Project Report – Faculty bar

Filippo Guggino, Riccardo Mancini, Essaid Zaccaria

January 14, 2020

Contents

1 Introduction

1.1 Problem description

Consider a system that manages the orders of customers in a bar. Orders are issued to a cashier, which has two queues: one for normal users and one for VIP users. VIP users have non-preemptive priority over normal users. Users belonging to the same queue are served according to a FIFO policy. Orders can be of simple or compound type. Simple orders are completed directly by the cashier, with a service rate $r_{cashier}$. Composite orders are first served by the cashier, with a service rate $r_{cashier}$, and then are passed to the kitchen, which queues them and serves them in FIFO order, with a service rate $r_{kitchen}$: Both normal and VIP users can issue both types of orders.

1.2 Objectives

In this analysis, the following objectives will be investigated:

- relationship between the service times and the overall experienced response times.
- advantages, in terms of response time, of being a VIP customer.
- advantages and disadvantages of introducing priority head-of-line queuing also in the kitchen.
- optimal value for the ratio between VIP and normal customers in order to achieve the best overall quality of service.
- demonstrate that queue lengths don't depend on the ratio between the service rates.

1.3 Performance indexes

In order to fulfill the above-stated objectives, the following performance indexes will be taken into consideration:

- response times for all four categories of users (normal/VIP, simple/compound order), in particular their mean and 90^{th} percentile.
- average queuing time for all queues.
- average queue length.
- advantage of VIP customers over normal customers as a ratio of their respective mean response times: $1 - \frac{E[R_{VIP}]}{E[R_{normal}]}$

1.4 Scenarios

The following scenarios will be taken into consideration:

- constant inter-arrival times, constant service times (only for code verification since in this scenario no queuing is possible and thus it is of no interest);
- exponential distribution of inter-arrival and service times;
- "business day": exponential distribution of inter-arrival and service times with varying average of the former at an hour granularity.

2 Model

2.1 Description

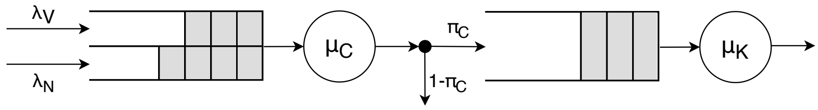


Figure 1: Schematic representation of the model of the bar.

In Figure 1 you can see the model of the bar, where:

- λ_V is the average arrival rate for VIP customers;
- λ_N is the average arrival rate for "normal" (i.e. non-VIP) customers;
- μ_C is the average service rate of the cashier ($r_{cashier}$);
- μ_K is the average service rate of the kitchen ($r_{kitchen}$);
- π_C is the ratio of compound orders over the total, i.e. the probability that an order is compound.

The cashier is modeled as a service center with two queues that are managed in an head-of-line-priority fashion.

The kitchen is modeled as a simple M/M/1 service center. Note that if VIP priority is introduced also in the kitchen, it will become identical to the cashier, mutatis mutandis.

2.2 Assumptions

The following assumptions were made when modeling the system:

- No renegation: customers cannot leave the queue.
- No jockeying: VIP customers cannot move to the normal customers' queue.
- No overtaking: customers cannot change their position in queue.
- Infinite queuing space: there is no upper bound in the number of customers in a queue.
- the inter-arrival times of VIP and normal customers are independent RVs.
- the type of an order (simple or compound) is independent of the other orders.
- the service rate is the same for normal and VIP customers.

2.3 Factors

The factors that can affect the modeled system are:

- cashier service rate: μ_C
- kitchen service rate: μ_K
- ratio of compound orders over the total: π_C
- average inter-arrival rate of VIP orders: λ_V
- average inter-arrival rate of Normal orders: λ_N
- Kitchen queue type: "FIFO" or "priority"

2.4 Stochastic model for the exponential scenario

In the exponential scenario, we can formulate a stochastic model of the system using queuing theory.

The cashier priority queuing is, in fact, a well known queuing strategy (L. Kleinrock, 1976)[1]. In the case of two static priorities, the following equations for the average waiting time hold:

$$E[W_V^C] = \frac{\lambda_V + \lambda_N}{(\mu_C - \lambda_V)\mu_C} \tag{1}$$

$$E[W_N^C] = \frac{\lambda_V + \lambda_N}{(\mu_C - \lambda_V - \lambda_N)(\mu_C - \lambda_V)} \tag{2}$$

Note that, for $\lambda_V = 0$ (or $\lambda_N = 0$) the above equations are the same as for an M/M/1 system.

In order to obtain the average response times, we just need to add $\frac{1}{\mu_C}$ due to the linearity of the mean operator.

Furthermore, note that if we didn't make any distinction between orders and just saw them entering the cashier SC and exiting it, we would not be able to distinguish it from an M/M/1 SC with arrival rate $\lambda_V + \lambda_N$ and service rate μ_C just by looking at the distribution of the inter-departure times (we don't care about the order).

This result has been confirmed by the validation runs on the simulator. Please note that this would not have hold if service rate were different between normal and VIP customers.

Therefore, we can apply queueing network theory which tells us that the average arrival rate at the kitchen is the same as an M/M/1 SC with arrival rate equal to $\pi_C(\lambda_V + \lambda_N)$. Thus, the response time of the kitchen (independently of the customer type) is:

$$E[R^K] = \frac{1}{\mu_K - \pi_C(\lambda_V + \lambda_N)} \tag{3}$$

Please note that, if the kitchen had a priority queuing, we would not be able to follow the same reasoning above. However, simulation results suggest that the mean waiting time could be computed as in Equations (1) and (2) considering $\pi_C \lambda_N$ and $\pi_C \lambda_V$ as the arrival rates.

Finally, we can write the average response times for the four classes of orders:

$$E[R_{N,S}] = E[W_N^C] + \frac{1}{\mu_C} \tag{4}$$

$$E[R_{N,C}] = E[W_N^C] + \frac{1}{\mu_C} + E[R^K] \tag{5}$$

$$E[R_{V,S}] = E[W_V^C] + \frac{1}{\mu_C} \tag{6}$$

$$E[R_{V,C}] = E[W_V^C] + \frac{1}{\mu_C} + E[R^K] \tag{7}$$

[1] Leonard Kleinrock (1976). Head-of-the-Line Priorities. *Queueing systems, volume 2: Computer applications* (pp. 119-126). New York, Wiley.

3 Implementation

3.1 Code overview

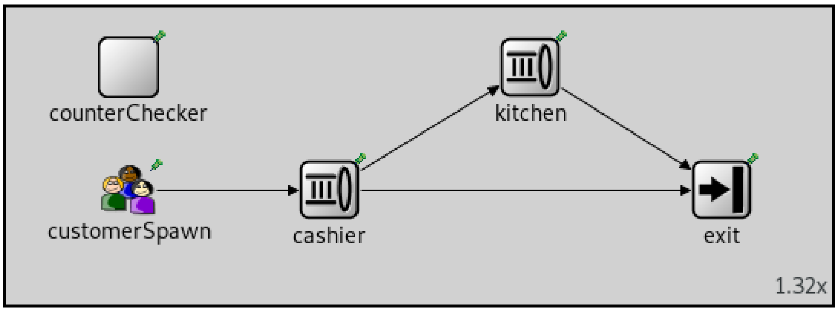


Figure 2: Diagram of the Omnet++ network.

The Omnet++ network is composed of 5 nodes:

- **customerSpawn**: generates the orders using a configurable distribution.
- **cashier**: serves orders and, upon completion, sends them to the *kitchen* if they are compound or to the *exit* otherwise. The type of order is determined in the *customerSpawn* node.
- **kitchen**: serves orders and sends them to the *kitchen* upon completion.
- **exit**: collects statistics about order response times and disposes of them.
- **counterChecker**: collects number of generated, exited and in-queue orders in order to check that no order is lost.

3.2 Verification

In order to verify that the implementation exactly reflects our model, we carried out the following tests.

Test 1: Single-type orders with constant inter-arrival times Only orders of a single type are generated with constant inter-arrival times in order to check that the correct path is followed by the orders and that statistics collection is working correctly. Note that in this case there is no queuing.

Test 2: Simple and compound normal orders with constant inter-arrival times Only normal orders are generated, both simple and compound, with constant inter-arrival times in order to check that the orders are correctly routed between kitchen and exit with the set probability.

Test 3: Normal and VIP simple orders with constant inter-arrival times Only simple orders are generated, both normal and VIP, with constant inter-arrival times in order to check that the priority queuing works correctly. In order to do that, two normal orders are sent at time t=1 and t=2 respectively, a VIP order is sent at time t=2.5 and the service time is 2. It is expected the last VIP order to be serviced second.

Test 3b: Normal and VIP compound orders with constant inter-arrival times Only compound orders are generated, both normal and VIP, with constant inter-arrival times in order to check that the priority queuing works correctly. In order to do that, two normal orders are sent at time t=1 and t=2 respectively, a VIP order is sent at time t=2.5 and the service time is

2. It is expected the last VIP order to be serviced second. The test is repeated for both FIFO and priority queue in the kitchen.

Test 4: Simple orders with exponential inter-arrival times Simple orders are generated with exponential inter-arrival times and the mean response time for each kind is compared with the mathematical model. Several scenarios are considered namely:

- total arrival rate is kept constant to 1 while the proportion of normal and VIP orders varies from 0% to 100%, with a step of 25%.
- cashier service rate is varied from 1.1 to 2 with a step of 0.1.

Test 4b: Compound orders with exponential inter-arrival times Compound and simple orders are generated with exponential inter-arrival times. Compound orders are 10% of the total, while simple orders are ignored (they are only generated in order to create a realistic scenario). The mean response time for each kind of compound orders is compared with the mathematical model. Several scenarios are considered namely:

- total arrival rate is kept constant to 1 while the proportion of normal and VIP orders varies from 0% to 100%, with a step of 25%.
- kitchen service rate is varied from 0.15 to 0.5 with a step of 0.05.
- cashier service rate is kept constant to 1.5.

The test is repeated for both FIFO and priority queue in the kitchen. In the case of the priority queue, the mathematical model could not be used so it was checked that:

- mean response time for normal orders is higher than VIP orders.
- mean waiting time for normal orders is higher than VIP orders.
- performance indexes are in a continuous relationship with the simulation factors.

Validation results All the validation tests listed above have been passed by the implementation. Tests 1 to 3/3b have been checked manually, while tests 4/4b have been visually checked using automatically-drawn plots of the performance indices against the cashier (or kitchen) service time (an example can be seen in Figure 3).

3.3 Calibration

To calibrate the system, the following factor ranges were defined:

- $\mu_C \in [1.5, 2.0]$
- $\mu_K \in [0.45, 0.6]$
- $\pi_C \in [0.1; 0.3]$
- $\lambda_V + \lambda_N \in [0.5, 1.5]$
- $\frac{\lambda_V}{\lambda_V+\lambda_N} \in [0; 1]$

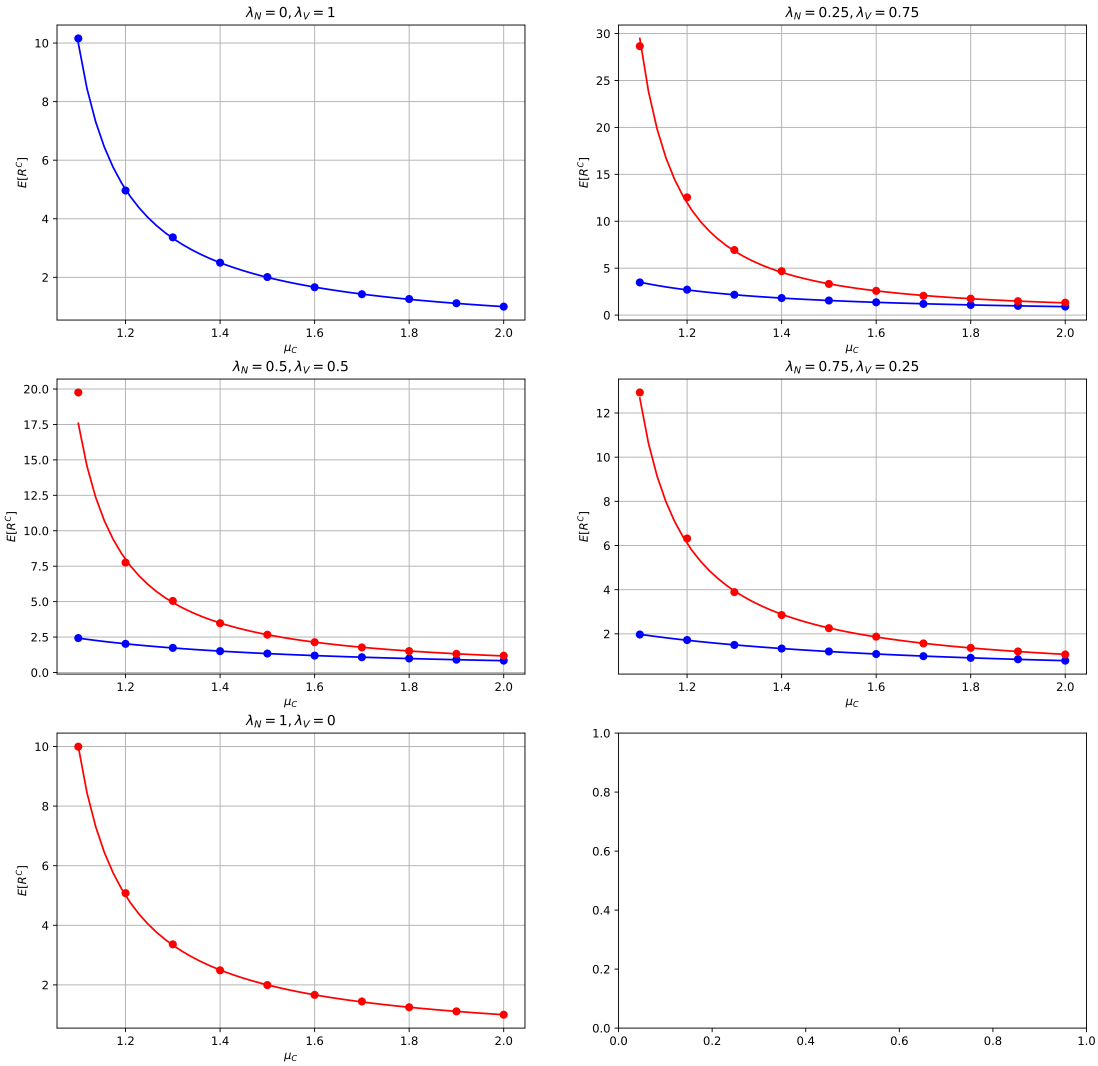
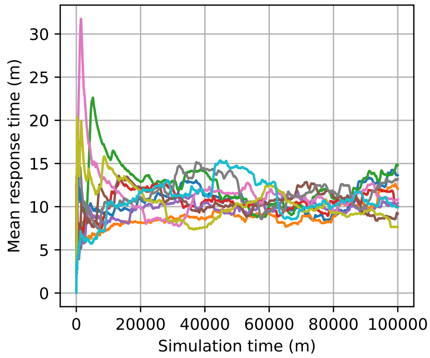


Figure 3: Verification of Test 4 showing the mean cashier response time for both VIP (in blue) and normal (in red) customers. The points represent the computed metrics, while the lines are the reference mathematical model.

4 Experiments

4.1 Setting warm-up and simulation times

In order to define the required warm-up period, we plotted the moving average of the response time as a function of the simulation time for 10 different runs. The runs were configured in order to represent a worst-case scenario: $\lambda_V = 1.3, \lambda_N = 0.1, \pi_C = 0.1, \mu_K = 0.2, \mu_C = 1.5$. The resulting plot can be seen in the picture on the right, where only simple normal orders are shown. From the plot we can see that the mean starts converging at around 30000m therefore we chose 50000m in order to have some safety margin. We can afford this since running the simulations is quite inexpensive.



After having chosen the warm-up period, the simulation time was chosen as to have both a high level of accuracy and a small execution time. The value of 500000m has hence been selected since it provides a confidence interval around 2% with 90% confidence in the above worst-case scenario, while still running in a reasonable time ($\sim 5s$ per run). In practice, the confidence intervals we obtain in the following analysis are much narrower.

4.2 Steady-state analyses

4.2.1 2kr analysis

In order to grasp the factors contribution to the customers experience, we computed a number of 2^kr (r = 5) factorial analysis. In these analyses, we took into consideration the following factors in the FIFO kitchen and exponential service and inter-arrival rates scenario:

- A = normal customers rate (VIP rate is set as to keep the total rate constant) $[0.1, 1.2]$
- B = odds of an order being a Compound one $[0.1, 0.3]$
- C = kitchen Rate $[0.45, 0.6]$
- D = cashier Rate $[1.5, 2]$

Please note that A is, in reality, a measure of the number customers over the total, ranging from 10% to 90%.

After having run the analyses, we visually checked the residuals' hypotheses for every metric through the related QQ-plot, scatter and lag plots. In the case of queue lengths, few tweaks needed to be made since the residuals QQ-plot was showing lighter tails. The solution was to run the analysis on the log of the queue length. This highlights the non-linear relationship of the queue length with the other factors.

After analyzing all of the metrics, we found no strange behaviour. Therefore, for the sake of brevity, we will only report the most interesting results we found:

- the cashier rate has a negative impact (-0.159 ± 0.01, with an explained 92% variability) on the advantage of VIP customers on normal customers since an increase of it implies lower queuing and thus a more "equal" waiting time between VIP and normal customers.
- parameter A has a negative contribution on the advantage of VIP customers on normal customers (-0.0463 ± 0.0008, with an explained 7.82% variability). This is somehow counter-intuitive since we expected the VIP advantage to increase when fewer VIP customers were present. However, this results highlights the phenomenon of "quasi-starvation" happening to normal customer with a very high number of VIP customers. In fact, on the one hand,

if VIP customers are the great majority, normal customers experience huge waiting times due to too many VIP arrivals jumping in queue in front of them. On the other hand, if normal customers are the great majority: VIP customers are serviced very fast, but normal user are no more "starved" by VIPs. This phenomenon is further investigated in Section 4.2.4.

- the cashier rate has a negligible influence on the kitchen queue length. This phenomenon will be further investigated in Section 4.2.5.

4.2.2 Kitchen Queue Comparison

In this section we want to assess whether enforcing a priority queue also in the kitchen brings any perks. In order to get that results we observe the trajectories of the *compoundResponseTimeRatio* statistic (i.e. the advantage of being a VIP customer over a normal customer when the order is compound) in both FIFO and priority queue cases, varying the ratio of compound orders.

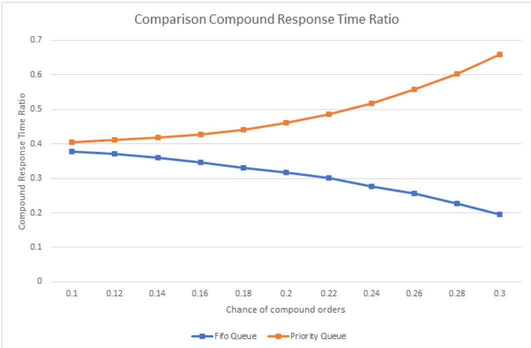


Figure 4: VIP advantage over normal for both FIFO and priority queue at the kitchen at different π_C ($\lambda_N = 1, \lambda_C = 0.2, \pi_C = 0.1..0.3$ step $0.02, \mu_K = 0.45, \mu_C = 1.5$).

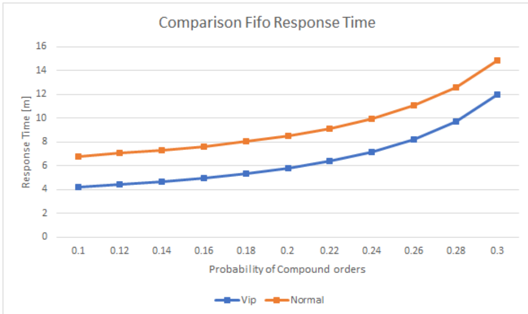


Figure 5: Response time for VIP and normal customers with FIFO queuing at the kitchen at different π_C ($\lambda_N = 1, \lambda_C = 0.2, \pi_C = 0.1..0.3$ step $0.02, \mu_K = 0.45, \mu_C = 1.5$).

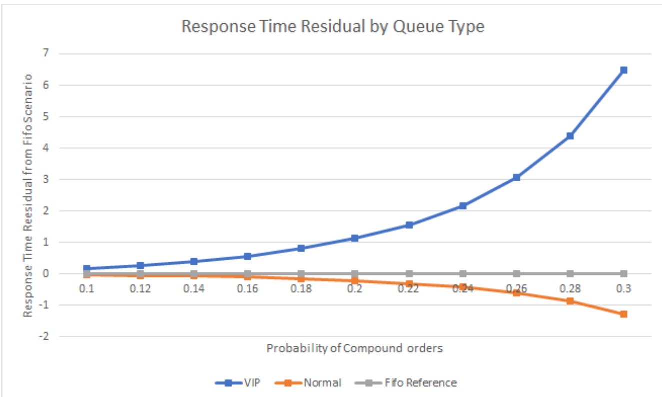


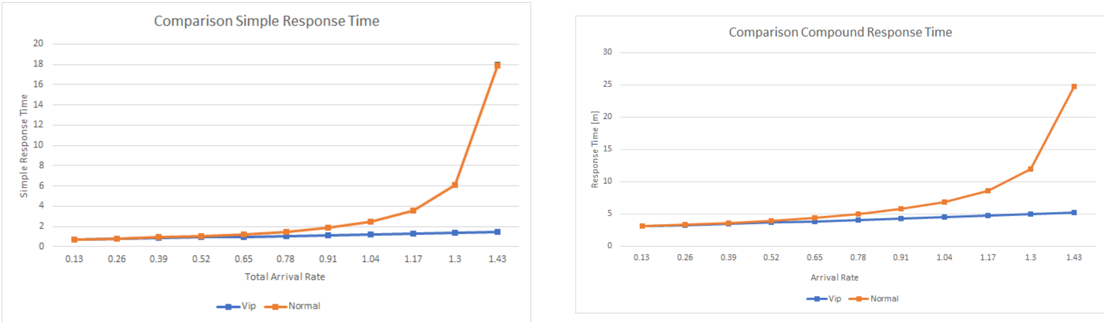
Figure 6: Residual Response time for Normal and VIP users from what they should expect from a FIFO Scenario in the kitchen($\lambda_N = 1, \lambda_V = 0.2, \pi_C = 0.1..0.3$ step $0.02, \mu_K = 0.45, \mu_C = 1.5$).

In Figure 4, we can clearly see that, in the FIFO case, by increasing the number of compound orders, VIP users tend to lose the "advantage" obtained during the cashier service. So, indeed, by using a priority queue in kitchen we can provide an even more privileged service.

Figure 5 highlights even more this phenomenon: in the FIFO scenario, the percentage benefit is reduced since the difference is constant (e.g., around 2m in this scenario), while the response times increase due to higher utilization.

Finally, we would like to be certain that introducing priority queuing at the kitchen does not affect the normal customers' response times too much. This is done in Figure 6, where we plotted the difference between recorded response times from FIFO and priority, so that to quantify the time advantage of introducing the priority queuing. From the figure, you can see how introducing priority queuing can bring to great time savings for the VIPs, while not hurting the normal customers too much (e.g., at 26% compound orders, the VIP mean response time is reduced by 3 minutes while the normal users experience just an half a minute difference).

4.2.3 System Response to different Workloads



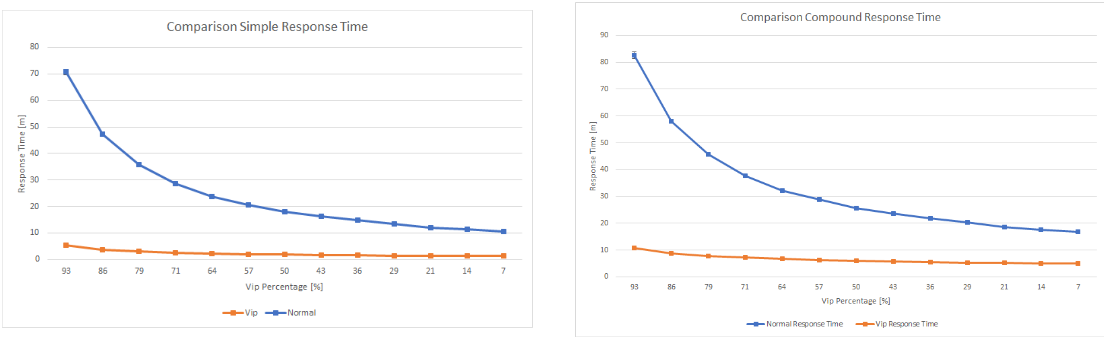
(a) (b)

Figure 7: Mean response times for both simple (a) and compound (b) orders of both VIP (blue) and normal (orange) customers by varying total arrival rate. ($\lambda_N = 0.1..1.1$ step $0.1, \lambda_V = 0.3 * \lambda_N, \pi_C = 0.2, \mu_K = 0.45, \mu_C = 1.5$). Note that the Inter-Arrival rate for the kitchen with this configuration is actually 0.2 * "value showed in the x axis".

In Figure 7, the response time for all four kind of orders are shown at different arrival rates. An interesting, but also expected from QT, aspect is that after a certain threshold of Arrival Rate, that is around 1.3, the Normal response time, for both Simple and Compound orders, explodes.

4.2.4 VIP Rate study

One thing that should be taken into consideration when fine-tuning the system is the amount of allowed VIP customers. Of course if all customers were VIP, they would gain little to no benefit and chances are that the few normal customers that arrive will experience a vary bad service. Therefore, it is interesting to study the response of the system at varying percentages of VIP customers in order to find an "optimal" value, i.e. one such that VIP benefits are preserved and normal customers are served in a reasonable time. Such study was carried out with a full factorial analysis, varying the percentage of VIP customers.



(a) (b)

Figure 8: Both VIP and Normal Response Time at the cashier (a) and kitchen (b) ($\lambda_N = 0.1..1.3$ step $0.1, \lambda_V = 1.4 - \lambda_N, \pi_C = 0.2, \mu_K = 0.45, \mu_C = 1.5$).

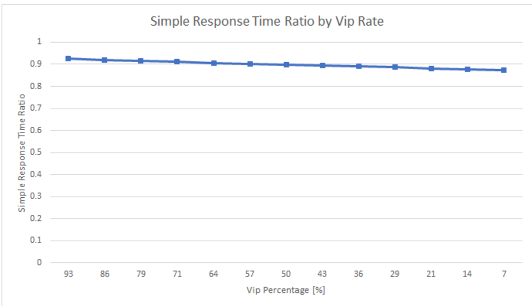


Figure 9: VIP advantage over Normal at the cashier($\lambda_N = 0.1..1.3$ step $0.1, \lambda_V = 1.4 - \lambda_N, \pi_C = 0.2, \mu_K = 0.45, \mu_C = 1.5$).

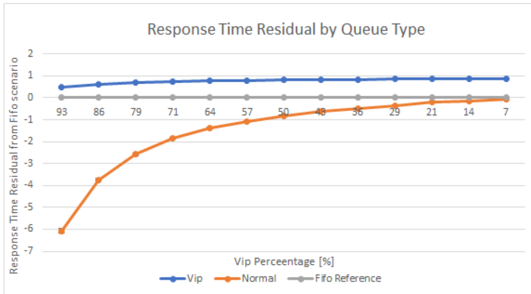


Figure 10: Residual Response time for Normal and VIP users from what they should expect from a FIFO Scenario ($\lambda_N = 0.1..1.3$ step $0.1, \lambda_V = 1.4 - \lambda_N, \pi_C = 0.2, \mu_K = 0.45, \mu_C = 1.5$).

In Figure 8, you can see how the response time changes for both VIP and normal customers, varying the percentage of the former over the total. From this plot we could choose the desired maximum number of VIP customers and see how the choice will influence on the average normal customer experience. For example, if we wanted the normal customers' mean response time to be below 15m, we could choose a percentage of VIP customers between 10% and 30%.

We can also plot the *simpleResponseTimeRatio* (Figure 9) with the same configuration and ensure that, in fact, VIP customers are at least 80% faster than Normal customers in that range. Furthermore, from the same plot, we can see how the advantage of VIP customers, expressed as a ratio, is fairly constant.

Finally, in Figure 10, we show how the service offered to customers improve (or not) compared to a FIFO scenario. This is done by computing the difference between the Responses Time got from the experiments and the theoretical FIFO response time ($\frac{1}{\mu_C - \lambda}$), i.e. the response time the customers would get if the cashier queue were FIFO. The plot clearly shows, once again, how the benefit of VIP customers does not change too much at different VIP customers ratios, while normal users are highly and negatively impacted.

4.2.5 Relationship between service rates proportion and queue lengths

In this section we will investigate the relationship between the proportion of the service rates and queue lengths. First of all, let's notice that the queue at the cashier is by no means influenced by the rate of the cashier, therefore we will just look at the behaviour of the kitchen queue in both FIFO and priority queuing strategies.

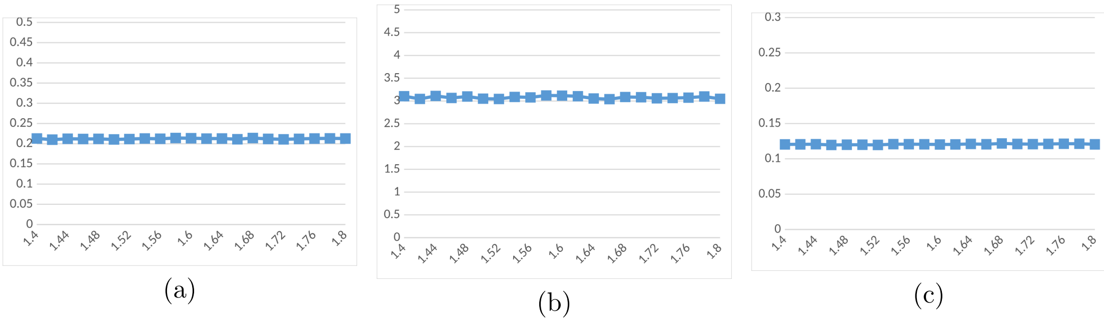


Figure 11: Kitchen queue length as a function of the cashier service rate in the FIFO (a) and priority (b: normal queue; c: VIP queue). $r = 30, \lambda_N = 1, \lambda_V = 0.2, \pi_C = 0.2, \mu_K = 0.3, \mu_C = \{1.4..1.8 \text{ step } 0.02\}$.

In Section 4.2.1 we showed how the kitchen mean queue length was not influenced by the cashier rate and Figure 11 just confirms the above result. Furthermore, recall that the same result was obtained also in the mathematical model for the FIFO case.

4.3 "Business day" analysis

After having studied the system response at the steady state, we decided to observe it also in a business day. We used the multipliers in Figure 12 to vary the arrival rate throughout the day so that to have peaks during meal times. In the following section, each scenario has been repeated 100 times and obtained confidence intervals are shown in the plots.

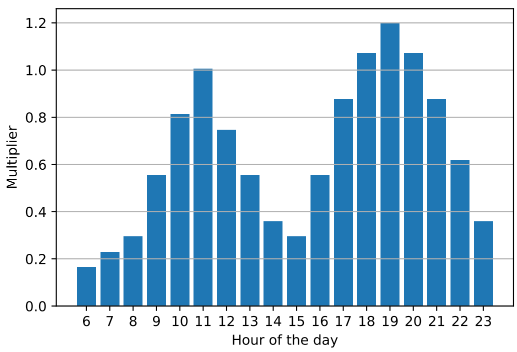


Figure 12: Multipliers used to modify the average arrival rate per business hour.

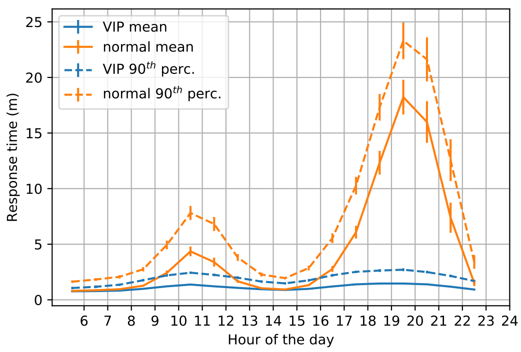


Figure 13: Comparison of the hourly average of the response time for VIP and normal customers with simple orders in a "normal" business day ($\mu_C = 1.5, \mu_K = 0.4, \lambda_{tot,max} = 1.65, \pi_V = 0.2, \pi_C = 0.2$).

Figure 13 shows a comparison between VIP and normal customers. As expected, the VIP customers are serviced very fast and experience almost no waiting time. Furthermore, note that the mean waiting time during dinner is much higher than lunch time since, with the used configuration, during the former the arrival rate exceeds the cashier service rate, conversely to what happens in the latter.

Figures 14 and 15 show the response of the system when varying, respectively, the cashier rate and the arrival rate. Both plots highlight that a small change ($\sim 10\%$) of either rate leads to an exponential change of the response time. Thus, even a small improvement in cashier speed during peak times could lead to huge improvements in customer satisfaction.

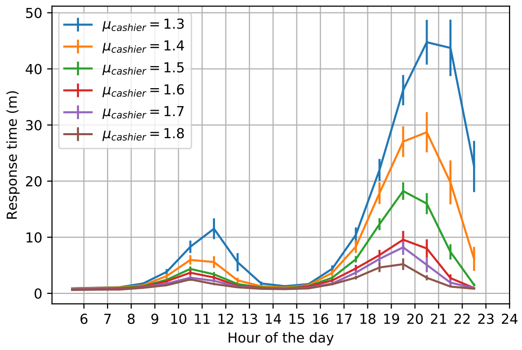


Figure 14: Hourly average of the response time for normal-simple orders in a business day at different cashier rates ($\mu_K = 0.4, \lambda_{tot,max} = 1.65, \pi_V = 0.2$).

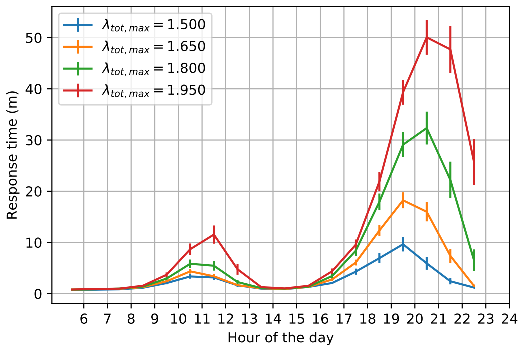


Figure 15: Hourly average of the response time for normal-simple orders in a business day at different arrival rates ($\mu_C = 1.5, \mu_K = 0.4, \pi_V = 0.2$).

5 Conclusions

Our study has shown how the customers' experienced waiting time is related to the factors of the systems. Obvious relationships have been confirmed, like those related to cashier and kitchen rate. Furthermore, we showed that, counter-intuitively, the mean queue length at the kitchen is not influenced by the speed of the cashier (in the considered exponential scenario). We also showed that the number of VIP customers over the total can negatively impact the satisfaction of normal customers.

The possibility of introducing priority queuing also in the kitchen has been explored. On the one hand, it obviously increases the benefits of the VIP user but, on the other hand, it reduces normal customers satisfaction. Nevertheless, we showed how this increased waiting time is negligible, therefore the suggestion is to adopt priority queuing also in the kitchen.

Finally, by studying a classic business day, we showed how small improvements in the cashier serving speed can have huge benefits on the overall customer experience.