## Chapter 6. Waiting Line Analysis

The Little’s Law represents the steady state relationship between Throughput, Flow Time and Inventory in form of RT=I. Using this formula, and also the using the unit load of each resource pool, Tp. And capacity of each resource pool (c/Tp), we can compute utilizations. Using utilization (U) of each resource pool, and c the number of resource units in each resource pool we can compute the number of low units in each resource pool. Summation of the member of flow units in each resource pool provide us with the total number of flow units in all resource poos, i.e., in the process. That is what we refer to it as Ip (number of flow units in the processes). If we add Tp of all the resource pools, we then get the theoretical flow time (ThFt), by using the Little Law R(ThFT) = Ip, we can find Ip in a different way. If the Ips obtained through the two procedures are not equal, we have made a mistake, either in computing ThFT or in not including c. However, a process does not contain the process alone, usually, a waiting line exists in front of each process (even is capacity is larger or much larger than demand Rp>>R), Why?

****



**1. The Application to Waiting Lines:**

In earlier chapters, we were introduced to formulas like:

1. Flow time $T=T\_{i}+T\_{p}$
* Explains that flow time ($T$) is equal to buffer time ($T\_{i}$) + processing time ($T\_{p}$)
1. Inventory $ I=I\_{i}+I\_{p}$
* Explains that Inventory ($I$) is equal to the number of flow units in the buffer ($I\_{i}$) + the total number of flow units in all processors ($I\_{p}$)
1. $I\_{i}=R x T\_{i}$
* Explains that the number of flow units in the buffer ($I\_{i}$) is equal to throughput ($R $x buffer time ($T\_{i}$)
1. $I\_{p}=R x T\_{p}$
* Explains that the total number of flow units in all processors ($I\_{p}$) is equal to throughput x processing time ($T\_{p}$)

**Hence we can conclude that:**

$$R=\frac{I}{T}=\frac{I\_{i}}{T\_{i}}=\frac{I\_{p}}{T\_{p}}$$

We also have to take into consideration the formulas that we learn in Capacity such as:

1. $U=R/R\_{p}$
* Explains that utilization is equal to throughput ($R$) divided capacity ($R\_{p}$)
1. $R\_{p}=c/T\_{p}$
* Explains that Capacity ($R\_{p}$) is equal to the number of processors ($c$) divided by processing time ($T\_{p}$)
1. $R=I\_{p}/T\_{p}$
* Explains that throughput ($R$) is equal to the total number of flow units in all processors ($I\_{p}$) divided by processing time ($T\_{p}$)

**Hence we conclude that:**

$$U=\frac{R}{R\_{p}}=\left(\frac{I\_{p}}{T\_{p}}\right)\left(\frac{c}{T\_{p}}\right)=\frac{I\_{p}}{c}$$

$$U=\frac{I\_{p}}{c}$$

**B. Characteristics of Waiting Lines**

1. Variability in arrival time and service time leads to
2. Idleness of resources
3. Waiting time of flow units

As utilization goes up, the number of people in the buffer goes up. The other one is variability; If every four minutes, exactly every four minutes, one customer arrives, and it takes exactly three minutes for that flow unit to pass the processor, then we never observe the waiting line.

In this case, utilization is one hundred percent, that is the maximum possible utilization, but there is absolutely no variability, neither in inter-arrival time nor flow time, then we never see flow units in the buffer.

1. We are interested in two measures
2. Average waiting time of flow units in the waiting line and in the system (Waiting line + Processor).
3. Average number of flow units waiting in the waiting line (to be then processed).

**C. Operational Performance Measures**

As explained earlier, if we know $R$, and we know $T\_{p}$, we can easily computer $I\_{p}$. Therefore all we need to compute is $T\_{i} and I\_{i}.$ And if we know one of them because we know throughput ($R$), then we can computer the other one.

1. Flow time $T=T\_{i}+T\_{p}$

1. Inventory $I=I\_{i}+I\_{p}$

We have an approximation formula for Ii and we have exact formula for some specific cases of waiting line.

**D. Utilization – Variability - Delay Curve**

The graph below explains the relationship between utilization and variability. As utilization increases, the waiting time increases and as variability increases, the waiting time increases and vice versa.



* + - 1. **Utilization and Variability:**
1. Our two measures of effectiveness (average number of flow units waiting and their average waiting time) are driven by

**1. Utilization:** The higher the utilization the longer the waiting line/time.

**2. Variability:** The higher the variability, the longer the waiting line/time.

B. High utilization $U=R/R\_{p} $or low safety capacity$ Rs=R\_{p}-R$, due to

**1. High inflow rate** $R$

**2. Low** processing rate $R\_{p}=c/T\_{p}$, which may be due to small-scale **c** and/or slow speed $1/T\_{p}$

**2. Drivers of Process Performance:**

Variability in the inter arrival time and processing time is measured using standard deviation (or Variance). Higher standard deviation (or Variance) means greater variability. The formula for variance is:

$$V\left(X\right)=E[(X- μ\_{x})^{2}$$

Standard deviation is computed by taking the square root of variance. The formula for standard deviation is:

$$σ\_{X}=\sqrt{V\left(x\right)}= \sqrt{E[(X-μ\_{x})^{2}]}$$

Standard deviation is not enough to understand the extent of variability. Does a standard deviation of 20 *for an average of 80* represents more variability or a standard deviation of 150 *for an average of 1000?*

Coefficient of Variation: The ratios of the standard deviation of inter arrival time (or processing time) to the mean (average).

We refer to coefficient of variation of inter-arrival time as $C\_{a}$, and coefficient of variation of processing time as $C\_{p}$, capital $C$. The number of servers we show it by small $c$, coefficient of variation, we show it by capital $C$. Capital $C$ small $a$ for coefficient of variation of inter-arrival time, and capital $C$ small $p$ for coefficient of variation of processing time.

$$C\_{a=}coefficient of variation for interarrival time$$

$$C\_{p=}coefficient of variation for processing time$$

**3. The Queue Length Approximation Formula:**

The approximation formula is used for the number of flow units in the buffer. Ii, the number of flow units in the buffer is equal to utilization to the power of two times one plus the number of servers divided by one minus the utilization and we refer to it as utilization affect or u-part. Utilization affect is multiplied by variability affect. And that is squared coefficient of variability of inter-arrival time plus square of coefficient of variability of processing time divided by two. And we refer to it as variability affect or v-part.

$$I\_{i}=\frac{U^{\sqrt{2\left(c+1\right)}}}{1-U} x \frac{Ca^{2}+Cp^{2}}{2}$$

And we have to keep in mind that:

* $U=\frac{R}{R\_{p}}, where R\_{p}=c/T\_{p}$
* $C\_{a}$ and $C\_{p}$ are the Coefficients of Variation
* Standard Deviation/Mean of the inter-arrival or processing times (assumed independent)

**4. Factors affecting Queue Length:**

1. Utilization effect; the queue length increases rapidly as $U$ approaches 1.

$$I\_{i}=\frac{U^{\sqrt{2\left(c+1\right)}}}{1-U}$$

1. Variability effect; the queue length increases as the variability in inter arrival and processing times increases.

$$\frac{Ca^{2}+Cp^{2}}{2}$$

1. While the capacity is not fully utilized, if there is variability in arrival or in processing times, queues will build up and customers will have to wait.

**E. Coefficient of Variations for Alternative Distributions:**

* $T\_{p}$: average processing time: $R\_{p}=c/T\_{p}$
* $T\_{a}$: average inter arrival time: $R\_{a}=1/T\_{a}$
* $S\_{p}$: standard deviation of the processing time
* $S\_{a}$: standard deviation of the inter arrival time

If processing time does not have a specific distribution such as exponential, Poisson, or constant, and if its average is$ T\_{p}$, and its standard deviation is $S\_{p}$, then its coefficient of variation, is $S\_{p}$ divided by $T\_{p}$.

The same is for inter-arrival time. If inter-arrival time is a general distribution, and by general distribution we mean it is not Poisson, not exponential, and it is not constant, and it has an average of $T\_{a}$and standard deviation of $S\_{a}$, then for inter-arrival time, coefficient of variation is $S\_{a}$divided by $T.$

This is explained by the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Inter arrival time or Processing Time distribution | General (G) | Poisson (M) | Exponential (M) | Constant (D) |
| Mean arrival time or Processing Time | $$T\_{p }(or T\_{a})$$ | $$T\_{p }(or T\_{a})$$ | $$T\_{p }(or T\_{a})$$ | $$T\_{p }(or T\_{a})$$ |
| Standard deviation ofInter arrival time or Processing Time | $$S\_{p }(or T\_{a})$$ | $$T\_{p }(or T\_{a})$$ | $$T\_{p }(or T\_{a})$$ | 0 |
| Coefficient of Variation of Inter arrival time or Processing Time | $$S\_{p }/T\_{p} (or S\_{a}/T\_{a})$$ | 1 | 1 | 0 |

Therefore exponential distribution shows the time between two events and Poisson distribution shows the number of events over a specific time period.

**Example 1:**

A sample of 10 observations on inter arrival times in minutes are:

10, 10, 2, 10, 1, 3, 7, 9, 2, 6 minutes.

$T\_{a}=$ average of sample = 6 minutes

$R\_{a}=\frac{1}{T\_{a}}=\frac{1}{6}$ arrivals per minute

$S\_{a}=$ standard deviation, found through the use of excel: 3.94

$C\_{a}=S\_{a}/T\_{a}$ = 3.94/6 = 0.66

**Example 1b:**

A sample of 10 observations on inter arrival times in minutes are:

7, 1, 7, 2, 8, 7, 4, 8, 5, 1 minutes.

$T\_{p}=$ average of sample = 5 minutes

$R\_{p}=\frac{1}{T\_{p}}=\frac{1}{5}$ arrivals per minute

$S\_{p}=$ standard deviation, found through the use of excel: 2.83

$C\_{p}=S\_{p}/T\_{p}$ = 2.83 /5 = 0.57

Using examples 1 and 1b, we are able to figure out how many people are in the waiting line and what is the safety capacity with the use of the queue length approximation formula. We can also compute what is the waiting time in the line and system. (Example 1c)

**Example 1c:**

$R\_{a}=\frac{1}{6}$ per minute or 10 per hour.

$R\_{p}=\frac{1}{5}$ per minute or 12 per hour.

$R\_{a}<R\_{p}.$ Since $R\_{a}=R, U=R/R\_{p}$ = (1/6)/(1/5) = 0.83

$C\_{a}$= 0.66 and $C\_{p}$= 0.57

By using the queue length approximation formula, we have the following:

$$I\_{i}=\frac{0.83^{\sqrt{2\left(1+1\right)}}}{1-0.83} x \frac{0.66^{2}+0.57^{2}}{2}$$

$I\_{i}=$ 1.56 passengers waiting in line.

To compute safety capacity, we use $Rs=R\_{p}-R$.

$Rs$ = (1/5) – (1/6) = 1/30 passengers per minute or (60)(1/30) = 2 passenger per hour.

We want to compute what is the waiting time for each person in the waiting line. The Little’s law states that *I = RT*, therefore $RT\_{i}=I\_{i}$. We can conclude that $T\_{i}=I\_{i}/R.$

$T\_{i}=$1.56/(1/6) = 9.4 minutes

The approximate waiting time in the line is 9.4 minutes. We can also calculate the waiting time in the whole system.

$$T=T\_{i}+T\_{p}$$

Because we know that $T\_{p}$ is 5 minute, the waiting time in the whole system is 14.4 minutes (5 + 9.4).

Using Little’s Law (*I = RT*) we are able to calculate the total number of people in the system.

$I=RT=\left(\frac{1}{6}\right)\left(14.4\right)=$ 2.4 people

We know that there are 1.56 people in the buffer. To calculate the total number of people in the processor, we add 1.56 to .83, the arrivals per minute to get 2.39.

Now let us suppose we have **two servers** instead of one, are we able to calculate the new waiting line and time? The following examples show how two servers change the waiting line and waiting time.

**Example 1c-a:**

$R\_{a}=\frac{1}{6}$ per minute or 10 per hour.

$R\_{p}=\frac{2}{5}$ per minute or 24 per hour.

$R\_{a}<R\_{p}.$ Since $R\_{a}=R, U=R/R\_{p}$ = (1/6)/(2/5) = (5/12) = 0.417

$C\_{a}$= 0.66 and $C\_{p}$= 0.57

By using the queue length approximation formula, we have the following:

$$I\_{i}=\frac{0.417^{\sqrt{2\left(2+1\right)}}}{1-0.417} x \frac{0.66^{2}+0.57^{2}}{2}$$

$I\_{i}=$ 0.076 passengers waiting in line.

To compute safety capacity, we use $Rs=R\_{p}-R$.

$Rs$ = (2/5) – (1/6) = 7/30 passengers per minute or (60)(7/30) = 14 passenger per hour.

We want to compute what is the waiting time for each person in the waiting line. The Little’s law states that *I = RT*, therefore $RT\_{i}=I\_{i}$. We can conclude that $T\_{i}=I\_{i}/R.$

$T\_{i}=$0.076/(1/6) = 0.46 minutes

The approximate waiting time in the line is 9.4 minutes. We can also calculate the waiting time in the whole system.

$$T=T\_{i}+T\_{p}$$

Because we know that $T\_{p}$ is 5 minute, the waiting time in the whole system is 5.46 minutes (5 + 0.46).

Using Little’s Law (*I = RT*) we are able to calculate the total number of people in the system.

$I=RT=\left(\frac{1}{6}\right)\left(5.46\right)=$ 0.91 people

We know that there are 0.076 people in the buffer. To calculate the total number of people in the processor, we add 0.076 to 0.417, the arrivals per minute to get 0.91.

**Example 2:**

A call center has 11 operators. The arrival rate of calls is 200 calls per hour. Each of the operators can serve 20 customers per hour. Assume inter-arrival time and processing time follow Poisson and Exponential, respectively. What is the **average waiting time** (time before a customer’s call is answered?)

By using the queue length approximation formula, we have the following:

$$I\_{i}=\frac{0.91^{\sqrt{2\left(11+1\right)}}}{1-0.91} x \frac{1^{2}+1^{2}}{2}$$

$$I\_{i}=6.89$$

Therefore:

$$I=RT=200\left(T\_{i}\right)=6.89$$

$T\_{i}=$ 2.1 minutes

**Example 2a:**

The average waiting time before a call is answered is 2.1 minutes. Now let us suppose the service time is constant. Will the waiting time be more or less than 2.1 minutes?

$$I\_{i}=\frac{0.91^{\sqrt{2\left(11+1\right)}}}{1-0.91} x \frac{1^{2}+0^{2}}{2}$$

$$I\_{i}=3.45$$

Therefore:

$$I=RT=200\left(T\_{i}\right)=3.45$$

$T\_{i}=$ 1.03 minutes

**Example 3:**

Vons contains 7 checkout stands. The arrival rate of customers is 175 per hour. Each of the checkout stands can serve 15 customers per hour. What is the average waiting time (time before a customer gets to the cashier?)

By using the queue length approximation formula, we have the following:

$$I\_{i}=\frac{0.92^{\sqrt{2\left(7+1\right)}}}{1-0.92} x \frac{1^{2}+1^{2}}{2}$$

$I\_{i}=8.95$ minutes

**Example 4:**

A small room has been assigned on a University campus, for honor students to have a quiet place to study by themselves. Since the room is so small, the only other use for it would be storage. The room is open 16 hours every day. Only one student can use the room at a time. The dean wants to know if the room is being fully utilized by students, or of the room should be changed to a storage space. The best SOM professor at the University was assigned to solve this problem. In his research, he found that students arrive at the room at a rate of 3 per hour following Poisson distribution. The students stay in the room an average of 15 minutes and deviation of 5 minutes.

1. What percentage of time is the room idle (rounded)?

$$=1-\frac{\left(3\*15\right)}{60}=25\%$$

1. What is the utilization rate of the room?

$$=\frac{\left(3\*15\right)}{60}=75\%$$

1. What is the average number of students in the waiting line?



1. How much time, on average, does a student spend in the waiting line?

$I=RT=3T=$ 2.25

$$T=.75 hrs=45 minutes$$

**Example 5:**

Yogurtland has one cashier. Every 5 minutes, a new customer comes in. It takes the cashier about 5 minutes to weigh the yogurt and ring up each customer. Yogurtland has been losing customers to the new frozen yogurt shop next door. In order to stop this, Yogurtland has implemented a new policy to pay customers waiting in line. The customer will get paid 1 dollar per minute of waiting time. A SOM professor has been recruited from a nearby university to analyze the cost of the new pay while you wait policy. Preliminary studies the SOM professor did, indicated that there are an average of .5 people waiting in line. Assume that arrivals follow Poisson and service time follows exponential distribution.

1. What is the capacity of the cashier per hour?

It takes a cashier 5 minutes per customer, therefore, 60 minutes / 5 minutes per customer = 12 customers.

1. What proportion of the time is the cashier busy?

$U=\frac{R}{R\_{p}}= $6/12 = 50%

1. On average, how long does a customer wait in line?



.05 \* 60 minutes = 3 minutes

1. What is the hourly cost of the new policy?

The hourly cost of the new policy is 1 \* 60 = $60. Because there are .5 customers in the line at all ties, .5 \* 60 = $30.

1. What is the most Yogurtland would be willing to pay for another cashier?

Answer = $30.

Another cashier would increase Rp to 24 customers per hour. If Rp is 24 per hour, there would be no waiting line because there are only 12 customer per hour. Yogurtland would only be willing to pay the amount that would be saved by having a new cashier.