运营管理(Operations Management)

排序计划
Operations Scheduling

等待是日常生活的一部分

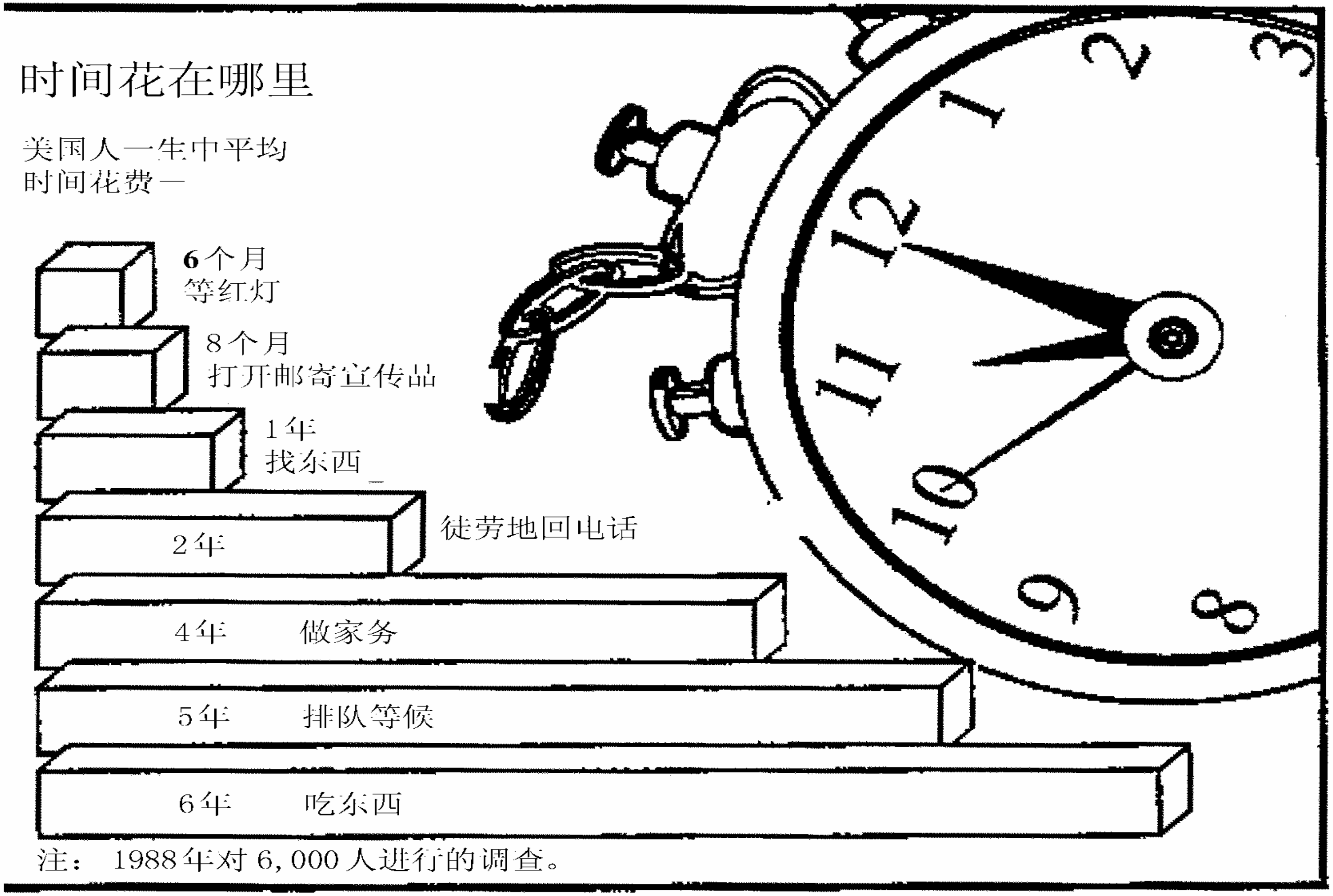


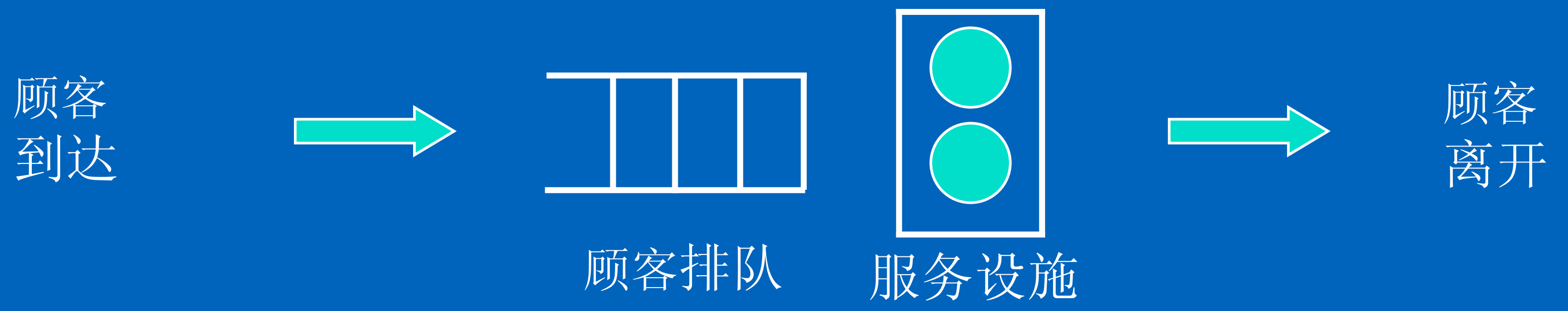
图 11.1 时间花在哪里

资料来源：版权，1989年1月30日《美国新闻与世界报告》第81页。

有趣吗？

"Isn't this fun?"

为什么会出现排队现象？



- 假定每小时平均有4位顾客到达，服务人员为每位顾客的平均服务时间为15分钟。如果顾客到达的间隔时间正好是15分钟，而服务人员为每位顾客的服务时间也正好是15分钟，那么，就只需要一名服务人员，顾客也根本用不着等待。
- 在以下情况将出现排队现象：
  - 平均到达率（顾客/小时）高于平均服务率（顾客/小时），就像红酸果案例中的情况一样。
  - 顾客到达的间隔时间不一样（随机）。
  - 服务时间不一样 （随机）。

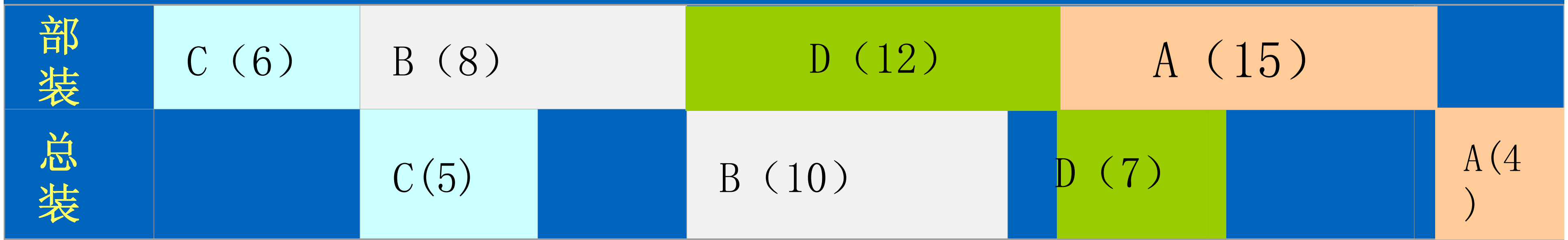
一个排序的例子

四种型号的电视机的装配工时定额

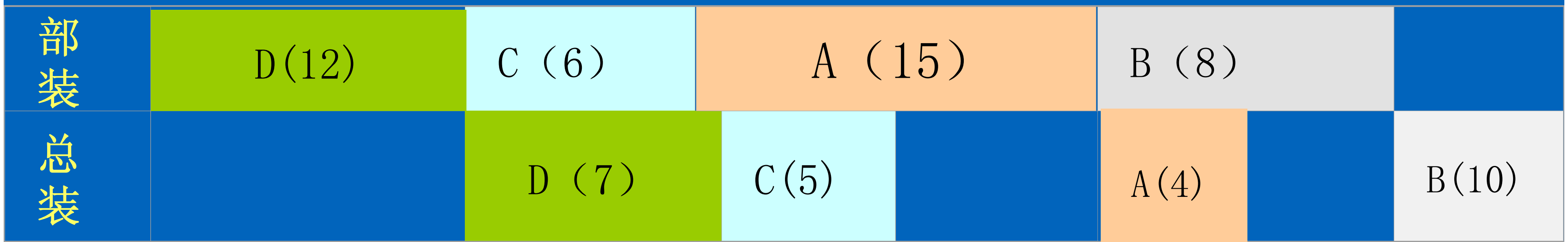
型号	部装定额工时（小时）	总装定额工时（小时）
A	15	4
B	8	10
C	6	5
D	12	7

一个排序的例子

（a）装配顺序为A→B→C→D，总装配时间为48小时

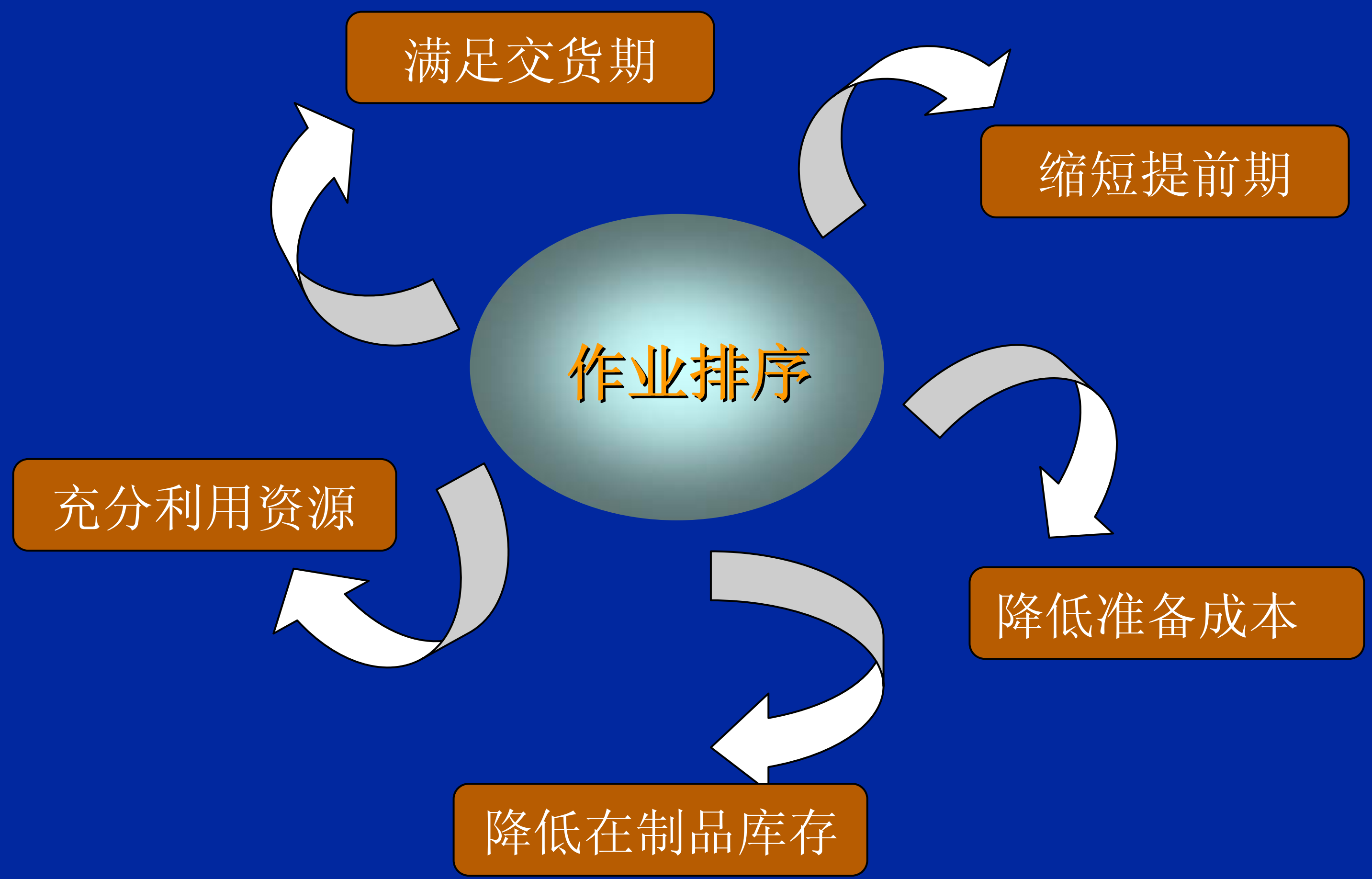


（b）装配顺序为C→B→D→A，总装配时间为45小时



（c）装配顺序为D→C→A→B，总装配时间为51小时

排序的目标



排序分类

- 排序问题的分类
  - 行业
    - 服务业的排序问题
    - 制造业的排序问题
  - 排序对象
    - 服务者排序问题
    - 服务对象排序问题
  - 服务者数量
    - 单服务者排序问题
    - 多服务者排序问题
  - 加工路线
    - 流水作业排序问题
    - 非流水作业排序问题
  - 服务对象到达
    - 静态作业排序问题
    - 动态作业排序问题

制造业中的排序问题(Schedule Problem in Manufacturing)

(一) n种工件在单台设备上加工的排序

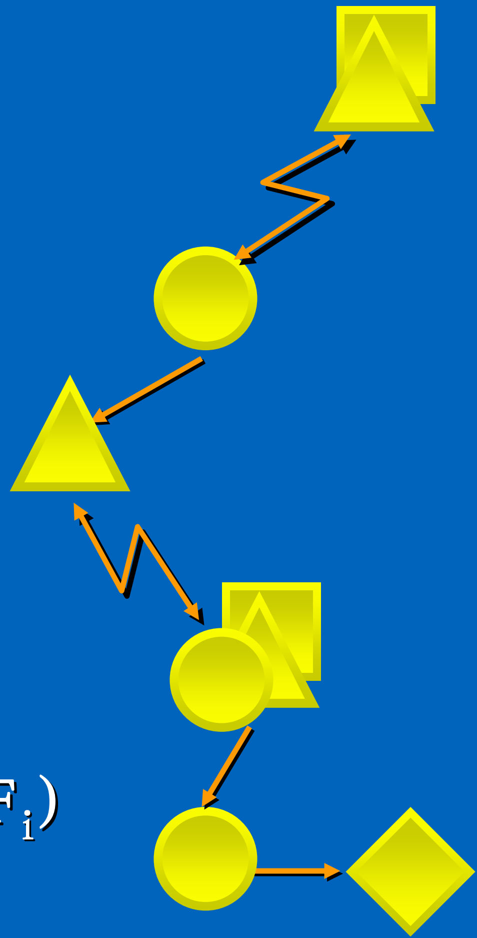
1、平均流程时间最短（SOT, SPT,

Shortest Operating Time

Shortest Processing Time）

F_i: 流程(Flow Time)　w_i+t_i

- w_i: i工件的等待时间
- t_i: i工件的加工时间
- 平均流程 $\overline{F}$
- 总流程: 最大流程　$F_{max}=max(F_i)$
- 优化目标: 平均流程↓
- $t_1 \leqslant t_2 \leqslant \ldots\ldots t_n$



制造业中的排序问题(Schedule Problem in Manufacturing)

	J_1	J_2	J_3	J_4	J_5	J_6
t_i	4	8	2	5	9	3
d_i	24	23	8	6	32	13

	J_3	J_6	J_1	J_4	J_2	J_5
t_i	2	3	4	5	8	9
D_i	0	0	0	8	0	0

$\bar{F}=13.8$　$D_{max}=8$

d_i: i工件规定交货时间(Due Time)

D_i:交货延期量(Delay Time)

制造业中的排序问题(Schedule Problem in Manufacturing)

2、使最大交货延期量最小（EDD规则，Early Delivery Date）

$d_1 \leqslant d_2 \leqslant \ldots\ldots d_n$

$\bar{F}=15.5 \quad D_{max}=0$

	J_4	J_3	J_6	J_2	J_1	J_5
t_i	5	2	3	8	4	9
F_i	5	7	10	18	22	31
d_i	6	8	13	23	24	32
D_i	0	0	0	0	0	0

3、混合法

1） 先按EDD排序

2） 找出$d_i > \max F_i$的，按SPT排

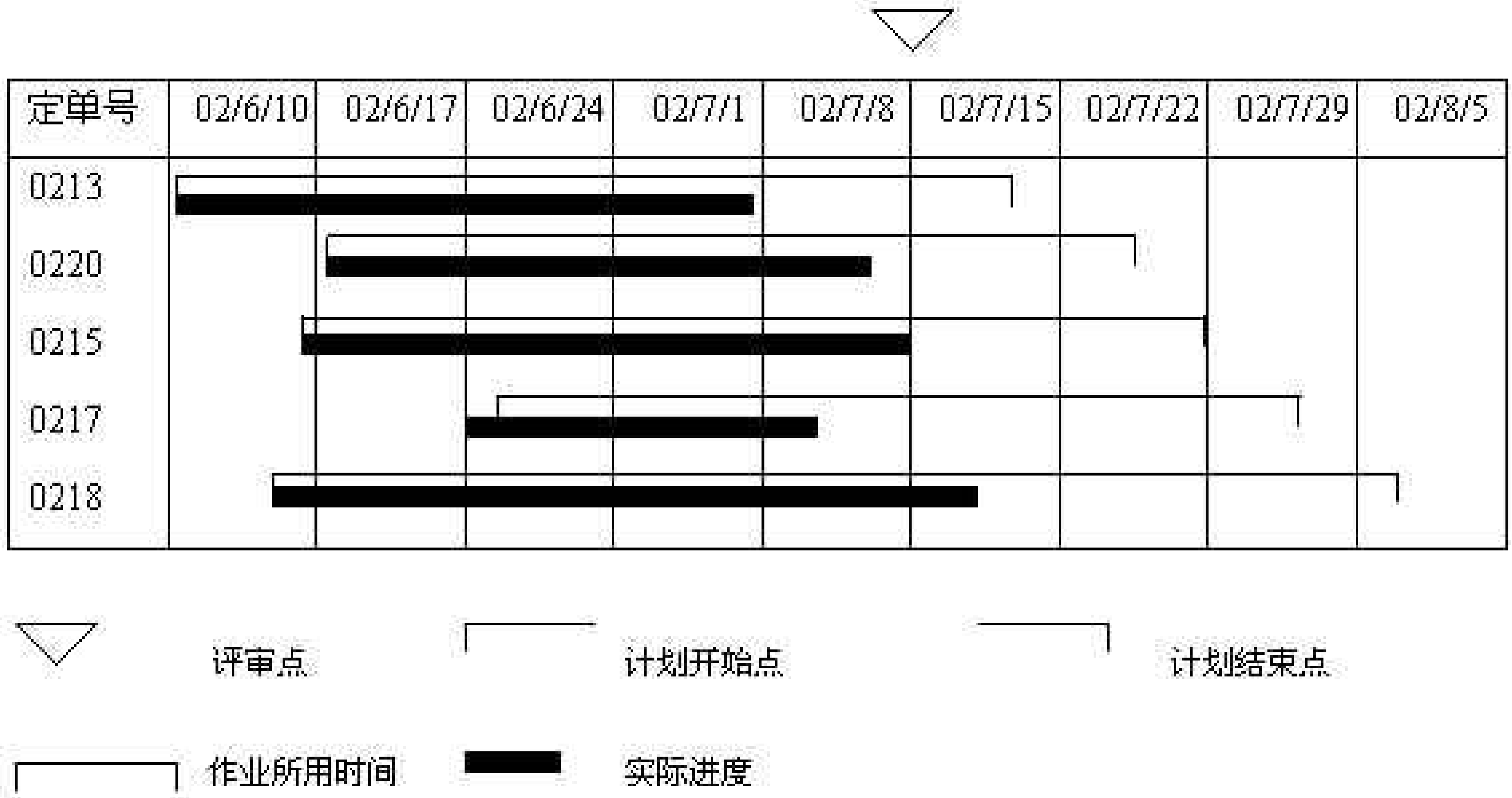
去掉找出的工件，剩下部分继续循环

………. $\bar{F}=14.8, D_{max}=0$

	J_4	J_3	J_6	J_1	J_2	J_5
t_i	5	2	3	4	8	9
F_i	5	7	10	14	22	31
d_i	6	8	13	24	23	32
D_i	0	0	0	0	0	0

单台设备的使用场合：维修、单工艺，加工中心等

甘特图



定单号
02/6/10
02/6/17
02/6/24
02/7/1
02/7/8
02/7/15
02/7/22
02/7/29
02/8/5
0213
0220
0215
0217
0218
评审点
计划开始点
计划结束点
作业所用时间
实际进度

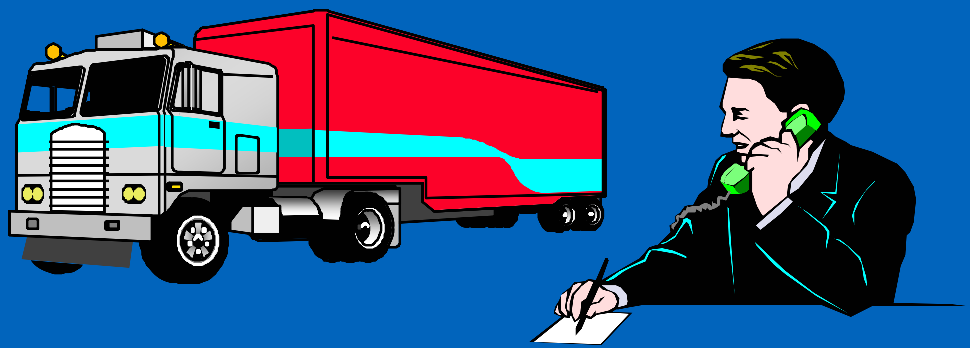
制造业中的排序问题(Schedule Problem in Manufacturing)

（二） n种工件在两台设备上的流水型排序问题 (Scheduling n Jobs on 2 Machines)

约翰逊——贝尔曼规则(Johnson-Bellman's Rule)

	J_1	J_2	J_3	J_4	J_5
t_{iA}	6	8	12	3	7
t_{iB}	11	9	5	3	4

$J_4 J_1 J_2 J_3 J_5$　$F_{max}=40$　最优解



（三） n种工件在3台机床上加工的流水型排序问题及解法

- 转换条件：若min $t_{iA} \geqslant$ max t_{iB}

  或：min $t_{iC} \geqslant$ max t_{iB}

可得最优解，否则较优解

零件 / 机床及工时		J_1	J_2	J_3	J_4
A	t_{iA}	15	8	6	12
B	t_{iB}	3	1	5	6
C	t_{iC}	4	10	5	7

制造业中的排序问题(Schedule Problem in Manufacturing)

∵min $t_{iA}=6 \geqslant$ max $t_{iB}=6$

∴可转换

零件 / 机床及工时		J_1	J_2	J_3	J_4
G	t_{iG}	18	9	11	18
H	t_{iH}	7	11	10	13

- 排序方案为：J_2 J_4 J_3 J_1

  $F_{max}=48$　　最优解

制造业中的排序问题(Schedule Problem in Manufacturing)

（四） n种工件在m台机床上加工的流水型排序问题及解法(Scheduling n Jobs on m Machines)

1、约翰逊规则的扩展法

- 组合原则为：

$$t_{iA}=\sum_{k=1}^{h} t_{ik} \quad (h=1,\ 2,\ \ldots\ldots,\ m-1)$$

$$t_{iB}=\sum_{k=1}^{h} t_{i,m-k+1}$$

- 共组合m－1次，每一次组合对应一个顺序，从m－1种顺序的加工周期中挑最小的。

制造业中的排序问题(Schedule Problem in Manufacturing)

2、关键零件法(Key Parts Model)

- 把总工时（$\sum_{j=1}^{m} t_{ij}$）最大的零件作为关键零件，记为J_c，
  - 若$\mathrm{Max}(\sum_{j=1}^{m} t_{ij}) = \sum_{j=1}^{m} t_{kj}$，则k零件为$J_c$

  1）除J_c外，凡$t_{i1} < t_{im}$的零件，按t_{i1}从小到大排在J_c前；

  2）除J_c外，凡满足$t_{i1} > t_{im}$的零件，按t_{im}从大到小排在J_c后。

  3）若$t_{i1}=t_{im}$，相应零件既可排在J_c前，又可排在J_c后，得到多个方案，从中选最优。

制造业中的排序问题(Schedule Problem in Manufacturing)

机床工时 \ 零件		J_1	J_2	J_3	J_4	J_5	J_6
M_1	t_{i1}	5	5	4	1	2	10
M_2	t_{i2}	5	5	5	3	6	10
M_3	t_{i3}	8	3	3	4	7	4
M_4	t_{i4}	2	8	2	1	5	6
M_5	t_{i5}	5	2	1	2	8	10
$\sum_{j=1}^{5} t_{ij}$		25	23	15	11	28	40

用表格法计算总流程

零件 / 机床	J_4	J_5	J_1	J_6	J_2	J_3
M_1	1	3^{1}	8^{2}	18	23^{5}	27^{6}
M_2	4	10	15^{2}	${}_{3}28$	33	38
M_3	8	${}_{2}17$	25	${}_{3}32$	${}_{1}36^{2}$	${}_{2}41^{5}$
M_4	9	${}_{8}22$	${}_{3}27^{3}$	${}_{5}38$	46^{2}	48^{2}
M5	11	${}_{11}30$	35	${}_{3}48$	50	51

- 右上角零件等待
- 左下角机床空闲

制造业中的排序问题(Schedule Problem in Manufacturing)

以上规则也有一些局限性

1) 生产系统(Production System)是动态的，规则不可能考虑到各种变化

2) 规则看不到上游或下游的情况（设备的忙闲等）

3) 看不到其他许多重要因素

- 如延期交货(Delayed Delivery)所造成的损失
- 因此可根据规则无法考虑的因素作调整，用专家系统(Expert System)或有交互功能的排序软件

服务业中的排序问题(Schedule Problem in Service Industry)

- 排序的对象是人而不是物；系统动态与随机性特点比制造业更突出；排序问题与排队模型结合在一起而产生作用

例：

医院病人等待问题

人流、信息流分离与否

可否挑选医生

价值观

顾客

生产者

长远

市场

服务业中的排序问题(Schedule Problem in Service Industry)

(一) 划分需求，进行排序

- 需求的分类

  随机

  计划（通过排序，平衡负荷(Load Balance)）

  航空公司将乘客划分为工作日商务乘客和周末旅游乘客。

- 美国某医院对需求分析显示:

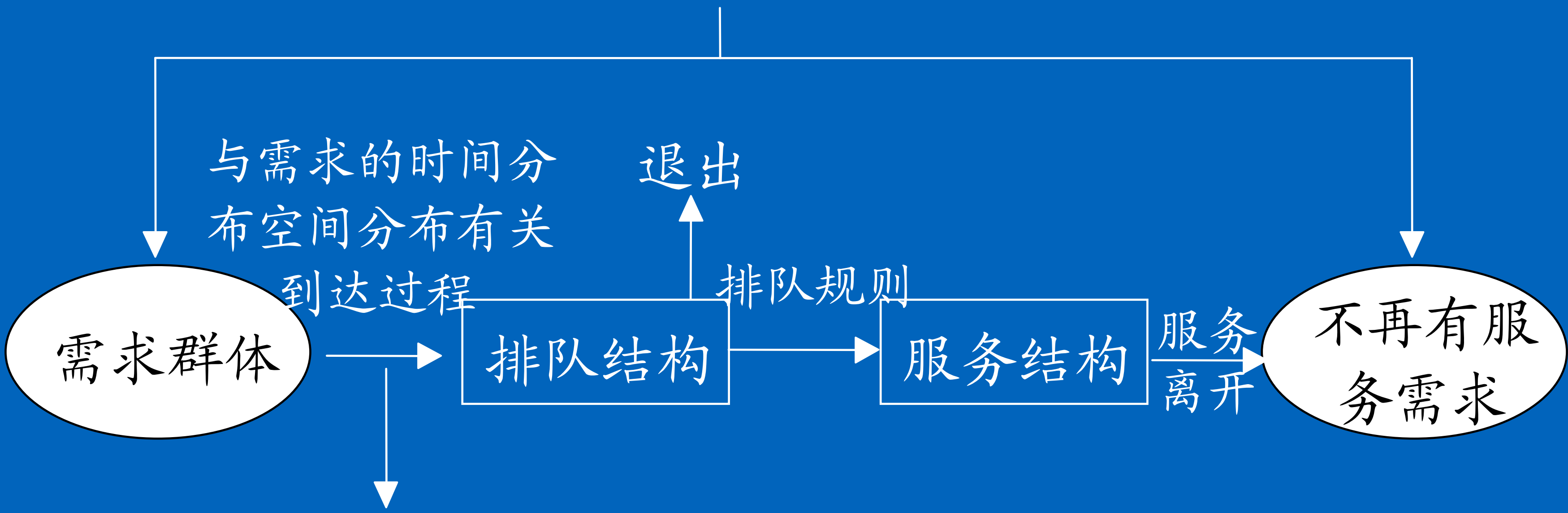
  非预约病人周一看病人数最多，而其他时间来的相对较少。

  非预约——随机

  预约——可控——安排在每周后几天，使负荷稳定，减少等待时间。

(二) 排队管理(Waiting Line Management)

1、排队系统的基本特征



服务业中的排序问题(Schedule Problem in Service Industry)

2、需求群体

不同群体需求不同，每一类需求的预期等待时间(Expected Waiting Time)不同

3、排队结构

1) 多条排队(Multi-Line)

2) 单一排队(Single-Line)

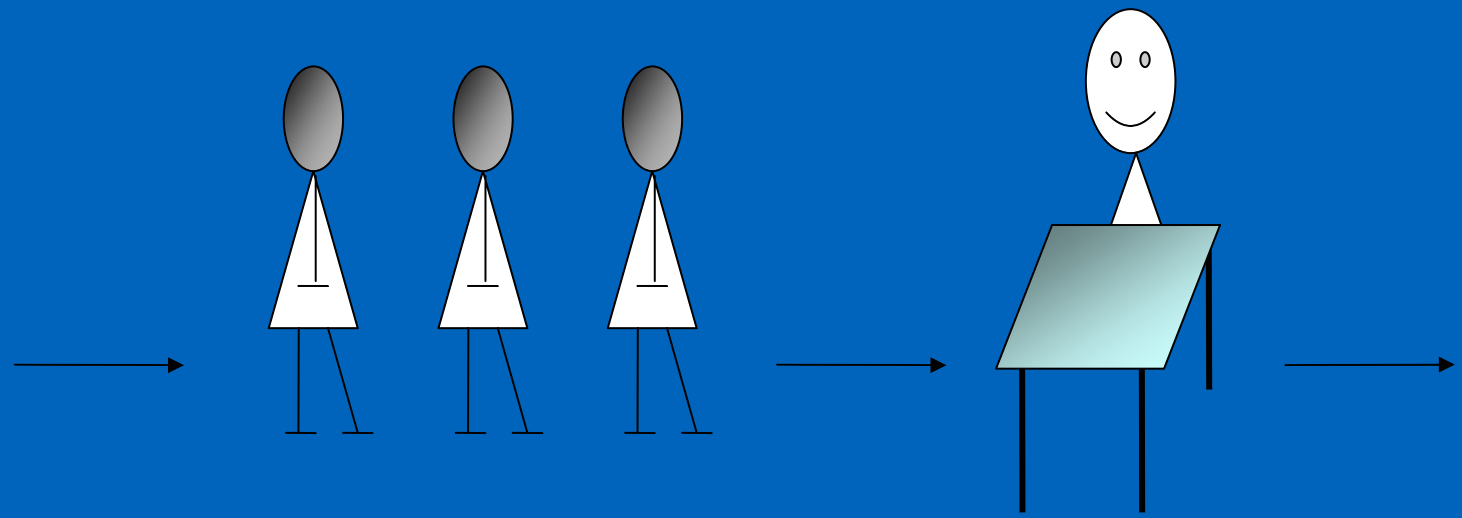
3) 领号

服务业中的排序问题(Schedule Problem in Service Industry)

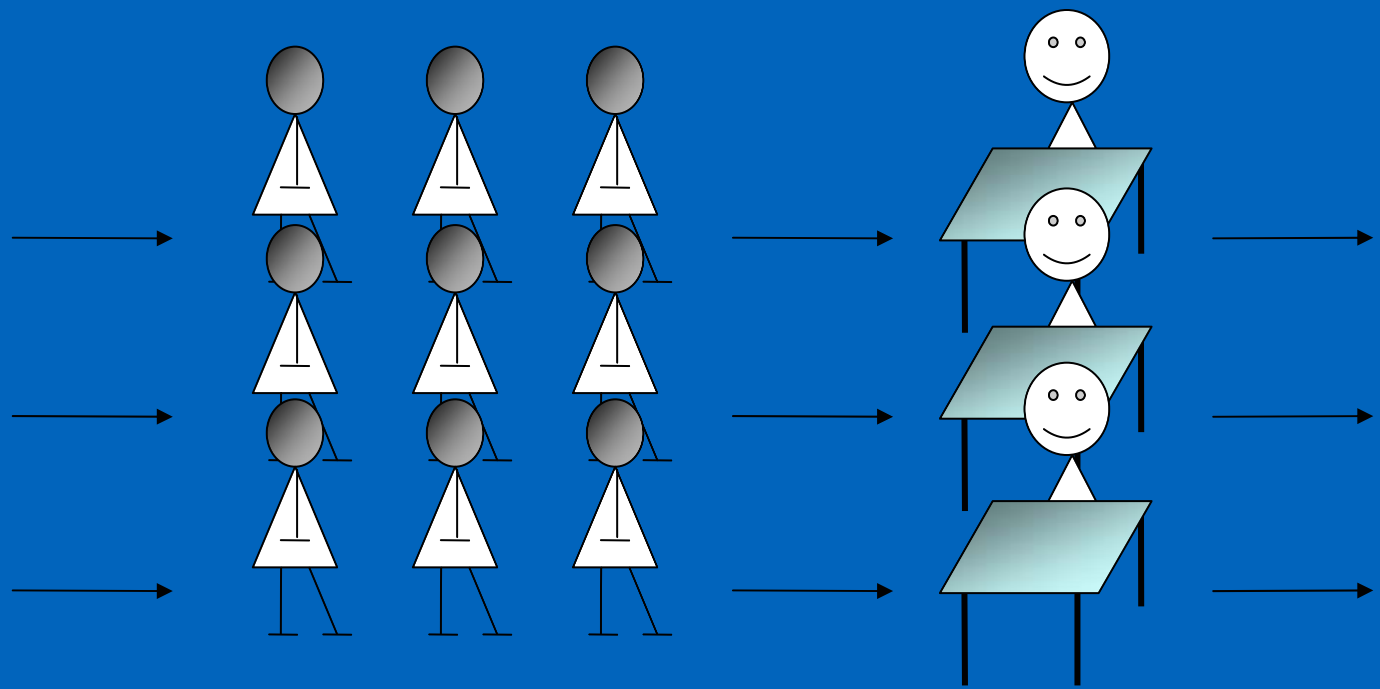
4、服务结构

1) 单队单服务台单阶段的服务排队系统
2) 多队多服务台单阶段的服务排队系统
3) 单队多服务台单阶段的服务排队系统
4) 单队单服务台多阶段的服务排队系统
5) 单队多服务台多阶段的服务排队系统

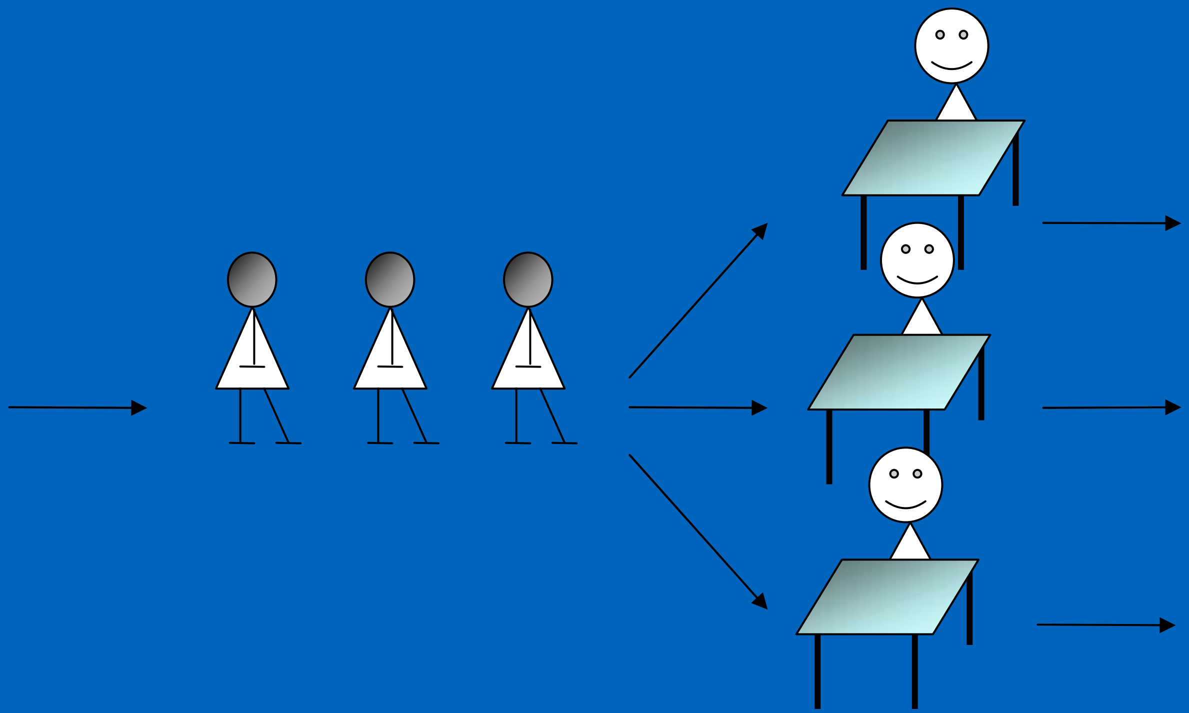
单队单服务台单阶段的服务排队系统



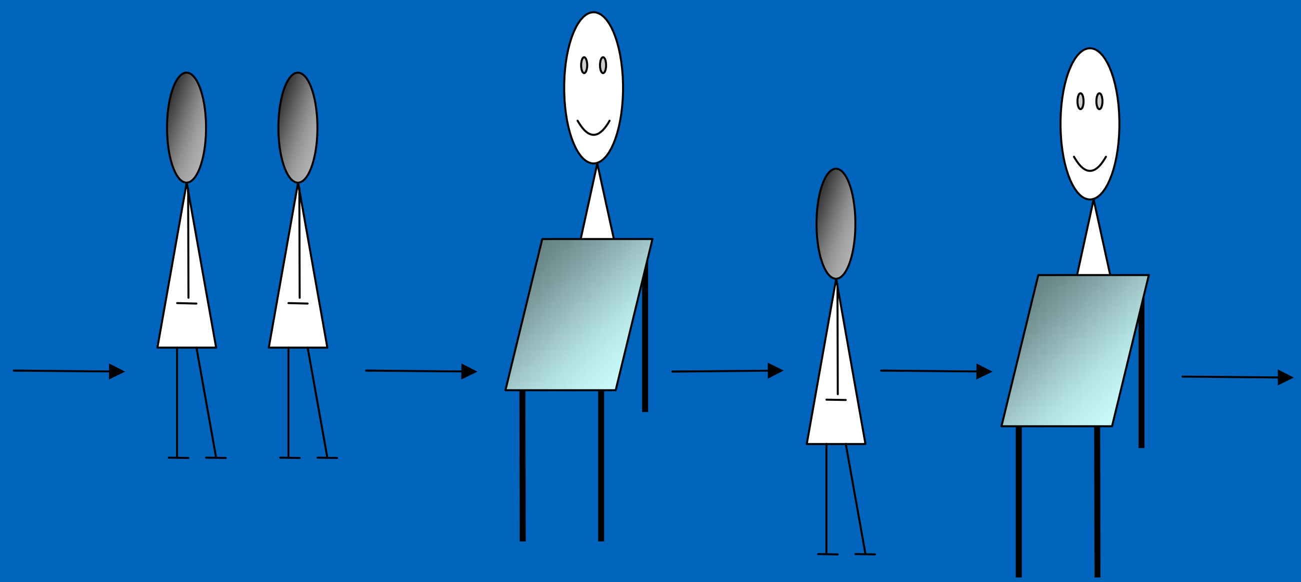
多队多服务台单阶段的服务排队系统



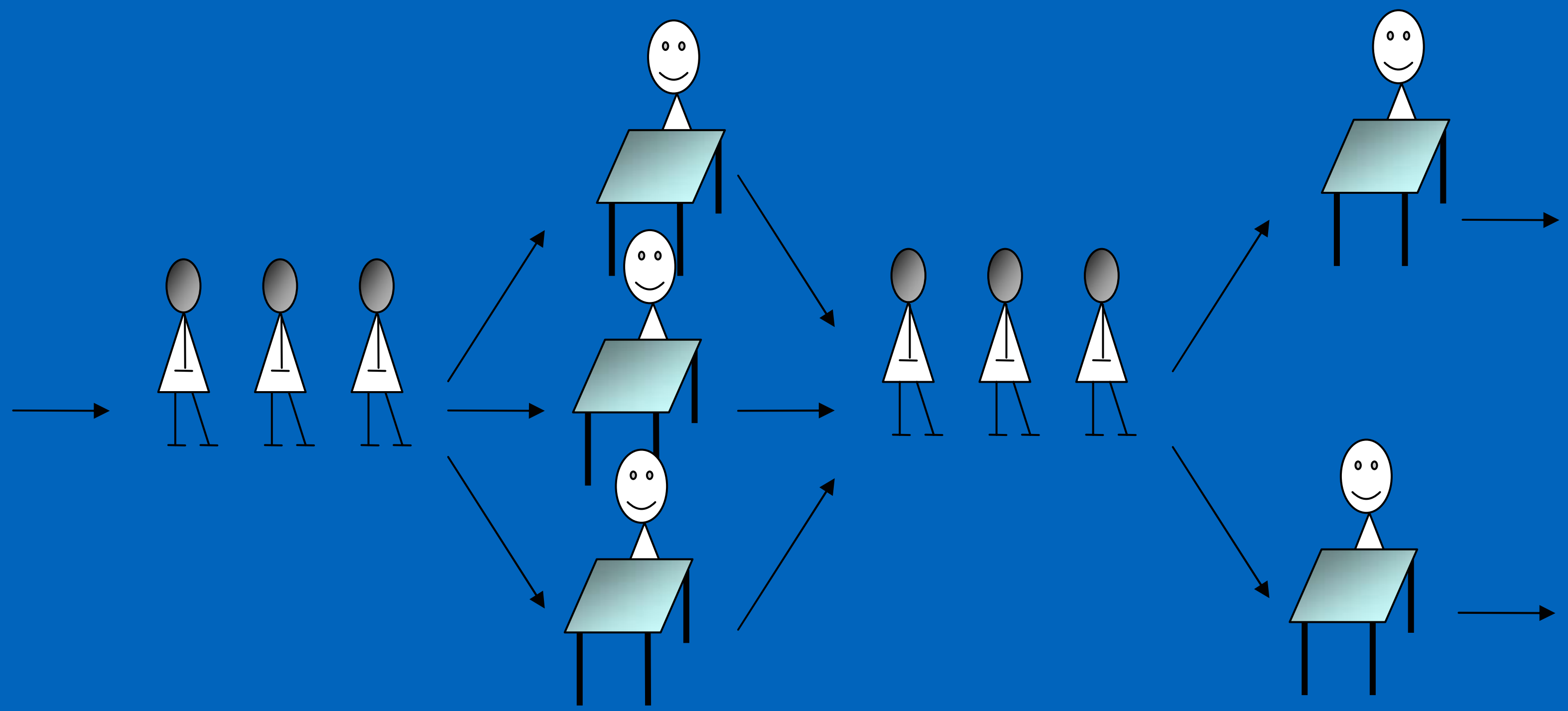
单队多服务台单阶段的服务排队系统



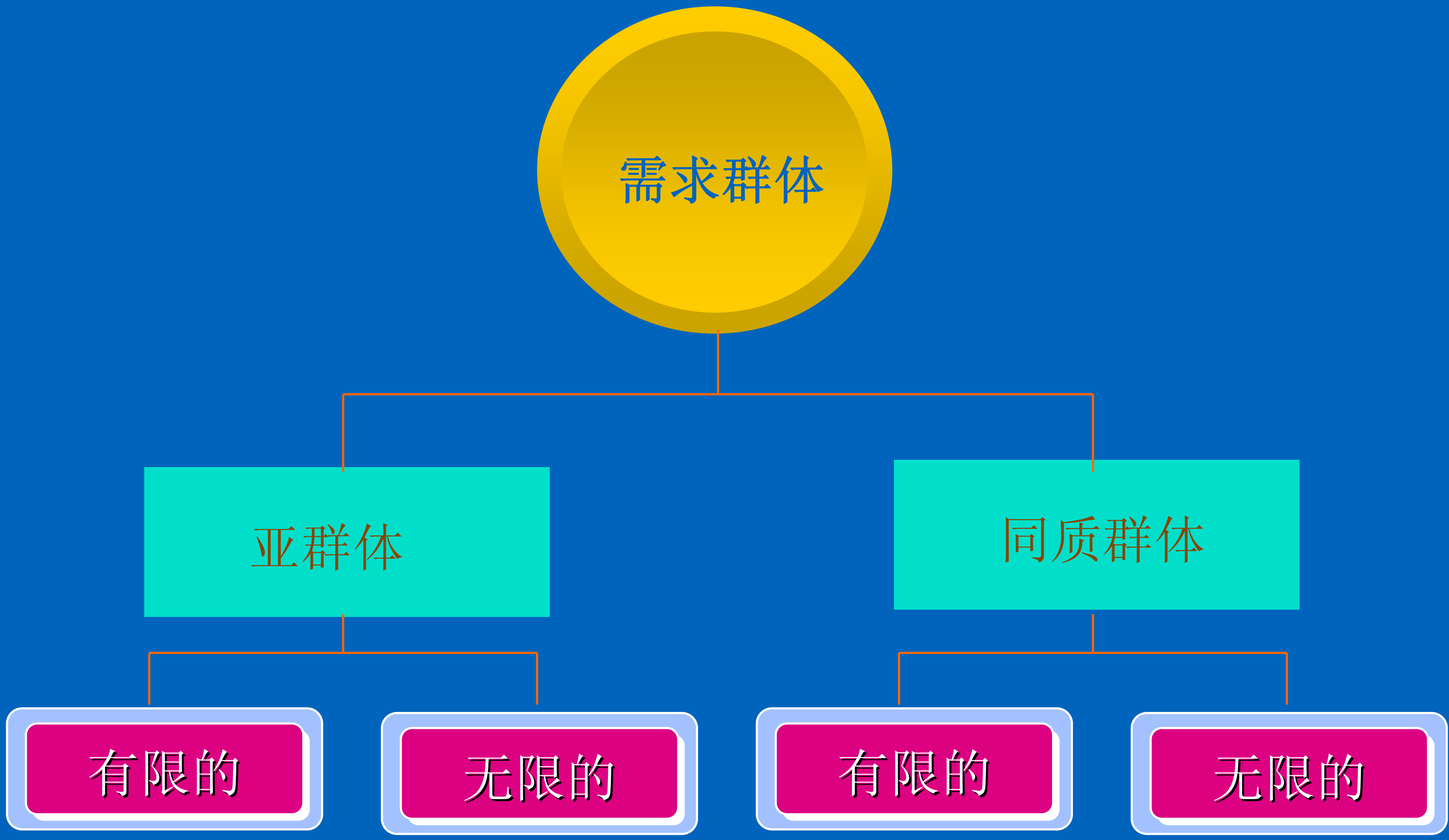
单队单服务台多阶段的服务排队系统



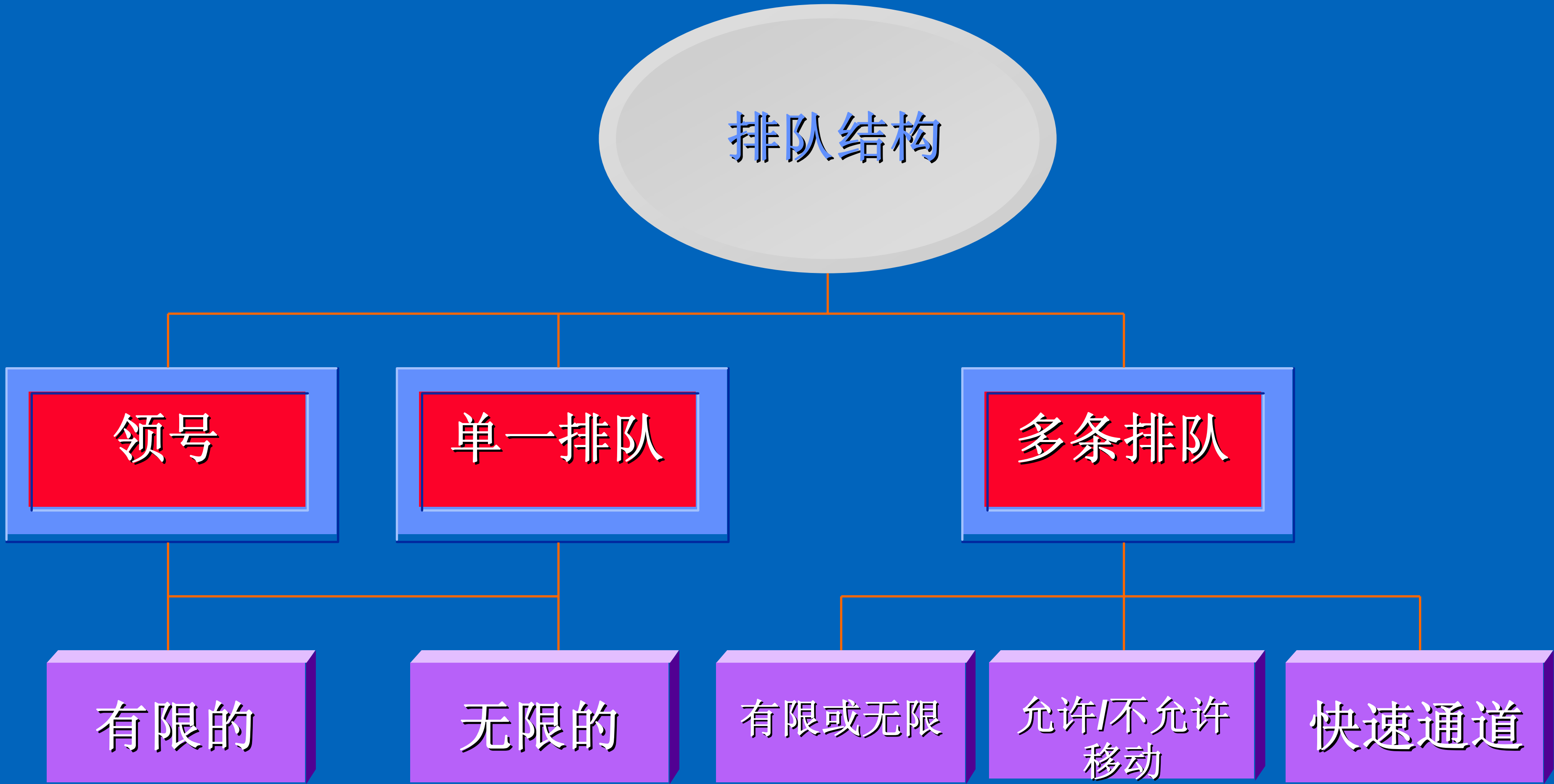
单队多服务台多阶段的服务排队系统



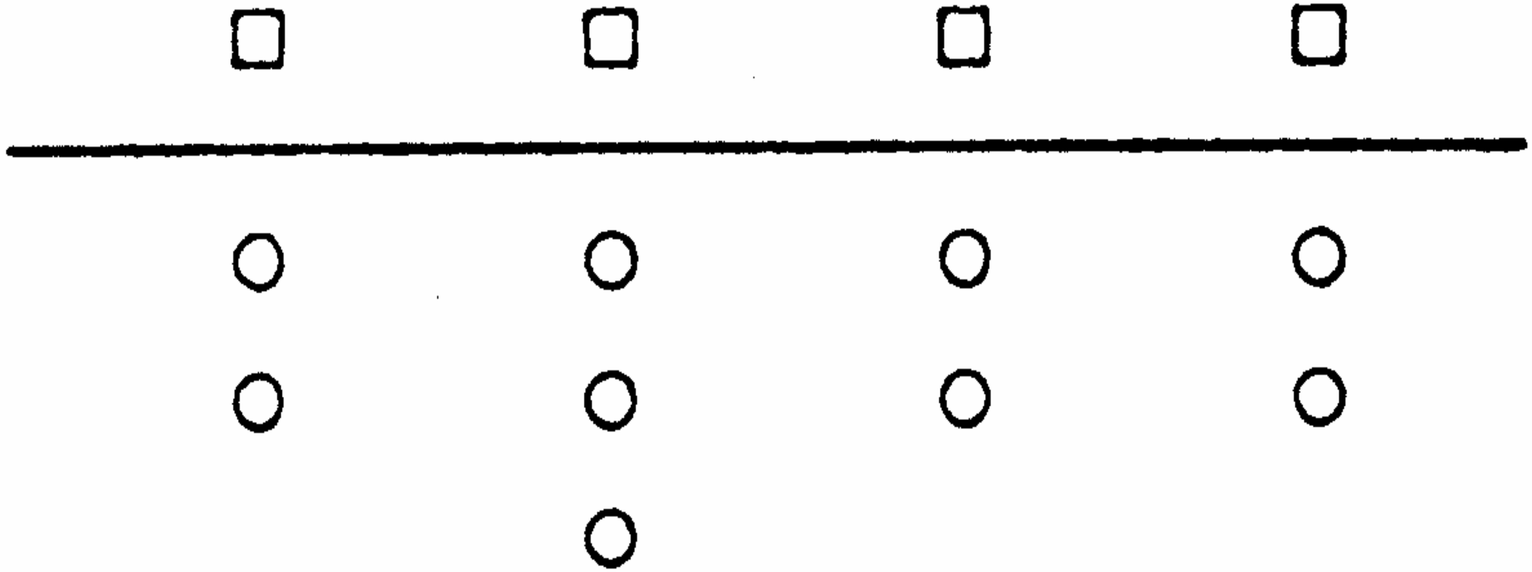
服务业中的排序问题(*Schedule Problem in Service Industry*)



服务业中的排序问题(Schedule Problem in Service Industry)



多条排队



a)

服务业中的排序问题(Schedule Problem in Service Industry)

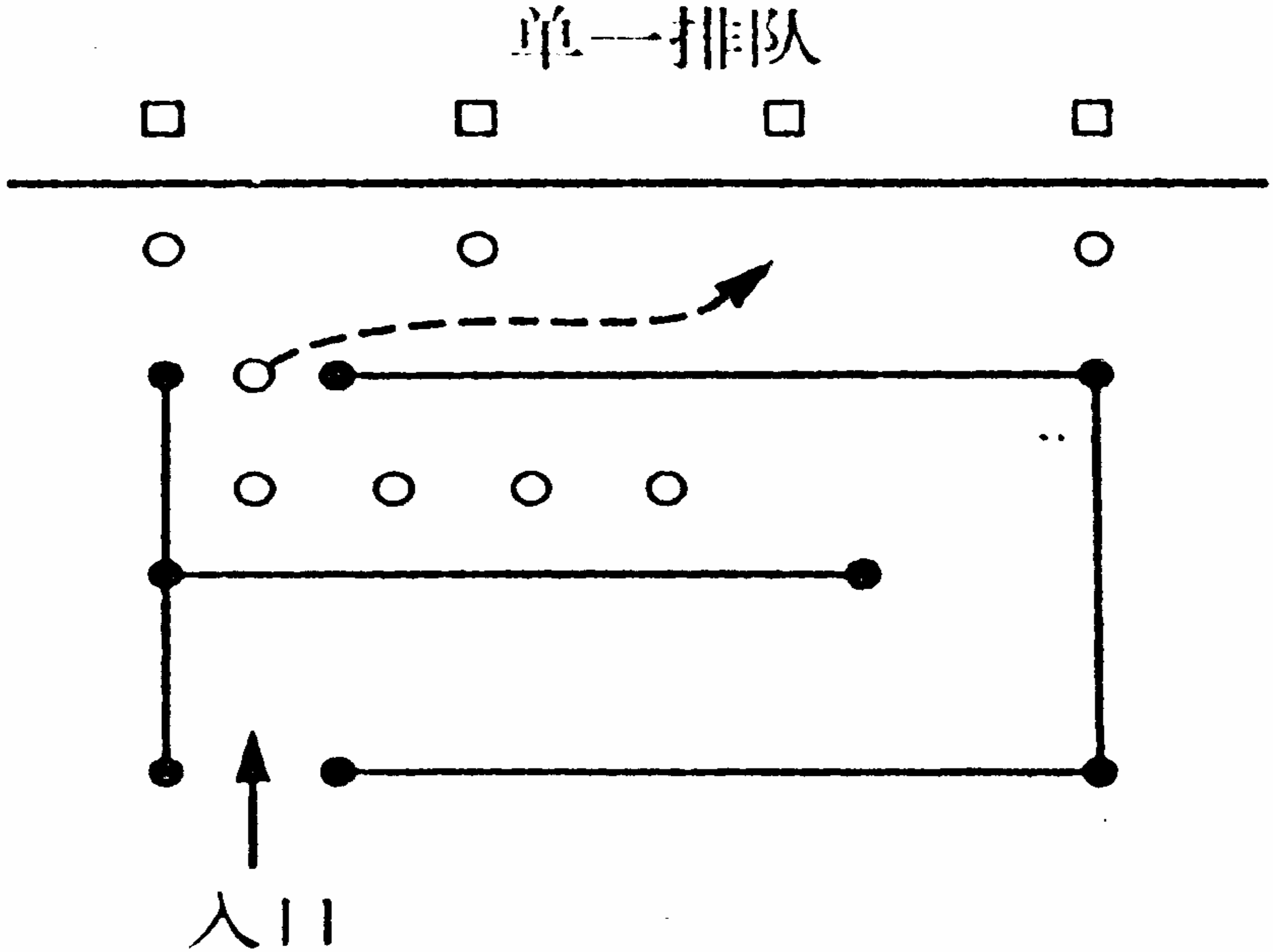
优点

- 多条排队

a) 提供差别服务(Differentiate Service)(超市快速结帐)

b) 顾客可选择

c) 有助于减少不加入队伍的现象



b)

服务业中的排序问题(*Schedule Problem in Service Industry*)

- 单一排队

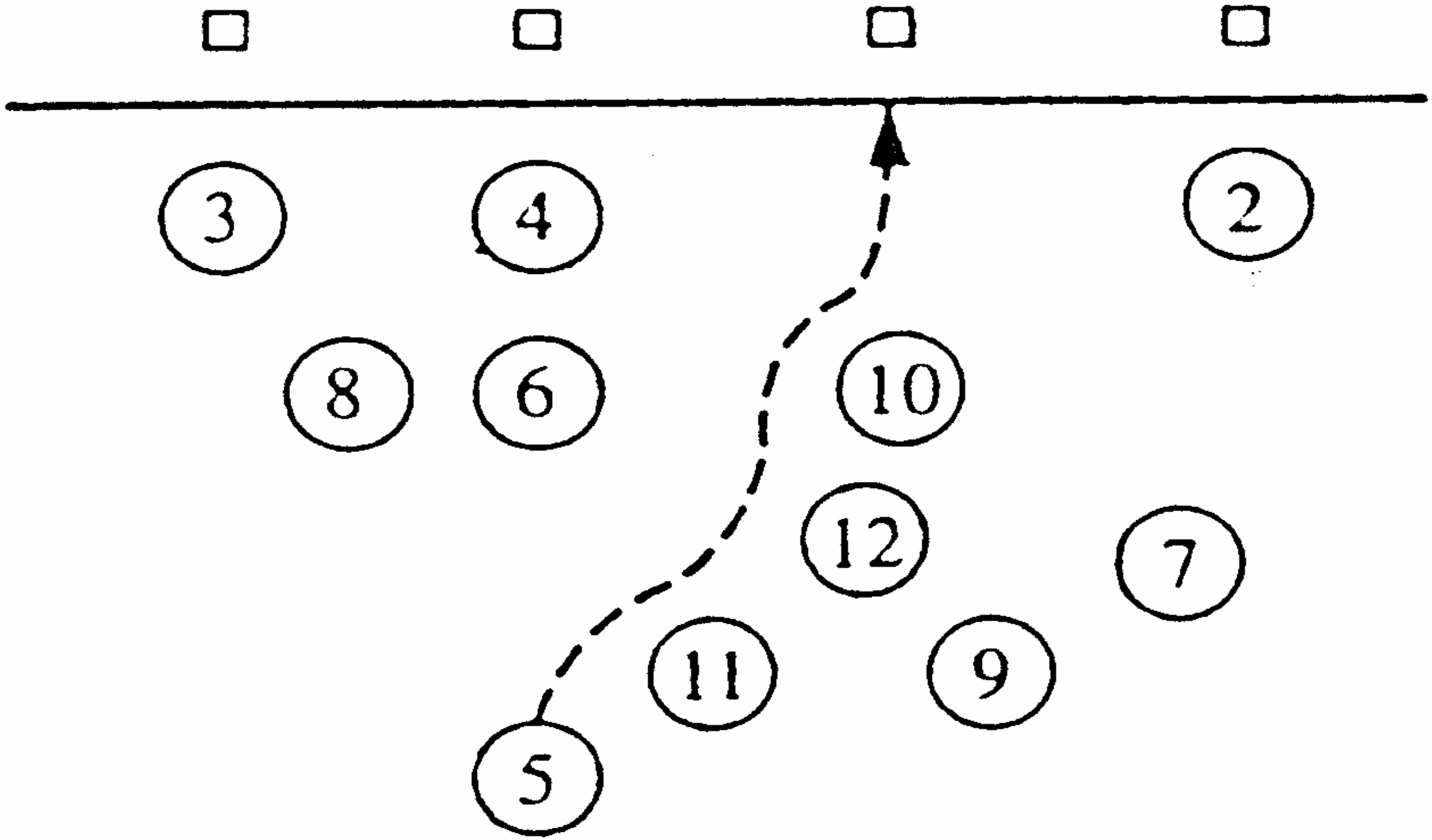
  a） 先到先服务(FCFS, First Come First Serve)

  b） 顾客不会看到别的队伍移动得快而着急

  c） 插队困难

  d） 提高了服务的私密性（一米线 ）

领导



c)

服务业中的排序问题(Schedule Problem in Service Industry)

- 领号

  a） 顾客可四处走动，但必须警觉是否叫到号。

  b） 商店可利用“领号”系统增加顾客冲动购物，浏览、多买

服务业中的排序问题(Schedule Problem in Service Industry)

- 有限排队(Waiting Line with Limited Capacity)
  - 若等待场所无法容纳所有需求服务的顾客，一些人会离去，
  - 这种情况称为有限排队

服务业中的排序问题(Schedule Problem in Service Industry)

5、排队规则

1) FCFS(First Come First Serve)

- 常用到SPT规则，但服务时间长的不断让位于后到达者，所以先用SPT分类，然后FCFS（超市，快速通道）

2) 最高优先权法则(Highest Priority Principle)

- 火警、救护车
- 为减少等待顾客离队可能，应让顾客知道预期等待时间的信息，并使顾客得到定期更新的信息。

服务业中的排序问题(Schedule Problem in Service Industry)

6、调整顾客到达率的措施

1）利用预约系统

控制顾客到达时间

实现最高程度的服务能力利用率

减少顾客等待的时间

提高服务水平

2）采用预订系统

对特定期间的服务需求做出较准确的估计

3）采用差异定价措施

6、调整服务能力的策略

1）进行有效的人员班次排序

2）利用临时工或兼职人员

3）招聘和培养多技能的员工

4）对组织结构、服务流程进行重组
（重复工作 如多封同一信件广告或流程重叠空档 造成能力浪费）

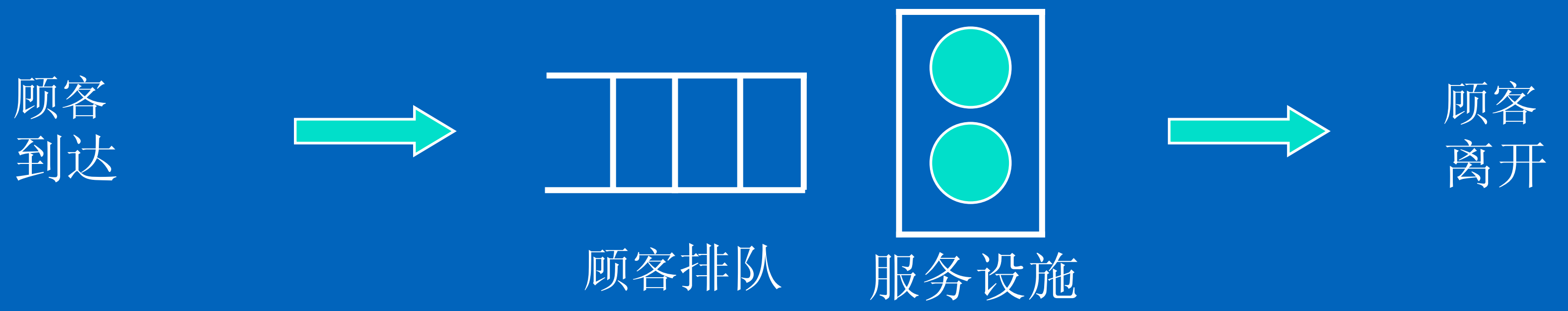
排队理论

- 是二十世纪初在研究电话系统时提出的方法
- 排队是指顾客或工件排队等待服务或加工

随机到达

- 我们假设，在某段时间内顾客到达是随机的，互不相关，但平均到达率不变。
  - 顾客到达有许多来源
  - 每个来源到达率很低
  - 顾客到达互不相关
- 例如，假定平均到达率是每小时16位顾客。但每个小时的实际到达人数是不一样的。

为什么会出现排队现象？



- 假定每小时平均有**4**位顾客到达，服务人员为每位顾客的平均服务时间为**15**分钟。如果顾客到达的间隔时间正好是**15**分钟，而服务人员为每位顾客的服务时间也正好是**15**分钟，那么，就只需要一名服务人员，顾客也根本用不着等待。
- 在以下情况将出现排队现象：
  - 平均到达率（顾客/小时）高于平均服务率（顾客/小时），就像红酸果案例中的情况一样。
  - 顾客到达的间隔时间不一样（随机）。
  - 服务时间不一样 （随机）。

衡量服务系统表现的尺度

- 输出率 =单位时间内系统所服务的平均顾客数。
- ρ = 利用率 = 服务时间所占的比例
- *L* = 系统中的平均顾客数 (包括正被服务的)
- *W* = 每位顾客在系统中平均逗留时间，即流程时间 (包括接受服务时间)
- L_q = 在队列中等待的平均顾客数
- W_q = 每位顾客的平均等待时间，即排队时间

影响等待时间的因素

- **顾客：以泊松分布方式到达**

  λ = 平均到达率 (单位时间到达的顾客数）

- **服务时间：**

  M **= 平均服务时间**

  σ = 服务时间的方差

  服务率（人/小时）： $\mu = \dfrac{1}{M}$

◆ 服务台数： c

排队理论是设计和改进生产系统的重要工具

- 排队理论的主要目标是：由系统的参数（顾客到达率、服务时间、服务台数等等）出发，估计系统的表现，特别是在时间方面的表现。
- 主要方法：
  - 公式
  - 近似公式
  - 电脑模拟

服务台利用率是衡量表现的重要指标

服务台利用率就是每个服务台实际工作时间所占比例。

- **服务台利用率 =** $\dfrac{\text{平均实际服务时间}}{\text{总计可利用时间}}$
- **实例：**
  - **平均到达率为每小时10位顾客；**
  - **每位顾客平均要求10分钟的服务（服务时间），这样在一个小时内，平均就有100分钟或者说 1.667个小时的工作量；**
  - **共有两个服务台，每个可利用的时间为一个小时，这样总计可利用时间为两个小时。 .**
  - **因此服务台利用率为1.667/2 = 83.3%.**

另一种计算服务台利用率的方法

- 服务台利用率 = $\dfrac{\text{平均到达率}}{\text{平均服务率}}$

- 实例：
  - 平均到达率为每小时10位顾客；
  - 每个服务台的平均服务时间为10分钟，这样一个服务台平均一小时可以为6位顾客提供服务；
  - 共有两个服务台；
  - 因此服务台利用率为 10/(2*6) = 0.833 或者说83.3%

三种重要的关系

- "管道原理": 在一个稳定系统中

  平均输出 = 平均输入，或者输出率 = λ

- 利特尔法则:

$$L = \lambda W$$

$$L_q = \lambda W_q$$

- 时间的可加性：在系统中逗留的时间等于服务时间加排队时间:

$$W = M + W_q = \frac{1}{\mu} + W_q$$

利特尔法则

- 根据利特尔法则：

  系统内的平均顾客数 = 平均到达率 × 平均逗留时间。

  实例1：某医疗机构的平均到达率为每个月50位患者，平均逗留时间为两个月。该医院内的平均患者数为多少？

  实例2：某银行平均每天收到支票金额50万元，支票的平均传送、处理时间为三天。假定年利率为10%，该银行每年在支票交易上的利息损失是多少？

- 利特尔法则不仅适用于整个系统，而且也适用于系统的任何一部分。

单个服务台；任意服务时间分布

- 一个稳定系统的条件：

  利用率小于 1：

$$\rho = \lambda M = \frac{\lambda}{\mu} < 1$$

- 计算公式：

$$W_q = \frac{\lambda(M^2 + \sigma^2)}{2(1-\rho)}$$

- σ 数值增大就意味着服务时间有更多的可变性。而更多的可变性则会造成更多的拥塞和延误，这样就要等待更长的时间。

实例

- 以下是两个计算机系统的有关数据：

	1 号系统	2 号系统
M (分钟)	3.0	3.1
σ (分钟)	2.1	1.0

- 顾客的到达率为每小时**15**位。也就是说 **λ = 15/60 = 0.25** 位顾客**/**分钟 **(**注意**:**使用相同的时间单位！**)**

实例（续）

- **1号计算机系统：**

$$\rho = \lambda M = 0.25 \times 3.0 = 0.75$$

$$W_q = \frac{0.25 \times (3.0^2 + 2.1^2)}{2(1-0.75)} = 6.71\text{分钟}$$

$$L_q = \lambda W_q = 0.25 \times 6.71 = 1.68\text{位顾客}$$

$$W = M + W_q = 3 + 6.71 = 9.71\text{分钟}$$

$$L = \lambda W = 0.25 \times 9.71 = 2.43\text{位顾客}$$

实例（续）

- 2号计算机系统：

$$\rho = \lambda M = 0.25 \times 3.1 = 0.775$$

$$W_q = \frac{0.25 \times (3.1^2 + 1^2)}{2(1 - 0.775)} = 5.89 \text{ 分钟}$$

$$L_q = \lambda W_q = 0.25 \times 5.89 = 1.47 \text{ 位顾客}$$

$$W = M + W_q = 3.1 + 5.89 = 8.99 \text{ 分钟}$$

$$L = \lambda W = 0.25 \times 8.99 = 2.25 \text{ 位顾客}$$

慢的反而好！

单个服务台；服务时间为指数分布

- 服务时间为指数分布，那么， ***M* = σ =1/μ**
- 利用率：

$$\rho = \lambda M = \frac{\lambda}{\mu} < 1$$

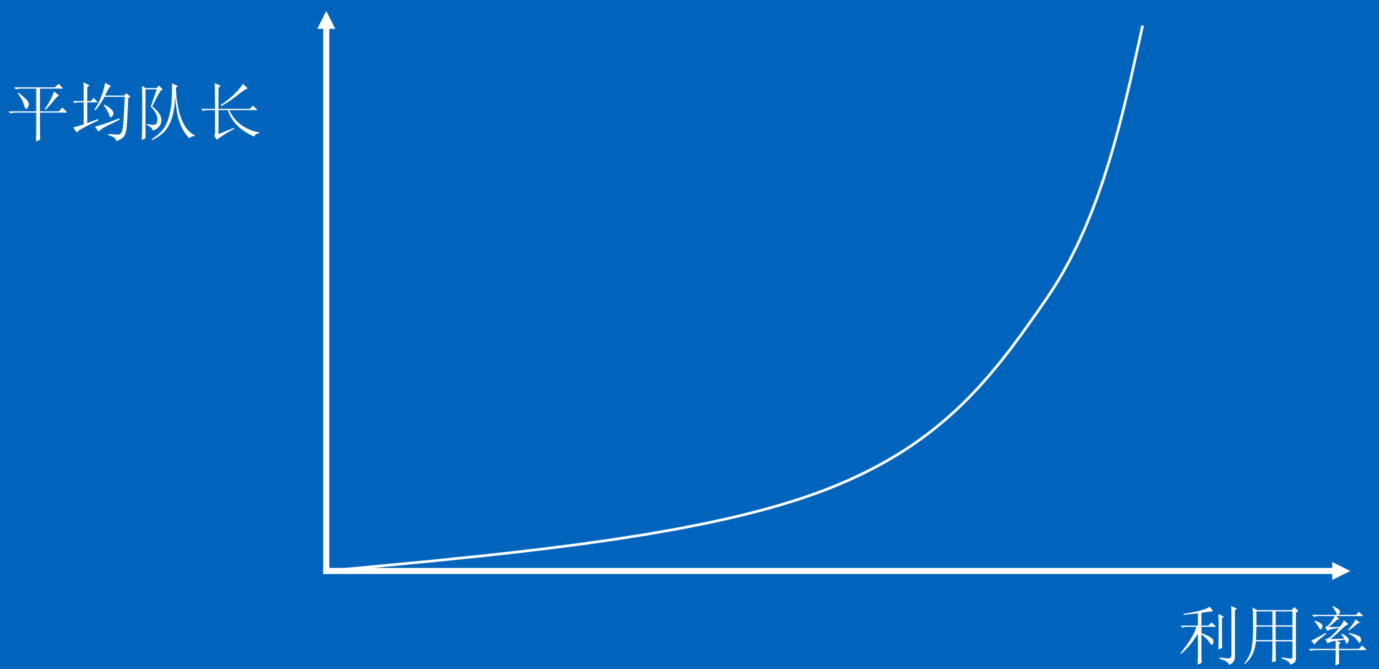
- 计算公式：

$$W_q = \frac{\rho}{\mu(1-\rho)}$$

- 如果把到达数量增加一倍(2λ)，再把服务率增加一倍 (2μ), 等待时间就会减少一半 。这就是排队论的规模经济学。

平均队长与利用率的关系

- 平均队长（和平均等待时间）与服务台利用率之间的关系不是线性的关系。资产利用率太高会造成服务质量急速下降，因而要权衡利弊。要保证服务质量，就必须保持“过剩的”生产或服务能力。



实例

一个服务台：指数服务时间

到达率 = 10/小时

利用率	平均 排队长度	平均 等待时间
0.2	0.05	0.005 小时 (0.3 分钟)
0.5	0.50	0.05 小时 (3.0 分钟)
0.8	3.20	0.32 小时 (19.2 分钟)
0.9	8.10	0.81 小时 (48.6 分钟)
0.95	18.05	1.805 小时 (200.3 分钟)

多个服务台;服务时间为指数分布

- 利用率：

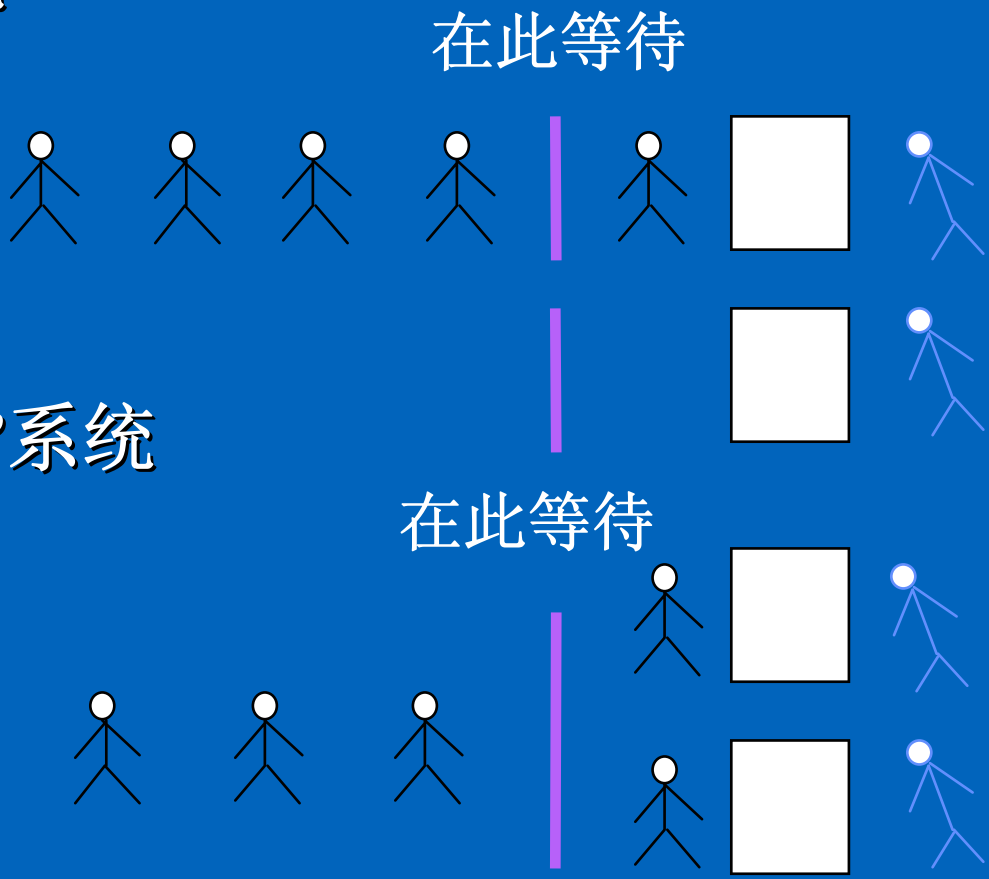
$$\rho=\frac{\lambda M}{c}\lambda=\frac{\lambda}{c\mu}<1$$

- 计算公式较为复杂。我们可使用教材（**Chase**第**8**版）中的表 **S5-4** 或电子表格程序。
- 如果把到达数量增加一倍(**2λ**)，再把服务人员增加一倍(**2*c***), 等待时间就会减少一半以上。这就是排队论的规模经济学。

λ	M	c	ρ	Wq
10	0.2	3	66.7%	0.09
20	0.2	6	66.7%	0.03
20	0.2	5	80%	0.11

为何“集中使用”可以提高服务质量？

- 独立的系统
- “集中使用”系统



有关排队模型的要点

- 集中使用的优点
  - 将资源组合在一起为所有的顾客提供服务，可以在等待时间不变的条件下，减少所需要的资源总量。如果是两列独立排队，那么客户可能要等那位指定的服务人员提供服务，这位服务人员可能当时正忙得抽不开身，而另一位服务人员却闲着没事干。在集中使用的系统中就不会出现这种现象。
- 大规模制造或服务设施的规模经济学
  - 在保持同样利用率的情况下减少平均等待时间
  - 在保持同样平均等待时间的情况下提高利用率

有关排队模型的要点

- 假如平均服务率大于平均到达率，那末等待是由到达和服务时间的变动性造成的。
- 瓶颈是由两方面原因造成的：利用率太高或可变性太大。
- 在系统规模不变时，利用率越高，等待时间越长。但两者并非线性关系。利用率越高，等待时间增长越快。
- 利特尔法则：假如平均到达率不变，平均等待时间与平均等待人数会按比例增长。

减少等待时间的办法

- 减少平均服务时间
- 减少服务时间的可变性
  - 如果能将服务时间的可变性从很大（服务时间的均值等于标准差）减到零(服务务时间为常数)，那么平均等待时间将可以减少大约一半。
- 增加服务人员
- 减少平均到达人数
- 通过顾客预约等办法来减少到达的可变性
- 集中使用服务资源！
- 更好地计划和调度

掌握客户的感受:等待心理学

客户所感受到的等待可能与实际的等待有很大的差别。研究结果表明：

- 服务越有价值，人们就越愿意多等一会儿。
- 服务开始之前的等待感觉要比服务过程中的等待要长。
- 心急会让人觉得等待的时间很长。
- 不公平的等待比公平的等待要长。
- 不确定的等待比已知的等待要长。
- 不明情况的等待要比知情的等待要长。
- 没事干的时候会让人觉得比有事干的时候要长。
- 独自等待会让人觉得比大家一起等待要长。