Waiting for the Bus

-- CS166 First Project

Executive Summary

This project model a bus route system using Python. Simulation is run and statistics of interest are computed to find out how many buses should be in the system to ensure optimized performance. Having 15 buses in the system would make the system reach a balanced optimal state in terms of waiting time, length of the queue, taking cost into consideration.

Introduction

In this project, a model is implemented by Python to simulate the bus route system. The goal is to find out how many buses should be in the system to ensure optimized performance, in terms of waiting time for passengers, length of the queue at each stop, etc. The performance of the system is analyzed and the optimum number is decided based on experiments.

Model Description

The model simulates a simplified version of a bus system. Assumptions and variables of the model are discussed in this section.

Classic queueing model is chosen to represent the situation. The system is broken down into customers (passengers), queues (bus stops), servers (buses). Specifically, M/G/1 queue is used. In M/G/1 queue, arrival rates are exponentially distributed and service rates follows any probability distribution. There is one server at a time (one bus serving for each queue at a time). The assumptions and justifications are listed in the section below.

Model Assumptions

1. The bus route is circular and operates 24 hours a day. It contains 15 different stops.

2. The rate at which passengers join the queue at a bus stop is constant and equal to 1 person per minute. The time between consecutive passengers joining the queue is assumed to follow an exponential distribution with $\lambda = 1$. Exponential distribution fits the scenario because it's reasonable to assume that the events occur at a fixed rate and occur independently (current passenger's decision to join the queue doesn't depend on decisions of past passengers)

$$Arrival\ Interval\ (min) \sim Exponential\ (1)$$

3. In the lifetime of a passenger in this system, the passenger arrives at a bus stop, waits until the next bus arrives, gets on the bus, and gets off the bus at the destination.
4. Each passenger chooses uniformly at a random destination that is at most 7 stops away from where they start.
5. The time it takes passengers to embark and disembark both follows normal distribution (passengers can only embark or disembark one by one):

$$Disembark\ time\ (min) \sim Normal\ (0.03n,\ 0.01\sqrt{n}\ )$$

$$Embark\ time\ (min) \sim Normal\ (0.05n,\ 0.01\sqrt{n}\ )$$

,where n is the number of passengers who reaches their destination

Normal distribution is chosen because we would expect it normally would take each passenger 0.03 minutes to embark and 0.05 minutes to disembark so the values should be centered around it with some variation.

6. The travel time of the bus between consecutive bus stops is also assumed to follow a normal distribution with mean of 2 and a rather big standard deviation of 0.5:

$$Travel\ time\ between\ stops\ (min) \sim Normal\ (2,\ 0.5)$$

7. The maximum capacity of each bus is 130 passengers. If the bus is full, no more passengers can embark and they have to wait for the next bus to arrive.

Model Specification

Based on the assumptions, the chosen model, variables, and update rules are shown below:

Model Variables

- For bus stop
    - Number of bus stops: 15
    - Queue length
    - Passengers at the bus stop
    - Arrival distribution
- For passenger
    - Number of passengers
    - The time for passengers to embark
    - The time for passengers to disembark
    - Source and Destination
    - Time waited for the bus
    - Time spent in the system
- For bus
    - Max capacity (130 passengers)
    - Travel time between stops
    - Bus location
    - Passengers on the bus

Update rules

- For bus stop
    - Passengers are added to queues with exponential distribution
    - Passengers are randomly assigned a destination from among the next 7 stations
- For bus
    - No passengers can embark while other passengers are still disembarking
    - If the bus is full, no more passengers can embark
    - When a bus arrives at a station, people disembark, and new passengers embark. Time taken for each step follows normal distribution
    - Busses move between stations after an amount of time following a normal distribution
- For passengers
    - Passengers can travel at most 7 stops

Implementation

The model is implemented using an object-oriented approach, with classes for bus, bus stop, passenger, bus system. The attributes keep track of the properties of each object and the method handles the update rules. A event and schedule class is also implemented to schedule the timestamp for each event, for example, starting to embark, arrive at the next bus stop, etc. Specific events rather than periods are scheduled.

Priority queue is implemented with heap for the schedule because it maintains the heap property to have the item with the highest priority (minimum or maximum) on the top, so it's always O(1) to retrieve the next event based on time stamp to run. So it's computationally efficient compared to simply a list of events.

Results and Discussion

Empirical analysis

The results are obtained from running the simulations 50 times for a number of buses from 1 to 20 after a time unit of 24 * 60 minutes.

Expected value and 95% confidence interval are chosen to analyze the results because the expected value would show the average value over simulations and the confidence interval would show the span of possible outcomes and how much confidence we have. All results contribute to answering the question 'How many buses should be on the route?'

The three following statistics are being compared:

1. Average passenger waiting time given the number of buses on the route

Waiting time is defined as the duration between a passenger joining a queue at a bus stop and when the passenger gets on the bus. Passengers will be more satisfied when waiting for shorter period of time.

2. Maximum queue length at all bus stops during a day given the number of buses on the route

Maximum queue length captures the most extreme case - the longest queue length encountered when a new passenger arrives. Passengers would be more likely to join the queue and feel more comfortable when this quantity is minimized.

3. Mean queue length across bus stops after a day given the number of buses on the route

Mean queue length across bus stops at the end of the simulation gives an idea of the queue length at equilibrium. This quantity supplements maximum queue length in helping to understand on average what a new passenger should expect the queue length to be when arriving.

Number of buses	Maximum queue length	95% CI	Mean queue length	95% CI	Mean waiting time	95% CI
2	69.60	[68.91, 70.29]	22.82	[22.21, 23.43]	23.43	[23.31, 23.55]
3	59.40	[58.51, 60.29]	18.32	[17.91, 18.73]	18.92	[18.82, 19.02]
4	54.78	[54.14, 55.42]	16.91	[16.54, 17.28]	16.91	[16.8, 17.03]
5	52.76	[52.21, 53.31]	15.48	[15.13, 15.82]	15.62	[15.49, 15.76]
6	51.26	[50.41, 52.11]	14.80	[14.36, 15.24]	14.58	[14.42, 14.74]
7	49.64	[48.87, 50.41]	14.33	[13.93, 14.73]	13.55	[13.33, 13.76]
8	48.04	[47.32, 48.76]	13.22	[12.7, 13.75]	12.65	[12.43, 12.88]
9	47.00	[46.22, 47.78]	12.52	[11.88, 13.16]	11.27	[10.9, 11.64]
10	45.84	[45.13, 46.55]	11.41	[10.7, 12.12]	10.51	[10.14, 10.88]
11	44.04	[43.14, 44.94]	10.21	[9.34, 11.08]	8.93	[8.52, 9.35]
12	41.76	[40.76, 42.76]	8.70	[7.84, 9.57]	8.12	[7.76, 8.47]
13	40.82	[39.58, 42.06]	8.69	[7.89, 9.49]	7.33	[6.9, 7.76]
14	38.40	[37.15, 39.65]	7.08	[6.31, 7.86]	6.19	[5.87, 6.51]
15	35.60	[34.2, 37.0]	6.64	[5.87, 7.4]	5.30	[4.95, 5.65]
16	35.12	[33.74, 36.5]	5.86	[5.24, 6.47]	5.04	[4.71, 5.37]
17	32.90	[31.31, 34.49]	4.93	[4.3, 5.56]	4.32	[4.02, 4.63]
18	31.38	[30.08, 32.68]	4.39	[3.84, 4.95]	3.97	[3.7, 4.23]
19	29.62	[28.12, 31.12]	3.49	[3.02, 3.96]	3.61	[3.25, 3.97]

Table 1. Simulation results of statistics of interest with varying bus numbers

As shown in the figures below, all of the three quantities decrease as the number of buses increases. The 95% confidence interval for mean queue length and mean waiting time increases when the number of buses increased from 2 to 13 and decreases thereafter. This might be because the system is more stable for such a number of buses (not too high or too low a utilization rate).

Based on the empirical results, the most optimal number of buses in the system is suggested to be 15, because the graph shows that the mean waiting time decreases steepest until having 15 buses in the system and decreases more slowly after the threshold. It's also reasonable for passengers if they only need to wait for 5.30 minutes on average for the bus (with a 95% confidence interval of [4.95, 5.65], meaning that there's a 95% chance that the waiting time will fall into such interval at equilibrium. The maximum queue length at 15 buses is 35.12 with a 95% confidence interval of [34.20, 37.00], and the mean queue length is 6.64 with a 95% confidence interval of [5.87, 7.40], which are all reasonable. Therefore, considering the costs of adding buses to the system and the satisfaction of passengers, 15 is the optimal number for such a system.

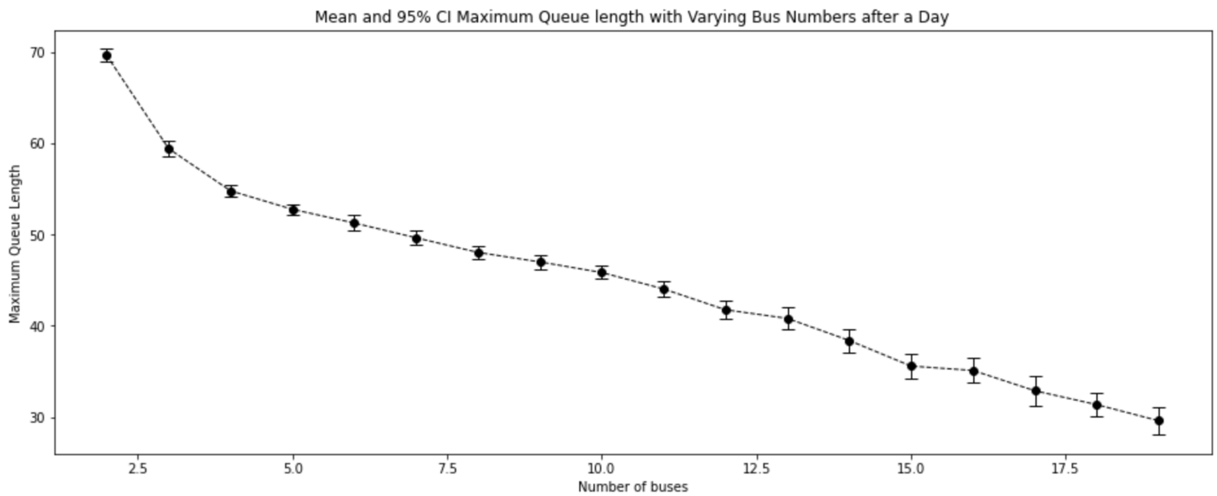

Figure 1. Mean and 95% confidence interval of maximum queue length with varying bus numbers after a day

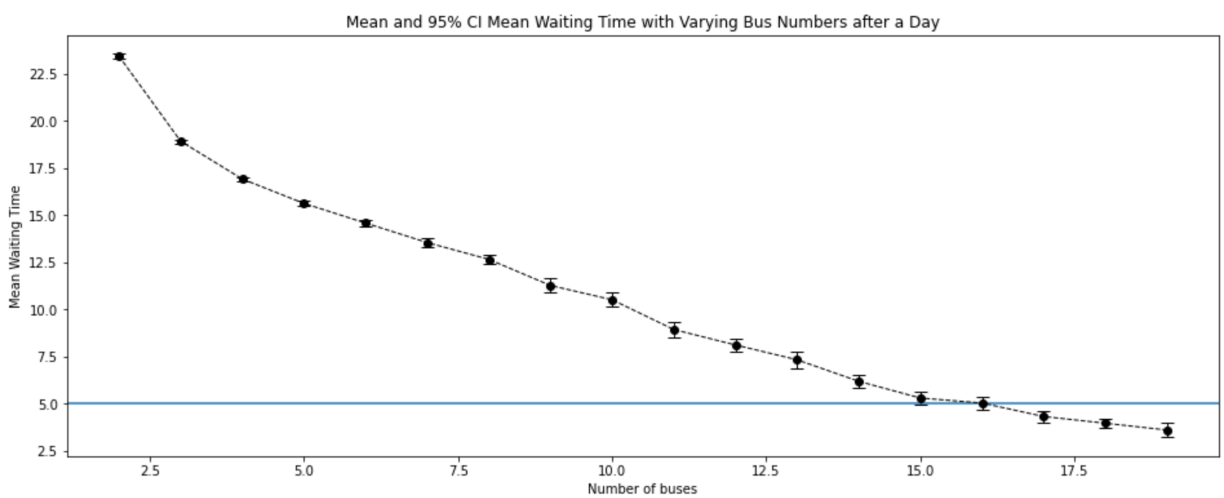

Figure 2. Mean and 95% confidence interval of mean waiting time with varying bus numbers after a day

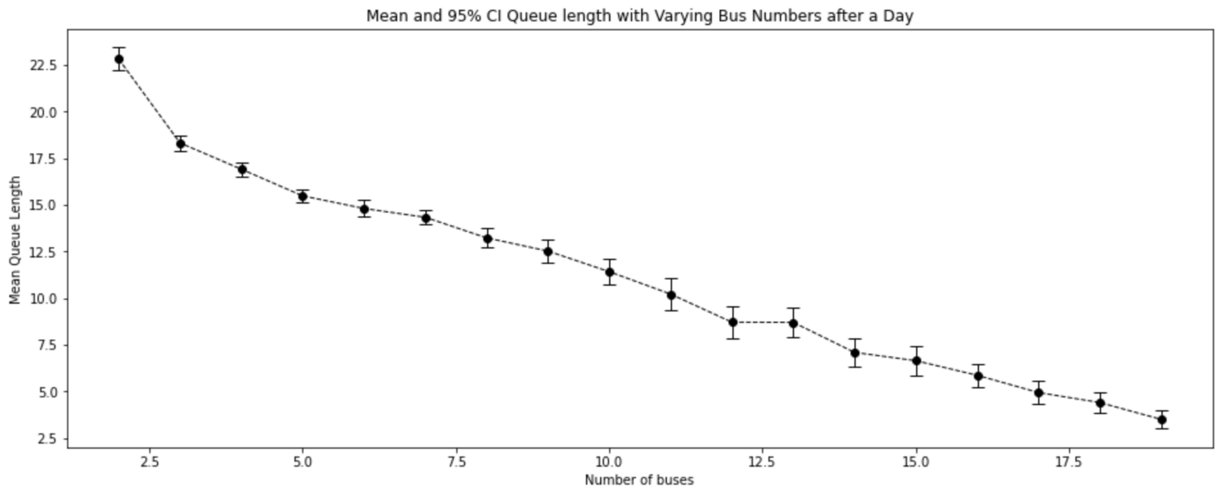

Figure 3. Mean and 95% confidence interval of mean queue length with varying bus numbers after a day

Theoretical analysis

Given that average service time of M/G/1 queues is τ with variance σ^2. Utilization is $\rho = \lambda\tau$.

After integration, the average waiting time at equilibrium should be $\frac{\rho\tau}{2(1-\rho)}(1 + \frac{\sigma^2}{\tau^2})$.

Average queue length should be $\lambda \times \textit{average waiting time} = \frac{\rho^2}{2(1-\rho)}(1 + \frac{\sigma^2}{\tau^2})$.

Because the travel time of buses between consecutive bus stops has a mean of 2 minutes and standard deviation of 0.5 minutes, and there are 15 bus stops in total. To simplify the comparison between theoretical result and empirical result, we would not consider embark and disembark time and the bus is never full. In this case, the total mean of the service distribution would be $\frac{15}{\textit{number of buses} * \textit{customers it serves}} \times 2$, so we assume $\tau = \frac{30}{\textit{number of buses} * 20}$, $\sigma = 3$, $\lambda = 1$, $\rho = \lambda\tau$. And average queue length and waiting time at equilibrium would be the same in this case, and as bus number increases, both of the statistics would decrease. To ensure that utilization is smaller than one, there should be at least two buses in the system.

The graph below shows a comparison between theoretical and empirical results. The have the same decreasing trend. Empirical result is higher than theoretical one because the embark and disembark time.

The assumption that the bus will serve 20 passengers at a time might also lead to the gap. But generally they agrees, showing that the simulation is producing predictable results.

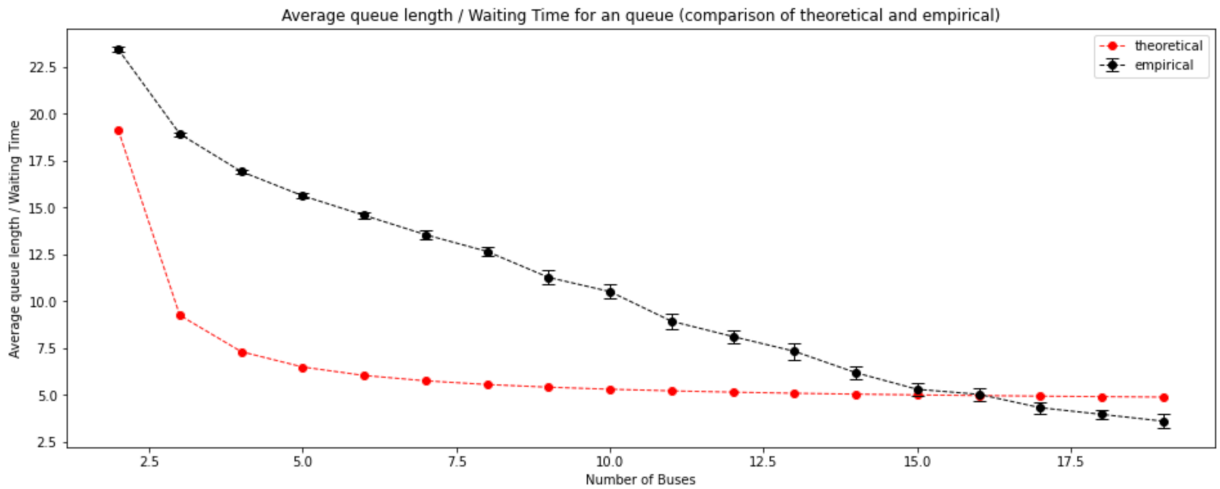

Figure 4. Comparison of theoretical and empirical result of average queue length and waiting time for queues after a day with varying number of buses

Critique of the model

Although the model correctly simulates the system behavior, we also acknowledge some drawback of the model that would decrease its representation of the reality. For example, the assumption of equal arrival rate throughout the day may be unrealistic because some periods of the day might be peak hours and there might be difference between bus stops. Also the time it takes for buses to travel from stops to stops might also be different because of difference in distances and speed. Therefore, extension of the model would consider those complex situations and make the simulation more accurate.