Production and Operations Management
2023/2024

Lisbon School of Economics & Management
ULisboa | Universidade de Lisboa

Waiting Lines

Module D

Queuing Theory

- The study of waiting lines
- Waiting lines are common situations
- Useful in both manufacturing and service areas

Philosophy of Waiting (Maister 1984)

1 Unoccupied time feels longer

2 Pre-service waiting feels longer

3 Anxiety makes waiting seem longer

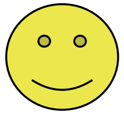
4 Uncertainty waiting is longer than known, finite waiting

5 Unexplained waiting feels longer than explained waiting

6 Unfair waiting is longer

7 Solo waiting is longer than group waiting

8 The more valuable the service the longer is worth waiting



Service Design

The psychology of waiting

Waiting is one of the most stressful experiences for human beings

Strategies to help organizations manage the waiting process:

- Make waiting more comfortable (chairs; air conditioning; refreshments)
- Establish virtual queues (pagers; Disney's Genie+)
- Distracting customers' attention (mirrors near elevators; videos)
- Start service early (take drink orders; let customers see the menu before they sit down)
- Explain the reasons for waiting (reduces uncertainty; creates understanding and empathy)
- Providing pessimistic wait time estimates (customers are pleasantly surprised by a shorter wait)
- Compensate for the extraordinary wait (free drinks; coupons)
- Don't make unrealistic promises (avoids anger later)

Be fair! (Customers are more willing to "share the pain" of waiting if others have similar wait times)

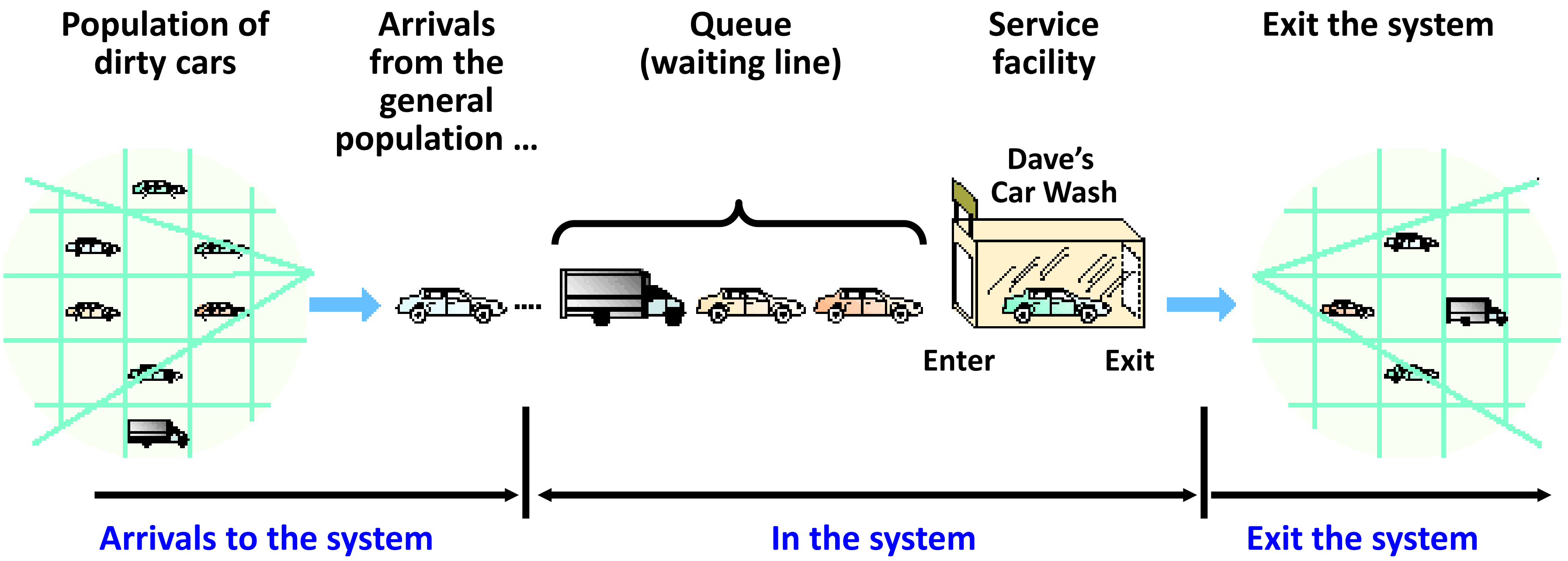
Common Queuing Situations

Situation	Arrivals in Queue	Service Process
Supermarket	Grocery shoppers	Checkout clerks at cash register
Highway toll booth	Automobiles	Collection of tolls at booth
Doctor’s office	Patients	Treatment by doctors and nurses
Computer system	Programs to be run	Computer processes jobs
Telephone company	Callers	Switching equipment to forward calls
Bank	Customer	Transactions handled by teller
Machine maintenance	Broken machines	Repair people fix machines
Harbor	Ships and barges	Dock workers load and unload

Characteristics of Waiting-Line Systems

1. **Arrivals** or inputs to the system
   - Population size, behavior, statistical distribution
2. **Queue discipline**, or the waiting line itself
   - Limited or unlimited in length, discipline of people or items in it
3. **The service facility**
   - Design, statistical distribution of service times

Parts of a Waiting Line



Arrival Characteristics
- Size of the population
- Behavior of arrivals
- Statistical distribution of arrivals

Waiting Line Characteristics
- Limited vs. unlimited
- Queue discipline

Service Characteristics
- Service design
- Statistical distribution of service

Arrivals Characteristics

1. **Size of the population**
   - Unlimited (infinite) or limited (finite)
2. **Pattern of arrivals**
   - Scheduled or random, often a Poisson distribution
3. **Behavior of arrivals**
   - Wait in the queue and do not switch lines
   - No balking or reneging

X: v.a represents the arrivals per unit of time

$X \sim Po\ (\lambda)$

λ: average number of arrivals per unit of time

Waiting-Line Characteristics

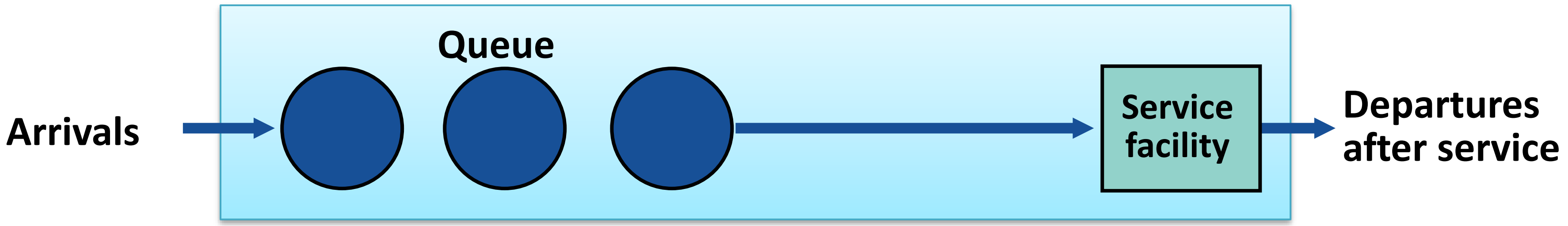
1. **Queue length**
   - Limited or unlimited
2. **Queue discipline**
   - first-in, first-out (FIFO) is most common
   - Other priority rules may be used in special circumstances

Service Characteristics

1. **Queuing system designs**
   - Single-channel system, multiple-channel system
   - Single-phase system, multiphase system
2. **Service time distribution**
   - Constant service time
   - Random service times, usually a negative exponential distribution

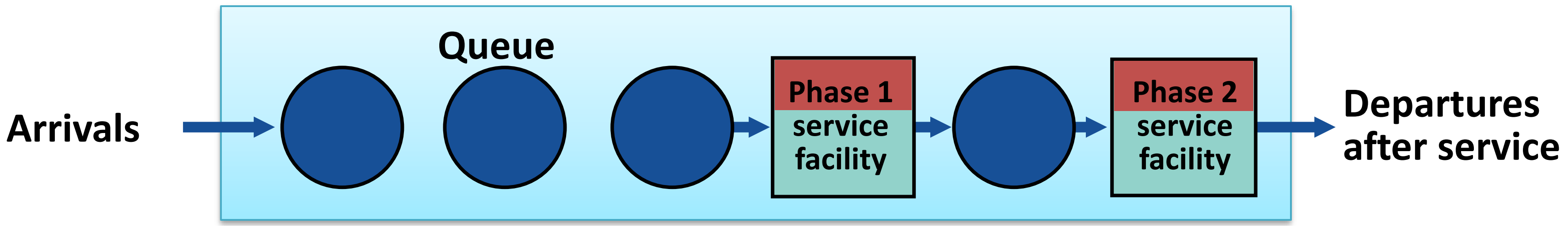
Queuing System Designs

A family dentist's office



Single-channel, single-phase system

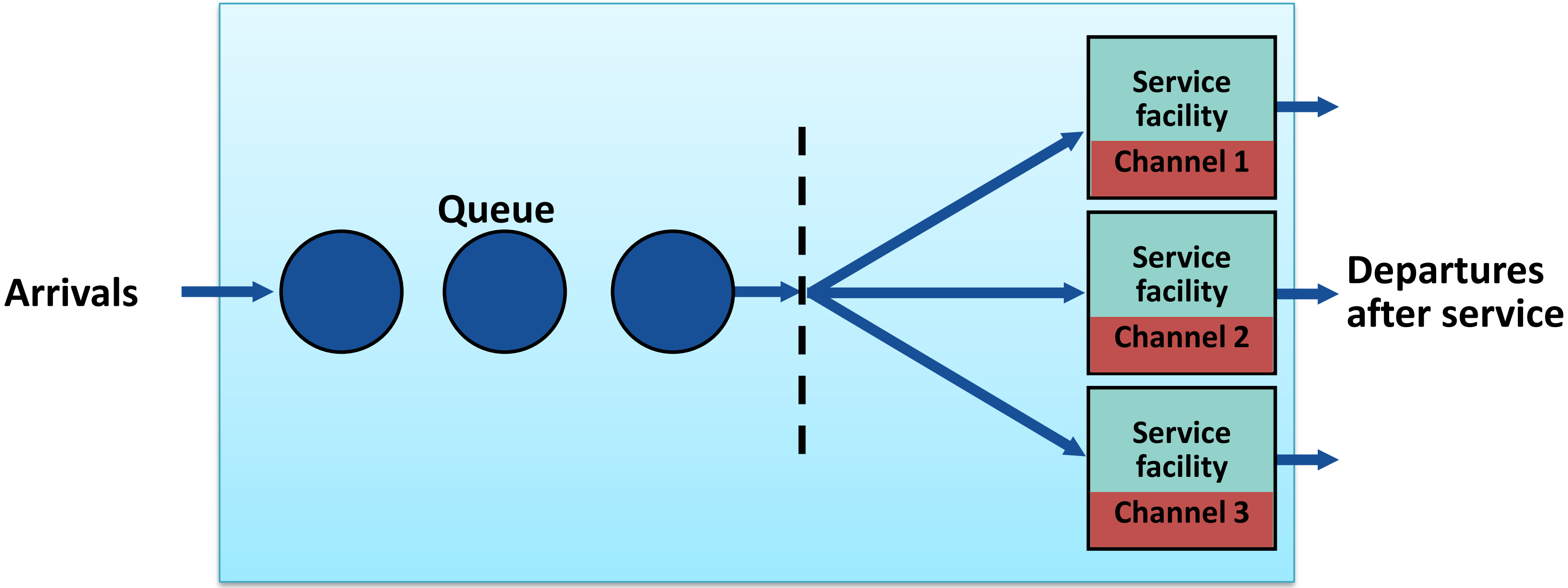
A McDonald's dual window drive-through



Single-channel, multiphase system

Queuing System Designs

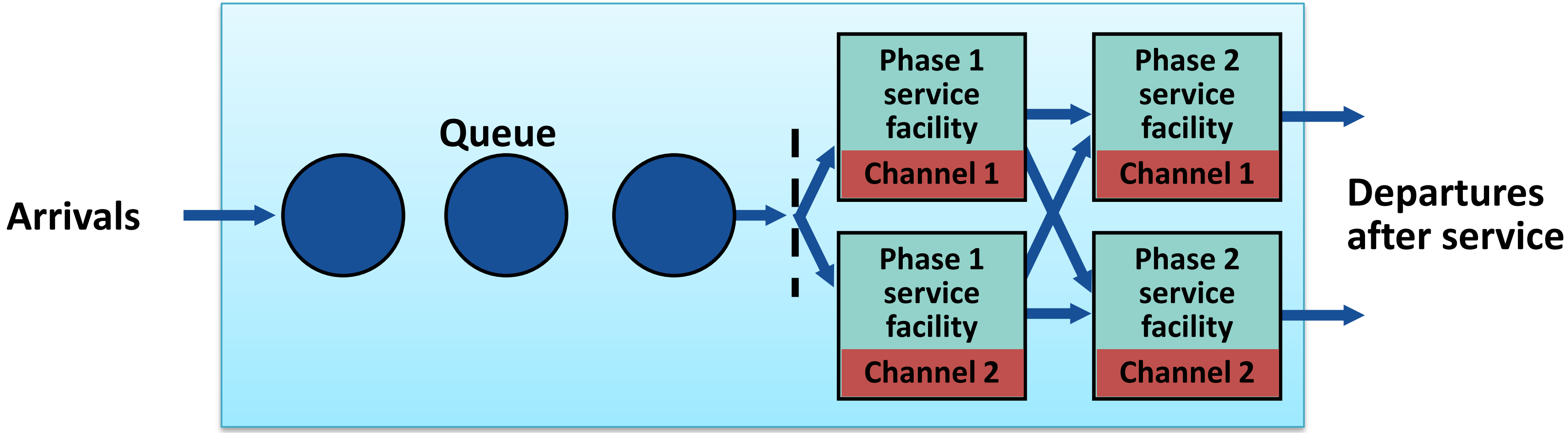
Most bank and post office service windows



Multi-channel, single-phase system

Queuing System Designs

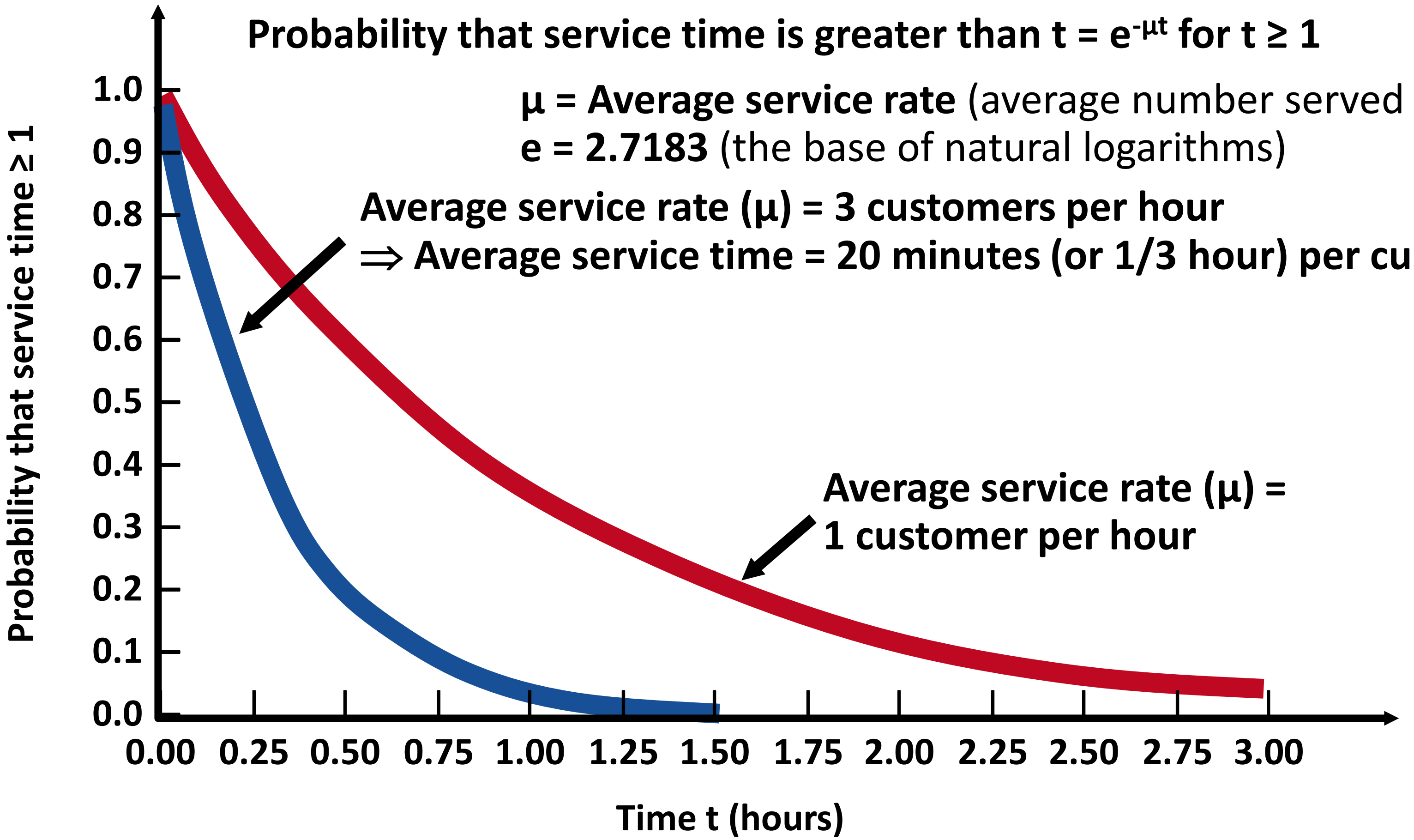
Some college registrations



Multi-channel, multiphase system

Negative Exponential Distribution

Distribution of service duration



Measuring Queue Performance

1. Average number of units (costumers) waiting in the queue (L_s)
2. Average time a unit (customer) spends waiting in the system (W_s)
3. Average number of units (costumers) waiting in the queue (L_q)
4. Average time a unit (customer) spends waiting in the queue (W_q)
5. Utilization factor for the system (ρ)
6. Probability of 0 units (customers) in the system (that is, the service is idle)(P_0)
7. Probability of more than k units (customers) in the system, where n is the number of units (customers) in the system ($P_{n>k}$)

Little's Law

A practical and useful relationship in queuing for any system in a *steady state:*

$$L = \lambda W$$

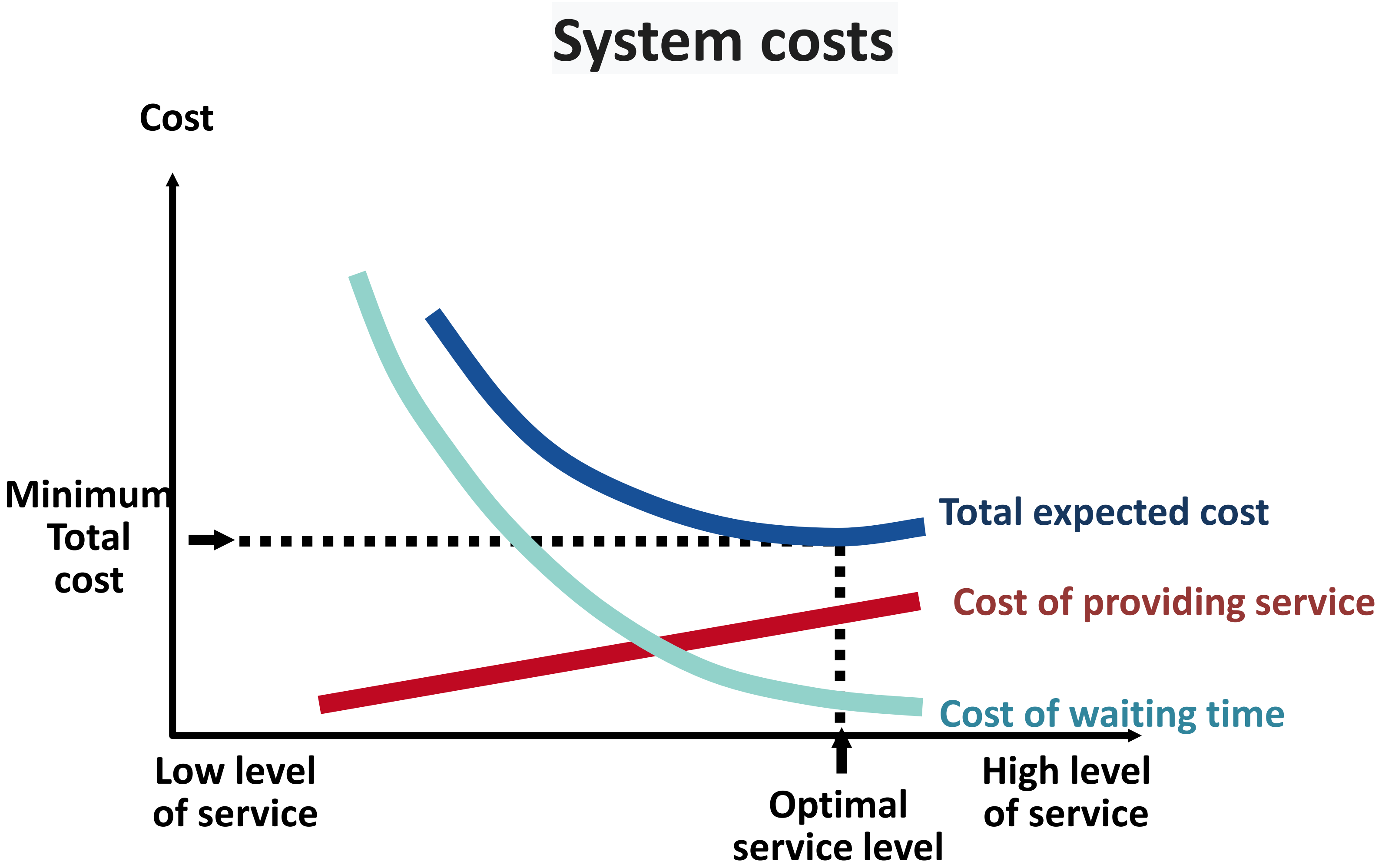
- Once two parameters are known, the other can be easily found
- Makes no assumptions about the probability distribution of arrivals and service times
- Applies to all queuing models, except the finite population model

Fundamental Relationships of Waiting Lines

$$L_q = \lambda W_q \qquad L_s = L_q + \lambda/\mu$$

$$L_s = \lambda W_s \qquad W_s = W_q + 1/\mu$$

Queuing Costs



Kendall's Notation

A / B / S

A = Arrival distribution

(**M** for Poisson, **D** for deterministic, and **G** for general)

B = Service time distribution

(**M** for exponential, **D** for deterministic, and **G** for general)

S = Number of servers

Queuing Models

1. M/M/1
2. M/M/S
3. M/D/1

The three queuing models here all assume:

- Poisson distribution arrivals
- First-In First-Out (FIFO) discipline
- A single-service phase

Model M/M/1 or Model A

Assumptions:

- **Arrivals are described by a Poisson probability distribution and come from an infinite population**
- Arrivals are served on a FIFO basis and every arrival waits to be served regardless of the length of the queue
- **Service times occur according to the negative exponential distribution**
- Arrivals are independent of preceding arrivals, but the average number of arrivals does not change over time
- **One server**
- Service times vary from one customer to the next and are independent of one another, but their average rate is known
- The service rate is faster than the arrival rate

Model M/M/1 or Model A

Measuring Queue Performance

λ = average arrival rate (average number of arrivals per time period)

μ = average service rate (average number of people or items served per time period)

$1/\mu$ = service time

1	$L_s = \dfrac{\lambda}{\mu - \lambda}$	average number of units (customers) in the system (waiting and being served)
2	$W_s = \dfrac{1}{\mu - \lambda}$	average time a unit spends in the system (waiting time plus service time)
3	$L_q = \dfrac{\lambda^2}{\mu(\mu - \lambda)}$	average number of units waiting in the queue
4	$W_q = \dfrac{\lambda}{\mu(\mu - \lambda)} = \dfrac{L_q}{\lambda}$	average time a unit spends waiting in the queue
5	$\rho = \dfrac{\lambda}{\mu}$	utilization factor for the system
6	$P_0 = 1 - \dfrac{\lambda}{\mu}$	probability of 0 units in the system (that is, the service unit is idle)
7	$P_{n>k} = \left(\dfrac{\lambda}{\mu}\right)^{k+1}$	probability of more than *k* units in the system, where *n* is the number of units in the system

Model M/M/1

Example 1:

At the “MontaEscapes” workshop, an exhaust is fitted every 20 minutes. On average, two customers arrive at the workshop per hour, looking for this type of service. The service is done on a FIFO basis. The number of arrivals follows a Poisson distribution, and the service time follows a negative exponential distribution.

Determine Ls, Ws, Lq, Wq, P0 and ρ.

λ = 2 clients/hour, arrivals according to Poisson distribution

$1/\mu$ = 20m ➔ μ = 3 clients/hour, service following a negative Exponential distribution

$$Ls = \frac{\lambda}{\mu - \lambda} = \frac{2}{3-2} = 2\ clients$$

Ls - 2 clients in the system, on average

$$Ws = \frac{1}{\mu - \lambda} = \frac{1}{3-2} = 1\ hour$$

Ws - 1 hour average time in the system

Model M/M/1

Example 1 (cont.):

λ = 2 clients/hour, arrivals according to Poisson distribution

1/μ = 20m ➔ μ = 3 clients/hour, service following a negative Exponential distribution

$$Lq = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{2^2}{3 \times (3-2)} = 1{,}33\ clients$$

Lq – 1.33 clients waiting in line, on average

$$Wq = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{2}{3 \times (3-2)} = \frac{2}{3}\ hours = 40\ min$$

Wq - 40 minute average waiting time per client

ρ = λ/μ = 2/3 = 0.667 = 66.67% of time the barber is busy

P0 = 1 - λ/μ = 1 - 2/3 = 0.333 = 33.33% probability there are 0 clients in the system

Example 1 (cont.):

Model M/M/1

Probability of more than *k* clients in the system

λ = 2 clients/hour, arrivals according to Poisson distribution

μ = 3 clients/hour, service following a negative Exponential distribution

k	$P_{n>k} = (2/3)^{k+1}$	
0	.667 ←	Note that this is equal to 1 - P_0 = 1 - 0.33
1	.444	
2	.296	
3	.198 ←	**Implies that there is a 19.8% chance that more than 3 clients are in the system**
4	.132	
5	.088	
6	.058	
7	.039	

Model M/M/1

Example 1 (cont.):

In order to calculate the costs associated with the system, it was estimated that the cost of waiting in line, in terms of customer dissatisfaction, is around 10 euros per hour, with the labor cost being 7 euros. per hour.

Determine the costs associated with the system.

client dissatisfaction cost = €10 per hour spent ***waiting*** in line.

W_q = 2/3 hours

Total arrivals = 16 clients per day (2 arrivals per hour x 8 working hours)

Labor cost = €56 per day (€7 per hour)

Average number of costumers in the queue per day= (2/3)×16 arrivals = 10.67 clients (Little's Law)

Customer waiting-time cost = 10 €/hour × 2/3 hours × 26 arrivals/day = €106,67 €/day

Total expected costs = €106,67 + €56 = **€162,67 per day**

Model M/M/S

Assumptions:

- Arrivals are served on a FIFO basis and every arrival waits to be served regardless of the length of the queue.
- Arrivals are independent of preceding arrivals but the average number of arrivals does not change over time.
- **Arrivals are described by a Poisson probability distribution and come from an infinite population.**
- Service times vary from one customer to the next, and are independent of one another, but their average rate is known.
- **Service times occur according to the negative exponential distribution.**
- The service rate is faster than the arrival rate.
- **S servers (there are multiple-servers).**

Model M/M/S – Queuing Formulas

S = number of available servers

1 $L_s = \frac{\lambda \mu (\lambda/\mu)^S}{(S-1)!(S\mu-\lambda)^2} P_0 + \frac{\lambda}{\mu}$	**3** $L_q = L_s - \frac{\lambda}{\mu}$
2 $W_s = \frac{\mu(\lambda/\mu)^S}{(S-1)!\,(S\mu-\lambda)^2} P_0 + \frac{1}{\mu} = \frac{L_s}{\lambda}$	**4** $W_q = W_s - \frac{1}{\mu} = \frac{L_q}{\lambda}$

L_q

All performance measures depend on the probability of the system being empty (P0)

Model M/M/S – Queuing Formulas

S = number of available servers

5

$$P_0 = \frac{1}{\left[\sum_{n=0}^{S-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{S!}\left(\frac{\lambda}{\mu}\right)^S \frac{S\mu}{S\mu - \lambda}} \quad para\, S\mu > \lambda$$

$$\rho = \frac{\lambda}{S\mu}$$

6

$$P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} P_0 \quad (n \leq S) \qquad P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{S!\, S^{n-S}} P_0 \quad (n > S)$$

Note: In the M/M/s model, there is no checks the following relationship:

P0 = 1-ρ

Model M/M/S

Example 2:

The person responsible for the "MontaEscapes" workshop decided to hire a second mechanic for the assembly section. On average, two customers arrive at the workshop per hour, who wait in a single line until one of the mechanics is available. Assembling an exhaust requires 20 minutes.

Determine the performance measures: Ls, Ws, Lq, Wq, P0 and ρ.

$P_0 =$

$\lambda =$

$\mu =$

$S =$

$L_s =$

$W_s =$

$L_q =$

$W_q =$

Model M/M/S

Example 2 (cont.):

$\lambda = 2$ Customers arriving hourly

$\mu = 3$ Customers are served hourly

$S = 2$ Servers $S\mu = 2 \times 3 > 2 = \lambda$

$$P_0 = \frac{1}{\left[\sum_{n=0}^{2-1} \frac{1}{n!}\left(\frac{2}{3}\right)^n\right] + \frac{1}{2!}\left(\frac{2}{3}\right)^2 \frac{2\times 3}{2\times 3 - 2}} = \frac{1}{2}$$

$\lambda =$ **2 clients arriving hourly** $\mu =$ **3 clients are served per hour**

S = 2 servers

$$L_s = \frac{2\times 3\times (2/3)^2}{(2-1)!\,(2\times 3 - 2)^2} \times \frac{1}{2} + \frac{2}{3} = \frac{3}{4} = 0{,}75$$ **clients, on average, in the system**

$$W_s = \frac{L_s}{\lambda} = \frac{3/4}{2} = \frac{3}{8}$$ **hours** $=$ $22{,}5$ **minutes, in average, on system**

Model M/M/S

Example 2 (cont.):

$$L_q = \frac{3}{4} - \frac{2}{3} = \frac{1}{12} = 0{,}083 \text{ clients, on average, in the waiting line}$$

$$W_q = \frac{3}{8} - \frac{1}{3} = \frac{1}{24} \text{ hours} = 2{,}5 \text{ minutes, on average, in the waiting line}$$

	One server	Two servers
P_0	0.33	0.5
L_s	2 cars	0.75 cars
W_s	60 minutes	22.5 minutes
L_q	1.33 cars	0.083 cars
W_q	40 minutes	2.5 minutes

Model M/M/S

Example 2 (cont.):

Lq values depending on the number of servers and as well as on the λ/μ values					
POISSON ARRIVALS, EXPONENTIAL SERVICE TIMES					
	NÚMERO DE SERVIDORES, S				
λ/μ	**1**	**2**	**3**	**4**	**5**
.10	.0111				
.25	.0833	.0039			
.50	.5000	.0333	.0030		
.75	2.2500	.1227	.0147		
.90	8.1000	.2285	.0300	.0041	
1.0		.3333	.0454	.0067	
1.6		2.8444	.3128	.0604	.0121
2.0			.8888	.1739	.0398
2.6			4.9322	.6581	.1609
3.0				1.5282	.3541
4.0					2.2164

Model M/M/S

Example 3:

The person responsible for the "Frutinhas" store is considering making more than one server available to its customers. These arrive at the store at a rate of 18 per hour, with one server serving 20 customers per hour.

Compare the average number of customers in the queue (Lq) and the average waiting time in the queue, per customer (Wq), for different options regarding the number of servers

$\lambda =$ **18 clients arrive per hour**

$\mu =$ **20 clients are served, per hour**

$$\frac{\lambda}{\mu} = 0{,}9$$

$$W_q = \frac{L_q}{\lambda}$$

Nr. of servers *(S)*	Nr. of clients on queue (*L_q*)	Average time on queue *(W_q)*
1	8.1	.45 *hrs*, 27 *min.*
2	.2285	.0127 *hrs*, ¾ *min.*
3	.03	.0017 *hrs*, 6 *seg.*
4	.0041	.0003 *hrs*, 1 *seg.*

Model M/D/1 (Constant-Service-Time)

Assumptions

- Arrivals are served on a FIFO basis and every arrival waits to be served regardless of the length of the queue
- The service fee is higher than the arrival rate
- Arrivals are described by a Poisson probability distribution and come from an infinite population
- One single server
- Service duration is constant
- The average number of arrivals is constant (each arrival is independent of the previous one)

Constant service times, usually attained through automation, help control the variability inherent in service systems. This can lower average queue length and average waiting time.

Model M/D/1 (Constant-Service-Time)

λ = average number of arrivals per unit of time

μ = average number of units served per unit of time

1/μ = average length of service per unit

Performance measures

1	$L_s = \frac{\lambda^2}{2\mu(\mu-\lambda)} + \frac{\lambda}{\mu} = L_q + \frac{\lambda}{\mu}$
2	$W_s = \frac{\lambda}{2\mu(\mu-\lambda)} + \frac{1}{\mu} = W_q + \frac{1}{\mu} = \frac{L_s}{\lambda}$
3	$L_q = \frac{\lambda^2}{2\mu(\mu-\lambda)}$
4	$W_q = \frac{\lambda}{2\mu(\mu-\lambda)} = \frac{L_q}{\lambda}$
5	$\rho = \frac{\lambda}{\mu}$

Model M/D/1 (Constant-Service-Time)

Example 4:

Currently, a driver to wash their car at the "DaEsquina" automatic washing station has to wait in line, on average, 15 minutes. It is estimated that the cost associated with customer dissatisfaction is 60 euros per hour. The person responsible is considering purchasing new equipment capable of providing the same service in 5 minutes, per car, with these arriving at the location at a rate of 8 per hour.
The cost of the new equipment will be amortized at a rate of 3 euros per car.

Calculate the system performance measures for the new equipment

$\lambda =$

$\mu =$

$L_q =$

$W_q =$

$L_s =$

$W_s =$

Model M/D/1 (Constant-Service-Time)

Example 4 (cont.):

$\lambda =$ **8 clients arrive per hour** $\quad \mu =$ **12 clients are served per hour**

$$L_q = \frac{(8)^2}{(2)(12)(12-8)} = \frac{2}{3} \approx 0,667$$ **clients, on average, in the waiting line (queue)**

$$W_q = \frac{2/3}{8} = \frac{1}{12}\,\mathbf{h} = 5$$ **minutes, on average, in the waiting line (queue)**

$$L_s = \frac{2}{3} + \frac{8}{12} = \frac{4}{3} \approx 1,33$$ **clients, on average in the system**

$$W_s = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}\,\mathbf{h} \approx 10$$ **minutes, on average, in the system**

Model M/D/1 (Constant-Service-Time)

Example 4 (cont.):

***Objective* –** Determine whether it is worth the investment

System	W_q (h)	Cost Dissatisfaction (u.m. / car)	Amortization Cost (u.m. / car)	Total Cost (u.m. / car)
Actual	15/60	60×(15/60)=15	-	15
Novo	5/60	60×(5/60)=5	3	8
			Saving	**7 u.m./car**

Model M/D/1 (Constant-Service-Time)

Example 5:

ABC, Inc., collects and compacts aluminum cans and glass bottles. Its truck drivers currently wait an average of 15 minutes before emptying their loads for recycling. The cost of driver and truck time while they are in queues is valued at €60,00 per hour. A new automated compactor can be purchased to process truckloads at a *constant* rate of 12 trucks per hour (that is, 5 minutes per truck). Trucks arrive according to a Poisson distribution at an average rate of 8 per hour. If the new compactor is put in use, the cost will be amortized at a rate of €3,00 per truck unloaded.

Conduct an analysis to evaluate the costs versus benefits of the purchase.

Model M/D/1 (Constant-Service-Time)

Example 5 (cont.):

Current waiting cost/trip = (1/4 hour waiting now)x(€60/hour cost) = **€15/trip**

New system: **λ** = 8 trucks/hour arriving; **µ** = 12 trucks/hour served

Average waiting time in queue: Wq = λ/[2µ(µ - λ)] = 8/[2x12x(12-8)] = **1/12 hour**
Waiting cost/trip with new compactor: (1/12 hour wait)x(€60/hour cost)=**€5/trip**
Savings with new equipment: €15(current system) - €5(new system) = **€10/trip**
Cost of new equipment amortized: **€3/trip**

Net savings: €7,00/trip

OPEN
MINDS.
GRAB
THE FUTURE.

iseg.ulisboa.pt