HO CHI MINH UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER ENGINEERING

SYSTEM PERFORMANCE EVALUATION
SEMESTER 201

INSTRUCTOR: TRAN VAN HOAI

# Project Report: Simulating queueing system by SimPy

| *Student* | *Student ID* |
|---|---|
| HO HOANG THIEN LONG | 1852161 |
| NGUYEN DUY TINH | 1852797 |
| DOAN ANH TIEN | 1852789 |

August 28, 2021

## Contents

# 1 Introduction

In this semester 201, we are conducting several projects involve Python coding environment as well as concepts such as queueing simulation, algorithms simulation,...

Our final project is to implement an environment in order to simulating the M/M/1 queueing system. Compared to the last project that using trend analyzation to look on the shortest path algoritms' performance, this time we will create the simulation, or specifically Discrete Event Simulation (DES) by using SimPy package.

This report contains the steps and paramaters which are neccessities in a performance evaluating process:

- Goal and system definition
- Services
- Performance metrics
- System parameters
- Workload parameters
- Factors and their ranges
- Evaluation technique

# 2 Project Scope

## 2.1 System definition and Goals

### 2.1.1 System definition

With the current status of industries, data processing is one of the inevitable needs of almost all systems from the most basic to the most complex; The processing must be based on pre-set criteria to suit the needs of users as well as to make the process more efficient.

Open queueing network is a system modelled by networked queues in which a job departing from one queue arrives at another queue. Open queueing network has external arrivals and departures and possesses the following characteristics:

- Number of jobs in system varies with time.
- Assumed that throughput equals to arrival time.
- The analysis goal of this queueing network is to characterize the distribution of number of jobs in the system.

(or possibly the same queue) Jackson networked queueing is one of the data processing methods widely used in the service industries and e-commerce. This type of queueing has external arrivals and departures (Open queueing network), which satisfies that:

- Any external arrivals to node i form a Poisson process.
- The service time is exponentially distributed and the service discipline is First Come First Served (FCFS).
- A job after finished at queue i, either moves to queue j with probability $p_{ij}$, or leaves the system with probability

$$1 - \sum_{j=1}^{m} p_{ij}$$

- Utilization of all queues < 1.

Before heading to main features of the project, we will look again at the requirements of the given project:

A network of 3 queues Q1 = $M/M_{(\mu 1)}/1$, Q2 = $M/M_{(\mu 2)}/1$, Q3 = $M/M_{(\mu 3)}/1$, in which $Q_1 \to Q_2$ with $p_{12} = 0.7$, $Q_1 \to Q_3$ with $p_{13} = 0.3$, arrival process to $Q_1$ with $\lambda$, jobs after going through $Q_2$, $Q_3$ will leave the network.

### 2.1.2 Boundary

This system follow some assumptions:

- Service rate (bps) and packet size (s), is fixed at some certain exponetial distributions.
- Probability of a packet going out from switch port 1 get into queue 2 or queue 3.
- All cases are simulated on a specific version of SimPy and its environment.
- All parameters are simulated on a specific version of random library.

### 2.1.3 Goals

In order to understand how a network works, we need a tool that can simulate the entire process of the system and also record the necessary metrics for evaluation. This model should be detailed, objective and highly symbolic.

To reach that, SimPy - a Python package provides Processes to model active components such as messages, customers, trucks, and planes. It has three classes to model facilities where congestion might occur: Resources for ordinary queues, Levels for the supply of quantities of material, and Stores for collections of individual items.

Besides, we also used SimComponent [1] - a pre-compiled source-code used to describe an object-oriented system; they include objects such as PacketSink, Packet, PacketGenerator, SwitchPort, ... along with its associated properties and methods.

In our experimental design, we want to observe and analyze how the average waiting time of a packet in process and the occupancy of queues change respect to the change of arrival time.

### 2.1.4 Scenario - Queueing network

In this report, three specific occurrences about the correspondence between arrival rate and the service rate of $Q_1$ is evaluated:

- Arrival rate is larger than service rate of $Q_1$
- Arrival rate is smaller than service rate of $Q_1$
- Arrival rate equals to service rate of $Q_1$

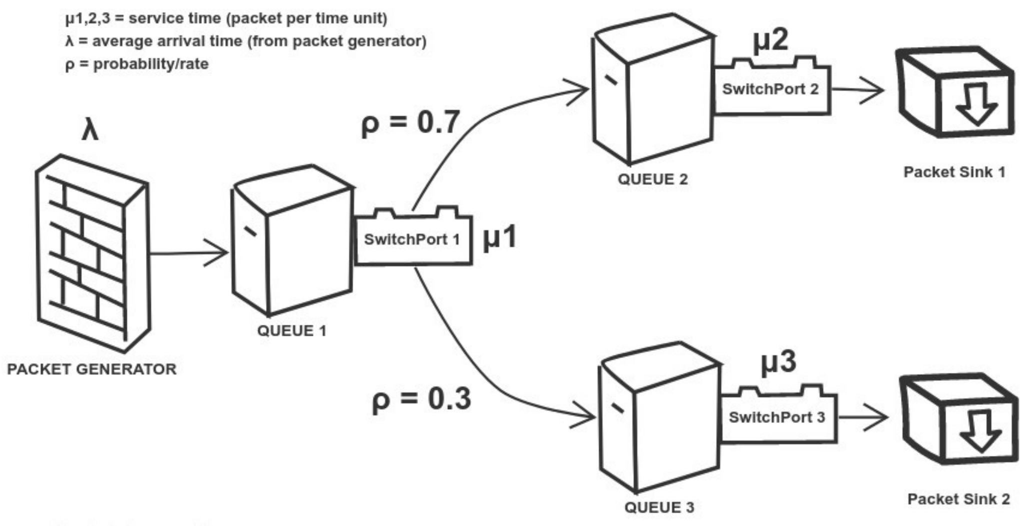

Figure 1: Queueing network system evaluated by the group

The components of our simulator design contains:

1. **Packet Generator:** A Packet Generator simulates the sending of packets with a specified inter-arrival time distribution (second) and a specified packet size (byte).

2. **Switch Port:** Models a switch output port ontaining a queue with a processor that process packets waiting in the queue with a service rate (bits per second).

3. **Packet Sink:** Receives packets and collects delay information into the waits[] list.

4. **Port Monitor:** A monitor for an SwitchPort. Looks at the number of items in the SwitchPort in service + in the queue and records that info in the sizes[] list for every a specific time interval.

## 2.2 Services

Overall, this model provides a simulation of the process sending packets through Switch Ports (including queues and transition handlers) based on a certain criterion, and finally transferred to the Packet Sink. By calculating quantities such as time, packets currently in the queue, etc. the model provides user processing metrics (from given input metrics) to evaluate the efficiency of the simulated system.

## 2.3 Performance metrics

To evaluate the system's efficiency, there are some parameters related to the quickness, low consumption and stablity of the system. They are:

- The waiting time of 1 packet spend in the system (represents the promptitude of the system): the time that a packet has to wait to be processed (i.e. the total time a packet has lost since it arrived until it reached the Packet Sink). It is easy to see that the total waiting time of all the packets is also the process delay, which reduces the performance of the system.
- Number of packets occupied in the system (represents the low space consumption of the system): this stands for the number of packets currently contained in a whole system. The congestion of the system should be considered because of the memory's possession of the packets, old packets should be freed before a new packet is inserted. The concurrent existence of a numerous packets makes more memory occupied, thereby risking overflowing or slowing down the system.

## 2.4 System parameters

System parameters are constant values associated with the properties of the environment that is currently used. In this problem, there is only one system parameter - randomness of expovariate function. In simpy environment, we considered $\rho$ as the probability of wheather a packet go to queue 2 or 3, after waiting in queue 1. This probability variable need a randomness function which affects the model by its algorithm.

### 2.5 Workload parameters

The workload parameters are those that describe the load imposed on the computer system by the jobs or transactions submitted to it. They are:

- Packet size: The size of each packet that is initially fixed (bytes). Mean packet size take on value [100] (byte).
- Arrival time: This is the time at which the transactions arrive into the system for service. This is expressed in second, minute, etc., depending on the nature of transactions. It is denoted by $1/\lambda$ (second). Mean arrival time takes on value [0.1, 0.25, 2] (s).
- Service rate: The fixed value that defines the number of bits that switch port can handle every second (bps). Mean service rate of [SwitchPort1, SwitchPort2, SwitchPort3] takes on value [4*8*(mean packet rate), 2*8*(mean packet rate), 3*8*(mean packet rate)], representing [4, 2, 3] (packets/s).

All parameters mentioned are exponential distributed.

### 2.6 Factors

In this problem the factors are:

- Arrival time (s).
- Randomness of expovariate function $\rho$.

### 2.7 Evaluation technique

The main technique throughout our research and implementation is simulation - specificially Discrete Event Simulation. We also used the Initial data deletion as a tool to proccess and compute average value of data.

## 3 Simulation

### 3.1 Implementation

For executing Python code and efficient co-working, we used Google Colaboratory (also known as Colab) - a free Jupyter notebook environment that runs in the cloud and stores its notebooks on Google Drive.

The source code executed on Google Colaboratory: https://bit.ly/3rhEetx

Figure 2: Google Colaboratory

The notebook contains 7 main blocks:

- Libraries installing block: Install nescessary libraries such as simPy, simComponets for simulation, matplotlib for plotting data.
- Parameters block: Declare essential parameters that affects the properties of the queueing system.
- System block: describe the queueing system in simPy simulation environment
- Initial data deletion block: contain the algorithms of intitial data deletion used for finding steady state, delete transient state of the system. Moreover, there is a function helps to compute variance of the data collected.
- Data plot block: is used to plot data.
- Main block: contains the functions that run the simulation, collect, process data as well as plotting.

The simulations runs for 100(s). The replications for initial data deletion method is 10. Replications mean how many times running the system independently.

## 3.2 Results of the experiments

### 3.2.1 Occupancy

#### Arrival rate is larger than service rate of $Q_1$

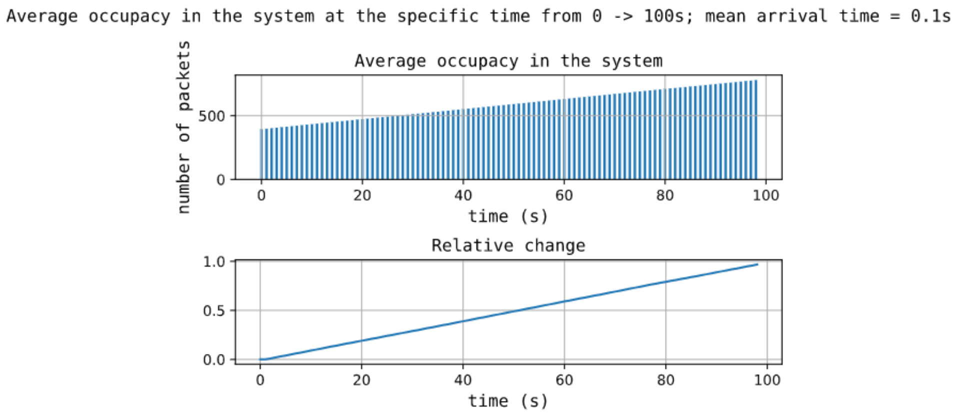

Figure 3: The average total occupancy with $1/\lambda = 0.1$s

Variance of average total occupancy: 13061.584775129495.
Variance of relative change: 0.08069456505542849.

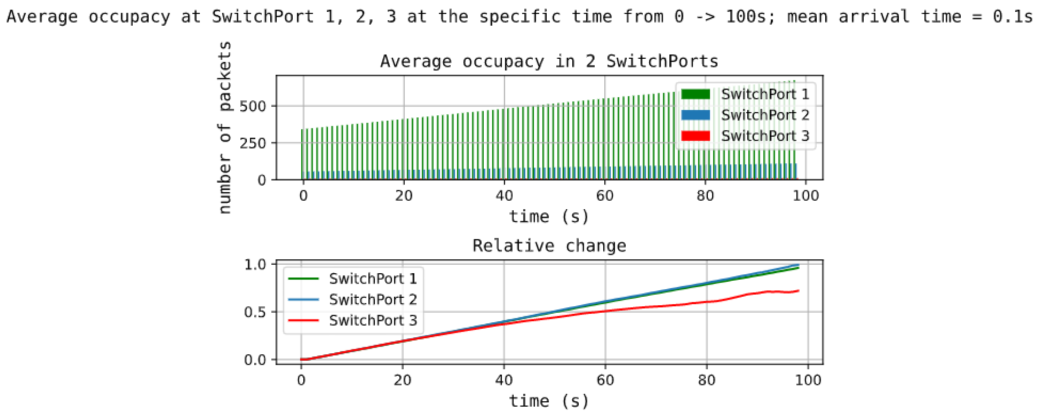

Figure 4: The average total occupancy at each switch port with $1/\lambda = 0.1$s

Variance of average occupancy in SwitchPort 1: 9387.392913267378
Variance of average occupancy in SwitchPort 2: 255.96471361835913
Variance of average occupancy in SwitchPort 3: 1.7477440294886335
Variance of relative change of SwitchPort 1: 0.08057048231486202
Variance of relative change of SwitchPort 2: 0.0848886578017927
Variance of relative change of SwitchPort 3: 0.04396670465355961

**Arrival rate equals to service rate of $Q_1$**

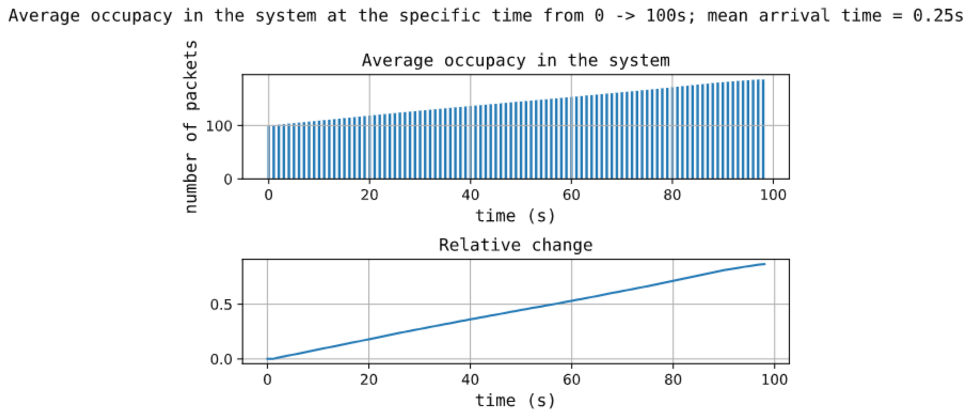

Figure 5: The average total occupancy with $1/\lambda = 0.25$s

Variance of average total occupancy: 645.3603894626448.
Variance of relative change: 0.06514776896000957.

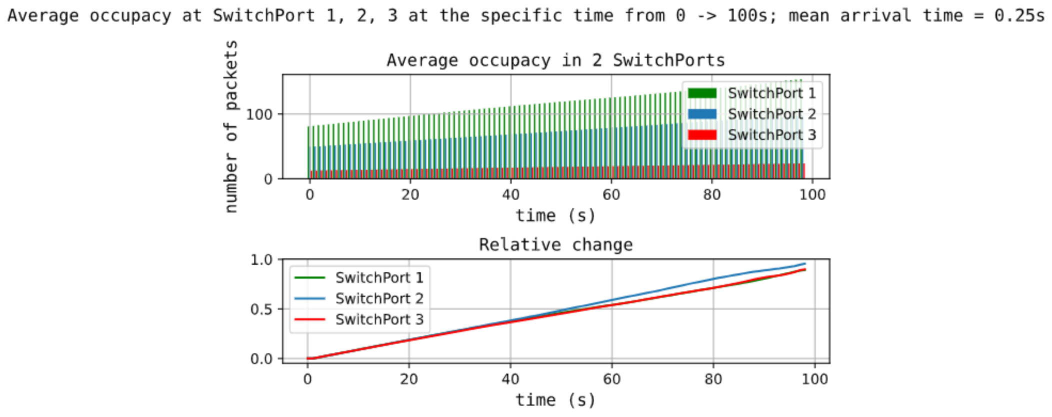

Figure 6: The average total occupancy at each switch port with $1/\lambda = 0.25$s

Variance of average occupancy in SwitchPort 1: 423.4453422445867
Variance of average occupancy in SwitchPort 2: 198.29628234224367
Variance of average occupancy in SwitchPort 3: 10.268109730741413
Variance of relative change of SwitchPort 1: 0.06487285319232057
Variance of relative change of SwitchPort 2: 0.0821783232897282
Variance of relative change of SwitchPort 3: 0.06613565243836367

**Arrival rate is smaller than service rate of $Q_1$**

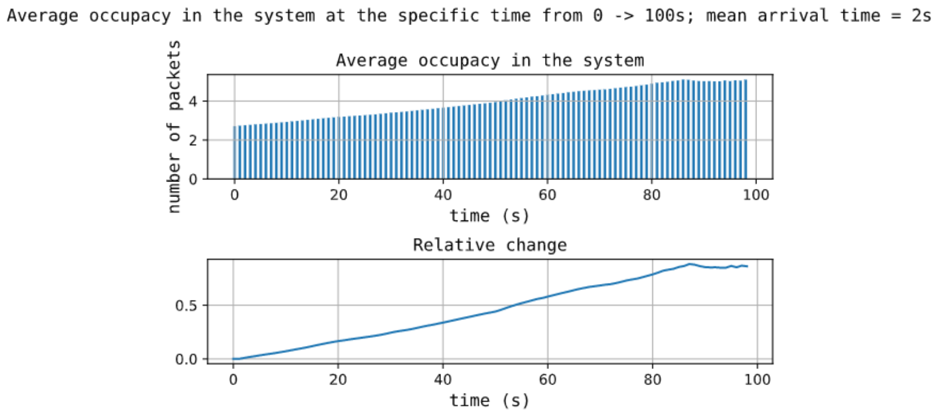

Figure 7: The average total occupancy with $1/\lambda$ = 2s

Variance of average total occupancy: 0.5974526698899131.
Variance of relative change: 0.08153975760027386.

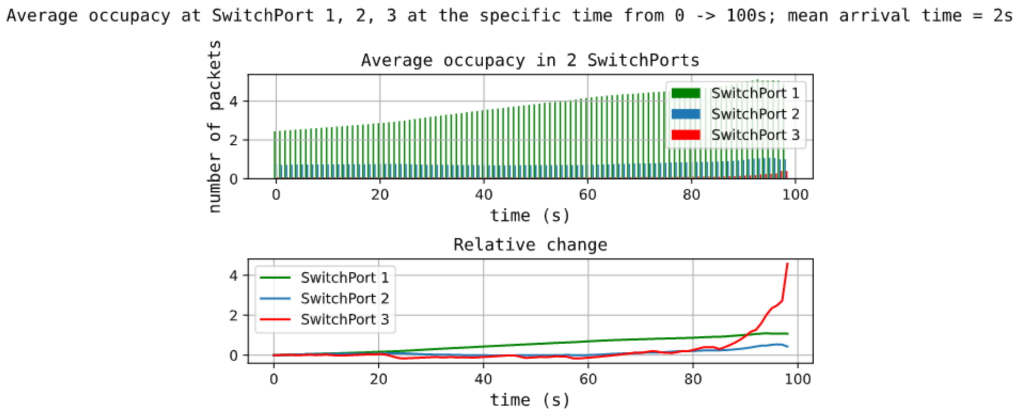

Figure 8: The average total occupancy at each switch port with $1/\lambda$ = 2s

Variance of average occupacy in SwitchPort 1: 0.6685281211693613
Variance of average occupacy in SwitchPort 2: 0.009672983137694937
Variance of average occupacy in SwitchPort 3: 0.003510952092998149
Variance of relative change of SwitchPort 1: 0.11310471047306869
Variance of relative change of SwitchPort 2: 0.018851770938791274
Variance of relative change of SwitchPort 3: 0.49304427682752594

### 3.2.2 Waiting Time

**Arrival rate is larger than service rate of $Q_1$**

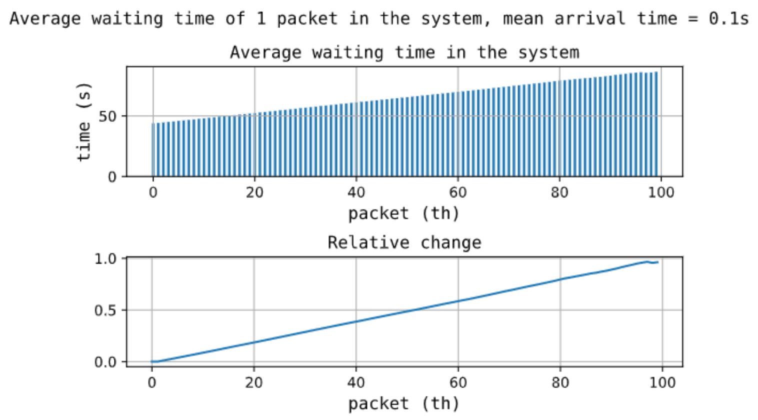

Figure 9: The average total occupancy with $1/\lambda = 0.1$s

Variance of average waiting time in the system: 161.02378759312344.
Variance of relative change: 0.0841039725368625.

**Arrival rate equals to service rate of $Q_1$**

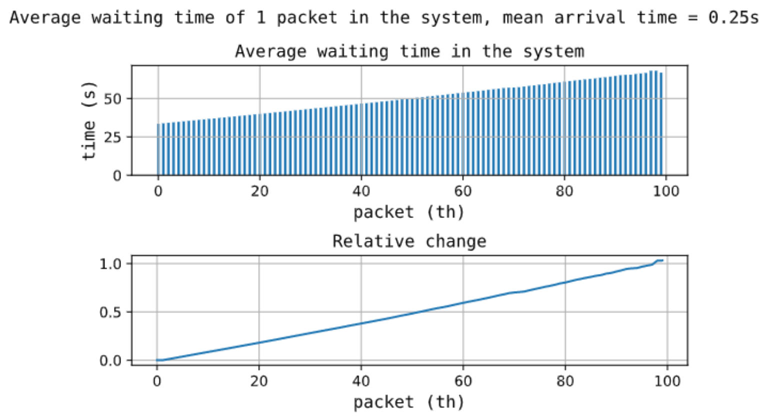

Figure 10: The average total occupancy with $1/\lambda = 0.25$s

Variance of average waiting time in the system: 101.3627430015386.
Variance of relative change: 0.09016635555580145.

**Arrival rate is smaller than service rate of $Q_1$**

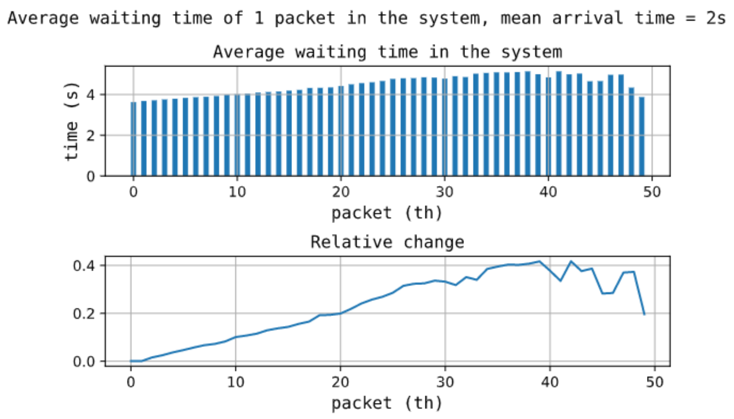

Figure 11: The average total occupancy with $1/\lambda = 2$s

Variance of average waiting time in the system: 0.22320267405506974.
Variance of relative change: 0.017528660845246263.

## 3.3 Analysis of results

The results above shows information about how many packets in average are currently occupied in the system, as well as in the 3 Switch Ports at the specific time and how long in average the packet spends in the system. Each case use 10 independent replications to compute mean value and is processed using initial data deletion method.

Overall, it can be seen that the mean occupacy of this system as well as for each Switch Prot increase with time for the case Arrival rate is larger than or equal the service rate of Q1, and it cannot reach steady state for the first 100s. This trends can also be observed in the first two cases of waiting times data, whose latter packet will stay in the queue longer than the former packet. The "going-in" rate is larger then the "going-out" rate.

The occupacy in Switch Port 1 > Switch Port 2 > Switch Port 3 for all cases.

The occupacy in the case Arrival rate equals to service rate of Q1 is approxilately 2.5 times smaller then the occupacy in the case Arrival rate bigger than service rate of Q1 for each second. The similar trend can be seen in the result of waiting times, in which the average waiting time in the case Arrival rate equals to service rate of Q1 is nearly 1.5 times smaller than the average waiting time in the case Arrival rate bigger than service rate of Q1.

In the case Arrival rate is smaller than service rate of Q1 sees a steady state in occupacy as well as waiting time at the second 87 the packet $38^{TH}$, respectively. Particularly, the steady state of occupacy goes around 4 packets, and the steady state of waiting time goes around 4 seconds.

# 4 Conclusion

We have built a simulation environment and the queueing system using SimPy and ran it for 100s to observe and analyze how the average waiting time of a packet in process and the occupancy of queues change respect to the change of arrival time. After doing some experiments respect to 3 cases and using initial data deletion method, we can see the trends of the system in 3 cases and get the average results, find steady of the system in the case arrival rate is smaller than service rate of Q1.

## References

[1] Greg Bernstein. *Basic Network Simulations and Beyond in Python*. URL: https://www.grotto-networking.com/DiscreteEventPython.html.