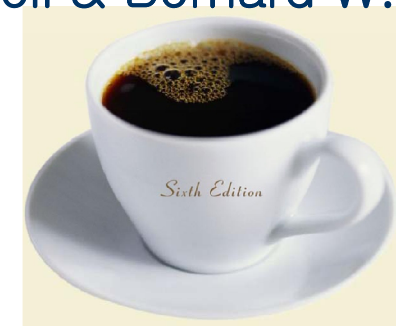
Chapter 5

Service Design

Operations Management - 6th Edition

Roberta Russell & Bernard W. Taylor, III



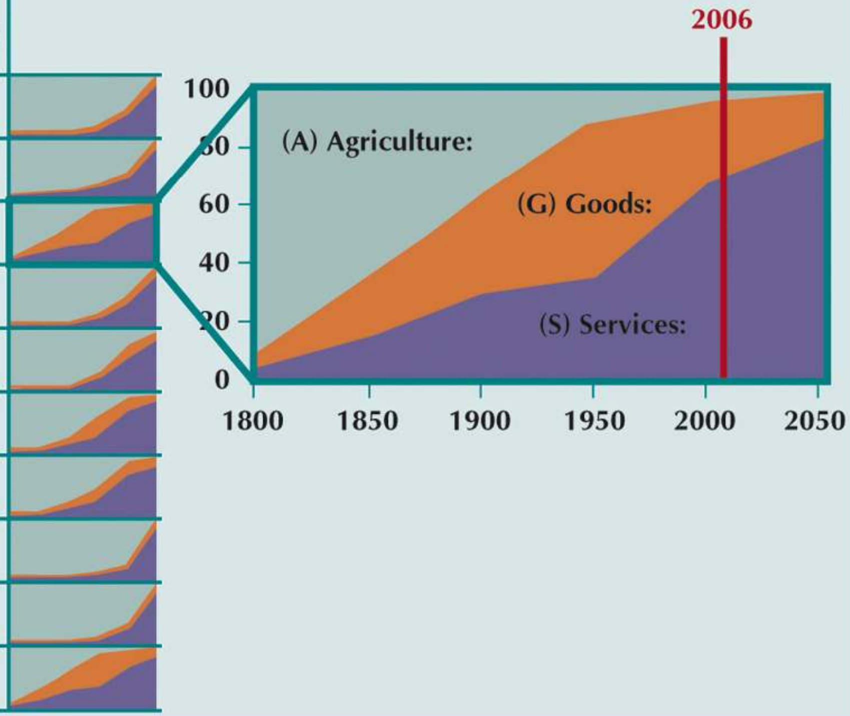
Beni Asllani
University of Tennessee at Chattanooga

Lecture Outline

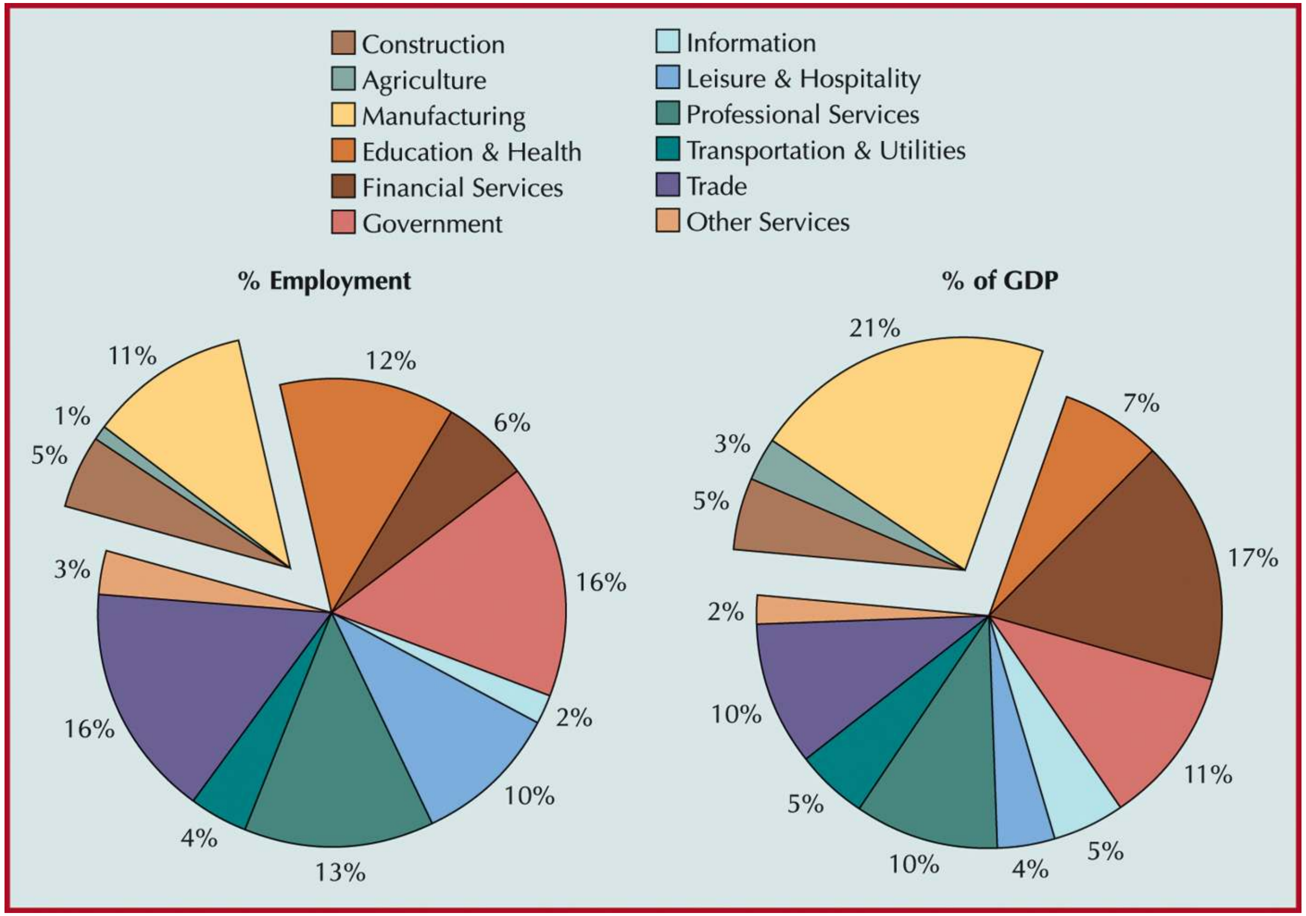
- Service Economy
- Characteristics of Services
- Service Design Process
- Tools for Service Design
- Waiting Line Analysis for Service Improvement

Service Economy

Nation	% World Labor	% Agriculture	% Goods	% Services
China	21.0	50	15	35
India	17.0	60	17	23
U.S.	4.8	3	27	70
Indonesia	3.9	45	16	39
Brazil	3.0	23	24	53
Russia	2.5	12	23	65
Japan	2.4	5	25	70
Nigeria	2.2	70	10	20
Bangledesh	2.2	63	11	26
Germany	1.4	3	33	64



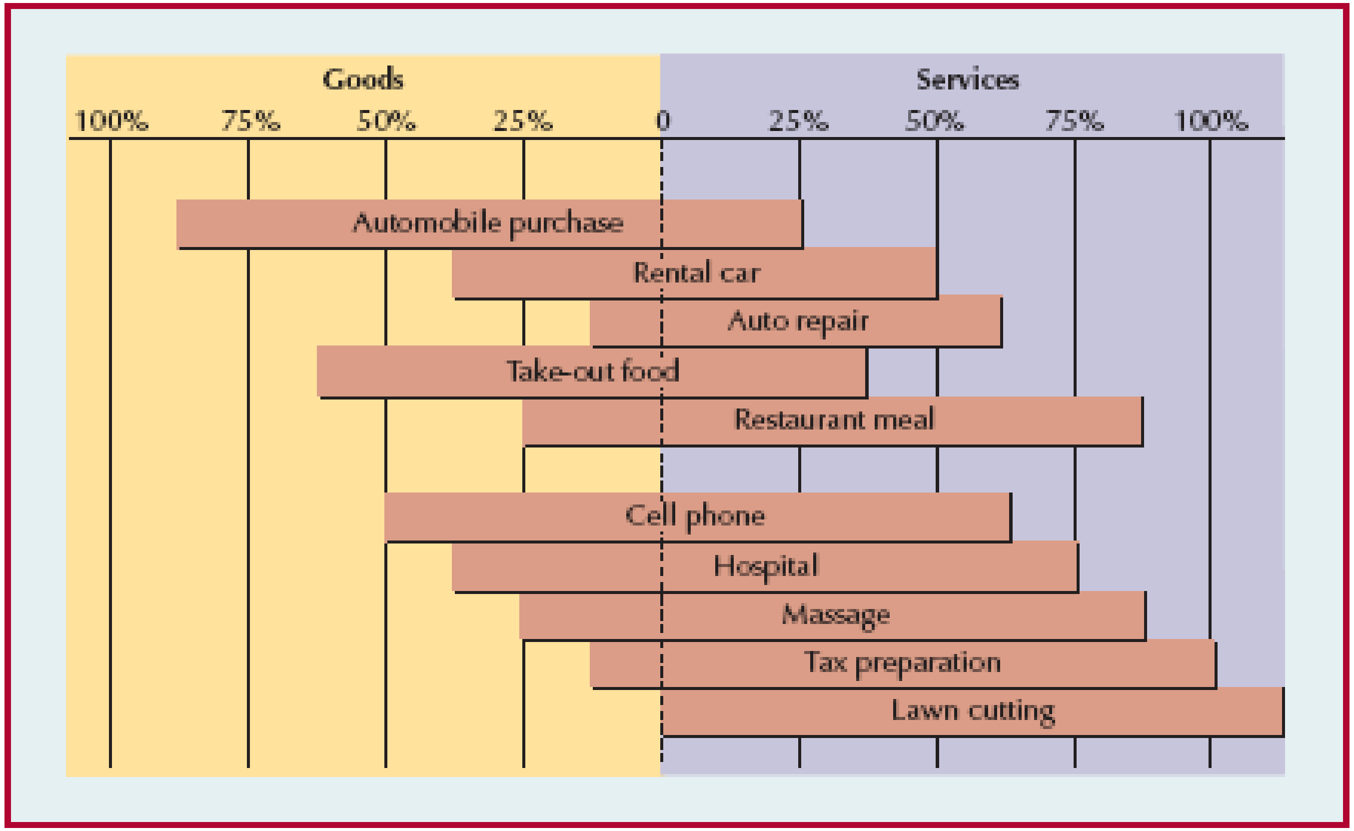
Source: U.S. Bureau of Labor Statistics, IBM Almaden Research Center


Construction
Agriculture
Manufacturing
Education & Health
Financial Services
Government
Information
Leisure & Hospitality
Professional Services
Transportation & Utilities
Trade
Other Services
% Employment
11%
12%
6%
1%
5%
16%
3%
2%
16%
10%
4%
13%
% of GDP
21%
7%
3%
5%
17%
2%
10%
11%
5%
10%
4%
5%

Characteristics of Services

- Services
  - acts, deeds, or performances
- Goods
  - tangible objects
- Facilitating services
  - accompany almost all purchases of goods
- Facilitating goods
  - accompany almost all service purchases

Continuum from Goods to Services

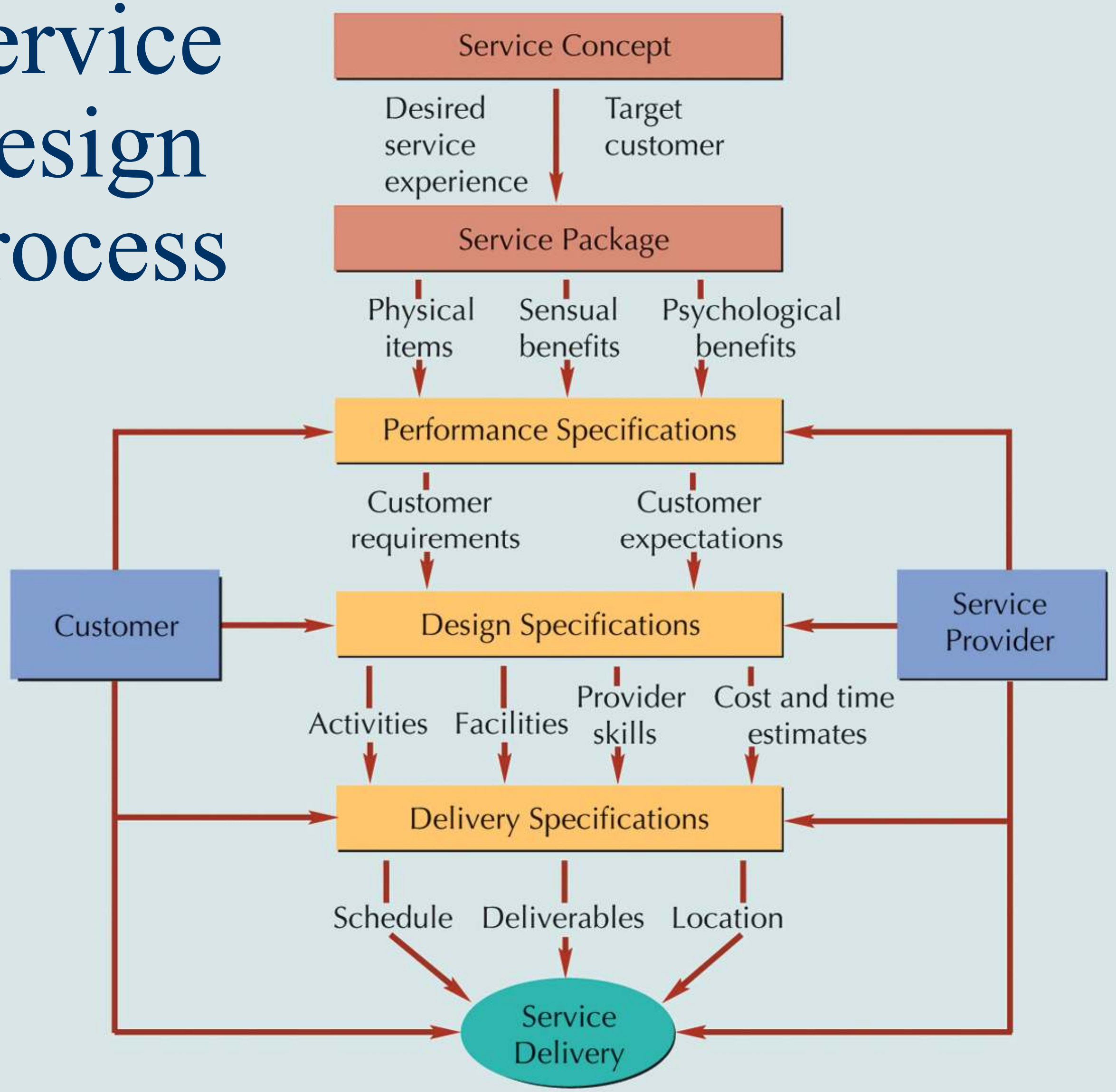


Source: Adapted from Earl W. Sasser, R.P. Olsen, and D. Daryl Wyckoff, *Management of Service Operations* (Boston: Allyn Bacon, 1978), p.11.

Characteristics of Services (cont.)

- Services are intangible
- Service output is variable
- Services have higher customer contact
- Services are perishable
- Service inseparable from delivery
- Services tend to be decentralized and dispersed
- Services are consumed more often than products
- Services can be easily emulated

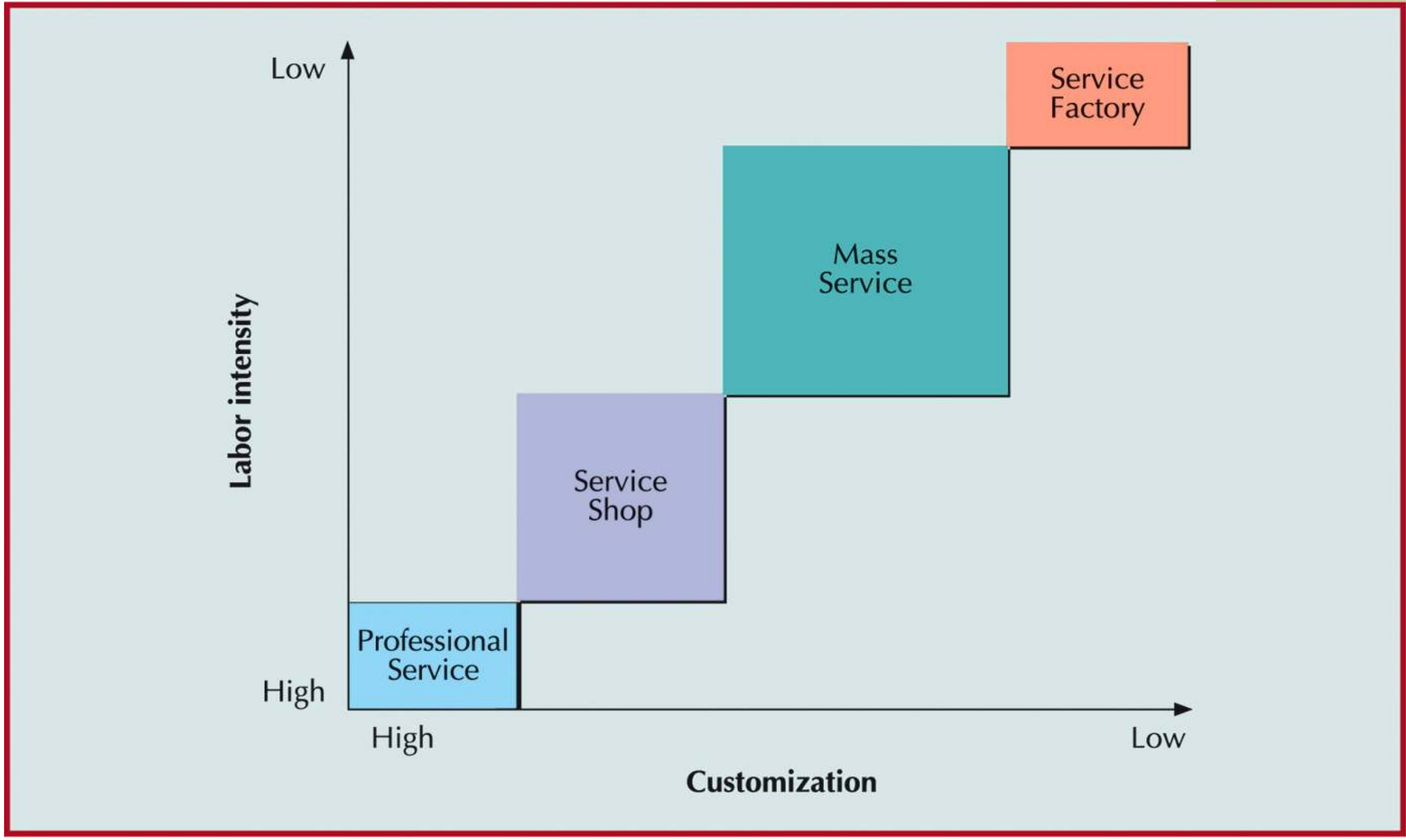
Service Design Process



Service Design Process (cont.)

- Service concept
  - purpose of a service; it defines target market and customer experience
- Service package
  - mixture of physical items, sensual benefits, and psychological benefits
- Service specifications
  - performance specifications
  - design specifications
  - delivery specifications

Service Process Matrix



High v. Low Contact Services

Design Decision	High-Contact Service	Low-Contact Service
◆ Facility locatio n	◆ Convenient to customer	◆ Near labor or transportation source
◆ Facility layout	◆ Must look presentable, accommodate customer needs, and facilitate interaction with customer	◆ Designed for efficiency

Source: Adapted from R. Chase, N. Aquilano, and R. Jacobs, *Operations Management for Compensative Advantage* (New York:McGraw-Hill, 2001), p. 210

High v. Low Contact Services (cont.)

Design Decision	High-Contact Service	Low-Contact Service
◆ Quality control	◆ More variable since customer is involved in process; customer expectations and perceptions of quality may differ; customer present when defects occur	◆ Measured against established standards; testing and rework possible to correct defects
◆ Capacity	◆ Excess capacity required to handle peaks in demand	◆ Planned for average demand

Source: Adapted from R. Chase, N. Aquilano, and R. Jacobs, *Operations Management for Compensative Advantage* (New York:McGraw-Hill, 2001), p. 210

High v. Low Contact Services (cont.)

Design Decision	High-Contact Service	Low-Contact Service
◆ Worker skills	◆ Must be able to interact well with customers and use judgment in decision making	◆ Technical skills
◆ Scheduling	◆ Must accommodate customer schedule	◆ Customer concerned only with completion date

Source: Adapted from R. Chase, N. Aquilano, and R. Jacobs, *Operations Management for Compensative Advantage* (New York:McGraw-Hill, 2001), p. 210

High v. Low Contact Services (cont.)

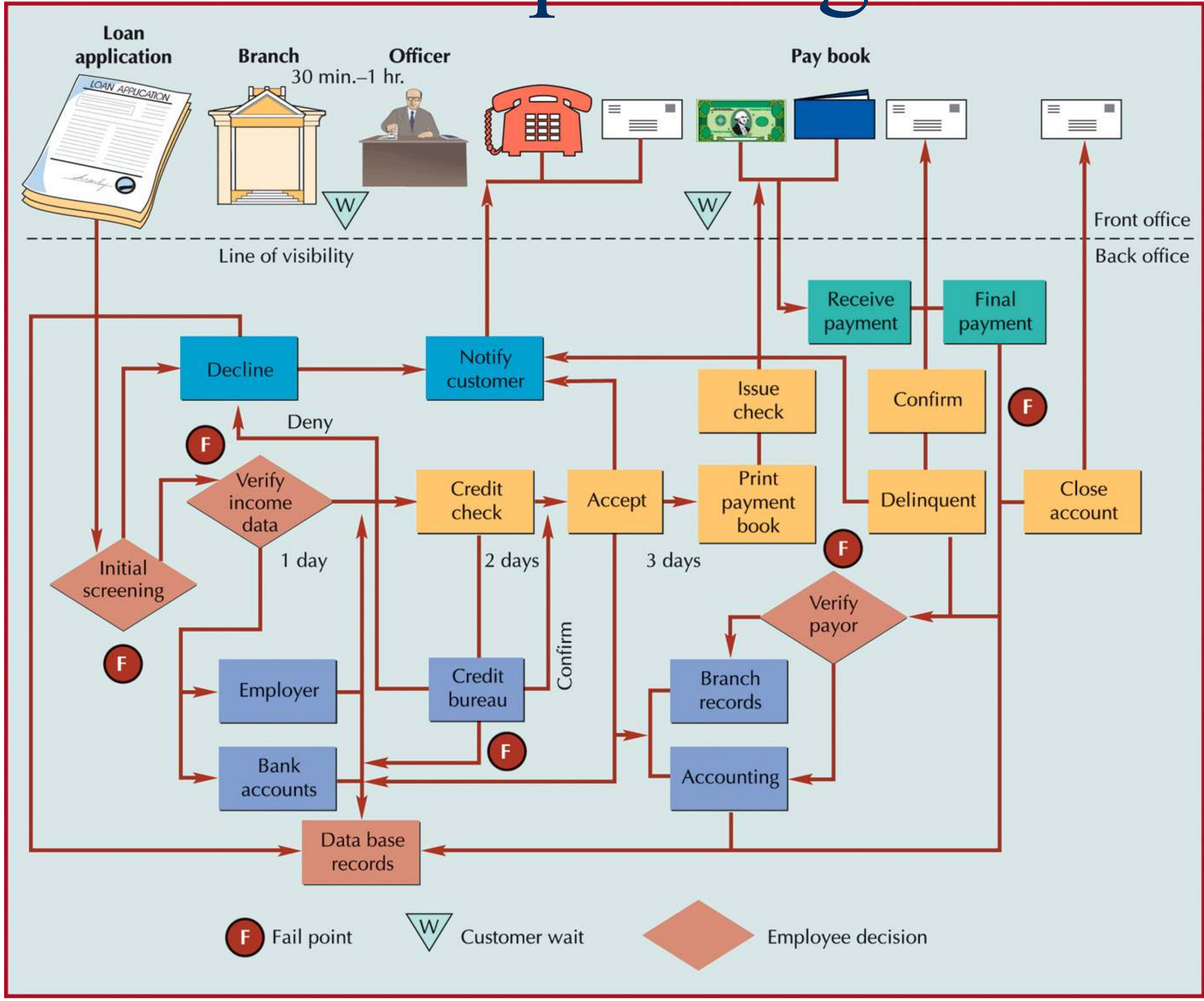
Design Decision	High-Contact Service	Low-Contact Service
◆ Service process	◆ Mostly front-room activities; service may change during delivery in response to customer	◆ Mostly back-room activities; planned and executed with minimal interference
◆ Service package	◆ Varies with customer; includes environment as well as actual service	◆ Fixed, less extensive

Source: Adapted from R. Chase, N. Aquilano, and R. Jacobs, *Operations Management for Compensative Advantage* (New York:McGraw-Hill, 2001), p. 210

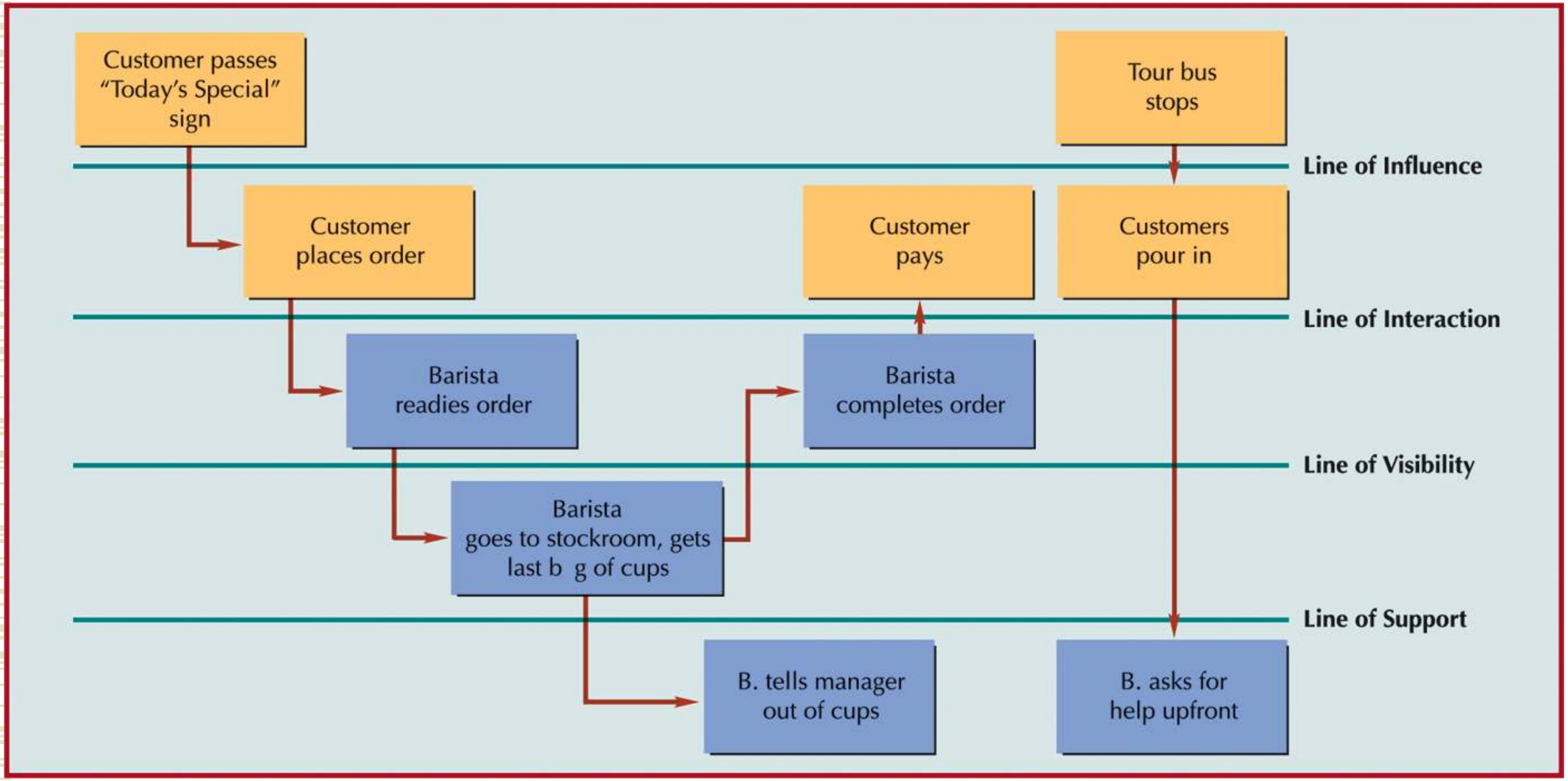
Tools for Service Design

- Service blueprinting
  - line of influence
  - line of interaction
  - line of visibility
  - line of support
- Front-office/Back-office activities
- Servicescapes
  - space and function
  - ambient conditions
  - signs, symbols, and artifacts
- Quantitative techniques

Service Blueprinting

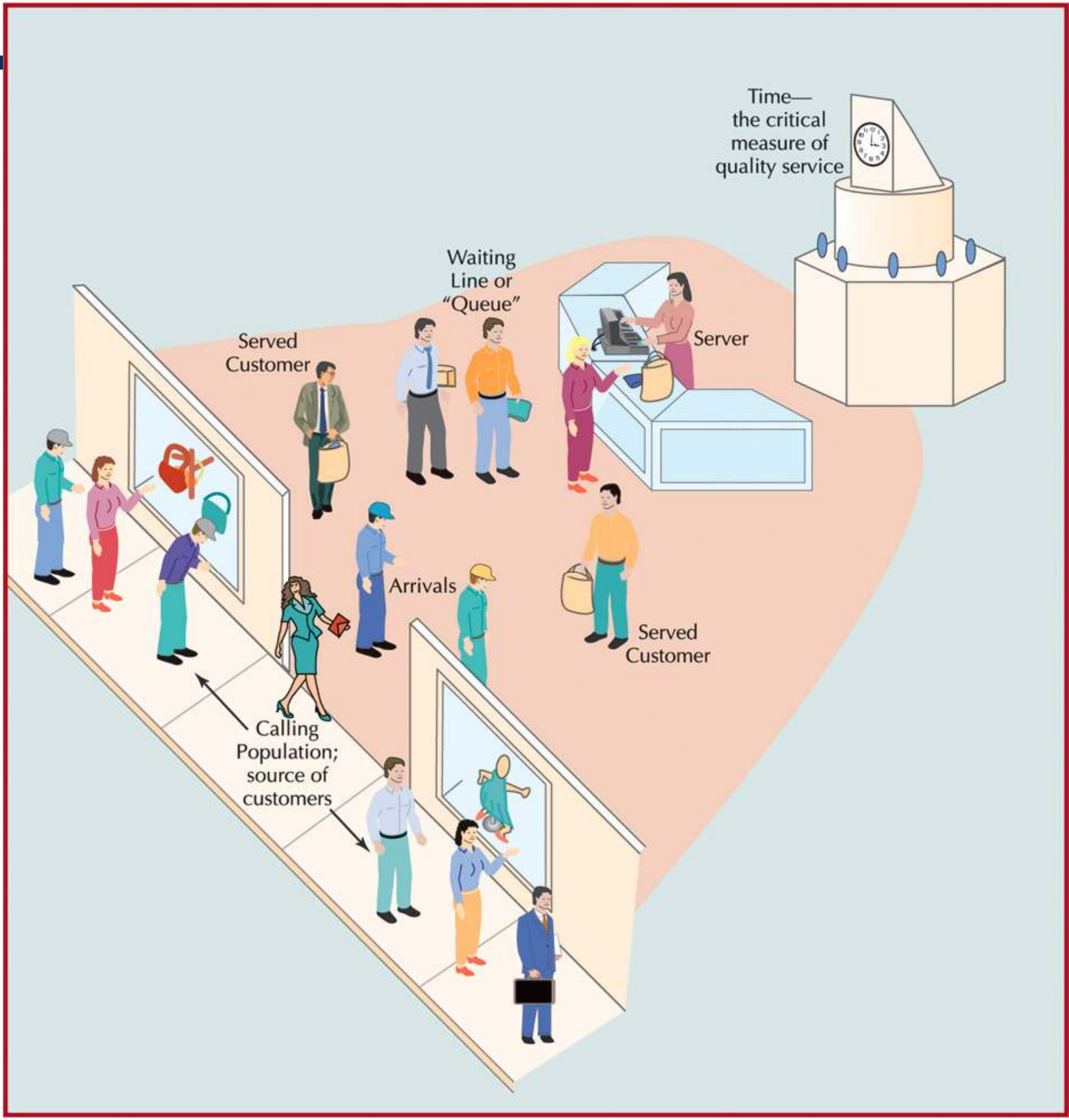


Service Blueprinting (Con't)



Elements of Waiting Line Analysis

- Operating characteristics
  - average values for characteristics that describe performance of waiting line system
- Queue
  - a single waiting line
- Waiting line system
  - consists of arrivals, servers, and waiting line structure
- Calling population
  - source of customers; infinite or finite


Time—
the critical
measure of
quality service
Waiting
Line or
"Queue"
Server
Served
Customer
Arrivals
Served
Customer
Calling
Population;
source of
customers

Elements of Waiting Line Analysis (cont.)

- Arrival rate (λ)
  - frequency at which customers arrive at a waiting line according to a probability distribution, usually Poisson
- Service time (μ)
  - time required to serve a customer, usually described by negative exponential distribution
- Service rate must be shorter than arrival rate ($\lambda < \mu$)
- Queue discipline
  - order in which customers are served
- Infinite queue
  - can be of any length; length of a **finite** queue is limited

Elements of Waiting Line Analysis (cont.)

(b) Multiple-server waiting line

(a) Single-server waiting line

- Channels
  - number of parallel servers for servicing customers
- Phases
  - number of servers in sequence a customer must go through

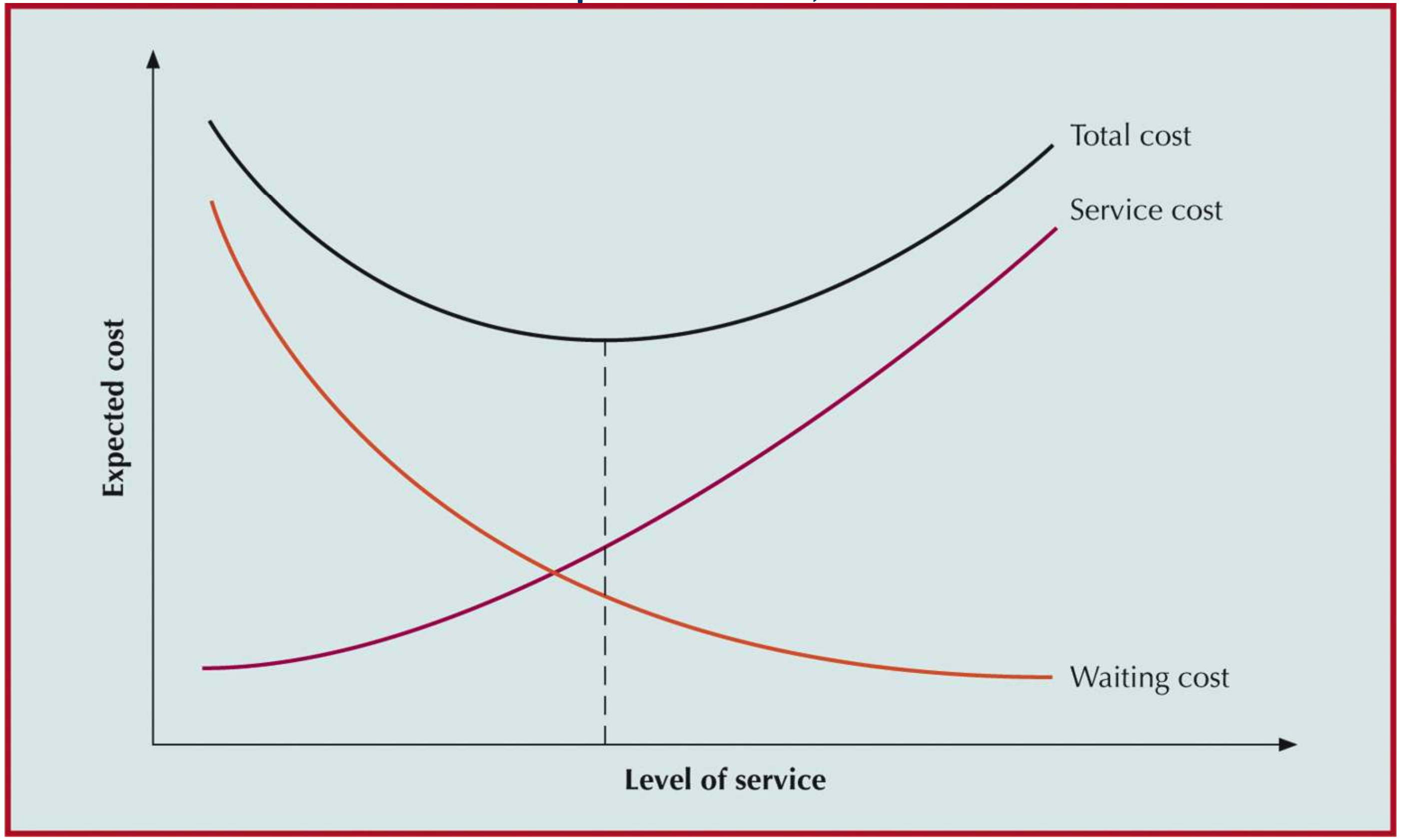
Operating Characteristics

- *Operating characteristics* are assumed to approach a *steady state*

Notation	Operating Characteristic
L	Average number of customers in the system (waiting and being served)
L_q	Average number of customers in the waiting line
W	Average time a customer spends in the system (waiting and being served)
W_q	Average time a customer spends waiting in line
P_0	Probability of no (i.e., zero) customers in the system
P_n	Probability of *n* customers in the system
ρ	Utilization rate; the proportion of time the system is in use

Traditional Cost Relationships

- as service improves, cost increases



Psychology of Waiting

- Waiting rooms
  - magazines and newspapers
  - televisions
- Bank of America
  - mirrors
- Supermarkets
  - magazines
  - "impulse purchases"
- Disney
  - costumed characters
  - mobile vendors
  - accurate wait times

Psychology of Waiting (cont.)

- Preferential treatment
  - Grocery stores: express lanes for customers with few purchases
  - Airlines/Car rental agencies: special cards available to frequent-users or for an additional fee
  - Phone retailers: route calls to more or less experienced salespeople based on customer's sales history
- Critical service providers
  - services of police department, fire department, etc.
  - waiting is unacceptable; cost is not important

Waiting Line Models

- *Single-server* model
  - simplest, most basic waiting line structure
- Frequent variations (all with Poisson arrival rate)
  - exponential service times
  - general (unknown) distribution of service times
  - constant service times
  - exponential service times with finite queue
  - exponential service times with finite calling population

Basic Single-Server Model

- Assumptions
  - Poisson arrival rate
  - exponential service times
  - first-come, first-served queue discipline
  - infinite queue length
  - infinite calling population
- Computations
  - λ = mean arrival rate
  - μ = mean service rate
  - n = number of customers in line

Basic Single-Server Model (cont.)

- probability that no customers are in queuing system

$$P_0 = \left(1 - \frac{\lambda}{\mu}\right)$$

- average number of customers in queuing system

$$L = \frac{\lambda}{\mu - \lambda}$$

- probability of *n* customers in queuing system

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \cdot P_0 = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right)$$

- average number of customers in waiting line

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

Basic Single-Server Model (cont.)

- average time customer spends in queuing system

$$W = \frac{1}{\mu - \lambda} = \frac{L}{\lambda}$$

- average time customer spends waiting in line

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)}$$

- probability that server is busy and a customer has to wait (utilization factor)

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

- probability that server is idle and customer can be served

$$I = 1 - \rho$$

$$= 1 - \frac{\lambda}{\mu} = P_0$$

Basic Single-Server Model Example

$$P_0 = \left(1 - \frac{\lambda}{\mu}\right) = \left(1 - \frac{24}{30}\right)$$

$= 0.20$ probability of no customers in the system

$$L = \frac{\lambda}{\mu - \lambda} = \frac{24}{30 - 24}$$

$= 4$ customers on the average in the queuing system

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{(24)^2}{30(30 - 24)}$$

$= 3.2$ customers on the average in the waiting line

Basic Single-Server Model Example (cont.)

$$W = \frac{1}{\mu - \lambda} = \frac{1}{30 - 24}$$

$= 0.167$ hour (10 minutes) average time in the system per customer

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{24}{30(30 - 24)}$$

$= 0.133$ hour (8 minutes) average time in the waiting line per customer

$$\rho = \frac{\lambda}{\mu} = \frac{24}{30}$$

$= 0.80$ probability that the server will be busy and the customer must wait

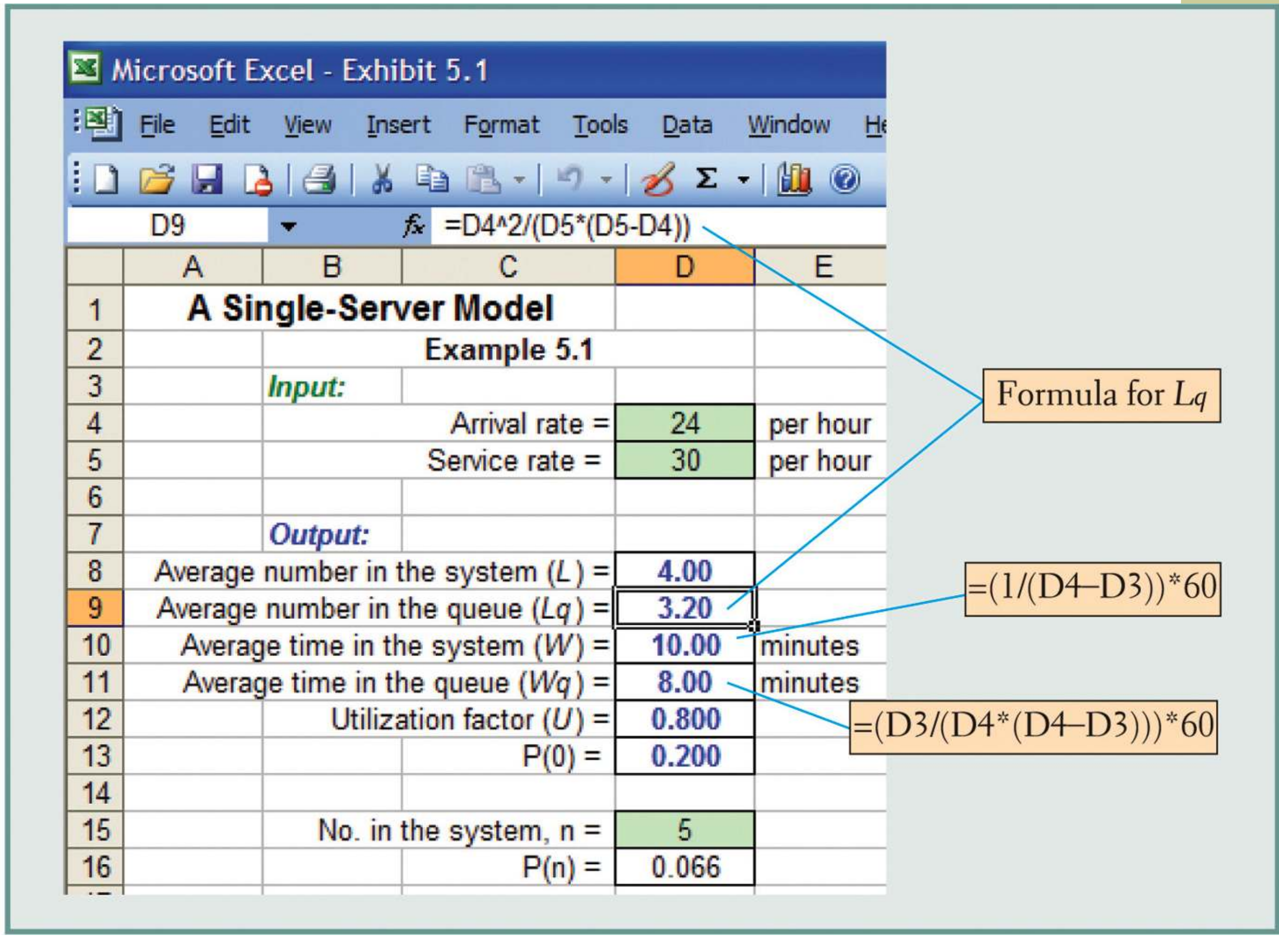
$$I = 1 - \rho = 1 - 0.80$$

$= 0.20$ probability that the server will be idle and a customer can be served

Service Improvement Analysis

- waiting time (8 min.) is too long
  - hire assistant for cashier?
    - increased service rate
  - hire another cashier?
    - reduced arrival rate
- Is improved service worth the cost?

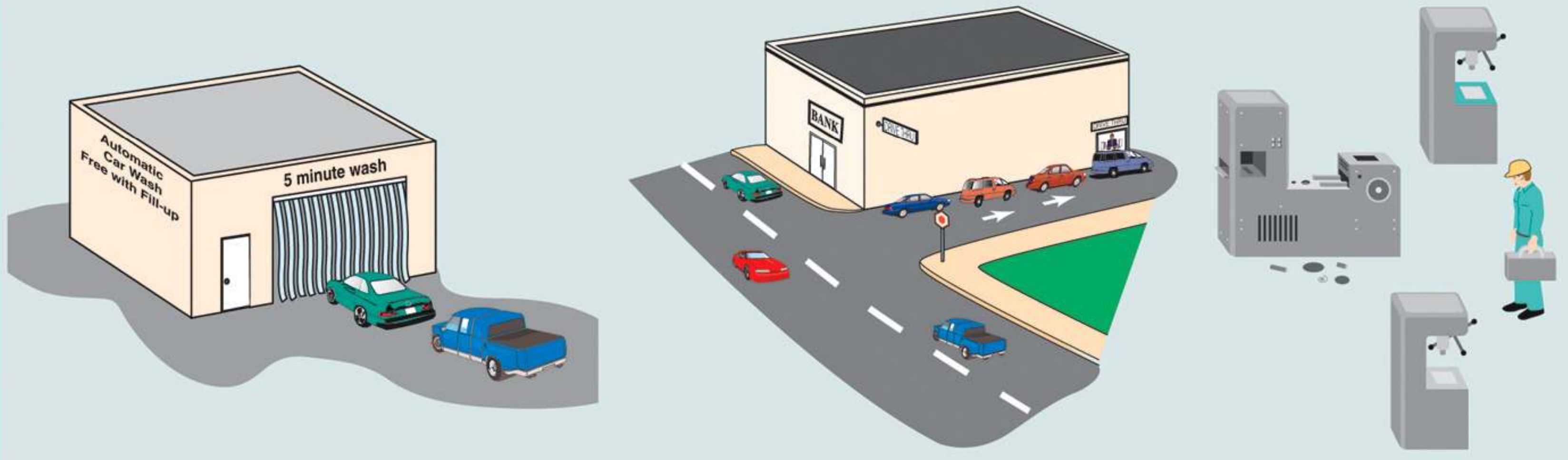
Basic Single-Server Model Example: Excel



Advanced Single-Server Models

- Constant service times
  - occur most often when automated equipment or machinery performs service
- Finite queue lengths
  - occur when there is a physical limitation to length of waiting line
- Finite calling population
  - number of "customers" that can arrive is limited

Advanced Single-Server Models (cont.)



(a) Single-Server, Constant Time

(b) Single-Server, Finite Queue

(c) Single-Server, Finite Calling Population

Basic Multiple-Server Model

- single waiting line and service facility with several independent servers in parallel
- same assumptions as single–server model
- $s\mu > \lambda$
  - s = number of servers
  - servers must be able to serve customers faster than they arrive

Basic Multiple-Server Model (cont.)

- probability that there are no customers in system

$$P_0 = \frac{1}{\left[\sum_{n=0}^{n=s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s\left(\frac{s\mu}{s\mu - \lambda}\right)}$$

- probability of *n* customers in system

$$P_n = \begin{cases} \dfrac{1}{s!s^{n-s}}\left(\dfrac{\lambda}{\mu}\right)^n P_0, & \text{for } n > s \\ \dfrac{1}{n!}\left(\dfrac{\lambda}{\mu}\right)^n P_0, & \text{for } n \leq s \end{cases}$$

Basic Multiple-Server Model (cont.)

- probability that customer must wait

$$P_w = \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s \frac{s\mu}{s\mu - \lambda} P_0$$

$$L_q = L - \frac{\lambda}{\mu}$$

$$L = \frac{\lambda\mu\,(\lambda/\mu)^s}{(s-1)!\,(s\mu - \lambda)^2} P_0 + \frac{\lambda}{\mu}$$

$$W_q = W - \frac{1}{\mu} = \frac{L_q}{\lambda}$$

$$W = \frac{L}{\lambda}$$

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

Basic Multiple-Server Model Example

$$\lambda = 10 \text{ students per hour}$$

$$\mu = 4 \text{ students per hour per service representative}$$

$$s = 3 \text{ service representatives}$$

$$s\mu = (3)(4) = 12 \ (> \lambda = 10)$$

Basic Multiple-Server Model Example (cont.)

$$P_0 = \frac{1}{\left[\sum_{n=0}^{n=s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s\left(\frac{s\mu}{s\mu - \lambda}\right)}$$

$$= \frac{1}{\left[\frac{0}{0!}\left(\frac{10}{4}\right)^0 + \frac{1}{1!}\left(\frac{10}{4}\right)^1 + \frac{1}{2!}\left(\frac{10}{4}\right)^2\right] + \frac{1}{3!}\left(\frac{10}{4}\right)^3 \frac{3(4)}{3(4) - 10}}$$

$= 0.045$ probability that no customers are in the health service.

Basic Multiple-Server Model Example (cont.)

$$L = \frac{\lambda\mu(\lambda/\mu)^s}{(s-1)!(s\mu-\lambda)^2}P_0 + \frac{\lambda}{\mu}$$

$$= \frac{(10)(4)(10/4)^3}{(3-1)![3(4)-10]^2}(0.045) + \frac{10}{4}$$

$= 6$ students in the health service

$$W = \frac{L}{\lambda}$$

$$= \frac{6}{10}$$

$= 0.60$ hour or 36 minutes in the health service

Basic Multiple-Server Model Example (cont.)

$$
\begin{aligned}
L_q &= L - \frac{\lambda}{\mu} \\
&= 6 - \frac{10}{4} \\
&= 3.5 \text{ students waiting to be served}
\end{aligned}
$$

$$
\begin{aligned}
W_q &= \frac{L_q}{\lambda} \\
&= \frac{3.5}{10} \\
&= 0.35 \text{ hour or 21 minutes waiting in line}
\end{aligned}
$$

Basic Multiple-Server Model Example (cont.)

$$P_w = \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s \frac{s\mu}{s\mu - \lambda} P_0$$

$$= \frac{1}{3!}\left(\frac{10}{4}\right)^3 \frac{3(4)}{3(4) - (10)}(0.045)$$

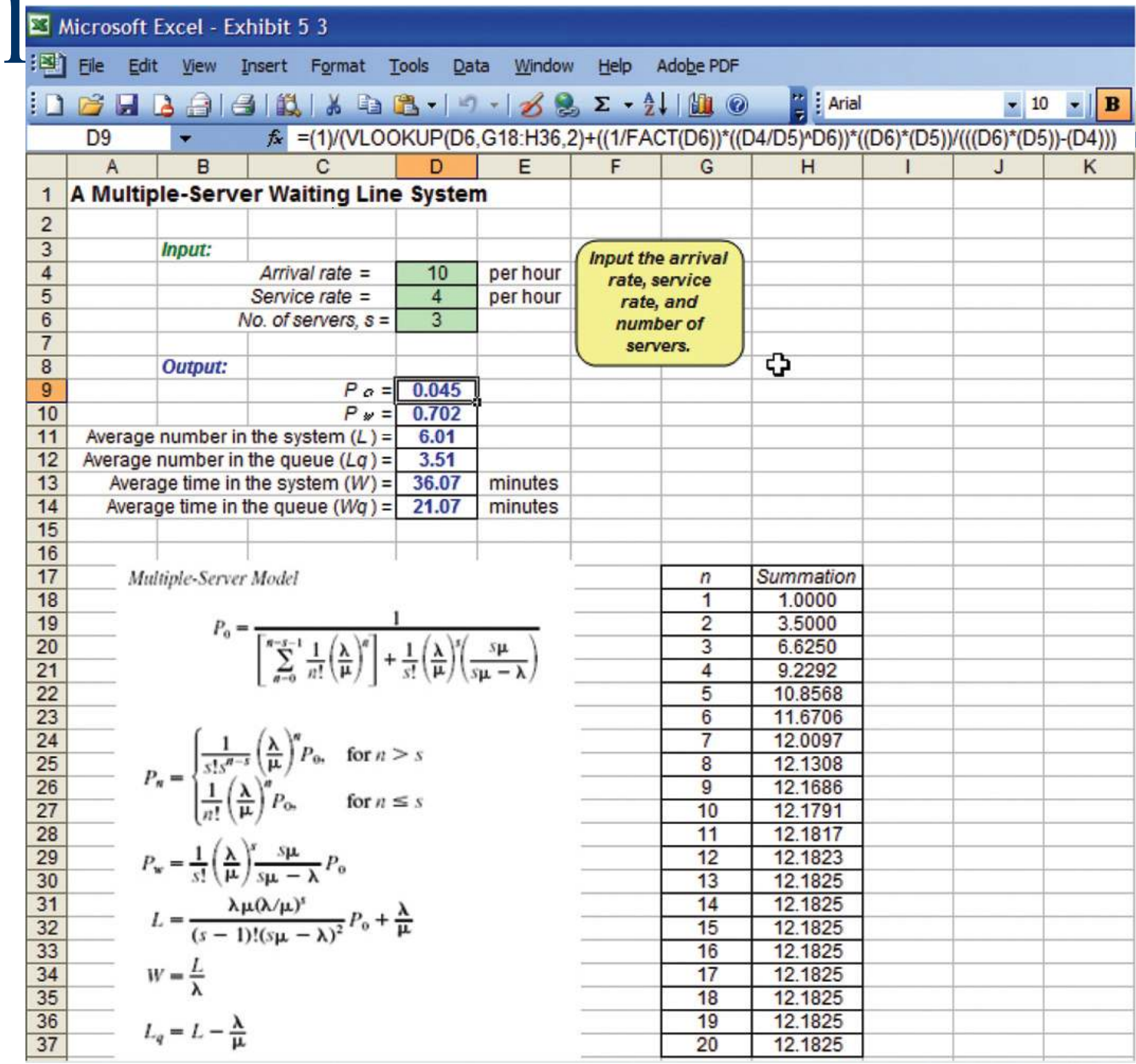
$= 0.703$ probability that a student must wait for service (i.e., that there are three or more students in the system)

Basic Multiple-Server Model Example (cont.)

- To cut wait time, add another service representative
  - now, $s = 4$
- Therefore:

P_0 = 0.073 probability that no students are in the health service
L = 3.0 students in the health service
W = 0.30 hour, or 18 minutes, in the health service
L_q = 0.5 students waiting to be served
W_q = 0.05 hour, or 3 minutes, waiting in line
P_w = 0.31 probability that a student must wait for service

Multiple-Server Waiting Line in Excel



Microsoft Excel - Exhibit 5 3

D9 =(1)/(VLOOKUP(D6,G18:H36,2)+((1/FACT(D6))*((D4/D5)^D6))*((D6)*(D5))/(((D6)*(D5))-(D4)))

A Multiple-Server Waiting Line System

Input:

Arrival rate =	10	per hour
Service rate =	4	per hour
No. of servers, s =	3	

Input the arrival rate, service rate, and number of servers.

Output:

P_0 =	0.045	
P_w =	0.702	
Average number in the system (*L*) =	6.01	
Average number in the queue (*Lq*) =	3.51	
Average time in the system (*W*) =	36.07	minutes
Average time in the queue (*Wq*) =	21.07	minutes

Multiple-Server Model

$$P_0 = \frac{1}{\left[\sum_{n=0}^{s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s\left(\frac{s\mu}{s\mu-\lambda}\right)}$$

$$P_n = \begin{cases} \frac{1}{s!s^{n-s}}\left(\frac{\lambda}{\mu}\right)^n P_0, & \text{for } n > s \\ \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0, & \text{for } n \le s \end{cases}$$

$$P_w = \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s \frac{s\mu}{s\mu-\lambda} P_0$$

$$L = \frac{\lambda\mu(\lambda/\mu)^s}{(s-1)!(s\mu-\lambda)^2} P_0 + \frac{\lambda}{\mu}$$

$$W = \frac{L}{\lambda}$$

$$L_q = L - \frac{\lambda}{\mu}$$

n	*Summation*
1	1.0000
2	3.5000
3	6.6250
4	9.2292
5	10.8568
6	11.6706
7	12.0097
8	12.1308
9	12.1686
10	12.1791
11	12.1817
12	12.1823
13	12.1825
14	12.1825
15	12.1825
16	12.1825
17	12.1825
18	12.1825
19	12.1825
20	12.1825