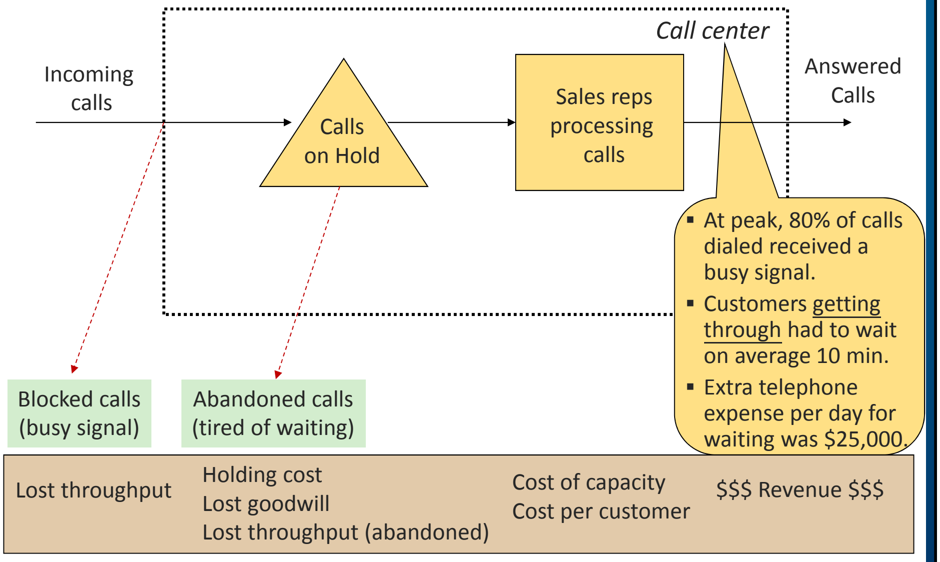
Chapter 8 Variability and Waiting Time Problems

- A Call Center Example
- Arrival Process and Service Variability
- Predicting Waiting Times
- Waiting Line Management

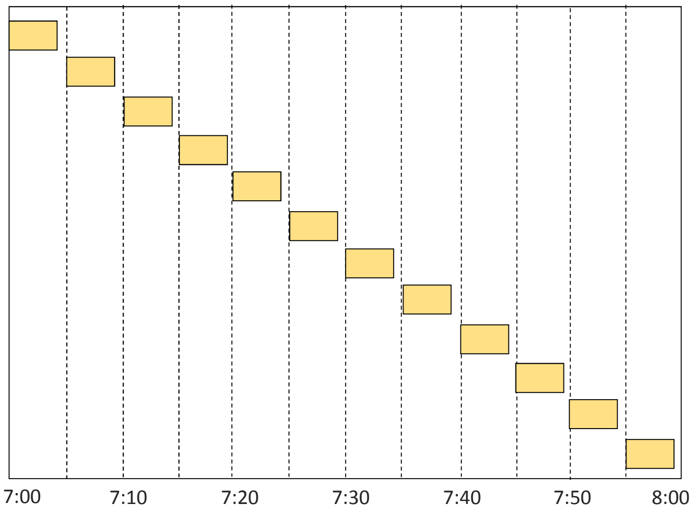
8.1 The Operation of a Typical Call Center



3

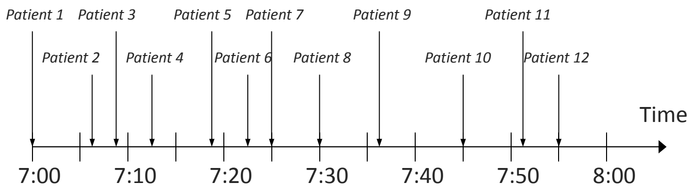
A "Perfect or Somewhat Odd" Call Center

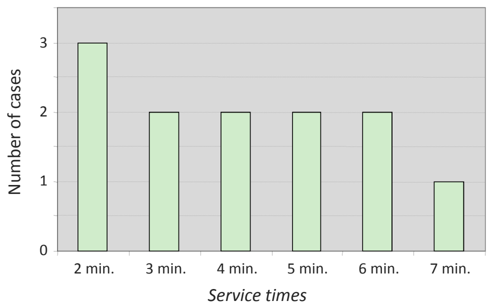
Patient	Arrival Time	Service Time
1	0	4
2	5	4
3	10	4
4	15	4
5	20	4
6	25	4
7	30	4
8	35	4
9	40	4
10	45	4
11	50	4
12	55	4



A More Realistic Service Process

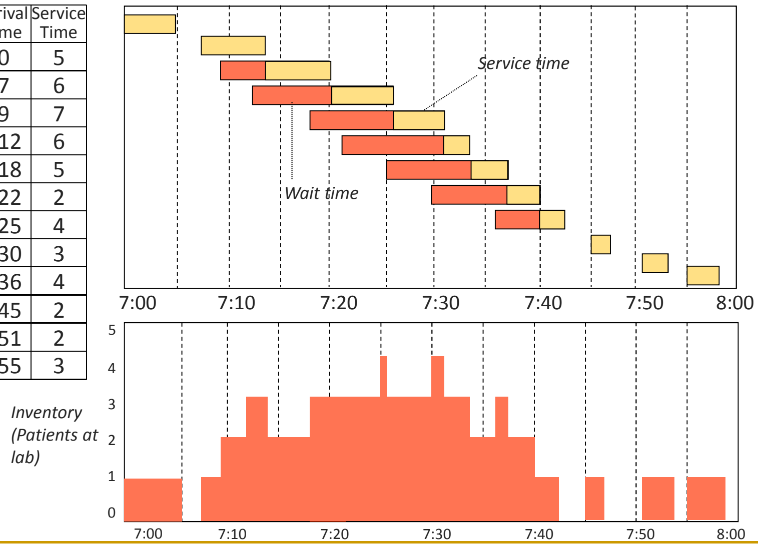
Patient	Arrival Time	Service Time
1	0	5
2	7	6
3	9	7
4	12	6
5	18	5
6	22	2
7	25	4
8	30	3
9	36	4
10	45	2
11	51	2
12	55	3





Variability Leads to Waiting Time

Patient	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



6

Observations

Most customers have to wait, although, on average, there is plenty of capacity in the call center.

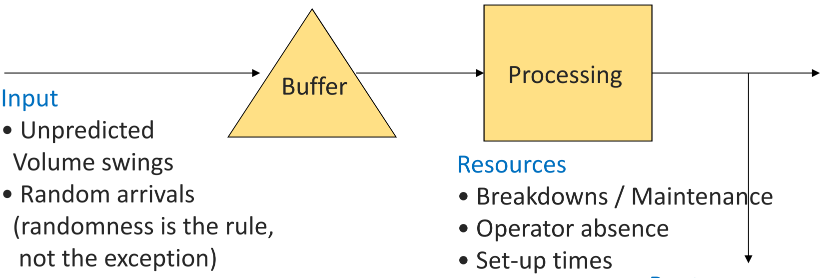
The call canter is unable to provide consistent service quality.

If customer abandon calls after long waits, the call center loses customer goodwill and revenue.

7

8.2 Variability: Where does it come from?

Tasks
- Inherent variation
- Lack of SOPs
- Quality (scrap / rework)



Input
- Unpredicted Volume swings
- Random arrivals (randomness is the rule, not the exception)
- Incoming quality
- Product Mix

Resources
- Breakdowns / Maintenance
- Operator absence
- Set-up times

Routes
- Variable routing
- Dedicated machines

8

Ignoring Variability Leads to Problems

Random arrivals and varying demands are common in services.

In the presence of variability, one cannot estimate the process performance based on averages.

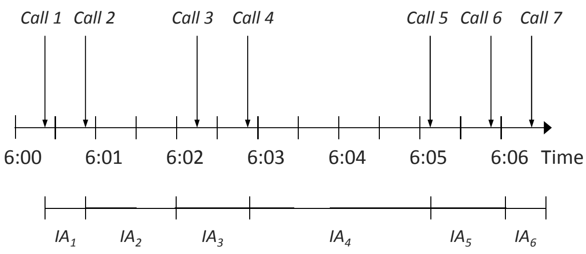
Q: Why does variability not average out over time?

A: You cannot inventory services — Capacity can never run ahead of demand.

9

8.3 Analyzing an Arrival Process

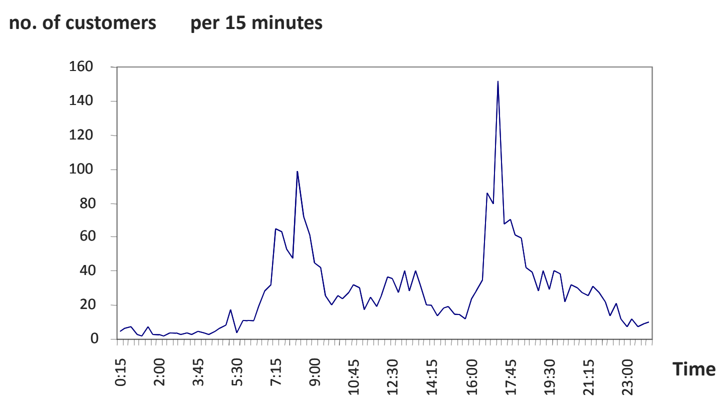
Call	Arrival Time, AT_i	Interarrival Time $IA_i = AT_{i+1} - AT_i$
1	6:00:29	00:23
2	6:00:52	01:24
3	6:02:16	00:34
4	6:02:50	02:24
5	6:05:14	00:36
6	6:05:50	00:38
7	6:06:28	



$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Mean}}$$

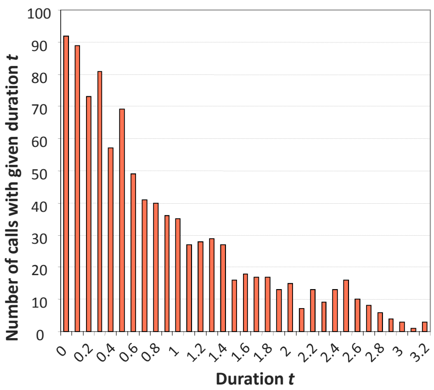
Seasonality over the Course of a Day

An arrival process is not stationary if the average number of arrivals in any given time interval is not fixed over the entire time period.



Exponential Distribution

Customers arriving independently from each other follow exponential inter-arrival times. $\Rightarrow$ Random (Poisson) arrivals



$$f(t) = \frac{1}{a} e^{-\frac{t}{a}}$$

a = average inter-arrival time

12

How to Analyze a Demand/Arrival Process

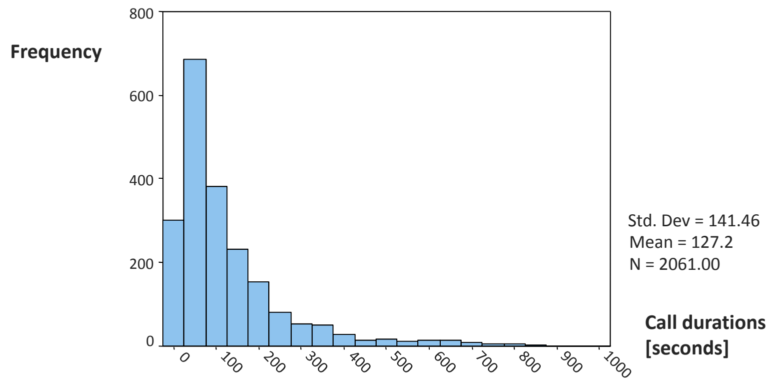
Stationary Arrivals?

- YES → Random Arrivals?
  - YES:
    - Compute a: average interarrival time
    - CV_a=1
    - All results of chapters 8 and 9 apply
  - NO:
    - Compute a: average interarrival time
    - CV_a= St.dev. of interarrival times/a
    - All results of chapter 8 apply
    - Results of chapter 9 do not apply, require simulation or more advanced tools
- NO → Break arrival process up into smaller time intervals

$$\text{mean} = a \qquad CV_a = \frac{\text{Standard deviation of interarrival time}}{\text{Average interarrival time}}$$

13

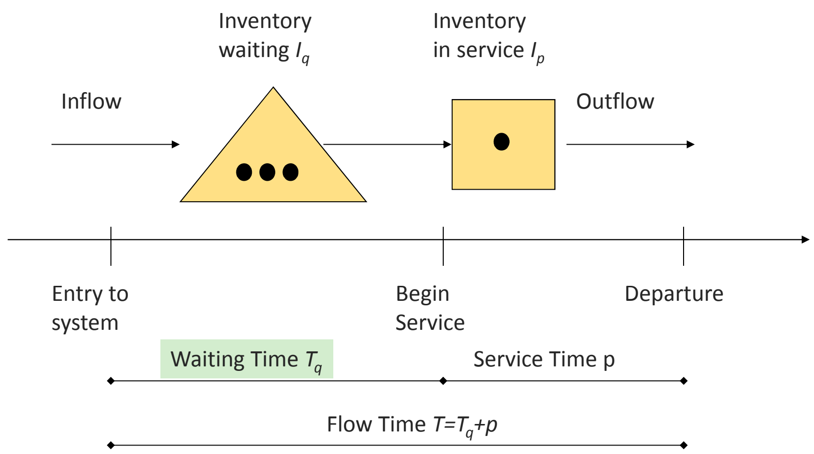
8.4 Service Times in Call Center



$$\text{mean} = p \qquad CV_p = \frac{\text{Standard deviation of activity time}}{\text{Average activity time}}$$

14

8.5 Predicting Average Waiting Time: One Server



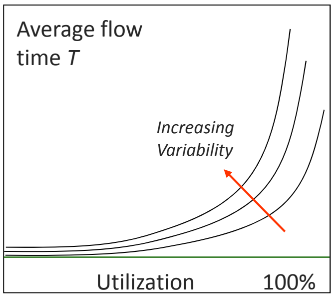
15

The Waiting Time Formula

$$\text{utilization} = \frac{\text{flow rate}}{\text{capacity}} = \frac{1/a}{1/p} = \frac{p}{a} < 100\%$$

$$\textit{Time} \text{ in queue} = \text{Activity Time} \times \left(\frac{utilization}{1 - utilization}\right) \times \left(\frac{CV_a^2 + CV_p^2}{2}\right)$$

Service time factor (Activity Time); Utilization factor $\left(\frac{utilization}{1-utilization}\right)$; Variability factor $\left(\frac{CV_a^2+CV_p^2}{2}\right)$

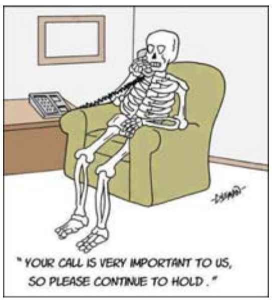


16

Reducing average waiting time does not guarantee customer satisfaction.

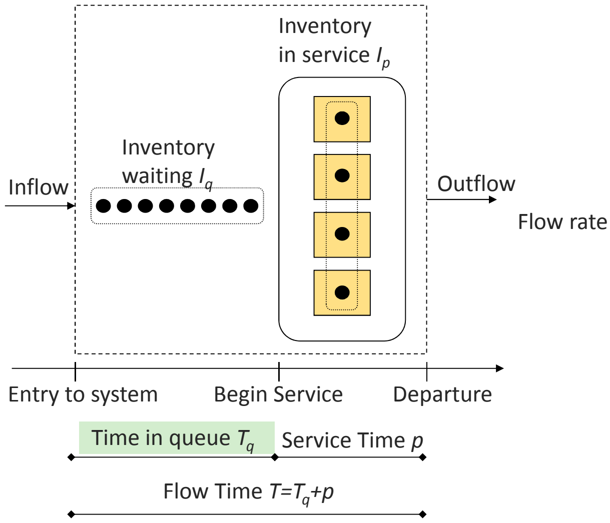
A small percentage of customers may experience long waits and complain bitterly.

Solution: service guarantee and/or service recovery



17

8.6 Multiple, Parallel Resources with One Queue



Waiting Time for Multiple, Parallel Resources

Under the assumption that

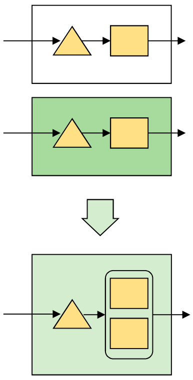
$$\text{Utilization} = \frac{\text{Flow rate}}{\text{Capacity}} = \frac{1/\text{interarrival time}}{m \times (1/\text{activity time})} = \frac{1/a}{m/p} = \frac{p}{a \times m} < 1$$

we approximate the average waiting time as

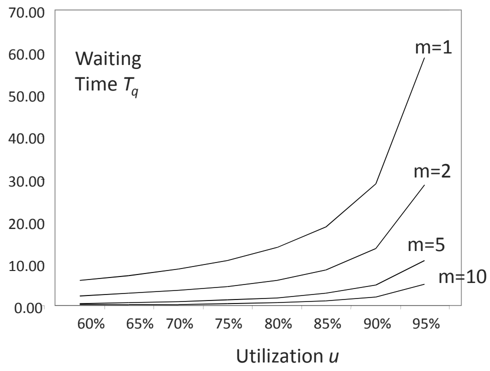
$$\text{Time in queue} = \left(\frac{\textit{Activity}\ \text{time}}{m}\right) \times \left(\frac{\textit{utilization}^{\sqrt{2(m+1)}-1}}{1-\textit{utilization}}\right) \times \left(\frac{CV_a^2 + CV_p^2}{2}\right)$$

The Power of Pooling

Independent Resources
2x(m=1)

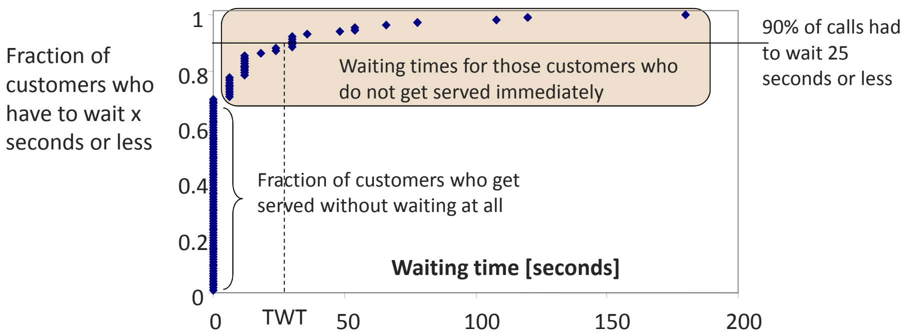


Pooled Resources
(m=2)



Pooling benefits are lower if queues are not truly independent

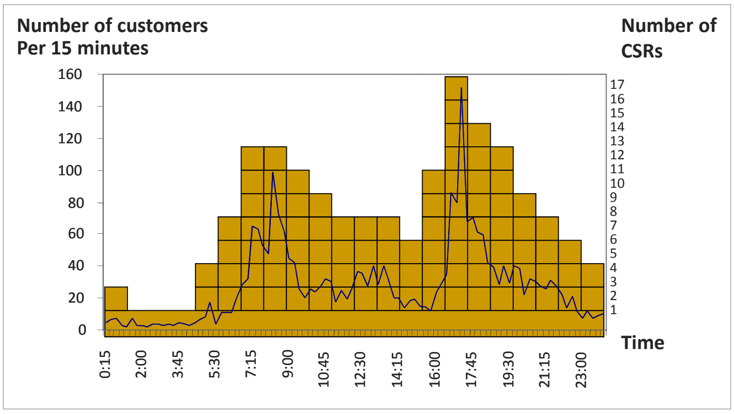
8.7 Service Levels in Waiting Systems



Target Wait Time (TWT) depends on your market position and the importance of incoming calls for your business

Service Level = Probability{Waiting Time≤TWT}

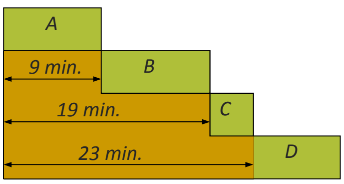
Staffing & Incoming Calls over the Course of a Day



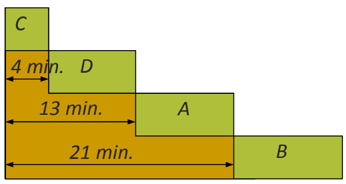
8.10 Priority Rules in Waiting Time Systems

- First-Come-First-Serve: easy to implement + perceived fairness
- Shortest Processing Time Rule: Minimizes average waiting time
- Sequence based on importance: emergency or profitable customers

A: 9 minutes B: 10 minutes C: 4 minutes D: 8 minutes



Total wait time: 9+19+23=51min



Total wait time: 4+13+21=38 min

8.11 Reducing Variability

Reduce Arrival Variability

- appointment/reservation: how to handle late arrivals or no-shows
- encourage customers to avoid peak hours.

Reduce Service Time Variability

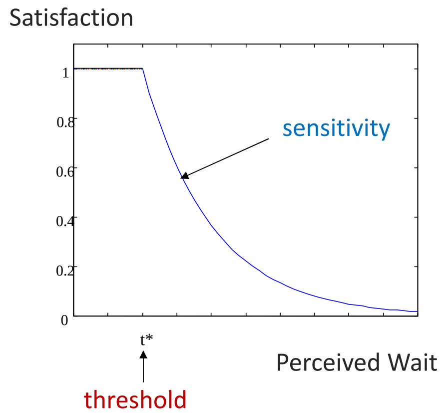
- training and technology
- limit service selection
- reduce customer involvement

24

Actual Wait Time vs. Perceived Wait Time

Perceived Wait Time

- Amount of time customers believe they have waited prior to receiving service.
- Has a greater effect on customer satisfaction than actual waiting time



25

Factors Affecting Perceived Wait Times

Server-Related Factors

- Passive vs. active waits
- Unfair vs. fair waits
- Uncomfortable vs. comfortable waits
- Unexplained vs. explained waits
- Unproductive vs. Productive waits

Customer-Related Factors

- Solo versus group waits
- Waits for more valuable versus less valuable services
- Customer's own tolerance

26

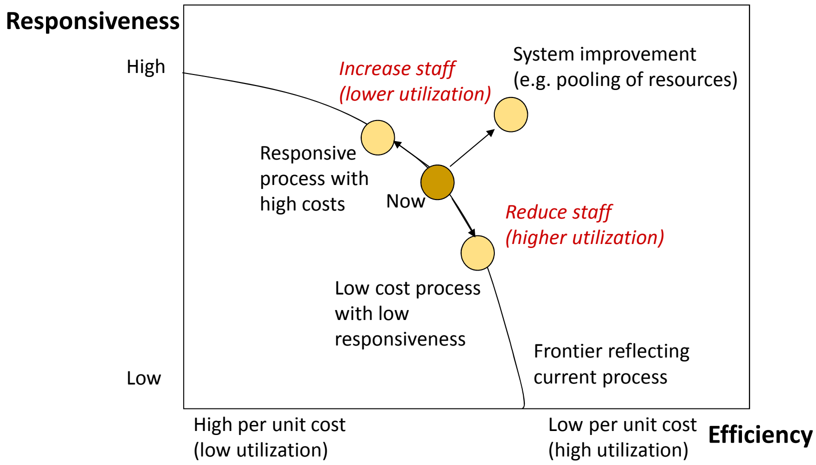
Suggestions for Managing Queues

MITSloan Management Review

1. Determine an acceptable waiting time for your customers
2. Try to divert your customer's attention when waiting
3. Inform your customers of what to expect
4. Keep employees not serving the customers out of sight
5. Segment customers
6. Train your servers to be friendly
7. Encourage customers to come during the slack periods
8. Take a long-term perspective (and redesign the system)

27

Balancing Efficiency with Responsiveness



29

- Variability is the norm, not the exception
  - understand where it comes from and eliminate what you can
- Variability leads to waiting times although utilization<100%

Operations benefit from flexibility in capacity

- Demand can exhibit seasonality → Time varying capacity
- Pooling resources can reduce waiting times

Managing customers' perceived wait times

30