}[L \leq b_0] > 0, \mathbb{P}[A \leq b_0] > 0$, then for any $\epsilon > 0$, there exists n_0, such that, for all $n \geq n_0, b \geq b_0$,

$$\log \mathbb{P}[N_b > n] \leq (1-\epsilon)\left[n \log \mathbb{P}[A \leq b] - \alpha \log n\right].$$

Remark 1. *Note that this result can be restated as*

$$\mathbb{P}[N_b > n] \leq \mathbb{P}[A \leq b]^{n(1-\epsilon)} n^{-\alpha(1-\epsilon)},$$

for n, b sufficiently large. Hence, the distribution $\mathbb{P}[N_b > n]$ is bounded by the product of a power law and a geometric term.

Proof. By assumption, there exists $0 < \epsilon < 1$ such that for all $x > x_\epsilon \geq b_0 > 0$,

$$\bar{F}(x) \leq \bar{G}(x)^{\alpha(1-\epsilon)}. \tag{2.3}$$

Next, it is easy to see that $\mathbb{P}[N_b > n | L_b] = (1 - \bar{G}(L_b))^n$, and thus,

$$\begin{aligned}
\mathbb{P}[N_b > n] &= \mathbb{E}[1 - \bar{G}(L_b)]^n \\
&= \mathbb{E}[1 - \bar{G}(L_b)]^{n(1-\epsilon+\epsilon)} \\
&\leq (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\mathbb{E}[1 - \bar{G}(L_b)]^{n\epsilon} 1(L_b \leq x_\epsilon) + \mathbb{E}[1 - \bar{G}(L_b)]^{n\epsilon} 1(L_b > x_\epsilon)\right] \\
&\leq (1 - \bar{G}(b))^{n(1-\epsilon)} \left[(1 - \bar{G}(x_\epsilon))^{n\epsilon} + \int_{x_\epsilon}^{b} \left(1 - \bar{G}(x)\right)^{n\epsilon} \frac{dF(x)}{F(b)}\right] \\
&\leq (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} + \int_{0}^{b} \left(1 - \bar{F}(x)^{\frac{1}{\alpha(1-\epsilon)}}\right)^{n\epsilon} \frac{dF(x)}{F(b)}\right],
\end{aligned}$$

where $\eta_{x_\epsilon} = 1 - \bar{G}(x_\epsilon)$, and the last inequality follows from (2.3); in case $x_\epsilon \geq b \geq b_0$, the integral in the second inequality is zero and the last inequality trivially holds. Now, by extending the preceding integral to ∞, we obtain

$$\begin{aligned}
\mathbb{P}[N_b > n] &\leq \frac{1}{F(b)} (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} F(b) + \int_{0}^{\infty} \left(1 - \bar{F}(x)^{\frac{1}{\alpha(1-\epsilon)}}\right)^{n\epsilon} dF(x)\right] \\
&= \frac{1}{F(b)} (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} F(b) + \mathbb{E}\left(1 - \bar{F}(L)^{\frac{1}{\alpha(1-\epsilon)}}\right)^{n\epsilon}\right] \\
&\leq \frac{1}{F(b)} (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} F(b) + \mathbb{E}\, e^{-\bar{F}(L)^{\frac{1}{\alpha(1-\epsilon)}} n\epsilon}\right],
\end{aligned}$$

where we use the elementary inequality $1 - x \leq e^{-x}$, $x \geq 0$, and thus

$$\mathbb{P}[N_b > n] \leq \frac{1}{F(b)}(1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} F(b) + \mathbb{E}\, e^{-U^{\frac{1}{\alpha(1-\epsilon)}} n\epsilon}\right],$$

by $\bar{F}(L) = U$, where U is uniformly distributed on $[0, 1]$ by Proposition 2.1 in Chapter 10 of [21]. Hence,

$$\begin{aligned}
\mathbb{P}[N_b > n] &\leq \frac{1}{F(b)}(1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} F(b) + \int_0^1 e^{-x^{\frac{1}{\alpha(1-\epsilon)}} n\epsilon} dx\right] \\
&= \frac{1}{F(b)}(1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} F(b) + \int_0^{n\epsilon} \frac{\alpha(1-\epsilon)}{(n\epsilon)^{\alpha(1-\epsilon)}} e^{-z} z^{\alpha(1-\epsilon)-1} dz\right] \\
&\leq \frac{1}{F(b)}(1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} F(b) + \frac{\alpha(1-\epsilon)}{(n\epsilon)^{\alpha(1-\epsilon)}} \int_0^{\infty} e^{-z} z^{\alpha(1-\epsilon)-1} dz\right] \\
&= \frac{1}{F(b)}(1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} F(b) + \frac{\alpha(1-\epsilon)}{(n\epsilon)^{\alpha(1-\epsilon)}} \Gamma(\alpha(1-\epsilon))\right],
\end{aligned}$$

which follows from the definition of the Gamma function $\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$. Therefore,

$$\begin{aligned}
\mathbb{P}[N_b > n] &\leq (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} + \frac{\alpha(1-\epsilon)}{F(b)(n\epsilon)^{\alpha(1-\epsilon)}} \Gamma(\alpha(1-\epsilon))\right] \\
&\leq (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} + \frac{\alpha(1-\epsilon)\epsilon^{-\alpha(1-\epsilon)}}{F(b_0) n^{\alpha(1-\epsilon)}} \Gamma(\alpha(1-\epsilon))\right] \\
&= (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\eta_{x_\epsilon}^{n\epsilon} + \frac{H_\epsilon}{n^{\alpha(1-\epsilon)}}\right],
\end{aligned}$$

since $b \geq b_0$, whereas, in the last inequality, we set $H_\epsilon = \alpha(1-\epsilon)\epsilon^{-\alpha(1-\epsilon)}\Gamma(\alpha(1-\epsilon))/F(b_0)$. Now, we can choose n_0, such that for any $\epsilon > 0$ and for all $n \geq n_0$, $\eta_{x_\epsilon}^{n\epsilon} \leq \epsilon H_\epsilon n^{-\alpha(1-\epsilon)}$, so that

$$\begin{aligned}
\mathbb{P}[N_b > n] &\leq (1 - \bar{G}(b))^{n(1-\epsilon)} \left[\epsilon \frac{H_\epsilon}{n^{\alpha(1-\epsilon)}} + \frac{H_\epsilon}{n^{\alpha(1-\epsilon)}}\right] \\
&= (1 - \bar{G}(b))^{n(1-\epsilon)} \frac{H_\epsilon}{n^{\alpha(1-\epsilon)}}(1 + \epsilon),
\end{aligned}$$

and by taking the logarithm in the preceding expression, we obtain

$$\begin{aligned}\log \mathbb{P}[N_b > n] &\leq \log\left(H_\epsilon(1+\epsilon)\right) + n(1-\epsilon)\log(1-\bar{G}(b)) - \alpha(1-\epsilon)\log n \\ &= \log\left(H_\epsilon(1+\epsilon)\right) + (1-\epsilon)\left[n\log(1-\bar{G}(b)) - \alpha\log n\right].\end{aligned}$$

Next, since $-n\log(1-\bar{G}(b)) > 0$ and $\alpha\log n > 0$, $n > 1$,

$$\begin{aligned}\frac{\log \mathbb{P}[N_b > n]}{-n\log(1-\bar{G}(b)) + \alpha\log n} &\leq \frac{\log\left(H_\epsilon(1+\epsilon)\right)}{-n\log(1-\bar{G}(b)) + \alpha\log n} - (1-\epsilon) \\ &\leq \frac{\log\left(H_\epsilon(1+\epsilon)\right)}{\alpha\log n} - (1-\epsilon)\end{aligned}$$

and $\alpha\log n$ being increasing in n, we can choose n_0 such that for any $n \geq n_0$,

$$\frac{\log\left(H_\epsilon(1+\epsilon)\right)}{\alpha\log n} \leq \epsilon.$$

Thus,

$$\frac{\log \mathbb{P}[N_b > n]}{-n\log(1-\bar{G}(b)) + \alpha\log n} \leq -(1-2\epsilon),$$

which completes the proof by replacing ϵ with $\epsilon/2$. □

Next, we determine the region where the power law asymptotics holds on the logarithmic scale.

Theorem 2.1. *If*

$$\log \mathbb{P}[L > x] \sim \alpha \log \mathbb{P}[A > x] \qquad \text{as } x \to \infty, \tag{2.4}$$

$\alpha > 0$, *then, for any* $\epsilon > 0$, *there exists positive* n_0, *such that for all* $n \geq n_0$, *for which* $n^{1+\epsilon}\mathbb{P}[A > b] \leq 1$, *we have*

$$\left|\frac{-\log \mathbb{P}[N_b > n]}{\alpha\log n} - 1\right| \leq \epsilon. \tag{2.5}$$

Note that this result appeared in Theorem 1 of [39]. The proof can be found in Section 4 of [2]; see also [1].

This result holds in the region $n^{1+\epsilon} \leq O(1/\bar{G}(b))$. Also, note that one can easily

characterize the logarithmic asymptotics of the very end of the exponential tail of $\mathbb{P}[N_b > n]$ for small b and large n. In particular, for fixed b, it can be shown that $\log \mathbb{P}[N_b > n] \sim n \log(1 - \bar{G}(b))$ as $n \to \infty$, see Theorem 1 in [39]. However, our objective is to determine the entire body of the distribution of $\mathbb{P}[N_b > n]$ uniformly in n and b.

Next, we extend Theorem 2.1 to the entire region $n \geq n_0, b \geq b_0$, which includes the geometric term $\mathbb{P}[A \leq b]^n$. For this theorem, we need slightly more restrictive assumptions. The reason why this is the case is that $\mathbb{P}[N_b > n]$ behaves like a power law in the region where $n = o(\log n/\bar{G}(b))$, while for $n >> \log n/\bar{G}(b)$, it follows essentially a geometric distribution; see Theorem 2.2 below. Hence, more restrictive assumptions are required since the geometric distribution is much more sensitive to the changes in its parameters (informally, $((1+\epsilon)x)^{-\alpha} \approx x^{-\alpha}$ but $e^{-(1+\epsilon)x} \not\approx e^{-x}$).

Definition 2.2.1. *A function $\ell(x)$ is slowly varying if $\ell(x)/\ell(\lambda x) \to 1$ as $x \to \infty$ for any fixed $\lambda > 0$.*

If not directly implied by our assumptions, $\ell(x)$ is assumed positive and locally bounded.

Theorem 2.2. *If $\mathbb{P}[L > x] = \ell(\mathbb{P}[A > x]^{-1})\mathbb{P}[A > x]^{\alpha}$, for $\alpha > 0$, $\ell(x)$ slowly varying, then for any $\epsilon > 0$, there exist n_0, b_0, such that for all $n \geq n_0, b \geq b_0$,*

$$\left| \frac{-\log \mathbb{P}[N_b > n]}{-n \log \mathbb{P}[A \leq b] + \alpha \log n} - 1 \right| \leq \epsilon. \tag{2.6}$$

The proof appears in [1] as well as its extended version; see Section 4 of [2].

Remark 2. *Note that the statement of this theorem can be formulated in an equivalent form*

$$\left| \frac{-\log \mathbb{P}[N_b > n]}{n\mathbb{P}[A > b] + \alpha \log n} - 1 \right| \leq \epsilon,$$

since $-n \log \mathbb{P}[A \leq b] \sim n\mathbb{P}[A > b]$ as $b \to \infty$.

Remark 3. *This theorem extends Theorem 1 in [39]. In particular, it proves the result uniformly in n and b, while Theorem 1 in [39] characterized the initial power law part of*

the distribution ($n \leq \bar{G}(b)^{-\eta}, 0 < \eta < 1$) and the very end with exponential tail (fixed b, $n \to \infty$).

2.3 Exact Asymptotics

In this section, we derive the exact approximation for $\mathbb{P}[N_b > n]$ that works uniformly for all n, b sufficiently large (Theorems 2.3 and 2.4). As noted earlier in the introduction, this characterization is explicit in that it is a product of a power law and the Gamma distribution

$$\mathbb{P}[N_b > n] \approx \frac{\alpha}{n^\alpha \ell(n \wedge \mathbb{P}[A > b]^{-1})} \int_{-n \log \mathbb{P}[A \leq b]}^{\infty} e^{-z} z^{\alpha-1} dz, \tag{2.7}$$

where $x \wedge y = \min(x, y)$ and $\ell(\cdot)$ is slowly varying. Implicitly, the argument of $\ell(x)$ is altered depending on whether $n\mathbb{P}[A > b] \leq C$ or $n\mathbb{P}[A > b] > C$ for some constant C. Hence, we can choose $C = 1$ since $\ell(n \wedge 1/\mathbb{P}[A > b]) \approx \ell(n \wedge C/\mathbb{P}[A > b])$ for large n, b. Note that when $-n \log \mathbb{P}[A \leq b] \downarrow 0$, the power law dominates, whereas when $-n \log \mathbb{P}[A \leq b] \to \infty$, the integral determines the tail with the geometric (exponential) leading term.

We would like to point out that approximation (2.7) actually works well when $\mathbb{P}[A > b]^{-1}$ is large rather than b; this can be concluded by examining the proofs of the theorems in this section. Hence, formula (2.7) can be accurate for relatively small values of b provided that A is light tailed. This may be the reason why we obtain accurate results in our simulation examples in Section 2.4 for small values of b.

First, in Theorem 2.3, we precisely describe the region where the distribution of N_b exhibits the power law behavior, $n\mathbb{P}[A > b] \leq C$, for any fixed constant C. Then, Theorem 2.4 covers the remaining region, $n\mathbb{P}[A > b] > C$, where $\mathbb{P}[N_b > n]$ approaches the geometric tail. Additional discussion of the results and the treatment of some special cases are presented at the end of this section; see Propositions 2.2 and 2.3.

Theorem 2.3. *Let $\mathbb{P}[L > x]^{-1} = \ell(\mathbb{P}[A > x]^{-1})\mathbb{P}[A > x]^{-\alpha}$, $\alpha > 0$, $x \geq 0$, and $C > 0$ be a fixed constant. Then, for any $\epsilon > 0$, there exists n_0 such that for all $n > n_0$, and*

$n\mathbb{P}[A > b] \leq C$,

$$\left| \frac{\mathbb{P}[N_b > n] n^\alpha \ell(n)}{\alpha \Gamma(-n \log \mathbb{P}[A \leq b], \alpha)} - 1 \right| \leq \epsilon, \tag{2.8}$$

where $\Gamma(x, \alpha)$ *is the incomplete Gamma function defined as* $\int_x^\infty e^{-z} z^{\alpha-1} dz$.

Remark 4. *Related result was derived in Theorem 3 of [39] where it was required that* $n \leq \bar{G}(b)^{-\eta}, 0 < \eta < 1$. *Note that here we broaden the region where the result holds by requiring* $n \leq C/\bar{G}(b)$, *which is larger than* $n \leq \bar{G}(b)^{-\eta}$. *Furthermore, this is the largest region where the exact power law asymptotics* $O(n^{-\alpha}/\ell(n))$ *holds since for* $n\bar{G}(b) > C, \Gamma(n\bar{G}(b), \alpha) \leq \Gamma(C, \alpha) \to 0$ *as* $C \to \infty$.

Remark 5. *Note here that the incomplete Gamma function* $\Gamma(\alpha, x) = \int_x^\infty z^{\alpha-1} e^{-z} dz$ *can be easily computed using the well known asymptotic approximation (see Sections 6.5.32 in [30]), as* $x \to \infty$,

$$\Gamma(\alpha, x) \sim x^{\alpha-1} e^{-x} \left[1 + \frac{\alpha - 1}{x} + \frac{(\alpha-1)(\alpha-2)}{x^2} + \dots \right].$$

Proof. This proof uses some of the ideas from the proof of Theorem 2.1 in [10]. However, it is much more involved since one has to incorporate the assumption $n\mathbb{P}[A > b] \leq C$, which ensures the power law body.

Let $\Phi(x) = \ell(x) x^\alpha$. Then, $\Phi(x)$ is regularly varying with index α and, thus, for any $c > 0$,

$$\lim_{x \to \infty} \frac{\Phi(cx)}{\Phi(x)} = c^\alpha < \infty,$$

and, in particular, we can choose $c = e$, which implies that there exists n_ϵ such that for $n/e^k > n_\epsilon$,

$$\frac{\Phi(n)}{\Phi(n/e^k)} \leq e^{k(\alpha+1)}. \tag{2.9}$$

Without loss of generality, we may assume that $\Phi(.)$ is eventually absolutely continuous, strictly monotone and locally bounded for $x > 0$ since we can always find an absolutely

continuous and strictly monotone function

$$\Phi^*(x) = \begin{cases} \alpha \int_1^x \Phi(s) s^{-1} ds, & x \geq 1 \\ 0, & 0 \leq x < 1, \end{cases} \tag{2.10}$$

which for x large enough satisfies

$$\bar{F}(x)^{-1} = \Phi(\bar{G}(x)^{-1}) \sim \Phi^*(\bar{G}(x)^{-1}).$$

This implies that, for any $0 < \epsilon < 1$ and $x \geq x_0$, we have

$$1/\Phi^{\leftarrow}\left((1+\epsilon)\bar{F}(x)^{-1}\right) \leq \bar{G}(x) \leq 1/\Phi^{\leftarrow}\left((1-\epsilon)\bar{F}(x)^{-1}\right), \tag{2.11}$$

where $\Phi^{\leftarrow}(\cdot)$ denotes the inverse function of $\Phi^*(\cdot)$; note that the monotonicity of $\Phi^*(x)$, for all $x \geq 1$, guarantees that its inverse exists. To simplify the notation in this proof, we shall use $\Phi(\cdot)$ to denote $\Phi^*(\cdot)$. Furthermore, $\Phi^{\leftarrow}(\cdot)$ is regularly varying with index $1/\alpha$ (see Theorem 1.5.12 in [20]), implying that

$$\Phi^{\leftarrow}((1+\epsilon)x) \sim \left(\frac{1+\epsilon}{1-\epsilon}\right)^{1/\alpha} \Phi^{\leftarrow}((1-\epsilon)x),$$

as $x \to \infty$. Therefore, for $\eta_\epsilon = \eta(\epsilon) = [(1+\epsilon)/(1-\epsilon)]^{2/\alpha}$ and x large,

$$\eta_\epsilon^{-1}\bar{G}(x) \leq 1/\Phi^{\leftarrow}\left((1+\epsilon)\bar{F}(x)^{-1}\right) \leq 1/\Phi^{\leftarrow}\left((1-\epsilon)\bar{F}(x)^{-1}\right) \leq \eta_\epsilon \bar{G}(x). \tag{2.12}$$

First, notice that the number of retransmissions is geometrically distributed given the data size L_b,

$$\begin{aligned} \mathbb{P}[N_b > n] &= \mathbb{E}[1-\bar{G}(L_b)]^n \\ &= \mathbb{E}[1-\bar{G}(L_b)]^n \mathbf{1}(L_b \leq x_0) + \mathbb{E}[1-\bar{G}(L_b)]^n \mathbf{1}(L_b > x_0). \end{aligned} \tag{2.13}$$

We begin with the *lower bound*. For $H > C$ and x_0 as in (2.13), we choose $x_n > x_0$ such that $\Phi^{\leftarrow}((1-\epsilon)\bar{F}(x_n)^{-1}) = n/H$, for n large, and thus,

$$\begin{aligned}
\mathbb{P}[N_b > n] &= \mathbb{E}[1-\bar{G}(L_b)]^n \\
&\geq \mathbb{E}\left[(1-\bar{G}(L_b))^n \mathbf{1}(L_b > x_n)\right] \\
&\geq \mathbb{E}\left[\left(1 - \frac{1}{\Phi^{\leftarrow}((1-\epsilon)\bar{F}(L_b)^{-1})}\right)^n \mathbf{1}(L_b > x_n)\right] \\
&= \int_{x_n}^{b} \left(1 - \frac{1}{\Phi^{\leftarrow}((1-\epsilon)\bar{F}(x)^{-1})}\right)^n \frac{dF(x)}{F(b)},
\end{aligned}$$

where we use our main assumption (2.11). Now, since $F(b) \leq 1$ and using the continuity of $F(x)$ and change of variables $z = n/\Phi^{\leftarrow}((1-\epsilon)\bar{F}(x)^{-1})$, we obtain,

$$\begin{aligned}
\mathbb{P}[N_b > n] &\geq \int_{n/\Phi^{\leftarrow}((1-\epsilon)\bar{F}(b)^{-1})}^{H} \left(1-\frac{z}{n}\right)^n \frac{\Phi'(n/z)}{\Phi^2(n/z)} \frac{(1-\epsilon)n}{z^2} dz \\
&\geq \int_{\eta_\epsilon n\bar{G}(b)}^{H} \left(1-\frac{z}{n}\right)^n \frac{\Phi'(n/z)}{\Phi^2(n/z)} \frac{(1-\epsilon)n}{z^2} dz,
\end{aligned}$$

where we use that $\eta_\epsilon \bar{G}(b) \geq 1/\Phi^{\leftarrow}((1-\epsilon)\bar{F}(b)^{-1})$ from (2.12), which holds for b large, or equivalently n large by our assumption $n\bar{G}(b) \leq C$. Now, we consider two distinct cases:

If $\eta_\epsilon n\bar{G}(b) < h$, where $h > 0$ is a small constant, then

$$\begin{aligned}
\mathbb{P}[N_b > n] &\geq (1-\epsilon) \int_{h}^{H} \left(1-\frac{z}{n}\right)^n \frac{\Phi'(n/z)}{\Phi^2(n/z)} \frac{n}{z^2} dz \\
&\geq (1-\epsilon)^{3/2} \frac{\alpha}{\Phi(n)} \int_{h}^{H} \left(1-\frac{z}{n}\right)^n z^{\alpha-1} dz,
\end{aligned}$$

where we use the properties of regularly varying functions that for all $h \leq z \leq H$ and large n,

$$\frac{\Phi(n)}{\Phi(n/z)} \geq (1-\epsilon)^{1/2} z^{\alpha},$$

and

$$\Phi'(n/z)/\Phi(n/z) = \frac{\alpha z}{n},$$

for $n > H$ [see (3.9)]. Next, using $1 - x \geq e^{-(1+\delta)x}$ for $\delta > 0$ and $0 \leq x \leq x_\delta$, for n large enough ($n > H/x_\delta$) we obtain

$$\begin{aligned}\mathbb{P}[N_b > n] &\geq (1-\epsilon)^{3/2}\frac{\alpha}{\Phi(n)}\int_h^H e^{-(1+\delta)z}z^{\alpha-1}dz \\ &\geq (1-\epsilon)^{3/2}e^{-\delta H}\frac{\alpha}{\Phi(n)}\int_h^H e^{-z}z^{\alpha-1}dz,\end{aligned}$$

and by choosing $\delta > 1/H$ so that $e^{-\delta H} \geq (1-\epsilon)^{1/2}$, we have

$$\begin{aligned}\mathbb{P}[N_b > n] &\geq (1-\epsilon)^2\frac{\alpha}{\Phi(n)}\int_h^H e^{-z}z^{\alpha-1}dz \\ &\geq (1-\epsilon)^2\frac{\alpha}{\Phi(n)}\left[\int_{n\bar{G}(b)}^H e^{-z}z^{\alpha-1}dz - \int_{n\bar{G}(b)}^h e^{-z}z^{\alpha-1}dz\right] \\ &\geq (1-\epsilon)^2\frac{\alpha}{\Phi(n)}\left[\int_{n\bar{G}(b)}^H e^{-z}z^{\alpha-1}dz - \int_0^h e^{-z}z^{\alpha-1}dz\right]\end{aligned}$$

$$\begin{aligned}&\geq (1-\epsilon)^2\frac{\alpha}{\Phi(n)}\left[\int_{n\bar{G}(b)}^\infty e^{-z}z^{\alpha-1}dz - \int_H^\infty e^{-z}z^{\alpha-1}dz - \frac{h^\alpha}{\alpha}\right] \\ &\geq (1-\epsilon)^2\frac{\alpha}{\Phi(n)}\left[\int_{n\bar{G}(b)}^\infty e^{-z}z^{\alpha-1}dz - 2e^{-H}H^{\alpha-1} - \frac{h^\alpha}{\alpha}\right] \\ &= (1-\epsilon)^2\frac{\alpha}{\Phi(n)}\Gamma(n\bar{G}(b),\alpha)\left(1 - \frac{2e^{-H}H^{\alpha-1} + h^\alpha/\alpha}{\Gamma(C,\alpha)}\right),\end{aligned}$$

where the second to last inequality follows from the approximation for the incomplete gamma function for large H [see Remark 5 of Theorem 4.8] and the last inequality uses the assumption $n\bar{G}(b) \leq C$. Now, picking H, h such that $2e^{-H}H^{\alpha-1} + h^\alpha/\alpha \leq \epsilon\Gamma(C,\alpha)$, yields

$$\mathbb{P}[N_b > n] \geq (1-3\epsilon)\frac{\alpha}{\Phi(n)}\Gamma(n\bar{G}(b),\alpha).$$

If $h \leq \eta_\epsilon n\bar{G}(b) \leq C$, then

$$\begin{aligned}\mathbb{P}[N_b > n] &\geq (1-\epsilon)\int_{\eta_\epsilon n\bar{G}(b)}^{H} e^{-z}\frac{\Phi'(n/z)}{\Phi^2(n/z)}\frac{n}{z^2}dz \\ &\geq (1-\epsilon)^2\frac{\alpha}{\Phi(n)}\int_{\eta_\epsilon n\bar{G}(b)}^{H} e^{-z}z^{\alpha-1}dz,\end{aligned}$$

which follows from the regularly varying properties in the region $h/\eta_\epsilon < n\bar{G}(b) \leq z \leq H$.

For the preceding integral, similarly as before, we have

$$\begin{aligned}\int_{\eta_\epsilon n\bar{G}(b)}^{H} e^{-z}z^{\alpha-1}dz &= \int_{n\bar{G}(b)}^{H} e^{-z}z^{\alpha-1}dz - \int_{n\bar{G}(b)}^{\eta_\epsilon n\bar{G}(b)} e^{-z}z^{\alpha-1}dz \\ &\geq \int_{n\bar{G}(b)}^{H} e^{-z}z^{\alpha-1}dz - \int_{n\bar{G}(b)}^{\eta_\epsilon n\bar{G}(b)} z^{\alpha-1}dz \\ &= \int_{n\bar{G}(b)}^{H} e^{-z}z^{\alpha-1}dz - (n\bar{G}(b))^\alpha\frac{\eta_\epsilon^\alpha - 1}{\alpha} \\ &\geq \int_{n\bar{G}(b)}^{\infty} e^{-z}z^{\alpha-1}dz - \int_{H}^{\infty} e^{-z}z^{\alpha-1}dz - C^\alpha\frac{\eta_\epsilon^\alpha - 1}{\alpha} \\ &\geq \int_{n\bar{G}(b)}^{\infty} e^{-z}z^{\alpha-1}dz - 2e^{-H}H^{\alpha-1} - \frac{4\epsilon C^\alpha}{\alpha} \\ &= \Gamma(n\bar{G}(b),\alpha)\left(1 - \frac{2e^{-H}H^{\alpha-1} + 4\epsilon C^\alpha/\alpha}{\Gamma(C,\alpha)}\right),\end{aligned}$$

where we use the approximation for the incomplete gamma function for large H, that $\eta_\epsilon^\alpha - 1 \to 4\epsilon$ as $\epsilon \to 0$ and $n\bar{G}(b) \leq C$. Now, letting H be such that $2e^{-H}H^{\alpha-1} + 4\epsilon C^\alpha/\alpha \leq \sqrt{\epsilon}\Gamma(C,\alpha)$ yields

$$\mathbb{P}[N_b > n] \geq (1-\epsilon)^2(1-\sqrt{\epsilon})\frac{\alpha}{\Phi(n)}\Gamma(n\bar{G}(b),\alpha).$$

Finally, since $\bar{G}(b) \leq -\log(1-\bar{G}(b))$, we obtain

$$\mathbb{P}[N_b > n] \geq (1-\epsilon)^2(1-\sqrt{\epsilon})\frac{\alpha}{\Phi(n)}\Gamma(-n\log(1-\bar{G}(b)),\alpha), \tag{2.14}$$

which proves the lower bound after replacing $(1-\epsilon)^2(1-\sqrt{\epsilon})$ with $1-\epsilon$.

Next, we derive the *upper bound*. Note that for x_0 as in (2.13),

$$\begin{aligned}\mathbb{P}[N_b > n] &= \mathbb{E}[1-\bar{G}(L_b)]^n \\ &\leq (1-\bar{G}(x_0))^n + \mathbb{E}[1-\bar{G}(L_b)]^n \mathbf{1}(L_b > x_0) \\ &\leq e^{-n\bar{G}(x_0)} + \mathbb{E}\left(1 - \frac{1}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(L_b)^{-1})}\right)^n \mathbf{1}(L_b > x_0), \end{aligned} \tag{2.15}$$

which follows from (2.11) and the elementary inequality $1-x \leq e^{-x}$. Now, for any $H > \max(C,1)$, we obtain

$$\begin{aligned}&\mathbb{P}[N_b > n] \leq e^{-n\bar{G}(x_0)} + \mathbb{E}\left(1 - \frac{1}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(L_b)^{-1})}\right)^n \mathbf{1}\left(L_b > x_0\right) \\ &\leq e^{-n\bar{G}(x_0)} + \mathbb{E}\left(1 - \frac{1}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(L_b)^{-1})}\right)^n \mathbf{1}\left(\frac{1}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(L_b)^{-1})} < \frac{H}{n}\right) \\ &+ \sum_{k=\lfloor \log H \rfloor}^{\lceil \log(n/n_\epsilon)\rceil} e^{-e^k} \mathbb{P}\left[e^k \leq \frac{n}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(L_b)^{-1})} \leq e^{k+1}\right] + e^{-n/n_\epsilon} \\ &\triangleq I_0 + I_1 + I_2 + I_3. \end{aligned} \tag{2.16}$$

First, we upper bound I_1 in (2.16), which equals

$$\begin{aligned} I_1 &= \frac{1}{F(b)} \int_0^b \left(1 - \frac{1}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(x)^{-1})}\right)^n \mathbf{1}\left(\frac{n}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(x)^{-1})} < H\right) dF(x) \\ &= \frac{1+\epsilon}{\Phi(n)F(b)} \int_{n/\Phi^{\leftarrow}((1+\epsilon)\bar{F}(b)^{-1})}^{H} \left(1-\frac{z}{n}\right)^n \frac{\Phi(n)}{\Phi(n/z)} \frac{\Phi'(n/z)}{\Phi(n/z)} \frac{n}{z^2} dz \\ &\leq \frac{1+\epsilon}{\Phi(n)F(b)} \int_{n\bar{G}(b)/\eta_\epsilon}^{H} \left(1-\frac{z}{n}\right)^n \frac{\Phi(n)}{\Phi(n/z)} \frac{\Phi'(n/z)}{\Phi(n/z)} \frac{n}{z^2} dz, \end{aligned}$$

where we use the change of variables $z = n/\Phi^{\leftarrow}((1+\epsilon)F(x)^{-1})$ and the absolute continuity of $F(x)$. For the last inequality, observe that $1/\Phi^{\leftarrow}((1+\epsilon)\bar{F}(b)^{-1}) \geq \bar{G}(b)/\eta_\epsilon$ from (2.12). Now, similarly as before, we consider two cases:

If $n\bar{G}(b) < \eta_\epsilon h_\epsilon \leq C$, where $h_\epsilon > 0$ is a small constant, I_1 is upper bounded by

$$I_1 \leq \frac{1+\epsilon}{F(b)\Phi(n)} \int_{h_\epsilon}^{H} \left(1-\frac{z}{n}\right)^n \frac{\Phi(n)}{\Phi(n/z)} \frac{\Phi'(n/z)}{\Phi(n/z)} \frac{n}{z^2} dz + \mathbb{P}\left(\frac{n}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(L_b)^{-1})} < h_\epsilon\right). \tag{2.17}$$

Now, since $\Phi(.)$ is absolutely continuous and regularly varying, it follows that for all $h_\epsilon \leq z \leq H$,

$$\frac{\Phi(n)}{\Phi(n/z)} \leq (1+\epsilon)^{1/2} z^\alpha,$$

for large n, and, by (3.9),

$$\frac{\Phi'(n/z)}{\Phi(n/z)} = \frac{\alpha z}{n},$$

for $n > H$.

Next, we compute the second term in (2.17) as

$$\begin{aligned}\mathbb{P}\left(\bar{F}(L_b) < \frac{1+\epsilon}{\Phi(n/h_\epsilon)}\right) &\leq \int_0^\infty \mathbf{1}\left(\bar{F}(x) < \frac{1+\epsilon}{\Phi(n/h_\epsilon)}\right) \frac{dF(x)}{F(b)} = \frac{1}{F(b)} \mathbb{P}\left[\bar{F}(L) < \frac{1+\epsilon}{\Phi(n/h_\epsilon)}\right] \\ &\leq \frac{1+\epsilon}{F(b)\Phi(n/h_\epsilon)} \leq \frac{(1+\epsilon)^2 h_\epsilon^\alpha}{\Phi(n)},\end{aligned}$$

which follows from the uniform distribution of $\bar{F}(L)$ and using $\Phi(n)/\Phi(n/h_\epsilon) \leq (1+\epsilon)^{1/2} h_\epsilon^\alpha$ for large n, along with $F(b)^{-1} \leq (1+\epsilon)^{1/2}$. Now, observe that the first term in (2.17) is upper bounded by

$$\frac{\alpha(1+\epsilon)^2}{\Phi(n)} \int_{h_\epsilon}^{H} \left(1-\frac{z}{n}\right)^n z^{\alpha-1} dz \leq \frac{\alpha(1+\epsilon)^2}{\Phi(n)} \int_{n\bar{G}(b)/\eta_\epsilon}^{H} \left(1-\frac{z}{n}\right)^n z^{\alpha-1} dz,$$

since $n\bar{G}(b) < h_\epsilon \eta_\epsilon$. Also, for the integral we obtain

$$\begin{aligned}\int_{n\bar{G}(b)/\eta_\epsilon}^{H}\left(1-\frac{z}{n}\right)^n z^{\alpha-1}dz &= \int_{n\bar{G}(b)}^{H}\left(1-\frac{z}{n}\right)^n z^{\alpha-1}dz + \int_{n\bar{G}(b)/\eta_\epsilon}^{n\bar{G}(b)}\left(1-\frac{z}{n}\right)^n z^{\alpha-1}dz\\ &\le \int_{n\bar{G}(b)}^{H}\left(1-\frac{z}{n}\right)^n z^{\alpha-1}dz + \int_{n\bar{G}(b)/\eta_\epsilon}^{n\bar{G}(b)} z^{\alpha-1}dz\\ &\le \int_{n\bar{G}(b)}^{H}\left(1-\frac{z}{n}\right)^n z^{\alpha-1}dz + (n\bar{G}(b))^\alpha(1-\eta_\epsilon^{-\alpha})/\alpha\\ &\le \int_{n\bar{G}(b)}^{H}\left(1-\frac{z}{n}\right)^n z^{\alpha-1}dz + 5C^\alpha\epsilon/\alpha,\end{aligned}$$

after observing that $1-\eta_\epsilon^{-\alpha} \to 4\epsilon$ as $\epsilon \to 0$. Now, by changing the variables $1-z/n = e^{-u/n}$, we have

$$\begin{aligned}I_1 &\le \frac{\alpha(1+\epsilon)^2}{\Phi(n)}\int_{n\bar{G}(b)}^{H}\left(1-\frac{z}{n}\right)^n z^{\alpha-1}dz + \frac{(1+\epsilon)^2 5C^\alpha\epsilon}{\Phi(n)} + \frac{(1+\epsilon)^2 h_\epsilon^\alpha}{\Phi(n)}\\ &\le \frac{\alpha(1+\epsilon)^2}{\Phi(n)}\int_{-n\log(1-\bar{G}(b))}^{-n\log(1-H/n)} e^{-u}(1-e^{-u/n})^{\alpha-1}n^{\alpha-1}e^{-u/n}du + \frac{(1+\epsilon)^2(5C^\alpha\epsilon+h_\epsilon^\alpha)}{\Phi(n)}\\ &\le \frac{\alpha(1+\epsilon)^2}{\Phi(n)}\int_{-n\log(1-\bar{G}(b))}^{\infty} e^{-u}u^{\alpha-1}du + \frac{(1+\epsilon)^2}{\Phi(n)}(5C^\alpha\epsilon+h_\epsilon^\alpha)\\ &\le \frac{\alpha(1+\epsilon)^2}{\Phi(n)}\int_{-n\log(1-\bar{G}(b))}^{\infty} e^{-u}u^{\alpha-1}du\left[1+\frac{5C^\alpha\epsilon+h_\epsilon^\alpha}{\alpha\Gamma(2C,\alpha)}\right] \qquad (2.18)\\ &\le \frac{\alpha(1+\epsilon)^2(1+\sqrt{\epsilon})}{\Phi(n)}\int_{-n\log(1-\bar{G}(b))}^{\infty} e^{-u}u^{\alpha-1}du,\end{aligned}$$

where, in the second inequality, we use $e^{-u/n} \le 1$, the inequality $1-e^{-x} \le x$, $x \ge 0$ and extend the integral to infinity. Last, we pick ϵ small, such that $5C^\alpha\epsilon + h_\epsilon^\alpha \le \sqrt{\epsilon}\alpha\Gamma(2C,\alpha)$.

Note that the preceding equation along with (2.14) imply that I_1 is lower bounded as $I_1 \ge (1-\epsilon)\alpha\Gamma(2n\bar{G}(b),\alpha)/\Phi(n) \ge (\alpha/2)\Gamma(2C,\alpha)/\Phi(n)$, for all $n > n_0$ and $\epsilon < 1/2$, by the inequality $1-x \ge e^{-2x}$ for $x \ge 0$ small, since by assumption $n\bar{G}(b) \le C$, i.e., $\bar{G}(b)$ is small.

If $h_\epsilon\eta_\epsilon \le n\bar{G}(b) \le C$, we have

$$I_1 \le \frac{1+\epsilon}{F(b)\Phi(n)}\int_{n\bar{G}(b)/\eta_\epsilon}^{H}\left(1-\frac{z}{n}\right)^n \frac{\Phi(n)}{\Phi(n/z)}\frac{\Phi'(n/z)}{\Phi(n/z)}\frac{n}{z^2}dz,$$

and, by the properties of regularly varying functions in the interval $n/H \le n/z \le 1/\bar{G}(b) \le$

n/h_ϵ, for $H > C$, and using the same arguments as in (2.18), we have

$$\begin{aligned} I_1 &\leq \frac{\alpha(1+\epsilon)^2}{\Phi(n)} \int_{n\bar{G}(b)/\eta_\epsilon}^{H} \left(1-\frac{z}{n}\right)^n z^{\alpha-1} dz \\ &\leq \frac{\alpha(1+\epsilon)^2}{\Phi(n)} \int_{-n\log(1-\bar{G}(b))}^{\infty} e^{-z} z^{\alpha-1} dz \left[1 + \frac{5C^\alpha \epsilon}{\alpha\Gamma(2C,\alpha)}\right] \\ &\leq \frac{\alpha(1+\epsilon)^2(1+\sqrt{\epsilon})}{\Phi(n)} \int_{-n\log(1-\bar{G}(b))}^{\infty} e^{-z} z^{\alpha-1} dz. \end{aligned}$$

Therefore, from both cases, it follows that for all $n > n_0$,

$$I_1 \leq \frac{\alpha(1+\epsilon)}{\Phi(n)} \Gamma(-n\log(1-\bar{G}(b)), \alpha), \tag{2.19}$$

after replacing $(1+\epsilon)^2(1+\sqrt{\epsilon})$ with $1+\epsilon$.

Next, we evaluate the second term in (2.16) as

$$\begin{aligned} I_2 &= \sum_{k=\lfloor \log H \rfloor}^{\lceil \log(n/n_\epsilon) \rceil} e^{-e^k} \mathbb{P}\left[e^k \leq \frac{n}{\Phi^{\leftarrow}((1+\epsilon)\bar{F}(L_b)^{-1})} \leq e^{k+1}\right] \\ &= \sum_{k=\lfloor \log H \rfloor}^{\lceil \log(n/n_\epsilon) \rceil} e^{-e^k} \mathbb{P}\left[(1+\epsilon)/\Phi\left(\frac{n}{e^{k+1}}\right) \leq \bar{F}(L_b) \leq (1+\epsilon)/\Phi\left(\frac{n}{e^k}\right)\right] \end{aligned}$$

$$\begin{aligned} &\leq \sum_{k=\lfloor \log H \rfloor}^{\lceil \log(n/n_\epsilon) \rceil} e^{-e^k} \int_0^\infty \mathbf{1}\left(\bar{F}(x) \leq \frac{1+\epsilon}{\Phi\left(n/e^k\right)}\right) \frac{dF(x)}{F(b)} \\ &\leq \sum_{k=\lfloor \log H \rfloor}^{\infty} e^{-e^k} \frac{1+\epsilon}{F(b)\Phi\left(n/e^k\right)}, \end{aligned}$$

which follows from the fact that the integral in the second inequality is equal to $\mathbb{P}[\bar{F}(L) \leq (1+\epsilon)/\Phi\left(n/e^k\right)]/F(b)$ and $\bar{F}(L)$ is uniform in $[0,1]$. Thus,

$$I_2 \leq \frac{1+\epsilon}{F(b)\Phi(n)} \sum_{k=\lfloor \log H \rfloor}^{\infty} e^{-e^k} \frac{\Phi(n)}{\Phi(n/e^k)} \leq \frac{1+\epsilon}{F(b)\Phi(n)} \sum_{k=\lfloor \log H \rfloor}^{\infty} e^{-e^k} e^{k(\alpha+1)},$$

where we make use of (2.9). Since the preceding sum is finite, we obtain that for large H

and all $n > n_0$,

$$I_2 \leq \frac{\epsilon}{2} I_1. \tag{2.20}$$

Last, we observe that, for fixed x_0, it follows that for $n > n_0$,

$$I_0 + I_3 = e^{-n\bar{G}(x_0)} + e^{-n/n_\epsilon} \leq \frac{\epsilon}{2} I_1. \tag{2.21}$$

Finally, using (2.19)-(2.21), we obtain for (2.16) that for all $n > n_0$,

$$\mathbb{P}[N_b > n] \leq (1+\epsilon)^2 \frac{\alpha}{\Phi(n)} \Gamma(-n \log(1 - \bar{G}(b)), \alpha),$$

which completes the proof after replacing $(1+\epsilon)$ with $(1+\epsilon)^{1/2}$. □

The following corollary is an immediate consequence of Theorem 2.3 and it represents a small generalization of Theorem 2.1 in [10].

Corollary 2.1. *If $\mathbb{P}[L > x]^{-1} = \ell(\mathbb{P}[A > x]^{-1})\mathbb{P}[A > x]^{-\alpha}$, $x \geq 0, \alpha > 0$, where $\ell(x)$ is slowly varying, then, as $n \to \infty$ and $n\mathbb{P}[A > b] \to 0$,*

$$\mathbb{P}[N_b > n] \sim \frac{\Gamma(\alpha+1)}{\ell(n) n^\alpha}. \tag{2.22}$$

Now, we characterize the remaining region where $n\mathbb{P}[A > b] > C$. Informally speaking, this is the region where $\mathbb{P}[N_b > n]$ has a lighter tail converging to the exponential when $n >> \bar{G}(b)^{-1}$. In the following theorem, we need more restrictive assumptions for $\ell(x)$; see the discussion before Theorem 2.2. In particular, we assume that $\ell(x)$ is slowly varying and eventually differentiable with $\ell'(x)x/\ell(x) \to 0$ as $x \to \infty$.

Theorem 2.4. *Assume that $\mathbb{P}[L > x]^{-1} = \ell(\mathbb{P}[A > x]^{-1})\mathbb{P}[A > x]^{-\alpha}, \alpha > 0$, $x \geq 0$, where $\ell(x)$ is slowly varying and eventually differentiable with $\ell'(x)x/\ell(x) \to 0$ as $x \to \infty$. Then, for any $\epsilon > 0$, there exist b_0, n_0, such that for all $n > n_0, b > b_0, n\mathbb{P}[A > b] > C$,*

$$\left| \frac{\mathbb{P}[N_b > n] n^\alpha \ell(\mathbb{P}[A > b]^{-1})}{\alpha \Gamma(-n \log \mathbb{P}[A \leq b], \alpha)} - 1 \right| \leq \epsilon. \tag{2.23}$$

Remark 6. *Observe that Theorems 2.3 and 2.4 cover the entire distribution* $\mathbb{P}[N_b > n]$ *for all large* n *and* b*. Interestingly, the formula for the approximation is the same except for the argument of the slowly varying part, which equals to* n *and* $\mathbb{P}[A > b]^{-1}$*, respectively. Furthermore, when* $n\mathbb{P}[A > b] = C$ *the formulas are asymptotically identical as* $\ell(n) = \ell(C\mathbb{P}[A > b]^{-1}) \sim \ell(\mathbb{P}[A > b]^{-1})$ *as* $n \to \infty$*.*

Remark 7. *Note that most well known examples of slowly varying functions satisfy the condition* $\ell'(x)x/\ell(x) \to 0$ *as* $x \to \infty$*, including* $\log^\beta x, \log^\beta(\log x), \beta > 0, \exp(\log x/\log\log x), \exp(\log^\gamma x)$*, for* $0 < \gamma < 1$ *[see Section 1.3.3 on p.16 in [20]].*

Proof. Recall that

$$\begin{aligned}\mathbb{P}[N_b > n] &= \mathbb{E}[1 - \bar{G}(L_b)]^n \\ &= \int_0^b \left(1 - \bar{G}(x)\right)^n \frac{dF(x)}{F(b)} \\ &= \int_0^{x_0} \left(1 - \bar{G}(x)\right)^n \frac{dF(x)}{F(b)} + \int_{x_0}^b \left(1 - \bar{G}(x)\right)^n \frac{dF(x)}{F(b)}. \end{aligned} \tag{2.24}$$

Now, given that $\ell(x)$ is eventually differentiable ($x \geq x_0$) and slowly varying with $\ell'(x)x/\ell(x) \to 0$ as $x \to \infty$, it follows that $d\bar{F}(x) = (1+o(1))\alpha\bar{G}(x)^{\alpha-1}\ell^{-1}(1/\bar{G}(x))d\bar{G}(x)$ as $x \to \infty$. Thus, for any $0 < \epsilon < 1$, we can select x_0 large enough such that

$$\begin{aligned}\mathbb{P}[N_b > n] &\leq \left(1 - \bar{G}(x_0)\right)^n - (1+\epsilon)^{1/2} \int_{x_0}^b (1 - \bar{G}(x))^n \frac{\alpha\bar{G}(x)^{\alpha-1} d\bar{G}(x)}{\ell(1/\bar{G}(x))F(b)} \\ &= \left(1 - \bar{G}(x_0)\right)^n + (1+\epsilon)^{1/2} \int_{\bar{G}(b)}^{\bar{G}(x_0)} (1-z)^n \frac{\alpha z^{\alpha-1} dz}{\ell(1/z)F(b)}, \end{aligned} \tag{2.25}$$

which follows from the absolute continuity of $G(x)$, i.e., $\bar{G}(A)$ is uniformly distributed in [0,1].

Now, for $\alpha \geq 1$, we consider two different cases: (a) $n\bar{G}(b) \geq \log n$ and (b) $C < n\bar{G}(b) < \log n$.

Case (a): $n\bar{G}(b) \geq \log n$. Observe that, for any fixed $H > \alpha + 6$, we can make $H\bar{G}(b)$ small enough by picking b_0 large. Now, by continuity of $G(x)$, there exists x_0 such that

$\bar{G}(x_0) = H\bar{G}(b)$; we can choose x_0 larger than in (2.25) by picking b_0 large enough. Next, using the elementary inequality $1-x \leq e^{-x}, x \geq 0$, we upper bound the preceding expression by

$$\begin{aligned}
\mathbb{P}[N_b > n] &\leq e^{-n\bar{G}(x_0)} + \frac{\alpha(1+\epsilon)^{1/2}}{F(b)} \int_{\bar{G}(b)}^{H\bar{G}(b)} (1-z)^n \frac{z^{\alpha-1}dz}{\ell(1/z)} \\
&\leq e^{-nH\bar{G}(b)} + \frac{\alpha(1+\epsilon)^{1/2}}{\ell(1/\bar{G}(b))F(b)} \sup_{\bar{G}(b)\leq z\leq H\bar{G}(b)} \frac{\ell(1/\bar{G}(b))}{\ell(1/z)} \int_{\bar{G}(b)}^{H\bar{G}(b)} (1-z)^n z^{\alpha-1} dz \\
&\leq e^{-nH\bar{G}(b)} + \frac{\alpha(1+\epsilon)}{\ell(1/\bar{G}(b))F(b)} \int_{\bar{G}(b)}^{H\bar{G}(b)} (1-z)^n z^{\alpha-1} dz \\
&\triangleq I_0 + I_1, \qquad (2.26)
\end{aligned}$$

where, for the third inequality, by the uniform convergence theorem (see [20]) of $\ell(x)$, $\bar{G}(b)^{-1}$ can be chosen large enough such that $\sup_{(H\bar{G}(b))^{-1}\leq y\leq \bar{G}(b)^{-1}} \ell(\bar{G}(b)^{-1})/\ell(y) \leq (1+\epsilon)^{1/2}$.

Now, we derive a lower bound for I_1 in (2.26). Using the monotonicity of $z^{\alpha-1}, \alpha \geq 1$ and since $F(b) \leq 1$, we obtain

$$\begin{aligned}
I_1 &\geq \frac{1}{\ell(1/\bar{G}(b))} \int_{\bar{G}(b)}^{H\bar{G}(b)} (1-z)^n z^{\alpha-1} dz \\
&\geq \frac{1}{\bar{G}(b)^{-\epsilon}} \bar{G}(b)^{\alpha-1} \int_{\bar{G}(b)}^{H\bar{G}(b)} (1-z)^n dz \\
&= \frac{1}{n+1} \bar{G}(b)^{\alpha-1+\epsilon} (1-\bar{G}(b))^{n+1} \left(1 - \left(\frac{1-H\bar{G}(b))}{1-\bar{G}(b)}\right)^{n+1}\right),
\end{aligned}$$

where in the second inequality, we use the property of slowly varying functions $\ell(x) \leq x^\epsilon$ for x large enough. Now, observe that for all $x \geq 0$ small enough, $1 - x \geq e^{-2x}$, yielding

$$I_1 \geq \frac{1}{4n} \bar{G}(b)^\alpha e^{-4n\bar{G}(b)},$$

where the last inequality follows from the fact that $n/(n+1) \geq 1/2$ for $n \geq 1$ and

$\bar{G}(b)^{\alpha-1+\epsilon} \geq \bar{G}(b)^{\alpha}$ since $\epsilon < 1$. We also note that

$$\left(\frac{1-H\bar{G(b)}}{1-\bar{G}(b)}\right)^{n+1} \leq \left(e^{-H\bar{G}(b)}/e^{-2\bar{G}(b)}\right)^{n} = e^{-(H-2)n\bar{G}(b)} \leq e^{-(H-2)C} \leq 1/2,$$

where we use our assumption $n\bar{G}(b) > C$ and choose H large enough. Finally, we obtain

$$I_1 \geq \frac{1}{4n}\bar{G}(b)^{\alpha}e^{-4n\bar{G}(b)}. \tag{2.27}$$

Now, we proceed with proving that I_0/I_1 in (2.26) is negligible as $n \to \infty$. To this end, observe that

$$\begin{aligned}\frac{I_0}{I_1} &\leq 4\frac{e^{-Hn\bar{G}(b)}n}{\bar{G}(b)^{\alpha}e^{-4n\bar{G}(b)}} \leq 4\frac{e^{-(H-4)n\bar{G}(b)}n^{\alpha+1}}{(n\bar{G}(b))^{\alpha}} \\ &\leq 4e^{-(H-4)n\bar{G}(b)}n^{\alpha+1} \leq 4e^{-(\alpha+2)\log n}n^{\alpha+1},\end{aligned}$$

where we use our assumption that $n\bar{G}(b) \geq \log n > 1$ for $n > 2$, whereas for the last inequality, we also use the fact that $H > \alpha + 6$. Thus, the preceding expression is upper bounded by

$$\frac{I_0}{I_1} \leq \frac{4}{n} \leq \epsilon, \tag{2.28}$$

for all $n \geq 4/\epsilon$.

Now, we upper bound I_1 in (2.26) by changing the variables $z = 1 - e^{-u/n}$,

$$\begin{aligned}I_1 &= \frac{\alpha(1+\epsilon)}{F(b)\ell(1/\bar{G}(b))}\int_{-n\log\left(1-\bar{G}(b)\right)}^{-n\log(1-H\bar{G}(b))}\frac{e^{-u(n+1)/n}(1-e^{-u/n})^{\alpha-1}}{n}du \\ &\leq \frac{\alpha(1+\epsilon)}{F(b)\ell(1/\bar{G}(b))}\int_{-n\log\left(1-\bar{G}(b)\right)}^{\infty}\frac{e^{-u}(1-e^{-u/n})^{\alpha-1}}{n}du,\end{aligned}$$

where for the inequality we use $e^{-u/n} \leq 1$ and extend the integral to infinity. Thus, for $\alpha \geq 1$, from the preceding expression using the inequality $1 - e^{-x} \leq x$, for $x \geq 0$, we obtain

$$\begin{aligned} I_1 &\leq \frac{\alpha(1+\epsilon)}{F(b)n\ell(1/\bar{G}(b))} \int_{-n\log\left(1-\bar{G}(b)\right)}^{\infty} e^{-u} \left(\frac{u}{n}\right)^{\alpha-1} du \\ &\leq \frac{\alpha(1+\epsilon)}{F(b)n^{\alpha}\ell(1/\bar{G}(b))} \int_{-n\log\left(1-\bar{G}(b)\right)}^{\infty} e^{-u} u^{\alpha-1} du \\ &= \frac{\alpha(1+\epsilon)}{F(b)n^{\alpha}\ell(1/\bar{G}(b))} \Gamma(-n\log\left(1-\bar{G}(b)\right), \alpha). \end{aligned} \tag{2.29}$$

Combining (2.28) and (2.29), we obtain for (2.26) that for all $n \geq n_0, b \geq b_0$,

$$\mathbb{P}[N_b > n] \leq \frac{\alpha(1+\epsilon)^2}{F(b)n^{\alpha}\ell(1/\bar{G}(b))} \Gamma(-n\log\left(1-\bar{G}(b)\right), \alpha),$$

which completes the proof after replacing $(1+\epsilon)$ with $(1+\epsilon)^{1/2}$.

Case (b): $C < n\bar{G}(b) < \log n$. In this region, for any fixed $H > 5$, we choose the smallest $m \geq 1$ such that $He^m - 4 \geq (\alpha+2)\log n/C$, i.e., $m = \lceil \log\left((\alpha+2)\log n/C + 4\right) - \log H \rceil$. Furthermore, it is important to note that this choice of m allows for $He^m\bar{G}(b)$ to be small enough, since $He^m\bar{G}(b) \leq He^m \log n/n = O(\log^2 n/n) \to 0$, as $n \to \infty$, by the assumption that $n\bar{G}(b) < \log n$. Then, by continuity of $G(x)$, there exists x_0 such that $\bar{G}(x_0) = He^m\bar{G}(b)$ (larger than x_0 in (2.25) for b_0 large enough) and using the elementary inequality $1 - x \leq e^{-x}, x \geq 0$, we upper bound the expression in (2.25) by

$$\begin{aligned} &\mathbb{P}[N_b > n] \\ &\leq e^{-n\bar{G}(x_0)} + \frac{\alpha(1+\epsilon)^{1/2}}{F(b)} \left[\int_{\bar{G}(b)}^{H\bar{G}(b)} (1-z)^n \frac{z^{\alpha-1}dz}{\ell(1/z)} + \sum_{k=0}^{m-1} \int_{He^k\bar{G}(b)}^{He^{k+1}\bar{G}(b)} e^{-nz} \frac{z^{\alpha-1}dz}{\ell(1/z)} \right] \\ &\leq e^{-n\bar{G}(x_0)} + \frac{\alpha(1+\epsilon)^{1/2}}{F(b)} \left[\frac{(1+\epsilon)^{1/2}}{\ell(1/\bar{G}(b))} \int_{\bar{G}(b)}^{H\bar{G}(b)} (1-z)^n z^{\alpha-1} dz \right. \\ &\qquad \left. + \sum_{k=0}^{m-1} \frac{(1+\epsilon)^{1/2}}{\ell(1/(e^k\bar{G}(b)))} (He^{k+1}\bar{G}(b))^{\alpha-1} \int_{He^k\bar{G}(b)}^{He^{k+1}\bar{G}(b)} e^{-nz} \right], \end{aligned}$$

where the last inequality follows from the monotonicity of $z^{\alpha-1}$ for $\alpha \geq 1$ and the uniform

convergence theorem of $\ell(x)$, $\sup_{(H\bar{G}(b))^{-1}\leq z\leq 1/\bar{G}(b))}\ell(1/\bar{G}(b))/\ell(z)\leq(1+\epsilon)^{1/2}$, while for the second term, note that $\sup_{(He^{k+1}\bar{G}(b))^{-1}\leq z\leq(He^k\bar{G}(b))^{-1}}\ell(1/(e^k\bar{G}(b))/\ell(z)\leq(1+\epsilon)^{1/2}$, $k=0\ldots(m-1)$, since $He^m\bar{G}(b)$ is small enough. Now, since $\ell(x)/\ell(x/e)\leq e$ for x large, it follows that

$$\begin{aligned}\mathbb{P}[N_b>n]&\leq e^{-nHe^m\bar{G}(b)}+\frac{\alpha(1+\epsilon)}{F(b)\ell(1/\bar{G}(b))}\int_{\bar{G}(b)}^{H\bar{G}(b)}(1-z)^n z^{\alpha-1}dz\\&\quad+\frac{\alpha(1+\epsilon)}{F(b)n\ell(1/\bar{G}(b))}\sum_{k=0}^{m-1}e^k e^{-nHe^k\bar{G}(b)}(He^{k+1}\bar{G}(b))^{\alpha-1}\\&\triangleq I_0+I_1+I_2.\end{aligned}\tag{2.30}$$

Now, we derive a lower bound for I_1 following similar arguments as in (2.27). Note that, for $x\geq 0$ small enough, $1-x\geq e^{-2x}$, and thus, for H large enough, we have

$$\begin{aligned}\frac{\ell(1/\bar{G}(b))F(b)}{\alpha(1+\epsilon)}I_1&\geq\bar{G}(b)^{\alpha-1}\int_{\bar{G}(b)}^{H\bar{G}(b)}(1-z)^n dz\\&\geq\bar{G}(b)^{\alpha-1}\frac{(1-\bar{G}(b))^{n+1}-(1-H\bar{G}(b))^{n+1}}{n+1}\\&=\bar{G}(b)^{\alpha-1}\frac{(1-\bar{G}(b))^{n+1}}{n+1}\left(1-\left(\frac{1-H\bar{G}(b)}{1-\bar{G}(b)}\right)^{n+1}\right)\\&\geq\frac{\bar{G}(b)^{\alpha-1}}{4n}e^{-4n\bar{G}(b)},\end{aligned}\tag{2.31}$$

where the expression in brackets is bounded from below by $1/2$ as in (2.27).

Now, we prove that I_0/I_1 in (2.30) is negligible as $n\to\infty$. To this end, observe that

$$\frac{I_0}{I_1}\leq\frac{4F(b)}{\alpha(1+\epsilon)}\frac{e^{-He^m n\bar{G}(b)}\ell(1/\bar{G}(b))n}{\bar{G}(b)^{\alpha-1}e^{-4n\bar{G}(b)}},$$

where we use (2.31). Next, since $\alpha\geq 1, F(b)\leq 1$, and using the standard property of slowly varying functions that $\ell(x)\leq x$ for large x (see Theorem 1.5.6 on page 25 of [20]), we obtain

$$\frac{I_0}{I_1}\leq 4\frac{e^{-(He^m-4)n\bar{G}(b)}n}{\bar{G}(b)^\alpha},$$

and since $n\bar{G}(b) > C$, we have

$$\begin{aligned}\frac{I_0}{I_1} &\leq 4\frac{e^{-(He^m-4)n\bar{G}(b)}n^{\alpha+1}}{(n\bar{G}(b))^{\alpha}} \\ &\leq 4C^{-\alpha}e^{-(He^m-4)C}n^{\alpha+1} \\ &\leq 4C^{-\alpha}e^{-(\alpha+2)\log n}n^{\alpha+1},\end{aligned}$$

where the last inequality follows from the fact that m was chosen so that $(He^m - 4) \geq (\alpha + 2)\log n/C$. Thus, the preceding expression can be rewritten as

$$\frac{I_0}{I_1} \leq \frac{4}{C^{\alpha}n} \leq \epsilon/2, \tag{2.32}$$

for all $n \geq 8C^{-\alpha}/\epsilon$.

Next, for the ratio I_2/I_1 we proceed similarly as before

$$\begin{aligned}\frac{I_2}{I_1} &= \frac{4\sum_{k=0}^{m-1} e^k e^{-nHe^k\bar{G}(b)}(He^{k+1}\bar{G}(b))^{\alpha-1}}{\bar{G}(b)^{\alpha-1}e^{-4n\bar{G}(b)}} \\ &\leq 4\sum_{k=0}^{m-1} e^k e^{-(He^k-4)n\bar{G}(b)}(He^{k+1})^{\alpha-1}, \\ &\leq 4H^{\alpha-1}\sum_{k=0}^{m-1} e^k e^{-(He^k-4)C}e^{\alpha(k+1)-k-1} \\ &\leq 4H^{\alpha-1}e^{-HC}\sum_{k=0}^{\infty} e^{-5(e^k-1)C+\alpha(k+1)-1} \leq \epsilon/2,\end{aligned} \tag{2.33}$$

where for the last inequality we use $H > 5$. Now, we further observe that the preceding sum is finite and thus, letting $H \to \infty$, the above ratio converges to 0, i.e., $I_2 \leq \epsilon I_1/2$ for large H.

Hence, since the upper bound for I_1 from (2.29) holds in this case as well, by putting (2.32) and (2.33) together, we obtain for (2.30) that for all $n \geq n_0, b \geq b_0$,

$$\mathbb{P}[N_b > n] \leq \frac{\alpha(1+\epsilon)^2}{F(b)n^{\alpha}\ell(1/\bar{G}(b))}\Gamma(-n\log\left(1-\bar{G}(b)\right), \alpha),$$

which completes the proof after replacing $(1+\epsilon)$ with $(1+\epsilon)^{1/2}$.

Last, we prove the lower bound for $n\bar{G}(b) > C$; here, we do not need to distinguish two cases. Thus, starting from (2.24) and proceeding with similar arguments as in the proof for the upper bound, we obtain

$$\begin{aligned}
\mathbb{P}[N_b > n] &\geq -(1-\epsilon)^{1/2} \int_{x_0}^{b} (1-\bar{G}(x))^n \frac{\alpha \bar{G}(x)^{\alpha-1} d\bar{G}(x)}{\ell(1/\bar{G}(x))F(b)} \\
&= (1-\epsilon)^{1/2} \int_{\bar{G}(b)}^{H\bar{G}(b)} (1-z)^n \frac{\alpha z^{\alpha-1} dz}{\ell(1/z)F(b)} \\
&\geq \frac{\alpha(1-\epsilon)}{F(b)\ell(1/\bar{G}(b))} \int_{-(n+1)\log\left(1-\bar{G}(b)\right)}^{-(n+1)\log(1-H\bar{G}(b))} \frac{e^{-u}(1-e^{-\frac{u}{n+1}})^{\alpha-1}}{n+1} du,
\end{aligned}$$

where we use the uniform convergence theorem of slowly varying functions (Theorem 1.2.1 on page 6 of [20]) and pick $x_0 < b$ such that $\bar{G}(x_0) = H\bar{G}(b)$. Next, using the inequality $1 - e^{-x} \geq (1-\delta)x$, for some $\delta > 0$ and all $x \geq 0$ small enough, we have

$$\begin{aligned}
\mathbb{P}[N_b > n-1] &\geq \frac{\alpha(1-\epsilon)(1-\delta)^{\alpha-1}}{F(b)\ell(1/\bar{G}(b))n} \int_{-n\log(1-\bar{G}(b))}^{-n\log(1-H\bar{G}(b))} e^{-u} \left(\frac{u}{n}\right)^{\alpha-1} du \\
&\geq \frac{\alpha(1-\epsilon)^2}{F(b)n^\alpha \ell(1/\bar{G}(b))} \int_{-n\log\left(1-\bar{G}(b)\right)}^{Hn\bar{G}(b)} e^{-u} u^{\alpha-1} du \\
&= \frac{\alpha(1-\epsilon)^2}{F(b)n^\alpha \ell(1/\bar{G}(b))} \left[\int_{-n\log\left(1-\bar{G}(b)\right)}^{\infty} e^{-u} u^{\alpha-1} du - \int_{Hn\bar{G}(b)}^{\infty} e^{-u} u^{\alpha-1} du \right] \\
&\triangleq I_1 - I_2,
\end{aligned}$$

where, in the second inequality, we choose $\delta > 0$ small enough such that $(1-\delta)^{\alpha-1} \geq (1-\epsilon)$ and note that $-n\log(1 - H\bar{G}(b)) \geq Hn\bar{G}(b)$. Next, we proceed with showing that I_2/I_1 is negligible for large n. Note that $-n\log(1-\bar{G}(b)) \leq 2n\bar{G}(b)$, which follows from the elementary inequality $e^{-2x} \leq 1 - x$ for all $x \geq 0$ small enough. Thus,

$$\begin{aligned}
\frac{I_2}{I_1} &\leq \frac{\int_{Hn\bar{G}(b)}^{\infty} e^{-u} u^{\alpha-1} du}{\int_{2n\bar{G}(b)}^{\infty} e^{-u} u^{\alpha-1} du} \\
&\leq \frac{2(Hn\bar{G}(b))^{\alpha-1} e^{-Hn\bar{G}(b)}}{(2n\bar{G}(b))^{\alpha-1} e^{-2n\bar{G}(b)}},
\end{aligned}$$

where we use the approximation for the incomplete gamma function for large H [see Remark 5 of Theorem 2.3]. Now, using the main assumption $n\bar{G}(b) > C$, we obtain

$$\frac{I_2}{I_1} \leq 2H^{\alpha-1}e^{-(H-2)C} \leq \epsilon,$$

for H large enough. Then, using the preceding observation, we obtain

$$\mathbb{P}[N_b > n] \geq \frac{\alpha(1-3\epsilon)}{n^\alpha \ell(1/\bar{G}(b))}\Gamma(-n\log\left(1-\bar{G}(b)\right), \alpha),$$

which completes the proof after replacing ϵ with $\epsilon/3$.

Now, if $\alpha < 1$, the proof uses almost identical arguments coupled with the fact that $u^{\alpha-1}$ is a decreasing function. We omit the details to avoid unnecessary repetitions. □

Remark 8. *From the preceding two theorems we observe that $\mathbb{P}[N_b > n]$ behaves as a true power law of index α when $-n\log\mathbb{P}[A \leq b] \to c$, $0 \leq c < \infty$, and has an exponential tail (geometric) when $n\mathbb{P}[A > b] \to \infty$ ($n >> \mathbb{P}[A > b]^{-1}$). More specifically:*
(i) If $-n\log\mathbb{P}[A \leq b] \to c$, then by Theorem 2.3, as $n \to \infty, n\mathbb{P}[A > b] \to c$,

$$\mathbb{P}[N_b > n] \sim \frac{\alpha}{\ell(n)n^\alpha}\Gamma(c, \alpha).$$

(ii) If $n\mathbb{P}[A > b] \to \infty$, then $-n\log\mathbb{P}[A \leq b] \to \infty$ and thus, as $n \to \infty, b \to \infty, n\mathbb{P}[A > b] \to \infty$,

$$\mathbb{P}[N_b > n] \sim \frac{\alpha}{\ell(1/\bar{G}(b))n}\bar{G}(b)^{\alpha-1}(1-\bar{G}(b))^n,$$

which follows from Theorem 2.4 and the asymptotic expansion of the Gamma function (see Remark 5 of Theorem 2.3).

Interestingly, one can compute the distribution of $\mathbb{P}[N_b > n]$ exactly for the special case when the parameter α takes integer values and $\ell(x) \equiv 1$.

Proposition 2.2. *If* $\mathbb{P}[L > x] = \mathbb{P}[A > x]^\alpha$, *for all* $x \geq 0$ *and* α *is a positive integer, then*

$$\mathbb{P}[N_b > n] = \frac{1}{\mathbb{P}[L \leq b]} \sum_{i=1}^{\alpha} \frac{\alpha!\, n!\, \mathbb{P}[A > b]^{\alpha-i}}{(\alpha - i)!(n+i)!} \mathbb{P}[A \leq b]^{n+i}.$$

Proof. It follows directly from (2.24) using integration by parts. □

Finally, in the following proposition, we describe the tail of $\mathbb{P}[N_b > n]$ for fixed and possibly small b. This complements the conclusion of Remark 8(ii), however, we need $\ell(x) \equiv 1$.

Proposition 2.3. *Let* b *be fixed. If* $\mathbb{P}[L > x] = \mathbb{P}[A > x]^\alpha$, $\alpha > 0$, $x \geq 0$, *then*

$$\mathbb{P}[N_b > n] \sim \frac{\alpha \mathbb{P}[A > b]^{\alpha-1}}{\mathbb{P}[L \leq b]} \frac{\mathbb{P}[A \leq b]^{n+1}}{n+1} \quad \text{as } n \to \infty.$$

Proof. See Section 4 in [1]. □

2.4 Simulation Experiments

In this section, we illustrate the validity of our theoretical results with simulation experiments. In all of the experiments, we observed that our uniform exact asymptotics is literally indistinguishable from the simulation. In the following examples, we present the simulation experiments resulting from 10^8 (or more) independent samples of $N_{b,i}, 1 \leq i \leq 10^8$. This number of samples was needed to ensure at least 100 independent occurrences in the lightest end of the tail that is presented in the figures ($N_{b,i} \geq n_{\max}$), thus providing a good confidence interval.

Example 1. This example illustrates the uniform exact asymptotics presented in Theorems 2.3 and 2.4, i.e., approximation (2.7), which combines the results from both theorems. We assume that L and A follow exponential distributions with parameters $\lambda = 2$ and $\mu = 1$, respectively. It is thus clear that $\bar{F}(x) = e^{-2x} = \bar{G}(x)^\alpha$, where $\alpha = 2$ and $\ell(x) \equiv 1$. Now, approximation (2.7) states that $\mathbb{P}[N_b > n]$ is given by $(1 - e^{-2b})^{-1} 2n^{-2} \Gamma(ne^{-b}, 2)$.

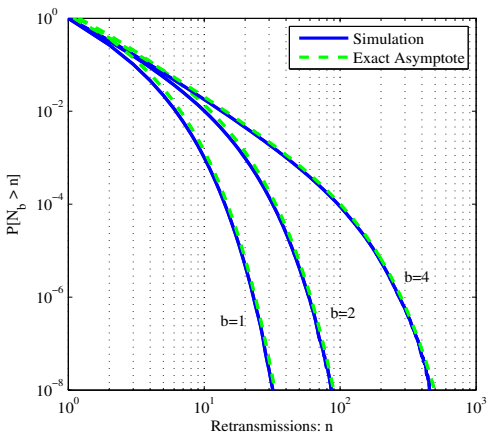


Figure 2.2: *Example 1(a). Exact asymptotics for* $\alpha > 1$.

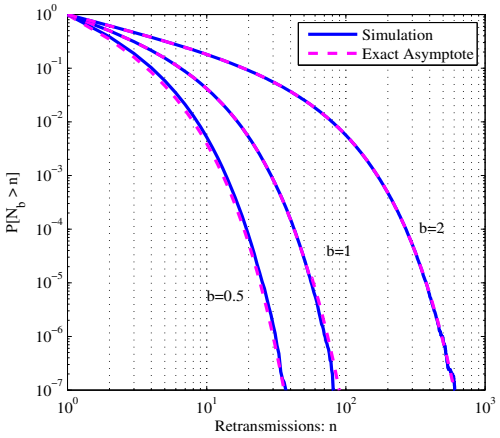


Figure 2.3: *Example 1(b). Exact asymptotics for* $\alpha < 1$.

Note that we added a factor $\mathbb{P}[L \leq b]^{-1} = (1 - e^{-2b})^{-1}$, as in Propositions 2.2 and 2.3, for increased precision when b is small; we add such a factor to approximation (2.7) in other examples as well. We simulate different scenarios when the data sizes L_b are upper bounded by b equal to 1, 2 and 4. The simulation results are plotted on log-log scale in Fig. 2.2.

From Fig. 2.2, we observe that the numerical asymptote approximates the simulation exactly for all different scenarios, even for very small values of n (large probabilities). We further validate our approximation by considering scenarios where L, A are exponentially distributed but $\alpha < 1$; in fact, this case tends to induce longer delays due to larger average data size compared to the channel availability periods. In this case, we obtain $\alpha = 0.5$ by assuming $\lambda = 1$ and $\mu = 2$. Again, the simulation results and the asymptotic formulas are basically indistinguishable for all n, as illustrated in Fig. 2.3.

For both cases, we deduce that for b small the power law asymptotics covers a smaller region of the distribution of N_b and, as n increases, the exponential tail becomes more evident and eventually dominates. As b becomes large - recall that $b \to \infty$ corresponds to the untruncated case where the power law phenomenon arises - the exponential tail becomes less distinguishable.

Example 2. This example demonstrates the exact asymptotics for the exponential tail

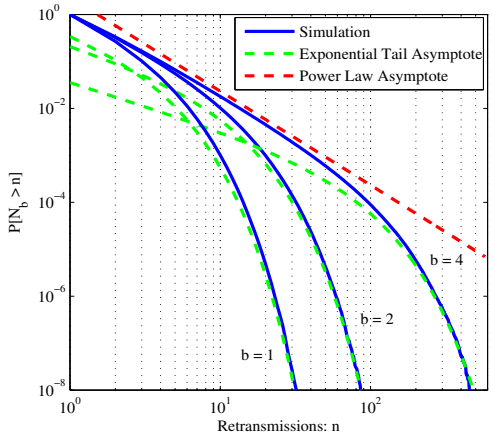


Figure 2.4: *Example 2. Power law vs. exponential tail asymptotics.*

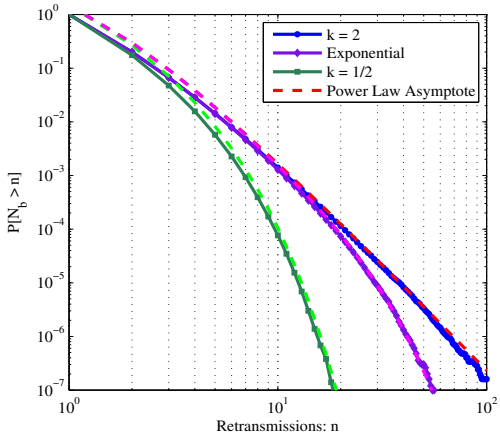


Figure 2.5: *Example 3. Power law region increases for lighter tails of L, A.*

as $n \to \infty$ and b is fixed, as in Proposition 2.3. Note that this proposition gives the exact asymptotic formula for the region $n \gg 1/\bar{G}(b)$ and lends merit to our Theorems 2.2 and 2.4. Informally, we could say that a point n_b such that $-n_b \log(1 - \bar{G}(b)) \approx n_b \bar{G}(b) = \alpha \log n_b$ represents the transition from power law to the exponential tail. We assume that L, A are exponentially distributed with $\lambda = 2$ and $\mu = 1$ (as in the first case of Example 1). Roughly speaking, we can see from Fig. 2.4 that the exponential asymptote appears to fit well starting from $n_b \approx \alpha e^b$, i.e., $n_b \approx 6, 15, 100$ for $b = 1, 2, 4$, respectively.

Example 3. This example highlights the impact of the distribution type of channel availability periods $\bar{G}(x) = \mathbb{P}[A > x]$. We consider some fixed b, namely $b = 8$, and assume that the matching between data sizes and channel availability, as defined in Theorems 2.3 and 2.4, is determined by the parameter $\alpha = 4$. We assume Weibull[1] distributions for L, A with the same index k and μ_L, μ_A respectively, such that $\alpha = (\mu_A/\mu_L)^k$. The simulations include three different cases for the aforementioned distributions: Weibull with index $k = 1$ (exponential) where $\mu_L = 1$ and $\mu_A = 4$, Weibull (normal-like) with index $k = 2$ ($\mu_L = 1, \mu_A = 2$) and Weibull with $k = 1/2$ ($\mu_L = 1, \mu_A = 16$). Fig. 2.5 illustrates the exact asymptotics from equation (2.7), shown with the lighter dashed lines; the main power law asymptote appears in the main body of all three distributions. We observe that heavier

distributions (Weibull with $k = 1/2$) correspond to smaller regions for the power law main body of the distribution $\mathbb{P}[N_b > n]$. On the other hand, the case with the lighter Gaussian like distributions for $k = 2$ follows almost entirely the power law asymptotics in the region presented in Fig. 2.5. This increase in the power law region can be inferred from our theorems, which show that the transition from the power law main body to the exponential tail occurs roughly at $n_b \approx \bar{G}(b)^{-1}$. Hence, the lighter the tail of the distribution of A, the larger the size of the power law region.

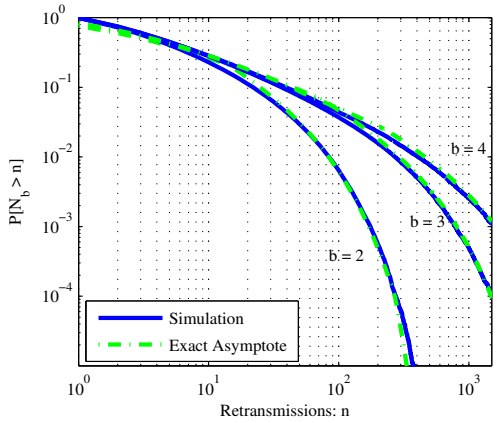


Figure 2.6: *Example 4(a). Exact asymptotics for the case where L follows the Gamma distribution.*

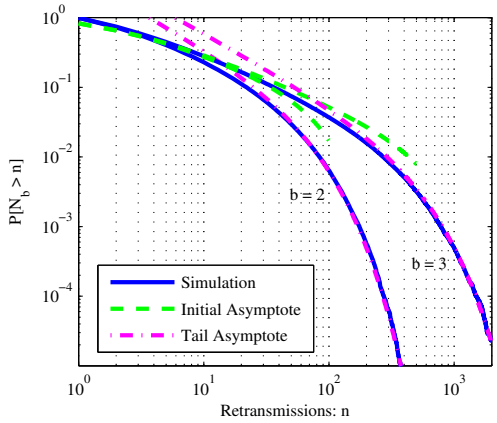


Figure 2.7: *Example 4(b). The asymptotes from Theorems 2.3 & 2.4 for Gamma distributed L.*

Example 4. In this last example, we study the case where there is a more general functional relationship between the distributions of availability periods A and data sizes L, as Theorems 2.3 and 2.4 assume. In particular, we consider the case $\bar{F}(x) = \bar{G}(x)^{\alpha}/\ell(\bar{G}(x)^{-1})$, where $\ell(x)$ is slowly varying. We validate the approximation (2.7) in this more general setting.

In particular, the availability periods A are exponentially distributed with parameter μ while the data sizes L follow the Gamma distribution with parameters (λ, k); the tail of the Gamma distribution function is defined as $\lambda^k \Gamma(k)^{-1} \int_x^{\infty} e^{-\lambda x} x^{k-1} dx = \Gamma(\lambda x, k)/\Gamma(k)$ and,

[1] In general, a Weibull distribution with index k has a complementary cumulative distribution function $\mathbb{P}[X > x] = e^{-(x/\mu)^k}$, where μ is the parameter that determines the mean.

therefore, the tail distribution of L can be approximated by $\bar{F}(x) \sim (\lambda^{k-1}/\Gamma(k))\ x^{k-1}e^{-\lambda x}$ for large x.

We can easily verify that $\bar{F}(x) = f(\mu^{-1}\log\bar{G}(x)^{-1})\bar{G}(x)^{\alpha}$, where $\alpha = \lambda/\mu$ and

$$f(x) = \lambda^{k-1}\Gamma(k)^{-1}\int_0^{\infty} e^{-z}(z/\lambda + x)^{k-1}dz.$$

Hence, the slowly varying function in Theorems 2.3 and 2.4 is $\ell(x) = 1/f(\mu^{-1}\log x)$. From the preceding integral representation for $f(x)$, it can be easily shown that $\ell(x) \approx \Gamma(k)\alpha^{1-k}\log^{1-k}x$, which is indeed slowly varying, and

$$\bar{F}(x) \approx (\alpha^{k-1}/\Gamma(k))\log(\bar{G}(x)^{-1})^{k-1}\bar{G}(x)^{\alpha}.$$

We take $\lambda = 2, k = 2$ and $\mu = 2$ and run simulations for $b = \{2, 3, 4\}$. In Fig. 2.6, we demonstrate the results using the approximation (2.7). Interestingly, our analytic approximation works nicely even for small values of n and b although the conditions in our theorems require n and b to be large.

In Fig. 2.7, we elaborate on the preceding example. To this end, we plot two asymptotes: (i) the 'Initial Asymptote' corresponding to the power law asymptote provided by Theorem 2.3 and (ii) the 'Tail Asymptote' from Theorem 2.4. Combining the two, we derive the approximation (2.7), as we have already shown in Fig. 2.6. Hereby, we see from Fig. 2.7 that both asymptotes are needed to approximate the entire distribution well, i.e., the 'Initial Asymptote' fits well the first part of the distribution, whereas the 'Tail Asymptote' is inaccurate in the beginning but works well for the tail. Recall that these two asymptotes differ only in the argument of the slowly varying function $\ell(\cdot)$, which is equal to n for the 'Initial Asymptote' and $\bar{G}(b)^{-1}$ for the tail.

2.5 Concluding Remarks

The uniform approximation presented in this chapter characterizes the entire body of the distribution $\mathbb{P}[N_b > n]$, and provides an explicit estimate of the region where the power law phenomenon arises. Therefore, it can be useful for diminishing the power law effects in order to achieve high throughput in modern engineering systems. Basically, a large power law region can lead to nearly zero throughput ($\alpha < 1$), implying that the system parameters should be carefully re-adjusted.

From an engineering perspective, our results further suggest that careful re-examination and possible redesign of retransmission based protocols in communication networks might be necessary. Specifically, current engineering trends towards infrastructure-less, error-prone wireless technology encourage the study of highly variable systems with frequent failures. In these types of systems, traditional approaches, e.g., blind data fragmentation, may be insufficient for achieving a good balance between throughput and resource utilization. Our analytical work could be applicable in network protocol design, e.g., data fragmentation techniques [13, 38, 37] and failure-recovery mechanisms. Overall, our generic model can be used towards improving the design of future complex and failure-prone systems in many different applications.

Chapter 3

Retransmissions over Correlated Channels

Frequent failures characterize many existing communication networks, e.g., wireless ad-hoc networks, where retransmission-based failure recovery represents a primary approach for successful data delivery. Recent work has shown that retransmissions can cause power law delays and instabilities even if all traffic and network characteristics are super-exponential. While the prior studies have considered an independent model, in this chapter, we extend the analysis to the practically important dependent case. We use modulated processes, e.g., Markov modulated, to capture the channel dependencies. We study the number of retransmissions and delays when the hazard functions of the distributions of data sizes and channel statistics are proportional, conditionally on the channel state. Our results show that the tails of the retransmission and delay distributions are determined by the state that generates the lightest asymptotics. Informally, the 'best case wins' and the system is insensitive to channel correlations. This insight is beneficial both for capacity planning and channel modeling since we do not need to account for the correlation details. However, this observation may be overly optimistic when the best state is infrequent, since the effects of 'bad' states may be prevalent for sufficiently long to downgrade the expected performance.

3.1 Introduction

Recovery mechanisms are employed in almost all engineering systems that are prone to failures. Restarting the interrupted jobs after a failure occurs is one of the most straightforward and widely used failure recovery mechanism. In modern communication networks, retransmission mechanisms are particularly important on all protocol layers to guarantee data delivery in the presence of channel failures. It was first recognized in [24, 25] that such mechanisms may result in power law delays even if the job sizes and failure rates are light tailed. In [9], it was shown that retransmission phenomena can lead to zero throughput and system instabilities, and therefore need to be carefully considered for the design of fault tolerant systems.

More specifically, it has been shown that power law delays arise in different layers of the networking architecture, where retransmission-based protocols are used, e.g., ALOHA type protocols in MAC layer [12, 14], end-to-end acknowledgements in transport layer [11, 9] as well as in other layers [9]. For other (non-retransmission) mechanisms that can give rise to heavy tails see [15] and the references therein. In particular, the proportional growth/multiplicative models can result in heavy tails [15, 16].

Previous studies consider an i.i.d model that was first introduced in [24] and further studied in [9, 10] to describe the channel dynamics. In practice, communication channels are highly correlated in the sense that they switch between states with different characteristics. We extend the previously studied model [9] to the dependent case where the availability periods depend on the channel state. In order to capture the channel dependencies, we introduce a modulating process. In general, the distributions of the channel availability periods depend on the current state of the channel.

The proposed model is as follows. Let $\{J_n\}_{n\geq 1}$ be a stationary and ergodic modulating process with finitely many states $\{1, 2, \ldots, K\}$. Now, let $\{A_n(k), n \geq 1, k = 1, \ldots, K\}$ be a family of independent random variables, independent of $\{J_n\}$, such that for fixed k, $\{A_n(k)\}_{n\geq 1}$ are identically distributed with $\bar{G}_k(x) = \mathbb{P}(A_1(k) > x)$. Then, we can construct

a modulating process A_n such that $A_n \triangleq A_n(J_n)$ and $\mathbb{P}(A_n > x|J_n = k) = \bar{G}_k(x)$. At each available period A_n, we transmit a generic data unit of size L; if $A_n > L$, the transmission is successful, else we wait until the next period A_{n+1} to retransmit. We study the asymptotic properties of the distribution of the number of retransmissions N when

$$\log \mathbb{P}(L > x) \approx \alpha_k \log \mathbb{P}(A_n > x|J_n = k), \tag{3.1}$$

$k = 1, \ldots, K$; see Section 3.1.1 for a more detailed description of the preceding model.

We show that when the channel is correlated, or less formally, when it alternates between different states, the tail asymptotics is determined by the properties of the 'best' channel state, e.g., the state that generates the lightest asymptotics in the corresponding independent channel model. Intuitively, as the channel switches between states, a large data unit is more likely to be transmitted when the channel is 'good'. Specifically, the 'best' availability periods correspond to the state with the largest α_k [as defined in (3.1)] among $1, \ldots, K$. Undoubtedly, this is an optimistic observation which further implies that instabilities can be eliminated as long as there exists at least one state with $\alpha > 1$.

From an engineering perspective, this optimistic best case scenario prediction and the apparent insensitivity to the structure of the channel correlations can be very promising in system analysis and design. The result implies that the initial i.i.d. model might be sufficient for modeling, and can also be extended to even more complex failure-prone networks. However, this is partially true as there are certain circumstances under which this claim underestimates the intricacies of the system.

Specifically, the light tail does not guarantee consistent good behavior for the entire body of the delay distribution. As discussed in [1, 2] in a different context, the delay distribution of bounded documents will always have a light exponential tail. However, the main body of the distribution can be a power law which will determine the performance in the relevant range of probabilities. Similarly here, the tail is determined by the lightest power law (largest value of α_k in (3.1)). However, when the corresponding channel state

is rare, the main body of the retransmission distribution can be dominated by the heavier power laws resulting from the 'bad' states. Hence, the system performance may be much worse than the predicted by the tail. Therefore, when the best case scenario is atypical, we need to pay closer attention to the channel correlations. We provide further discussions and some preliminary analysis of this situation in Section 3.3.

Our results can be useful from an engineering perspective, both for modeling and system design. The analysis can be extended to more complex networks or multi-channel systems that are characterized by frequent failures and correlated states. The results may be applied in designing new protocols, or developing new fragmentation schemes [13, 38, 37] specifically for correlated channels. A dynamic fragmentation technique is more likely to work better for a channel with high variability. In addition, the explicit approximation presented in Chapter 2 could be combined with the the analytic results of this chapter in order to accurately estimate the optimal sizes of the packet fragments.

The chapter is organized as follows. In the following Section 3.1.1, we formally describe the model and introduce the necessary definitions and notation, while in Section 3.2, we present our main theorems, on both the logarithmic and the exact scale. Next, Section 3.3 includes our simulation experiments that verify our analytic approximations. Last, in Section 3.4, we discuss the engineering implications of our results and provide some insight on the situation when the 'best case' scenario occurs rarely, while Section 3.5 contains some of the proofs.

3.1.1 Description of the Channel

In this section, we formally describe our model and provide necessary definitions and notation. Consider transmitting a generic data unit of random size L over a channel with failures. Without loss of generality, we assume that the channel is of unit capacity. The channel dynamics is modeled as follows. Let $J \triangleq \{J_n\}_{n\geq 1}$ be a stationary and ergodic modulating process with finitely many states $\{1, 2, \ldots, K\}$. Let $\pi_k = \mathbb{P}(J_n = k)$ denote the probability that the process is in state k, $k = 1 \ldots K$; see Figure 3.1 for an illustration of

the channel.

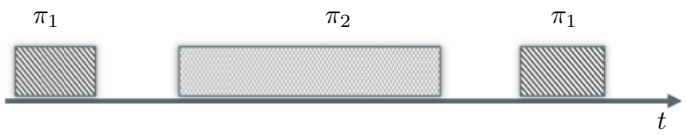


Figure 3.1: Correlated channel dynamics.

Now, define a family of independent random variables $\{A_n(k), n \geq 1, k = 1, \ldots, K\}$, independent of the modulating process $\{J\}$. In addition, for fixed k, $\{A_n(k)\}_{n\geq 1}$ are identically distributed with $G_k(x) = \mathbb{P}(A_1(k) \leq x)$ and $\bar{G}_k(x) = 1 - G_k(x)$. Then, we construct a modulating process A_n such that $A_n \triangleq A_n(J_n)$ and $\mathbb{P}(A_n > x | J_n = k) = \bar{G}_k(x)$.

At each period of time that the channel becomes available, say A_i, we attempt to transmit a generic data unit of size L. If $A_i > L$, we say that the transmission is successful; otherwise, we wait until the next period A_{i+1} when the channel is available and attempt to retransmit the data from the beginning. A sketch of the model depicting the system is drawn in Figure 3.2.

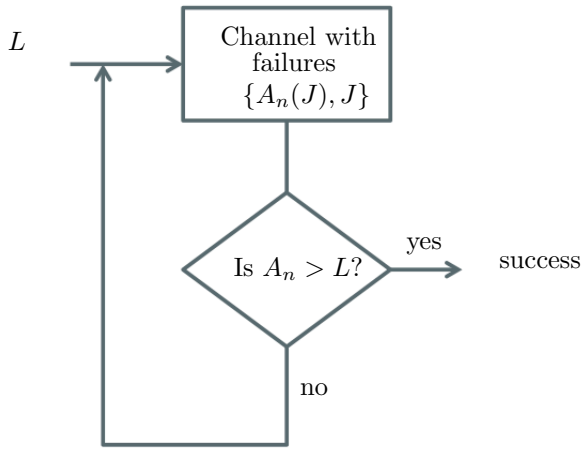


Figure 3.2: Packets sent over a channel with failures.

We are interested in computing the number of attempts N (retransmissions) that are required until L is successfully transmitted, which is formally defined as follows.

Definition 3.1.1. *The total number of retransmissions for a generic data unit of length L is defined as*

$$N \triangleq \inf\{n : A_n > L\}.$$

We denote the complementary cumulative distribution functions for $\{A(k)\}_{k=1\ldots K}$ and L, respectively, as

$$\bar{G}_k(x) := \mathbb{P}(A_n > x | J_n = k)$$

and

$$\bar{F}(x) := \mathbb{P}[L > x].$$

We assume that L and A are continuous (equivalently, $\bar{F}(x)$ and $\bar{G}_k(x)$ are absolutely continuous) and have infinite support, i.e., $\bar{G}_k(x) > 0$ and $\bar{F}(x) > 0$ for all $x \geq 0$. We use the following standard notations. For any two real functions $a(t)$ and $b(t)$ and fixed $t_0 \in \mathbb{R} \bigcup \{\infty\}$, we use $a(t) \sim b(t)$ as $t \to t_0$ to denote $\lim_{t\to t_0} a(t)/b(t) = 1$. Similarly, we say that $a(t) \gtrsim b(t)$ as $t \to t_0$ if $\underline{\lim}_{t\to t_0}\ [a(t)/b(t)] \geq 1$; $a(t) \lesssim b(t)$ has a complementary definition.

3.2 Main Results

In this section, we present our main analytic results. In Theorem 3.1, we characterize the tail distribution of the number of retransmissions on the logarithmic scale. In particular, under the assumption that the hazard functions of the data sizes and channel statistics are proportional, we show that the distribution exhibits a power law tail and the index is determined by the channel state with the longest availability periods, e.g., the maximum α_k as defined in (3.1); the asymptotics is the same as in the case of the uncorellated channel with $\bar{G}(x) = \bar{G}_m(x)$, see Theorem 2 in [9].

Next, in Theorem 3.2, under more restrictive assumptions, we prove the result on the exact scale. We derive the exact asymptotic tail of the number of retransmissions N, which also depends on the steady state probability π_m of the 'best' state m. The result shows

that the smallest the values of π_m, the longer it takes for the lightest tail to appear. In fact, from the simulation results of Section 3.3, it becomes evident that the distribution transits between power laws of different indices before the lighter one dominates. This is intuitive, since if the 'good' periods are not frequent ($\pi_m \to 0$), we would expect a lot of failures resulting from the other states with shorter availability periods.

Last, Theorem 3.3 presents the logarithmic asymptotics of the total delay distribution.

Theorem 3.1. *Let $\{J, J_n\}_{n\geq 1}$ be a stationary and ergodic process that takes values on the positive integers $k = 1 \dots K$. Assume that for the family of random variables $\{A, A_n(k)\}$,*

$$\log \mathbb{P}(L > x) \sim \alpha_k \log \mathbb{P}(A > x | J = k) \text{ as } x \to \infty,$$

and $\mathbb{P}(|\sum_{i=1}^{n} 1\{J_i = k\} - \pi_k n| \geq \epsilon n) = O(n^{-(\alpha_m+\epsilon)/K})$, for positive ϵ and $\alpha_m \triangleq \max_{k=1\dots K} \alpha_k > 0$, then

$$\lim_{n\to\infty} \frac{\log \mathbb{P}(N > n)}{\log n} = -\alpha_m.$$

Remark 9. *Throughout this chapter, we will use m to denote the index of the state with the largest α among all states $1, 2, \dots, K$. This corresponds to the state m which dominates the tail distribution of N and is responsible for the lighter asymptote for large n. Without loss of generality, we assume that there is a unique m that achieves the maximum $\alpha_m = \max_{k=1,\dots,K} \alpha_k$, i.e., $\alpha_m > \max_{k\neq m} \alpha_k$. Otherwise, if there are more than one indices that attain the maximum, we can merge the corresponding underlying states of the process $\{J_n\}_{n\geq 1}$ into a single one with the same property.*

Remark 10. *Note that the condition $\mathbb{P}(|\sum_{i=1}^{n} 1\{J_i = k\} - \pi_k n| \geq \epsilon n) = O(n^{-(\alpha_m+\epsilon)/K})$ is satisfied for a large class of modulating processes J_n, i.e., semi-Markov processes where the sojourn time distribution decays faster than $O(1/n^{(\alpha_m+\epsilon)/K+1})$.*

Proof. By assumption, there exists $0 < \epsilon < 1$ such that for all $x > x_0$,

$$\bar{F}(x)^{\frac{1}{\alpha_k(1-\epsilon)}} \leq \bar{G}_k(x) \leq \bar{F}(x)^{\frac{1}{\alpha_k(1+\epsilon)}}, k = 1 \dots K. \tag{3.2}$$

Recall that $\{A_n(k)\}_{n\geq 1}$ are conditionally independent given $\{J, J_n\}_{n\geq 1}$ and thus $\mathbb{P}(A_n(k) > x|J_n = k) = \mathbb{P}(A_1 > x|J_n = k) = \bar{G}_k(x)$. Note that $A_i(J_i)$ is independent of the past and future states of the modulating process $\{J_j\}_{j\neq i}$. Let $N_n^k := \sum_{i=1}^n 1\{J_i = k\}$ be the number of times that $\{J_i = k\}$ is in the interval $[1, n]$; thus, $\sum_{k=1}^K N_n^k = n$.

First, we establish the *lower bound*. It is easy to see that

$$\begin{aligned}
\mathbb{P}[N > n|L] &= \mathbb{P}[L > A_1, L > A_2 \ldots, L > A_n] \\
&= \mathbb{E}[\mathbb{P}(L > A_j, 1 \leq j \leq n|J_1, \ldots, J_n)] \\
&= \mathbb{E}\left[\prod_{j=1}^n \mathbb{P}(L > A_j|J_j)\right] \\
&= \mathbb{E}\left[\prod_{j=1}^n \prod_{k=1}^K \mathbb{P}(L > A_j|J_j = k)^{\mathbf{1}\{J_j=k\}}\right] \\
&= \mathbb{E}\left[\prod_{j=1}^n \prod_{k=1}^K \mathbb{P}(L > A|J = k)^{\mathbf{1}\{J_j=k\}}\right] \\
&= \mathbb{E}\left[\prod_{k=1}^K \mathbb{P}(L > A|J = k)^{N_n^k}\right]. \qquad (3.3)
\end{aligned}$$

For the ergodic and stationary process $\{J_n\}_{n\geq 1}$, by the strong law of large numbers, it follows that

$$\frac{N_n^k}{n} \to \pi_k \quad \text{as } n \to \infty,$$

for all $k = 1 \ldots K$. Thus, for any $\epsilon > 0$, we can choose n_0, such that $N_n^k \leq (1+\epsilon)\pi_k n$, for all $n \geq n_0$ and $k = 1 \ldots K$. Therefore,

$$\begin{aligned}
\mathbb{P}(N > n|L) &\geq \mathbb{E}\left[\prod_{k=1}^K \mathbb{P}(L > A|J = k)^{(1+\epsilon)\pi_k n} \mathbf{1}\left\{N_n^k \leq (1+\epsilon)\pi_k n\right\}\right] \\
&= \prod_{k=1}^K \mathbb{P}(N_n^k \leq (1+\epsilon)\pi_k n)\mathbb{P}(L > A|J = k)^{(1+\epsilon)\pi_k n}
\end{aligned}$$

$$\geq (1-\epsilon)^K \prod_{k=1}^{K} (1-\bar{G}_k(L))^{(1+\epsilon)\pi_k n},$$

where we note that $\mathbb{P}(N_n^k \leq (1+\epsilon)\pi_k n) \to 1$ as $n \to \infty$. Now, using our main assumption (3.2) and the elementary inequality $1 - x \geq e^{-(1+\epsilon)x}$ for small x, we obtain

$$\begin{aligned}
\mathbb{P}(N > n) &= \mathbb{E}[P(N > n|L)] \\
&\geq (1-\epsilon)^K \mathbb{E} \prod_{k=1}^{K} \left(1 - \bar{F}(L)^{\frac{1}{\alpha_k(1+\epsilon)}}\right)^{\pi_k n(1+\epsilon)} \mathbf{1}\{L \geq x_0\} \\
&\geq (1-\epsilon)^K \mathbb{E} \prod_{k=1}^{K} \exp\left(-\pi_k n(1+\epsilon)^2 \bar{F}(L)^{\frac{1}{\alpha_k(1+\epsilon)}}\right) \mathbf{1}\{L \geq x_0\},
\end{aligned}$$

for x_0 as in (3.2). Now, observe that the integral in the preceding expression is

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(-\sum_{k=1}^{K} \pi_k n(1+\epsilon)^2 \bar{F}(L)^{\frac{1}{\alpha_k(1+\epsilon)}}\right) \mathbf{1}\{L \geq x_0\}\right] \\
&\geq \mathbb{E}\left[\exp\left(-\sum_{k=1}^{K} \pi_k n(1+\epsilon)^2 U^{\frac{1}{\alpha_k(1+\epsilon)}}\right)\right] - \exp\left(-\sum_{k=1}^{K} \pi_k n(1+\epsilon)^2 \bar{F}(x_0)^{\frac{1}{\alpha_k(1+\epsilon)}}\right) \\
&\triangleq I_1 - I_0, \qquad (3.4)
\end{aligned}$$

which follows from $\bar{F}(L) = U$, with U being uniformly distributed in $(0, 1)$.

The first term in (3.4) is computed as

$$\begin{aligned}
I_1 &= \int_0^1 \exp\left(-\sum_{k=1}^{K} \pi_k n(1+\epsilon)^2 u^{\frac{1}{\alpha_k(1+\epsilon)}}\right) du \\
&\geq \int_0^{\epsilon} \exp\left(-n(1+\epsilon)^2 u^{\frac{1}{\alpha_m(1+\epsilon)}} \pi_m \left(1 + \sum_{k=1, k\neq m}^{K} \frac{\pi_k}{\pi_m} u^{\frac{1}{\alpha_k(1+\epsilon)} - \frac{1}{\alpha_m(1+\epsilon)}}\right)\right) du \\
&\geq \int_0^{\epsilon} \exp\left(-n(1+\epsilon)^2(1+\delta) u^{\frac{1}{\alpha_m(1+\epsilon)}} \pi_m\right) du,
\end{aligned}$$

after observing that for ϵ small enough, $\sum_{k=1, k\neq m}^{K} (\pi_k/\pi_m) u^{1/(\alpha_k(1+\epsilon)) - 1/(\alpha_m(1+\epsilon))} \leq \delta$.

Next, by change of variables $z = n\pi_m(1+\epsilon)^2(1+\delta)u^{1/(\alpha_m(1+\epsilon))}$, we have

$$\begin{aligned} I_1 &\geq \frac{(1-\delta_\epsilon)^{\alpha_m(1+\epsilon)}\alpha_m(1+\epsilon)}{\pi_m^{\alpha_m(1+\epsilon)}n^{\alpha_m(1+\epsilon)}}\int_0^{n\epsilon^{1/\alpha_m}(1-\delta_\epsilon)\pi_m} e^{-z}z^{\alpha_m(1+\epsilon)-1}dz \\ &\geq \frac{(1-\delta_\epsilon)^{\alpha_m(1+\epsilon)}}{\pi_m^{\alpha_m(1+\epsilon)}n^{\alpha_m(1+\epsilon)}}\Gamma(\alpha_m(1+\epsilon)+1), \end{aligned}$$

where we use the definition of the gamma function for large n and set $(1-\delta_\epsilon)^{-1} = (1+\epsilon)^2(1+\delta)$.

Now, for $h_{\epsilon,\delta} = (1-\epsilon)^K(1-\delta_\epsilon)^{\alpha_m(1+\epsilon)}\Gamma(\alpha_m(1+\epsilon)+1)/\pi_m^{\alpha_m(1+\epsilon)}$, and, since x_0 is fixed, the second term in (3.4) is negligible, i.e., $I_0 \to 0$ as $n \to \infty$. Taking the logarithm yields

$$\log \mathbb{P}(N > n) \geq \log h_{\epsilon,\delta} - \alpha_m(1+\epsilon)\log n,$$

and by picking n_0 such that $\log h_{\epsilon,\delta} \geq -\alpha_m\epsilon \log n$, we have

$$\log \mathbb{P}(N > n) \geq -\alpha_m(1+2\epsilon)\log n.$$

After replacing ϵ with $\epsilon/2$, we derive

$$\frac{\log \mathbb{P}(N > n)}{\log n} \geq -(1+\epsilon)\alpha_m. \tag{3.5}$$

The remainder of the proof for the upper bound is deferred to Section 3.5. □

Next, as briefly stated in the beginning of this section, we present our analytic approximation on the exact scale. In the following Theorem, we need more restrictive assumptions and, in particular, we assume that the matching between the distribution of the data sizes and the channel characteristics is described by a regularly varying function of index α. A regularly function is defined as a function of the form $\ell(x)x^{1/\alpha}$, where $\ell(\cdot)$ is slowly varying.

Definition 3.2.1. *A function $\ell(x)$ is slowly varying if $\ell(x)/\ell(\lambda x) \to 1$ as $x \to \infty$ for any fixed $\lambda > 0$.*

We assume that functions $\ell(x)$ are positive and bounded on finite intervals.

Theorem 3.2. *Let* $m := \arg\max_{k=1\ldots K} \alpha_k$. *If* $\bar{F}^{-1}(x) = \Phi_k(\bar{G}_k^{-1}(x))$, *where* $\Phi_k(x) = \ell(x)x^{1/\alpha_k}$, *for all* $x \geq 0, \alpha_k > 0, k = 1 \ldots K$, *and* $\mathbb{P}(|\sum_{i=1}^{n} 1\{J_i = k\} - \pi_k n| \geq \epsilon n) = O(n^{-(\alpha_m+\epsilon)/K})$, $\epsilon > 0$, *then as* $n \to \infty$,

$$\mathbb{P}[N > n] \sim \frac{\Gamma(\alpha_m + 1)}{\Phi_m(n)\pi_m^{\alpha_m}}.$$

Proof. The proof is deferred to Section 3.5. □

Last, we show that the distribution of the total transmission time also follows a power law of the same index α_m on the logarithmic scale.

Theorem 3.3. *Under the same conditions as in Theorem 3.1 and* $\mathbb{E}[A^{1+\theta}] < \infty$ *for some* $\theta > 0$, *then*

$$\lim_{t\to\infty} \frac{\log \mathbb{P}(T > t)}{\log t} = -\alpha_m.$$

Proof. See Section 3.5. □

3.3 Simulations

In this section, we present our simulation experiments. The results are derived from $N = 10^8$ independent samples of our simulated model.

Example 1. In this example, we simulate a channel with two states, 1 and 2. At each state, the availability periods are i.i.d. random variables exponentially distributed with $\mu_1 = 1/4$ and $\mu_2 = 1$. Also, the data unit sizes are continuous random variables, following the exponential distribution with mean 2 ($\lambda = 1/2$). Therefore, by definition, we have $\alpha_1 = 2$ and $\alpha_2 = 0.5$. The transition probability from state i to state j is defined as p_{ij} so that the steady state probabilities are given by $\pi_i = p_{ji}/(p_{ij} + p_{ji})$. In Fig. 2.2, we present the asymptotics of the number of retransmissions on the logarithmic scale for three values of steady state probabilities: $\pi_m = 0.1, \pi_m = 0.5$ and $\pi_m = 0.9$; recall that m is the index of the state with the larger α. We plot the exact asymptotics from Theorem 3.2, where we note

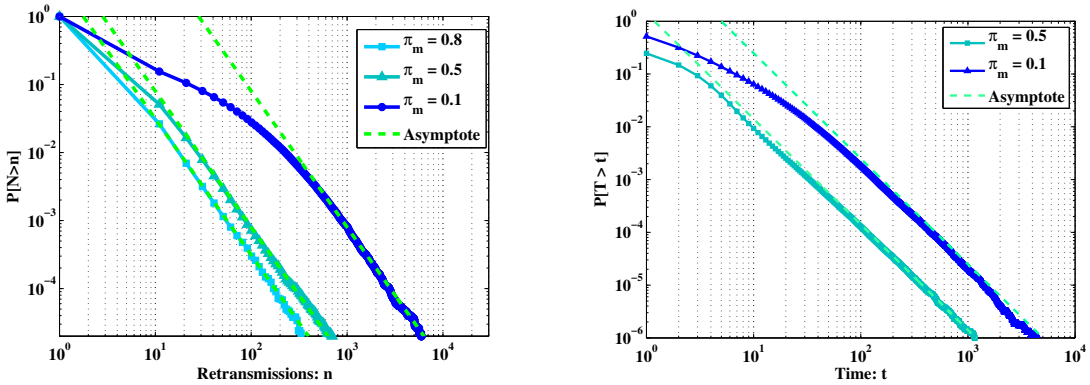


Figure 3.3: Example 1. Asymptotics of $\mathbb{P}[N > n]$ and transmission delay $\mathbb{P}[T > t]$ for a two-state channel.

that the constant term $\Gamma(\alpha_m+1)/\pi_m^{\alpha_m}$ increases the precision of our logarithmic asymptotics (Theorem 3.1). We observe that our simulation results are in excellent agreement with the theoretical asymptote.

Next, for the same channel and two values of π_m, namely 0.1, and 0.5, Fig. 3.3 demonstrates the logarithmic asymptotics for $\mathbb{P}[T > t]$ obtained from Theorem 3.3.

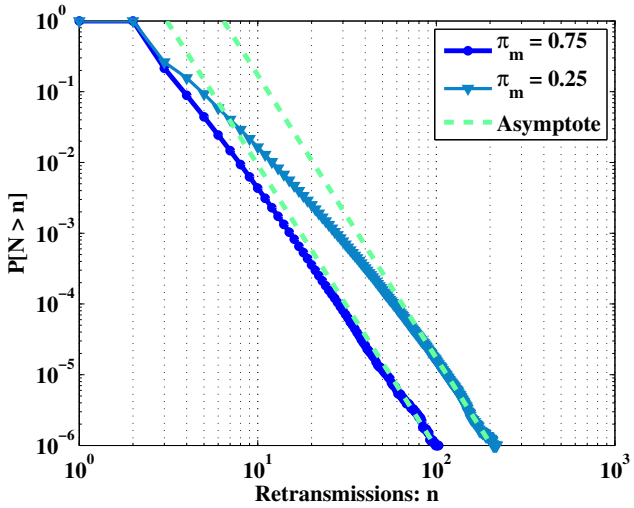


Figure 3.4: Example 2. Asymptotics of $\mathbb{P}[N > n]$ for a three-state channel.

Example 2. In this example, for completeness, we consider a three state channel, with transition probabilities such that $\pi_m = 0.25$ and $\pi_m = 0.75$. The availability periods at each

state are exponentially distributed with $\mu_1 = 2, \mu_2 = 1/2$ and $\mu_3 = 1/4$ and the data sizes are exponential with unit mean. From Fig. 3.4, we observe that the lightest asymptotics (power law with exponent $\alpha = 4$) dominates the tail of $\mathbb{P}[N > n]$. The power law tail appears earlier when $\pi_m = 0.75$ and is not affected by other states for large values of n.

Example 3. In our last example, we simulate a two-state channel where the packet sizes and the availability periods are normally distributed and the channel alternates between the two states with probability 1/2. Suppose that A and L take absolute values of zero mean normal random variables, with $\sigma_L = 5$ and $\sigma_{A_1} = 2, \sigma_{A_2} = 4, 6, 8$, for states 1 and 2 respectively. Therefore, the asymptotic assumption of Theorem 3.1 is satisfied. In Fig 3.5, we plot the logarithmic asymptotics for three different values of $\alpha = \sigma_A^2/\sigma_L^2$, as marked on the graph.

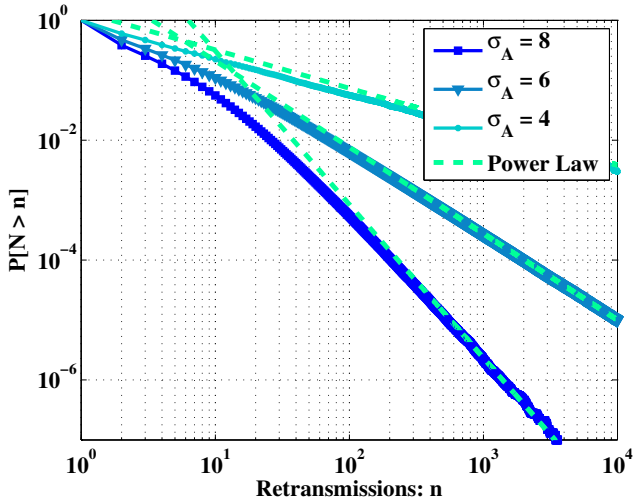


Figure 3.5: Example 3. Logarithmic asymptotics for a two-state channel where data sizes and channel statistics are normally distributed.

3.4 Concluding Remarks

In this section, we discuss the engineering implications of our results. Previously, we showed that when the channel is correlated, meaning that it switches between dependent states, the 'best case' scenario wins. This implies that the delay asymptotics and the stability condi-

tions are determined by the state that generates the lightest tail supposing that the channel is uncorrelated. This insensitivity to the detailed structure of the correlations as well as the optimistic best case predictions are beneficial both for modeling and dimensioning/capacity planning of such systems. In particular, our insights show that the independent channel model might be sufficient. Furthermore, the analysis of the independent model is more likely to be extended to more complex multi-channel and networking systems with failures.

3.4.1 A word of caution

However, in this subsection, we emphasize that a design relying on the best case scenario may be overly optimistic and even completely wrong if the best state of the channel is atypical, i.e., it occurs rarely. To illustrate this point, we study the following simplified model that demonstrates the impact of the correlated channel states on the distribution of N. In particular, we consider a channel with two states ($K = 2$), such that $\alpha_1 > \alpha_2$, and we assume that π_1 is very small ($\pi_1 \ll \pi_2, \pi_1 \approx 0$), i.e., state 1 is much less frequent than state 2. In this case, the tail of the distribution is still a power law with index $\alpha_m = \alpha_1$. However, there exists another power law asymptote that appears earlier and dominates the body of the distribution for smaller values of n.

We herein characterize these two asymptotes with two equivalent explicit formulas that approximate the retransmission distribution for large n (informal derivations are presented in the Appendix):

$$\mathbb{P}[N > n] \approx \frac{\alpha_2}{n^{\alpha_2}\pi_2^{\alpha_2}} \sum_{i=0}^{\infty} \frac{(-n^{1-\delta}P_2)^i \Gamma(\alpha_2 + i\delta)}{i!} \tag{3.6}$$

$$\mathbb{P}[N > n] \approx \frac{\alpha_1}{n^{\alpha_1}\pi_1^{\alpha_1}} \sum_{i=0}^{\infty} \frac{(-n^{1-1/\delta}P_1)^i \Gamma(\alpha_1 + i/\delta)}{i!}. \tag{3.7}$$

Note that the sum in (3.6) is absolutely convergent since $\Gamma(\alpha_2 + i\delta) \leq \lceil \alpha_2 + i\delta \rceil !$. As

we can easily infer from the first expression, when $n^{1-\delta}P_2 \ll 1$, the leading term dominates and the initial part of the distribution is determined by the heavier power law $O(n^{-\alpha_2})$ with exponent $\alpha_2 < \alpha_1$. This is indeed the asymptote that works well for small values of n, specifically when $n^{1-\delta} \ll 1/P_2$, as will be evident in the forthcoming example. Accordingly, the leading term of the second asymptote from (3.7) yields the correct tail asymptotics from Theorem 3.2, which holds as $n \to \infty$.

In order to illustrate the results, we plot the exact asymptotes from equations (3.6) and (3.7), in Fig. 3.6. In this example, we take $\alpha_1 = 2$ and $\alpha_2 = 1/2$ while the steady state probabilities are $\pi_1 = 0.01$ and $\pi_2 = 0.99$. Specifically, we use five error terms for both asymptotes. We observe that the precision of the first asymptote deteriorates after $n \approx 10^2$ unless we increase the number of terms in expression (3.6). The leading term is a power law of index 1/2, which leads to the heavier asymptote. On the other hand, the tail asymptote derived in this section, even with few terms, fits perfectly for large values of n, which lends credit to our main Theorem 3.2.

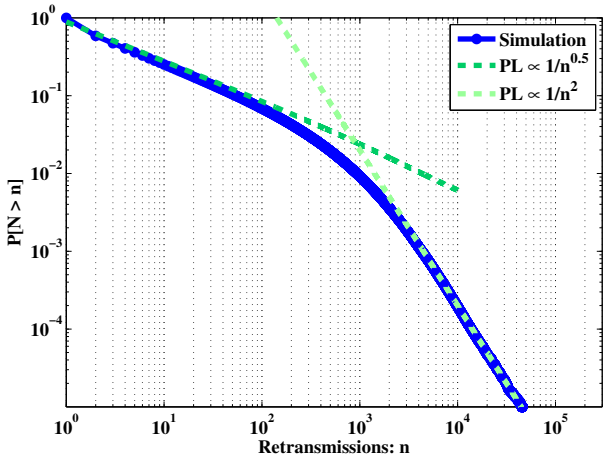


Figure 3.6: Example (a). Exact asymptotes from (3.6) and (3.7) for a two state channel where $\alpha_1 = 2$ and $\alpha_2 = 1/2$.

Next, we extend the preceding observation to a three-state channel where the availability periods are exponentially distributed with parameters $\mu_1 = 2$, $\mu_2 = 1$ and $\mu_3 = 0.5$, and the packet sizes are also exponential of unit mean. The steady state probabilities are $\pi_1 = 0.98$,

$\pi_2 = 0.01$ and $\pi_3 = 0.01$.

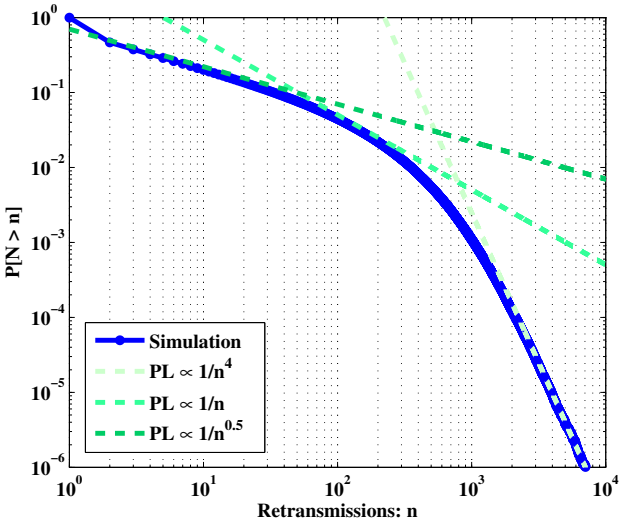


Figure 3.7: Example (b). Asymptotics of $\mathbb{P}[N > n]$ for a three-state channel.

This scenario also corresponds to a channel that is dominated by the dynamics of the 'bad' state. In this case, the lighter asymptotic tail starts to fit later in the distribution. In Fig. 3.7, we also observe transitions between power laws with different indices, determined by the values of α_k in (3.1). Although the 'best case' eventually dominates, the other states are still present and affect the main body of the distribution of N. In the region $n < 10^2$, the distribution is actually a power law with exponent $\alpha = 0.5$, which would correspond to an unstable system if the channel was uncorrelated.

The above discussion is potentially useful in engineering design. Our results imply that the tail distribution of the delay in a dynamic channel will be light as long as there is at least one state that generates light tail asymptotics. This claim might be unrealistic since, as shown in this section, an optimistic design will expose the system to high variability if the 'good' state is relatively rare. The main body of the distribution is characterized by different power laws and a mixture of distributions in between that are, in principle, much heavier than the tail. This situation must be treated with caution in order to guarantee system stability and smooth performance for all n. As illustrated in the last two examples, the main

body can exhibit power law asymptotics with index $\alpha < 1$; in practice, this corresponds to a system with zero throughput. If the design does not account for this behavior, it is highly likely that the system will achieve poor performance for a considerably long period of time.

In most cases, the system will transit to the lightest tail early enough to eliminate the extreme phenomena resulting from heavier distributions. The best strategy under unpredictable situations is to utilize channel feedback, possibly combined with dynamic fragmentation based on the number of unsuccessful retransmission attempts.

3.5 Proofs

In this section, we present the proof for the upper bound of Theorem 3.1 as well as the proofs of Theorems 3.2, 3.3.

Proof of Theorem 3.1. Here, we prove the *upper bound*. Similarly as before

$$\begin{aligned}
\mathbb{P}[N > n|L] &= \mathbb{E}\left[\prod_{k=1}^{K} \mathbb{P}(L > A|J=k)^{N_n^k}\right] \\
&\leq \mathbb{E}\left[\prod_{k=1}^{K} \mathbb{P}(L > A|J=k)^{(1-\epsilon)\pi_k n} \mathbf{1}\left\{N_n^k \geq (1-\epsilon)\pi_k n\right\} + \prod_{k=1}^{K} \mathbf{1}\left\{N_n^k \leq (1-\epsilon)\pi_k n\right\}\right] \\
&= \prod_{k=1}^{K} \mathbb{P}(N_n^k \geq (1-\epsilon)\pi_k n)\mathbb{P}(L > A|J=k)^{(1-\epsilon)\pi_k n} + \prod_{k=1}^{K} \mathbb{P}\left(N_n^k \leq (1-\epsilon)\pi_k n\right),
\end{aligned}$$

where $\mathbb{P}(N_n^k \geq (1-\epsilon)\pi_k n) \to 1$, implied by ergodicity and stationarity, whereas, from our main assumption, $\prod_{k=1}^{K} \mathbb{P}(N_n^k \leq (1-\epsilon)\pi_k n) = O(1/n^{\alpha_m+\epsilon})$. Thus,

$$\begin{aligned}
\mathbb{P}(N > n) &= \mathbb{E}[\mathbb{P}(N > n|L)] \\
&\leq (1+\epsilon)^K \mathbb{E} \prod_{k=1}^{K} \left(1 - \bar{F}(L)^{\frac{1}{\alpha_k(1-\epsilon)}}\right)^{\pi_k n(1-\epsilon)} \mathbf{1}\{L \geq x_0\} \\
&+ (1+\epsilon)^K \prod_{k=1}^{K} \left(1 - \bar{F}(x_0)^{\frac{1}{\alpha_k(1-\epsilon)}}\right)^{\pi_k n(1-\epsilon)} + \frac{\epsilon}{n^{\alpha_m+\epsilon}},
\end{aligned}$$

where x_0 is such that (3.2) holds. Now using the elementary inequality $1 - x \leq e^x$, and by

picking n_0 large so that $(1+\epsilon)^K \left(1 - \bar{F}(x_0)^{\frac{1}{\alpha_k(1-\epsilon)}}\right)^{n(1-\epsilon)} \leq \epsilon/n^{\alpha_m+\epsilon}$, for $n \geq n_0$, we obtain

$$\begin{aligned}\mathbb{P}(N > n) &\leq (1+\epsilon)^K \mathbb{E} \prod_{k=1}^{K} \exp\left(-\pi_k n(1-\epsilon)\bar{F}(L)^{\frac{1}{\alpha_k(1-\epsilon)}}\right) \mathbf{1}\{L \geq x_0\} + \frac{2\epsilon}{n^{\alpha_m+\epsilon}} \\ &= (1+\epsilon)^K \mathbb{E}\left[\exp\left(-\sum_{k=1}^{K} \pi_k n(1-\epsilon) U^{\frac{1}{\alpha_k(1-\epsilon)}}\right)\right] \mathbf{1}\{L \geq x_0\} + \frac{2\epsilon}{n^{\alpha_m+\epsilon}},\end{aligned}$$

where the first integral, by picking x_0 large, is derived similarly as before. Thus

$$\begin{aligned}\mathbb{P}(N > n) &\leq \frac{h_\epsilon \Gamma(\alpha_m(1-\epsilon)+1)}{(\pi_m n)^{\alpha_m(1-\epsilon)}} + \frac{2\epsilon}{n^{\alpha_m+\epsilon}} \\ &\leq \frac{h_\epsilon \Gamma(\alpha_m(1-\epsilon)+1)}{(\pi_m n)^{\alpha_m(1-\epsilon)}} + \frac{2\epsilon}{(\pi_m n)^{\alpha_m(1-\epsilon)}}.\end{aligned}$$

Next, we set $H_\epsilon = (h_\epsilon \Gamma(\alpha_m(1-\epsilon)+1) + 2\epsilon)/\pi_m^{\alpha_m(1-\epsilon)}$, and after taking the logarithm, we obtain

$$\log \mathbb{P}(N > n) \leq \log H_\epsilon - \alpha_m(1-\epsilon)\log n,$$

and by picking n_0 such that $\log H_\epsilon \leq \alpha_m \epsilon \log n$, we have

$$\log \mathbb{P}(N > n) \leq -\alpha_m(1-2\epsilon)\log n.$$

Last, replacing ϵ with $\epsilon/2$ yields

$$\frac{\log \mathbb{P}(N > n)}{\log n} \leq -(1-\epsilon)\alpha_m. \tag{3.8}$$

Letting $\epsilon \to 0$ in both (3.5) and (3.8) completes the proof. □ □

Proof of Theorem 3.2. Without loss of generality, we may assume that a regularly varying $\Phi(.)$ is eventually absolutely continuous, strictly monotone and locally bounded for $x > 0$

since we can always find an absolutely continuous and strictly monotone function such that

$$\Phi^*(x) = \begin{cases} \alpha \int_1^x \Phi(s) s^{-1} ds, & x \geq 1 \\ 0, & 0 \leq x < 1, \end{cases} \tag{3.9}$$

which for x large enough satisfies $\Phi(x) \sim \Phi^*(x)$. For the rest of the proof, we will use $\Phi(x)$ to denote $\Phi^*(x)$.

We also note that for positive h, H we have the following properties, for all $h \leq z \leq H$ and large n,

$$\frac{\Phi(n)}{\Phi(n/z)} \geq (1-\epsilon) z^\alpha \qquad \text{and} \qquad \Phi'(n/z)/\Phi(n/z) = \frac{\alpha z}{n}, \tag{3.10}$$

for $n > h$.

Therefore, for any $0 < \epsilon < 1$ and $x \geq x_0$, we have

$$1/\Phi_k^{\leftarrow}\left((1+\epsilon)\bar{F}(x)^{-1}\right) \leq \bar{G}_k(x) \leq 1/\Phi_k^{\leftarrow}\left((1-\epsilon)\bar{F}(x)^{-1}\right), \tag{3.11}$$

where $\Phi_k^{\leftarrow}(x)$ denotes the inverse function of $\Phi_k^*(\cdot)$; since $\Phi_k^*(x)$ is monotone, for all $x \geq 1$, its inverse exists.

We begin with the *lower bound*. Following the same arguments as in the proof of Theorem 3.1, we obtain

$$\begin{aligned}
\mathbb{P}[N > n|L] &= \mathbb{P}[L > A_1, L > A_2 \ldots, L > A_n] \\
&= \mathbb{E}[\mathbb{P}(L > A_j, 1 \leq j \leq n | J_1, \ldots, J_n)] \\
&= \mathbb{E}\left[\prod_{k=1}^K \mathbb{P}(L > A | J = k)^{N_n^k}\right] \\
&\geq \mathbb{E}\left[\prod_{k=1}^K \mathbb{P}(L > A | J = k)^{(1+\epsilon)\pi_k n} \mathbf{1}\left\{N_n^k \leq (1+\epsilon)\pi_k n\right\}\right] \\
&\geq (1-\epsilon)^K \prod_{k=1}^K (1 - \bar{G}_k(L))^{(1+\epsilon)\pi_k n},
\end{aligned}$$

which follows from recalling that $\mathbb{P}(N_n^k \leq (1+\epsilon)\pi_k n) = 1$ for $n \geq n_0$. Next, using our main assumption and the elementary inequality $1 - x \geq e^{-(1+\epsilon)x}$ for small x, we obtain

$$\begin{aligned}\mathbb{P}[N > n] &= \mathbb{E}[P(N > n|L)] \\ &\geq (1-\epsilon)^K \mathbb{E} \prod_{k=1}^{K} \left(1 - \frac{1}{\Phi_k^{\leftarrow}((1-\epsilon)\bar{F}(L)^{-1})}\right)^{\pi_k n(1+\epsilon)} \mathbf{1}\{L \geq x_n\}\end{aligned}$$

where, we can choose x_n such that $\Phi^{\leftarrow}((1-\epsilon)\bar{F}(x_n)^{-1}) = n/H$, for n large and $H > 0$, and thus

$$\begin{aligned}\mathbb{P}[N > n] &\geq (1-\epsilon)^K \mathbb{E} \prod_{k=1}^{K} \exp\left(-\frac{n\pi_k(1+\epsilon)^2}{\Phi_k^{\leftarrow}((1-\epsilon)U^{-1})}\right) \mathbf{1}\{U \leq \bar{F}(x_n)\} \\ &= (1-\epsilon)^K \int_0^{\bar{F}(x_n)} \exp\left(-n\sum_{k=1}^{K} \frac{\pi_k(1+\epsilon)^2}{\Phi_k^{\leftarrow}((1-\epsilon)u^{-1})}\right) du \\ &\geq (1-\epsilon)^K \int_0^{\bar{F}(x_n)} \exp\left(-\frac{n(1+\epsilon)^3\pi_m}{\Phi_m^{\leftarrow}((1-\epsilon)u^{-1})}\right) du,\end{aligned}$$

where we observe that $\bar{F}(L) = U$, where U is uniform in $(0,1)$, and that for large n, $\sum_{k\neq m} \pi_k \Phi_m^{\leftarrow}((1-\epsilon)u^{-1})/\pi_m \Phi_k^{\leftarrow}((1-\epsilon)u^{-1}) \leq \epsilon$, for small u. Next, by changing the variables $z = n/\Phi_m^{\leftarrow}((1-\epsilon)u^{-1})$, we obtain for small $h > 0$,

$$\mathbb{P}[N > n] \geq (1-\epsilon)^{K+1} \int_h^H e^{-(1+\epsilon)^3\pi_m z} \frac{\Phi_m'(n/z)}{\Phi_m^2(n/z)} \frac{n}{z^2} dz,$$

and by the properties of regularly varying functions [see (3.10)], it follows that

$$\begin{aligned}\mathbb{P}[N > n] &\geq \frac{(1-\epsilon)^{K+1}\alpha_m}{\Phi_m(n)} \int_h^H e^{-(1+\epsilon)^3\pi_m z} z^{\alpha_m - 1} dz \\ &\geq \frac{(1-\epsilon)\alpha_m}{\pi_m^{\alpha_m}\Phi_m(n)} \int_h^{H\pi_m} e^{-y} y^{\alpha_m - 1} dy,\end{aligned}$$

which is derived after replacing $(1-\epsilon)^{K+1}/(1+\epsilon)^{3\alpha_m}$ with $(1-\epsilon)$ and by change of variables.

Last, letting $H \to \infty$ and $h \to 0$, we obtain

$$\mathbb{P}[N > n] \geq (1-\epsilon)\frac{\Gamma(\alpha_m+1)}{\pi_m^{\alpha_m}\Phi(n)}, \tag{3.12}$$

which proves the lower bound.

For the *upper bound*, similarly as before, we obtain

$$\begin{aligned}\mathbb{P}[N > n|L] &\leq \mathbb{E}\left[\prod_{k=1}^{K}(1-\bar{G}_k(L))^{(1-\epsilon)\pi_k n}\mathbf{1}\left\{N_n^k \geq (1-\epsilon)\pi_k n\right\}\right] + \mathbb{E}\left[\prod_{k=1}^{K}\mathbf{1}\left\{N_n^k \leq (1-\epsilon)\pi_k n\right\}\right] \\ &\leq (1+\epsilon)^K\prod_{k=1}^{K}(1-\bar{G}_k(L))^{(1-\epsilon)\pi_k n} + \prod_{k=1}^{K}\mathbb{P}\left(N_n^k \leq (1-\epsilon)\pi_k n\right),\end{aligned}$$

where, by ergodicity and stationarity, we recall that $\mathbb{P}(N_n^k \geq (1-\epsilon)\pi_k n) \to 1$ whereas, from our main assumption, $\prod_{k=1}^{K}\mathbb{P}(N_n^k \geq (1-\epsilon)\pi_k n) = O(1/n^{\alpha_m+\epsilon})$. Thus,

$$\begin{aligned}\mathbb{P}(N > n) &= \mathbb{E}[\mathbb{P}(N > n|L)] \\ &\leq (1+\epsilon)^K\mathbb{E}\prod_{k=1}^{K}\left(1 - \frac{1}{\Phi_k^{\leftarrow}((1+\epsilon)\bar{F}(L)^{-1})}\right)^{\pi_k n(1-\epsilon)}\mathbf{1}\{L \geq x_0\} \\ &+ (1+\epsilon)^K\mathbb{E}\prod_{k=1}^{K}\left(1-\bar{G}_k(x_0)\right)^{\pi_k n(1-\epsilon)} + o\left(\frac{1}{n^{\alpha_m+\epsilon}}\right) \\ &:= I_1 + o\left(\frac{1}{n^{\alpha_m+\epsilon}}\right),\end{aligned}$$

where we pick x_0 such that $(1+\epsilon)^K\left(1-\bar{G}(x_0)\right)^{\pi_k n(1-\epsilon)} \leq o(1/n^{\alpha_m+\epsilon})$, for $n \geq n_0$. Now using the elementary inequality $1-x \leq e^x$, we have

$$\begin{aligned}I_1 &\leq (1+\epsilon)^K\mathbb{E}\exp\left(-\sum_{k=1}^{K}\frac{\pi_k n(1-\epsilon)}{\Phi_k^{\leftarrow}((1+\epsilon)\bar{F}(L)^{-1})}\right) \\ &\leq (1+\epsilon)^K\int_0^1\exp\left(-\sum_{k=1}^{K}\frac{\pi_k n(1-\epsilon)}{\Phi_k^{\leftarrow}((1+\epsilon)/u)}\right)du \\ &\leq (1+\epsilon)^K\int_0^1\exp\left(-\frac{\pi_m n(1-\epsilon)^2}{\Phi_m^{\leftarrow}((1+\epsilon)/u)}\right)du,\end{aligned}$$

where we argue similarly as in the preceding proof for the lower bound. Then, changing the variables $z = n/\Phi_m^{\leftarrow}((1+\epsilon)/u)$ yields

$$\begin{aligned} I_1 &\leq (1+\epsilon)^K \int_{1/\Phi_m(n/h)}^{1/\Phi_m(n/e^m)} \exp\left(-\frac{\pi_m n(1-\epsilon)^2}{\Phi_m^{\leftarrow}((1+\epsilon)/u)}\right) du \\ &+ \sum_{k=m}^{\lceil \log(n/n_\epsilon)\rceil} e^{-\pi_m e^k} \mathbb{P}\left[e^k \leq \frac{n}{\Phi_m^{\leftarrow}((1+\epsilon)U^{-1})} \leq e^{k+1}\right] + e^{-n/n_\epsilon} \\ &:= I_{11} + I_{12} + I_{10}. \end{aligned}$$

First, we compute I_{11} as

$$\begin{aligned} I_{11} &\leq (1+\epsilon)^{K+1} \int_h^{e^m} e^{-\pi_m(1-\epsilon)^2 z} \frac{\Phi_m'(n/z)n}{\Phi_m^2(n/z)z^2} dz \\ &\leq \frac{(1+\epsilon)^{K+1}\alpha_m}{\Phi_m(n)} \int_h^{e^m} e^{-\pi_m(1-\epsilon)^2 z} z^{\alpha_m - 1} dz \\ &\leq \frac{(1+\epsilon)\alpha_m}{\pi_m^{\alpha_m}\Phi_m(n)} \int_{h(1-\epsilon)^2\pi_m}^{e^m} e^{-z} z^{\alpha_m - 1} dz, \end{aligned}$$

where we replace $(1+\epsilon)^{K+1}/(1-\epsilon)^{2\alpha_m}$ with $(1+\epsilon)$. Now, I_{12} becomes

$$I_{12} \leq \sum_{k=m}^{\lceil \log(n/n_\epsilon)\rceil} \frac{e^{-\pi_m e^k}}{\Phi_m(n/e^{k+1})} \leq \sum_{k=m}^{\lceil \log(n/n_\epsilon)\rceil} \frac{e^{-\pi_m e^k}(1+\epsilon)^{k+1}}{\Phi_m(n)} \leq o\left(\frac{1}{\Phi_m(n)}\right),$$

since the preceding sum is finite and $\Phi(n)/\Phi(n/e^k) \leq (1+\epsilon)^k$ for all $n \geq n_0$.

Last, $I_{10} = o(1/n^{\alpha_m+\epsilon}) = o(1/\Phi_m(n))$ from our main assumption. Therefore,

$$\mathbb{P}[N > n]\Phi_m(n)\pi_m^{\alpha_m} \leq (1+\epsilon)\alpha_m \int_0^{e^m} e^{-z} z^{\alpha_m - 1} dz + o(1). \tag{3.13}$$

Note that passing $m \to \infty$ in the first term of (3.13), yields that for all $n \geq n_0$,

$$\frac{\mathbb{P}[N > n]\Phi_m(n)\pi_m^{\alpha_m}}{\Gamma(\alpha_m + 1)} \leq 1 + 2\epsilon.$$

After replacing 2ϵ with $\epsilon/2$, we obtain the upper bound

$$\mathbb{P}[N > n] \leq (1+\epsilon)\frac{\Gamma(\alpha_m+1)}{\pi_m^{\alpha_m}\Phi_m(n)}, \tag{3.14}$$

which, along with (3.12), finishes the proof. □ □

Proof of Theorem 3.3. In the following proof, we use the notation $(x \wedge y) = \min(x, y)$ to refer to the minimum of x and y. First we prove the *upper bound*.

For any $0 < \delta < 1$, we have

$$\begin{aligned}
\mathbb{P}[T > t] &= \mathbb{P}\left[\sum_{i=1}^{N-1} A_i + L > t\right] \\
&\leq \mathbb{P}\left[\sum_{i=1}^{N-1} (A_i \wedge L) > (t - t^{1-\delta}), L \leq t^{1-\delta}\right] + \mathbb{P}[L > t^{1-\delta}] \\
&\leq \mathbb{P}\left[\sum_{i=1}^{N} (A_i \wedge t^{1-\delta}) > (t - t^{1-\delta})\right] + \mathbb{P}[L > t^{1-\delta}] \\
&\leq \mathbb{P}\left[\sum_{i=1}^{N} (A_i \wedge t^{1-\delta}) > (1-\epsilon)t, N \leq t^{1-\delta}\right] + \mathbb{P}\left[N > t^{1-\delta}\right] + \mathbb{P}[L > t^{1-\delta}] \\
&\triangleq I_1 + I_2 + I_3,
\end{aligned}$$

where in the third inequality, we use $t - t^{1-\delta} \approx (1-\epsilon)t$ for large t. First, I_3 is upper bounded by

$$I_3 \leq \frac{\mathbb{E}[L^{\alpha_m}\mathbf{1}\{L > t^{1-\delta}\}]}{t^{\alpha_m(1-\delta)}} = o(1/t^{\alpha_m(1-\delta)}), \tag{3.15}$$

since the condition $\mathbb{E}(A^{1+\theta}) < \infty$, together with our main assumption, imply that for x_0 as in (3.2),

$$\begin{aligned}
\mathbb{E}[L^{\alpha_m}] &= x_0 + \int_{x_0}^{\infty} \mathbb{P}[L^{\alpha_m} > x]dx \leq x_0 + \int_{x_0}^{\infty} \mathbb{P}[A > x^{1/\alpha_m}|J = m]^{\alpha_m(1-\epsilon)}dx \\
&\leq x_0 + \int_{x_0}^{\infty} \frac{\left(\mathbb{E}A^{1+\theta}\right)^{\alpha_m(1-\epsilon)}}{x^{(1+\theta)(1-\delta)(1-\epsilon)}}dx < \infty,
\end{aligned}$$

as $t \to \infty$ for $1 + \theta > 1/(1-\delta)(1-\epsilon)$.

Next, for I_1, we have

$$\begin{aligned} I_1 &= \mathbb{E}\mathbb{P}\left[\sum_{i=1}^{t^{1-\delta}} (A_i \wedge t^{1-\delta}) > (1-\epsilon)t | J_1, J_2, \ldots, J_{\lfloor t^{1-\delta} \rfloor}\right] \\ &\leq \mathbb{E}e^{-\theta(1-\epsilon)t}\mathbb{E}\exp\left[\sum_{i=1}^{t^{1-\delta}} \theta(A_i \wedge t^{1-\delta}) > (1-\epsilon)t | J_1, J_2, \ldots, J_{\lfloor t^{1-\delta} \rfloor}\right] \\ &= \mathbb{E}e^{-\theta(1-\epsilon)t} \prod_{i=1}^{t^{1-\delta}} \mathbb{E}\exp\left[\theta(A_i \wedge t^{1-\delta}) | J_i\right], \end{aligned}$$

which follows by applying the exponential Chebyshev's inequality for $\theta > 0$. Now, observe that $\mathbb{E}[A_i|J_i] \leq \max_{k=1,\ldots,K} \mathbb{E}[A|J=k] =: \mu_m$ and using the inequality $e^x \leq 1 + xe^y, 0 \leq x \leq y$, we upper bound the exponential moments of $X_i := (A_i \wedge t^{1-\delta}|J_i)$ by

$$\mathbb{E}\exp\left[\theta(A_i \wedge t^{1-\delta})|J_i\right] \leq 1 + e^{\theta t^{1-\delta}}\theta\mathbb{E}(A_i|J_i) \leq 1 + e^{\theta t^{1-\delta}}\theta\mu_m \leq \exp\left(\theta\mu_m e^{\theta t^{\delta-1}}\right),$$

which renders

$$\begin{aligned} I_1 &\leq e^{-\theta(1-\epsilon)t}\exp\left(t^{1-\delta}\theta\mu_m e^{\theta t^{\delta-1}}\right) \\ &= e^{-(1-\epsilon)t^{\delta}}e^{\mu_m e} \leq o\left(t^{-\alpha_m(1-\delta)}\right), \end{aligned} \tag{3.16}$$

where we pick $\theta = t^{\delta-1}$.

From Theorem 3.2, we recall that for $0 < \delta < 1$,

$$\lim_{t\to\infty} \frac{\log \mathbb{P}\left[N > t^{1-\delta}\right]}{\log t} = -(1-\delta)\alpha_m,$$

which, along with (3.15) and (3.16), and passing $\delta \to 0$, imply

$$\limsup_{t\to\infty} \frac{\log \mathbb{P}[T > t]}{\log t} \leq -\alpha_m. \tag{3.17}$$

Next, we establish the *lower bound*. It follows easily that

$$\begin{aligned}
\mathbb{P}[T > t] &= \mathbb{P}\left[\sum_{i=1}^{N-1} A_i + L > t\right] \\
&\geq \mathbb{P}\left[\sum_{i=1}^{N-1} A_i > t, N \geq t^{1+\delta} + 1\right] \\
&\geq \mathbb{P}\left[N \geq t^{1+\delta} + 1\right] - \mathbb{P}\left[\sum_{i=1}^{t^{1+\delta}} A_i \leq t\right] \\
&\geq \mathbb{P}\left[N \geq t^{1+\delta} + 1\right] - \mathbb{E}\left[\mathbb{P}\left[\sum_{i=1}^{t^{1+\delta}} A_i \leq t | J_1, J_2, \ldots, J_{\lfloor t^{1-\delta} \rfloor}\right]\right] \\
&:= I_1 - I_2.
\end{aligned}$$

Now, we can show that $I_2 \leq o(t^{-\alpha_m(1+\delta)})$, by similar arguments as in the proof of (3.16) for the proof for the upper bound; we omit the details.

Regarding I_1, we recall from Theorem 3.1 that for $0 < \delta < 1$, we have

$$\lim_{t\to\infty} \frac{\log \mathbb{P}\left[N > t^{1+\delta} + 1\right]}{\log t} = -(1+\delta)\alpha_m,$$

and thus, by passing $\delta \to 0$,

$$\liminf_{t\to\infty} \frac{\log \mathbb{P}[T > t]}{\log t} \geq -\alpha_m. \tag{3.18}$$

Finally, combining (3.17) and (3.18) concludes the proof. □

Appendix

In this Appendix, we provide the informal derivation of formulas (3.6) and (3.7) of Section 3.4.1.

Proof of (3.6) *and* (3.7). Starting from equation (3.3), assuming that $\bar{F}(x) \approx \bar{G}_k(x)^{\alpha_k}, k = 1, 2$, and using similar arguments as in the derivation of (3.3)-(4.7), we informally argue that

$$\begin{aligned}
\mathbb{P}[N > n] &\approx \mathbb{E}[(1 - \bar{G}_1(L))^{\pi_1 n}(1 - \bar{G}_2(L))^{\pi_2 n}] \\
&\approx \mathbb{E}[(1 - \bar{F}(L)^{1/\alpha_1})^{\pi_1 n}(1 - \bar{F}(L)^{1/\alpha_2})^{\pi_2 n}] \\
&\approx \mathbb{E}[(1 - U^{1/\alpha_1})^{\pi_1 n}(1 - U^{1/\alpha_2})^{\pi_2 n}] \\
&\approx \mathbb{E}[e^{-\pi_1 n U^{1/\alpha_1} - \pi_2 n U^{1/\alpha_2}}],
\end{aligned}$$

since $\bar{F}(L) = U$, where U is uniformly distributed in (0,1). Thus,

$$\begin{aligned}
\mathbb{P}[N > n] &\approx \int_0^1 e^{-\pi_1 n u^{1/\alpha_1} - \pi_2 n u^{1/\alpha_2}} du \qquad (3.19) \\
&= \frac{\alpha_2}{n^{\alpha_2}\pi_2^{\alpha_2}} \int_0^{n\pi_2} e^{-z - \frac{\pi_1 n z^{\alpha_2/\alpha_1}}{(n\pi_2)^{\alpha_2/\alpha_1}}} z^{\alpha_2 - 1} dz,
\end{aligned}$$

which follows by changing the variables $z = \pi_2 n u^{1/\alpha_2}$. Now, let $\delta := \alpha_2/\alpha_1 < 1$ and

$P_2 := \pi_1/\pi_2^{\delta}$, so that

$$\mathbb{P}[N > n] \approx \frac{\alpha_2}{n^{\alpha_2}\pi_2^{\alpha_2}} \int_0^{n\pi_2} e^{-z-n^{1-\delta}P_2 z^{\delta}} z^{\alpha_2-1} dz$$
$$= \frac{\alpha_2}{n^{\alpha_2}\pi_2^{\alpha_2}} \int_0^{n\pi_2} e^{-z} z^{\alpha_2-1} \left(1 - n^{1-\delta}P_2 z^{\delta} + \frac{(n^{1-\delta}P_2)^2 z^{2\delta}}{2} - \cdots + \frac{(-n^{1-\delta}P_2)^i z^{i\delta}}{i!} + \dots\right) dz,$$

by the Taylor expansion of the function $e^x = \sum_{i=0}^{1} x^i/i!$. Now by extending the integral to infinity we have

$$\mathbb{P}[N > n] \approx \frac{\alpha_2}{n^{\alpha_2}\pi_2^{\alpha_2}} \int_0^{\infty} e^{-z} z^{\alpha_2-1} \left(1 - n^{1-\delta}P_2 z^{\delta} + \frac{(n^{1-\delta}P_2)^2 z^{2\delta}}{2} - \cdots + \frac{(-n^{1-\delta}P_2)^i z^{i\delta}}{i!} + \dots\right) dz$$
$$= \frac{\alpha_2}{n^{\alpha_2}\pi_2^{\alpha_2}} \left(\int_0^{\infty} e^{-z} z^{\alpha_2-1} dz - n^{1-\delta}P_2 \int_0^{\infty} e^{-z} z^{\alpha_2+\delta-1} dz + \frac{(n^{1-\delta}P_2)^2}{2} \int_0^{\infty} e^{-z} z^{\alpha_2+2\delta-1} dz - \right.$$
$$\left. \cdots + \frac{(-n^{1-\delta}P_2)^i}{i!} \int_0^{\infty} e^{-z} z^{\alpha_2+i\delta-1} dz + \dots\right)$$
$$= \frac{\alpha_2}{n^{\alpha_2}\pi_2^{\alpha_2}} \left(\Gamma(\alpha_2) - n^{1-\delta}P_2\Gamma(\alpha_2+\delta) + \frac{(n^{1-\delta}P_2)^2\Gamma(\alpha_2+2\delta)}{2} - \cdots + \frac{(-n^{1-\delta}P_2)^i\Gamma(\alpha_2+i\delta)}{i!} + \dots\right),$$

which follows immediately from the definition of the gamma function $\Gamma(\alpha) = \int_0^{\infty} e^{-z} z^{\alpha-1} dz$. This yields the explicit form

$$\mathbb{P}[N > n] \approx \frac{\alpha_2}{n^{\alpha_2}\pi_2^{\alpha_2}} \sum_{i=0}^{\infty} \frac{(-n^{1-\delta}P_2)^i \Gamma(\alpha_2+i\delta)}{i!}.$$

By the same approach, starting from (3.19) and changing of variables as $z = n\pi_1 u^{\alpha_1}$, we obtain

$$\mathbb{P}[N > n] \approx \int_0^1 e^{-\pi_1 n u^{1/\alpha_1} - \pi_2 n u^{1/\alpha_2}} du$$
$$= \frac{\alpha_1}{n^{\alpha_1}\pi_1^{\alpha_1}} \int_0^{n\pi_1} e^{-z - \frac{\pi_2 n z^{\alpha_1/\alpha_2}}{(n\pi_1)^{\alpha_1/\alpha_2}}} z^{\alpha_1-1} dz.$$

Now, for $P_1 := \pi_2/\pi_1^{1/\delta}$ and using Taylor expansion of e^x, we have

$$\begin{aligned}\mathbb{P}[N > n] &\approx \frac{\alpha_1}{n^{\alpha_1}\pi_1^{\alpha_1}} \int_0^{n\pi_1} e^{-z-n^{1-1/\delta}P_1 z^{1/\delta}} z^{\alpha_1-1} dz \\ &\approx \frac{\alpha_1}{n^{\alpha_1}\pi_1^{\alpha_1}} \int_0^{\infty} e^{-z} z^{\alpha_1-1} \left(1 - n^{1-1/\delta}P_1 z^{1/\delta} + \frac{(n^{1-1/\delta}P_1)^2 z^{2/\delta}}{2} - \cdots + \frac{(-n^{1-1/\delta}P_1)^i z^{i/\delta}}{i!} + \dots\right) dz,\end{aligned}$$

for large n. By identical arguments as before,

$$\begin{aligned}\mathbb{P}[N > n] \approx \frac{\alpha_1}{n^{\alpha_1}\pi_1^{\alpha_1}} \Bigg(&\Gamma(\alpha_1) - n^{1-1/\delta}P_1\Gamma(\alpha_1 + 1/\delta) + \frac{(n^{1-1/\delta}P_1)^2\Gamma(\alpha_1 + 2/\delta)}{2} - \dots \\ &\cdots + \frac{(-n^{1-1/\delta}P_1)^i\Gamma(\alpha_1 + i/\delta)}{i!} + \dots\Bigg),\end{aligned}$$

which yields the explicit form

$$\mathbb{P}[N > n] \approx \frac{\alpha_1}{n^{\alpha_1}\pi_1^{\alpha_1}} \sum_{i=0}^{\infty} \frac{(-n^{1-1/\delta}P_1)^i\Gamma(\alpha_1 + i/\delta)}{i!}.$$

□

Chapter 4

Instability of Sharing Systems in the Presence of Retransmissions

Retransmissions represent a primary failure recovery mechanism on all layers of communication network architecture. Similarly, fair sharing, e.g. processor sharing (PS), is a widely accepted approach to resource allocation among multiple users. In previous chapters, it has been shown that retransmissions in failure-prone, e.g. wireless ad hoc, networks can cause heavy tails and long delays. In this chapter, we discover a new phenomenon: PS-based scheduling induces complete instability in the presence of retransmissions, regardless of the job sizes and the traffic intensity. This phenomenon occurs even when the job sizes are bounded/fragmented, e.g., deterministic. Our work demonstrates that scheduling one job at a time, such as first-come-first-serve, achieves a larger stability region and should be preferred in these systems.

4.1 Introduction

High variability and frequent failures characterize the majority of large-scale systems, e.g., infrastructure-less wireless networks, cloud/parallel computing systems, etc. The nature of these systems imposes the employment of failure recovery mechanisms to guarantee their

good performance. One of the most straightforward and widely used recovery mechanisms is to simply restart all the interrupted jobs from the beginning after a failure occurs. In communication systems, restart mechanisms lie at the core of the network architecture where retransmissions are used on all protocol layers to guarantee data delivery in the presence of channel failures, e.g., automatic repeat request (ARQ) protocol [19], contention based ALOHA type protocols in the medium access control (MAC) layer, end-to-end acknowledgements in the transport layer, HTTP downloading scheme in the application layer, and others.

Furthermore, sharing is a primary approach to fair scheduling and efficient management of the available resources. Fair allocation of the network resources among different users can be highly beneficial for increasing throughput and utilization. For instance, CDMA is a multiple access method used in communication networks, where several users can transmit information simultaneously over a single channel via sharing the available bandwidth. Another example is Processor Sharing (PS) scheduling [36] where the capacity is equally shared between multiple classes of customers. In *Generalized* PS (GPS) [32], service allocation is done according to some fixed weights. The related *Discriminatory* PS (DPS) [17, 23, 31] is used in computing to model the Weighted Round Robin (WRR) scheduling, while it is also used in communications, as a flow level model of heterogenous TCP connections. Similarly, fair queuing (FQ) is a scheduling algorithm where the link capacity is fairly shared among active network flows; in weighted fair queuing (WFQ), which is the discretized version of GPS, different scheduling priorities are assigned to each flow.

In general, PS-based scheduling disciplines have been widely used in computer and communication networks. Early investigations of PS queues were motivated by applications in multiuser computer systems [22]. The M/G/1 PS queue has been studied extensively in the literature [35]. In the case of the M/M/1 PS system, the conditional Laplace transform of the waiting time was derived in [22]. The importance of scheduling in the presence of heavy tails was first recognized in [18], and later, in [29], the M/G/1 PS queue was studied assuming subexponential job sizes; see also [29] for additional references.

In [34], it was proven that, although there are policies known to optimize the sojourn time tail under a large class of heavy-tailed job sizes (e.g., PS and SRPT) and there are policies known to optimize the sojourn time tail in the case of light-tailed job sizes, e.g., first-come-first-serve (FCFS), no policies are known to optimize the sojourn time tail across both light and heavy-tailed job size distributions. Indeed, such policies must "learn" the job size distribution in order to optimize the sojourn time tail. In the heavy-tailed scenarios, any scheduling policy that assigns the server exclusively to a very large job, e.g., FCFS, may induce long delays, in which case, sharing guarantees better performance.

With regard to retransmissions, it was first recognized in [24, 25, 26, 9] that restart mechanisms may result in heavy-tailed (power law) delays even if the job sizes and failure rates are light-tailed. In [9], it was shown that the power law delays arise whenever the cumulative hazard functions of the data and failure distributions are proportional.

In this chapter, we study the effects of sharing on the system performance when restarts are employed in the presence of failures. We revisit the well-studied M/G/1 PS queue with a new focus on server failures and restarts. We use the following generic model, which was first introduced in [24] in the application context of computing. The system dynamics is described as a process $\{A_n\}_{n\geq 1}$, where A_n correspond to the periods when the system is available. $\{A_n\}_{n\geq 1}$ is a sequence of i.i.d random variables, independent of the job sizes. In each period of time that the system is available, say A_n, we attempt to execute a job of random size B. If $A_n > B$, we say that the job is successfully completed; otherwise, we restart the job from the beginning in the following period A_{n+1} when the channel is available.

In this work, our main contributions are the following. First, we prove that the M/G/1 PS queue is always unstable, regardless of how light the load is and how small the job sizes may be, see Theorems 4.1 and 4.2 in Section 4.3. This is a new phenomenon, since, contrary to the conventional belief, sharing the service even between very small deterministic jobs can render the system completely unstable when retransmissions/restarts are employed. This instability is strong, in the sense of system having zero throughput. The intuition

is the following. If a large number of jobs arrives in a short period of time, then under the elongated service time distribution induced by sharing, coupled with retransmissions, the queue will keep accumulating jobs that will equally share the capacity, which further exacerbates the problem. Every time a failure occurs, the system resets and the service requirement for each job elongates as the queue size increases. The expected delay until the system clears becomes increasingly long and, consequently, the queue will continue to grow leading to instability. This result also applies to the *discriminatory* PS (DPS) queue, where the service is not shared equally but according to some fixed weights. Next, we remove the Poisson assumption and extend our results to general renewal arrivals in Section 4.4. This demonstrates that instability arises from the interplay between sharing and retransmission/restart mechanisms, rather than any specific characteristics of the arrival process and/or service distribution.

We would also like to emphasize that job fragmentation cannot stabilize the system regardless of how small the fragments are made, since Theorem 4.2 shows instability for any minimum job size $\beta > 0$. Similarly, the system cannot be stabilized by checkpointing regardless of how small the intervals between successive checkpoints are chosen. In our experimental results, we make an interesting observation on the system behavior before it saturates. There exists a transient period, during which the queue appears as if it were stable. Although it may occasionally accumulate a substantial number of jobs, it returns to zero and starts afresh. However, there exists a time when the queue reaches a critical size after which the service rate of the jobs reduces so much that neither of them can depart. Hence, as the queue continues to increase in size, the system becomes unstable.

To contrast these results, in Section 4.3.2, we study the stability of a non-preemptive policy that serves one job at a time under more specific assumptions of Poisson failure rates. To this end, Theorem 4.3 shows that when jobs are bounded, serving one job at a time, e.g., FCFS, always has a non-empty stability region and, thus, performs better than PS.

In order to gain further insight into the system, we then focus on its transient behavior and study the properties of the completion time of a finite number of jobs with no future

arrivals. Specifically, we compare two work-conserving policies: scheduling one job at a time, e.g., FCFS, and PS. Overall, we discover that serving one job at a time exhibits uniformly better performance than PS; compare Theorems 4.7 and 4.8, respectively. Furthermore, under more technical assumptions, and for light-tailed job/failure distributions, we show that PS performs distinctly worse compared to the heavy-tailed ones, and that PS is always unstable.

From an engineering perspective, our results indicate that traditional approaches in existing systems may be inadequate in the presence of failures. This new phenomenon demonstrates the need of revisiting existing techniques to large-scale failure-prone systems, where PS-based scheduling may perform poorly. For example, since PS is unstable even for deterministic jobs, packet fragmentation, which is widely used in communications, cannot alleviate instabilities. Indeed, fragmentation can only postpone the time when the instability occurs, but cannot eliminate the phenomenon; see Example 1 in Section 4.6. Therefore, serving one job at a time, e.g., FCFS, is highly advisable in such systems; see Section 4.3.2.

The chapter is organized as follows. In Section 4.2, we introduce the model along with the necessary definitions and notation. Next, in Section 4.3, we present our main results on the M/G/1 PS queue, which are further extended in Section 4.4 to general renewal arrivals. On the other hand, in Section 4.3.2 we study the stability of non-preemptive policies that serve one job at a time, e.g., FCFS. Later, in Section 4.5, we analyze the transient behavior of the system under two different scheduling policies, e.g., serving one job at a time and PS. Last, Section 4.6 presents our simulation experiments that validate our main theoretical findings, while Section 4.7 concludes the chapter.

4.2 Definitions and Notation

First, we provide the necessary definitions and notation assuming that the jobs are served individually. Consider a generic job of random size B, $B > 0$ a.s., requesting service in a failure-prone system. Without loss of generality, we assume that the system is of unit

capacity. Its dynamics is described as a process $\{A_n\}_{n\geq 1}$ of i.i.d availability periods, where at the end of each period A_n, the system experiences a failure, as shown in Figure 4.1. The channel/server statistics $\{A_n\}_{n\geq 1}$ are independent of the job size B.

Furthermore, we assume that the first failure occurs at time $A_0 \geq 0$, which is independent of $\{A_n\}_{n\geq 1}$ and B. When A_0 is equal in distribution to the excess/residual distribution of A_1, $\{A_n\}_{n\geq 0}$ will be in stationarity. Throughout the chapter, we will use different assumptions on A_0, e.g., $A_0 \equiv 0$, which will be explicitly stated in the corresponding results. Let A be a generic random variable that is equal in distribution to A_1. We denote the complementary cumulative distribution functions for A and B, respectively, as

$$\bar{G}(x) \triangleq \mathbb{P}(A > x) \qquad \text{and} \qquad \bar{F}(x) \triangleq \mathbb{P}(B > x).$$

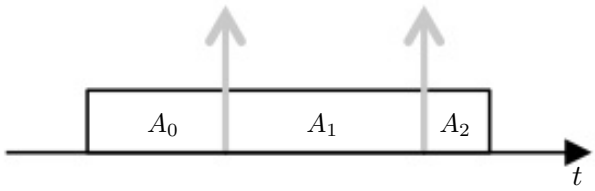


Figure 4.1: System with failures.

At each period of time that the system becomes available, say A_n, we attempt to process a generic job of size B. If $A_n > B$, we say that the job is completed successfully; otherwise, we wait until the next period A_{n+1} when the channel is available and restart the job. A sketch of the model depicting the system is drawn in Figure 4.2.

The number of restarts, N, and the total service time, S, for a job of size B, whose service begins immediately after a first failure A_0 and is served in isolation without preemption are defined as follows.

Definition 4.2.1. *The number of restarts for a generic job of size B is defined as*

$$N \triangleq \inf\{n \geq 1 : A_n > B\}.$$

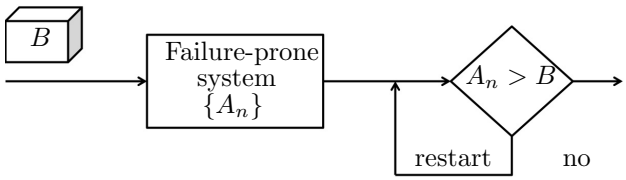


Figure 4.2: Jobs executed in a system with failures.

Definition 4.2.2. *The service time is the total time until a generic job of size B is successfully completed and is denoted as*

$$S \triangleq \sum_{i=1}^{N-1} A_i + B.$$

Note that the preceding definitions will be different when jobs are sharing a server, as in the Processor Sharing discipline. In general, a job B will successfully complete service during an availability period A_{n+1} if there exists $t \leq A_{n+1}$, such that

$$\int_{T_n}^{T_n+t} C_u^B du = B,$$

where C_u^B is the service rate that job B receives at time u and T_n is the time of the n^{th} failure, $T_n = \sum_{i=0}^{n} A_i, n \geq 0$. Note that in general C_u^B depends on the number of jobs at time T_n, the arrival process and the service discipline. We use B_j to denote the service requirement of job j where $\{B_j\}_{j\geq 1}$ is an i.i.d process equal in distribution to B. The failure times $\{A_n\}_{n\geq 0}$, job requirements $\{B_j\}_{j\geq 1}$ and the arrival process are mutually independent.

Throughout the chapter, we use the following standard notation. For any two real functions $f(x)$ and $g(x)$ and fixed $x_0 \in \mathbb{R} \cup \{\infty\}$, we say $f(x) \sim g(x)$ as $x \to x_0$, to denote $\lim_{x\to x_0} f(x)/g(x) = 1$.

4.3 M/G/1 Queue with Restarts

In this section, we discuss the stability of the M/G/1 queue under two scheduling disciplines: Processor Sharing (PS) and non-preemptive one job at a time policy. Throughout this

section, we assume that the arrival process is Poisson with rate $\lambda > 0$. In the following subsection, we show in Theorem 4.2 that the M/G/1 PS queue is unstable under considerable generality. Next, in subsection 4.3.2, we derive the necessary and sufficient condition for the system to be stable when the jobs are processed one at a time and the failure rates are Poisson.

4.3.1 Instability of Processor Sharing Queue

In this section, we show in Theorems 4.1 and 4.2 that the M/G/1 PS queue is unstable when jobs need to restart after failures. We consider a general renewal failure process as defined in Section 4.2. First, in Proposition 4.1, we show that for some initial condition on the queue size, the probability that no job completes service approaches 1, under the mild assumption that jobs are bounded from below by some positive constant β. This is a natural assumption for communication or computing applications where jobs, e.g., files, packets, threads, must have a header to contain the required information, such as destination address, thread id, etc. Hence, the job sizes, in practice, cannot be smaller than a positive constant.

Next, in Theorem 4.1, without any initial condition on the queue size, we prove that after some finite time, no job ever leaves the system; this result is stronger than standard stability theorems. Then, in Corollary 4.1, we draw a weaker conclusion that the queue size grows to infinity, which is also stated in Theorem 4.2. Nevertheless, the latter does not require the assumption on the minimum job size.

We begin with the following proposition. As previously mentioned, in this section we assume a general renewal failure process $\{A_n\}_{n\geq 0}$, as defined in Section 4.2. In the following proposition, we assume that the first failure occurs at $t = 0$, i.e. $A_0 = 0$. The remaining results (Theorems 4.1 and 4.2) allow for an arbitrary delay until the first failure, $0 \leq A_0 < \infty$; this assumption includes the stationary version of $\{A_n\}_{n\geq 0}$, when A_0 has the excess distribution of A.

Proposition 4.1. *Assume that a failure occurs at time $t = 0$, i.e. $A_0 \equiv 0$, and there are $Q_0 \geq k$ jobs in the M/G/1 PS queue. If $\mathbb{E}A < \infty$ and $\mathbb{P}[B \geq \beta] = 1, \beta > 0$, then there*

exists $\theta > 0$*, such that for all* $k \geq 1$

$$\mathbb{P}[\textit{no job ever completes service}] \geq 1 - O(\mathbb{E}A\mathbf{1}(A \geq \beta k) + e^{-\theta k}). \tag{4.1}$$

Proof. Let $T_1 = \sum_{i=1}^{ck} A_i$ be the cumulative time that includes the first ck failures for $t > 0$; to simplify notation we write $\sum_x^y$ to denote $\sum_{\lceil x \rceil}^{\lfloor y \rfloor}$, where $\lceil x \rceil$ is the smallest integer $\geq x$ and $\lfloor y \rfloor$ is the largest integer $\leq y$. Now, define the event $\mathcal{A}_1 \equiv \mathcal{A}_1(k) \triangleq \{A_1 < \beta k, A_2 < \beta k, \ldots, A_{ck} < \beta k\}$. On this event, no job can leave the system since $Q_0 \geq k$ and all of them are at least of size β. Thus, if they were served in isolation, they could not have completed service in the first ck attempts.

Now, let E_1 denote the event that there is no departure in the first ck service attempts and there are at least k arrivals in $(0, T_1]$; we use $Z_{(t_0,t_1]}$ to denote the number of Poisson arrivals in the interval $(t_0, t_1]$, whereas we simply write Z_t for intervals $(0, t]$. Formally,

$$E_1 \supset \underline{E}_1 \triangleq \{Z_{T_1} \geq k, \mathcal{A}_1\},$$

on the set $\{Q_0 \geq k\}$. Note that $\underline{E}_1$ is clearly a subset of E_1, since there may be many other scenarios when no jobs leave the queue either because jobs are larger than β or more than k jobs are sharing the server. Now, observe that

$$\begin{aligned}
\mathbb{P}(\underline{E}_1) &\geq \mathbb{P}(Z_{T_1} \geq k, T_1 \geq 2k/\lambda, \mathcal{A}_1) \\
&\geq \mathbb{P}(Z_{2k/\lambda} \geq k, T_1 \geq 2k/\lambda, \mathcal{A}_1) \\
&\geq \mathbb{P}(Z_{2k/\lambda} \geq k)\mathbb{P}(T_1 \geq 2k/\lambda, \mathcal{A}_1),
\end{aligned}$$

since Poisson arrivals are independent of the failure process. Thus,

$$\mathbb{P}(\underline{E}_1) \geq \mathbb{P}(Z_{2k/\lambda} \geq k) \left(\mathbb{P}(\mathcal{A}_1) - \mathbb{P}(T_1 < 2k/\lambda)\right).$$

First, note that

$$\begin{aligned}\mathbb{P}(Z_{2k/\lambda} \geq k) &= 1 - \mathbb{P}(Z_{2k/\lambda} < k) = 1 - \mathbb{P}(2k - Z_{2k/\lambda} > k) \\ &\geq 1 - e^{-\theta k}\mathbb{E}e^{\theta(2k - Z_{2k/\lambda})} = 1 - e^{\theta k}\mathbb{E}e^{-\theta Z_{2k/\lambda}},\end{aligned}$$

by Cramer's bound for $\theta > 0$. Next, observe that $Z_{2k/\lambda}$ is Poisson with mean $2k$ and thus

$$\mathbb{P}(Z_{2k/\lambda} \geq k) \geq 1 - e^{\theta k}e^{2(e^{-\theta}-1)k} = 1 - e^{-\theta_1 k},$$

where $\theta_1 = 2(1 - e^{-\theta}) - \theta > 0$, for θ small.

Second, observe that

$$\begin{aligned}\mathbb{P}(T_1 < 2k/\lambda) &= \mathbb{P}\left(\sum_{i=1}^{ck} A_i < 2k/\lambda\right) = \mathbb{P}\left(\sum_{i=1}^{ck}(A_i - \mathbb{E}A) < 2k/\lambda - ck\mathbb{E}A\right) \\ &\leq \mathbb{P}\left(\sum_{i=1}^{3k/\lambda\mathbb{E}A} (\mathbb{E}A - A_i) > k/\lambda\right),\end{aligned}$$

by picking $c \triangleq 3/(\lambda\mathbb{E}A)$. Now, let $X_i \triangleq \mathbb{E}A - A_i$, which are bounded from above since $X_i \leq \mathbb{E}A < \infty$, from our main assumption. Therefore, Cramer's large deviation bound implies that

$$\mathbb{P}(T_1 < 2k/\lambda) \leq \mathbb{P}\left(\sum_{i=1}^{3k/\lambda\mathbb{E}A} X_i > k/\lambda\right) \leq H_2 e^{-\theta_2 k},$$

for some $H_2, \theta_2 > 0$.

Therefore,

$$\begin{aligned}\mathbb{P}(\underline{E}_1) &\geq (1 - e^{-\theta_1 k})\left(\mathbb{P}(\mathcal{A}_1) - H_2 e^{-\theta_2 k}\right) \\ &\geq \mathbb{P}(A < \beta k)^{ck} - (e^{-\theta_1 k} + H_2 e^{-\theta_2 k} - H_2 e^{-(\theta_1+\theta_2)k}) \\ &\geq (1 - \mathbb{P}(A \geq \beta k))^{ck} - He^{-\theta k},\end{aligned}$$

where $\theta = \min(\theta_1, \theta_2)$ and $H > 0$ such that $H < (1 + H_2)$. Next, using $1 - x \geq e^{-2x}$ for small x, we have for all $k \geq k_0$

$$\begin{aligned}\mathbb{P}(\underline{E}_1) &\geq e^{-2ck\mathbb{P}(A\geq\beta k)} - He^{-\theta k} \\ &\geq 1 - 2ck\mathbb{P}(A \geq \beta k) - He^{-\theta k} \\ &\geq e^{-4ck\mathbb{P}(A\geq\beta k) - 2He^{-\theta k}}.\end{aligned}$$

Next, at time $\mathcal{T}_1 = T_1$, on event $\underline{E}_1$, the queue has at least $2k$ jobs, i.e., $Q_{\mathcal{T}_1} \geq 2k$, and no jobs have departed. Similarly as before, let $T_2 = \sum_{i=ck+1}^{3ck} A_i$ be the cumulative time that includes the next $2ck$ failures, and define $\mathcal{A}_2 \equiv \mathcal{A}_2(k) = \{A_{ck+1} < 2\beta k, A_{ck+2} < 2\beta k, \ldots, A_{3ck} < 2\beta k\}$. Now, if E_2 is the event that there is no departure in the next $2ck$ attempts and there are at least $2k$ arrivals in $(\mathcal{T}_1, \mathcal{T}_2]$, then $E_2 \supset \underline{E}_2 \triangleq \{Z_{T_2} \geq 2k, \mathcal{A}_2\}$ on $\{Q_{\mathcal{T}_1} \geq 2k\}$; note that $\underline{E}_2$ is independent of $\underline{E}_1$. Then, the probability that no job departs in $(0, \mathcal{T}_2]$, where $\mathcal{T}_2 = T_1 + T_2$, is lower bounded by

$$\begin{aligned}\mathbb{P}(\text{no job departs in } (0, \mathcal{T}_2]) &\geq \mathbb{P}(E_1 \cap E_2) \\ &\geq \mathbb{P}(Z_{T_1} \geq k, \mathcal{A}_1, Q_{\mathcal{T}_1} \geq 2k, Z_{(\mathcal{T}_1, \mathcal{T}_2]} \geq 2k, \mathcal{A}_2) \\ &\geq \mathbb{P}(Z_{T_1} \geq k, \mathcal{A}_1, Z_{T_2} \geq 2k, \mathcal{A}_2) = \mathbb{P}(\underline{E}_1)\mathbb{P}(\underline{E}_2), \qquad (4.2)\end{aligned}$$

since $\{Q_{\mathcal{T}_1} \geq 2k\} \supseteq \{Z_{T_1} \geq k, \mathcal{A}_1\}$ on the set $\{Q_0 \geq k\}$; the remaining statements in this proof should all be considered on $\{Q_0 \geq k\}$.

Next, via identical arguments as before, we obtain

$$\begin{aligned}\mathbb{P}(\underline{E}_2) &\geq \mathbb{P}(Z_{T_2} \geq 2k, T_2 \geq 4k/\lambda, \mathcal{A}_2) \\ &\geq \mathbb{P}(Z_{4k/\lambda} \geq 2k)\left(\mathbb{P}(\mathcal{A}_2) - \mathbb{P}(T_2 < 4k/\lambda)\right) \geq e^{-8ck\mathbb{P}(A\geq 2\beta k) - 2He^{-2\theta k}}.\end{aligned}$$

Therefore, at time $\mathcal{T}_2$, on event $\underline{E}_1 \cap \underline{E}_2$, there are at least $4k$ jobs.

In general, for any n, we can extend the reasoning from (4.2) to obtain

$$\begin{aligned}\mathbb{P}(\text{no job departs in } (0,\mathcal{T}_n]) &\geq \mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n)\\ &\geq \mathbb{P}(Z_{T_1} \geq k, \mathcal{A}_1, Z_{T_2} \geq 2k, \mathcal{A}_2, \ldots, Z_{T_n} \geq 2^{n-1}k, \mathcal{A}_n)\\ &= \mathbb{P}(\underline{E}_1 \cap \underline{E}_2 \cap \cdots \cap \underline{E}_n),\end{aligned}$$

where $\mathcal{T}_n = \sum_{i=1}^{n} T_i$, $T_n = \sum_{i=(2^{n-1}-1)ck+1}^{(2^n-1)ck} A_i$, E_n is the event that there are no departures during $2^{n-1}ck$ attempts and there are at least $2^{n-1}k$ arrivals in $(\mathcal{T}_{n-1}, \mathcal{T}_n)$, and $\underline{E}_n = \{Z_{T_n} \geq 2^{n-1}k, \mathcal{A}_n\}$. Similarly as before,

$$\mathbb{P}(\underline{E}_n) \geq e^{-2^{n+1}ck\mathbb{P}(A \geq 2^{n-1}\beta k) - 2He^{-\theta 2^{n-1}k}}.$$

Hence, using the preceding inequality and the independence of $\underline{E}_i$'s, we obtain

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) &\geq \mathbb{P}(\underline{E}_1 \cap \underline{E}_2 \cap \cdots \cap \underline{E}_n) = \mathbb{P}(\underline{E}_1)\mathbb{P}(\underline{E}_2)\cdots\mathbb{P}(\underline{E}_n)\\ &\geq \prod_{i=1}^{n} e^{-2^{i+1}ck\mathbb{P}(A \geq 2^{i-1}\beta k) - 2He^{-2^{i-1}\theta k}}\\ &= e^{-4\sum_{i=0}^{n-1} 2^i ck\mathbb{P}(A \geq 2^i\beta k) - 2H\sum_{i=0}^{n-1} e^{-2^i\theta k}}\\ &\geq e^{-4\sum_{i=0}^{\infty} 2^i ck\mathbb{P}(A \geq 2^i\beta k) - 2He^{-\theta k}\sum_{i=0}^{\infty} e^{-(2^i-1)\theta k}}.\end{aligned}$$

Now, observe that $\sum_{i=0}^{\infty} e^{-(2^i-1)\theta k} < \infty$, and thus we can pick H such that

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) \geq e^{-4\sum_{i=0}^{\infty} 2^i ck\mathbb{P}(A \geq 2^i\beta k) - He^{-\theta k}}.$$

Furthermore, we observe that

$$\begin{aligned}\sum_{i=0}^{\infty} 2^i ck\mathbb{P}(A \geq 2^i\beta k) &\leq \frac{c}{\beta}\sum_{i=0}^{\infty} \beta k \int_{2^i}^{2^{i+1}} \mathbb{P}(A \geq x\beta k)dx\\ &\leq \frac{c}{\beta}\beta k \int_{1}^{\infty} \mathbb{P}(A \geq x\beta k)dx = \frac{c}{\beta}\int_{\beta k}^{\infty} \mathbb{P}(A \geq y)dy = \frac{c}{\beta}\mathbb{E}A\mathbf{1}(A \geq \beta k),\end{aligned}$$

and thus

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) \geq e^{-4c\beta^{-1}\mathbb{E}A\mathbf{1}(A\geq\beta k)-He^{-\theta k}} \geq 1 - H(\mathbb{E}A\mathbf{1}(A \geq \beta k) + e^{-\theta k}).$$

Last, note that, on $\{Q_0 \geq k\}$,

$$\begin{aligned}\mathbb{P}(\text{no job ever completes service}) \geq \mathbb{P}(\cap_{i=1}^{\infty} E_i) &= \lim_{n\to\infty} \mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) \\ &\geq 1 - H(\mathbb{E}A\mathbf{1}(A \geq \beta k) + e^{-\theta k}),\end{aligned}$$

where the first inequality follows by definition and the second equality from monotone convergence.

Hence, we proved that the statement holds for all $k \geq k_0$. Last, for $k < k_0$, we can choose $H > 1/(\mathbb{E}A\mathbf{1}(A \geq \beta k_0) + e^{-\theta k_0})$, such that $\mathbb{P}(\text{no job ever completes service}|Q_0 \geq k) \geq 0 \geq 1 - H(\mathbb{E}A\mathbf{1}(A \geq \beta k_0) + e^{-\theta k_0}) \geq 1 - H(\mathbb{E}A\mathbf{1}(A \geq \beta k) + e^{-\theta k})$ and thus (4.1) holds trivially. □

We proceed with our main theorem which shows that, after some finite time, no job will ever depart.

Theorem 4.1. *In the M/G/1 PS queue, if* $\mathbb{E}A < \infty$, $0 \leq A_0 < \infty$ *a.s., and* $\mathbb{P}[B \geq \beta] = 1, \beta > 0$, *then*

$$\lim_{t\to\infty} \mathbb{P}(\textit{no job ever completes service after time } t) = 1.$$

Remark 11. *Note that Theorem 4.1 is stronger than standard stability theorems, since it also implies that eventually no job ever leaves the system.*

Proof. For any $k \geq 1$, let T_k be the first time that there are k jobs in the queue and a failure occurs. T_k is almost surely finite since it is upper bounded by the time $\bar{T}_k$ that there are at least k arrivals in an open interval of size β just before a failure; note that $0 \leq A_0 < \infty$ a.s. The probability of this event is $\mathbb{P}(Z_\beta \geq k) > 0$.

Let $\mathcal{B} \triangleq \{B_1^{T_k}, \ldots, B_{Q_{T_k}}^{T_k}\}$ denote the job sizes that are present in the queue at time T_k.

From Proposition 4.1, we have

$$\mathbb{P}(\text{no job leaves after } T_k | Q_{T_k}, \mathcal{B}) \geq 1 - H(\mathbb{E}A\mathbf{1}(A \geq \beta k) + e^{-\theta k}) \geq 1 - \epsilon, \tag{4.3}$$

for all $k \geq k_0$, since $\theta > 0$ and $\mathbb{E}A\mathbf{1}(A \geq \beta k) \to 0$ as $k \to \infty$.

Now, for any fixed time t, we obtain

$$\begin{aligned}\mathbb{P}(\text{no job leaves after time } t) &\geq \mathbb{P}(T_k \leq t, \text{no job leaves after } T_k) \\ &= \mathbb{E}[\mathbb{P}(T_k \leq t | Q_{T_k}, \mathcal{B})\mathbb{P}(\text{no job leaves after } T_k | Q_{T_k}, \mathcal{B})] \\ &\geq \mathbb{P}(T_k \leq t)(1-\epsilon),\end{aligned}$$

which follows from (4.3); the equality follows from the fact that the event {no job leaves after T_k} is independent of the past, e.g., $T_k \leq t$, given $Q_{T_k}, \mathcal{B}$. Next, recall that T_k is almost surely finite, i.e. $\lim_{t\to\infty} \mathbb{P}(T_k \leq t) = 1$, and thus taking the limit as $t \to \infty$ yields

$$\underline{\lim}_{t\to\infty} \mathbb{P}(\text{no job leaves after time } t) \geq 1 - \epsilon.$$

Last, letting $\epsilon \downarrow 0$ finishes the proof. □

Corollary 4.1. *Under the conditions in Theorem 4.1, we have as $t \uparrow \infty$,*

$$Q_t \uparrow \infty \quad a.s.$$

Proof. Note that the number of arrivals $Z_t \uparrow \infty$ as $t \uparrow \infty$ a.s. Thus, without loss of generality, we can assume that $Z_t(\omega) \uparrow \infty$ as $t \uparrow \infty$ for every ω (by excluding the set of zero probability). Then, for any $v > 0$,

$$U_v \triangleq \{\text{no job ever completes service after time } v\} \subset \{Q_t \uparrow \infty \text{ as } t \uparrow \infty\}.$$

Now, if $\omega \in U_v$, then for $t \geq v, Q_t(\omega)$ is non-decreasing. Furthermore, since there are no

departures, the rate of increase of Q_t is equal to the arrival rate, and thus $Q_t \uparrow \infty$. Hence,

$$\mathbb{P}(Q_t \uparrow \infty \text{ as } t \uparrow \infty) \geq \mathbb{P}(\text{no job ever completes service after time } v)$$

which, by Theorem 4.1, implies

$$\mathbb{P}(Q_t \uparrow \infty \text{ as } t \uparrow \infty) = \lim_{v \to \infty} \mathbb{P}(\text{no job ever completes service after time } v) = 1.$$

□

Finally, we show instability, in general, without the condition $\mathbb{P}[B \geq \beta] = 1$. However, the conclusion is slightly weaker than in Theorem 4.1, and is the same as in Corollary 4.1. Basically, one cannot guarantee that no job ever completes service, since jobs can be arbitrarily small.

Theorem 4.2. *In the M/G/1 PS queue, if* $\mathbb{E}A < \infty$ *and* $0 \leq A_0 < \infty$ *a.s., we have as* $t \uparrow \infty$*,*

$$Q_t \uparrow \infty \quad \textit{a.s.}$$

Proof. First, by assumption, we can pick $\beta > 0$ such that $\mathbb{P}[B \geq \beta] > 0$. Then, for any time t, let Q_t^β be the number of jobs whose size is at least β and q_t^β be the number of jobs that are smaller than β. Hence,

$$Q_t = Q_t^\beta + q_t^\beta \geq \underline{Q}_t^\beta,$$

where $\underline{Q}_t^\beta$ is the queue in a system with the same arrival process where only jobs of size $B \geq \beta$ are served and the smaller ones are discarded. By Corollary 4.1, $\underline{Q}_t^\beta \uparrow \infty$ a.s., and, therefore, we obtain $Q_t \uparrow \infty$ a.s. □

Extension to DPS

In modern system design, PS cannot capture the heterogeneity of users and services, which is associated with unequal sharing of resources. Hence, we discuss the DPS queue which is a multi-class generalization of the PS queue: all jobs are served simultaneously at rates that are determined by a set of weights $w_i, i = 1, \ldots, K$. If there are n_j jobs in class j, each class-k job receives service at a rate $c_k = w_k / \sum_{j=1}^{K} w_j n_j$.

DPS has a broad range of applications. In computing, it is used to model Weighted-Round-Robin (WRR) scheduling. In communication networks, DPS is used for modeling heterogenous, e.g., with different round trip delays, TCP connections. Despite the fact that the PS queue is well understood, the analysis of DPS has proven to be very hard; yet, our previous results on PS are easily extended to DPS in the corollary below.

Corollary 4.2. *Under the conditions in Theorems 4.1 and 4.2, the discriminatory processor sharing (DPS) queue is also always unstable, with the same conclusion as in Theorems 4.1 and 4.2, respectively.*

Proof. Without loss of generality, assume that the set of weights is ordered such that $w_1 \leq w_2 \cdots \leq w_K$. In the M/G/1 DPS queue, the service allocation at any given time t for a single customer in class k is given by

$$c_k(t) = \frac{w_k}{\sum_{i=1}^{K} w_i n_i(t)} \leq \frac{w_k}{w_1 \sum_{i=1}^{K} n_i(t)} \leq \frac{w_K}{w_1 Q_t}.$$

Note that $c(t) = w_K/(w_1 Q_t)$ is the service rate in a PS queue with capacity $c = w_K/w_1 \geq 1$. Therefore, each class-k job, $k = 1 \ldots K$, in the DPS queue is served at a lower rate than the rate c of the PS queue. Hence,

$$Q_t^{DPS} \geq Q_t^{PS(c)},$$

and since, under the conditions in Theorem 4.1, the PS queue is always unstable, it follows that the DPS queue is also unstable. □

4.3.2 Stability of One Job at a Time Non-Preemptive Policy

In this section, we study the stability of service disciplines where jobs are processed one at a time in a non-preemptive fashion, e.g., FCFS. The stability results will be derived for exponentially distributed availability period A with rate μ. This assumption is needed to ensure the memoryless property of the system after each job completion.

Under such policies, the expected service time for a single job from Definition 4.2.2 is given by

$$\mathbb{E}[S] = \mathbb{E}\left[\sum_{i=1}^{N-1} A_i + B\right].$$

Note that $N \triangleq \inf\{n \geq 1 : A_n > B\}$ is a well defined stopping time for the process $(A, \{A_n\}_{n\geq 1})$, and thus the expected service time follows from Wald's identity as

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}\left[\sum_{i=1}^{N} A_i - A_N + B\right] \\ &= \mathbb{E}[N]\mathbb{E}[A] - \mathbb{E}[A_N] + \mathbb{E}[B].\end{aligned}$$

Now, assuming that the availability period A is exponentially distributed with rate μ (Poisson failures), the expected service time is given by

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}[N]\mathbb{E}[A] - (\mathbb{E}[A] + \mathbb{E}[B]) + \mathbb{E}[B] \\ &= (\mathbb{E}[N] - 1)\mathbb{E}[A], \end{aligned} \tag{4.4}$$

since $\mathbb{E}[A_N] = \mathbb{E}\left[\mathbb{E}[A|A > B]\right] = \mathbb{E}[A + B] = \mathbb{E}[A] + \mathbb{E}[B]$, due to the memoryless property of the exponential distribution.

The necessary and sufficient condition for the stability of the non-preemptive M/G/1

queue with failures is

$$\lambda \mathbb{E}[S] < 1.$$

Next, we derive an explicit formula for $\mathbb{E}[N]$ by observing that

$$\mathbb{P}[N > n|B] = \mathbb{P}(A \leq B|B)^n = G(B)^n.$$

Thus, using the exponential distribution of A, the expected number of restarts is

$$\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N|B]] = \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbb{P}[N > n|B]\right] = \mathbb{E}\left[\sum_{n=0}^{\infty} G(B)^n\right] = \mathbb{E}\left[\bar{G}(B)^{-1}\right] = \mathbb{E}[e^{\mu B}].$$

Hence, plugging the preceding expression in (4.4), we obtain

$$\mathbb{E}[S] = (\mathbb{E}[e^{\mu B}] - 1)\mu^{-1},$$

which yields the following theorem.

Theorem 4.3. *If $\{A_n\}_{n\geq 0}$ is Poisson with rate μ, arrivals are Poisson with rate $\lambda > 0$, and B is a typical job size, then the queue, for any non-preemptive policy that serves one job at a time, e.g., FCFS, is stable iff*

$$\lambda \mathbb{E}[S] = \lambda \mu^{-1}(\mathbb{E}[e^{\mu B}] - 1) < 1. \tag{4.5}$$

Note that, for exponential job sizes, the mean service time is finite and equal to $1/(1/\mathbb{E}[B] - \mu)$ if and only if $\mathbb{E}[B] < 1/\mu$, and the stability region is given by $\lambda/(1/\mathbb{E}[B] - \mu) < 1$. On the other hand, if B does not have exponential moments, then $\mathbb{E}[S] = \infty$, i.e. any non-preemptive policy will be unstable. Furthermore, the stability region for the system with failures is strictly smaller than in the traditional M/G/1 queue, since $\mathbb{E}[S] > \mu^{-1}\mu\mathbb{E}[B] = \mathbb{E}[B]$. In addition, since $e^x - 1 - x$ is increasing in x for $x > 0$, the stability region

shrinks as the jobs grow in size. Alternatively, as the job sizes are decreasing, e.g., applying fragmentation/checkpointing techniques, the stability region of a system with failures can approach the one of the traditional M/G/1 queue. Specifically, if $B = \beta$ is deterministic, $\lambda\mu^{-1}(e^{\mu\beta} - 1) \sim \lambda\beta$ as $\beta \to 0$, where $\lambda\beta < 1$ is the stability region of the ordinary M/G/1 queue without failures.

Remark 12. *Note that the preceding result can be derived alternatively by noticing that for deterministic job sizes, $B = \beta$, the service time S behaves exactly the same as a busy period in an M/D/∞ queueing system with arrival rate μ and service time B, which yields $\mathbb{E}[S] = (e^{\mu\beta}-1)/\mu$. This line of argument extends to random job sizes B, as in Theorem 4.3.*

4.4 GI/G/1 PS Queue with Restarts

In the previous section, we show that PS is unstable assuming Poisson arrivals. Here, we show that this result can be extended to more general arrival distributions, e.g., renewal processes. However, to avoid technical complications we assume that the failure process is Poisson or rate μ, i.e. the availability periods A_i are exponential. To this end, we use $M_{(t_0,t_1]}$ to denote the number of Poisson failures in $(t_0, t_1]$ and write M_t for intervals of the form $(0, t]$. Let $(\tau, \{\tau_n\}_{n\geq 1})$ be an i.i.d. sequence, where τ_n represent the interarrival times of the renewal process. Similarly as in the definition of the general failure process in Section 4.2, we assume that the first arrival occurs at time $\tau_0 \geq 0$. When τ_0 has the residual distribution of τ_1, then $\{\tau_n\}_{n\geq 0}$ will be in stationarity.

The main purpose of this section is to show that there is nothing special about the Poisson arrival assumption that leads to instability. Instead, the instability results from the interplay between sharing and retransmission/restart mechanisms. First, we prove the following proposition using similar arguments as in Proposition 4.1. However, we embed the proof at the points of arrivals instead of failures. In the following proposition, we assume that the first arrival occurs at $t = 0$, i.e. $\tau_0 = 0$. The remaining results allow for an arbitrary delay until the first arrival, $0 \leq \tau_0 < \infty$; these results imply the stationary version

of $\{\tau_n\}_{n\geq 0}$, when τ_0 has the excess distribution of τ_1.

Proposition 4.2. *Assume that a new job arrives at time $t = 0$, i.e. $\tau_0 = 0$, and there are $Q_0 \geq k$ jobs in the GI/G/1 PS queue with remaining service $\geq \beta$. If failures are Poisson, $\mathbb{E}A < \infty$, $\mathbb{E}\tau^{1+\delta} < \infty, 0 < \delta < 1$ and $\mathbb{P}[B \geq \beta] = 1, \beta > 0$, then for all $k \geq 1$*

$$\mathbb{P}[\textit{no job ever completes service}] \geq 1 - O(\mathbb{E}A\mathbf{1}(A \geq \beta k) + k^{-\delta}).$$

Proof. Let $T_1 = \sum_{i=1}^{k} \tau_i$ be the cumulative time that includes the first k arrivals for $t > 0$ and M_{T_1} be the number of failures in $(0, T_1)$. Now, define the event $\mathcal{A}_1 \equiv \mathcal{A}_1(k) \triangleq \{A_1 < \beta k, A_2 < \beta k, \ldots, A_{M_{T_1}} < \beta k\}$. On this event, no job can leave the system since $Q_0 \geq k$ and all of them are at least of size β. Thus, if they were served in isolation, they could not have completed service in the first M_{T_1} attempts.

Now, with a small abuse of notation, let E_1 denote the event that there is no departure in the first M_{T_1} attempts and there are at most ck failures in $(0, T_1]$. Formally,

$$E_1 \supset \underline{E}_1 \triangleq \{M_{T_1} \leq ck, \mathcal{A}_1\},$$

on the set $\{Q_0 \geq k\}$. Now, observe that

$$\begin{aligned}
\mathbb{P}(\underline{E}_1) &= \mathbb{P}(M_{T_1} \leq ck, A_1 < \beta k, A_2 < \beta k, \ldots, A_{M_{T_1}} < \beta k) \\
&\geq \mathbb{P}(M_{T_1} \leq ck, A_1 < \beta k, A_2 < \beta k, \ldots, A_{ck} < \beta k) \\
&\geq \mathbb{P}(A_1 < \beta k)^{ck} - \mathbb{P}(M_{T_1} > ck).
\end{aligned}$$

Next, note that

$$\begin{aligned}
\mathbb{P}(M_{T_1} > ck) &= \mathbb{P}\left(M_{T_1} > ck, T_1 \leq \frac{3k\mathbb{E}\tau}{2}\right) + \mathbb{P}\left(M_{T_1} > ck, T_1 > \frac{3k\mathbb{E}\tau}{2}\right) \\
&\leq \mathbb{P}\left(M_{\frac{3k\mathbb{E}\tau}{2}} > ck\right) + \mathbb{P}\left(T_1 > \frac{3k\mathbb{E}\tau}{2}\right),
\end{aligned}$$

where the first term is negligible for $c > 2\mu\mathbb{E}\tau$ since the expected number of failures is $3k\mu\mathbb{E}\tau/2$. Now, observe that

$$\mathbb{P}(T_1 > \frac{3k\mathbb{E}\tau}{2}) = \mathbb{P}\left(\sum_{i=1}^{k}\tau_i > \frac{3k\mathbb{E}\tau}{2}\right) = \mathbb{P}\left(\sum_{i=1}^{k}(\tau_i - \mathbb{E}\tau) > \frac{3k\mathbb{E}\tau}{2} - k\mathbb{E}\tau\right).$$

Now, let $X_i \triangleq \tau_i - \mathbb{E}\tau$, and by choosing $h = 2^{-\delta}(\mathbb{E}\tau)^{1+\delta}$ and $y = \mathbb{E}\tau/4$ in Lemma 1 of [9], we obtain

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{k}X_i > k\mathbb{E}\tau/2\right) &\leq k\mathbb{P}(X_1 > k\mathbb{E}\tau/4) + \frac{hk}{2^{-\delta}(k\mathbb{E}\tau)^{1+\delta}} \\ &\leq k\mathbb{P}(\tau_1 > k\mathbb{E}\tau/4 + \mathbb{E}\tau) + \frac{1}{k^{\delta}} \\ &\leq k\frac{\mathbb{E}\tau^{1+\delta}}{(k\mathbb{E}\tau/4 + \mathbb{E}\tau)^{1+\delta}} + k^{-\delta} \leq 2k^{-\delta}.\end{aligned}$$

Therefore,

$$\mathbb{P}(\underline{E}_1) \geq (1 - \mathbb{P}(A \geq \beta k))^{ck} - 2k^{-\delta},$$

where using $1 - x \geq e^{-2x}$ for small x, we have for all $k \geq k_0$

$$\begin{aligned}\mathbb{P}(\underline{E}_1) \geq e^{-2ck\mathbb{P}(A\geq\beta k)} - 2k^{-\delta} &\geq 1 - 2ck\mathbb{P}(A \geq \beta k) - 2k^{-\delta} \\ &\geq e^{-4ck\mathbb{P}(A\geq\beta k) - 4k^{-\delta}}.\end{aligned}$$

Next, at time $\mathcal{T}_1 = T_1$, on event $\underline{E}_1$, the queue has at least $2k$ jobs, i.e., $Q_{\mathcal{T}_1} \geq 2k$, and no jobs have departed. Similarly as before, let $T_2 = \sum_{i=k}^{3k}\tau_i$ be the cumulative time that includes the next $2k$ arrivals, and define $\mathcal{A}_2 \equiv \mathcal{A}_2(k) = \{A_{M_{T_1}+1} < 2\beta k, A_{M_{T_1}+2} < 2\beta k, \ldots, A_{M_{T_1+T_2}} < 2\beta k\}$. The probability that no job departs in $(0, \mathcal{T}_2]$, where $\mathcal{T}_2 = T_1 + T_2$,

is lower bounded by

$$\begin{aligned}\mathbb{P}(\text{no job departs in } (0,\mathcal{T}_2]) &\geq \mathbb{P}(M_{T_1} \leq ck, \mathcal{A}_1, Q_{\mathcal{T}_1} \geq 2k, M_{(\mathcal{T}_1,\mathcal{T}_2]} \leq 2ck, \mathcal{A}_2) \\ &\geq \mathbb{P}(M_{T_1} \leq ck, \mathcal{A}_1, M_{(\mathcal{T}_1,\mathcal{T}_2]} \leq 2ck, \mathcal{A}_2), \end{aligned} \tag{4.6}$$

since $\{Q_{\mathcal{T}_1} \geq 2k\} \supseteq \{M_{T_1} \leq ck, \mathcal{A}_1\}$ on the set $\{Q_0 \geq k\}$; to avoid repetitions, the following statements are all on $Q_0 \geq k$.

Now, if E_2 is the event that there is no departure in the next M_{T_2} attempts and there are at most $2ck$ failures in $(\mathcal{T}_1, \mathcal{T}_2]$, then $E_2 \supset \underline{E}_2 \triangleq \{M_{T_2} \leq 2ck, \mathcal{A}_2\}$; note that $\underline{E}_2$ is independent of $\underline{E}_1$ due to Poisson memoryless property. Via identical arguments as before, we obtain

$$\begin{aligned}\mathbb{P}(\underline{E}_2) &\geq \mathbb{P}(M_{T_2} \leq 2ck, A_{ck+1} < \beta k, \ldots, A_{3ck} < \beta k) \\ &\geq e^{-8ck\mathbb{P}(A \geq 2\beta k) - 4(2k)^{-\delta}}.\end{aligned}$$

Therefore, at time $\mathcal{T}_2$, on event $\underline{E}_1 \cap \underline{E}_2$, there are at least $4k$ jobs.

In general, for any n, we can extend the reasoning from (4.6) to obtain

$$\begin{aligned}\mathbb{P}(\text{no job departs in } (0,\mathcal{T}_n]) &\geq \mathbb{P}(M_{T_1} \leq ck, \mathcal{A}_1, M_{T_2} \leq 2ck, \mathcal{A}_2, \ldots, M_{T_n} \leq 2^{n-1}k, \mathcal{A}_n) \\ &= \mathbb{P}(\underline{E}_1 \cap \underline{E}_2 \cap \cdots \cap \underline{E}_n),\end{aligned}$$

where $\mathcal{T}_n = \sum_{i=1}^{n} T_i, T_n = \sum_{i=(2^{n-1}-1)k+1}^{(2^n-1)k} \tau_i$, E_n denotes the event that there is no departure in M_{T_n} attempts and there are at most 2^{n-1} failures in $(\mathcal{T}_{n-1}, \mathcal{T}_n]$, and $\underline{E}_n = \{M_{T_n} \leq 2^{n-1}ck, \mathcal{A}_n\}$. Similarly,

$$\mathbb{P}(\underline{E}_n) \geq e^{-2^{n+1}ck\mathbb{P}(A \geq 2^{n-1}\beta k) - 4(2^{n-1}k)^{-\delta}}.$$

Hence, we obtain

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) &\geq \prod_{i=1}^{n} e^{-2^{i+1}ck\mathbb{P}(A\geq 2^{i-1}\beta k)-4(2^{i-1}k)^{-\delta}} \\ &= e^{-4\sum_{i=0}^{n-1} 2^i ck\mathbb{P}(A\geq 2^i\beta k)-4k^{-\delta}\sum_{i=0}^{n-1}(2^i)^{-\delta}} \\ &\geq e^{-4\sum_{i=0}^{\infty} 2^i ck\mathbb{P}(A\geq 2^i\beta k)-4k^{-\delta}\sum_{i=0}^{\infty} 2^{-\delta i}}.\end{aligned}$$

Now, observe that $\sum_{i=0}^{\infty} 2^{-\delta i} < \infty$, and thus we can pick $H > 0$ such that

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) \geq e^{-4\sum_{i=0}^{\infty} 2^i ck\mathbb{P}(A\geq 2^i\beta k)-Hk^{-\delta}}.$$

The remainder of the proof follows identical arguments as Proposition 4.1. Thus, on $\{Q_0 \geq k\}$,

$$\mathbb{P}(\text{no job ever completes service}) \geq 1 - H(\mathbb{E}A\mathbf{1}(A \geq \beta k) + k^{-\delta}).$$

□

Theorem 4.4. *In the GI/G/1 PS queue, if failures are Poisson,* $0 \leq \tau_0 < \infty$ *a.s.,* $\mathbb{E}\tau^{1+\delta} < \infty, 0 < \delta < 1$ *and* $\mathbb{P}[B \geq \beta] = 1, \beta > 0$*, then*

$$\lim_{t\to\infty} \mathbb{P}(\textit{no job ever completes service after time } t) = 1.$$

Proof. Similarly to the proof of Theorem 4.1, we observe the system at time V_k when there are k jobs in the queue and a failure occurs. Since the arrivals are non Poisson, we need additional reasoning to ensure that $V_k < \infty$ a.s. In this regard, let us consider a time interval $T_1 = \sum_{i=1}^{k} \tau_i$ when the first k arrivals occur. Then, let t_k be such that $\mathbb{P}(T_1 < t_k) > 0$ and divide t_k into smaller intervals of size β. Now, consider the probability that $\{T_1 < t_k\}$ and there is at least one failure in each of the small intervals of size β. Since the failures are Poisson, this event has a positive, albeit extremely small, probability. If this event occurs,

then $V_k \leq T_1 < \infty$ a.s. Otherwise, repeat the procedure on the next interval $T_2 = \sum_{i=k+1}^{2k} \tau_i$. Since the arrivals are renewal and failures are Poisson, the desired event in interval T_2 is independent and has the same probability as in T_1. Hence, after a geometric number of attempts, the queue will have at least k jobs at the time of failure, implying that $V_k < \infty$ a.s.

Now, the remainder of the proof follows the same arguments as in Theorem 4.1 of Section 4.3. We omit the details. □ □

Similarly to Theorem 4.2 of Section 4.3, we drop the condition $\mathbb{P}[B \geq \beta] = 1$ and prove general instability.

Theorem 4.5. *In the GI/G/1 PS queue, if failures are Poisson,* $0 \leq \tau_0 < \infty$ *a.s., and* $\mathbb{E}\tau^{1+\delta} < \infty, 0 < \delta < 1$*, we have as* $t \uparrow \infty$*,*

$$Q_t \uparrow \infty \quad a.s.$$

The proof is similar to the proof of Theorem 4.2 and thus is omitted. Furthermore, the equivalent results could be stated for the DPS scheduler as well. Last, the preceding findings could be further extended to both non Poisson arrivals and non Poisson failures. However, the proofs would be much more involved and complicated; here, we avoid such technicalities.

4.5 Transient Behavior - Scheduling a Finite Number of Jobs

In the previous sections, we focus on the steady state behavior of the M/G/1 queue with restarts and prove that PS is always unstable for failure distributions with finite first moment. We also show instability for the GI/G/1 PS queue, assuming Poisson failures. In this section, in order to gain further insight into this system, we study its transient behavior. In this regard, we consider a queue with a finite number of jobs and no future arrivals and compute the total time until all jobs are completed. In Subsections 4.5.1 and 4.5.2, we

analyze the system performance when the jobs are served one at a time and when Processor Sharing (PS) is used, respectively. More precisely, for a finite number of jobs with sizes $B_i, 1 \leq i \leq m$, and assuming no future arrivals, we study the completion time Θ_m, until all m jobs complete their service. Throughout this section, we assume that service starts at $t = 0$ and $A_0 \equiv 0$; furthermore, we assume that the distribution functions $\bar{G}(x)$ and $\bar{F}(x)$ are absolutely continuous for all $x \geq 0$.

Note that in the case of traditional work conserving scheduling systems the completion time does not depend on the scheduling discipline and is always simply equal to $\sum_{i=1}^{m} B_i$. However, in channels with failures there can be a stark difference in the total completion time depending on the scheduling policy. This difference can be so large that in some systems the expected completion time can be infinite while in others finite, or even having many high moments.

Overall, we discover that, with respect to the distribution of the total completion time Θ_m, serving one job at a time exhibits uniformly better performance than PS; see Theorems 4.7 and 4.8. Furthermore, when the cumulative hazard functions of the job and failure distributions are proportional, i.e. $\log \bar{F}(x) \sim \alpha \log \bar{G}(x)$, we show that PS performs distinctly worse for the light-tailed job/failure distributions as opposed to the heavy-tailed ones, see parts (i) and (ii) of Theorem 4.8.

Before presenting our main results, we state the following theorem on the logarithmic asymptotics of the time $\bar{S} = \sum_{i=1}^{N} A_i = S + (A_N - B)$, where S is from Definition 4.2.2. Note that $\bar{S}$ includes the remaining time $(A_N - B)$ until the next channel availability period, thus representing a natural upper bound for S. In the following, let $\vee \equiv \max$.

Theorem 4.6. *If* $\log \bar{F}(x) \sim \alpha \log \bar{G}(x)$ *as* $x \to \infty$, $\alpha > 1$, $\mathbb{E}[B^{\alpha+\delta}] < \infty$, *and* $\mathbb{E}[A^{1 \vee \alpha}] < \infty$ *for some* $\delta > 0$, *then*

$$\lim_{t \to \infty} \frac{\log \mathbb{P}[\bar{S} > t]}{\log t} = -\alpha. \tag{4.7}$$

Proof. By Theorem 6 in [9], when specialized to the conditions of this theorem, we obtain that $\log \mathbb{P}[S > t] \to -\alpha \log t$ as $t \to \infty$. This immediately yields the lower bound for

$\bar{S} = S + (A_N - B) \geq S$. For the upper bound, $\bar{S} = S + (A_N - B)$ and the union bound result in

$$\mathbb{P}[\bar{S} > 2x] \leq \mathbb{P}[S > x] + \mathbb{P}[A_N - B > x].$$

Hence, in view of Theorem 6 in [9], we only need to bound $\mathbb{P}[A_N - B > x]$. To this end, observe that

$$\begin{aligned}
\mathbb{P}[A_N - B > x] = \mathbb{P}[A_N > B + x] &= \sum_{i=1}^{\infty} \mathbb{P}[A_i > B + x, N = i] \\
&= \sum_{i=1}^{\infty} \mathbb{P}[A_i > B + x, A_1 < B, \ldots, A_{i-1} < B] \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left[\mathbb{P}\left(A_i > B + x | B\right) \mathbb{P}\left(A_1 < B | B\right)^{i-1}\right] \\
&= \mathbb{E}\left[\frac{\bar{G}(B + x)}{\bar{G}(B)}\right] \leq \bar{G}(x)\mathbb{E}[N],
\end{aligned}$$

since $\mathbb{E}[N] = \mathbb{E}(1/\bar{G}(B))$. Now, the condition $\alpha > 1$ guarantees that $\mathbb{E}[N] < \infty$ whereas $\mathbb{E}[A^{\alpha}] < \infty$ implies that $\bar{G}(x) = O(1/x^{\alpha})$. Thus, (4.7) is satisfied. □ □

4.5.1 One Job at a Time Non-Preemptive Policy

In this subsection, we consider the failure-prone system that was introduced in Section 4.2, with unit capacity. The jobs are served one at a time, e.g., FCFS. Herein, we analyze the performance of this system assuming that, initially, there are m jobs in the queue and there are no future arrivals. Specifically, we study the total completion time, which is defined below.

Definition 4.5.1. *The total completion time is defined as the total time until all the jobs are successfully completed and is denoted as*

$$\Theta_m \triangleq \sum_{i=1}^{m} S_i,$$

where m is the total number of jobs in the system and S_i is the service requirement for each

job.

In the following theorem, we prove that the tail asymptotics of the total completion time, from Definition 4.5.1, under this policy is a power law of the same index as the service time of a single job.

Theorem 4.7. *If* $\log \bar{F}(x) \sim \alpha \log \bar{G}(x)$ *as* $x \to \infty$, $\alpha > 1$, $A_0 = 0, \mathbb{E}[B^{\alpha+\delta}] < \infty$, *and* $\mathbb{E}[A^{1\vee\alpha}] < \infty$ *for some* $\delta > 0$, *then*

$$\lim_{t\to\infty} \frac{\log \mathbb{P}[\Theta_m > t]}{\log t} = -\alpha.$$

Proof. Recall that the service requirement for a job B_i was previously defined as $S_i = \sum_{j=1}^{N_i-1} A_j + B_i$.

For the *lower* bound, we observe that

$$\mathbb{P}[\Theta_m > t] \geq \mathbb{P}[S_1 > t],$$

since the total completion time is at least equal to the service time of a single job. By taking the logarithm and using Theorem 6 in [9], we have

$$\frac{\log \mathbb{P}[\Theta_m > t]}{\log t} \geq -(1+\epsilon)\alpha. \tag{4.8}$$

For the *upper* bound, we compare Θ_m with the completion time in a system where the server is kept idle between the completion time of the previous job and the next failure. Clearly,

$$\Theta_m \leq \bar{\Theta}_m \triangleq \sum_{i=1}^{m} \bar{S}_i,$$

where $\bar{S}_i \triangleq \sum_{j=1}^{N_i} A_j$ are the service times that include the remaining availability period A_{N_i}.

Then, we argue that

$$\mathbb{P}[\Theta_m > t] \leq \mathbb{P}\left[\sum_{i=1}^{m} \bar{S}_i > t\right] \leq m\mathbb{P}\left[\bar{S}_1 > \frac{t}{m}\right],$$

which follows from the union bound. By taking the logarithm and using Theorem 4.6, we have

$$\frac{\log \mathbb{P}[\Theta_m > t]}{\log t} \leq -\alpha(1-\epsilon) + \frac{\log m}{\log t} \leq -(1-2\epsilon)\alpha, \tag{4.9}$$

where we pick t large enough such that $\log t \geq \log m/(\alpha\epsilon)$.

Letting $\epsilon \to 0$ in both (4.8) and (4.9) finishes the proof. □ □

4.5.2 Processor Sharing Discipline

In this subsection, we analyze the Processor Sharing discipline where m jobs share the (unit) capacity of a single server. We present our main theorem on the logarithmic scale, which shows that the tail asymptotics of the total completion time is determined by the shortest job in the queue. In particular, under our main assumptions, this time is a power law, but it exhibits a different exponent depending on the job size distribution, as our results demonstrate; see Theorem 4.8 and the proof.

- If the jobs are subexponential (heavy-tailed) or exponential, the total delay is simply determined by the time required for any single job to complete its service, as if it were the only one present in the queue.

- If the jobs are superexponential (light-tailed), the total delay is determined by the service time of the *shortest* job. This job generates the heaviest asymptotics among all the rest.

Our main result, stated in Theorem 4.8 below, shows that on the logarithmic scale the distribution of the total completion time Θ_m^{PS} is heavier by a factor $m^{\gamma-1}$ for superexponential

jobs relative to the subexponential or exponential case, when the cumulative hazard functions F and G are proportional. Therefore, in systems with failures and restarts, sharing the capacity among light-tailed jobs induces long delays, whereas, for heavy-tailed ones, PS appears to perform as good as serving the jobs one at a time. Interestingly enough, this deterioration in performance is determined by the time it takes to serve the shortest job in the system.

Note that in a PS queue with no future arrivals, the shortest job will depart first. Immediately after this, the server will continue serving the remaining $m-1$ jobs, and, similarly, the shortest job, i.e. the second shortest among the original m jobs, will depart before all the others. This pattern will continue until the departure of the largest job, which is served alone.

Theorem 4.8. *Assume that the cumulative hazard function $-\log \bar{F}(x)$ is regularly varying with index $\gamma \geq 0$. If $\log \bar{F}(x) \sim \alpha \log \bar{G}(x)$ as $x \to \infty$, $\alpha > 1$, $A_0 = 0, \mathbb{E}[B^{\alpha+\delta}] < \infty$, and $\mathbb{E}[A^{1\vee\alpha}] < \infty$ for some $\delta > 0$, then*

1. *if $\gamma \leq 1$, i.e. B is subexponential or exponential, then*

$$\lim_{t\to\infty} \frac{-\log \mathbb{P}[\Theta_m^{PS} > t]}{\log t} = \alpha,$$

2. *if $\gamma > 1$, i.e. B is superexponential, then*

$$\lim_{t\to\infty} \frac{-\log \mathbb{P}[\Theta_m^{PS} > t]}{\log t} = \frac{\alpha}{m^{\gamma-1}} < \alpha.$$

Remark 13. *When $\alpha > 1$, we easily verify that $\mathbb{E}[\Theta_m^{PS}] < \infty$ in case (i); if the jobs are superexponential, e.g., case (ii), then $\mathbb{E}[\Theta_m^{PS}] = \infty$ if $\alpha < m^{\gamma-1}$.*

Proof. Let $B^{(1)} \leq B^{(2)} \leq \cdots \leq B^{(m)}$ be the order statistics of the jobs $B_1, B_2, \ldots, B_m$.

The assumption that $-\log \bar{F}(x)$ is regularly varying with index γ implies that

$$\log \bar{F}(\lambda x) \sim \lambda^{\gamma} \log \bar{F}(x), \tag{4.10}$$

for any $\lambda > 0$.

We begin with the *lower* bound.

(i) Subexponential or exponential jobs ($\gamma \leq 1$).

The total completion time is lower bounded by the time required for a single job to depart when it is exclusively served, e.g., if the total capacity of the system is used. Hence, it follows that

$$\mathbb{P}[\Theta_m^{PS} > t] \geq \mathbb{P}[S_1 > t], \tag{4.11}$$

where S_1 is the service time of a single job of random size B_1, when there are no other jobs in the system. Now, recalling Theorem 6 in [9], it holds that

$$\lim_{t\to\infty} \frac{\log \mathbb{P}[S_1 > t]}{\log t} = -\alpha.$$

By taking the logarithm in (5.5), the lower bound follows immediately.

(ii) Superexponential jobs ($\gamma > 1$).

The total completion time is lower bounded by the delay experienced by the shortest job, and hence,

$$\mathbb{P}[\Theta_m^{PS} > t] \geq \mathbb{P}[S_1^{PS} > t],$$

where S_1^{PS} is the service time of job $B^{(1)}$. First, note that the distribution of $B^{(1)}$ is given by

$$\begin{aligned}\mathbb{P}(B^{(1)} > x) &= \mathbb{P}(B_1 > x, B_2 > x, \ldots, B_m > x) \\ &= \mathbb{P}(B_1 > x)\mathbb{P}(B_2 > x)\cdots\mathbb{P}(B_m > x) \\ &= \mathbb{P}(B_1 > x)^m = \bar{F}(x)^m, \end{aligned} \tag{4.12}$$

since $B_i, i = 1, \ldots, m$, are independent and identically distributed. Now, using (4.12) and

(4.10), together with our main assumption, we observe that

$$\log \mathbb{P}(mB^{(1)} > x) = m \log \bar{F}\left(\frac{x}{m}\right) \sim m^{1-\gamma} \log \bar{F}(x) \sim \alpha m^{1-\gamma} \log \bar{G}(x);$$

note that we compute the distribution of $mB^{(1)}$ since $B^{(1)}$ receives $1/m$ fraction of the service. Then, Theorem 6 in [9] applies with $\alpha/m^{\gamma-1} \leq \alpha$, i.e.

$$\lim_{t\to\infty} \frac{\log \mathbb{P}[S_1^{PS} > t]}{\log t} = -\frac{\alpha}{m^{\gamma-1}}.$$

Next, we derive the *upper* bound. To this end, we consider a system where the server is kept idle after the completion of each job until the next failure occurs. At this time, all the remaining jobs are served under PS until the next shortest one departs. If there are more than one jobs of the same size, only one of these departs. Under this policy, it clearly holds that

$$\Theta_m^{PS} \leq \sum_{i=1}^{m} \bar{S}_i^{PS},$$

where $\bar{S}_i^{PS}$ corresponds to the service time of the i^{th} smallest job and includes the time until the next failure.

Using the union bound, we obtain

$$\mathbb{P}[\Theta_m^{PS} > t] \leq \mathbb{P}\left[\sum_{i=1}^{m} \bar{S}_i^{PS} > t\right] \leq (1+\epsilon) \sum_{i=1}^{m} \mathbb{P}\left(\bar{S}_i^{PS} > \frac{t}{m}\right). \tag{4.13}$$

It is easy to see that the service time of the i^{th} smallest job $B^{(i)}$ depends on the number of jobs that share the server, i.e. $m - i + 1$, since $m - i$ jobs have remained in the queue.

Now, the distribution of the i^{th} shortest job is derived as

$$\begin{aligned}\mathbb{P}(B^{(i)} > x) &= \sum_{k=0}^{i-1} \binom{m}{k} \mathbb{P}(B_1 \le x)^k \mathbb{P}(B_1 > x)^{m-k} \\ &\sim \binom{m}{i-1} \mathbb{P}(B_1 > x)^{m-i+1} \sim \bar{F}(x)^{m-i+1}. \end{aligned} \tag{4.14}$$

Next, starting from (4.14), it easily follows that

$$\begin{aligned}\log \mathbb{P}\left((m-i+1)B^{(i)} > x\right) &\sim \log \bar{F}\left(\frac{x}{m-i+1}\right)^{m-i+1} \\ &\sim (m-i+1)^{1-\gamma} \log \bar{F}(x) \\ &\sim \alpha(m-i+1)^{1-\gamma} \log \bar{G}(x),\end{aligned}$$

where we use (4.10) and our main assumption and define $\alpha_i \triangleq \alpha/(m-i+1)^{\gamma-1}$; here, we compute the distribution of $(m-i+1)B^{(i)}$ since the $B^{(i)}$ job receives $1/(m-i+1)$ fraction of the service.

Now, recalling Theorem 4.6, we have

$$\frac{\log \mathbb{P}[\bar{S}_i^{PS} > t]}{\log t} \to \alpha_i \text{ as } t \to \infty,$$

and thus (4.13) yields

$$\frac{\log \mathbb{P}[\Theta_m^{PS} > t]}{\log t} \le -(1-\epsilon) \min_{i=1\ldots m} \alpha_i,$$

for all $t \ge t_0$.

(i) Subexponential or exponential jobs ($\gamma \le 1$).

Observe that $\min_{i=1\ldots m} \alpha_i = \alpha$, and thus

$$\frac{\log \mathbb{P}[\Theta_m^{PS} > t]}{\log t} \le -(1-\epsilon)\alpha. \tag{4.15}$$

(ii) Superexponential jobs ($\gamma > 1$).

In this case, $\min_{i=1...m} \alpha_i = \alpha/m^{\gamma-1}$, and thus

$$\frac{\log \mathbb{P}[\Theta_m^{PS} > t]}{\log t} \leq -(1-\epsilon)\frac{\alpha}{m^{\gamma-1}}. \tag{4.16}$$

Letting $\epsilon \to 0$ in (4.15) and (4.16), we obtain the upper bound. □ □

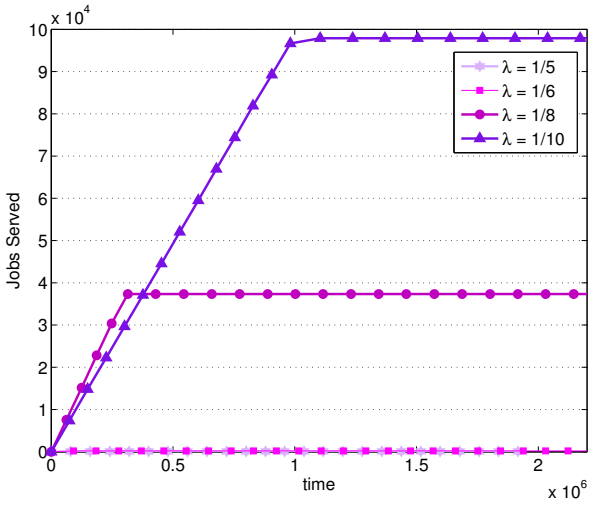


Figure 4.3: Example 1. Jobs completed over time.

4.6 Simulation

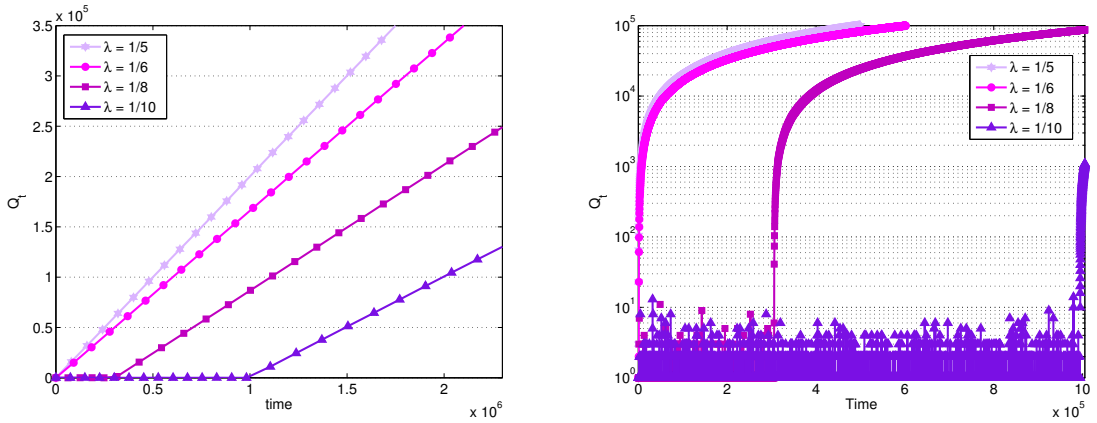
In this section, we present our simulation experiments in order to demonstrate our theoretical findings. All the experiments result from $N = 10^8$ (or more) samples of each simulated scenario; this guarantees the existence of at least 100 occurrences in the lightest end of the tail that is presented in the figures. First, we illustrate the instability results from Sections 4.3 and 4.4.

Example 1. *M/G/1 PS is unstable.* In this example, we show that the PS queue becomes unstable by simulating the M/G/1 PS queue for different arrival rates $\lambda > 0$, which all satisfy the stability condition for the non-preemptive M/G/1 queue, when jobs are served one at a time. In this regard, we assume constant job size $\beta = 1$ and Poisson

failures of rate $\mu = 1/20$. Therefore, by evaluating (4.5), we obtain

$$\lambda \mathbb{E}[S] = \lambda \mu^{-1}(e^{\mu} - 1) = 20(e^{0.05} - 1)\lambda = 1.025\lambda < 1,$$

or equivalently the stability region for the non-preemptive queue is given by $\Lambda = \{\lambda \leq 0 : \lambda < 0.9752\}$. Hence, in this example, we use λ from the preceding stability region, $\lambda \in \Lambda$.

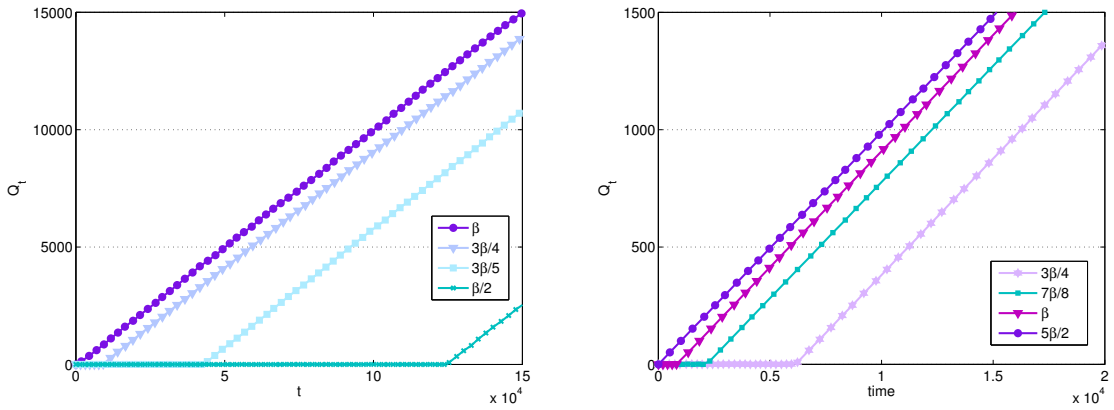


(a) Queue size over time (b) Queue size for small t

Figure 4.4: Example 1. Queue size evolution. Subfigure (b) zooms in the time range $[0, 10^6]$ of Fig. 4.4; Q_t (y-axis) is shown on the logarithmic scale.

In Fig. 4.3, we plot the number of jobs that have received service up to time t. We observe that the cumulative number of served jobs always converges to a fixed number and does not increase any further. This happens after some critical time when the queue starts to grow continuously and is unable to drain. For larger values of λ, the system saturates faster meaning that the cumulative throughput at the saturated state is lower.

Furthermore, we observe from the simulation that the system behaves as if it were stable until some critical time or queue size after which it is unable to drain. From Fig. 4.3, we can see that the case $\lambda = 10^{-1}$ saturates at time $t = 10^6$ and the total number of served jobs reaches 10^5. Hence, the departure rate until saturation time is $10^5/10^6 = 10^{-1}$, which is exactly equal to the arrival rate $\lambda = 10^{-1}$, corresponding to the departure rate of a stable queue. This further emphasizes the importance of studying the stability of these systems since, at first glance, they may appear stable.



(a) Example 1 (b) Example 2

Figure 4.5: (a) Example 1. Queue size over time parameterized by fragment length; $\beta = 2, \lambda = 0.1$. (b) Example 2. Queue size over time parameterized by job size; $\beta = 4$.

Fig. 4.4 demonstrates the queue size evolution over time. Similarly as in Fig. 4.3, we observe that for any arrival rate λ, there is a critical time after which the queue continues to grow and never empties. This time varies depending on the simulation experiment; yet, on average, we observe that the queue remains stable for longer time when λ is smaller. Now, we zoom in on the queue evolution on the logarithmic scale in Fig. 4.4(b). Again, we observe that the queue looks stable until some critical time/queue size.

Last, in Fig. 4.6, we plot the queue evolution for different job sizes, namely $\beta = 1, 1.2, 1.5$ and 2. We observe that larger job fragments cause instability much faster than the smaller units. For example, $\beta = 2$ leads to instability almost immediately, while $\beta = 1.5$ renders the queue unstable after 10^4 time units. Similarly, reducing the fragment size by 60% delays the process by an additional 3×10^4 units. Last, cutting the jobs in half causes instability after approximately 13×10^4 time units. This implies that one should apply fragmentation with caution in order to select the appropriate fragment size that will maintain good system performance for the longest time.

Example 2. *General arrivals.* In this example, we consider non Poisson arrivals. We assume that the failure distribution is exponential with mean $\mathbb{E}A = 10$ and that jobs interarrival times follow the Pareto distribution with $\alpha = 2$ and mean $\mathbb{E}\tau = 10.1$. Similarly

as in the previous example, Fig. 4.6 shows the queue evolution with time for different job sizes β.

Next, we validate the results on the transient analysis from Section 4.5.

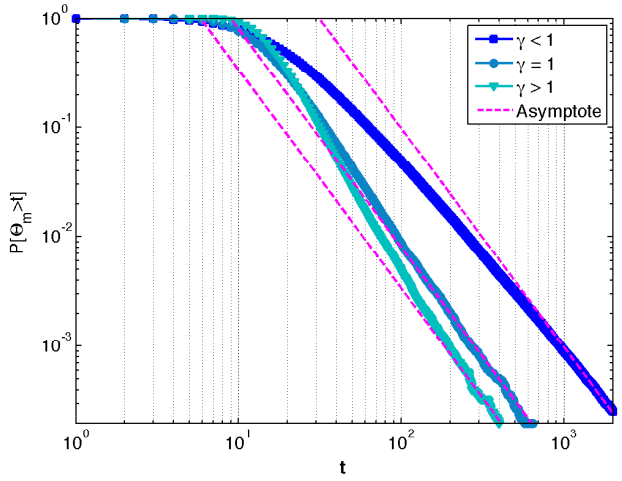


Figure 4.6: Example 3. Non-preemptive policy: Logarithmic asymptotics when $\alpha = 2$ for exponential, superexponential ($\gamma > 1$) and subexponential ($\gamma < 1$). distributions.

Example 3. *Non-preemptive policy: Always the same index* α. In this example, we consider a queue of $m = 10$ jobs, which are served First Come First Serve (FCFS), i.e. one at a time. The logarithmic asymptotics from Theorem 4.7 implies that the tail is always a power law of index $\alpha = 2$.

In Fig. 4.6, we plot the distribution of the total completion time in a queue with 10 jobs that are processed one at a time. On the same graph, we plot the logarithmic asymptotics (dotted lines) that correspond to a power law of index $\alpha = 2$. We consider the following three scenarios:

1. Weibull distributions with $\gamma = 2$. The failures A are distributed according to $\bar{G}(x) = e^{-(x/\mu)^2}$ with mean $\mathbb{E}[A] = \mu\Gamma(1.5) = 1.5$, and jobs B also follow Weibull distributions with $\bar{F}(x) = e^{-(x/\lambda)^2}, \lambda = \mu/\sqrt{2}$. In this case, it is easy to check that the main assumption of Theorem 4.7 is satisfied, i.e.

$$\log \bar{F}(x) = -(x/\lambda)^2 = \alpha \log \bar{G}(x), \quad \alpha = (\mu/\lambda)^2.$$

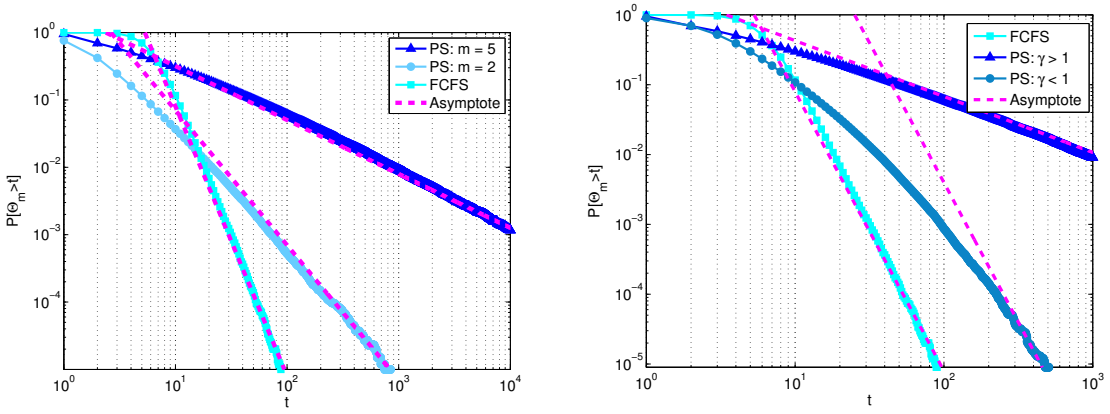
2. Exponential distributions. Failures are exponential with $\mathbb{E}[A] = 2$, $\bar{G}(x) = e^{-x/2}$, and the jobs B are also exponential of unit mean, i.e. $\bar{F}(x) = e^{-x}$. Then, trivially,

$$\log \bar{F}(x) = 2 \log \bar{G}(x).$$

3. Weibull distributions with $\gamma = 0.5$. A's are Weibull with $\bar{G}(x) = e^{-\sqrt{x}/2}$, i.e., $\mathbb{E}[A] = 8$. Also, we assume Weibull jobs B with $\bar{F}(x) = e^{-\sqrt{x}}$. Thus,

$$\log \bar{F}(x) = -\sqrt{x} = 2 \log \bar{G}(x).$$

In all three cases, we obtain $\alpha = 2$. Yet, we observe that the tail asymptotics is the same regardless of the distribution of the job sizes. For the subexponential jobs (case 3: Weibull with $\gamma < 1$), the power law tail appears later compared to the case of superexponential jobs. This is because the constant factor of the exact asymptotics is different for each case, and it depends on the mean size of A, $\mathbb{E}[A]$.



(a) Example 4 (b) Example 5

Figure 4.7: (a) Example 4. Logarithmic asymptotics for different number of superexponential jobs when $\alpha = 4$ under PS and FCFS discipline. (b) Example 5. Logarithmic asymptotics under FCFS, PS with subexponential and superexponential jobs.

Example 4. *PS: The effect of the number of jobs.* In this example, we consider a PS queue with $m = 5$ and $m = 2$ superexponential jobs, and compare it against a FCFS queue

with $m = 5$ jobs. We assume superexponential job sizes B's and A's, namely Weibull with $\gamma = 2$; see case 1 of Example 3. Here α is taken equal to 4. The logarithmic asymptotics is given in Theorems 4.7 and 4.8.

In Fig. 4.6, we demonstrate the total completion time Θ_m^{PS}, for different number of jobs, when $\gamma = 2$. Theorem 4.8(ii) states that $\alpha(m) = \alpha/m^{\gamma-1}$ and, thus, for $\gamma = 2$ we have $\alpha(m) = \alpha/m$, e.g., we expect power law asymptotes with index α/m for the different values of m. On the same figure, we also plot the FCFS completion time Θ_m, which is always a power law of index $\alpha = 4$, as we previously observed in Example 3. It can be seen that PS generates heavier power laws, for superexponential jobs. In particular, PS with $m = 2$ results in power law asymptotics with $\alpha(2) = 2$, while PS with $m = 5$ jobs leads to infinite expected delay since $\alpha(5) = 4/5 < 1$.

Example 5. *PS: The effect of the distribution type.* In this example, for completeness, we evaluate the impact of the job distribution on the total completion time under both heavy and light-tailed job sizes. To this end, we consider the PS queue from Example 4, with $m = 5$ jobs, and compare it against FCFS. In Fig. 4.6, we re-plot the logarithmic asymptotics of the total completion time $\mathbb{P}(\Theta_m^{PS} > t)$, for different distribution types of the failures/jobs and index $\alpha = 4$, as before. In particular, we consider Weibull distributions as in Example 3 with $\gamma = 1/2 < 1$ and $\gamma = 2 > 1$ for the subexponential and superexponential cases, respectively.

On the same graph, we plot the distribution of the completion time Θ_m in FCFS, which is always a power law of the same index, as illustrated in Example 3. By fixing the number of jobs to be $m = 5$, Fig. 4.6 shows that when the jobs are superexponential, PS yields the heaviest asymptotics among all three scenarios; for subexponential jobs, PS generates asymptotics with the same power law index as in FCFS, albeit with a different constant factor.

Example 6. *Limited queue: Throughput vs. overhead tradeoff.* In practice, job and buffer sizes are bounded and therefore the queue may never become unstable. However, our results indicate that the queue may lock itself in a 'nearly unstable' state, where it is at

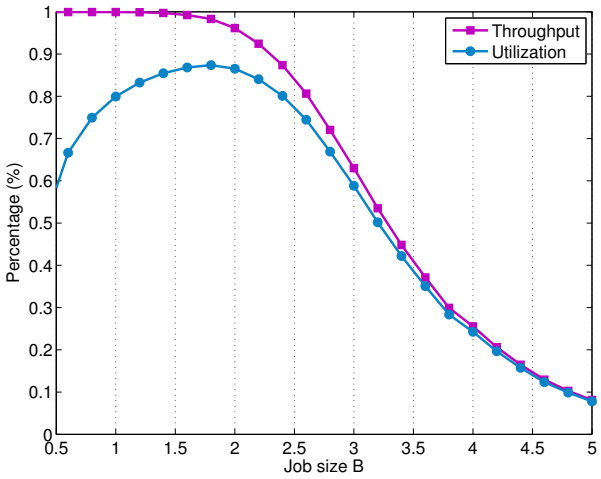


Figure 4.8: Example 6. Throughput vs. utilization tradeoff.

its maximum size and the throughput is very low. Here, we would like to emphasize that, unlike in the case of unlimited queue size, job fragmentation can be useful for increasing the throughput and the efficiency of the system. In this case, one has to be careful about the overhead cost of fragmentation. Basically, each fragment requires additional information, called the 'header' in the context of communications, which contains details on how it fits into the bigger job, e.g., destination/routing information in communication networks. Hence, if the fragments are too small, there will be a lot of overhead and waste of resources. In view of this fact, one would like to optimize the fragment sizes by striking a balance between throughput and utilization.

In this example, we demonstrate the tradeoff between throughput and generated overhead, assuming limited queue size q^*. If the newly arriving job does not fit in the queue, i.e. the number of jobs currently in the queue is equal to q^*, it is discarded. We define throughput as the percentage of the jobs that complete service among all jobs that arrive at the M/G/1 PS queue. It basically corresponds to the total work that is carried out in the system. On the other hand, we define utilization as the useful work that is served over the aggregated load in the system. Specifically, we consider jobs that require a minimum size b, where b represents the overhead, e.g., the packet header, thread id, etc. The remaining

job size, $\beta - b$, represents the useful information.

We consider different job sizes β from 0.4 up to 5 bytes, with overhead $b = 0.2$. We simulate the M/G/1 PS queue with maximum queue size $q^* = 10$ jobs for a fixed time $T = 10^8$ time units. The arrivals are Poisson with rate 1/10 and the failures are exponential of the same rate. Clearly, in the case of fixed job sizes β, throughput γ is lower bounded by the throughput of the system when it performs at the limit, i.e. when the queue is full. This state corresponds to the worst overall performance and can be easily computed. On average, for a fixed period of time T, q^* jobs will complete service every $\mathbb{E}[S_{q^*}]$ time units, while the total jobs that arrive in the system is λT. In this case, the lower bound for the throughput is given by

$$\underline{\gamma} = q^* \frac{T}{\mathbb{E}[S_{q^*}]} \frac{1}{\lambda T} = \frac{q^*}{\lambda \mathbb{E}[S_{q^*}]},$$

and in the particular case of exponential failures, using (4.5) we derive

$$\underline{\gamma} = \frac{q^*}{\lambda \mu^{-1}(e^{\mu q^* \beta} - 1)}.$$

Using this observation, throughput will be suboptimal when $\gamma < 1$. Thus, for job sizes larger than $\beta_* = \log(\mu q^* \lambda^{-1} + 1)/(\mu q^*)$, the throughput starts decreasing.

In Fig. 4.8, we observe that for small job sizes, the throughput is 100% and it deteriorates as the job size β increases. In particular, when the job size exceeds 1.5, the throughput drops exponentially. Utilization exhibits a different behavior; it is low when the job size is small, i.e. the useful job size is comparable to the overhead b, and reaches its peak at $\beta \approx 1.7$. After this, it starts decreasing following similar trend as the throughput. In this case, $\beta - b \approx 1.5$ appears to be the optimal size for the job fragments. This phenomenon of combining limited queue size with job fragmentation may require further investigation.

4.7 Concluding Remarks

Retransmissions/restarts represent a primary failure recovery mechanism in large-scale engineering systems, as it was argued in the introduction. In communication networks, retransmissions lie at the core of the network architecture, as they appear in all layers of the protocol stack. Similarly, PS/DPS based scheduling mechanisms, due to their inherent fairness, are commonly used in computing and communication systems.

However, our results show that PS/DPS scheduling in systems with retransmissions is always unstable. Furthermore, this instability cannot be resolved by job fragmentation techniques or checkpointing. On the contrary, serving one job at a time, e.g., FCFS, can be stable and its performance can be further enhanced with fragmentation. Interestingly, systems where jobs are served one at a time can highly benefit from fragmentation and, in fact, their performance can approach closely the corresponding system without failures.

Overall, using PS in combination with retransmissions in the presence of failures deteriorates the system performance and induces instability. In addition, our findings suggest that further examination of existing techniques is necessary in the failure-prone environment with retransmission/restart failure recovery and sharing, e.g., see Example 6.

Chapter 5

Future Work & Conclusion

This chapter provides new directions and insights based on the main results of the preceding chapters. First, we motivate future work in the area of networking as well as cloud computing services. To this end, we present our preliminary findings in Sections 5.1 and 5.2. Next, we conclude the thesis by summarizing the impact of our results on modern engineering design and discussing the extension of our analytical work, as well as the tools developed herein, to other research areas in Section 5.3.

5.1 Towards Stabilizing Sharing Systems

In this section, we present a thorough discussion of the limited queue size scenario that was presented in Chapter 4; see Example 6 of Section 4.6. To this end, we study the properties of the queue size evolution process and analyze the system throughput, which provides a typical performance measure.

We assume that the queue is limited in size by k and jobs are all of fixed size $\beta > 0$. If the newly arriving job does not fit in the queue, i.e. the number of jobs currently in the queue is equal to k, it is discarded. The analysis to follow will make use of more restrictive assumptions on the queue behavior. These assumptions are required in order to construct events that represent renewal points in the process. Hence, we observe the queue at the

time of one of the following events: (a) the queue empties or (b) a failure occurs. We study the dynamics of the queue size $\{Q_n\}_{n\geq 0}$, where Q_n is the queue size at the time of the n^{th} event. We further assume that only jobs that are present in the queue at the time of the n^{th} event, i.e., Q_n jobs, are allowed to receive service until the next event occurs. Last, we assume that an empty queue $Q_0 = 0$ transits immediately to state $Q_1 = 1$.

Under the preceding assumptions, we compute the queue size after the first event occurs in the following cases:

- if $Q_0 = 0$, then the system transits to $Q_1 = 1$ with probability one.
- if $Q_0 > 0$, then (a) if jobs are completed before the next failure occurs, i.e., $\beta Q_0 \leq A$, then the queue size is equal to the number of arrivals in the interval of size βQ_0, i.e., $Z_{\beta Q_0}$; (b) if a failure occurs before jobs are completed, i.e., $\beta Q_0 > A$, then the queue size equals the number of new arrivals in the interval A plus the initial workload Q_0, i.e., $Z_A + Q_0$.

In the *bounded* queue case, the queue size is the minimum of Q_1 and k; we use the notation $x \wedge y$ to denote the minimum of x and y. Also, if $Q_0 = k$, then the queue either remains the same, if a failure occurs before jobs are completed, or it empties; note that jobs depart all at once, since they are all of size β. Hence, we formulate the queue size evolution as follows.

(i) Unbounded Queue (Q)

$$Q_1 = \left[\mathbf{1}(\beta Q_0 \leq A) Z_{\beta Q_0} + \mathbf{1}(\beta Q_0 > A)(Q_0 + Z_A)\right] \mathbf{1}(Q_0 > 0) + \mathbf{1}(Q_0 = 0)$$

(ii) Bounded Queue ($Q \leq k$)

$$\begin{aligned} Q_1 = &\left[\mathbf{1}(\beta Q_0 \leq A) Z_{\beta Q_0} \wedge k + \mathbf{1}(\beta Q_0 > A)(Q_0 + Z_A) \wedge k\right] \mathbf{1}(0 < Q_0 < k) \\ &+ \mathbf{1}(Q_0 = 0) + (k\mathbf{1}(\beta k > A))\mathbf{1}(Q_0 = k) \end{aligned}$$

Next, we compute the transition probabilities for the special case when $k = 2$ in the bounded queue scenario. We assume Poisson arrivals with rate λ and exponential failures with rate ν. Due to the memoryless property of Poisson arrivals and failures, the system is Markovian and thus it is sufficient to compute the transition matrix, denoted as $\mathcal{P}$.

Proposition 5.1. *Consider the M/G/1 PS queue limited by 2 jobs. If arrivals and failures are Poisson with rate λ and ν, respectively, and jobs are all of size $\beta > 0$, then the transition matrix is given by*

$$\mathcal{P} = \begin{bmatrix} 1 & 0 & 0 \\ e^{-(\lambda+\nu)\beta} & \lambda\beta e^{-(\lambda+\nu)\beta} + \nu(1 - e^{-(\lambda+\nu)\beta})/(\lambda+\nu) & 1 - \mathcal{P}_{10} - \mathcal{P}_{11} \\ e^{-2\nu\beta} & 0 & 1 - e^{-2\nu\beta} \end{bmatrix}$$

Table 5.1: Transition matrix $\mathcal{P}$

Proof. We begin with the first row and compute the probability of transition from $Q_0 = 0$ to $Q_1 = i, i = 0 \dots 2$. By assumption, we have

$$\mathbb{P}(Q_1 = 1|Q_0 = 0) = 1, \quad \mathbb{P}(Q_1 = 0|Q_0 = 0) = \mathbb{P}(Q_1 = 2|Q_0 = 0) = 0.$$

We continue with the second row. Observe that the queue resets to zero if there is no failure before the job departs.

$$\mathbb{P}(Q_1 = 0|Q_0 = 1) = \mathbb{E}[\mathbf{1}(\beta \le A)\mathbf{1}(Z_\beta = 0)] = \mathbb{P}(A > \beta)\mathbb{P}(Z_\beta = 0) = e^{-\nu\beta}e^{-\lambda\beta} = e^{-(\lambda+\nu)\beta}$$

Note that the queue remains the same when either (a) the job is completed and a new one arrived in the meantime, or (b) there is a failure before the completion of this job.

$$\begin{aligned}\mathbb{P}(Q_1 = 1|Q_0 = 1) &= \mathbb{E}[\mathbf{1}(\beta \le A)\mathbf{1}(Z_\beta = 1)] + \mathbb{E}[\mathbf{1}(\beta > A)\mathbf{1}(Z_A = 0)] \\ &= e^{-\nu\beta}\mathbb{P}(Z_\beta = 1) + \mathbb{E}[e^{-\lambda A}\mathbf{1}(A < \beta)] = \lambda\beta e^{-(\lambda+\nu)\beta} + \nu(1 - e^{-(\lambda+\nu)\beta})/(\lambda+\nu)\end{aligned}$$

Last, the queue ends up at capacity when either (a) the job is completed before the next failure and more than two new jobs arrived, or (b) a failure occurred and one or more new jobs arrived.

$$\begin{aligned}\mathbb{P}(Q_1 = 2|Q_0 = 1) &= \mathbb{E}[\mathbf{1}(\beta \leq A)\mathbf{1}(Z_\beta \geq 2)] + \mathbb{E}[\mathbf{1}(\beta > A)\mathbf{1}(Z_A \geq 1)] \\ &= e^{-\nu\beta}\mathbb{P}(Z_\beta \geq 2) + \mathbb{E}[(1 - e^{-\lambda A})\mathbf{1}(A < \beta)] \\ &= e^{-\nu\beta}(1 - e^{-\lambda\beta} - \lambda\beta e^{-\lambda\beta}) + 1 - e^{-\nu\beta} - \nu(1 - e^{-(\lambda+\nu)\beta})/(\lambda + \nu) \\ &= \lambda/(\lambda + \nu) - \lambda e^{-(\lambda+\nu)\beta}/(\lambda + \nu) - \lambda\beta e^{-(\lambda+\nu)\beta}.\end{aligned}$$

Finally, we compute the probability of transition from $Q_0 = 2$ to $Q_1 = i, i = 0 \ldots 2$. In this case, the queue either returns to zero, or it remains at the same state, i.e., if a failure occurs before the next departure.

$$\mathbb{P}(Q_1 = 0|Q_0 = 2) = \mathbb{P}(2\beta \leq A) = e^{-2\nu\beta}$$

$$\mathbb{P}(Q_1 = 2|Q_0 = 2) = \mathbb{P}(A < 2\beta) = 1 - e^{-2\nu\beta}$$

$$\mathbb{P}(Q_1 = 1|Q_0 = 2) = 0.$$

By combining the preceding results, we obtain the transition matrix in Table 5.1. □

The transition probabilities can be similarly derived for any k; these derivations require tedious calculations and are out of the scope of this thesis.

Drift Analysis

In this subsection, we present additional results on the drift for the queue evolution $\{Q_n\}_{n\geq 0}$ process.

Proposition 5.2. *Assume that there are Q_0 jobs of fixed size $\beta > 0$ in the queue, arrivals are Poisson with rate λ and failures are exponential with rate μ. Then, the drift of the*

queue size, denoted as $d(Q_0)$, is given by

$$\mathbb{E}[Q_1|Q_0] - Q_0 = \lambda\mu^{-1} - e^{-\mu\beta Q_0}(Q_0 + \lambda\mu^{-1})$$

and achieves its minimum at $Q_{0*} = \frac{1}{\beta\mu}(1 - \lambda\beta)$.

Proof.

$$\begin{aligned} d(Q_0) := \mathbb{E}[Q_1|Q_0] - Q_0 &= \mathbb{E}[Z_{\beta Q_0}\mathbf{1}(A > \beta Q_0)] + \mathbb{E}[(Z_A + Q_0)\mathbf{1}(A \leq \beta Q_0)] - Q_0 \\ &= \lambda\beta Q_0\mathbb{P}(A > \beta Q_0) + \lambda\mathbb{E}[A\mathbf{1}(A \leq \beta Q_0)] - Q_0\mathbb{P}(A > \beta Q_0) \\ &= \lambda\beta Q_0 e^{-\mu\beta Q_0} + \lambda(\mu^{-1} - \beta Q_0 e^{-\mu\beta Q_0} - \mu^{-1}e^{-\mu\beta Q_0}) - Q_0 e^{-\mu\beta Q_0} \\ &= \lambda\mu^{-1} - e^{-\mu\beta Q_0}(Q_0 + \lambda\mu^{-1}). \end{aligned}$$

This function achieves its minimum at

$$\begin{aligned} \frac{d(d(Q_0))}{dQ_0} &= e^{-\mu\beta Q_0}\mu\beta(Q_0 + \lambda\mu^{-1}) - e^{-\mu\beta Q_0} = 0 \\ \frac{d^2(d(Q_0))}{dQ_0^2} &= -e^{-\mu\beta Q_0}(\lambda\mu^{-1} + 1) < 0. \\ \Rightarrow Q_{0*} &= \frac{1}{\beta\mu}(1 - \lambda\beta). \end{aligned}$$

□

From a design perspective, we can choose $k \leq k_{max}$ such that for given λ, β and μ, the queue size has negative drift. After some time, if the system is allowed to exceed k_{max}, it will always have positive drift and the queue will start growing to infinity.

5.1.1 Throughput

In the preceding section, we computed the transition matrix and the drift of the queue evolution process. Here, we analyze the throughput of the M/G/1 PS queue with limited capacity. This is defined as the percentage of the jobs that complete service among all jobs

that arrive at the queue. Informally, this corresponds to the total work that is carried out in the system. First, we prove the following proposition.

Proposition 5.3. *In the M/G/1 PS queue, where $Q \leq k$ and failures are exponential with rate μ, there exists λ_0 such that for all $\lambda \geq \lambda_0$,*

$$\gamma(k) \sim \mu\beta k e^{-\mu\beta k} \quad \text{as } k \to \infty. \tag{5.1}$$

Remark 14. *This result implies that the throughput decays exponentially at k with rate $\mu\beta$, and it achieves its maximum when $\mu\beta k = 1$.*

Proof. Note that when the queue always resets to zero, the system starts afresh, and thus the process is renewal. It is therefore sufficient to compute the average work between renewal points. The expected service time for Poisson failures and k jobs of size β is given (Theorem 4.3 of Chapter 4) by

$$\mathbb{E}[S_k] = \mu^{-1}(e^{\mu\beta k} - 1) \sim \mu^{-1}e^{\mu\beta k} \text{as } k \to \infty \tag{5.2}$$

Now, let T_k be the time to reach k jobs after the queue resets to 0. Then, the renewal cycle is equal to $T_k + S_k$, where S_k is the service requirement for k jobs. The workload that is served in a renewal cycle equals the workload that has arrived minus the work that is lost. The rate of lost work is the expected work that arrives when the queue is at capacity, i.e., $\lambda\beta\mathbb{E}[S_k]$ over the expected length of the renewal cycle $\mathbb{E}[T_k] + \mathbb{E}[S_k]$. Then, the throughput is computed as follows

$$\gamma(k) = \lambda\beta - \frac{\lambda\beta\mathbb{E}[S_k]}{\mathbb{E}[T_k] + \mathbb{E}[S_k]} = \lambda\beta\frac{\mathbb{E}[T_k]}{\mathbb{E}[T_k] + \mathbb{E}[S_k]}. \tag{5.3}$$

It can be proved that for λ large enough, $\mathbb{E}[T_k]$ is of the order $O(k)$, e.g., the queue grows linearly in k with rate equal to the rate of arrival. Thus, one can show that $\mathbb{E}[T_k] \sim k/\lambda$ as $k \to \infty$. Therefore, (5.2) yields

$$\gamma(k) \sim \lambda\beta \frac{k/\lambda}{k/\lambda + \mu^{-1}e^{\mu\beta k}} = \lambda\beta \frac{1}{1 + \frac{\mu^{-1}e^{\mu\beta k}}{k/\lambda}} \sim \mu\beta k e^{-\mu\beta k}. \tag{5.4}$$

□

Simulation Examples

In order to verify our analytical findings, we simulate the M/G/1 PS queue with limited capacity k. This setup allows us to evaluate the configuration of different parameters, e.g., arrival/failure rate, job size, etc. Similarly as in previous chapters, our simulation experiments result from at least 10^8 samples of the simulated scenarios.

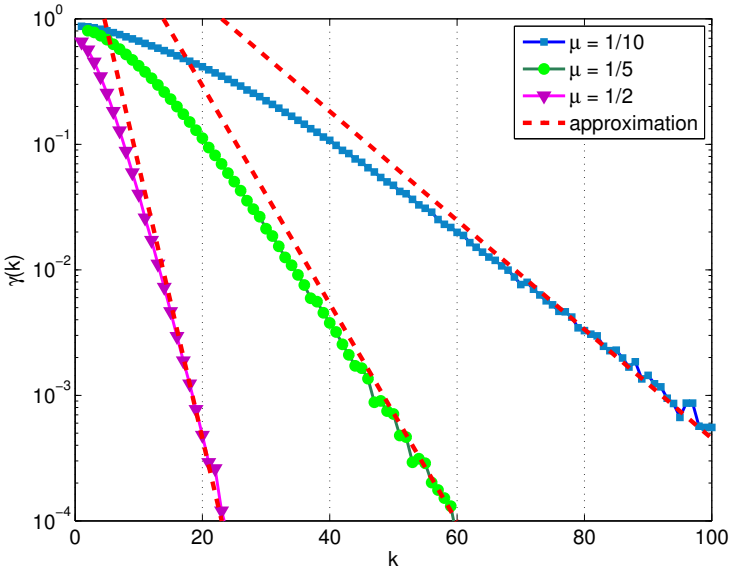


Figure 5.1: Example 1(a). Throughput $\gamma(k)$; the dotted line is the approximation from (5.4)

Example 1. *Throughput decay.* Our first example focuses on the tail behavior of the throughput as k increases. We simulate a PS queue with limited capacity k, arrival rate $\lambda = 10$, job size $\beta = 2$ and variable failure rates. As it was previously discussed, the system can be stabilized for $k\mu\beta = 1$, e.g., FCFS, but its performance will deteriorate for

larger k. In the infinite queue case, the time when the queue starts growing to infinity is not deterministic; yet, there exists a region where the system remains stable. Fig. 5.1 demonstrates the exponential decay of the throughput with respect to k for different failure rates μ.

As we can easily infer from Fig. 5.1, the throughput decays exponentially at different rates, which depend on μ. In particular, the tail is exponential with rate $\mu\beta = 2\mu$, i.e., the decay is faster for higher failure rates. The approximation from (5.4) fits the throughput tail even for smaller k, e.g., $k < 50$. In fact, when the system experiences frequent failures, sharing the service among more than one jobs is not always preferable since the probability of failure increases as the service requirement is prolonged. Under these conditions, serving one job at a time guarantees stability. On the other hand, if failures are not very frequent, then serving only one job at a time can be suboptimal, especially when the arrival rate is high. This is why it is crucial to understand the significance of carefully adjusting the system parameters in sharing systems.

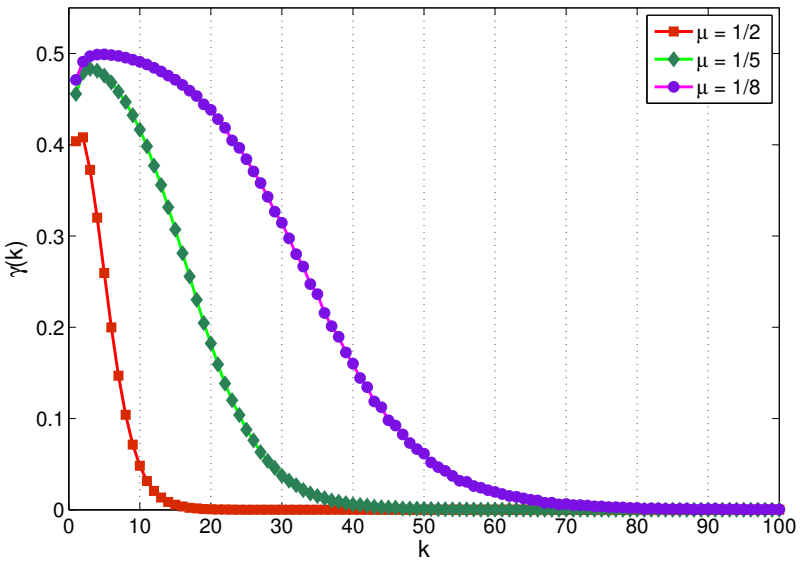


Figure 5.2: Example 1(b). Throughput $\gamma(k)$ for different failure rates μ.

Next, we simulate the PS queue assuming jobs of size $\beta = 1$ and arrival rate $\lambda = 1$. We

plot the throughput on the exact scale to clearly demonstrate where the maximum values are achieved. In Fig. 5.2, we observe that the throughput is not always optimal for $k = 1$ but instead it may reach a higher value as k increases. This results from the fact that more jobs are dropped while serving the existing one, especially when the queue has very low capacity. Hence, it is preferable to allow more than one jobs to share the service by increasing the queue capacity to $k_* = 1/(\mu\beta)$, or at least by picking k around this value. In Fig. 5.2, we observe that initially the throughput increases as k grows, then it remains high around $k_* = 1/(\mu\beta) = 1/\mu$ and eventually starts decaying exponentially to zero. The expected criticality point is attained around $k_* = 2, 5$ and 8 for $\mu = \{1/2, 1/5, 1/8\}$, respectively.

Example 2. *Queue Instability.* In our second example, we discuss the results from Section 5.1. To this end, we plot the drift function $d(Q)$ for different values of queue sizes. We set $\lambda = \mu = 1$ and $\beta = 1/10$, which yield $d(Q_0) = 1 - e^{-Q_0/10}(Q_0 + 1)$. Fig. 5.3 demonstrates the detrimental effects of allowing more than a few jobs, e.g., 40 or more, to share the service in this system.

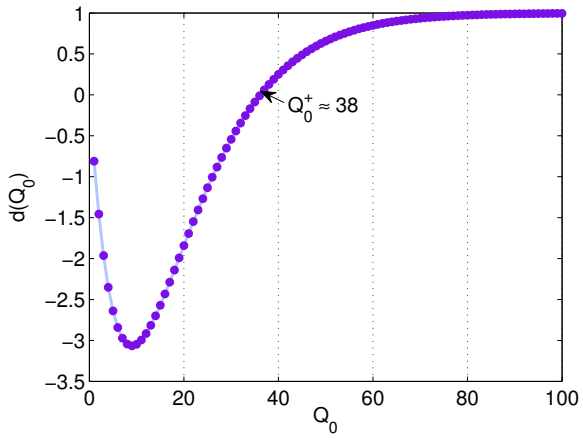


Figure 5.3: Example 2. Drift for different initial queue sizes Q_0.

Specifically, the function achieves its minimum around $Q_0 = 10$ and beyond this point, it is monotonically increasing. It is evident that the drift hits zero when $Q_0^+ = 38$, as marked on the Figure, and it remains positive for $Q_0 > 38$, eventually converging to 1. Beyond this point, the queue will become unstable since its size will continuously grow.

During the transient period before this happens, the system can jump to a state with more than Q_0^+ jobs and positive drift or it could remain in the seemingly "stable" state before it accumulates this critical number of jobs.

5.2 Reliability Tradeoffs in Cloud Computing

In this section, we study the tradeoffs between parallel processing and redundancy in large-scale distributed systems, with a special emphasis on cloud computing services. In modern engineering design, failure recovery is based on data replication, i.e., a large number of servers execute identical copies of the original job in parallel. Another widely accepted technique is to split the each job smaller ones and distribute the fragments to the available servers, e.g., each fragment is assigned to exactly one server. In both cases, the server needs to restart the job execution after a failure occurs. We evaluate the performance of these approaches under different assumptions on the failure and job distributions.

5.2.1 Fragmentation

Here, we study the asymptotics of the total number of restarts for a job of random size B which is split into k fragments. The total completion time of a job B is determined by the maximum delay of each of the fragments of size B/k. In this analysis, we show that under our main assumptions, the number of restarts behaves as a power law, but it exhibits a different index which depends on the distribution type.

We restate the following theorem that appears in [9] and describes the tail asymptotics of the number of restarts N for a single job; the notation is borrowed from previous chapters, e.g., see also Section 1.2.

Theorem 5.1. *Assume that* $\log \bar{F}(x) \sim \alpha \log \bar{G}(x)$ *for all* $x \geq 0, \alpha > 0$, *then*

$$\lim_{n\to\infty} \frac{-\log \mathbb{P}[N > n]}{\log n} = \alpha.$$

Remark 15. *Note that when* $\alpha < 1$, *then* $\mathbb{E}N = \infty$, *i.e.,* N *does not have any moments.*

The preceding theorem provides the logarithmic asymptotics for the number of restarts of a job of size B without fragmentation. In the following theorem, we assume that the initial job is divided into k fragments and derive the logarithmic asymptotics of the tail distribution for the total delay, denoted as $\mathcal{T}_k$.

Theorem 5.2. *Assume that the hazard function* $-\log \bar{F}(x)$ *is regularly varying with index* $\beta \geq 0$. *If* $\log \bar{F}(x) \sim \alpha \log \bar{G}(x), \mathbb{E}[A^{1+\theta}] < \infty$ *for all* $x \geq 0$, *then*

$$\lim_{t\to\infty} \frac{-\log \mathbb{P}(\mathcal{T}_k > t)}{\log t} = k^\beta \alpha.$$

Remark 16. *When* $\alpha > 1$, *we easily verify that* $\mathbb{E}[\mathcal{T}_k] < \infty$ *and fragmentation reduces the delay tail* k^β *times. Basically, if jobs are (super)exponential* ($\beta \geq 1$) *then we gain* $k^\beta \alpha - 1$ *extra moments. For subexponential jobs, fragmentation is not as beneficial and, in fact, as* $\beta \downarrow 0$, *the asymptotic behavior approaches the no-fragmentation case.*

Proof. We begin with the *lower* bound. The total delay for k fragments $\mathcal{T}_k$ is lower bounded by the service time of any single fragment. Hence, it follows that

$$\mathbb{P}[\mathcal{T}_k > t] \geq \mathbb{P}[T_1 > t], \tag{5.5}$$

where T_1 is the delay of one job fragment of size B/k. Now, observe that

$$\log \mathbb{P}(B/k > x) = k^\beta \log \bar{F}(x) \sim \alpha k^\beta \log \bar{G}(x),$$

which follows from our main assumption and the fact that $-\log \bar{F}(x)$ is regularly varying with index β, i.e., $\log \bar{F}(\lambda x) \sim \lambda^\beta \log \bar{F}(x)$, for any $\lambda > 0$. Then, Theorem 6 in [9] applies with $k^\beta \alpha$ and thus

$$\lim_{t\to\infty} \frac{-\log \mathbb{P}[T_1 > t]}{\log t} = k^\beta \alpha.$$

By taking the logarithm in (5.5), the lower bound follows immediately.

Next, we derive the *upper* bound. Here, by the union bound we obtain

$$\mathbb{P}[\mathcal{T}_k > t] \leq k\mathbb{P}[T_1 > t],$$

which yields the desired logarithmic asymptotics. □

5.2.2 Replication

In this section, we study the effects of data replication in large distribution centers under different assumptions on the failure/job statistics. In particular, we are interested in computing the tail asymptotics of the delay when jobs are replicated, i.e., k independent copies of the jobs are assigned to k different servers.

First, we prove the following Theorem which demonstrates the tail insensitivity of the delay to the number of copies k.

Theorem 5.3. *Assume that* $\log \bar{F}(x) \sim \alpha \log \bar{G}(x)$ *as* $x \to \infty$, *and* $\mathbb{E}[A^{1+\theta}] < \infty$, *then*

$$\lim_{t\to\infty} \frac{-\log \mathbb{P}(\mathcal{T}_k > t)}{\log t} = \alpha.$$

Sketch of the proof. Note that the delay is defined as the minimum delay induced by each of the k copies. Therefore,

$$\mathbb{P}[\mathcal{T}_k > t|B] = \mathbb{P}[\min_{i=1\ldots k} T_i > t|B] = \mathbb{P}[T_1 > t|B]^k,$$

since the k servers are independent conditionally on the job size B. Note also that the k servers are identical, i.e., they have the same failure distribution.

Next, taking the expectation with respect to B and the logarithm in the preceding expression yields

$$\lim_{t\to\infty} \frac{-\log \mathbb{P}[\mathcal{T}_k > t]}{\log t} = \alpha, \tag{5.6}$$

which finishes the proof. □

In the following Proposition, we derive the exact asymptotics for the number of restarts of a job of size B, denoted as $\mathcal{N}_k$. The proof follows similar arguments as the preceding Theorem and thus is omitted.

Proposition 5.4. *If* $\log \bar{F}(x) = \alpha \log \bar{G}(x)$ *for all* $x \geq 0$, *then as* $n \to \infty$

$$\mathbb{P}(\mathcal{N}_k > n) \sim \frac{\Gamma(\alpha+1)}{k^\alpha n^\alpha}.$$

The preceding results indicate that the delay does not depend on whether the distributions of failures/jobs are light or heavy-tailed. Specifically, replication leads to an improvement of $\alpha - 1$ extra moments for the tail of the delay distribution, albeit with a different constant factor which depends on the number of replicas.

5.2.3 Simulation Examples

In this subsection, we present a set of simulation examples that validate the preceding findings.

Example 1. *Light-tailed distributions.* In this example, we simulate a light-tailed scenario with the following assumptions. The failures are Weibull distributed, i.e., $\bar{G}(x) = e^{-(x/\mu)^\beta}$ whereas the jobs follow the same distribution albeit with a different scale parameter, i.e., $\bar{F}(x) = e^{-(x/\lambda)^\beta}$. Note that, under these conditions, our main assumption holds since

$$\log \bar{F}(x) = -x^\beta/\lambda^\beta = (\mu/\lambda)^\beta \log \bar{G}(x) = \alpha \log \bar{G}(x).$$

First, we run simulations for $\beta = 2$ and $\alpha = (20/19)^\beta = 1.108$ for different number of fragments/replicas k. From Fig. 5.4, we conclude that fragmentation has significant impact on the power law tail of $\mathbb{P}(\mathcal{T}_k > t)$ and $\mathbb{P}(\mathcal{N}_k > n)$ as k increases. Also, we can easily verify that the index of the power law tail in the case of replication remains the same regardless of the number of replicas k. However, we observe that both distributions move to the left as k increases. For the case of replication, this improvement comes from the smaller

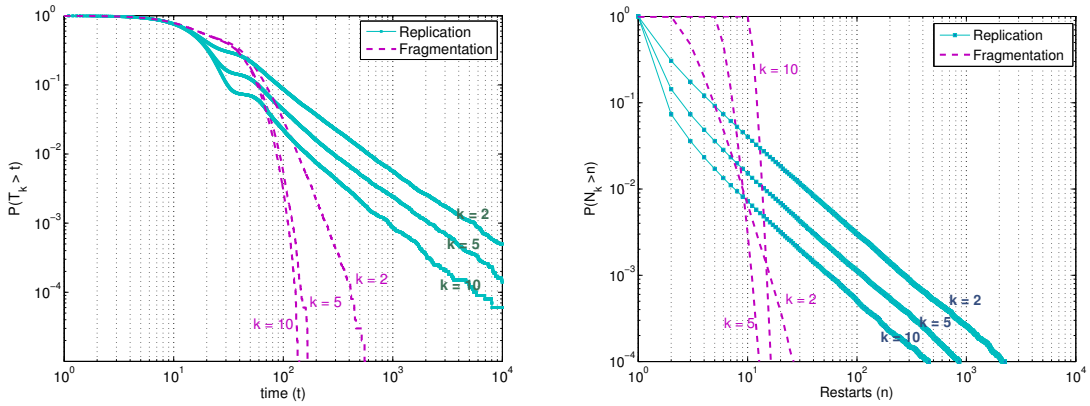


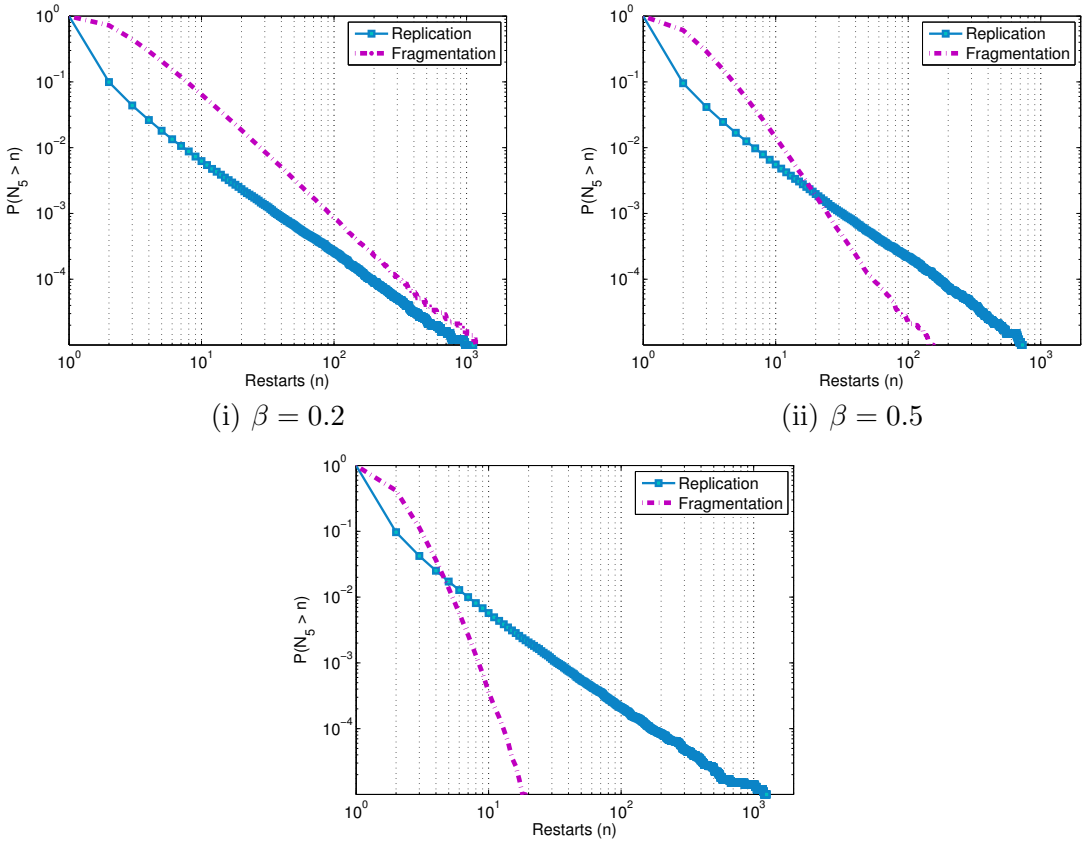
Figure 5.4: Example 1. $\mathbb{P}(\mathcal{T}_k > t)$ and $\mathbb{P}(\mathcal{N}_k > n)$ for different number of fragments/replicas k.

constant factor $\Gamma(\alpha+1)/k^{\alpha}$. On the other hand, when fragmentation is applied, the tail index becomes $k^2\alpha$ and the slope changes.

Example 2. *Heavy vs. Light-tailed distributions.* In this example, we consider the family of Weibull-type distributions of Example 1 and evaluate the impact of the shape parameter β on the tail behavior of the delay both for fragmentation and replication. In particular, we compare the performance of these techniques for light-tailed ($\beta \geq 1$) and heavy-tailed distributions ($\beta < 1$). To this end, we fix the number of fragments/replicas to $k = 5$ and set $\alpha \approx 1.4$ as follows: (i) $\beta = 0.2$: $\lambda = 1, \mu = 5$, (ii) $\beta = 0.5$: $\lambda = 1, \mu = 2$ and (iii) $\beta = 1$ (exponential case): $\lambda = 1, \mu = 1.4$.

We plot the distribution of N_5 for fragmentation/replication and different β parameters. As shown in Fig. 5.5, the performance of fragmentation deteriorates as β decreases. In the special case when $\beta \to 0$, fragmentation becomes as good as replication, as shown in Theorem 5.2. When β is close to zero, we observe that fragmentation is always worse than replication. This is due to the fact that the power law tail is approximately the same but the constant term is different for replication; see Proposition 5.4. In addition, we make the following observations:

- In the case of replication, the tail in unaffected by changes in β. In the case of fragmentation, larger values for β guarantee lighter power law tails, and, specifically



(i) $\beta = 0.2$ (ii) $\beta = 0.5$

(iii) $\beta = 1$

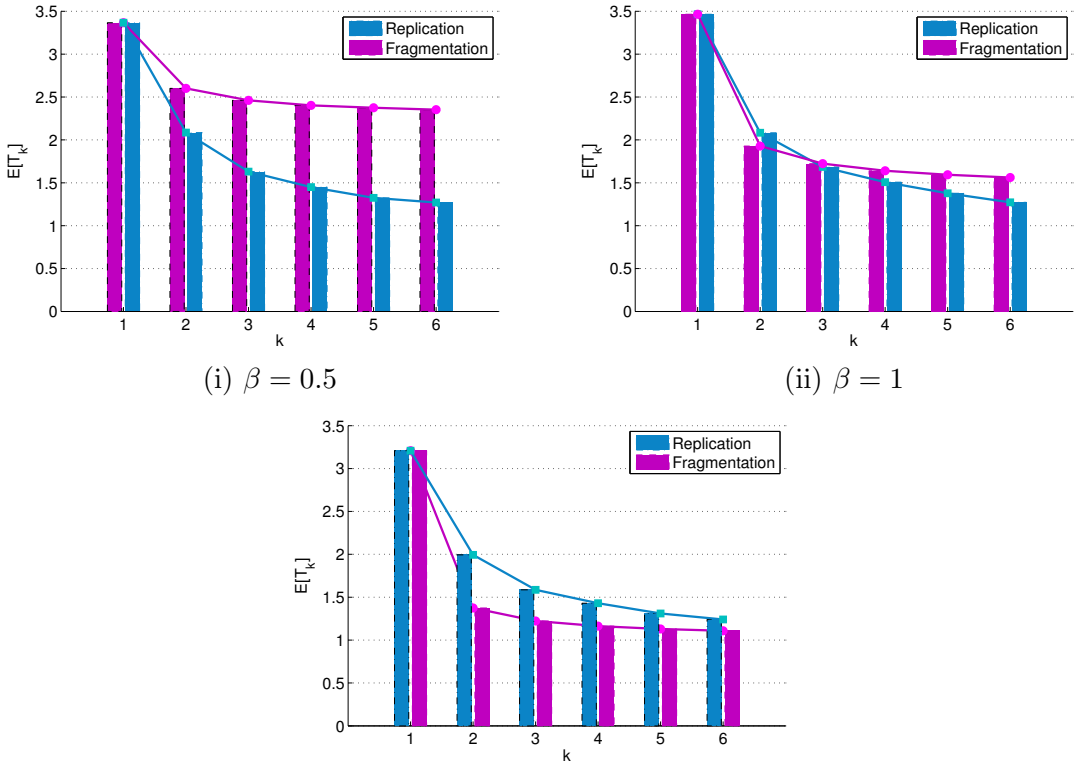
Figure 5.5: Example 2. $\mathbb{P}(\mathcal{N}_5 > n)$ for different distribution types.

in this example, we obtain $k = 5^{0.8} \approx 3.6$ times better tail index as β increases from 0.2 to 1, which is verified by the change of the slope.

- For β large enough, there exists a region when replication outperforms fragmentation but eventually fragmentation becomes more efficient. Initially, the delay distribution with replication lies below the main body of $\mathbb{P}(\mathcal{N}_k > n)$ for fragmentation but, as β increases, the break-even point moves to the left, implying that lighter distributions will benefit from fragmentation faster. Eventually, fragmentation outperforms replication since it guarantees a lighter tail.

In general, there is no superior technique that guarantees optimality across the entire delay distribution. Therefore, the situation must be treated with caution so that discrep-

ancies in performance are prevented. Once there is a good understanding of the behavior of each technique in a given range of probabilities, one could choose between the two in order to obtain the best possible results.



(i) $\beta = 0.5$ (ii) $\beta = 1$

(iii) $\beta = 2$

Figure 5.6: Example 3. $\mathbb{E}[\mathcal{T}_k]$ for different types of distributions.

Example 3. *Mean Value Analysis.* In our last example, we study the expected delay for fragmentation and replication and compare the values across different distribution types. To this end, we compute the mean delay for $\alpha = 1.44$ and various β values. Specifically, we consider heavy-tailed ($\beta = 0.5$), exponential ($\beta = 1$) and light-tailed distributions ($\beta = 2$). Fig. 5.6 demonstrates the mean delay for different number of fragments/replicas. When $\beta = 0.5$, replication is always faster than fragmentation, despite the heavier tail asymptotics; see Example 2(ii). For the exponential case ($\beta = 1$), the average delay is the same for small k, e.g., $k = 1, 2$ but as k increases, replication outperforms fragmentation. Last, for larger values of β, replication demonstrates longer average delays for small k, but eventually, it

achieves the same performance as fragmentation. Overall, from a mean value perspective, as k increases, replication is at least as good as fragmentation.

5.2.4 Replication Or Fragmentation?

In the preceding results, we discover a dichotomy between replication and fragmentation based on the distribution of failures and job sizes, as well as the number of fragments/replicas. Overall,

(i) *Superexponential or exponential jobs ($\beta \geq 1$)*

If $\beta \geq 1$ then fragmentation leads to a power law tail with index $k^\beta \alpha > \alpha$, where α is the power law index corresponding to data replication.

(ii) *Subexponential jobs ($\beta < 1$)*

For subexponential jobs, fragmentation yields power law delays with $k^\beta \alpha \to \alpha$ as $\beta \to 0$, meaning that replication might lead to better improvement (due to the lower constant factor) and should be preferred in that case.

It is therefore unclear which technique is preferable unless there is prior information on the distribution of the job sizes/failure statistics. Conditionally on this, one should perform careful computations to decide whether fragmentation or replication should be applied to deal with instabilities in large-scale distributed systems.

5.3 Concluding Remarks

The main contribution of this thesis is summarized as follows:

- The instability result from Chapter 4 reveals a new phenomenon: processor sharing is always unstable when restarts are employed. We also emphasize that job fragmentation cannot stabilize the system regardless of how small the fragments are made. Indeed, fragmentation can only postpone the time when the instability occurs, but cannot eliminate the phenomenon; serving one job at a time (e.g., FCFS) is highly

advisable in such systems. Similarly, the system cannot be stabilized by checkpointing regardless of how small the intervals between successive checkpoints are chosen. From an engineering perspective, our results indicate that traditional approaches in existing systems may be inadequate in the presence of failures. This new phenomenon demonstrates the need of revisiting existing techniques in large-scale failure-prone systems, where PS-based scheduling may perform poorly.

- The uniform approximation that was presented in Chapter 2 characterizes the entire body of the distribution for the number of retransmissions, which takes the form of the product of power law and Gamma distributions, thus allowing for an accurate estimation of the power law region. It also provides an assessment tool for efficiency and is applicable in modern network protocol design, e.g., retransmission based protocols in communication networks, where traditional approaches, e.g., blind data fragmentation, may be insufficient for achieving a good balance between throughput and resource utilization. Last, our model is generic and thus can be used towards improving the design of future complex and failure-prone systems in a variety of applications.

- The study of retransmissions over correlated channels in Chapter 3 shows that when the channel is correlated, or less formally, when it alternates between states of different quality, the tail asymptotics is determined by the properties of the 'best' channel state, e.g., the state that generates the lightest asymptotics in the corresponding independent model. This insensitivity to the detailed structure of the correlations as well as the optimistic best case predictions are useful both for modeling and dimensioning/capacity planning of such systems. However, a design relying on the best case scenario may be overly optimistic and even completely wrong if the best state of the channel is atypical. Last, the explicit approximation presented in Chapter 2 could be combined with these results in order to improve fragmentation techniques.

Furthermore, the observations and insights provided in Sections 5.2 and 5.1, could drive future research in the area of cloud computing where resource allocation and high through-

put algorithms are still in a developing stage. The main objective of this thesis is to gear the attention towards potential problems that arise in designing or analyzing modern large-scale systems. We have shown that traditional methodologies may fail due to the complexity and variability of such systems. Nevertheless, our novel approach and intuition has enabled us to uncover the underlying laws that govern their behavior and discover new phenomena via developing new analytical techniques or exploiting existing ones.

The aforementioned results also demonstrate the applicability of this work in various fields, beyond electrical engineering. We strongly believe that the tools and methodologies developed herein can be extended to other areas and disciplines, such as economics, statistics, operations research, computer science, applied math, etc., where phenomena of a similar flavor could be investigated. Hence, we are optimistic that the contents of this thesis will inspire modern engineering design and possibly help to solve existing problems in large-scale complex systems.

Bibliography

[1] Jelenković, P.R., Skiani, E.D.: Uniform approximation of the distribution for the number of retransmissions of bounded documents. In: Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE joint international conference on Measurement and Modeling of Computer Systems, SIGMETRICS '12, pp. 101-112 (June 2012).

[2] Jelenković, P.R., Skiani, E.D.: Distribution of the number of retransmissions of bounded documents. Advances in Applied Probability **47(2)**, June 2015. arXiv:1210.8421

[3] Jelenković, P.R., Skiani, E.D.: Retransmissions over correlated channels. SIGMETRICS Performance Evaluation Review **41**(2), 15–25, 2013.

[4] Jelenković, P.R., Skiani, E.D.: Is sharing with retransmissions causing instabilities? Proceedings of the The 2014 ACM international conference on Measurement and Modeling of Computer Systems (SIGMETRICS '14), pp. 167–179. SIGMETRICS Performance Evaluation Review **42**(1), 167–179, June 2014.

[5] Jelenković, P.R., Skiani, E.D.: Instability of Sharing Systems in the Presence of Retransmissions. Queueing Systems, 2015.

[6] Jelenković, P.R., Skiani, E.D.: Retransmission delays over Correlated Channels. *INFORMS Applied Probability Conference 2013*, Costa Rica, July 2013 (invited talk).

[7] Jelenković, P.R., Skiani, E.D.: Scheduling on a Channel with Failures and Retransmissions. *INFORMS 2013*, Minneapolis, USA, October 2013 (invited talk).

[8] Jelenković, P.R., Skiani, E.D.: Instability of Sharing Systems in the Presence of Retransmissions. *INFORMS 2014*, San Fransisco, November 2014 (invited talk).

[9] Jelenković, P.R., Tan, J.: Can retransmissions of superexponential documents cause subexponential delays? In: Proceedings of IEEE INFOCOM'07, pp. 892–900, 2007.

[10] Jelenković, P.R., Tan, J.: Characterizing heavy-tailed distributions induced by retransmissions. Advances in Applied Probability **45**(1), 106-138, 2013. (extended version) arXiv: 0709.1138v2.

[11] P. R. Jelenković and J. Tan. Are end-to-end acknowledgements causing power law delays in large multi-hop networks? In *14th Informs Applied Probability Conference*, July 2007.

[12] P. R. Jelenković and J. Tan. Is ALOHA causing power law delays? In *Proeedings of the 20th International Teletraffic Congress,* Ottawa, Canada *; Lecture Notes in Computer Science,* No 4516, Springer-Verlag, pages 1149–1160, June 2007.

[13] P. R. Jelenković and J. Tan. Dynamic packet fragmentation for wireless channels with failures. In *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, May 2008.

[14] P. R. Jelenković and J. Tan. Stability of finite population ALOHA with variable packets. *Technical Report EE2009-02-20, Department of Electrical Engineering, Columbia University*, New York, 2009, eprint arxiv: 0902.4481v2.

[15] P. R. Jelenković and J. Tan. Modulated branching processes, origins of power laws, and queueing duality. *Mathematics of Operations Research*, 35(4):807–829, 2010.

[16] P. R. Jelenković and M. Olvera-Cravioto. Implicit renewal theorem for trees with general weights. *Stochastic Processes and their Applications*, 122(9):3209–3238, 2012.

[17] Altman, E., Avrachenkov, K., Ayesta, U.: A survey on discriminatory processor sharing. Queueing Syst. Theory Appl. **53**(1-2), 53–63, 2006.

[18] Anantharam, V.: Scheduling strategies and long-range dependence. Queueing Systems **33**(1/3), 73–89, 1999.

[19] D. P. Bertsekas and R. Gallager. *Data Networks.* Prentice Hall, 2nd edition, 1992.

[20] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation.* Cambridge University Press, 1987.

[21] S. M. Ross. *A First Course in Probability.* Prentice Hall, 6th edition, 2002.

[22] Coffman Jr., E.G., Muntz, R.R., Trotter, H.: Waiting time distributions for processor-sharing systems. Journal of the ACM **17**(1), 123–130, 1970.

[23] Fayolle, G., Mitrani, I., Iasnogorodski, R.: Sharing a processor among many job classes. Journal of the ACM **27**(3), 519–532, 1980.

[24] Fiorini, P.M., Sheahan, R., Lipsky, L.: On unreliable computing systems when heavy-tails appear as a result of the recovery procedure. SIGMETRICS Performance Evaluation Review **33**(2), 15–17, 2005.

[25] Sheahan, R., Lipsky, L., Fiorini, P.M., Asmussen, S.: On the completion time distribution for tasks that must restart from the beginning if a failure occurs. SIGMETRICS Performance Evaluation Review **34**(3), 24–26, 2006.

[26] S. Asmussen, P. Fiorini, L. Lipsky, T. Rolski, and R. Sheahan. Asymptotic behavior of total times for jobs that must start over if a failure occurs. *Mathematics of Operations Research*, **33**(4), 932–944, November 2008.

[27] S. Asmussen and A. Rønn-Nielsen. Failure recovery via RESTART: Wallclock models. Research Report No. 4, Thiele Centre for Applied Mathematics in Natural Science, Aarhus University, March 2010.

[28] S. Asmussen, L. Lipsky and S. Thompson. Failure recovery in computing and data transmission: Restart and checkpointing. Springer Lecture Notes in Computer Science **8499**, 253–272, 2014.

[29] Jelenković, P., Momčilović, P.: Large deviation analysis of subexponential waiting times in a processor sharing queue. Mathematics of Operations Research **28**(3), 587–608, 2003.

[30] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1964.

[31] Kleinrock, L.: Time-shared systems: A theoretical treatment. J. ACM **14**(2), 242–261, 1967.

[32] Parekh, A.K., Gallagher, R.G.: A generalized processor sharing approach to flow control in integrated services networks: The single-node case. IEEE/ACM Trans. Netw. **1**(3), 344–357, 1993.

[33] Parekh, A.K., Gallagher, R.G.: A generalized processor sharing approach to flow control in integrated services networks: The multiple node case. IEEE/ACM Trans. Netw. **2**(2), 137–150, 1994.

[34] Wierman, A., Zwart, B.: Is tail-optimal scheduling possible? Operations Research **60**(5), 1249–1257, 2012.

[35] Yashkov, S.: Mathematical problems in the theory of shared-processor systems. Journal of Soviet Mathematics **58**(2), 101–147, 1992.

[36] Yashkov, S., Yashkova, A.: Processor sharing: A survey of the mathematical theory. Automation and Remote Control **68**(9), 1662–1731, 2007.

[37] J. Nair, M. Andreasson, L. Andrew, S. Low, and J. Doyle. File fragmentation over an unreliable channel. In *Proceedings of IEEE INFOCOM'10*, pages 965–973, March 2010.

[38] J. Nair, and S. Low. Optimal job fragmentation. In *SIGMETRICS Performance Evaluation Review*, **37**(2), 21–23, October 2009.

[39] J. Tan and N. Shroff. Transition from heavy to light tails in retransmission durations. In *Proceedings of IEEE INFOCOM'10*, pages 1334–1342, March 2010.