NAME: Najeebat Ahmed opeyemi

MATRIC NO: 16/sci03/001

COURSE CODE: CSC 410

1.) The concepts of operational laws as applied to computer and network system performance evaluation.

A number of laws are derived which establish relationships between throughput, response time, device utilization, space-time products and various other factors related to computer system performance. These laws are obtained through the operational method of computer system analysis. The operational method, which is formally introduced in this paper, differs significantly from the conventional stochastic modeling approach and is based on a set of concepts that correspond naturally and directly to observed properties of real computer systems. The operational laws presented in this paper apply with complete precision to all collections of observational data.

2.) Eight operational laws that are widely employed in computer system

a. Utilization Law

Utilization law can be used to verify the internal consistency of a set of empirical data collected during some observation interval.

Note that the Throughput Law can be applied to

any device or processor in a system. If, for example, the law were applied to device j with j # i, one

would obtain

X= U(j) / S(j).D(j)

Since throughput as defined in equation (i) has a unique value regardless of the device used in equation

U(I)/s(I).D(I)=U(j)/s(j).D(j)

. This equation can be used to verify the internal consistency of a set of empirical data collected during some observation interval. As in the case of the Throughput Law,

the Utilization Equality is applicable to all systems and is a universal law of computer performance

b. Forced Flow Law

Forced Flow Law relates the system throughput to individual device throughputs.

According to the Forced Flow law, the following is proven

• In an open model, System throughput is equal to the number of jobs leaving the system per unit time

• In a closed model, System throughput is equal to the number of jobs traversing out to in link per unit time.

• If observation period T is such that Ai = Ci, then the Device satisfies the assumption of job flow balance.

• Each job makes Vi requests for ith device in the system.

• Ci = C0 Vi or Vi = is called visit ratio

• System Throughput (X) = =

• Device Throughput (Xi) = = which can be simplified to Xi = X Vi

Where Ci and C0 = The Jobs completed

T = Total Time

Ai = Number of arrivals

Vi =

c. 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. Almost any queuing system and even any sub-system can be assessed using the law. 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.

The Formula for Little’s Law:

Mean Number in the device = Arrival Rate Mean Time in the device

Which can be denoted as Qi = λi Ri

If the job flow is balanced, the arrival rate is equal to the throughput and we can write:

Qi = Xi Ri

d. General Response Time Law

According to General Response Time Law, there is one terminal per user and the rest of the system is shared by all users.

When applying Little's law to the central subsystem, the formula is Q = X R

Where, Q = Total number of jobs in the system

R = system response time

X = system throughput

The General Response Time Formula is R = X1R1 + X2R2 + … + XMRM

When dividing both sides by X and using forced flow law, the formula is

R = V1R1 + V2R2 + … + VMRM

e. Interactive Response Time Law

According to Interactive Response Time Law:

• If Z = think-time, R = Response time

The total cycle time of requests is R+Z

Each user generates about T/(R+Z) requests in T

• If there are N users:

System Throughput = =

This can be further simplified into )

Or R =

f. Bottleneck Analysis

From forced flow law, the formula is Ui α Di

According to the Bottleneck Analysis,

• The device with the highest total service demand Di has the highest utilization and is called the bottleneck device.

• Note: Delay centers can have utilizations more than one without any stability problems. Therefore, delay centers cannot be a bottleneck device.

• Only queueing centers used in computing Dmax.

• The bottleneck device is the key limiting factor in achieving higher throughput.

• Improving the bottleneck device will provide the highest payoff in terms of system throughput.

• Improving other devices will have little effect on the system performance.

• Identifying the bottleneck device should be the first step in any performance improvement project.

Throughput and response times of the system are bound as follows:

And​

​Here, D = is the sum of total service demands on all devices except terminals.

These are known as asymptotic bounds

Bottle Neck Analysis Proof:

The asymptotic bounds are based on the following observations:

The utilization of any device cannot exceed one. This puts a limit on the maximum obtainable throughput.

The response time of the system with N users cannot be less than a system with just one user. This puts a limit on the minimum response time.

The interactive response time formula can be used to convert the bound on throughput to that on response time and vice versa.

For the bottleneck device b we have: Ub = XDmax

Since Ub cannot be more than one, we have: XDmax 1

​​​​​​X

With just one job in the system, there is no queueing and the system response time is simply the sum of the service demands:

R(1) = D1 + D2 + … + DM = D

Here, D is defined as the sum of all service demands. ! With more than one user there may be some queueing and so the response time will be higher. That is:

Applying the interactive response time law to the bounds:

R(N) D

Combining these bounds we get the asymptotic bounds

g. Space-Time Product Throughput Law

The Space-Time Product Throughput Law states that the throughput is equal to the average amount of memory in use are divided by average space-time product.

The formula is:

Where X = Throughput (i.e. number of job completions per unit time)

​M = Average amount of memory in use during the observation interval

​Y = Average space-time product (i.e. space-time product completed per job

h. Space-Time Product Response Time Law

The Space-Time Product Response Time Law is obtained by a similar relationship between average space-time product and average response time in the asymptotic case.

The formula is:

​Where R = response time (i.e. average amount of time in system state per interaction)

N = Number of interactive terminals. It is assumed that N is constant throughout the observation interval.

Y = Average space-time product (i.e. space-time product completed per job

M = Average amount of memory in use during the observation interval

Z = Average think time (i.e. average amount of think time per transition from think state to system state).

3.) The differences between the Forced Flow Law and the Residence Time Law from a systems perspective

Forced Flow Law

* Relates the system throughput to individual device throughputs.
* In an open model, System throughput = # of jobs leaving the system per unit time
* In a closed model, System throughput = # of jobs traversing OUT to IN link per unit time.
* If observation period T is such that Ai = Ci⇒ Device satisfies the assumption of job flow balance.
* Each job makes Vi requests for ith device in the system
* Ci = C0 Vi or Vi =Ci/C0 Vi is called visit ratio





**WHILE**

**Residence Time Law**





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

Some basic queuing models:

Arrival process: The arrival process is simply how customers arrive. They may come into a queue alone or in groups, and they may arrive at certain intervals or randomly.

Behavior: How do customers behave when they are in line? Some might be willing to wait for their place in the queue; others may become impatient and leave. Yet others might decide to rejoin the queue later, such as when they are put on hold with customer service and decide to call back in hopes of receiving faster service.

How customers are serviced: This includes the length of time a customer is serviced, the number of servers available to help the customers, whether customers are served one by one or in batches, and the order in which customers are serviced, also called service discipline.

Service discipline: refers to the rule by which the next customer is selected. Although many retail scenarios employ the “first come, first served” rule, other situations may call for other types of service. For example, customers may be served in order of priority, or based on the number of items they need serviced (such as in an express lane in a grocery store). Sometimes, the last customer to arrive will be served first (such s in the case in a stack of dirty dishes, where the one on top will be the first to be washed).

Waiting room: The number of customers allowed to wait in the queue may be limited based on the space available.

Basic queuing discipline:

First In First Out (FIFO): First In, First Out, commonly known as FIFO, is an asset-management and valuation method in which assets produced or acquired first are sold, used, or disposed of first. For tax purposes, FIFO assumes that assets with the oldest costs are included in the income statement's cost of goods sold (COGS). The remaining inventory assets are matched to the assets that are most recently purchased or produced.

Last In First Out (LIFO): Last in, first out (LIFO) is a method used to account for inventory that records the most recently produced items as sold first. Under LIFO, the cost of the most recent products purchased (or produced) are the first to be expensed as cost of goods sold (COGS)—which means the lower cost of older products will be reported as inventory.

Shortest Remaining Time(SRT): Shortest remaining time, also known as shortest remaining time first, is a scheduling method that is a preemptive version of shortest job next scheduling. In this scheduling algorithm, the process with the smallest amount of time remaining until completion is selected to execute.

5. How to resolve basic queuing problems

Let Customers Know How Long The Wait Is: The uncertainty of how long it will take to wait is often the cause of queue anxiety. Because of this the customers are impatient and this is a major cause of queuing problems as people want to jump the queues or altogether leave the queue.

Assess and improve your queue management strategy: How do you currently handle a long line of customers? Think about what works well and what doesn’t. Assessing the tactics used to manage the queue in the particular organization will really help solve the queuing problems being encountered.

Design Your Environment To Be Able To Accommodate Queues: Studies have shown that one of the most common issues found in lines is queue anxiety. A well-designed queuing area can help organize waiting lines, remove the possibility of queue jumpers and generally ease customer flow management.

Implement Digital Queuing Software: Long queues can inspire customer’s irritation even disgust. But anyone can learn how to reduce queues the use of a nifty technology called a queue management system (QMS).Automating the queuing process creates more labor efficient customer lines, decreases the overall amount of walkaways as well as ultimately reducing queue times. When it’s their turn, a teller calls them to the counter to be served. They can see where they are in line by observing HDTVs hung on the walls of the organization and therefore customers are free to sit or wander and maybe grab a coffee across the street as they wait. They’re not corralled into the line like sheep. By giving customers back their time (their autonomy) one enable customers to wait in leisure. Now that’s effective queue management.

Occupy Customers in The Queue: Boredom in the queue can often lead to longer perceived waiting times. Queue solutions is to provide a distraction to people in the queue and help them continue shopping while waiting, easing up frustrations etc. Display entertaining programming on HDTVs. Prompt customers to answer surveys to report on their experience. Engaging customers is the best way to reduce the tension inherent in queueing. Because it’s typically the psychology behind queueing rather than the queues themselves that makes queues feel unbearable.

. Keep The Rules Of Queuing Fair And Consistent: One of the most important characteristics of any queue problem solving method is the queuing discipline used. Simply put, the queuing discipline is the rule used to decide who goes next in a queue.

Two of the most commonly used rules are:

o First in, First out.

o Last in, First out.

Bottom line, people expect queues to be fair. It’s not like they’re happy to be stuck waiting in line, to begin with. But when everyone abides by the same rules, we can’t help but follow them too.

Reduce Response times: So when it comes to providing service, be quick as possible. It's not possible to solve every problem immediately, but customers don’t expect that from you. What they do expect is that you give them some kind of response quickly. Having all information at your fingertips is the next step as these steps will help improve the flow of the queues and have less waiting times.

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.