Due: April 26, 2017

Task 1 (M/M/1/K System) Calculate the mean and variance of the bounded Pareto distribution for the parameters given above. Show all the math for the derivation.

$$f_X = \frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} x^{-\alpha-1}$$

$$\begin{aligned}
\bar{X} = E[X] &= \int_{-\infty}^{+\infty} x * f_X(x) * dx \\
&= \int_k^p x * f_X(x) * dx \\
&= \frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \int_k^p x \cdot x^{-\alpha-1} dx \\
&= \frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \left[\frac{1}{-\alpha+1} x^{-\alpha+1}\right|_k^p \\
&= \frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \left(\frac{1}{\alpha-1}\right)\left(k^{-\alpha+1} - p^{-\alpha+1}\right) \\
&= 3000 \quad k = 332, p = 10^{10}, \alpha = 1.1
\end{aligned}$$

$$\begin{aligned}
E[X^2] &= \int_{-\infty}^{+\infty} x^2 * f_X(x) * dx \\
&= \int_k^p x^2 * f_X(x) * dx \\
&= \frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \int_k^p x^2 \cdot x^{-\alpha-1} dx \\
&= \frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \left[\frac{1}{-\alpha+2} x^{-\alpha+2}\right|_k^p \\
&= \frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \left(\frac{1}{\alpha-2}\right)\left(k^{-\alpha+2} - p^{-\alpha+2}\right) \\
&= 7.25 * 10^{11} \quad k = 332, p = 10^{10}, \alpha = 1.1
\end{aligned}$$

$$\begin{aligned}
\sigma_X^2 &= E[X^2] - (E[X])^2 \\
&= \frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \left(\frac{1}{\alpha-2}\right)\left(k^{-\alpha+2} - p^{-\alpha+2}\right) - \left(\frac{\alpha k^\alpha}{1-\left(\frac{k}{p}\right)^\alpha} \left(\frac{1}{\alpha-1}\right)\left(k^{-\alpha+1} - p^{-\alpha+1}\right)\right)^2 \\
&= 7.25 * 10^{11} - 3000^2 \quad k = 332, p = 10^{10}, \alpha = 1.1 \\
&= 725098000000
\end{aligned}$$

Task 2 Read the paper [CL97] in the reading list, and answer the following questions:

(a) How do the authors define "heavy-tailed" distributions?

`A distribution whose tail is scale free; i.e., follows a power law.`

(b) What is an α-Stable distribution, and why are they introduced?

An α-Stable distribution is a distribution with power-law tails that have the same α (location parameter) as the parent distribution. These distributions are used to analyze the convergence properties of power-law distributions.

(c) What is a "swamping observation" and what is its significance?

An outlier that will change the observed $\bar{\mu}$ to be 2x actual. In power-law distributions, a swamping observation is relatively common for values of α between 1.0 and 1.7, which makes it difficult to perform consistent simulation results.

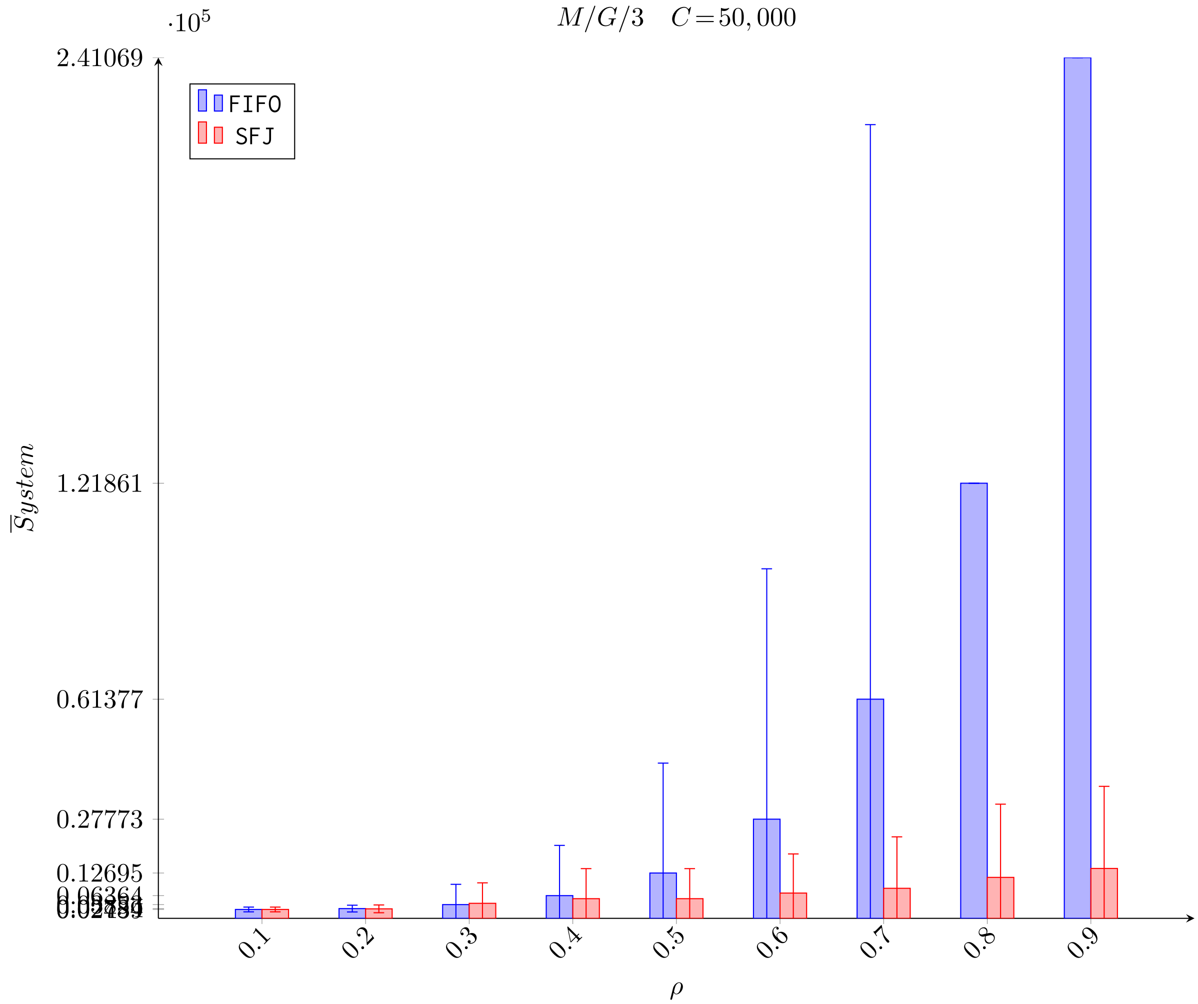
(d) Under what conditions do the authors claim that a simulation may always be in transient state?

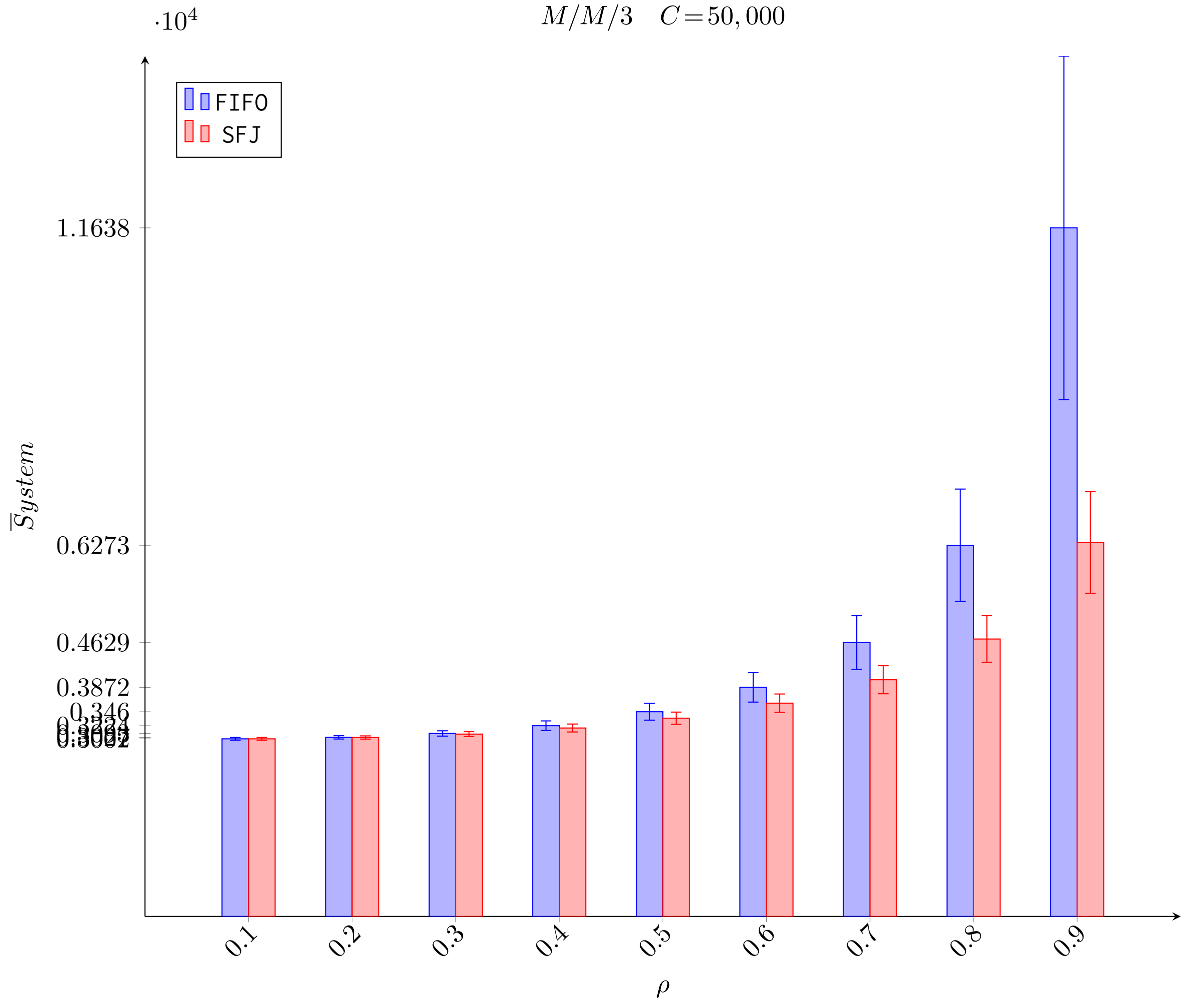
The authors explain that power-law distributions slowly converge to steady state, and are highly variable during that transition period. For some distributions with $0 \leq \alpha \lesssim 1.7$, the number of observations to reach a steady state is so high that the system should be considered permanently transient .

(e) Overall, what are the conclusions that the authors draw regarding simulations with heavy-tailed workload?

Unbounded power-law distributions are infeasible to simulate due to the time required to reach steady state. By bounding the distribution, the effect of swamping observations is minimized and steady state can be achieved. The value of the bounds can be calculated by statistical inference from the time of the simulation.

Task 3) This task involves the 3-server queueing system. Plot the average customer system time against the value of ρ, for $\rho = 0.1$ to $\rho = 0.9$, in increments of 0.1 (include confidence intervals). Note that $\rho = \frac{\lambda \bar{x}}{3}$ for this system. Use $C = 50,000$ customers. Compile two pairs of plots, one pair for each of the service disciplines (FCFS and SJF), where one plot of each pair corresponds to the M/M/3 system and one plot corresponds to the M/G/3 system. Discuss the relative performance of the plots; did you expect these results based on what we have covered in class?

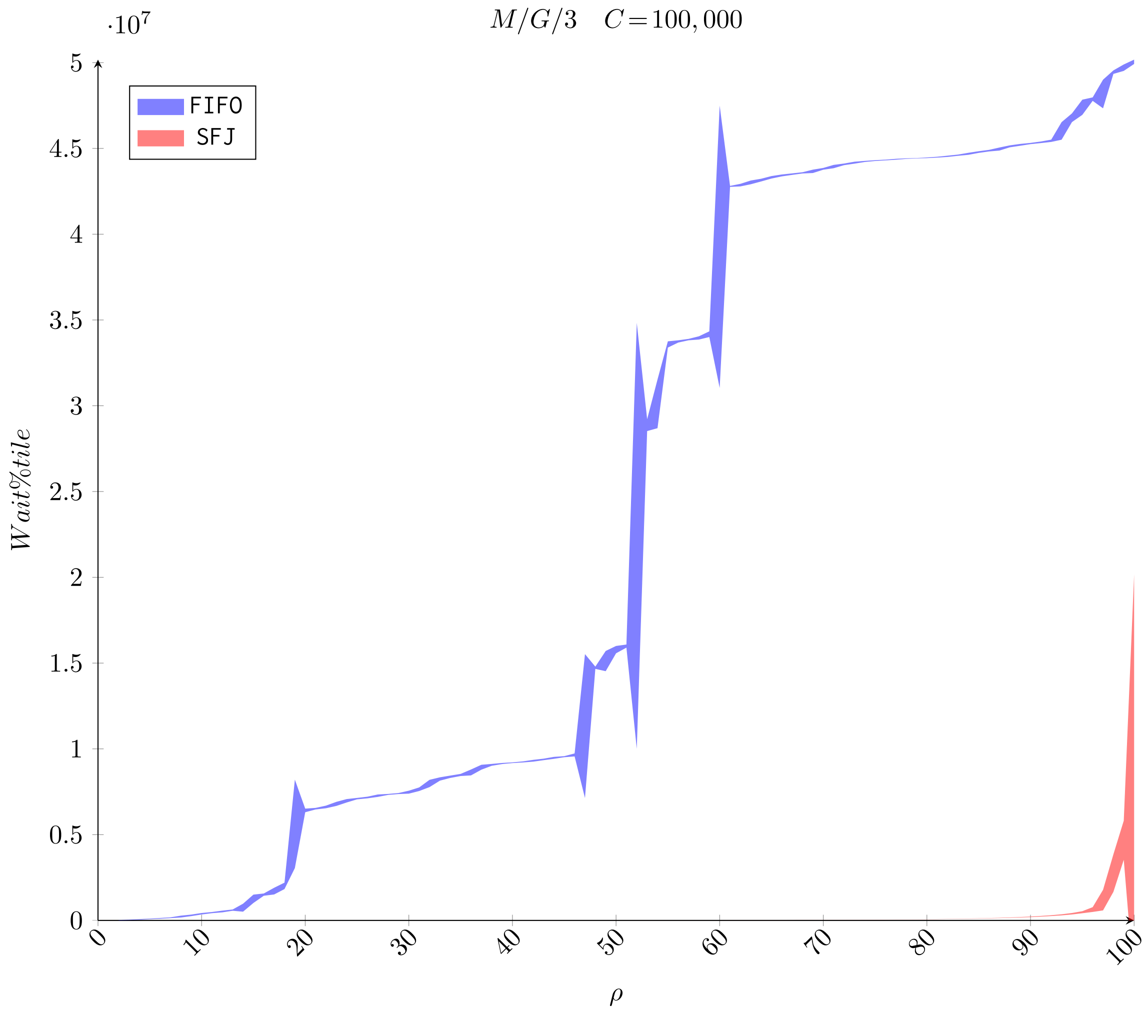


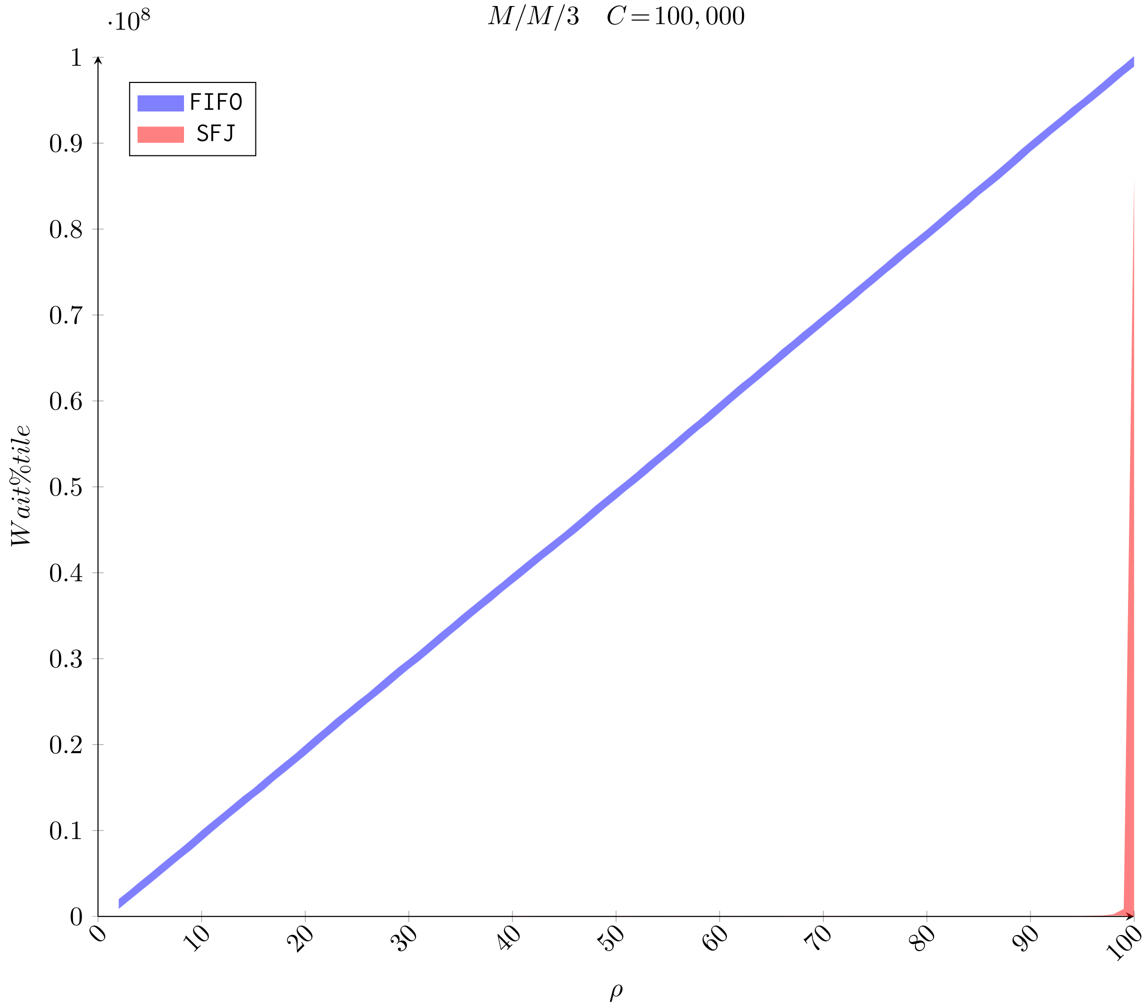


Note on graphing: error bars are omitted on the last two M/G/3 FIFO datapoints for readability. The variability of the Pareto distribution results in high variability,

and the results are in line with expectation. The SJF discipline reduces overall system time by prioritizing shorter jobs, and this is especially true with long-tailed distributions where a large number of jobs will be starved due to high service times.

Task 4 This task involves the single-server queueing system. You will assume that $\rho = 0.5$ and that $C = 100,000$. Let $T(x)$ denote the total system time for a customer whose service time is x, and $W(x) = T(x) - x$ its waiting time. We define the slowdown for a customer with service time x as $S(x) = \frac{W(x)}{x}$. The slowdown can be thought of as a measure of fairness. The objective of this task is to see how the size of the service time affects the slowdown of the customer. For this, you need to compute the slowdown for each of the C customers, and for each of the two service disciplines (FCFS and SJF). Since there are too many values of $S(x)$, making it difficult to plot, you will plot the slowdown against the percentile of the service time distribution. Use $B = 100$ bins, and place each customer to the appropriate bin according to its service time (i.e., the 1% of customers with the smallest service times go into the first bin, the 1% of customers with the next smallest service times go to the second bin, etc). For each bin, take an average of the slowdown for the customers in that bin. Then, plot the slowdown against the number of bins; compile two plots, one for FCFS and one for SJF. What can you say about the fairness of each of these service disciplines? How do the results compare to the slowdown of the processor sharing discipline which we discussed in class?





Note on implementation and graphing: Error bands shown in shaded region. The work in queue at time the simulation finishes is not included in metrics. For SJF, this work in queue would - by definition - have the longest wait times. The FIFO discipline

is inherently fair given random I.I.D. arrival and service times. The bounded pareto service distribution gives a very jagged percentile graph with high variance due to the outliers, but M/M/1 shows a completely level distribution. SFJ shows very low

service times for all but most unfortunate, and the customers in the queue at the time of simulation end would be the most unfortunate of all. Why fairness? Fairness

describes the concept of rough equal shared resource usage, and high slowdown reflects that sharing is not equal.