Project 2: Queueing

WAYNE NYABUTO
&
EDEN SEYOUM

Introduction

In this experiment, we explored the response rates for a simulation of a server executing jobs with job sizes and interarrival times taken from varying exponential and hyper-exponential distributions with a few different variances. This gives us insight into how variance of job sizes (for experiment 1) and interarrival times (for experiment 2) affect response time.

We sought to answer two primary questions:

1. How does the response time vary for different values of **λ** and different values of Var(x) given that the interarrival times are exponentially distributed with mean 1 and job sizes are drawn from a distribution with a two-step hyper-exponential distribution?
2. If instead we had exponentially distributed job sizes with mean 1 and variability of interarrival times, how would the response time in each case change?

We are interested in answering these questions so that we can have an idea of what mean response times of systems would be given different distributions of job sizes and arrival times. In practice this would be important for scheduling (maybe the order in which jobs are added to a system).

Experimental Setup

In the first part of our experiment we tested the mean response time of our job sequence of 10^6 jobs with interarrival times from an exponential distribution with var(x) ∈ {1, 10, 20, 50} and **λ** ∈ {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 07, 0.8, 0.9} in that order. We tested job sizes from a two-step hyper exponential, X ~ H2(μ1, μ2, p).

To find the parameters for the hyper exponential:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \sum_{i=1}^{n} p_i \int_0^{\infty} x \mu_i e^{-\mu x} \, dx = \sum_{i=1}^{n} \frac{p_i}{\mu_i}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \sum_{i=1}^{n} p_i \int_0^{\infty} x^2 \mu_i e^{-\mu x} \, dx = \sum_{i=1}^{n} \frac{2}{\mu_i^2} p_i$$

$$Var[X] = E[X^2] - E[X]^2 = \sum_{i=1}^{n} \frac{2}{\mu_i^2} p_i - [\sum_{i=1}^{n} \frac{p_i}{\mu_i}]^2$$

$$\frac{p}{\mu_i} = \frac{1-p}{\mu_2}$$

Given that E[x] = 1, we also have

$$\frac{p}{\mu_1} + \frac{1-p}{\mu_2} = 1$$

$$\text{Var}[X] = \frac{2p}{{\mu_1}^2} + \frac{2(1-p)}{{\mu_2}^2} - 1$$

Using these simultaneous equations, we found the following values for each value of var(x).

	μ1	μ2	p
var(x) = 1	1	1	0.5
var(x) = 10	1.9045	0.0955	0.9523
var(x) = 20	1.9512	0.0488	0.9756
var(x) = 50	1.9802	0.0198	0.9901

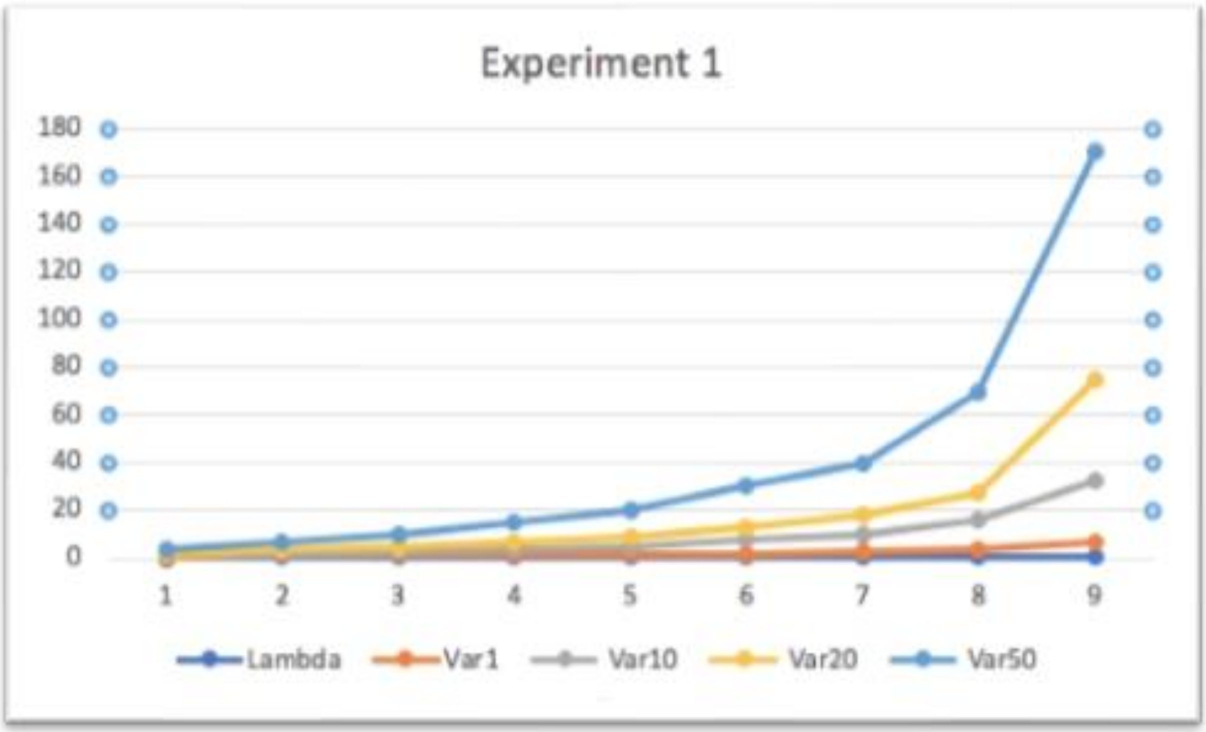
To simulate server accesses, we created "requests" with functions that produced integers draw from the described distributions. We kept track of the job size and response time for each job being process. We also kept track of the current time, next arrival time and remaining job size. This allowed us to know add to the queue (implemented as a linked list) when the server is busy and process a job when it is idle. The interarrival times are randomly drawn from an exponential distribution x~ Exp(**λ**). The exponential distribution for the interarrival times was implemented from a sequence of pseudo random generating of a random number between 0 and 1 and inserting that number into the inverse of the function $f(x) = \boldsymbol{\lambda} e^{-\lambda x}$.

In the second part of our experiment we tested how the response times would change if instead we used an exponential distribution for the job sizes and changed the variability of the interarrival times. For these round of the experiment we used λ values higher than 1 to keep up with rate of arrival which is on average 1 (to maintain the stability of the system); we used $\boldsymbol{\lambda} \in \{1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9\}$.

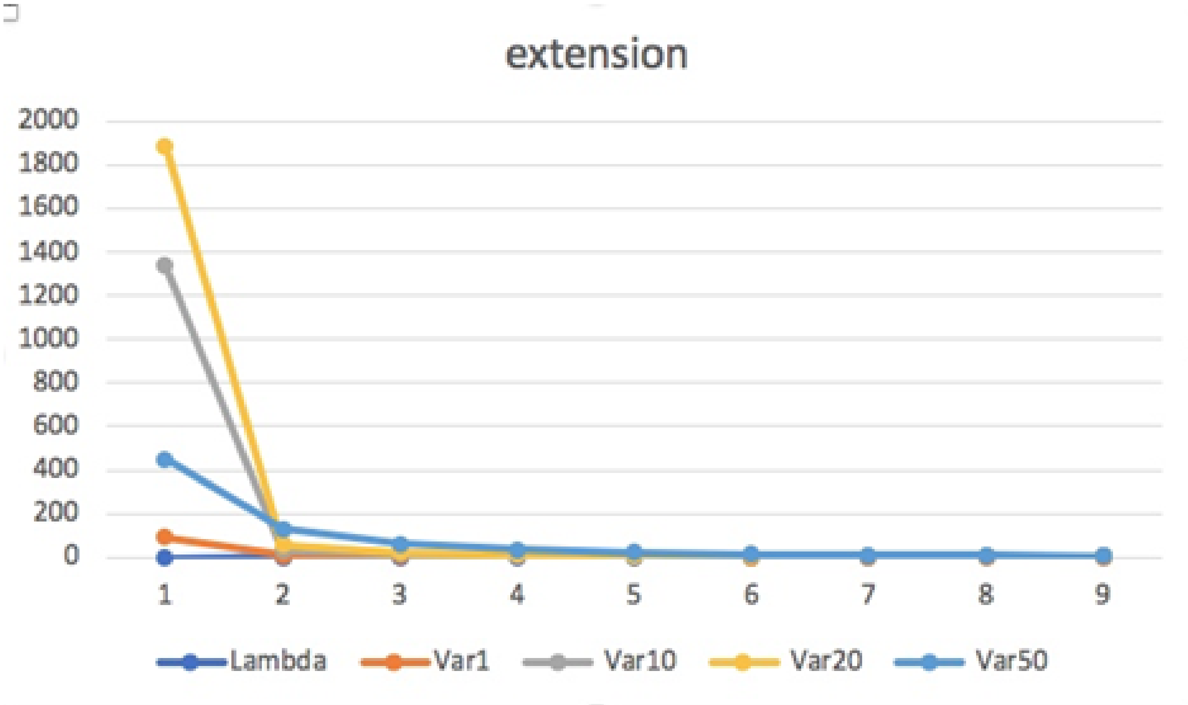
Results and Discussion

Experiment 1

For all four variances, mean response time increases exponentially with increase in lambda values. This trend can be explained as follows:
Lambda values have no effect on job sizes. They only affect generation of interarrival times. Interarrival times are exponentially distributed and decrease with the value of lambda. Lower interarrival times mean a larger rate of arrival and therefore a larger response time.
Higher variances in job sizes also exhibit higher mean response time. This can be explained as follows:
Higher variances have higher mew1 values and lower mew2 values. Also, the probability of generating jobs with mew1 is a lot higher than that of generating with mew2. This would mean that most of the time larger job sizes are generated with the larger mew1.
The probability of using mew1 increases with variance. Therefore, higher variance results in higher job sizes and therefore results in higher processing times and perpetually higher mean response rate.



Experiment 2



The graph has higher values of mean response time for lower values of λ.
Here lambda is used in job size generation.
For larger values of λ, it is more likely that our generated jobs would have smaller sizes. These smaller job sizes result in shorter mean response times.
The graph exhibits higher mean response time for higher values of variance of the interarrival time. (ignoring the extremely low lambdas)
Higher variances have higher mew1 values and lower mew2 values. Also, the probability of generating jobs with mew1 is a lot higher than that of generating with mew2. This would mean that most of the time interarrival times are generated with the larger mew1.
The probability of using mew1 increases with variance. Therefore, higher variance results in generation of higher interarrival times and therefore results in lower rate of arrival and perpetually lower mean response rate.

Conclusion

In this experiment we tried to find the effect of different job size distributions, exponential or hyper-exponential on the response time with different values for variance. In the first experiment, where the interarrival times for the jobs to be processed is from an exponential distribution and we have job size variability, we found that response times are higher with higher values of **λ**, where interarrival time is distributed with x~ Exp(**λ**).

For our second experiment we have smaller response time for higher values of **λ** where job size is distributed with x~ Exp(**λ**) and we have interarrival time variability.

In the future we would like to explore the same variable and conditions affect mean response time in a multi-server system.