The heavy-tailed heavy-traffic machine-repairman problem: A GI/G/2 Queue

Yujiao Sun

Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China
sunyujiao@126.com

Ping Wang

Texas A & M University at Galveston, 200 Seawolf Parkway, Galveston, TX 77553, USA

Xiangpei Hu

Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China

Abstract

Recent research in queuing focuses more on the impacts of heavy-tailed service times, neglecting the fact that stochastic inter-arrival times are often heavy-tailed as well. We investigate the impacts of heavy-tailed gamma inter-arrival time distributions on the waiting-time distribution of a GI/G/2 queue, and simulate the tail properties of waiting-time distribution using real data.

Keywords: Heavy-tailed Arrival Times, GI/G/m Queuing, Stochastic Process

INTRODUCTION

The heavy tailed behaviors in various queuing systems have been well studied in early literature. Most of them investigate situations when either inter-arrival times or service times are heavy tailed. For situations when both times are heavy tailed, only a few studies have been conducted for the single-server systems (e.g., Boxma and Cohen 1999, Zwart 2001). In this study we will analyze a multi-server queuing system with both heavy-tailed inter-arrival times and service times.

The conventional queuing theory for systems assumes Poisson inter-arrival times and exponential service times. Such assumptions have been challenged with real-world examples, and consequently system performance such as waiting times has been reexamined (e.g., Medhi 2003).

Earlier treatments of the heavy tailed queuing systems are to approximate tail properties of either inter-arrival times or service times as exponential. For example, Halfin and Whitt (1981) prove a heavy-traffic theorem for the standard GI/M/m queue with heavy-tailed inter-arrival times, but generate approximations for GI/G/m queues only for the case of approximated exponential service times. Some queuing systems with heavy-tailed, non-Poisson inter-arrival times have also been investigated (see Abate and Whitt 1997; Boxma and Cohen 1998; Asmussen et al. 2000). For example, Whitt (2000) shows that the steady-state waiting-time distribution of an M/GI/m queue has a heavy tail (with appropriate definition), no matter what service-time distribution follows. Moreover, it has been recognized that some queuing systems can be better approximated by power or subexponential distributions (e.g., Greiner et al. 1999, Harris et al. 2000, and Asmussen et al. 1999). For example, the GI/M/1 models with power-tailed inter-arrival times in computer system environments are studied by Greiner et al. (1999), and a number of queuing applications with subexponential service times are investigated by Asmussen et al. (1999) and Foss and Korshunov (2013).

Queuing applications with heavy-tailed inter-arrival and service times have been investigated for a single-server system. Boxma and Cohen (1999) prove a heavy-traffic limit theorem for the distribution of the stationary actual waiting time with power-tailed service times and/or inter-arrival times. Baltrūnas et al. (2004) consider a subexponential service time in a GI/GI/1 queue, and Zwart (2011) considers a subexponential service time in a GI/G/1 queue. Their main objectives are to derive the tail behavior of the (busy period) stationary waiting time. Generally speaking, recent research focuses more on the impacts of heavy-tailed service times, neglecting that stochastic inter-arrival times are also often heavy-tailed. Figure 1 presents a schematic illustration of selected literature on heavy-tailed queuing systems.

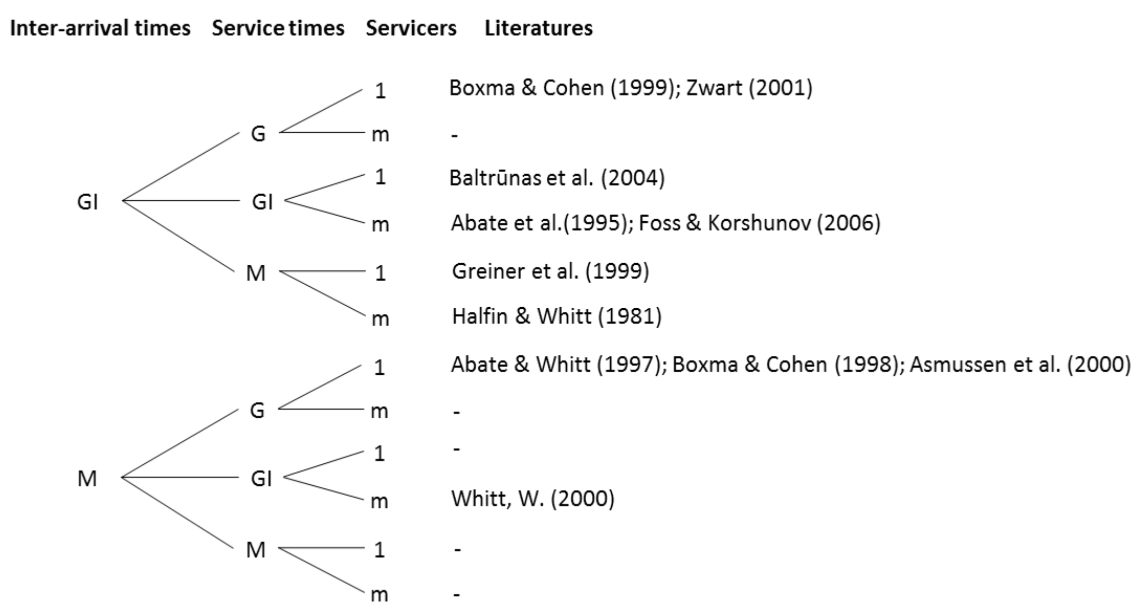


Figure 1 – Selected literature on heavy tailed queuing systems

In this study we will focus on a two-server queuing system with heavy-tailed inter-arrival times and service times. The remaining sections are organized as follows:

We first present an example of quay crane operations at container terminals, and conceptualize the loading process as a heavy-tailed heavy traffic machine-repairman problem. Analysis on a set of real data shows that the inter-arrival time can be better approximated as Gamma distributions and the service time for each individual server has the heavy-tailed property in general. After that, we conduct analysis on the tail properties of the joint distribution between the two servers. Numerical results are then provided to illustrate the tail properties of waiting time in this heavy-tailed machine repairman problem. Research findings and future research opportunities will be discussed at the end.

A HEAVY-TAILED HEAVY-TRAFFIC QUEUING PROBLEM

In this section we first present the containers' loading process to visualize a heavy-tailed service queuing system of interest. In a container terminal, there are two quay cranes (QCs) to load containers to a vessel, and 11 trucks are assigned to serve both QCs. Each truck picks up a container from the yard, and then transports it to the quayside for QC operations. The two QCs can load containers simultaneously, and each QC can only service one truck at a time. When a truck arrives to the quayside but none of both QCs is available, it has to wait. Such a waiting characterizes the queuing system as a heavy traffic one. Without loss of generality, we assume the loading process follows the first-come-first-serve principle and there are no restrictions for a container to be loaded by a particular QC.

This QCs' operation process can be viewed as a machine-repairmen problem, as shown in Figure 2. The trucks can be viewed as the machines that are subjects to be loaded (i.e., repaired) by a QC (i.e., a repairman) when they arrive at the quayside (i.e., they breakdown). We denote the number of machines by N, and for the data we are going to use, $N = 11$. Therefore, waiting time may exists before the repair starts. We use the terms inter-arrival times and service times from queuing theory instead of lifetimes and repair times as used in the conventional machine repairman problem.

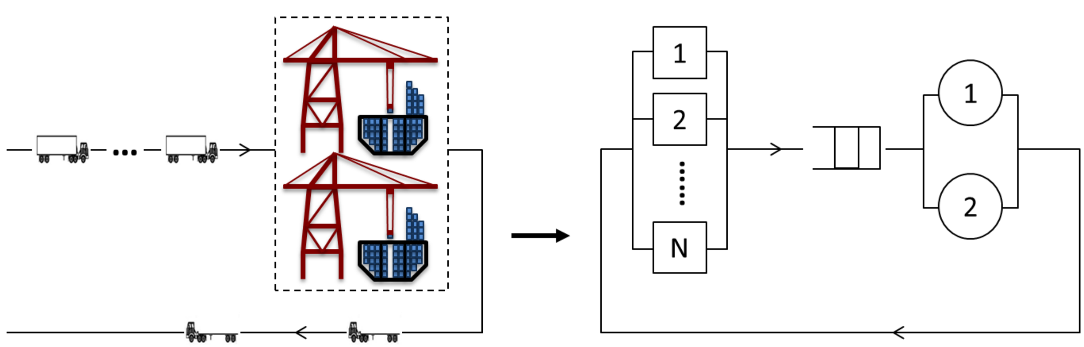


Figure 2 – The QCs' operation process and machine-repairmen problem

Tail Properties Analysis

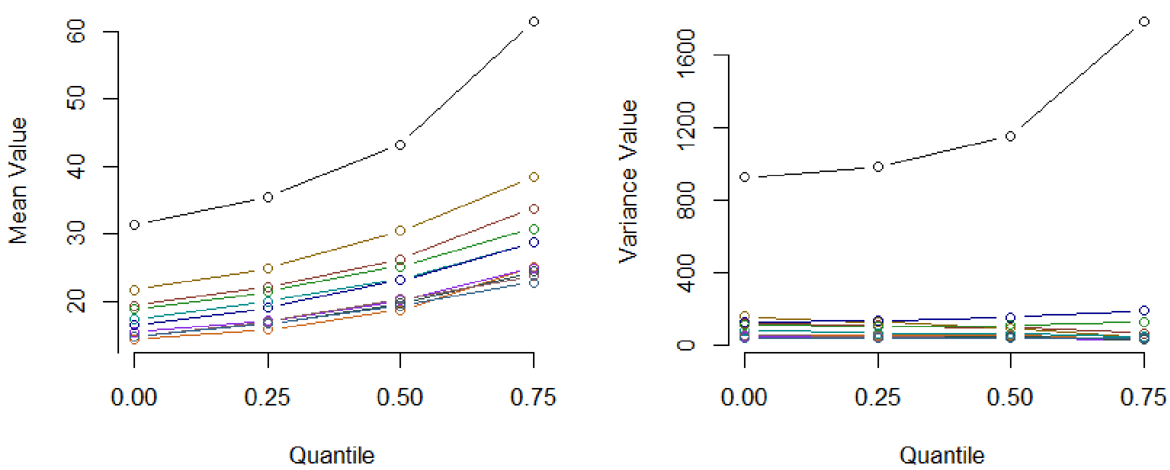
When considering the tail heaviness we analyzed both the moments and the tail index, specifically the first four moments for the shape of inter-arrival distributions. The first moment is the sample mean. The second moment is the sample variance. The third moment is skewness, which illustrates the lopsidedness of distributions. And the forth moment is kurtosis, and it is a measure of the heaviness of the tail of distributions.

We also use Hill's estimator to obtain the tail index of machines' lifetime distribution, repairmen's service time distribution, and waiting time distribution. Suppose the sample size of the time observations is n, and $x(i)$ denote the i^{th}-order statistic. The Hill's estimator with threshold equals k is proposed as the following equation (Hill 1975).

$$\alpha(k) = \frac{1}{k}\sum_{j=1}^{k} \ln\big(x(n-j+1)\big) - \ln\big(x(n-k)\big) \tag{1}$$

Moments of Truck Inter-Arrival Times

Eleven sets of truck inter-arrival time data are used to generate the four moments. Figure 3 shows the first four moments (i.e., mean, variance, skewness, and kurtosis) of eleven truck inter-arrival time distributions at the quintiles of 0%, 25%, 50% and 75% of the length of the data sets. The 11th distribution has outstanding different mean and variance from other distributions as presented in Figure 3. All 11 distributions are right tailed with positive skews. Two distributions are platykurtic with kurtosis less than 3, while other distributions are all leptokurtic with kurtosis greater than 3, indicating that those tails are heavy.



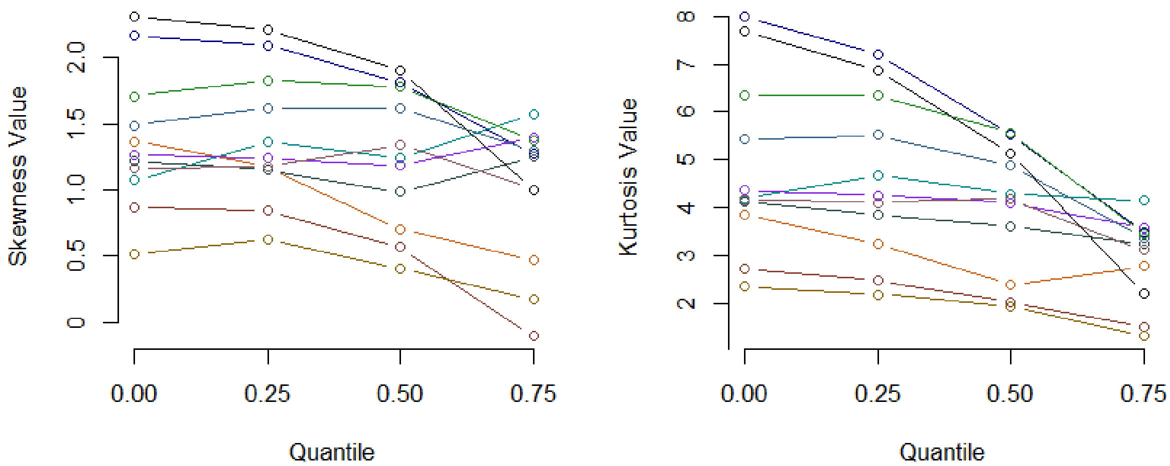
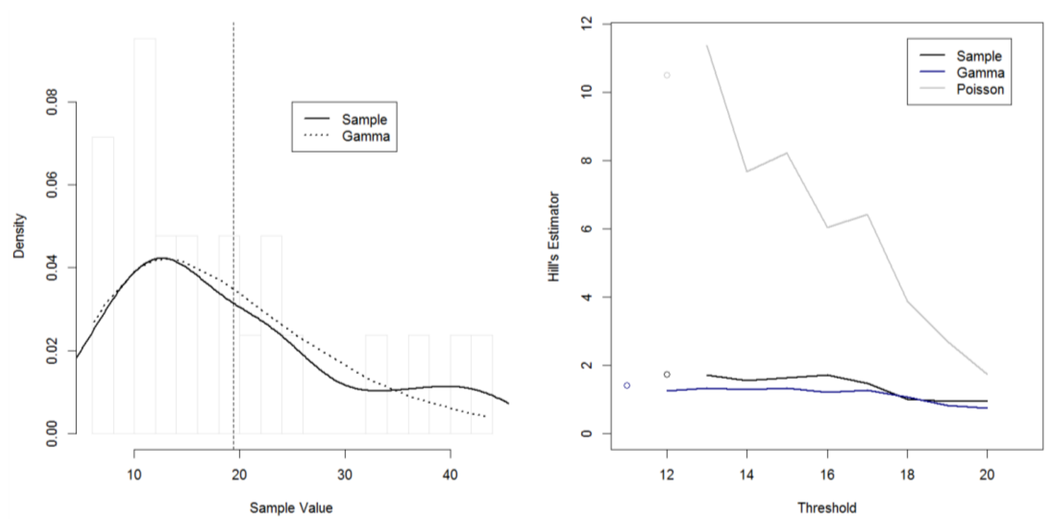


Figure 3 – Moments of eleven truck inter-arrival time distributions

Tail Indices of Heavy-Tailed Truck Inter-Arrival Times

The operating data collected in a container terminal show that the distributions of trucks' inter-arrival times fail to possess Poisson properties. For a truck's inter-arrival times at sample size of 21, Figure 4 presents the distribution plot and Hill's estimators of the truck's inter-arrival times. Figure 4 (a) plots the sample against Gamma, while Figure 4(b) plots the Hill's estimators of the sample versus that of the simulated Gamma and Poisson with the same sample mean and variance. The mean inter-arrival time is close to 19 minutes, as shown in Figure 4(a) by the dotted vertical line. When calculating Hill's estimators we set the thresholds at eight integers between 13 and 20. Figure 4(b) shows that the tail starts to be heavily distributed for values far below the mean.



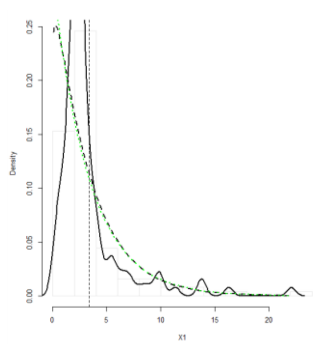
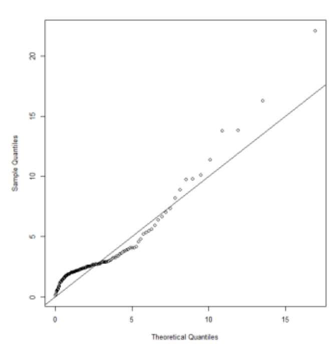
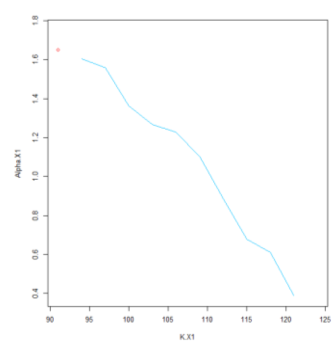
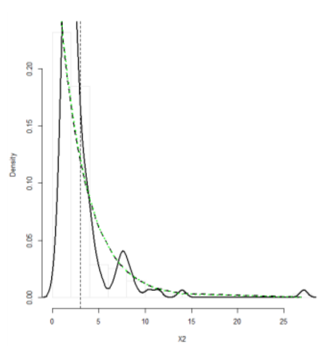
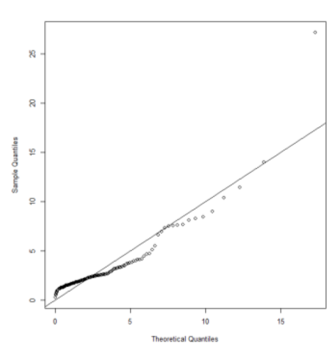
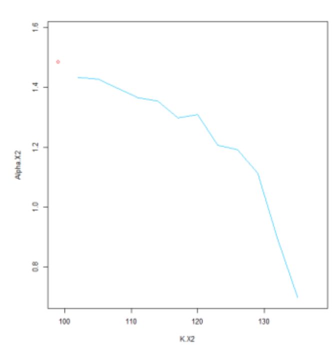
(a) (b)

Figure 4 – The distribution and Hill's estimator of the inter-arrival time of a truck

Both exploratory results of the inter-arrival times show that the inter-arrival process can be better approximated by Gamma, rather than by Poisson. Therefore, we assume that the inter-arrival times are independent Gamma distributions.

Tail Distributions and Hill's Estimators for Service Times

The data of two QCs corresponding to the 11 trucks are collected to analyze the distributions of service times. Two separate plots of both service times, denoted as X1 and X2, versus exponential distributions (not shown in this paper) indicate that both service times are far from exponential. The joint service times is generated with R as a random sum (W) of the two discrete service times. Figure 5 provides the distributions, Gamma Q-Q plots, and Hill's estimators for both of the two QCs' service times, as well as those of the random sum. While the Gamma Q-Q plots of the distributions of QCs' service times versus Gamma are not strictly convincing, the joint distribution of random sum seems to be appropriate if approximated as Gamma. Besides, the Hill's estimators of the three distributions are below two (2) for threshold immediately greater than the index of mean value, as indicated by the small circle on the upper left corner of each plot of Hill's estimator, providing a strong evidence that the three distributions are heavy-tailed for values beyond their means.

	Distribution	Gamma Q-Q plot	Hill estimator α
X1			
X2			

W

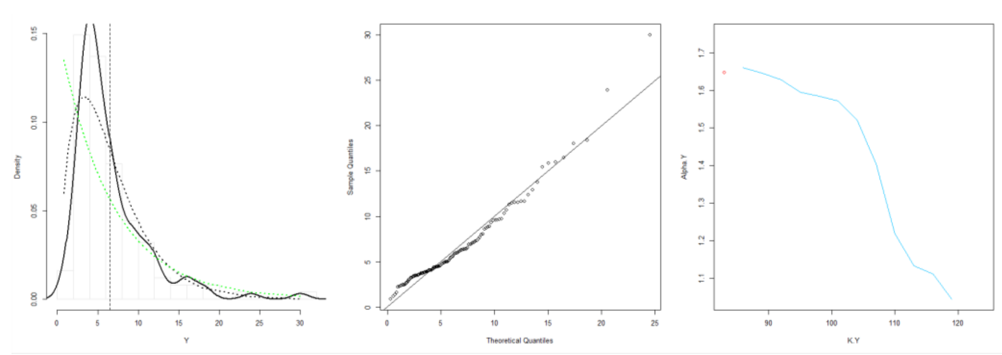
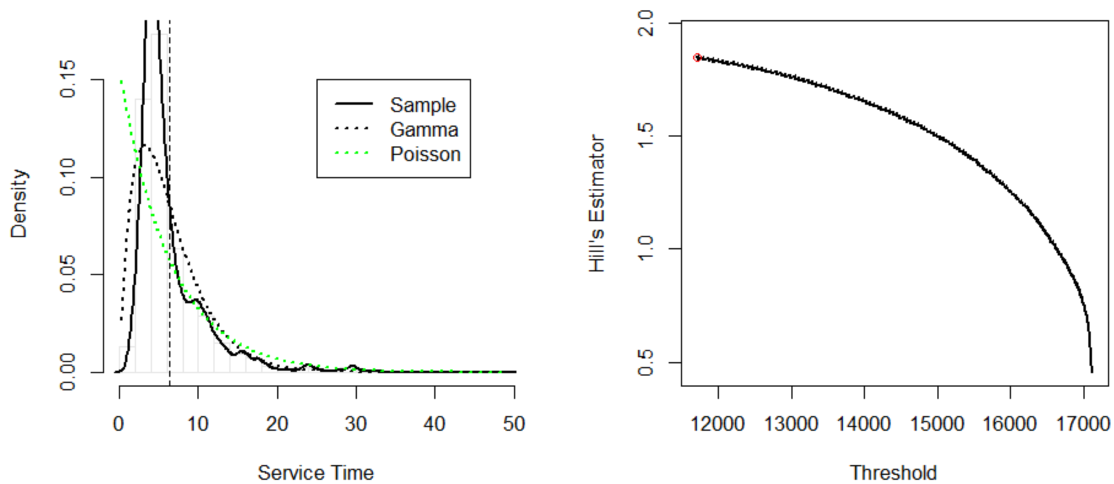


Figure 5 – Tail distributions and Hill's estimators for QCs

Convolution Simulation of Service Times

For such a complicated queuing system, it is impossible to derive a closed-form representation of the waiting time distribution. In order to better understand the joint distribution of the two servers, we simulate the joint service time as the convolution sum of two continuous operation with the quasi Monte Carlo method. 1,000 data points are generated for each service times, which are used to calculate the convolution sum Y, i.e., Y=X1+X2. The Hill's estimator of the convolution sum Y is much smoother than that of the random sum W (Figure 6). Q-Q plots for simulated convolution sum versus a Gamma distribution and an exponential distribution are also presented. It is evident that the convolution sum cannot be approximated as exponential. While the distribution of the convolution sum does not strictly overlap with the line of *y=x* in the Gamma Q-Q plot, it is still of the linear pattern and the trajectory of its tail portion is parallel to the line of *y=x*, indicating that the tail distribution is still a gamma distribution but might have a constant shift for its location parameter.



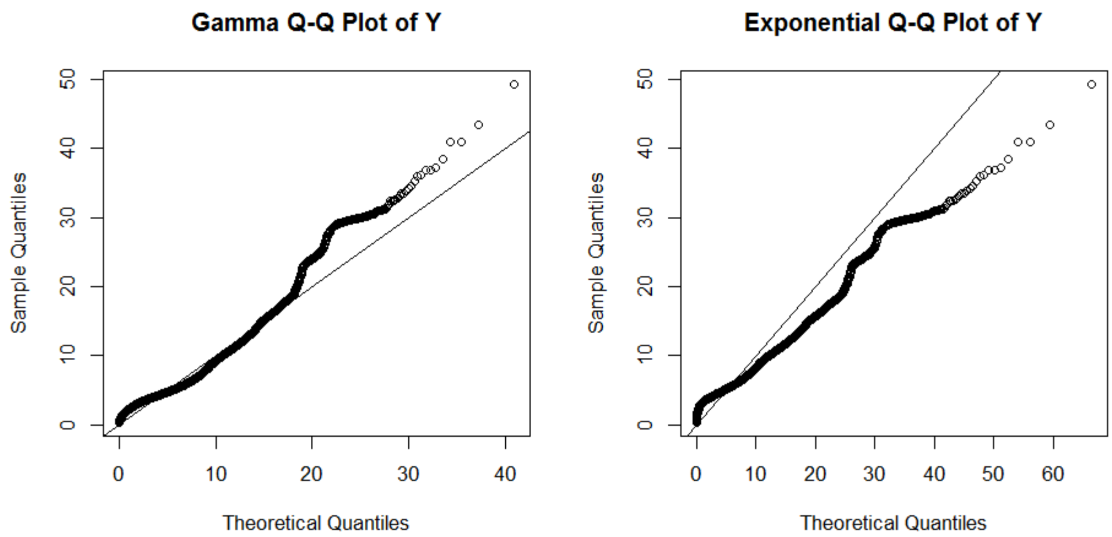
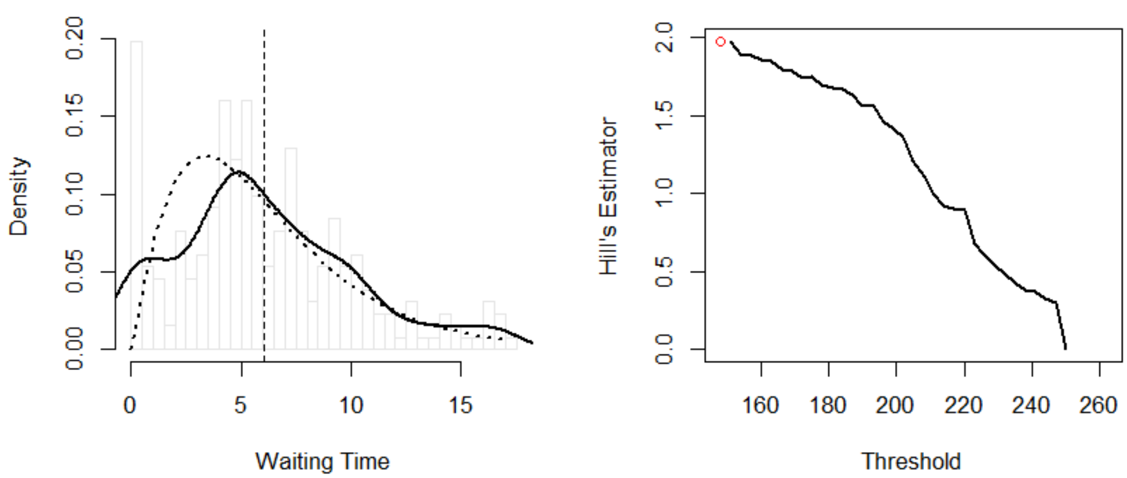


Figure 6 – Convolution simulation of service times

Tail Behavior of the Waiting Time Distribution

In the last part of tail behavior analysis, we simulate waiting time distribution based on the previous analysis of both inter-arrival time and service time distributions. Both inter-arrival and service times are assumed to be heavy-tailed with Gamma distributional properties. Figure 7 illustrates the tail properties of the waiting time distribution. The Hill's estimators for threshold greater than the index of mean value are all less than 2. This result means that the waiting time has a heavy-tailed distribution when inter-arrival times and service times are both Gamma distributions. Further Q-Q plots indicate that the distribution of waiting time fits the Gamma model very well.



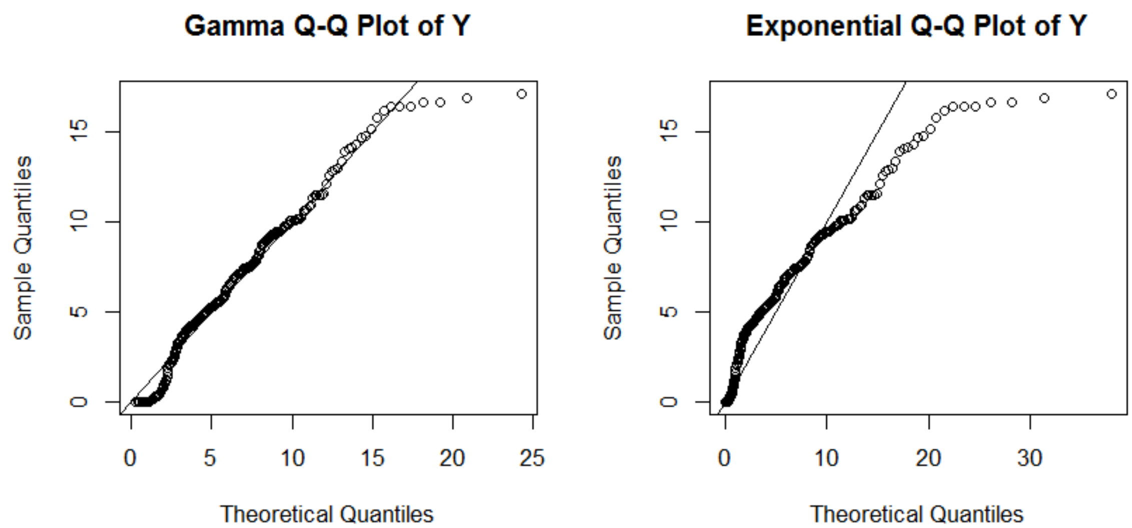


Figure 7 – Tail behavior of waiting time distribution

CONCLUSIONS

This study shows the tail behavior of waiting-time distribution of a GI/G/2 queuing system when both inter-arrival time and service time are heavy-tailed Gamma distributions. We use the real data from a container terminal to visualize the problem as a machine-repairman problem and simulate the tail behavior of the waiting time. We contribute to the heavy-tailed queuing literature by presenting an approach of approximating the tail distribution of the joint service time as the convolution sum of both individual service times when both service times have different trajectories of tail heaviness. It also contributes to the literature of machine repairman problem by analyzing the tail properties of waiting time when both inter-arrival and service times in a GI/G/m system have finite means but infinite variances. Based on our data, we find that tail properties of inter-arrival time have higher influence on the tail properties of waiting time than that of the service times. Two future research opportunities might be of interesting. First, with necessary assumptions on the distributional parameters of both inter-arrival and service time distributions, this study can be extended by analytically investigate the tail properties of waiting time and/or sojourn time distributions. It will be also interesting to see whether or not different tail heaviness of both inter-arrival time and services affect the system performance in different ways.

Acknowledgements
We would like to thank the China Scholarship Council for a scholar exchange grant that led to this collaborative research.

Bibliography

Abate, J., Choudhury, G. L., & Whitt, W. 1995. Exponential approximations for tail probabilities in queues, I: waiting times. Operations Research, 43(5), 885-901.

Abate, J., & Whitt, W. 1997. Asymptotics for M/G/1 low-priority waiting-time tail probabilities. Queueing Systems, 25(1-4), 173-233.

Asmussen, S., Binswanger, K., & Højgaard, B. 2000. Rare events simulation for heavy-tailed distributions. Bernoulli, 6(2), 303-322.

Baltrūnas, A., Daley, D. J., & Klüppelberg, C. 2004. Tail behaviour of the busy period of a GI/GI/1 queue with subexponential service times. Stochastic processes and their applications, 111(2), 237-258.

Boxma, O. J., & Cohen, J. W. 1998. The M/G/1 queue with heavy-tailed service time distribution. Selected Areas in Communications, IEEE Journal on, 16(5), 749-763.

Boxma, O. J., & Cohen, J. W. 1999. Heavy-traffic analysis for the GI/G/1 queue with heavy-tailed distributions. Queueing systems, 33(1-3), 177-204.

Foss, S., & Korshunov, D. 2006. Heavy tails in multi-server queue. Queueing Systems, 52(1), 31-48.

Greiner, M., Jobmann, M., & Lipsky, L. 1999. The importance of power-tail distributions for modeling queueing systems. Operations Research, 47(2), 313-326.

Halfin, S., & Whitt, W. 1981. Heavy-traffic limits for queues with many exponential servers. Operations research, 29(3), 567-588.

Harris, C. M., Brill, P. H., & Fischer, M. J. 2000. Internet-type queues with power-tailed interarrival times and computational methods for their analysis. INFORMS Journal on Computing, 12(4), 261-271.

Hill, B. M. 1975. A simple general approach to inference about the tail of a distribution. The annals of statistics, 3(5), 1163-1174.

House, A. W. H. 2004. The convolution sum for discrete-time LTI systems. *Discrete-Time Systems and Signals*. Available at http://www.eecg.toronto.edu/~ahouse/mirror/engi7824/course_notes_7824_part6.pdf.

Medhi, J. 2002. Stochastic models in queuing theory. Academic Press, San Diego, CA.

Whitt, W. 2000. The impact of a heavy-tailed service-time distribution upon the M/GI/s waiting-time distribution. Queueing Systems, 36(1-3), 71-87.

Zwart, A. P. 2001. Tail asymptotics for the busy period in the GI/G/1 queue. Mathematics of Operations Research, 26(3), 485-493.