Operations & Business Process Management

Prof. Apurva Jain

MSIS 503

Contents

Next

Where are we…(flow of the class)

What is the flow?: Flow-units, Resources, Activities,

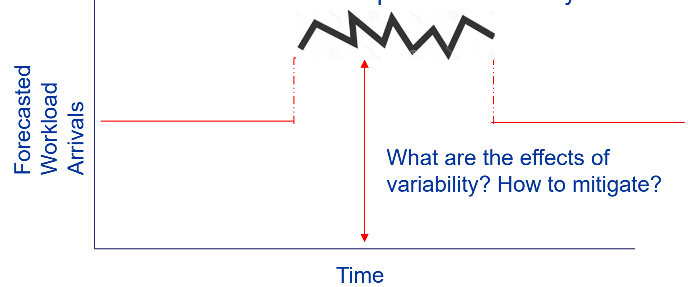
What are the symptoms?: Cost, Time, Quality

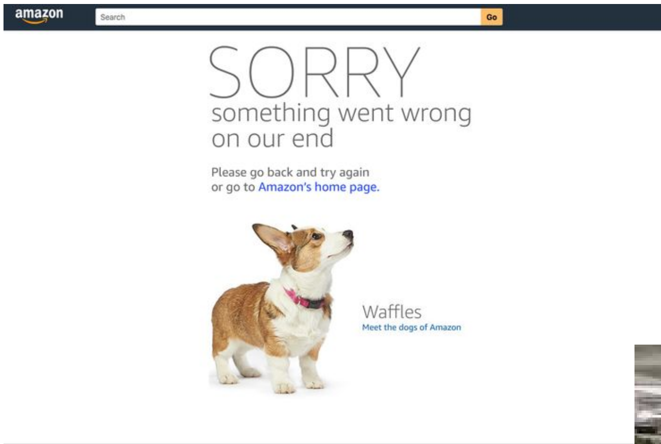
Activities
Which tasks to automate?
Which activities add value?
- scorecard
- value/waste

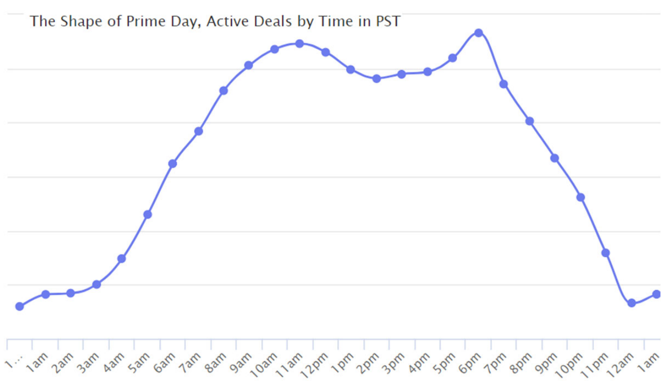
Flow-units
How many flow-units?
How to manage a variety of units?
- forecasting models
- errors, software
- improvement ideas

Resources
How many, where & what type of resources?
Ideas: locate bottleneck, add capacity, reallocate work or reassign resources, cut service times & provide flexibility

How do we incorporate uncertainty?

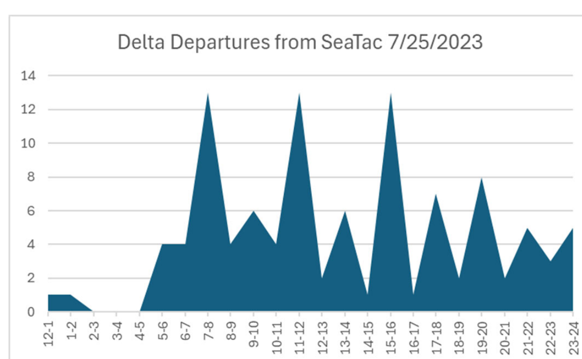






How many servers at 7 pm?
How many at 10 pm?

How many checkpoints at 3 pm?
How many at 5 pm?



Why is there waiting at non-peak times?

SEA TSA Wait Times

SEA, Seattle, WA

AM Times		PM Times	
12 am - 1 am	3.1 mins	11 am - 12 pm	14.4 mins
1 am - 2 am	11.8 mins	12 pm - 1 pm	17.8 mins
2 am - 3 am	20.5 mins	1 pm - 2 pm	19.6 mins
3 am - 4 am	29.3 mins	2 pm - 3 pm	15.4 mins
4 am - 5 am	21.8 mins	3 pm - 4 pm	11.6 mins
5 am - 6 am	19.1 mins	4 pm - 5 pm	18.6 mins
6 am - 7 am	13.1 mins	5 pm - 6 pm	17.1 mins
7 am - 8 am	12 mins	6 pm - 7 pm	12.8 mins
8 am - 9 am	10.1 mins	7 pm - 8 pm	11.6 mins
9 am - 10 am	18.4 mins	8 pm - 9 pm	24.8 mins
10 am - 11 am	17.6 mins	9 pm - 10 pm	18.7 mins
		10 pm - 11 pm	12.7 mins

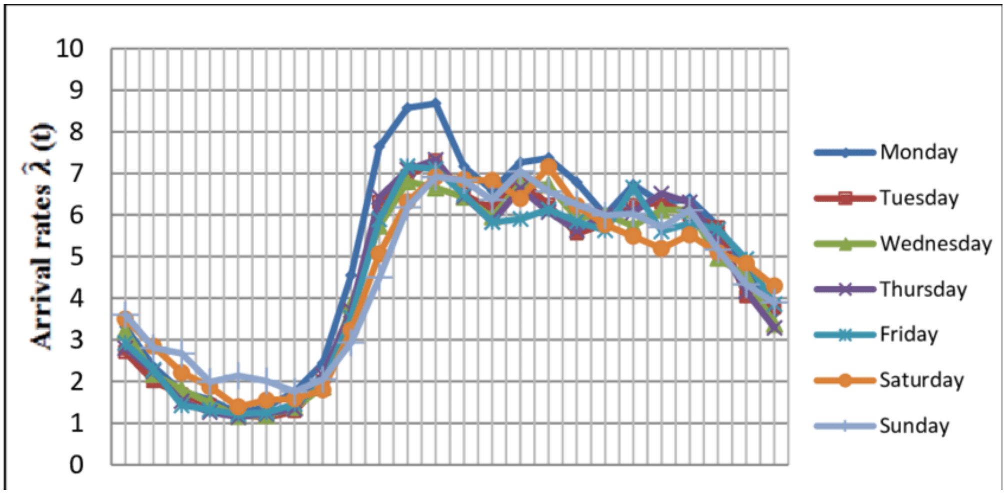
WHAT'S THE HURRY

A look at average wait times:

- Hospital emergency room: 4 hours, 7 minutes
- California Department of Motor Vehicles: 42 minutes, 32 seconds
- Main security line at Hartsfield-Jackson Atlanta International Airport during Monday morning rush: 25 minutes
- PRIMARY-CARE PHYSICIAN: 22 minutes
- McDonald's drive-through window: 2 minutes, 54 seconds

Sources: Press Ganey Associates; California DMV; Transportation Security Administration; QSR Magazine

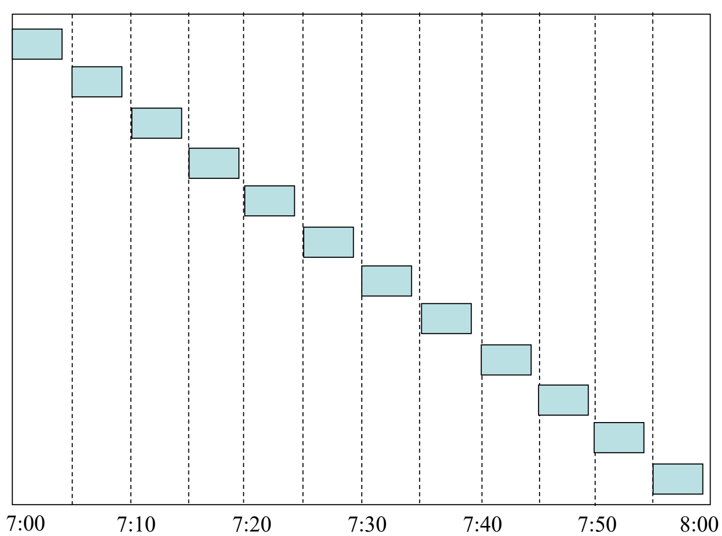
In many contexts, there is variability in the arrival rates.



We must think about how to determine capacity when there is variability.

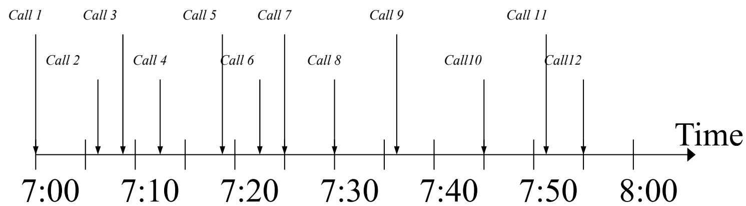
What should be the capacity?

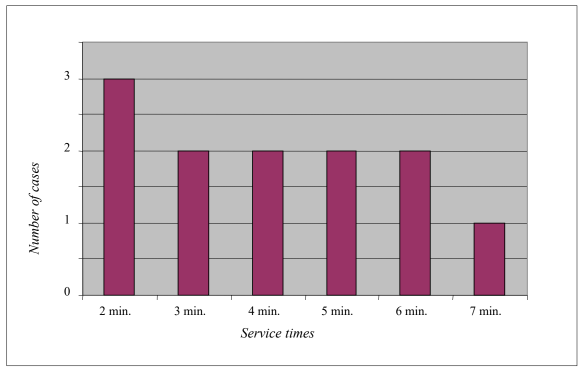
Call	Arrival Time	Service Time
1	0	4
2	5	4
3	10	4
4	15	4
5	20	4
6	25	4
7	30	4
8	35	4
9	40	4
10	45	4
11	50	4
12	55	4



A More Realistic Service Process

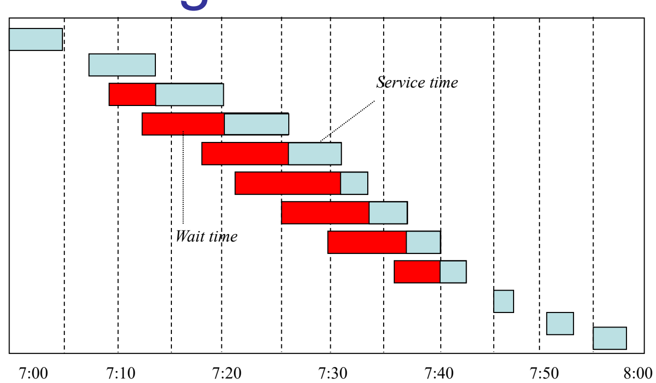
Call	Arrival Time	Service Time
1	0	5
2	7	6
3	9	7
4	12	6
5	18	5
6	22	2
7	25	4
8	30	3
9	36	4
10	45	2
11	51	2
12	55	3

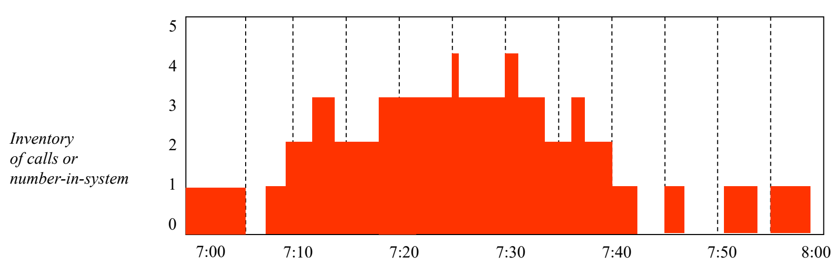


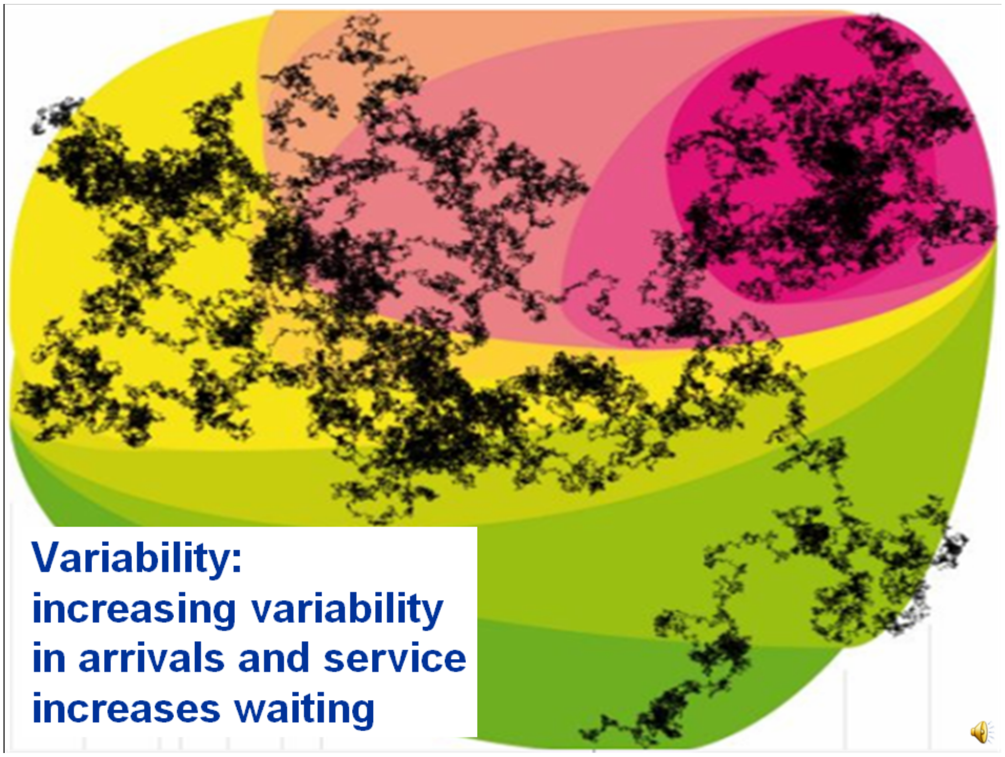


Variability Leads to Waiting Time

Calls	Arrival Time	Service Time
1	0	5
2	7	6
3	9	7
4	12	6
5	18	5
6	22	2
7	25	4
8	30	3
9	36	4
10	45	2
11	51	2
12	55	3

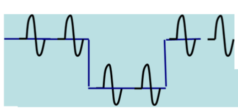






To make capacity decisions, we must understand different types of variability, measure variability and estimate its impact on waiting.

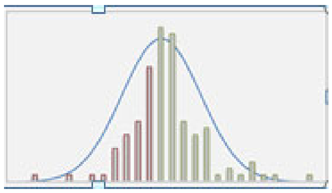
Classifying Types of Variability

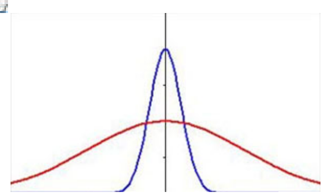


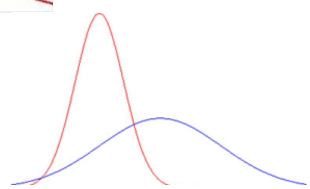
	Arrivals	Service
Long/ Medium-term (predictable)	Peak / Off-Peak High / Low Season	Turnover
Short-term (unpredictable)	Natural Randomness Customer Behavior Errors in Forecasting Unforeseeable events	Customization requests Uneven skill levels Non-standard process Resource availability

Do you think about variability in your industry?

Measuring Variability







- Use historical data to compute average and standard deviation

- Standard deviation divided by average gives a normalized measure of variability.

- The ratio described above is known as Coefficient of Variation

To measure service variability, observe changes in service time.

Service time

Observations (min): **5 6 7.............................2 2 3**

Average → $average_s = 4$

Standard Deviation → $std.dev._s = 1.73$

Coefficient of variation of service times C_s

$$C_s = std.dev._s \,/\, average_s = 1.73/4 = 0.42$$

Consistent Service (fixed service time) $C_s=0$

Uncertain Service (variable service time) $C_s=0.5$

To measure arrival variability, observe changes in inter-arrival times.

stopwatch (min:sec) readings for arrivals:

0:00 7:00 9:00 12:00 18:00……

Interarrival time Observations (min): **7 2 3 6 ……**

Average: average$_a$ = 5

Standard Deviation: std.dev.$_a$ = 1.95

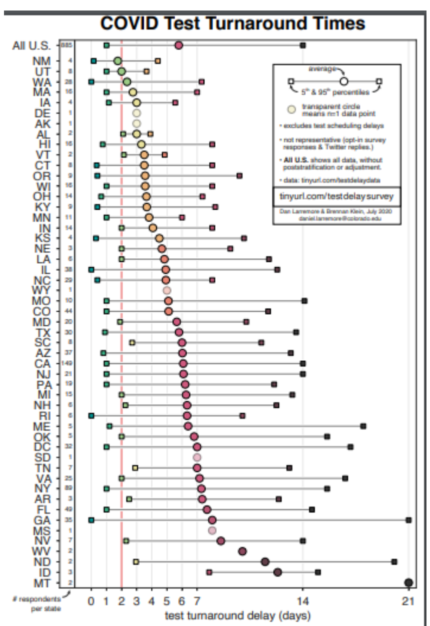
Coefficient of variation of interarrival times Ca

Ca = std.dev.$_a$ / average$_a$ = 1.95/5 = 0.39

Scheduled arrivals (fixed inter-arrival time) C_a=0

Walk-ins (variable inter-arrival time) C_a=1

Pandemic: Testing



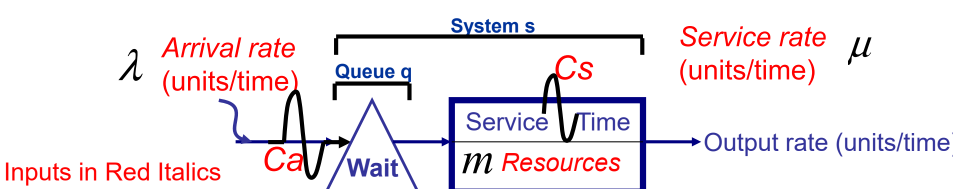
Coronavirus testing delay: https://www.nytimes.com/2020/08/04/us/virus-testing-delays.html In Brunswick, Ga., lab technicians are working around the clock. Machines churn out results in 50 minutes to four and a half hours, but there are so many orders that the labs cannot keep up. Regular shipments of chemicals needed to test for the virus do not last even a week, so pathologists have begun to carefully dole out their supplies.

"We literally ration tests," said Dr. Patrick Godbey, the president of the College of American Pathologists and the director of two labs in the Brunswick area. He estimated that for every test his labs are able to perform, they have to send three to national commercial laboratory companies.

"Why keep a backlog?" he said. "Tomorrow is going to be just as busy."

Ref: Pinterest ks lam

Incorporating variability in flow model



Utilization =
Output rate /
(m*service rate)

Flow Performance measures related to time:

Wait Time: average time a customer (flow-unit) waits before entering service

Service Time: average time a customer spends in service
(time it takes for one server (resource) to work on one flow-unit)

Lead Time: wait time + service time
(also known as time-in-system or turnaround time)

Queueing Model to compute wait time

Uses WaitTimes spreadsheet to compute wait on a single station.

#	Arrival Time	Service Time
1	0	5
2	7	6
3	9	7
4	12	6
5	18	5
6	22	2
7	25	4
8	30	3
9	36	4
10	45	2
11	51	2
12	55	3

Average → $\text{average}_a = 5$ | Standard Deviation → $\text{std.dev.}_a = 1.95$

Coefficient of variation of interarrival times Ca

$C_a = \text{std.dev.}_a / \text{average}_a = 1.95/5 = 0.39$

Average → $\text{average}_s = 4$ | Standard Deviation → $\text{std.dev.}_s = 1.73$

Coefficient of variation of service times C_s

$C_s = \text{std.dev.}_s / \text{average}_s = 1.73/4 = 0.42$

Enter Inputs here:	
Arrival rate=	0.2
Capacity of one resource (service rate)=	0.25
Number of resources=	1
Utilization=	0.8
Coefficient of variation of arrivals Ca=	0.39
Coefficient of variation of service Cs=	0.42
Read Outputs here:	
Waiting Time=	2.628
Service Time=	4
Total lead time=	6.628
Number in waiting line=	0.5256
Number in system=	1.3256

Queueing Model to compute wait

Uses WaitTimes spreadsheet to compute wait on a single station.

If 1 customer arrives per minute, service time is 1.5 minutes, there are two serving resource, and C_a=1, C_s=1, what is the wait time?

Enter Inputs here:	
Arrival rate=	1
Capacity of one resource (service rate)=	0.66666667
Number of resources=	2
Utilization=	0.75
Coefficient of variation of arrivals Ca=	1
Coefficient of variation of service Cs=	1

Read Outputs here:	
Waiting Time=	1.97707815
Service Time=	1.5
Total lead time=	3.47707815
Number in waiting line=	1.97707815
Number in system=	3.47707815

How to solve problems

Read the text of the problem and list all resource-type or stations and determine service time for each resource-type. Note arrival rate.

If any information about variability is given (if words like standard deviation, Poisson, exponential are mentioned), note or compute Ca and Cs. If numbers are not clearly given, default values for both are 1.

For each station, enter five inputs in spreadsheet (arrival rate, service rate, number of resources, Ca, Cs) and read wait times and other outputs. Adding wait times and service times across all stations gives total lead time or turnaround time.

Consider possible improvements in variability and capacity and compute the effect on wait times.

Practice problem:
The local office of an investment consulting firm in Seattle receives 12 walk-in clients in the 5-hour walk-in window every day, and the arrival is Poisson. Its only team works together on a client and takes 20 minutes to process each client, with the processing time being exponentially distributed. What is the wait time? Turnaround time (lead time)?

Practice: Wait Time Computations

The local office of an investment consulting firm in Seattle receives 12 walk-in clients in the 5-hour walk-in window every day, and the arrival is Poisson. Its only team works together on a client and takes 20 minutes to process each client, with the processing time being exponentially distributed. What is the wait time? Turnaround time (lead time)?

Practice: Wait Time Computations

The local office of an investment consulting firm in Seattle receives *12 walk-in clients in the 5-hour* walk-in window every day, and the arrival is *Poisson*. Its *only team* works together on a client and takes *20 minutes* to process each client, with the processing time being *exponentially* distributed. What is the wait time? Turnaround time (lead time)?

Enter Inputs here:	
Arrival rate=	2.4
Capacity of one resource (service rate)=	3
Number of resources=	1
Utilization=	0.8
Coefficient of variation of arrivals Ca=	1
Coefficient of variation of service Cs=	1
Read Outputs here:	
Waiting Time=	1.33333333
Service Time=	0.33333333
Total lead time=	1.66666667
Number in waiting line=	3.2
Number in system=	4

What would be the impact of reducing variability?

By making either Ca or Cs number smaller, we can see that wait time decreases.

Practice: Using Spreadsheet Inputs

Continuing from the previous slide ---------------------->.

How will the wait time change if

(i) Cs decreases to 0.5, OR

(ii) Resources increase to 2, OR

(iii) Service rate increases to 6?

Enter Inputs here:	
Arrival rate=	2.4
Capacity of one resource (service rate)=	3
Number of resources=	1
Utilization=	0.8
Coefficient of variation of arrivals Ca=	1
Coefficient of variation of service Cs=	1
Read Outputs here:	
Waiting Time=	1.33333333
Service Time=	0.33333333
Total lead time=	1.66666667
Number in waiting line=	3.2
Number in system=	4

Wait time
(i) decreases
(ii) decreases
(iii) decreases

Wait time
(i) increases
(ii) increases
(iii) increases

Wait time
(i) increases
(ii) decreases
(iii) decreases

Wait time
(i) increases
(ii) decreases
(iii) increases

Practice:
Using Variability Inputs Ca & Cs

Continuing from the previous question→..

Enter Inputs here:		
	Arrival rate=	2.4
	Capacity of one resource (service rate)=	3
	Number of resources=	1
	Utilization=	0.8
	Coefficient of variation of arrivals Ca=	1
	Coefficient of variation of service Cs=	1
Read Outputs here:		
	Waiting Time=	1.33333333
	Service Time=	0.33333333
	Total lead time=	1.66666667
	Number in waiting line=	3.2
	Number in system=	4

The standard deviation of service time is computed to be 0.3 hour. What will be the average waiting time?

Continuing from the previous part, the standard deviation of interarrival times is computed to be 0.5 hour. What will be the average waiting time?

Practice: Capacity to meet Time Targets

Target is: average time-in-system<20 min.
Two improvement proposals are on table:

A. Add another team.

B. Invest in technology to cut service time by 50%.

Practice: Capacity

…. Continuing the investment consulting firm example from earlier… Suppose we want to analyze the case for adding another team at the cost of $100 per hour. Our estimate of how much a customer's wait time cost is $25 per hour. Should we add another team?

Recall earlier question:

Enter Inputs here:	
Arrival rate=	2.4
Capacity of one resource (service rate)=	3
Number of resources=	1
Utilization=	0.8
Coefficient of variation of arrivals Ca=	1
Coefficient of variation of service Cs=	1
Read Outputs here:	
Waiting Time=	1.33333333
Service Time=	0.33333333
Total lead time=	1.66666667
Number in waiting line=	3.2
Number in system=	4

Number of resources changes to 2.

Waiting Time=	0.07360165
Service Time=	0.33333333
Total lead time=	0.40693498

Practice: Using Spreadsheet Inputs

Continuing from the previous slide. -----------------------------→

Enter Inputs here:	
Arrival rate=	2.4
Capacity of one resource (service rate)=	3
Number of resources=	1
Utilization=	0.8
Coefficient of variation of arrivals Ca=	1
Coefficient of variation of service Cs=	1
Read Outputs here:	
Waiting Time=	1.33333333
Service Time=	0.33333333
Total lead time=	1.66666667
Number in waiting line=	3.2
Number in system=	4

How will the wait time change if
(i) Cs decreases to 0.5, OR
(ii) Resources increase to 2, OR
(iii) Service rate increases to 6?

i) Waiting Time decreases to 0.8333

ii) decreases to 0.0736

iii) decreases to 0.1111

Practice: Using Variability Inputs

Continuing from the previous question..

The standard deviation of service time is computed to be 0.3 hour. What will be the average waiting time?

Because the service time standard deviation is available, we should not use the default value of Cs=1.
The average service time is given as 20 minutes = 1/3 hour
Cs = standard deviation of service time / average service time = 0.3/(1/3)=0.9
Using Cs=0.9 (in place of Cs=1) in the spreadsheet and keeping everything else the same, we get average wait time = 1.2067 hours

Continuing from the previous part, the standard deviation of interarrival times is computed to be 0.5 hour. What will be the average waiting time?

Because the interarrival time standard deviation is available, we should not use the default value of Ca=1.
The average interarrival time is given as 5/12 hour (from 12 arrivals in 5 hours)
Ca = standard deviation of interarrival time / average interarrival time = 0.5/(5/12)=1.2
Using Ca=1.2 (in place of Ca=1) in the spreadsheet and keeping everything else the same, we get average wait time = 1.5 hours

Practice: Capacity to meet Time Targets

Target is: average time-in-system<20 min.
Two improvement proposals are on table:

A. Add another team.

Number of resources changes to 2.

From Spreadsheet

Waiting Time=	0.07360165
Service Time=	0.33333333
Total lead time=	0.40693498

Time-in-system is 0.4069 hr = 24.41 min.

B. Invest in technology to cut service time by 50%.

Number of resources remains 1. Service rate changes to 6 per hr.

From Spreadsheet

Waiting Time=	0.11111111
Service Time=	0.16666667
Total lead time=	0.27777778

Time-in-system is 0.278 hr = 16.68 min.
Option B is needed to meet the time target.

Practice: Capacity

…. Continuing the investment consulting firm example from earlier…
Suppose, we want to analyze the case for adding another team at the cost of $100 per hour. Our estimate of how much a customer's wait time cost is $25 per hour. Should we add another team?

With the additional team:

Cost of providing service will increase by $100 per hour.

An average customer will wait less; its wait time will decrease by
=1.333 hr – 0.073 hr = 1.26 hr
Therefore the cost of waiting for an average customer decreases by
= 1.26 hr*$25 per hr = $31.50 per customer
Because arrival rate is 2.4 per hr; on an average 2.4 customer per hr pass through the system. Therefore waiting-cost decrease per hour
= $31.50 per customer * 2.4 customers per hour = $75.6 per hr

Increase in service cost is more than decrease in waiting cost.
We should not add another team.

A Moment of Reflection

"Did I keep you waiting long?"

What is the main reason for unpredictability in your workflow?

Is there a way to measure it?

Operations & Business Process Management

Prof. Apurva Jain

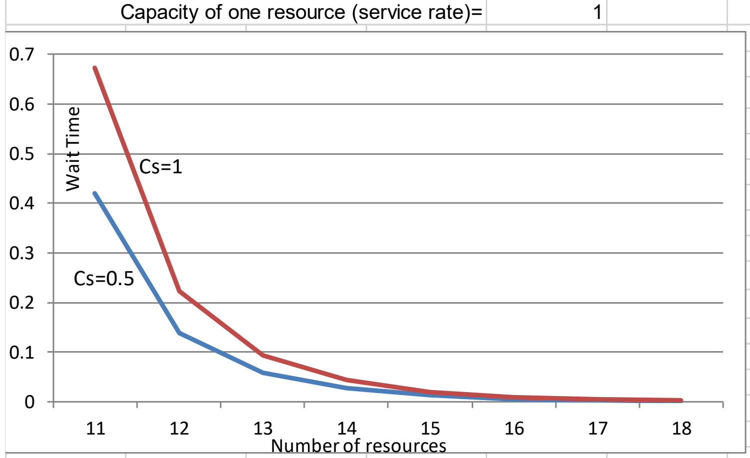
MSIS 503

Contents

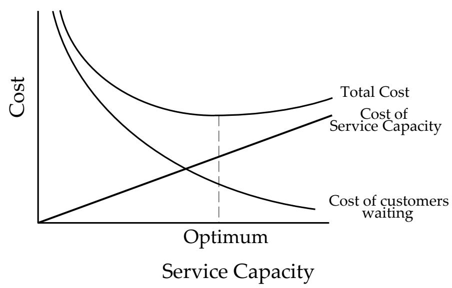
Next

Graph from the spreadsheet

Arrival rate=	9.99
Capacity of one resource (service rate)=	1



To Optimize Capacity & Evaluate Options Understand Capacity-Waiting Trade-off



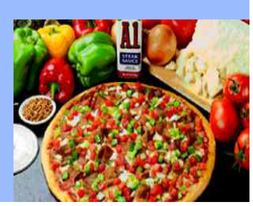
Various Estimates of “Cost” of Waiting

- $ value of customer wait time
- $ value of lost business due to excessive wait
- $ value of “payments” to customers due to waiting
- A target average waiting time
- A probabilistic waiting target (x% of customers must receive service within y minutes)

Cost of Waiting is defined as the $ value of making a customer wait an extra time-unit. The graph show the trade-off between the cost of capacity and cost of waiting.

All orders may not be called in turn
Some pizza takes longer to burn
But while you wait, keep this in mind
If number six is called, number five can not be far behind.

Sign at Pietro’s Pizza



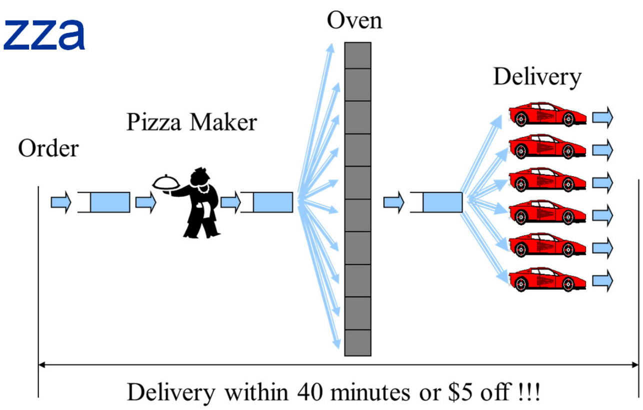
4 – Puget Sound Energy
Puget Sound Energy customer service guarantee.
Customer Service Guarantees
In addition to tracking our performance, we offer two service guarantees. We commit to keeping scheduled appointments and to restoring power outages as soon as we can. Here's how our service guarantees work:
•If we don't keep an appointment to install new service, reconnect existing service or inspect natural gas equipment, you'll receive a $50 credit on your PSE bill.

- Check out our electric appointment service guarantee and natural gas appointment service guarantee for details.

•If your electric service is out for 120 consecutive hours or longer, you may be eligible to receive a $50 credit on your PSE bill.

- Read our electric service restoration guarantee details.

Pronto Pizza



Arrival rate	Pizza station	Oven station	Delivery
12 per hr	1 pizza maker 3 min / pizza	10 racks 12 min/pizza	6 drivers (5+5)min / pizza delivery

Where is the bottleneck?

____

What is the lead time?

____

Pronto Pizza: What does intuition say?

I doubt if we will ever have to pay $5 off.
Your guess: have to pay to…..% of customers.

Almost 0%

1%

3%

5%

Pronto Pizza: Computing Lead Time

	Pizza station	Oven station	Delivery
Arrival rate	12/hr	12/hr	12/hr
Service rate	20/hr	5/hr	6/hr
Resources	1	10	6
Wait Time			
Lead Time			

What is the average total lead-time (turnaround time)

= ______________________ min

(Ignoring the very small delivery delay)

Pronto Pizza: Computing Lead Time

	Pizza station	Oven station	Delivery
Arrival rate	12/hr	12/hr	12/hr
Service rate	20/hr	5/hr	6/hr
Resources	1	10	6
Wait Time	0.075hr*60min = 4.5min	0.0001hr =0.006min	0.002hr =0.12min
Lead Time	4.5+3 = 7.5min	0.006+12= 12.006 min	0.12+5 =5.12min

What is the average total lead-time (turnaround time)

= __7.5+12.006+5.12 = 24.626 min

Pronto Pizza: Cost of Waiting

Target average total lead-time =40 min means that
(lead-time at Pizza station)+(0+12)+(0+5) = 40 min
(Ignoring the very small delivery delays at oven and delivery)
Target lead-time at Pizza station = 40-17=23
Prob. (total lead time > 40) = Prob.(lead time at Pizza station>23)

Probability that lead time at station is > t is given by $e^{-\text{(service rate - arrival rate)} * t}$

Probability that lead time at pizza station is >23 min:

$$= e^{-(20-12)(23/60)}$$

$$= 0.0465$$

Cost of $5-off guarantee = 0.0465*$5
= $ 0.2325 per customer

= $ 0.2325 * 12 customer per hr = $2.79 / hr

Pronto Pizza:
Optimize Capacity / Evaluate Options

Given the current context, if we can increase pizza station capacity to 25/hr at the cost of $2 /hr, would you go for it?

Pronto Pizza:
Optimize Capacity / Evaluate Options

Given the current context, if we can increase pizza station capacity to 25/hr at the cost of $2 /hr, would you go for it?

New service rate = 25/hr,
New probability of Pizza station lead time >23 min
$= e^{-(25-12)*(23/60)} = 0.0069$
Cost of $5-off guarantee = 0.0069*$5
= $ 0.0344 per customer

= $ 0.0.0344 * 12 customer per hr = $0.41 / hr

Save 2.79-0.41 = $ 2.38 per hour in guarantee cost at the extra cost of $ 2 per hour. Go for it.

Examples of “Cost” of waiting

Match Bullets from the two lists:

- Call centers track “abandon rate” ~ 2- 5%
- Royal Cornwall Hospital, UK aims at fixing a specialist appointment within 18 weeks!
- A Chipotle study confirms the fast food industry maxim “seven-second reduction in waiting times increases a chain’s market share by 1%.”
- Ameritrade guarantee “if a trade takes more than 10 sec. to execute, no commission.”

Various Estimates of “Cost” of Waiting

- $ value of customer wait time
- $ value of lost business due to excessive wait
- $ value of “payments” to customers due to waiting
- A target average waiting time
- A probabilistic waiting target (x% of customers must receive service within y minutes)

SFGATE

BAY AREA & STATE

Thousands of people are paying a startup $94 to wait in line for them at the DMV

Filipa Ioannou | on January 24, 2019

https://www.sfgate.com/bayarea/amp/dmv-line-waiting-service-sf-yogov-california-id-13559898.php

A Moment of Reflection

Is there a way to measure the cost of customer waiting in your industry / company?

Operations & Business Process Management

Prof. Apurva Jain
MSIS 503

Contents

1. Identify Flow
2. Automate Flow
3. Forecast Demand
4. Balance Capacity
   - 4.1 Capacity Definitions
   - 4.2 Balance
   - 4.3 Flow Improvement
5. Mitigate Variability's Impact
   - 5.1 Variability & Wait Measurement
     - Ca, Cs, Wait Time Spreadsheet
   - 5.2 Cost vs. Time Trade-off
     - Capacity Wait trade-off
     - Turnaround time calculation
   - 5.3 Idea Matrix, Simulation

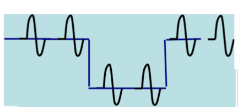
Next

6. Configure Capacity for variety
7. Collaborate across Flow..

..

To reduce wait, we should reduce variability. What are the main ideas for variability reduction?

Variability Reduction



	Arrivals	Service
Long/ Medium-term	Smooth the peaks: Incentives, Information	Staffing Policies: Continuity, Reduce turnover
Short-term	Consistent arrivals: Appointments Reservations	Standardize Service: Automation / Scripting Customer participation

Variability Reduction: Examples

Check the right box

In construction, a remodeling firm advertises a kitchen special in winter.

In non-profit work, a charity has an app to schedule evening meetings between clients and volunteer mentors.

In a finance advising firm, clients are asked to fill out a form before the first appointment .

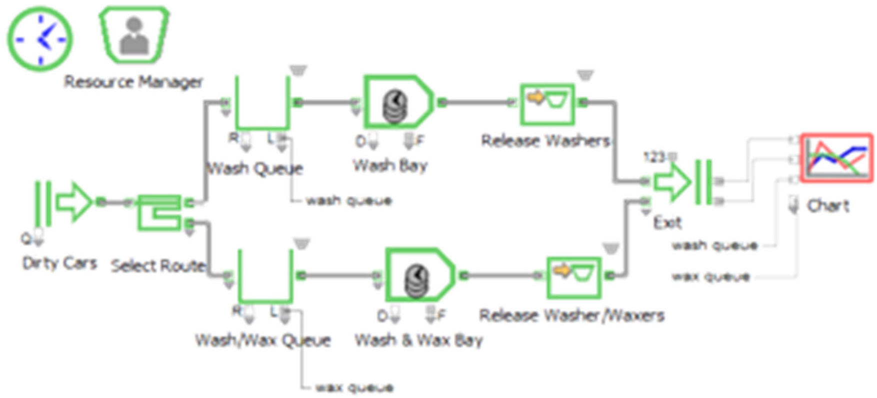
Even if we cannot reduce actual wait, We can manage the perception of wait.

1. Occupied Time
2. Get Started
3. Reduce Anxiety
4. Known waits
5. Explain waits
6. Fair waits

Waiting time greater than:	Actual probability	Perceived probability
3 minutes	0.58	0.72
5 minutes	0.29	0.45
7 minutes	0.12	0.49
10 minutes	0.03	0.40

Survey data from Boston red line

Digital Twins / Simulation Software



Improvement Ideas

- Is there significant waiting in your flow? Then, think about a way to measure variability in arrivals and services, even if it is simply noticing the *range*.
- As long as there is significant variability, we must plan for extra capacity as a buffer. If capacity is just equal to workload, variability will result in long waits.
- Suggest adding some extra capacity at stations where utilization and variability are high. If possible, use technical methods to optimize how much capacity to add.
- Suggest decreasing utilization at stations where variability is high. By reducing service time, adding resources or by reallocation of work.
- Generate ideas for making arrivals more consistent; provide incentives or information for customers to change their arrival behavior.
- Identify technology solutions that can provide information about future customer arrivals.
- Where service variability is high, make sure that there is consistency in resource type and availability. Ideas for standardization will help a lot.

Key points and takeaways

- In service contexts like contact centers and retail, waiting time is one of the most important drivers of customer satisfaction. Capacity decisions in these contexts must account for waiting.
- Waiting is driven by variability. Variability can be either in arrivals or in service. We measure variability in terms of Ca and Cs.
- Given inputs of arrival rate, service rate, number of servers, Ca and Cs, we can use the spreadsheet (or formulas) to compute:
  - wait time, lead-time, number-in-waiting and number-in-system
- Given a target wait time, the graph and the spreadsheet model can be used to make capacity decisions.
- If a "cost of waiting" can be estimated, we can make capacity decisions by considering the trade-off between the cost of waiting and the cost of providing capacity. Pronto pizza problem provides an example how to compute turnaround time and how to make a capacity decision in a setting with multiple stations.
- To improve waiting times, we must reduce variability. A classification of variability reduction ideas can be used as a framework to generate our own. We can also manage the customer's perception of waiting.

Operations & Business Process Management

Prof. Apurva Jain
MSIS 503

Customer satisfaction is sensitive to waiting. Waiting is driven by variability. To address waiting problems in flow, measure and reduce variability.

Contents