**DATE:** 29/04/2020

**NAME:** ADEKUNLE GLORIA. J

**MATRIC. NO:** 16/SCI01/001

**COURSE CODE:** CSC410

**COURSE TITLE:** COMPUTER SYSTEM PERFORMANCE EVALUATION

ASSIGNMENT

1. Explain the concepts of operational laws as applied to computer and network system performance evaluation.

2. Exhaustively describe at least eight operational laws that are widely employed in computer system performance evaluation.

3. Distinguish between the Forced Flow Law and the Residence Time Law from a systems perspective (not by definition).

4. Discuss some basic queuing models and basic queuing disciplines.

5. Discuss how to resolve some basic queuing problems.

6. You have been presented with some systems performance evaluation report; the first was done using measurement technique only, the second was done with simulation technique only, the third was done using analytical technique only, then the fourth was done using measurement and simulation technique only, while the fifth was done using simulation and analytical technique only and the sixth was done using measurement and analytical technique only. You are the director of information technology infrastructure in your firm and your firm is about to acquire the information infrastructure concerned in this report

i. What specific motive will you consider imperative in general?

ii. Why do you consider this metric important?

iii. If they are not in this report, what will you do in such a report?

iv. If they are part of the report, what is the first action to take with that report and why?

v. Which reports or combination of report will you adopt and why?

**SOLUTION**

1) Operational laws are equations which may be used as an abstract representation of a model of the average behavior of any system. They are general and make no assumptions about the behavior of the random variable that characterize the system. Using these abstractions, these laws can be applied to any device and gradually build a more complex system.

Operational laws are based on observable variables i.e. values which we could derive from watching system over a finite period of time. We assume that the system receives requests from its environments. Each request generates a job or customer within the system. When the job has been processed the system responds to the environment with the completion of the corresponding request. If we observed such an abstract system, we might measure the following quantities:

T, the length of time we observe the system;

A, the number of request arrivals we observe;

C, the number of request completions we observe;

B, the total amount of time during which the system is busy (B ≤ T );

N, the average number of jobs in the system.

From these observed values we can derive the following four important quantities:

λ = A/T, the arrival rate;

X = C/T, the throughput or completion rate,

U = B/T, the utilization;

S = B/C, the mean service time per completed job.

Assume that the system is job flow balanced. This means that the number of arrivals is equal to the number of completions during an observation period, i.e. A = C

This is a testable assumption because an analyst can always test whether the assumption holds

**NB:** that if the system is job flow balanced the arrival rate will be the same as the completion rate, that is, λ = X.

2) a) **Little’s Law**- Little’s Law is a theorem that determines the average number of items in a stationary queuing system based on the average waiting time of an item within a system and the average number of items arriving at the system per unit of time.

The law provides a simple and intuitive approach for the assessment of the efficiency of queuing systems. The concept is hugely significant for business operations because it states that the number of items in the queuing systems primarily depends on two key variables, and it is not affected by other factors such as the distribution of the service or service order.

Little’s Law can only be used in queuing systems. In addition, the theorem can be applied in different fields, from running a small coffee shop to the maintenance of the operations of a military airbase.

 **Formula for Little’s Law**

 L = λ x W

L – The average number of items in a queuing system

λ – The average number of items arriving at the system per unit of time

W – The average waiting time an item spends in a queuing system

b) **Space-Time Product Laws**- it states that the throughput is equal to average amount of memory in use divided by average space-time product. Space-time products are often used to evaluate program performance and assign accounting charges to programs in virtual memory systems. Essentially, a program’s space-time product is equal to its execution time multiplied by the average amount of money allocated to it during its execution. Since space-time products, response time, and throughput are all used as indicators of system performance, it is interesting to examine the manner in which these quantities are related.

c) **Forced-Flow Law**- The Forced-Flow Law (FFL) relates throughputs at individual resources within a system to the overall system throughput. It is the average no of visits that a system level job makes to that resource. The general residence time law is the sum of the product of its average residence time at each resource and the number of visits it makes to that resource.

**The Forced-Flow Law:**

λk=Vkλ.

The average arrival rate to resource k is the total system arrival rate times the expected number of visits made to resource k. Because of the Conservation Law, we could also state the FFL in terms of output rates, Λ and Λk.

d) **Conservation Law**- What goes in must come out. Consider a system with arrival rate of λ and an output rate of Λ. If the system is not overloaded and no customers are created or destroyed inside the system, then λ = Λ.

Creating customers in a system is called forking. Destroying customers is called joining. It’s possible to model systems with these behaviors, but usually difficult, so we won’t see any examples until the end of class.

e) **Utilization Law**- It states that, the utilization of a resource is equal to the product of the throughput of that resource and the average service requirement at that resource. If we know the amount of processing that each job requires at a resource, then we can calculate the utilization of the resource. The total amount of service that a system job generates at the ith resource is called the service demand,

 Di: Di = SiVi

The utilization of a resource, the percentage of time that the ith resource is in use processing to a job, is denoted Ui

 Ui = XiSi = XD

f) **Interactive Response Time Law**- The name of this law dates back to the time when most of the systems which were being modelled were mainframes processing both interactive jobs and batch jobs. The think time, Z, was quite literally the length of time that a programmer spent thinking at his terminal before submitting another job. More generally interactive systems are those in which jobs spend time in the system not engaged in processing, or waiting for processing: this may be because of interaction with a human user, or may be for some other reason.

The think time represents the time between processing being completed and the job becoming available as a request again. Thus the residence time of the job, as calculated by Little’s law as the time from arrival to completion, is greater than the system’s response time. The interactive response time law reflects this: it calculates the response time, R as follows:

 R=N=X-Z

The response time in an interactive system is the residence time minus the think time.

Note that if the think time is zero, Z = 0 and R = W, then the interactive response time law simply becomes little’s law.

g) **General Residence Time Law**- One method of computing the mean residence or response time per job in a system is to apply little’s law to the system as a whole. However, if the mean number of jobs in the system, N, or the system level throughput, X, are not known an alternative method can be used. Applying little’s law to the ith resource we see that Ni = XiWi, where Ni is the mean number of jobs at the resource and Wi is the average response time of the resource. From the forced flow law we know that Xi = XVi. Thus we can deduce that

 Ni/X = ViWi.

 The total number jobs in the system is clearly the sum of the number of jobs at each resource, i.e. N = N1 + • • • + NM if there are M resources in the system. We know from Little’s law that W = N/X and from this we arrive at the general residence time, or general response time law:

The average residence time of a job in the system will be the sum of the product of its average residence time at each resource and the number of visits it makes to that resource.

h) **Bottle leg analysis** - The resource within a system which has the greatest service demand is known as the bottleneck resource or bottleneck device, and its service demand is maxi {Di}, denoted Dmax. The bottleneck resource is important because it limits the possible performance of the system, this will be the resource which has the highest utilization in the system.

i) **Residence time, service demand, contention -** The residence time of a job within a system will always be at least as large as the total amount of processing that each job requires.

The total amount of processing that a job requires is D, the total service demand,

M

D = \ Di

i =1

In general, there will be some contention in the system meaning that jobs have to wait for processing so the residence time will be larger than this, i.e. W ≥ D

j) **Throughput, utilization and overall performance** - The throughput of a system will always be limited by the throughput at the slowest resource (think of the Forced Flow Law); this is the bottleneck device.

By the utilization law, at this resource, let’s call it b, Ub = XDmax .

Therefore, since Ub ≤ 1

X ≤ 1/Dmax

It follows that if we wish to improve throughput we should first concentrate on this resource—improving throughput at other resources in the system might have little effect on the overall performance.

3) The forced flow time law; It is often natural to regard a system as being made up of a number of devices or resources. Each of these resources may be treated as a system in its own right as far as the operational laws are concerned, with the rest of the system forming the environment of that resource. A request from the environment generates a job within the system; this job may then circulate between the resources until all necessary processing has been done; as it arrives at each resource it is treated as a request, generating a job internal to that resource. The forced flow law, which states that the flows (throughputs) in all parts of a system must be proportional to one another. Suppose that during an observation interval we count not only system completions, but also the number of completions at each resource. We define the visit count of a resource to be the ratio of the number of completions at that resource to the number of system completions, or, more intuitively, to be the average number of visits that a system-level request makes to that resource. If we let a variable with the subscript k refer to the k-th resource (a variable with no subscript continues to refer to the system as a whole), then we can write this definition as: V,, the visit count of resource k: ck V,

S 7 If during an observation interval we measure 10 system completions and 150 completions at a specific disk, then on the average each system-level request requires 150/10 = 15 disk operations. If we rewrite this definition as ck = vk c and recall that the completion count divided by the length of the observation interval is defined to be the throughput, then the throughput of resource k is given by:

The Forced Flow Law: Xk = V,X. And while the response time law on the other hand;One method of computing the mean residence or response time per job in a system is to apply Little’s law to the system as a whole. However, if the mean number of jobs in the system, N, or the system level throughput, X, are not known an alternative method can be used. Applying Little’s law to the ith resource we see that

Ni = XiWi ,

Where Ni is the mean number of jobs at the resource and Wi is the average response time of the resource. From the forced flow law we know that

Xi = XVi . Thus we can deduce that,

 Ni/X = ViWi

The total number jobs in the system is clearly the sum of the number of jobs at each resource, i.e. N = N1 + • • • + NM if there are M resources in the system. We know from Little’s law that W = N/X and from this we arrive at the general residence time, or general response time law:

W = X M i=1 WiVi

4) Basic Queuing Models;

a. Calling population: the population of potential customers, may be assumed to be finite or infinite.

I. Finite population model: if arrival rate depends on the number of customers being served and waiting, e.g., model of one corporate jet, if it is being repaired, the repair arrival rate becomes zero.

II. Infinite population model: if arrival rate is not affected by the number of customers being served and waiting, e.g., systems with large population of potential customers.

b. System Capacity: a limit on the number of customers that may be in the waiting line or system.

I. Limited capacity, e.g., an automatic car wash only has room for 10 cars to wait in line to enter the mechanism.

II. Unlimited capacity, e.g., concert ticket sales with no limit on the number of people allowed to wait to purchase tickets.

C. Arrival Process: In terms of inter-arrival times of successive customers.

I. Random arrivals: inter-arrival times usually characterized by a probability distribution.

II. Poisson arrival process (with rate λ), where An represents the inter-arrival time between customer n − 1 and customer n, and is exponentially distributed (with mean 1/λ).

III. Scheduled arrivals: inter-arrival times can be constant or constant plus or minus a small random amount to represent early or late arrivals.

D. Arrival Processes - Finite population models: Customer is pending when the customer is outside the queueing system, e.g., machine-repair problem: a machine is “pending” when it is operating, it becomes “not pending” the instant it demands service form the repairman.

Basic Queuing Disciplines-

a) First-in-first-out (FIFO)- this means the oldest inventory items are recorded as sold first but do not necessarily mean that the exact oldest physical object has been tracked and sold. In other words, the cost associated with the inventory that was purchased first is the cost expensed first.

b) Last-in-first-out (LIFO)- this describes a method for accounting for inventories. Under this system, the last unit added to an inventory is the first to be recorded as sold.

c) Service in random order (SIRO)- Under this type of queue structure, the customer is chosen for service randomly and hence all the customers are equally likely to be selected. Therefore, the time of arrival of the customer has no consequence on the selection of the customer.

d) Shortest processing time first (SPT) - Its principle is to order jobs according to their duration and schedule them by beginning by the shorter.

e) Service According to Priority (PR)

5) This could be by evaluating the system’s performance by either the measurement, simulation or analytical technique;

• Measurement; measurement of the real system can be undertaken if the real system is accessible and the process of measurement will not disrupt the system or services. The impact of these disturbances or instructions are very serious.

• Simulation; this technique drops from discreet event simulation philosophies a simplified model of the system and its nodes are implemented in a software and the performance of interest is measured as of a real system. But measurement side effects are usually not present.

• Analytical; a mathematical model of the system is analyzed numerically. This technique is viewed as a special form of simulation, it is often much quicker than simulation, but most times wide assumption needs to be made for numerical processes to be applicable. Analytical methods are usually used to gain insight during development phases and also to learn fundamental facts about a system.

These are some of the algorithms that can be used to solve queuing problem;

6) I. Don’t trust the result of a simulation model until they have been validated by other performance evaluation techniques

II. it’s important so as to know if the system is correctly defined and if the goals are clearly stated.

iii. Such report is invalid

iv. Go through the report to check if there are any mistakes

V. The forth report (measurement and simulation) because measurement technique requires more effort and it is the most accurate; simulation method requires relatively lesser effort and it is averagely accurate while analytical method is very quick and often very less accurate.