POLITECNICO MILANO 1863

Traffic Theory
Priority Scheduling with General Services
Project Report

Abdullah Nasr
Bahram Hedayati

1. Introduction

This report presents the findings and analysis of the Priority Scheduling with General Services project, which aimed to investigate and implement a 3-class priority scheduling for a M/D/1 queueing system with preemption-resume using OMNeT++ simulation platform. Initially, we define the topic briefly, followed by presenting the theoretical findings. Subsequently, the simulation results will be demonstrated and a comparative analysis between the simulated outcomes and the theoretical findings will be conducted.

2. Theoretical Results

The queuing system, which is implemented in this project, categorizes customers into three distinct classes based on their priority levels. When an arrival from a higher priority class enters the system while the server is serving a customer from a lower priority class, preemption occurs. The ongoing service is interrupted, and the server switches to serve the higher priority customer. After serving the higher priority customer, the server resumes service for the interrupted lower priority customer. It is obvious that when an arrival from the priority class-1 is in the server, it will stay there until the end of its service time. The same queuing system is also implemented for the non-preemptive case in which each arrival will be served completely after going to server and no other arrivals cannot trigger ejecting it from the server. The non-preemptive case can be run by switching `**.*.preemption` flag in omnetpp.ini to false. Figure 1 illustrates a simple scheme of what we explained in this section which is implemented in the OMNeT++.

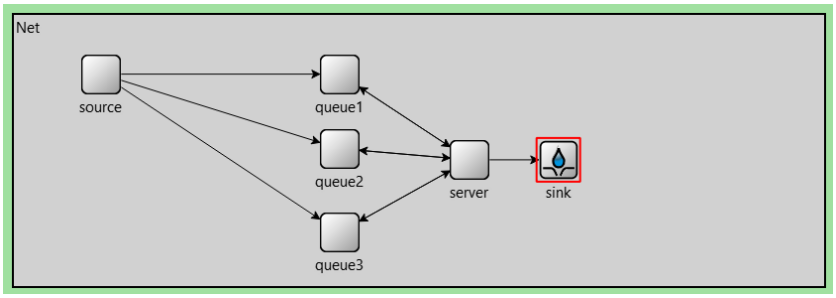

Figure 1: a simple scheme of M/D/1 queuing system with preemption

Furthermore, the arrival rates and service rates, Utilization Factors and residual service time Parameters which are used in the computations and the simulation listed in the table below. It can be seen that as $\sum_{i=1}^{3} \rho i = 0.75 < 1$, the system is stable.

Class	Arrival Rate (λ_i) in user/second	Service Rate ($\mu_i=1/d_i$) in user/second	Utilization Factor (ρ_i)	Residual Service Time ($Z_{B,i}$) in second
1	0.25	1	0.25	0.5
2	0.05	0.2	0.25	2.5
3	0.025	0.1	0.25	5

Table 1: Arrival Rates, Service Rates, Utilization Factors and Residual Service Times

According to the fact that we have learned during the course, general service time breaks the memoryless property of exponential service time. In addition, we have deterministic service time for serving each category of arrivals. Thus, the residual service time matters. The residual service time is computed as half of the service time for each priority class.

- $Z_{B,1} = 0.5$
- $Z_{B,2} = 2.5$
- $Z_{B,3} = 5$

Waiting time for class-i users includes two different time periods and can be obtained using $W^q(i) = T_1 + T_2$ where:

- T_1: Time spent waiting by a class-i user until he is taken into service for the first time.
- T_2: Service Time of all higher-priority users (classes < i) who arrive after class-i user's first entrance into service.
- Also, we should compute Extended Service Time which is the time from first entrance in service to departure from the service and it is computed as $T^{serv}(i) = T_2 + \frac{1}{\mu}$.

To compute the per-class average queuing time, we get the beneficial of following formulas:

$$W^q(1) = \frac{\rho 1\ ZB,1}{1 - \rho 1} = \frac{(0.25 \times 0.5)}{0.75} = \mathbf{0.167\ seconds}$$

$$W^q(2) = \frac{\rho 1\ ZB,1 + \rho 2\ ZB,2}{(1 - \rho 1 - \rho 2)(1 - \rho 1)} + \frac{\frac{\rho 1}{\mu 2}}{1 - \rho 1} = \frac{(0.25 \times 0.5) + (0.25 \times 2.5)}{0.5 \times 0.75} + \frac{\frac{0.25}{0.2}}{0.75} = \mathbf{3.67\ seconds}$$

$$W^q(3) = \frac{\rho 1\ ZB,1 + \rho 2\ ZB,2 + \rho 3\ ZB,3}{(1 - \rho 1 - \rho 2 - \rho 3)(1 - \rho 1 - \rho 2)} + \frac{\frac{\rho 1 + \rho 2}{\mu 3}}{1 - \rho 1 - \rho 2}$$

$$= \frac{(0.25 \times 0.5) + (0.25 \times 2.5) + (0.25 \times 5)}{0.25 \times 0.5} + \frac{5}{0.5} = 16 + 10 = \mathbf{26\ seconds}$$

Then, we can compute the average response time per class using the average queueing time values we obtain in the previous step.

$$W(1) = W^q(1) + \frac{1}{\mu 1} = 0.167 + 1 = \mathbf{1.167\ seconds}$$

$$W(2) = W^q(2) + \frac{1}{\mu 2} = 3.67 + 5 = \mathbf{8.67\ seconds}$$

$$W(3) = W^q(3) + \frac{1}{\mu 3} = 26 + 10 = \mathbf{36\ seconds}$$

With these values, we can compute the generic average queueing and response times.

$\lambda = \lambda_1 + \lambda_2 + \lambda_3 = 0.25 + 0.05 + 0.025 =$ **0.325 user / second**

$$W^G(q) = \frac{\lambda 1}{\lambda} Wq(1) + \frac{\lambda 2}{\lambda} Wq(2) + \frac{\lambda 3}{\lambda} Wq(3) = \frac{0.25}{0.325} 0.167 + \frac{0.05}{0.325} 3.67 + \frac{0.025}{0.325} 26 = \mathbf{2.67\ seconds}$$

$$W^G = \frac{\lambda 1}{\lambda} W(1) + \frac{\lambda 2}{\lambda} Wq(2) + \frac{\lambda 3}{\lambda} Wq(3) = \frac{0.25}{0.325} 1.167 + \frac{0.05}{0.325} 8.67 + \frac{0.025}{0.325} 36 = \mathbf{5\ seconds}$$

In order to compute the Extended service time for each class, we can write:

$T^{serve}(1) = 0 + \frac{1}{\mu 1} =$ ***1 second***

$T^{serve}(2) = \frac{\frac{\rho 1}{\mu 2}}{1 - \rho 1} + \frac{1}{\mu 2} = 1.67 + 5 =$ ***6.67 seconds***

$T^{serve}(3) = \frac{\frac{\rho 1 + \rho 2}{\mu 3}}{1 - \rho 1 - \rho 2} + \frac{1}{\mu 3} = 10 + 10 =$ ***20 seconds***

In summary, we computed the theoretical results of these performance figures and show their values in table 2. All the values are in second.

- **W^q(i):** Per-class average queueing time for i = 1, 2, 3.
- **W(i):** Per-class average response time for i = 1, 2, 3.
- **T^{serve} (i):** Per-class extended service time for i = 1, 2, 3.
- **W^G(q):** Generic average queueing time.
- **W^G:** Generic average response time.

Class	$W^q(i)$	W(i)	T^{serve} (i)	$W^G(q)$	W^G
1	0.167	1.167	1	2.67	5
2	3.67	8.67	6.67		
3	26	36	20		

Table 2: Theoretical Results

3. Implementation and Results

After designing the topology of the queueing system, which was illustrated in figure 1, the simulation ran. The results related to the values of performance figures converges to the theoretical ones after running many events. Table 3 shows the experimental results (last step) which are received per-class from the simulation after running more than 1 million steps. The snapshot of simulation to show some parts of the output of simulation seen in table 3 is provided on the last page of the report (figure 2).

Class	$W^q(i)$	W(i)	T^{serve} (i)	$W^G(q)$	W^G
1	0.167081	1.16708	1	2.67286	4.9768
2	3.67923	8.67923	6.6675		
3	25.7993	35.7993	19.9342		

Table 3: Experimental Results

Finally, to have a clear comparison between the obtained values of theoretical and experimental performance figures, we put them on the same table below. As a conclusion, looking at the results obtained from the simulation, we can see that the values of Average Response Time, Extended Service Time and Server Utilization Factor are so close to the related theoretical values. It is happened also for Generic Average Queuing Time and Generic Average Response Time. Therefore, both types of the results are compatible with each other.

Performance Figures		Theoretical Values	Experimental Values
Average Queueing Time	$W^q(1)$	0.167	0.167081
	$W^q(2)$	3.67	3.67923
	$W^q(3)$	26	25.7993
Average Response Time	W(1)	1.167	1.16708
	W(2)	8.67	8.67923
	W(3)	36	35.7993
Extended Service Time	$T^{serve}(1)$	1	1
	$T^{serve}(2)$	6.67	6.6675
	$T^{serve}(3)$	20	19.9342
Server Utilization Factor	ρ_1	0.25	0.250217
	ρ_2	0.25	0.248569
	ρ_3	0.25	0.250727
Generic Average Queueing Time	$W^G(q)$	2.67	2.67286
Generic Average Response Time	W^G	5.00	4.9768

Table 4: Comparison between theoretical and experimental results

Figure 2: The output of simulation. Highlighted values are reported in the related tables