LNMB-course Algorithmic Methods in Queueing Theory (AlQT)
Rob van der Mei and Stella Kapodistria

Assignment 2022-23

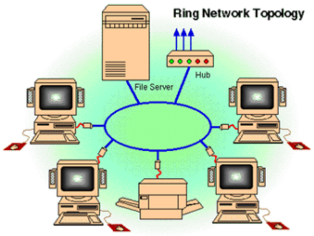
Deadline: April 3 2023

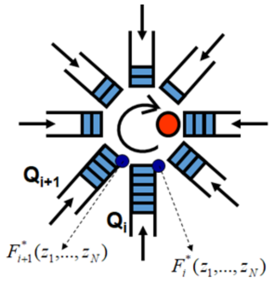
General information: This is the assignment for the course Algorithmic Methods in Queueing Theory (AIQT). The assignment requires the implementation of the approaches discussed during the course. The assignment can be made in groups of at most two persons, and it is also required to submit a report with your findings.

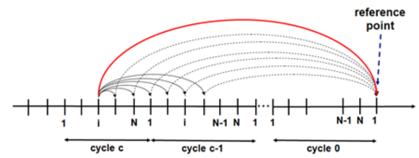
The report of the assignment should be sent by email to rob.van.der.Mei@cwi.nl and to s.kapodistria@tue.nl with the subject "Assignment for course AIQT" before 23:59 on April 3 2023. To this purpose, you need to create a single pdf file with the answers to the assignment. Since, the assignment requires programming, you can use your favorite programming language or mathematical/statistical software, however the original source code, in original format, should be submitted together with the report, i.e. in the same email as the solutions of the assignment.

The findings reported in the assignment should be presented in a clear, concise way and the code should be documented and should provide sufficient information for confirmation and replication of the results. Moreover, all graphs/tables should have a caption explaining what is depicted and all axes should be labeled.

Lastly, but still important, create a cover for the report and include the names of all the members of the group, the students affiliations and the course name.







<u>Exercise 1</u>: Waiting-time analysis in polling models

Consider the following three-queue polling model. Arrivals occur according to independent Poisson processes with *relative* arrival rates 1:2:2 for queues 1, 2 and 3, respectively. The service times at queue are exponentially distributed with mean 3, the service times at queues are gamma-distributed with mean 1 and squared coefficient of variation 3, and the service times at queue 3 are constant with mean 4. Queue 1 and queue 2 receive gated service and queue 3 is served exhaustively. The server visits the queue in cyclic order 1 → 2 → 3 → 1 → 2 → 3, and so on. The switchover times from queue 1 to queue 2, are exponentially distributed with mean 0.5, the switch-over times from queue 2 to queue 3 is uniformly distributed on the interval [0.2, 0.8], and the switch-over times from queue 3 to queue 1 are constant with mean 0.9.

1.1. What are the (absolute) arrival rates when the load of the system is set to $\rho = 0.85$? What is the mean switch-over time per cycle? What is the second moment of the total switch-over time per cycle? What are the first two moments of the service times of an *arbitrarily* chosen customer in the system (i.e., weighted proportionally to the arrival rates)?

1.2. A necessary and sufficient condition for the stability of the system is $\rho < 1$. Remarkably, this condition does not depend on the switch-over time distributions. Give an intuitive explanation for this (no proofs needed).

1.3. Formulate the pseudo-conservation law (PCL) for this model. Make sure that all the parameters that you mention are well-defined.

1.4. Assume $\rho = 0.85$. Write a program that calculates the expected waiting time at *each* of the three the queues by using the Buffer Occupancy Method (BOM). Write down the formula's and equations you use.

<u>Hint</u>: Validate the correctness of your results by comparing the left-hand side and right-hand side of the PCL formulated in Exercise 1.3.

1.5. For the same model, work out the details of Descendant Set Approach (DSA) and write a program that calculates the expected waiting time queue 1 (the mean waiting time at queues 2 and 3 are not needed). Check the correctness of the results by comparing the outcome to the results from Exercise 3.4.

Throughout, we assume that *all three queues are served exhaustively*.

1.6. Intuitively, what would you expect to happen to the mean waiting time at each of the three queues?

1.7. Calculate the mean waiting times at each of the queues by using the BOM, and check the correctness of the results by substituting the results into the PCL formulated in Exercise 1.3. Do the results meet your expectations formulate in Exercise 1.6? Interpret what you see.

1.8. Formulate *Boon's interpolation method* for the mean delay at each of the three queues, for arbitrary values of the load to the system ρ.

1.9. Formulate *Groenendijk's PCL-based approximation method* for the mean delay at each of the three queues, for arbitrary values of the load to the system ρ.

1.10. Plot a graph with three curves, in which you plot the mean waiting time at queue 1 as a function of the load ρ, where ρ is varied as 0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9, 0.95, 0.98 and 0.99. Curve 1 is based on the exact results obtained via the BOM or the DSA, and curves 2 and 3 are calculated using by Boon's approximation and Groenendijk's approximation, respectively. Discuss your findings.

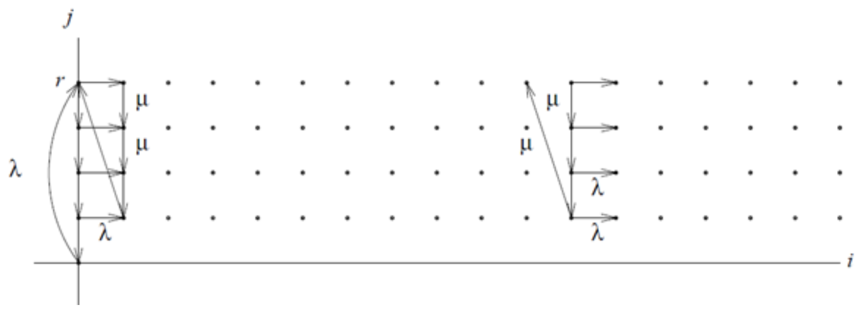
During the lectures, it was argued that the basic Poisson-driven cyclic polling models with nonzero switch-over times and with gated or exhaustive service at each queue fall in Resing's class of "easy" polling models for which the joint queue-length process at successive polling instants at (say) queue 1 can be described as a multi-type branching process with immigration in each state.

1.11. Now suppose that the server routing is not cyclic, but *periodic*, in the sense that the visit order is prescribed by a fixed polling table T, for example, for T = (1, 2, 1, 3, 2). Does this this model fall in the class of "easy" polling models? Motivate your answer.

1.12. As another alternative, let us assume that the visit order is *random*, in the sense that after leaving a queue, the server is routed *to* queue j with some probability p_j (regardless of the queue where the server departs from). Does this this model fall in the class of "easy" models? Motivate your answer.

<u>Exercise 2</u>: Matrix-geometric method for the M/Erlang-r/1 queue

Consider a single-server queue. Customers arrive according to a Poisson process with rate λ and they are served in order of arrival. The service times are Erlang-r distributed with mean r/μ. As discussed during the course, the evolution of the model can be described as a two-dimensional continuous-time Markov chain, as illustrated by Figure 1 below.



Prof. Marcel Neuts

Figure 1. Transition-rate diagram for the M/E_r/1 model.

2.1 The Q-matrix of this QBD process has a block structure. Write down the Q-matrix.

2.2 What are the stability conditions for the QBD process (by using Neuts' drift condition)?

2.3 What are the balance equations for this process?

2.4 Formulate the equation that characterizes the R-matrix. Write a program that solves this problem, and calculates the R-matrix. Add the source code to the assignment.

2.5 Write a program that calculates the equilibrium distribution of the QBD process via the matrix-geometric method. Add the source code to the assignment.

2.6 Assume that the mean service time r/μ = 1. Use the programs developed in Exercises 2.4 and 2.5 to calculate the expected delay as a function of the load ρ, where ρ is varied as 0.1, 0.2, 0.3, 0.5, 0.7, 0.8, 0.9, 0.95, 0.98 and 0.99 (by varying the arrival rate λ). Do so for r = 1 (which corresponds to the case of exponential service times), 2, 5 and 10; note that this creates four curves. Verify that the results are equal to the known results for the Pollaczek-Khintchine formula for the M/G/1 queue.

<u>Exercise 3:</u>

Consider the symmetric join the shortest queue model: Assume a queueing system consisting of three identical servers, in which each server has its own dedicated (infinite) queue and serves the customers in that queue according to FCFS. The duration of a service time is exponentially distributed with rate $\mu = 1$. Customers arrive to the system according to a Poisson process at rate λ and join the shortest queue. In case of a tie, the arriving customer joins either queue with the same probability (1/3).

We can model the above described system as a three-dimensional stochastic process by considering the corresponding queue lengths, i.e. $\{(X_1(t), X_2(t), X_3(t)),\ t \geq 0\}$, with $X_i(t)$ the number of customers in the queue (including the customer in service, if any) of the i-th server, $i = 1, 2, 3$. Note that the three-dimensional stochastic process $\{(X_1(t), X_2(t), X_3(t)),\ t \geq 0\}$ is a Markov chain defined on $\mathbb{N}_0 \times \mathbb{N}_0 \times \mathbb{N}_0$. The system as a whole is stable if $\lambda < 3$ or equivalently if $\rho = \frac{\lambda}{3} < 1$.

Assignment questions:

1) For $\rho = 0.5$, write a numerical routine based on the power series algorithm (PSA) and calculate the probability of a zero waiting time.

2) For $\rho = 0.5$, write a numerical routine based on the PSA and calculate the expected waiting time of a customer.

3) Comment on whether your numerical routine can be used in light/heavy traffic and (in case you would need to make adaptations) explain how you would adapt your numerical routine to work efficiently in these regimes.

4) For $\rho = 0.5$, describe the steps of a numerical routine based on the compensation approach (CA) for the calculation of the probability of an empty system. Describe how this process deviates from the one described in classroom for the two-dimensional queueing model.

Provide a pseudo-code for the numerical routines and substantiate the choices made (e.g., choice of hyperparameter values, termination criteria, etc).

For inspiration, you might want to read the paper titled “The power-series Algorithm applied to the shortest-queue model” by J. P. C. Blanc. Published at *Operations Research*, Vol. 40, No. 1, (Jan. - Feb., 1992), pp. 157–167. `http://www.jstor.org/stable/171192`.