Dashboard > My courses > 21VT-2DV608 > Module 2: Performance Engineering > Performance Engineering - Retake

Performance Engineering - Retake

Performance Engineering

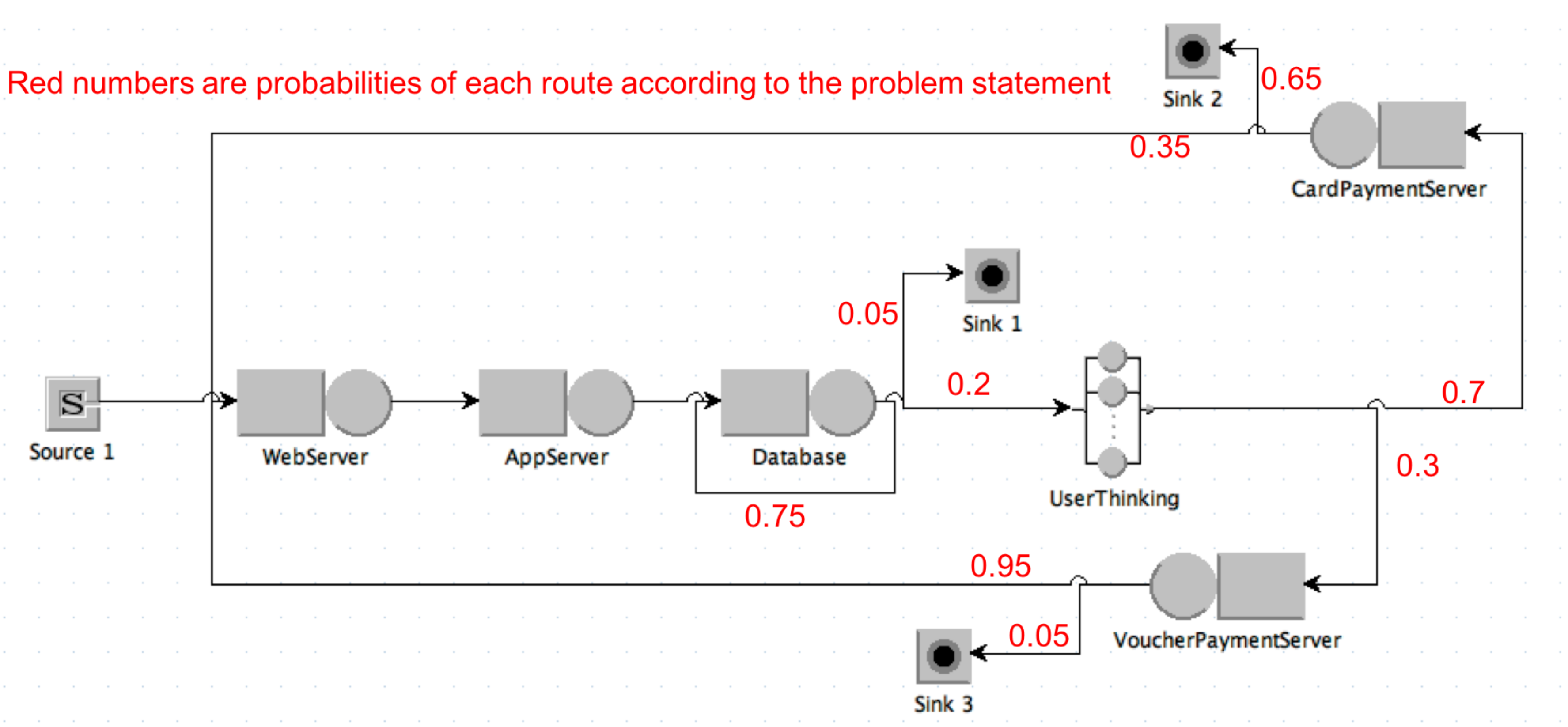
We are going to upgrade a software system that executes web sessions from users from the Internet. The system offers an e-commerce site to the users.

We have observed the running system for 3 full days. During these three days, we have seen that 466560 user sessions have been completed. Actually, we have seen that sessions arrived at an average rate of 1.8 sessions per second, and, therefore, we can deduce that all sessions that arrived were completed.

The system is composed of five service centers: A *WebServer,* an *ApplicationServer*, a *Database*, a *CardPaymentServer,* and a *VoucherPaymentServer.* There is one resource for each of the service centers. The description of the system and measured information is the following:

- Each user's session starts executing in the *WebServer*. The average demand of this server during a complete user session is 104.16ms, and this server receives 1.7361 visits on average during a complete user session. After the *WebServer* executes, the user session continues in the *ApplicationServer.*The *ApplicationServer* produces the list of articles to the user. We have observed that the utilization of the *ApplicationServer* is 78.12%. We also saw that there were, on average, 3.565 jobs in the *ApplicationServer;* and that, the average time between a job arrived at the *ApplicationServer* and its execution was completed (including the time that it spent waiting for service in a queue) was 1.1408 seconds.
- To create the list of articles, each execution in the *ApplicationServer* needs to request, on average, 4 times the execution of the *Database* (it reads from the Database the information about articles, pricing, etc.). The service time of a *Database* execution is 35ms, on average.
- After the output of the *ApplicationServer* is produced, the user may decide to immediately leave the system (when he/she did not get the expected type of articles), which happens 20% of times, or to buy something the 80% of times. If he/she decides to buy something, the user thinks which article to purchase for 20 seconds, on average. After deciding the article, the 70% of times the user decides to purchase an article and pay with a credit card, and the 30% of times the user decides to purchase and article and pay with a Voucher. The credit card payments are executed in the *CardPaymentServer,* while the payments using a voucher are executed in the *VoucherPaymentServer.*
- The *CardPaymentServer* executes the logic for payment with credit card: receives the payment information from the user, it performs basic tests about the correctness of the data (for example, the existence of card number, no special characters in the name or address, etc.) and contacts the banking system. The average service time of this service center is 200ms.
- The *VoucherPaymentServer* executes the logic for payment with a Voucher. We have observed that the utilization of the *VoucherPaymentServe* is 22.5%.
- After the payment, the 95% of users who paid with a Voucher feel that they would like to continue purchasing items and return to the *WebServer* to start another purchase (the other 5% leave the system). However, only 35% of users who paid with actual money in their cards start another shopping (the other 65% leave the system).

The following figure shows the structure of a Queueing Network that represents the system and the routing probabilities between elements.



Red numbers are probabilities of each route according to the problem statement

You have to submit a PDF document following the template with your answers to the following 3 exercises:

A) Use the operational laws to calculate the average service time Sk of the *WebServer, ApplicationServer* and the *VoucherPaymentServer.*

B) Model the system that has only one resource in each service center using Queueing Networks (in JMT or your preferred Queueing Network simulation engine). Simulate the model to calculate the system response time, the utilization of each service center, and the throughput of each service center. Use the simulation results to calculate how much time a user is waiting on average, during his/her complete user session, for a reply from the system (clarification: the user is not waiting for a reply from the system when he/she is thinking what to buy, and he/she is waiting a reply from the system when the execution is any of the other service centers)

C) Assume that the population is going to be confined at home due to a pandemic. In that case, people will increase their online shopping because they cannot go shopping physically. Therefore, we expect that the workload to our e-commerce site will increase to an average rate of 12 user sessions per second. Moreover, we also expect that people embrace more consumerism during confinement, increasing the proportion of people who continues buying after a payment with credit card from the previous 35% to a new 55%.

In that situation, you suspect that some components of your system will saturate.

- Calculate the minimum number of resources of each server that we will need to avoid saturation.
- Simulate the new system to calculate the system response time, utilization of each service center, throughput of each service center, and the amount of time that a user is waiting, on average, for a reply from the system.

Hint1: In the cases that, from a service center (e.g., *Database*), a job can go to more than one service center, use Probabilistic Routing.

Hint2: Use the exponential distribution for all times and rates (frequencies) you need to model.

Hint3: Use the figure above, the structure and probabilities are the same.

Hint4: If the question about *"how much time a user is waiting"* does not sound familiar to you, check slides 24-25 in Part 3. We calculated the proportion of time that a user is waiting in that exercise, which is something similar to what it is asked here.

Good luck!

template-PE-retake.docx	31 March 2021, 1:08 AM
template-PE-retake.pdf	31 March 2021, 1:08 AM

Submission status

Attempt number	This is attempt 1.
Submission status	Submitted for grading
Grading status	Graded
Due date	Wednesday, 31 March 2021, 11:59 PM
Time remaining	Assignment was submitted 31 mins 46 secs early
Last modified	Wednesday, 31 March 2021, 11:27 PM
File submissions, max 100 MB per file	xz222bb-PE-retake.pdf Opt-out URKUND31 March 2021, 11:27 PM
Submission comments	Comments (0)

Feedback

Grade	35.0 (35.0 %)
Graded on	Wednesday, 7 April 2021, 10:55 AM
Graded by	Diego Perez Palacin
Feedback files	xz222bb-PE-retake.pdf 7 April 2021, 10:55 AM

« PREVIOUS ACTIVITY
Performance Engineering

NEXT ACTIVITY »
Architecting and Design Forum

Jump to...

Lnu

Lnu.se/Student

Data privacy

Moodle App

Help

Help

IT Support

Links

LnuPlay

Zoom login

Moodle Docs for this page