

• 第二部分：相关数学理论

第5章 排队论及其应用

5.1 排队论基础

5.2 M/M/m/(n)排队系统

5.3 通信业务量分析

5.4 随机接入系统业务分析

4

第5章 排队论及其应用

- 通信网规划设计和优化遵循的原则：
能够满足各项性能指标要求又节省费用的设计或优化方案。
- 对设计人员的要求：掌握相应的理论基础知识和网络分析的计算方法，以便对通信网的性能进行分析与指标计算，为**设计和优化**提供理论数据。
- 应用的数学理论：排队论。它起源于最早的电话系统，可应用于很多领域，目前通信网仍是其中一个重要的应用领域。

2

5.1 排队论基础

5.1.1 基本概念

5.1.2 有关的概率模型及最简单流

5.1.3 排队系统的主要性能指标

5

本章学习要求

- 重点掌握和理解排队论的基本概念、M/M/m(n)排队系统的模型分析方法，了解它们在网络中的实际应用；
- 掌握通信网业务量的基本概念，理解、掌握和运用 Erlang B公式和C公式；能够运用这些知识分析和计算实际网络的性能指标；
- 掌握随机接入系统的工作原理及其业务分析方法。

3

5.1.1 基本概念

排队论（Queuing Theory）：

- 是一个独立的数学分支，有时也把它归到运筹学中。
- 排队论是专门研究由于**随机因素**的影响而产生的拥挤现象(排队、等待)的科学，也称为随机服务系统理论或拥塞理论（Congestion Theory）。
- 它是在研究各种排队系统概率规律性的基础上，解决有关**排队系统的最优设计和最优控制问题**。

design & analysis

6

5.1.1 基本概念

排队论的起源:

- 排队论起源于20世纪初。当时,美国贝尔(Bell)电话公司发明了自动电话以后,如何合理配置电话线路的数量,以尽可能地减少用户重复呼叫次数问题出现了。
- 1909年,丹麦工程师爱尔兰(A.K.Erlang)发表了具有重要历史地位的论文“概率论和电话交换”,从而求解了上述问题。
- 1917年,A.K.Erlang又提出了有关通信业务的拥塞理论,用统计平衡概念分析了通信业务量问题,形成了概率论的一个新分支。
- 后经C.Palm等人的发展,由近代概率论观点出发进行研究,奠定了话务量理论的数学基础。

7

5.1.1 基本概念

1. 排队的概念

通信网中的排队现象:

- 无形的排队:如打电话
- 有形的排队:如数据分组的传送

顾客

- 把要求服务的一方统称为“顾客”,如电话用户产生的呼叫和待传送的分组信息。

服务机构

- 把提供服务的一方统称为服务机构,如电话交换设备、信息传输网络等。

服务窗口或服务人员

- 把服务机构内的具体设施统称为“服务窗口”或“服务人员”,如中继线路、信道等。

排队系统(随机服务系统)

- 由要求随机性服务的顾客和服务机构两方面构成的系统称为随机服务系统或排队系统。

10

5.1.1 基本概念

经过通信、计算机和应用数学三个领域的研究学者的努力,排队论得到了迅速的发展和应用。

应用:

- 网络的设计和优化方法;
- 移动通信系统中的切换呼叫的处理方法;
- 随机接入系统的流量分析方法;
- ATM业务流的数学模型及其排队分析方法等。

经典排队论

- 把相继到达“顾客”的到达时间间隔和服务时间都相互独立的排队论内容称为经典(或古典)排队论。
- 经典排队论仍是新的排队论的基础,而且通信领域的许多问题可以用它来解决。

8

5.1.1 基本概念

产生排队的原因:

- 顾客需求的随机性和服务设施的有限性。

应用的理论:

- 概率论和随机过程理论

研究目的:

- 研究排队系统内服务机构与顾客需求之间的关系,以便合理地设计和控制排队系统,使之既能满足一定的服务质量要求又能节省服务机构的费用。

11

- (Kleinrock) "We study the phenomena of standing, waiting, and serving, and we call this study Queueing Theory."
- "Any system in which arrivals place demands upon a finite capacity resource may be termed a queueing system."

00:07:39

5.1.1 基本概念

2. 排队系统的组成

一个排队系统由三个基本部分组成:

- 输入过程
- 排队规则
- 服务机构

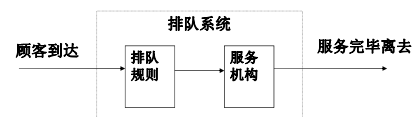


图5.1 排队系统的基本组成

12

5.1.1 基本概念

(1) 输入过程

描述顾客按怎样的规律到达排队系统，包括以下三方面：

- **顾客总数**：指顾客的来源（简称顾客源）数量，顾客源数可以是**无限**的，也可以是**有限**的。
- **顾客到达方式**：描述顾客是怎样到达系统的，是成批（集体）到达（每批数量是随机的还是确定性的）还是单个到达。
- **顾客流的概率分布（或顾客到达的时间间隔分布）**：所谓顾客流，就是顾客在随机时刻一个个（一批批）到达排队系统的序列。

13

5.1.1 基本概念

- **非拒绝系统**：又称**非拒绝方式**、**非截止型系统**。系统排队队长无限制，允许顾客排队等待（一般认为顾客数是无限的）。例如，公用电话。
要求该类系统稳定性参数 ρ 要满足 $\rho < 1$ 。

——— 到达率低于服务率

- ✓ **即时拒绝系统**：也称为**立接制系统**、**损失制系统**；
- ✓ **延时拒绝系统**：也称为**混合制系统**；
- ✓ **延时拒绝系统**和**非拒绝系统**：也称为**等待制系统**、**缓接制系统**。

16

5.1.1 基本概念

(2) 排队规则

排队规则包括：

- 1排队系统类型
- 2服务规则

1排队系统类型：

- ✓ **拒绝系统**
- ✓ **非拒绝系统**

表明服务机构是否允许顾客排队等待服务。

14

• 对应——排队规则

- **损失制**—— 顾客到达系统时，如果系统中所有服务窗均被占用，则到达的顾客随即离去
- **等待制**—— 顾客到达系统时，如果所有服务窗均被占用，则系统能够提供足够的排队空间让顾客排队等待
- **混合制**—— 是损失制与等待制混合组成的排队系统，此系统仅允许**有限个**顾客等候排队，其余顾客被拒绝



00:07:32

5.1.1 基本概念

- **拒绝系统**：又称**拒绝方式**、**截止型系统**。

n ：系统允许排队的队长（也称截止队长）；

m ：窗口数。

分为两种情况：

- ✓ **即时拒绝系统** L ： $n=m$ 的系统。此时，顾客到达后或立即被拒绝，或立即被服务，不存在排队等待服务的情况。电话网就是即时拒绝系统。
- ✓ **延时拒绝系统**： $m < n$ 的系统。此时容许一定数量的顾客排队等待，当系统内顾客总数达到**截止队长**时，新来的顾客就被拒绝而离去。带有缓冲存储的数据通信、分组交换等就属于这一类。

15

5.1.1 基本概念

• 2服务规则

- **先到先服务**（FCFS）或**先入先出**（FIFO）
这是常见的情况。若无其他说明时，常按这种方式来分析。
- **后到先服务**（LCFS）
- **优先制服务**
在通信网中，这种情况也较为常见。
- **随机服务**
通信网中一般是顺序服务，但有的也采用优先制服务方式。

16

5.1.1 基本概念

(3) 服务机构

服务机构包括以下三方面内容。

➤ 窗口或服务数量

- ✓ 当 $m = 1$ 时，称为单窗口排队系统。
- ✓ 当 $m > 1$ 时，称为多窗口排队系统。

➤ 服务方式及排队方式

➤ 服务时间分布

19

5.1.1 基本概念

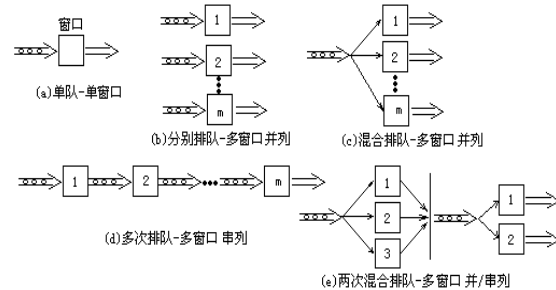


图5.2 服务方式与排队方式

22

5.1.1 基本概念

• 服务方式及排队方式

➤ 服务方式

是指在某一时刻系统内接受相同服务的顾客数，即是单个顾客接受服务（**串联**服务方式）还是成批顾客同时接受服务（**并联**服务方式）。

- ✓ **串联服务方式**：即 m 个窗口的串联排队系统。此时， m 个窗口服务的内容互不相同，某一时刻只能有一个顾客接受其中一个窗口的单项服务，每个顾客要依次经过这 m 个窗口接受全部的服务。
- ✓ **并联服务方式**：即 m 个窗口的并联排队系统。此时， m 个窗口服务的内容相同，系统一次可以同时服务 m 个顾客。

20

5.1.1 基本概念

• 服务时间分布

- 服务时间和顾客到达时间一样，多数情况下是随机型的。要知道它的经验分布或概率分布。
- 一般说来，服务时间的概率分布有定长分布、指数分布、Erlang分布等。

3. 排队系统的三个基本参数

任何排队系统都有三个基本参数 m 、 λ 、 μ ，称为排队模型的三要素。

23

5.1.1 基本概念

➤ 排队方式：包括混合排队和分别排队两种方式。

- ✓ **混合排队方式**：顾客排成一个队列，接受任意一个空闲窗口的服务。
- ✓ **分别排队方式**：顾客排成 m 个队列，同时分别接受 m 个窗口的相同服务。

- 当 $m = 1$ 时，在该系统中，如果允许排队，顾客则只能排成一列队列接受服务。
- 当 $m > 1$ 时，在该系统中，如果允许排队，则有混合排队和分别排队两种排队方式。**排队方式的选择取决于两种服务方式。**

21

5.1.1 基本概念

m 参数

称为窗口数或服务数量，表征系统的资源量。它表示系统中有多少服务设施可同时向顾客提供服务。

λ 参数

顾客到达率或系统到达率，即单位时间内到达系统的平均顾客数。其单位为 **个/时间** 或 **份/时间**。

- ✓ λ 反映了顾客到达系统的快慢程度，也反映了需要服务的一方对提供服务的一方的要求。
- ✓ λ 越大，说明系统的负载越重。
- ✓ λ 的倒数称为平均到达时间间隔 \bar{T} ，即

$$\bar{T} = 1/\lambda$$

24

5.1.1 基本概念

系统的有效到达率 λ_e :

实际能够进入系统并接受服务的到达率, 即单位时间内进入系统的平均顾客数, 有

$$\lambda_e = (1 - P_n) \lambda \quad (5.1)$$

P_n 为阻塞概率 (或拒绝概率)。对于**非拒绝系统**, $P_n = 0$
则 $\lambda_e = \lambda$

- 在无限顾客源的情况下, 顾客的到达按**集体到达**方式考虑的, 式 (5.1) 中的系统到达率 λ 为常数。

25

5.1.1 基本概念

ρ 对系统稳定性的影响:

- 若 $\rho < 1$, 即 $\lambda < m\mu$ 时, 说明平均到达系统的顾客数小于平均离开系统的顾客数。这时系统是稳定的, 可以采取非拒绝方式或拒绝方式。
- 若 $\rho \geq 1$, 即 $\lambda \geq m\mu$ 时, 说明平均到达系统的顾客数多于平均离开系统的顾客数。
✓ 必须采用拒绝方式, 人为地限制系统内的顾客数量, 保证系统的稳定性。

26

5.1.1 基本概念

- 在**有限顾客源**的情况下, 顾客的到达按单个到达方式考虑。
- ✓ 设每个顾客的到达率 λ_0 相同, λ_0 是有限顾客源中每个顾客在单位时间内到达系统的平均数。
- ✓ N : 顾客源总数; L_s : 系统的平均队长。
($N - L_s$): 系统外的顾客平均数为,
此时**系统的有效到达率**为

$$\lambda_e = (N - L_s) \lambda_0 \quad (5.2)$$

k 状态 (系统中有 k 个顾客时) 的系统到达率为

$$\lambda_k = (N - k) \lambda_0 \quad (5.3)$$

26

5.1.1 基本概念

4. 排队系统分类的表示方法

目前较为广泛采用的分类表示方法是D.G.Kendall提出的分类方法。

$$X / Y / m (n, N)$$

X: 顾客到达时间间隔分布

Y: 服务时间分布

m: 窗口或服务员工数目 (此处特指并列排队系统)

n: 截止队长, 省略这一项表示 $n \rightarrow \infty$, 即为非拒绝系统

N: 表示潜在的顾客总数, 对于潜在的无限顾客源, 即 $N \rightarrow \infty$ 时, 可省去这一项

29

5.1.1 基本概念

μ 参数

是一个服务员 (或窗口) 的服务速率, 即**单位时间**内由一个服务员 (或窗口) 进行服务所离开系统的平均顾客数。

- ✓ $\bar{\tau}$: μ 的倒数 $1/\mu$, 是单个窗口对顾客的平均服务时间, 也是一个呼叫的**平均持续时间**。

$$\bar{\tau} = 1 / \mu$$

ρ : 为排队强度, 又称**稳定性参数**。

$$\rho = \lambda / m\mu \quad (5.4)$$

27

5.1.1 基本概念

表示不同输入过程 (顾客流) 和服务时间分布的符号有:

M: 泊松 (Poisson) 流 (或指数分布)。两者都具有**马尔可夫随机过程性质**

D: 定长分布

E_k: k 阶Erlang分布

GI: 一般相互独立的随机分布

G: 一般随机分布

- ✓ **M/M/1**系统: 指顾客流为泊松流、服务时间为指数分布、单窗口排队系统。

- ✓ **M/D/m**系统: 指顾客流为泊松流、服务时间为定长分布、有 m 个窗口的排队系统。

- ✓ 一般如没有特别说明, 则认为顾客总体数是无限源、属于非拒绝方式的排队系统。

30

5.1 排队论基础

5.1.1 基本概念

5.1.2 有关的概率模型及最简单流

5.1.3 排队系统的主要性能指标

31

5.1.2 有关的概率模型及最简单流

2. 排队系统中常用的几个定义

- 系统状态：指一个排队系统中的顾客数（包括正在被服务的顾客数）。
- $N(t)$ ：在时刻 t 排队系统中的顾客数，即系统在时刻 t 的瞬时状态。
- $P_k(t)$ ：在时刻 t 系统中有 k 个顾客的概率。
- λ_k ：当系统中有 k 个顾客时，新到顾客的到达率或系统到达率（单位时间内新顾客的到达数）。

$$\lambda_k = \begin{cases} \lambda & \text{无限顾客源, } k \geq 0, \lambda \text{ 为常数} \\ (N-k)\lambda_0 & \text{有限顾客源, } k \geq 0 \end{cases}$$

34

5.1.2 有关的概率模型及最简单流

1. 排队系统中常用的概率模型

(1) 泊松分布

设随机变量 X 所有可能取的值为 $0, 1, 2, \dots$ ，而取各个值的概率为

$$P_k = P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots \quad (5.5)$$

其中 $\lambda > 0$ 是常数，则称 X 服从参数为 λ 的泊松分布。

$$\text{其均值为} \quad E(X) = \lambda \quad (5.6)$$

$$\text{方差为} \quad D(X) = \lambda \quad (5.7)$$

32

5.1.2 有关的概率模型及最简单流

- μ_k ：当系统中有 k 个顾客时，整个系统的平均服务率（单位时间内服务完毕离去的平均顾客数）。有：

$$\mu_k = \begin{cases} \mu & m = 1 \\ k\mu & m > 1, \quad k \leq m \\ m\mu & m > 1, \quad k \geq m \end{cases}$$

- 稳定状态：当一个排队系统运转一段时间后，系统的状态将独立于初始状态及经历的时间，这时称系统处于稳定状态。排队论中主要研究系统处于稳定状态下的工作情况。稳定状态时工作情况与时刻 t 无关，这时

$P_k(t)$ 可写成 P_k ， $N(t)$ 可写成 N 。

35

5.1.2 有关的概率模型及最简单流

(2) 指数分布

一般，若随机变量 t 取具有概率密度函数为

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases} \quad (5.8)$$

其中 $\lambda > 0$ 为常数，则称 t 服从参数为 λ 的指数分布，其分布函数 $F(t)$ 为

$$F(t) = \int_{-\infty}^t f(t) dt = \int_{-\infty}^t \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t}$$

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases} \quad (5.9)$$

$$\text{其均值为} \quad E(t) = \frac{1}{\lambda} \quad (5.10)$$

$$\text{方差为} \quad D(t) = \frac{1}{\lambda^2} \quad (5.11)$$

33

5.1.2 有关的概率模型及最简单流

3. 最简单流

随机事件流：

- 通常把随机时刻出现的事件组成的序列称为随机事件流，例如用 $N(t)$ 表示 $(0, t)$ 时间内要求服务的顾客人数就是一个随机事件流。

(1) 最简单流

定义：如果一个事件流 $\{N(t), t > 0\}$ ，这里以输入流为例，满足下述三个条件则称该输入为最简单流：

- ✓ 平稳性
- ✓ 无后效性
- ✓ 疏稀性

36

5.1.2 有关的概率模型及最简单流

- **平稳性**：在时间间隔 t 内，到达 k 个顾客的概率只与 t 有关，而与这间隔的起始时刻无关。
即以任意时刻 t_0 为起点， (t_0, t_0+t) 时间内出现的顾客数只与时间长度 t 有关而与起点 t_0 无关，
 $N(t)$ ： (t_0, t_0+t) 内出现的顾客数
 $P_k(t)$ ： $N(t) = k$ 的概率，则

$$P_k(t) = P(N(t) = k) \quad k = 0, 1, 2, \dots$$

$$\sum_{k=0}^{\infty} P_k(t) = 1$$

37

5.1.2 有关的概率模型及最简单流

(2) 泊松过程的顾客到达时间间隔分布

顾客到达时间间隔分布：

- 顾客到达的时间间隔小于 t 的概率，即 t 内有顾客的概率分布。
- 两相邻顾客到达的时间间隔是一连续型随机变量，用 T 表示。在时间内没有顾客到达的概率为

$$P_0(t) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

则 T 的分布函数为

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - P_0(t) = 1 - e^{-\lambda t} \quad (5.16)$$

其概率密度函数：

$$f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t} \quad (5.17)$$

40

5.1.2 有关的概率模型及最简单流

- **无后效性**：顾客到达时刻相互独立，即顾客各自独立地随机到达系统。此假设使顾客数 k 的随机过程具有马尔柯夫性。即在 (t_0, t_0+t) 时间内出现 k 个顾客与 t_0 以前到达的顾客数无关。
- **稀疏性**：在无限小时间间隔 Δt 内，到达两个或两个以上顾客的概率可认为是零，且在有限时间区间内到达的顾客数是有限的。即在充分小的时间区间 Δt 内，发生两个或两个以上事件的概率是比 Δt 高阶的无穷小量，即 $\Delta t \rightarrow 0$ 时，有

$$P_{k \geq 2}(\Delta t) = \sum_{k=2}^{\infty} P_k(\Delta t) = o(\Delta t)$$

$$P_1(\Delta t) = \lambda \Delta t$$

$$P_0(\Delta t) = 1 - \lambda \Delta t$$

38

5.1.2 有关的概率模型及最简单流

顾客到达的平均时间间隔为

$$E(T) = \int_0^{\infty} t f(t) dt = \frac{1}{\lambda} \quad (5.18)$$

- 一个随机过程为“泊松到达过程”或“到达时间间隔为指数分布”实际上是一回事。

(3) 服务时间分布

- **服务时间**：也叫占用时间，指一个顾客接受服务时实际占用一个窗口的时间，也就是服务结束的间隔时间，用 τ 表示。在电话呼叫中，就是一个呼叫的平均持续时间。
- **服务过程**：也就是顾客离去的过程。当一个服务完毕的顾客离开系统时，下一个顾客立即得到服务，服务完毕后离去，二者离去的间隔时间 τ 即为服务时间。

41

5.1.2 有关的概率模型及最简单流

在上述三个条件下，可以推出

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \dots \quad (5.12)$$

在时间 t 内有 k 个顾客到达的均值：(5.13)

$$E(k) = \lambda t \quad (5.14)$$

$$\lambda = \frac{E}{t}$$

注意： $P_k(t)$

- ✓ 是在时间 t 内有 k 个顾客到达的概率；
- ✓ 或是一个排队系统中在时间 t 内有 k 个顾客在等待或正在处理的概率；
- ✓ 或是总的 C 条信道中有 k 条信道被占用概率 $(C \geq k)$ 。

39

5.1.2 有关的概率模型及最简单流

若顾客的离去过程也满足最最简单流条件，则离去过程（即服务过程）也为泊松过程，即有

$$P_k(t) = \frac{(\mu t)^k}{k!} e^{-\mu t} \quad (5.19)$$

离去时间间隔分布（服务时间间隔分布）为指数分布，即服务时间间隔的概率密度函数为：

$$f_{\tau}(t) = \mu e^{-\mu t} \quad (5.20)$$

完成服务的平均时间：

$$E(\tau) = \int_0^{\infty} t f(t) dt = \frac{1}{\mu} \quad (5.21)$$

42

5.1.2 有关的概率模型及最简单流

例5.1 设电话呼叫按30次/h的泊松过程进行，求：5分钟间隔内，(1) 不呼叫的概率；(2) 呼叫3次的概率。

解：按题意， $\lambda=30$ 次/h=0.5次/min， $t=5$ min， $k=0$ 及 $k=3$

(1) 5分钟不呼叫的概率为

$$P_0(5) = e^{-0.5 \times 5} = 0.082 = 8.2\%$$

(2) 5分钟内呼叫3次的概率为

$$P_3(5) = \frac{2.5^3}{3!} e^{-2.5} = 0.214 = 21.4\%$$

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

- 一般来说，大量的稀有事件流，如果每一事件流在总事件流中起的作用很小，而且相互独立，则总的合成流可以认为是最简单流。
- 大量研究表明，将电话呼叫当做最简单流处理，得到的分析结果是正确的。

43

5.1.3 排队系统的主要性能指标

1. 排队长度 k

排队长度：简称队长，是某观察时刻系统内滞留的顾客数，包括正在被服务的顾客。 k 是**非负离散型随机变量**。

- 通常用来描述队长 k 的指标有两个：

- k 的概率分布 P_k ：通常采用系统稳定状态下与时间无关的 P_k 。
- k 的统计平均值 L_s 和平均等待队长 L_q 。
- 知道了队长分布，就可以确定队长超过某个数量的概率，从而能为**设计排队空间**的大小提供依据。

46

5.1 排队论基础

5.1.1 基本概念

5.1.2 有关的概率模型及最简单流

5.1.3 排队系统的主要性能指标

44

5.1.3 排队系统的主要性能指标

- ✓ **平均队长 L_s** ： k 的统计平均值。
- ✓ **平均等待队长 L_q** ：系统内排队等待的平均顾客数。
- ✓ \bar{r} ：正在接受服务的平均顾客数（或平均占用窗口数）。有下式成立：

$$L_s = L_q + \bar{r} \quad (5.26)$$

$$\text{其中, } L_s = \sum_{k=0}^{\infty} k P_k \quad (\text{非拒绝系统}) \quad (5.27)$$

$$\text{或 } L_s = \sum_{k=0}^{\infty} k P_k \quad (\text{拒绝系统}) \quad (5.28)$$

$$L_q = \sum_{k=m+1}^{\infty} (k-m) P_k \quad (\text{非拒绝系统}) \quad (5.29)$$

$$\text{或 } L_q = \sum_{k=m+1}^{\infty} (k-m) P_k \quad (\text{拒绝系统}) \quad (5.30)$$

47

5.1.3 排队系统的主要性能指标

最优化问题一般涉及两种类型：

- 排队系统的最优设计（静态优化）：例如，固话网中的中继电路群数目的确定，分组交换网中的存储空间容量的配置，等等。
- 排队系统的最优控制（动态优化）：例如，固话网中的中继电路群数目的增加与否、无线信道中的信道分配策略等。

排队系统的性能指标描述了排队的概率规律性。通过计算一些性能指标，研究排队系统的最优化问题。

45

5.1.3 排队系统的主要性能指标

2. 等待时间和系统逗留时间

等待时间

- 从顾客到达排队系统的时刻算起，到它开始接受服务的时刻为止的这段时间。
- **平均等待时间 W_q** ：等待时间的统计平均值。

系统逗留时间

- 从顾客到达系统时刻算起，到它接受服务完毕离开系统时刻为止的这段时间。
- **平均系统逗留时间（或系统时间） W_s** ：系统逗留时间的统计平均值。

46

5.1.3 排队系统的主要性能指标

服务时间 τ

- 是一个顾客被服务的时间，即顾客从开始被服务起到离开系统的时间间隔。
- 平均服务时间 $\bar{\tau}$ ： τ 的统计平均值。

$$\bar{\tau} = \frac{1}{\mu} \quad (5.31)$$

$$W_s = W_q + \bar{\tau} \quad (5.32)$$

$$W_s = W_q + \frac{1}{\mu} \quad (5.33)$$

一个有效到达率为 λ_e 的排队系统，在平均的意义上，有

$$\lambda_e W_s = L_s \quad (5.34) \quad \text{Little公式}$$

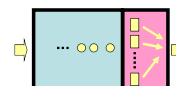
$$\lambda_e W_q = L_q \quad (5.35) \quad \text{适用于任何排队系统。}$$

49

Little公式的普遍性

- Little公式成立的条件只有一个，那就是排队系统要达到统计平衡状态，除此之外，它适用于任何排队系统。
- 它关心的只是排队系统的三个统计平均量，对顾客到达的间隔时间和服务时间的分布以及排队规则不作任何要求
- 但值得注意的是，Little公式中的三个统计平均量必须是针对同一顾客群而言。

$$\begin{aligned} - L_s &= \lambda_s W_s \\ - L_q &= \lambda_q W_q \\ - L_{\text{服}} &= \lambda_{\text{服}} W_{\text{服}} \end{aligned}$$



00:07:32

附：Little公式

- 考虑一个能够达到平稳的排队系统， λ 为到达率， W 为每个顾客在系统中耗费的平均时间， L 为系统中的平均顾客数，则有 $L = \lambda W$
- 证明：
假设在一段比较长的时间区间 $(0, t)$ 内，系统一直处于统计平衡状态， L ， W 都存在， λ 表示单位时间进入到系统中的顾客数
- 全部顾客耗费在系统中的时间
= 到达的顾客数 \times 平均等待时间
= $\lambda t \times W$
= 系统中的平均顾客数 $\times t$
= $L \times t$

$$\lim_{t \rightarrow \infty} \frac{\lambda t \cdot W}{L \cdot t} = 1 \quad \therefore L = \lambda W$$

00:07:32 50

Little公式的物理意义

- L 是一个时间（time average）平均的概念，是不同时刻队列长度在很长一段时间内的平均

$$L = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t l(x) dx$$

- W 是顾客平均的概念，是许许多多不同顾客等待时间的平均

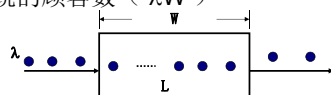
$$W = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n w_i$$

- 一般来讲 W 比较容易从统计中获得， L 比较容易从理论分析中获得

53 00:07:32

Little公式的直观理解

- 在统计平衡状态下，某一顾客离开排队系统时，回头看到的队列长度的平均值（ L ）应该等于顾客在排队等待过程中平均进入排队系统的顾客数（ λW ）



排队系统

00:07:32 51

Little公式应用

- 一个没有等待位置的餐厅，平均每小时服务完30人，平均就餐时间为30分钟，请问平均就餐人数？

$$\lambda = 30$$

$$W = 0.5$$

$$L = \lambda W = 0.5 \times 30 = 15 \text{ (人)}$$

00:07:32

5.1.3 排队系统的主要性能指标

3. 系统效率 η

- 设某时刻有 r 个窗口被占用, 若共有 m 个窗口, 则 r/m 就是窗口占用率。它的统计平均值为平均窗口占用率, 就是系统效率, 即

$$\eta = \frac{\bar{r}}{m} \quad (5.36)$$

4. 空闲概率 P_0 和拒绝概率 P_n

- P_0 : 为系统内无顾客的情况, 即系统空闲状态概率。通过, 可知系统的忙闲情况。
- 拒绝系统 P_n (或 P_c): 为系统内顾客已满、拒绝新到顾客进入系统的状态概率, 也称为阻塞概率 (或损失概率)。

55

5.2 M/M/m/(n)排队系统

5.2.1 M/M/1排队系统

5.2.2 M/M/m/(n)排队系统

58

总结: 排队论已知条件与所求目标

- 顾客到达间隔时间 $A(x)$
- 服务时间 $B(x)$
- 排队模型



- 平均系统队长 L_s
- 平均等待队列长度 L_q
- 平均服务队列长度 $L_{服}$
- 平均系统时间 W_s
- 平均队列时间 W_q
- 平均服务时间 $t_{服}$
- $A, Q, P_{损}$

00:07:58

M/M...排队模型综述

00:09:32

第5章 排队论及其应用

5.1 排队论基础

5.2 M/M/m/(n)排队系统

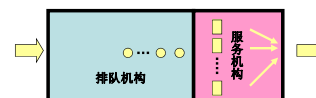
5.3 通信业务量分析

5.4 随机接入系统业务分析

57

排队模型回顾

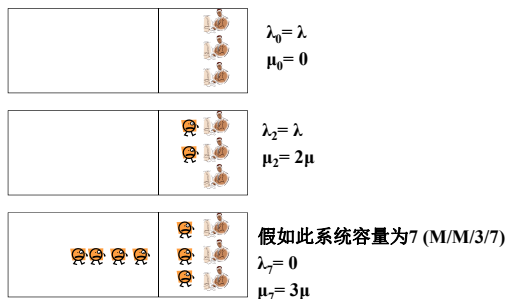
- 顾客到达排队系统请求服务



- 如果有空闲的服务窗, 则直接到服务机构接受服务
- 如果服务窗全部被占用, 则排队等候
- 如果排队系统中顾客数没有满, 则进入排队系统

00:07:32

M/M/..排队系统的几种可能状态



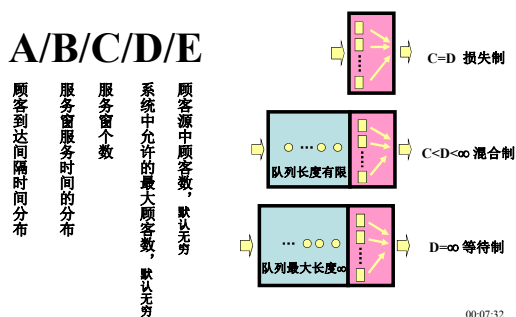
00:07:32
61

M/M/...的排队模型

- M/M/...的排队系统，系统中顾客数变化是一种生灭过程
 - 0状态代表系统有0个顾客
 - 1状态代表系统中有1个顾客
 - 2状态代表系统中有2个顾客
 - ...
- 生灭过程的增长率和消亡率怎么确定？
 - 增长率取决于到达率 λ 和当前系统状态
 - 消亡率取决于服务率 μ 和当前系统状态

生灭过程是用来描述输入过程为最简单流、服务时间为指数分布的这一类最简单的排队模型，即M/M/m (n, N)过程。

排队模型—Kendall记号



00:07:32
62

增长率和消亡率的分析

- 假定顾客到达为强度为 λ 的泊松流，服务窗的服务率为 μ ，服务时间服从负指数分布。在 Δt （极短）时间内，
 - 若顾客到达间隔时间为负指数分布，则在 Δt （极短）时间内有1个顾客到达的概率为 $\lambda\Delta t + 0(\Delta t)$ ，没有顾客到达的概率为 $1 - \lambda\Delta t + 0(\Delta t)$
 - 若服务时间服从参数为 μ 的负指数分布，则在 Δt （极短）时间内，1个正在忙的服务窗服务完当前顾客的概率是 $\mu\Delta t + 0(\Delta t)$ ，没有服务完的概率是 $1 - \mu\Delta t + 0(\Delta t)$

00:07:32
63

M/M/...的排队模型

- 考虑整个排队系统中顾客数的变化
 - 有顾客到达，系统中顾客数加1
 - 有顾客服务完毕，系统中顾客数减1
 - 总之，顾客的到达和离开致使系统顾客数有变化
- M/M/...的排队系统顾客数变化有什么特点？
 - 顾客到达间隔与顾客服务时间均服从负指数分布
 - 因为顾客到达间隔时间是相互独立的，顾客接受服务也是相互独立的，因此，之前的顾客到达情况、服务情况不影响当前顾客数变化概率
 - 因为到达间隔时间和服务时间都具有无记忆性，因此，下一个顾客的到达间隔时间已经过去了多久、当前正在服务的顾客的服务时间已经过去了多久不影响当前顾客数的变化概率

00:07:32
65

增长率和消亡率的分析

- i状态下，i状态代表排队系统中有i个顾客，假定此时有j个顾客正在接受服务($j \leq i$)，当 $i \leq m$ $j=i$ ，当 $i > m$ 时 $j=m$ ，m为服务窗个数

$$p_{i,i+1}(\Delta t) = P(\Delta t \text{内到达了1个，离开了0个}) + P(\Delta t \text{内到达了k个，离开了k-1个, } k \geq 2)$$

$$= \lambda \Delta t e^{-\lambda \Delta t} \cdot (e^{-\mu \Delta t})^i + 0(\Delta t)$$

$$= \lambda \Delta t + 0(\Delta t)$$

$$p_{i,i-1}(\Delta t) = P(\Delta t \text{内到达了0个，离开了1个}) + P(\Delta t \text{内到达了k个，离开了k+1个, } k \geq 1)$$

$$= e^{-\lambda \Delta t} \cdot C_j^i \mu \Delta t e^{-\mu \Delta t} + 0(\Delta t)$$

$$= j \mu \Delta t + 0(\Delta t)$$

00:07:32
66

增长率和消亡率的分析

$$\begin{aligned}
 p_i(\Delta t) &= P(\Delta t \text{内到达了 } i \text{ 个, 离开了 } i \text{ 个}) \\
 &+ P(\Delta t \text{内到达了 } k \text{ 个, 离开了 } k \text{ 个, } k \geq 1) \\
 &= e^{-\lambda \Delta t} \cdot (e^{-\mu \Delta t})^i + o(\Delta t) \\
 &= 1 - (\lambda + j\mu)\Delta t + o(\Delta t)
 \end{aligned}$$

当 $i = 0$ 时

$$\begin{aligned}
 p_{00}(\Delta t) &= P(\Delta t \text{内到达了 } 0 \text{ 个}) \\
 &+ P(\Delta t \text{内到达了 } k \text{ 个, 离开了 } k \text{ 个, } k \geq 1) \\
 &= e^{-\lambda \Delta t} + o(\Delta t) \\
 &= 1 - \lambda \Delta t + o(\Delta t)
 \end{aligned}$$

00:07:32
67

若 S 仅包含有限个元素 $S = \{0, 1, 2, \dots, n\}$, 同时也满足以上条件, 则称为有限状态生灭过程。

• 概率分布

在 t 时刻处于 k 状态的概率分布 P_k ,

即 $t \rightarrow t + \Delta t$, $\{k-1, k, k+1\} \rightarrow k$ 时的 $P_k(t + \Delta t)$ ($\Delta t \rightarrow 0$)

设系统在 $t + \Delta t$ 时刻处于 k 状态, 根据上述生灭过程的定义, 这一事件可分解为如下四个互不相容的事件之和:

> 系统在 $t \rightarrow (t + \Delta t)$, $k \rightarrow k$ 状态, 则在 Δt 内, 其转移概率为

$$P_k(t)[(1 - \lambda_k \Delta t)(1 - \mu_k \Delta t) + \lambda_k \Delta t \mu_k \Delta t]$$

即

$$P_k(t)(1 - \lambda_k \Delta t - \mu_k \Delta t) + o(\Delta t)$$

70

增长率和消亡率的分析

- 由此, M/M/...型排队模型, 在状态 i 时的增长率和消亡率为:

$$\lambda_i = \lim_{\Delta t \rightarrow 0} \frac{P_{i,i+1}(\Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\lambda \Delta t + o(\Delta t)}{\Delta t} = \lambda$$

$$\mu_i = \lim_{\Delta t \rightarrow 0} \frac{P_{i,i-1}(\Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{j\mu \Delta t + o(\Delta t)}{\Delta t} = j\mu \quad j \text{ 是正在忙的服务窗个数}$$

- $j=i, i \leq m$ 系统顾客数少于等于服务窗数时, 所有顾客都在接受服务
- $j=m, i > m$ 系统顾客数大于服务窗个数时, 所有服务窗都在服务, 正在接受服务的顾客数 = 服务窗个数

00:07:32
68

> 系统在 $t \rightarrow (t + \Delta t)$, $(k-1) \rightarrow k$ 状态, 则在 Δt 内, 其转移概率为

$$> P_{k-1}(t) \lambda_{k-1} \Delta t (1 - \mu_{k-1} \Delta t)$$

即

$$P_{k-1}(t) \lambda_{k-1} \Delta t + o(\Delta t)$$

> 系统 $t \rightarrow (t + \Delta t)$, $(k+1) \rightarrow k$ 状态, 则在 Δt 内, 其转移概率为

$$P_{k+1}(t) \mu_{k+1} \Delta t (1 - \lambda_{k+1} \Delta t)$$

即

$$P_{k+1}(t) \mu_{k+1} \Delta t + o(\Delta t)$$

> 系统在 $t \rightarrow (t + \Delta t)$, 其他的状态(即不是 $k-1$ 、 k 、 $k+1$ 状态) $\rightarrow k$ 状态, 其转移概率为 $o(\Delta t)$ 。

71

生灭过程

• 生灭过程定义

设有某个系统, 具有状态集 $S = \{0, 1, 2, \dots\}$, 若系统的状态随时间 t 变化的过程 $\{N(t); t \geq 0\}$ 满足以下条件, 则称为一个生灭过程。

设在时刻 t 系统处于状态 k 的条件下, 再经过长为 Δt ($\Delta t \rightarrow 0$) 的时间, 即当 $t \rightarrow t + \Delta t$ 时, 有

- ✓ 转移到 $k+1$ ($0 \leq k < +\infty$) 状态的转移概率为 $\lambda_k \Delta t + o(\Delta t)$
- ✓ 转移到 $k-1$ ($1 \leq k < +\infty$) 状态的转移概率为 $\mu_k \Delta t + o(\Delta t)$
- ✓ 转移到 $S - \{k-1, k, k+1\}$ 状态的转移概率为 $o(\Delta t)$

其中 $\lambda_k > 0, \mu_k > 0$, 且均为与时间 t 无关的固定常数。

69

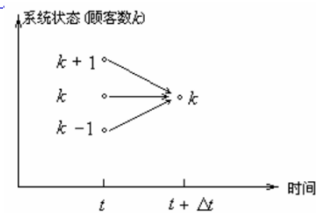


图5.3 Δt 内状态转移示意图 ($\Delta t \rightarrow 0$)

$$\begin{aligned}
 > P_k(t + \Delta t) &= P_k(t)(1 - \lambda_k \Delta t - \mu_k \Delta t) \\
 &+ P_{k-1}(t) \lambda_{k-1} \Delta t + P_{k+1}(t) \mu_{k+1} \Delta t + o(\Delta t)
 \end{aligned}$$

>

72

由全概率公式，得生灭过程的系统稳定状态方程（简称系统方程）：

$$\begin{cases} \lambda_{k-1}P_{k-1} + \mu_{k+1}P_{k+1} = (\lambda_k + \mu_k)P_k & k=1, 2, \dots \\ \lambda_0P_0 = \mu_1P_1 \end{cases} \quad (5.24)$$



图5.4 生灭过程的状态转移图

即有：系统“进入某状态的概率等于离开该状态的概率”

73

第一节 单服务窗损失制排队模型 M/M/1/1

00:47:32

由 $\sum_{k=0}^{\infty} P_k = 1$

得生灭过程在 $t \rightarrow \infty$ 时的稳定状态概率

$$\begin{cases} P_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\mu_k\mu_{k-1}\cdots\mu_1}} \\ P_k = \frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_0}{\mu_k\mu_{k-1}\cdots\mu_1} P_0 & k=1, 2, \dots \end{cases} \quad (5.25)$$

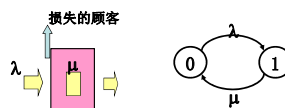
其中， $\lambda_k = \begin{cases} \lambda & \text{无限顾客源, } k \geq 0, \lambda \text{ 为常数} \\ (N-k)\lambda_0 & \text{有限顾客源, } k \geq 0 \end{cases}$

$$\mu_k = \begin{cases} \mu & m=1 \\ k\mu & m>1, \quad k \leq m \\ m\mu & m>1, \quad k \geq m \end{cases}$$

74

排队模型分析

- M/M/1/1
- 顾客到达间隔时间为负指数分布，参数为 λ ， $a(t) = \lambda e^{-\lambda t}$
- 服务窗服务时间为负指数分布，参数为 μ ， $b(t) = \mu e^{-\mu t}$



- 系统最大顾客数1决定了系统状态为{0,1}
 - 状态0 \Leftrightarrow 系统中顾客数为0 \Leftrightarrow 服务窗空闲
 - 状态1 \Leftrightarrow 系统中有1个顾客，此顾客正在接受服务 \Leftrightarrow 系统顾客满 \Leftrightarrow 服务窗忙

00:47:32

5.2.1 单服务窗排队模型

- 第一节 损失制M/M/1/1
- 第二节 等待制M/M/1
- 第三节 混合制M/M/1/m
- 第四节 可变服务率的M/M/1
- 第五节 可变输入率的M/M/1
- 第六节 具有不耐烦顾客的M/M/1
- 第七节 单服务窗封闭式M/M/1/m/m
- 第八节 有差错服务的M/M/1

00:47:32

求解平稳分布

- 根据马氏链、生灭过程列出平衡方程：

$$\begin{cases} \mu p_1 = \lambda p_0 \\ p_0 + p_1 = 1 \end{cases} \Rightarrow p_0 = \frac{\mu}{\lambda + \mu} = \frac{1}{1 + \rho} \quad \text{令 } \rho = \frac{\lambda}{\mu}$$

$$p_1 = \frac{\lambda}{\lambda + \mu} = \frac{\rho}{1 + \rho}$$

本书从现在开始用 $\{p_0, p_1, p_2, \dots\}$ 表示平稳分布

00:47:32

M/M/1/1的各个目标参量

- $P_{\text{损}} = p_1 = \frac{\lambda}{\lambda + \mu}$
- 单位时间内损失的顾客数

$$\bar{\lambda}_L = \lambda p_1 = \lambda \cdot \frac{\lambda}{\lambda + \mu} = \frac{\lambda^2}{\lambda + \mu} = \frac{\lambda \rho}{1 + \rho}$$
- 单位时间内平均进入系统的顾客数

$$\lambda_e = \lambda p_0 = \frac{\lambda}{1 + \rho}$$
- 相对通过能力Q（即单位时间内被服务完的顾客数与请求服务顾客数之比值）

$$Q = p_0 = \frac{1}{1 + \rho}$$
- 绝对通过能力A（单位时间内被服务完顾客的均值）

$$A = \lambda_e = \lambda p_0 = \frac{\lambda}{(1 + \rho)}$$

00:07:32

5.2.2 M/M/m/(n)排队系统

例5.3 有一条电话线，平均每分钟有0.8次呼叫，即 $\lambda = 0.8$ 次/min，如果每次通话时间平均需要1.5分钟，则 $\mu = 1/1.5 = 0.667$ 次/min。求该条电话线每小时能接通多少次电话，又有多少次因呼叫不通而挂断？

解：由题意可知，这是一个M/M/1(1)损失制系统，其 $\lambda = 0.8$ 次/min， $\mu = 0.667$ 次/min。

所以

$$p_0 = \frac{\mu}{\mu + \lambda} = \frac{0.667}{0.667 + 0.8} = 0.455 = 45.5\%$$

82

M/M/1/1例题

- 设某条电话线，平均每分钟有0.6次呼唤，若每次通话时间平均为1.25分钟，求相应的Q，A与 $P_{\text{损}}$

（电话业务我们通常采用M/M/...排队模型）

00:07:32

5.2.2 M/M/m/(n)排队系统

即说明系统处于无顾客状态的概率为45.5%，换句话说，可以接通电话的概率为45.5%。因每分钟平均呼叫0.8次，故每分钟可以接通的次数为

$$0.8 \times 0.455 = 0.364 \text{ (次)}$$

每小时能接通的次数为

$$60 \times 0.364 = 22 \text{ (次)}$$

又电话线损失概率为

$$P_{\text{损}} = p_1 = 1 - p_0 = 1 - 0.455 = 0.545 = 54.5\%$$

故每分钟不能接通的次数为

$$0.8 \times 0.545 = 0.436 \text{ (次)}$$

每小时不能接通的次数为

$$60 \times 0.436 = 26 \text{ (次)}$$

83

M/M/1/1例题

- 设某条电话线，平均每分钟有0.6次呼唤，若每次通话时间平均为1.25分钟，求相应的Q，A与 $P_{\text{损}}$
- 解：按题意知

$$\lambda = 0.6 \quad \mu = \frac{1}{1.25} = 0.8 \quad \text{那么} \quad \rho = \frac{\lambda}{\mu} = \frac{3}{4} = 0.75$$

$$Q = \frac{1}{1 + \rho} = \frac{1}{1 + 0.75} = 0.57 = p_0 \quad \text{即在稳态时有57\%的呼唤得到服务}$$

$$A = \lambda_e = \frac{\lambda}{1 + \rho} = \frac{0.6}{1 + 0.75} = 0.34 \quad \text{即每条电话线路平均每分钟有0.34次通话}$$

$$P_{\text{损}} = 1 - Q = 1 - 0.57 = 0.43 \quad \text{即约43\%的呼唤不能接通}$$

00:07:32

补充：业务负载

- 业务强度（traffic intensity）/业务负载（traffic load）

a. 单位时间内的业务到达量（offered load）

= 单位时间内到达系统的平均呼叫数 × 平均通话时间

$$\text{长度}$$

$$= \lambda \cdot \frac{1}{\mu}$$

b. 单位时间内的业务承载量（carried load）

= 单位时间内得到服务的平均呼叫数 × 平均通话时间

$$\text{长度}$$

$$= \lambda_{\text{e}} \cdot \frac{1}{\mu} = \lambda Q \cdot \frac{1}{\mu} = Q \cdot (\text{offered load})$$

如果通话时间长度的单位为“小时”的话，则话务量单位为“小时·呼”，也叫“爱尔兰（erl）”。

话务量总是针对一段时间而言，如：一天或一小时。

00:07:32

补充：业务负载举例

- 例如：
某电话用户10~12点之间共拨打电话5次，总通话时间为30分钟，求此用户线平均每小时的业务量
- 解：
 $\lambda = 5/(2 \times 60)$ $\mu = 5/30$
承载的业务量为 $a = \lambda/\mu = 30/(2 \times 60) = 0.25\text{erl}$
一个服务窗每小时最多提供 1erl 的业务承载量
- 传统电话网：
 - 普通用户 0.1~0.2erl
 - 集团交换机 0.1~0.6erl

00:07:32

单服务窗排队模型

- 第一节 损失制M/M/1/1
- 第二节 等待制M/M/1
- 第三节 混合制M/M/1/m
- 第四节 可变服务率的M/M/1
- 第五节 可变输入率的M/M/1
- 第六节 具有不耐烦顾客的M/M/1
- 第七节 单服务窗封闭式M/M/1/m/m
- 第八节 有差错服务的M/M/1

00:07:32

补充：关于业务负载的几个典型参数

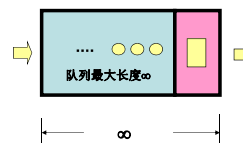
- 分组交换网：
 - 考虑两个路由器之间的一条传输线路，假定每秒钟平均传输10个数据包，数据包平均长度400字节，线路传输速度为64kbps。
- $\lambda = 10$ $\mu = 64,000/400 \times 8$
- 则业务强度为：
 $\rho = 10 \times 400 \times 8 / 64,000 = 0.5 = 50\%$
 如果线路速度为150Mbps，则
 $\rho = 10 \times 400 \times 8 / 150,000,000 = 0.0002 = 0.02\%$

00:07:32

2.1 单服务窗等待制排队模型

M/M/1

- 顾客到达——参数为 λ 的泊松流
- 顾客服务时间——负指数分布，服务率为 μ



00:07:32

服务强度

- 资源利用率（utilization ratio）/服务强度
= 承载业务量/线路数（服务窗个数）

在M/M/1排队模型中，

$$\text{服务强度} = \frac{\lambda_{\text{服}} \cdot \frac{1}{\mu}}{1} = \frac{\rho}{1 + \rho} = \rho_i$$

就是服务窗忙的概率，通信中就是传输线路有数据传输的概率或者通话线路被占用的概率

00:07:32

2.2 M/M/1排队模型分析

- $\lambda_k = \lambda$ $k=0,1,2,3,\dots$
- $\mu_k = \mu$ $k=1,2,3,4,\dots$

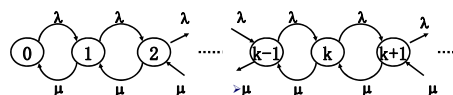


图 M/M/1排队系统的状态转移图

00:07:32

2.3 平稳分布

平衡方程

$$\begin{cases} p_1 = \frac{\lambda}{\mu} p_0 = \rho p_0 \\ p_2 = \frac{\lambda}{\mu} p_1 = \rho^2 p_0 \\ \dots \\ p_k = \frac{\lambda}{\mu} p_{k-1} = \rho^k p_0 \\ \dots \end{cases}$$

正则条件:

$$1 = \sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \rho^k p_0 = p_0 \cdot \frac{1}{1-\rho}$$

$$\Rightarrow p_0 = 1 - \rho$$

$$\Rightarrow p_k = \rho^k (1 - \rho)$$

$$\rho = \frac{\lambda}{\mu} < 1 \text{ 时平稳分布存在}$$

00:07:32
91

5.2.1 M/M/1排队系统

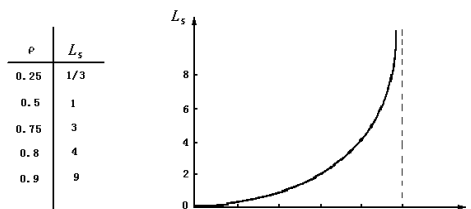


图5.7 M/M/1排队系统中平均队长 L_s 随 ρ 变化的关系

94

5.2.1 M/M/1排队系统

1. 求解 p_0, p_k

$$\begin{cases} \lambda_k = \lambda & k \geq 0 \\ \mu_k = \mu & k \geq 1 \end{cases} \quad (5.37)$$

$$\begin{cases} p_0 = 1 - \rho & \rho < 1 \\ p_k = (1 - \rho) \rho^k & k \geq 1 \end{cases} \quad (5.38)$$

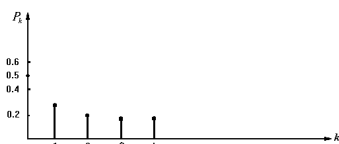


图5.6 M/M/1排队系统的状态概率分布 ($\rho=0.5$)

92

2.4 目标参量

- 3. 系统内排队等候的平均顾客数

$$L_q = L_s - L_{\text{服}} = \sum_{k=1}^{\infty} (k-1) p_k$$

$$L_{\text{服}} = 0 \times p_0 + 1 \times \sum_{k=1}^{\infty} p_k = 1 - p_0 = \rho$$

$$L_q = L_s - L_{\text{服}} = L_s - \rho = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

- 4. 顾客平均排队等候时间

$$W_q = \frac{L_q}{\lambda} = \frac{\rho^2}{\lambda(1-\rho)} = \frac{\lambda}{\mu(\mu-\lambda)}$$

00:07:32
95

2.4 目标参量

- 1. 平均系统队长

$$\begin{aligned} L_s &= \sum_{k=0}^{\infty} k p_k = \rho(1-\rho) \sum_{k=0}^{\infty} k \rho^{k-1} \\ &= \rho(1-\rho) \left(\frac{\rho}{1-\rho} \right)' = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda} \end{aligned}$$

- 2. 顾客在系统内平均逗留时间

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda}$$

00:07:32
96

2.4 目标参量

- 5. 系统内多于k个顾客的概率

$$p = 1 - \sum_{l=0}^k p_l = 1 - (1 - \rho^{k+1}) = \rho^{k+1} \begin{cases} p_0 = 1 - \rho & \rho < 1 \\ p_k = (1 - \rho) \rho^k & k \geq 1 \end{cases}$$

- 6. 记 L_s 为系统内顾客数, 则其方差 $\sigma_s^2 = D L_s$ 为

$$\sigma_s^2 = \sum_{k=0}^{\infty} (l_s - L_s)^2 p_k = \dots = \frac{\rho}{(1-\rho)^2}$$

- 7. 记 L_q 为系统内排队等候的顾客数, 则其方差为

$$\sigma_q^2 = D l_q = E(l_q^2) - L_q^2 = \sum_{k=0}^{\infty} k^2 p_{k+1} - L_q^2 = \dots = \frac{\rho^2(1+\rho-\rho^2)}{(1-\rho)^2}$$

00:07:32
96

5.2.1 M/M/1 排队系统

例5.2 在某数据传输系统中，有一个数据交换节点。信息包按泊松流到达此节点。已知平均每小时到达20个信息包，此节点的处理时间服从指数分布，平均处理每个信息包需要2.5分钟，试求该节点的有关性能指标。

解：根据题意可知，这是一个M/M/1系统，且已知

$$\lambda = \frac{20}{60} = \frac{1}{3} \text{ 个/min}, \mu = \frac{1}{2.5} \text{ 个/min.}$$

$$\text{故有 } \rho = \frac{\lambda}{\mu} = \frac{5}{6}$$

97

例题2.1 M/M/1

某音乐厅设有一个售票处，营业时间为8时到16时，假定顾客流和服务时间均为负指数分布，且顾客到来的平均间隔时间为2.5分钟，窗口为每位顾客服务平均需1.5分钟，试求：

- 顾客不需等待的概率 p_0 ；
- 平均排队长度 L_q ；
- 顾客在系统中平均逗留时间 W_s ；
- 平均排队等待人数 L_q ；
- 平均排队等待时间；
- 系统内顾客人数超过4个的概率 $p=P(L_q > 4)$ ；

5.2.1 M/M/1 排队系统

① L_s ：系统内信息包逗留平均数为

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\frac{1}{3}}{\frac{1}{2.5} - \frac{1}{3}} = 5 \text{ (个)}$$

② L_q ：系统内信息包排队等待平均数为

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{(\frac{1}{3})^2}{\frac{1}{2.5}(\frac{1}{2.5} - \frac{1}{3})} = 4.166 \text{ (个)}$$

③ W_s ：每一信息包在系统内平均逗留时间为

$$W_s = \frac{1}{\mu - \lambda} = \frac{1}{\frac{1}{2.5} - \frac{1}{3}} = 15 \text{ (min)}$$

98

例题2.1 M/M/1

某音乐厅设有一个售票处，营业时间为8时到16时，假定顾客流和服务时间均为负指数分布，且顾客到来的平均间隔时间为2.5分钟，窗口为每位顾客服务平均需1.5分钟，试求：

- 在六天工作日内系统中没有顾客的小时数；
- 若决定当顾客平均逗留时间超过半小时时，就应增加一个售票窗口，试问这相当于要求顾客的平均到达率是原有的几倍？

00:07:32

5.2.1 M/M/1 排队系统

④ W_q ：每一信息包在系统内平均排队等待时间为

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\frac{1}{3}}{\frac{1}{2.5}(\frac{1}{2.5} - \frac{1}{3})} = 12.5 \text{ (min)}$$

⑤ P_0 ：

$$P_0 = 1 - \rho = 1 - \frac{5}{6} = \frac{1}{6} = 16.7\%$$

➤ 结论：M/M/1系统的主要参数均取决于排队强度 ρ 。作为M/M/1系统的稳定参数，必须使 $\rho < 1$ 。

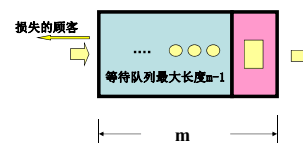
M/M/1系统存在的主要问题是服务质量和系统效率之间的矛盾。

99

3.1 单服务窗混合制排队模型

M/M/1/m

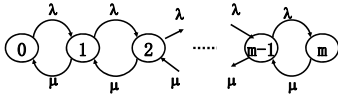
- 顾客到达间隔时间——参数为 λ 的负指数分布
- 服务时间——参数为 μ 的负指数分布
- 排队系统容量： m
- 如果顾客到达系统发现系统满员，则不得不开，是系统损失了的顾客



00:07:32

3.2 M/M/1/m排队模型分析

- $\lambda_k = \lambda \quad k=0,1,2,3,\dots,m-1$
- $\mu_k = \mu \quad k=1,2,3,4,\dots,m$



00:07:32
103

3.4 目标参量 ($\rho \neq 1$)

- 4 服务机构平均顾客数 $L_{\text{服}}$

$$L_{\text{服}} = 1 \times (1 - p_0) + 0 \times p_0 = 1 - \frac{1 - \rho}{1 - \rho^{m+1}} = \frac{\rho - \rho^{m+1}}{1 - \rho^{m+1}}$$

- 5 系统内平均顾客数 L_s

$$\begin{aligned} L_s &= L_q + L_{\text{服}} = \left(\frac{\rho}{1 - \rho} - \frac{m\rho^{m+1} + \rho}{1 - \rho^{m+1}} \right) + \frac{\rho - \rho^{m+1}}{1 - \rho^{m+1}} \\ &= \frac{\rho}{1 - \rho} - \frac{(m+1)\rho^{m+1}}{1 - \rho^{m+1}} \end{aligned}$$

00:07:32
106

3.3 平稳分布

平衡方程

$$\begin{cases} p_1 = \frac{\lambda}{\mu} p_0 = \rho p_0 \\ p_2 = \frac{\lambda}{\mu} p_1 = \rho^2 p_0 \\ \dots \\ p_m = \frac{\lambda}{\mu} p_{m-1} = \rho^m p_0 \end{cases}$$

正则条件:

$$\begin{aligned} 1 &= \sum_{i=0}^m p_i = \sum_{i=0}^m \rho^i p_0 = p_0 \cdot \frac{1 - \rho^{m+1}}{1 - \rho} \\ \Rightarrow p_0 &= \frac{1 - \rho}{1 - \rho^{m+1}} \\ \Rightarrow p_i &= \rho^i \frac{1 - \rho}{1 - \rho^{m+1}} \quad i = 1, 2, \dots, m \\ \text{当 } \rho &= 1 \text{ 时,} \\ p_0 &= \left[\sum_{i=0}^m \rho^i \right]^{-1} = \frac{1}{m+1} \end{aligned}$$

00:07:32
104

3.4 目标参量 ($\rho \neq 1$)

- 6 单位时间内平均损失的顾客数

$$\bar{\lambda}_L = \lambda p_m$$

- 7 单位时间内平均进入系统的顾客数

$$\lambda_e = \lambda(1 - p_m)$$

- 8 平均等待时间

$$\begin{aligned} W_q &= \frac{L_q}{\lambda_e} = \left(\frac{\rho}{1 - \rho} - \frac{m\rho^{m+1} + \rho}{1 - \rho^{m+1}} \right) \cdot \frac{1 - \rho^{m+1}}{\lambda(1 - \rho^m)} = \frac{\rho}{\mu(1 - \rho)} - \frac{m\rho^m}{\mu(1 - \rho^m)} \\ W_s &= \frac{L_s}{\lambda_e} = \left(\frac{\rho}{1 - \rho} - \frac{(m+1)\rho^{m+1}}{1 - \rho^{m+1}} \right) \cdot \frac{1 - \rho^{m+1}}{\lambda(1 - \rho^m)} = \frac{1}{\mu(1 - \rho)} - \frac{m\rho^m}{\mu(1 - \rho^m)} \end{aligned}$$

00:07:32
107

3.4 目标参量 ($\rho \neq 1$)

- 1 $P_{\text{损}}$

$$P_{\text{损}} = p_m = \rho^m \cdot p_0 = \frac{1 - \rho}{1 - \rho^{m+1}} \rho^m$$

- 2 相对通过能力 Q

$$Q = 1 - p_m = 1 - \frac{1 - \rho}{1 - \rho^{m+1}} \rho^m = \frac{1 - \rho^m}{1 - \rho^{m+1}}$$

- 3 等待队列的平均长度

$$L_q = \sum_{k=0}^{m-1} k p_{k+1} = \dots = \frac{\rho}{1 - \rho} - \frac{m\rho^{m+1} + \rho}{1 - \rho^{m+1}}$$

00:07:32
105

3.4 目标参量 ($\rho \neq 1$)

- 9 服务窗平均服务强度

$$p_m = \rho^m \cdot p_0 = \frac{1 - \rho}{1 - \rho^{m+1}} \rho^m$$

$$\bar{\rho} = \frac{\lambda_{\text{服}}}{\mu} = \frac{\lambda_e}{\mu} = \frac{\lambda(1 - p_m)}{\mu} = \frac{\rho - \rho^{m+1}}{1 - \rho^{m+1}} = 1 - p_0 \neq L_{\text{服}}$$

结论:

- 任何一个单服务窗的平均服务强度等于平均服务队长

00:07:32
108

3.4 目标参量 ($\rho=1$)

- 当 $\rho=1$ 时,

$$p_k = \rho^k p_0 = p_0 = \frac{1}{m+1}$$

$$L_q = \sum_{k=0}^{m-1} k p_{k+1} = p_0 \sum_{k=0}^{m-1} k = \frac{m(m-1)}{2(m+1)}$$

$$L_s = \sum_{k=0}^m k p_k = \frac{m}{2}$$

$$W_q = \frac{L_q}{\lambda_e} = \frac{L_q}{\lambda(1-p_m)} = \frac{m-1}{2\lambda}$$

$$W_s = \frac{L_s}{\lambda_e} = \frac{m+1}{2\lambda}$$

00:07:32
109

4 可变服务率的M/M/1排队模型

由k氏方程可知:

$$\text{状态0: } \lambda p_0 = \mu_1 p_1, p_1 = \frac{\lambda}{\mu_1} p_0 = \rho_1^1 p_0$$

$$\text{状态1: } \lambda p_1 = \mu_1 p_2, p_2 = \frac{\lambda}{\mu_1} p_1 = \rho_1^2 p_0$$

$$\text{状态n-1: } \lambda p_{n-1} = \mu_1 p_n, p_n = \frac{\lambda}{\mu_1} p_{n-1} = \rho_1^n p_0$$

$$\text{状态n: } \lambda p_n = \mu_2 p_{n+1}, p_{n+1} = \frac{\lambda}{\mu_2} p_n = \rho_1^n \rho_2 p_0$$

...

$$\text{状态n+r-1: } \lambda p_{n+r-1} = \mu_2 p_{n+r}, p_{n+r} = \rho_1^n \rho_2^r p_0$$

00:07:32

3.5 例题

- 设某自行车修理处只有一个修理工, 修理处内最大容量可以停放7量自行车, 又自行车按平均每小时3辆的速率到达修理处要求修理, 而修理工平均修理一辆自行车需要15分钟, 试求各相应目标参量。

00:07:32
110

4 可变服务率的M/M/1排队模型

$$p_k = \begin{cases} \rho_1^k p_0, & 1 \leq k \leq n \\ \rho_1^n \rho_2^{k-n} p_0, & k > n \end{cases}$$

$$\rho_1, \rho_2 < 1 \text{ 时, 由正则性 } 1 = \sum_{k=0}^{\infty} p_k, p_0 = \left(\frac{1-\rho_1^{n+1}}{1-\rho_1} + \frac{\rho_1^n \rho_2}{1-\rho_2} \right)^{-1}$$

$$L_s = \sum_{k=0}^{\infty} k p_k = p_0 \left[\sum_{k=0}^n k \rho_1^k + \sum_{k=n+1}^{\infty} k \rho_1^n \rho_2^{k-n} \right]$$

$$= p_0 \left\{ \frac{\rho_1 [1 + (n-1) \rho_1^n - n \rho_1^{n-1}]}{(1-\rho_1)^2} + \frac{\rho_1^n (n - (n-1) \rho_2)}{(1-\rho_2)^2} \right\}$$

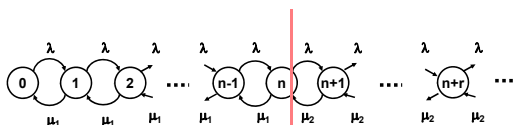
$$L_q = L_s - L_{服} = L_s - (1-p_0)$$

$$W_s = \frac{L_s}{\lambda}, W_q = \frac{L_q}{\lambda}$$

00:07:32
113

4 可变服务率的M/M/1排队模型

- 服务率会因为系统中的顾客数不同而变化
- 举例1 (有2种服务率的情况)
 - 等待制排队系统, 服务率大于到达率时系统才能进入统计平衡状态



顾客数小于等于n时, 采用服务率 μ_1 顾客数大于n时, 采用服务率 μ_2

00:07:32
111

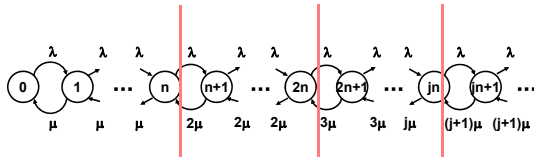
4 可变服务率的M/M/1排队模型

- 例4.1 设某自选商场只有一个收款窗口, 顾客按泊松流来到收款窗口, 平均每分钟到达1人, 收款员视顾客排队长度改变其收款速度, 假定当排队长度超过5人时, 他平均用40秒服务完一位顾客; 反之用平均50秒的速率服务完一位顾客。试求各相应的目标参量。

00:07:32
114

4 可变服务率的M/M/1排队模型

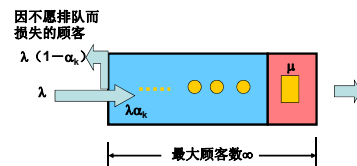
- 举例2（服务率根据系统内顾客数成倍增长的情况）
 - 系统内顾客数为 $jn+1$ 时，服务率发生变化



00:07:32
115

5 可变输入率的M/M/1排队模型

- 顾客到达排队系统，因为不愿进入排队系统而离开，顾客进入系统、离开系统的概率与系统内顾客人数有关。



00:07:32
118

4 可变服务率的M/M/1排队模型

- 平均服务率 $\bar{\mu} = \sum_{k=1}^{\infty} \mu_k p_k$
- 平均服务时间 $\bar{t}_{服} = \frac{1}{\bar{\mu}}$
- 平均服务强度 $\bar{\rho} = \frac{\lambda}{\bar{\mu}} = \lambda \bar{t}_{服} = L_{服} = 1 - p_0$
- L_s $L_s = \sum_{k=0}^{\infty} k p_k$
- $L_q = L_s - L_{服}$ $L_q = L_s - L_{服} = L_s - (1 - p_0)$
- W_s
- W_q

00:07:32
116

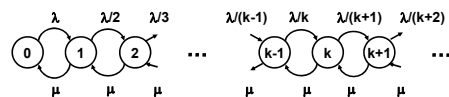
5 可变输入率的M/M/1排队模型

举例:

- 顾客进入系统的概率为 $\alpha_k = \frac{1}{k+1}$
- 实际进入到排队系统的顾客输入率为

$$\lambda_k = \lambda \alpha_k = \frac{\lambda}{k+1}$$

$$p_0 = \left[\sum_{k=0}^{\infty} \frac{\rho^k}{k!} \right]^{-1} = e^{-\rho}; p_k = \frac{\rho^k}{k!} e^{-\rho}$$



00:07:32
119

4 可变服务率的M/M/1排队模型

- 例4.2
设在可变服务率的M/M/1排队模型中取到达率为2，基本服务率为3， $m=2$ ，且服从第二种情况，求 L_k ， L_s ， W_s ， W_q 。

00:07:32
117

5 可变输入率的M/M/1排队模型

- 平均输入率 $\bar{\lambda} = \sum_{k=0}^{\infty} \lambda_k p_k$
- 平均服务强度 $\bar{\rho} = \frac{\bar{\lambda}}{\mu}$
- 损失概率，系统内有 k 个顾客时，损失概率为 $(1-\alpha_k)$

$$P_{损} = \sum_{k=0}^{\infty} (1-\alpha_k) p_k = 1 - \sum_{k=0}^{\infty} \frac{p_k}{1+k} = 1 - \frac{1-e^{-\rho}}{\rho}$$

$$P_{量} = 1 - P_{损} = \frac{1-e^{-\rho}}{\rho}$$

单位时间内平均进入系统队列的顾客数:

$$H = \lambda(1-P_{损}) = \mu(1-e^{-\rho}) = \bar{\lambda}$$

单位时间内平均损失的顾客数: $H_{损} = \lambda P_{损} = \lambda - \bar{\lambda}$

00:07:32
120

5 可变输入率的M/M/1排队模型

- 例5.1

00:07:32
121

6 具有不耐烦顾客的M/M/1排队模型

假定: $\Delta_k = k\delta$

状态0: $\lambda p_0 = \mu p_1, \quad p_1 = \frac{\lambda}{\mu} p_0 = \rho p_0$

状态1: $\lambda p_1 = (\mu + \delta) p_2, \quad p_2 = \frac{\lambda^2}{\mu(\mu + \delta)} p_0 = \frac{\rho^2}{(1 + \beta)} p_0$

$$\Delta_1 = \delta \quad \beta \triangleq \frac{\delta}{\mu}$$

状态2: $\lambda p_2 = (\mu + 2\delta) p_3, \quad p_3 = \frac{\rho^3}{(1 + \beta)(1 + 2\beta)} p_0$

状态k-1: $\lambda p_{k-1} = [\mu + (k-1)\delta] p_k$

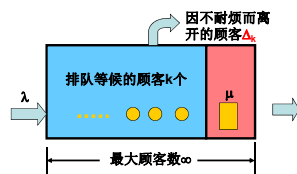
$$p_k = \frac{\rho^k}{(1 + \beta)(1 + 2\beta) \dots [1 + (k-1)\beta]} p_0$$

00:07:32

124

6 具有不耐烦顾客的M/M/1排队模型

- 顾客在排队等候的过程中, 会因为不耐烦而离开排队系统, 使系统顾客数减1, 成为系统损失的顾客



00:07:32
122

6 具有不耐烦顾客的M/M/1排队模型

正则性条件 $\sum_{k=0}^{\infty} p_k = 1$, 容易得到:

$$p_0 = \left\{ 1 + \rho + \frac{\rho^2}{(1 + \beta)} + \frac{\rho^3}{(1 + \beta)(1 + 2\beta)} + \dots + \frac{\rho^k}{(1 + \beta)(1 + 2\beta) \dots [1 + (k-1)\beta]} + \dots \right\}^{-1}$$

由于 $p_k < \frac{\rho^k}{(k-1)!\beta^{k-1}} p_0 \quad (k > 1)$

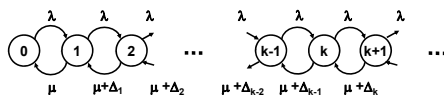
所以, 顾客排队等待的概率 $P_{\text{wait}} = 1 - p_0 = \sum_{k=1}^{\infty} p_k < \sum_{k=1}^{\infty} \frac{\rho^k}{(k-1)!\beta^{k-1}} p_0 = \rho p_0 e^{\frac{\rho}{\beta}}$

00:07:32
125

6 具有不耐烦顾客的M/M/1排队模型

- 假如离开的顾客流是泊松流, 强度与系统内排队等候的顾客数有关 Δ_k , 则系统内顾客数变化是生灭过程

$$\begin{cases} \lambda_i = \lambda \\ \mu_i = \mu + \Delta_{i-1} \end{cases}$$



00:07:32
126

6 具有不耐烦顾客的M/M/1排队模型

平均排队等待队长为: $L_q = \sum_{k=0}^{\infty} k p_{k+1} < \sum_{k=1}^{\infty} \frac{\rho^{k+1}}{(k-1)!\beta^{k-1}} p_0 = \frac{\rho^2}{\beta} p_0 e^{\frac{\rho}{\beta}}$

平均系统队长为: $L_s = \sum_{k=0}^{\infty} k p_k < \sum_{k=1}^{\infty} \frac{k \rho^k}{(k-1)!\beta^{k-1}} p_0 = \rho \left(1 + \frac{\rho}{\beta} \right) p_0 e^{\frac{\rho}{\beta}}$

单位时间内离队而去的不耐烦顾客的均值: $L_i = \sum_{k=0}^{\infty} k \delta p_{k+1} = \delta L_q$

单位时间内被服务完的平均顾客数为:

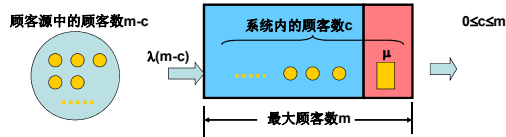
$$A = \lambda - \delta L_q, \quad Q = 1 - \frac{\delta}{\lambda} L_q$$

00:07:32
126

7 单服务窗闭合式排队模型

M/M/1/m/m

- 顾客到达排队系统间隔时间服从负指数分布
- 顾客接受服务的时间服从负指数分布，参数为 μ
- 假定顾客源中单个顾客的到达率为 λ



00:07:32
127

7 单服务窗闭合式排队模型M/M/1/m/m

- 此排队模型虽然系统内顾客最大数有限，但是不会出现顾客损失的情况，所有的顾客都可以进入到排队系统等候、接受服务

- 目标参量

$$A = \mu(1 - p_0)$$

$$\because Q = 1 \therefore \lambda_e = A$$

$$\lambda(m-k) = \mu(1 - p_0)$$

$$k = m - \frac{1 - p_0}{\rho}; L_s = m - \frac{1 - p_0}{\rho}$$

$$L_q = L_s - L_{\text{服}} = m - \frac{1 - p_0}{\rho} - (1 - p_0) = m - (1 - p_0)(1 + \frac{\mu}{\lambda})$$

$$W_s = \frac{L_s}{\lambda_e} = \frac{m}{\mu(1 - p_0)} - \frac{1}{\lambda}$$

$$W_q = \frac{L_q}{\lambda_e} = \frac{m}{\mu(1 - p_0)} - \frac{1}{\lambda} - \frac{1}{\mu}$$

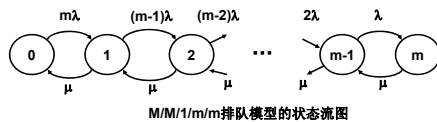
1个工人负责维修m条中继线路
假设k是单位时间内故障线路均值

00:07:32
130

7 单服务窗闭合式排队模型

M/M/1/m/m

- 到达率分析
 - 顾客源中单个顾客的到达率为 λ
 - 当系统中有k个顾客的时候，顾客源中有(m-k)个顾客，到达率为(m-k) λ ，



00:07:32

128

7 单服务窗闭合式排队模型

M/M/1/m/m

- 例题
- 设机器维修工人一人负责看管3台机器，每台机器平均正常工作5天后出现一次故障，维修工人平均每天可以修复半台机器。
 - 试求平均停机台数 L_s ，
 - 平均停机等待修理台数 L_q ，
 - 机器发生故障停机到修复的平均耗时 W_s ，
 - 出故障机器平均等待检修的时间 W_q ，
 - 以及机器运转的效率与维修工人的劳动效率

00:07:32
131

7 单服务窗闭合式排队模型

M/M/1/m/m

- 求平稳分布

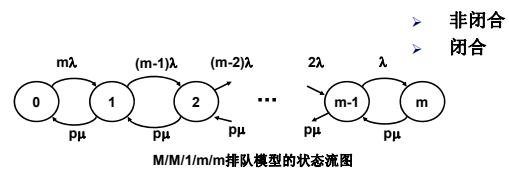
$$\begin{cases} p_k = \frac{\lambda_0 \lambda_1 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_k} p_0 = \frac{m!}{(m-k)!} p_0 \\ \sum_{k=0}^m p_k = 1 \end{cases}$$

- 得:

$$p_0 = \left[\sum_{k=0}^m \frac{m!}{(m-k)!} \rho^k \right]^{-1}$$

00:07:32
132

8 有差错服务的M/M/1排队模型



$$\rho^* = \frac{\lambda}{p\mu}$$

到达率分析

- 顾客源中单个顾客的到达率为 λ ;
- 当系统中有k个顾客的时候，顾客源中有(m-k)个顾客，到达率为(m-k) λ ;
- 排除故障的正确服务率为 $p\mu$ 。

00:07:32
132

8 有差错服务的M/M/1排队模型

- 例8.1

00:07:32
133

5.2.2 M/M/m 排队系统

解决M/M/1系统的**服务质量**与**系统效率**之间的矛盾，必须压缩排队长度、减小等待时间。通常可采用两种措施：

- **增加窗口数**。可提高总服务率但意味着投资加大。
- **截止排队长度**。通过降低系统质量，换取系统效率和稳定性。

136

5.2 M/M/m/(n)排队系统

5.2.1 M/M/1排队系统

5.2.2 M/M/m/(n)排队系统

134

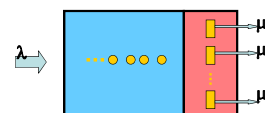
overview

- 1 多服务窗损失制排队模型M/M/n/n
- 2 多服务窗等待制排队模型M/M/n
- 3 多服务窗混合制排队模型M/M/n/m
- 4 窗口能力不等的多服务窗排队模型
- 5. 无限多个服务窗排队模型
- 6. 具有不耐烦顾客的M/M/n排队模型
- 7 多服务窗闭合式排队模型M/M/n/m/m
- 8. 多服务窗损失制排队模型M/M/n/n/m
- 9. 多服务窗有备用品排队模型M/M/n/m+N/m

5.2.2 多服务窗排队模型 M/M/n...

多服务窗排队模型总述

- M/M/n...的排队模型
 - 服务窗个数为多个
 - 假定单个服务窗的服务率为 μ ，则系统在某状态下的消亡率为 $j\mu$ ， j 是此状态下正在忙的服务窗个数

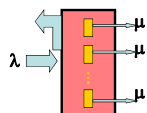


1 多服务窗损失制排队模型

M/M/n/n

- 顾客到达的间隔时间——负指数分布，参数为 λ
- 顾客接受服务的时间——负指数分布，参数为 μ
- 系统有 n 个服务窗
- 系统最多容纳顾客 n 个

因系统满员造成顾客损失



1 多服务窗损失制排队模型

M/M/n/n

- 目标参数

$$-P_{\text{损}} = p_n = \frac{\rho_1^n / n!}{\sum_{k=0}^n \frac{\rho_1^k}{k!}} \rightarrow \frac{1}{p_0}$$

$$B(n, \frac{\lambda}{\mu}) = B(n, \rho_1) = \frac{\rho_1^n / n!}{\sum_{k=0}^n \frac{\rho_1^k}{k!}}$$



— 称为爱尔兰损失公式，又称爱尔兰B公式，欧洲人称为爱尔兰第一公式

1 多服务窗损失制排队模型

M/M/n/n

- $E = \{0, 1, 2, \dots, n\}$
- $\begin{cases} \lambda_k = \lambda & k = 0, 1, 2, \dots, n-1 \\ \mu_k = k\mu & k = 1, 2, 3, \dots, n \end{cases}$

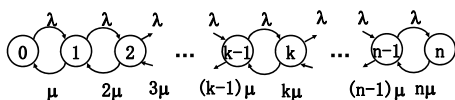


图 M/M/n/n排队模型的状态流图

1 多服务窗损失制排队模型

M/M/n/n

- 爱尔兰B公式的广泛性：
 - 我们有一个具有泊松输入的损失制排队系统称为爱尔兰损失制系统，这种损失制系统对于任何服务时间分布，它在统计平衡条件下的状态概率都相同与M/M/n/n相同。
 - 即M/M/n/n排队系统的平稳分布=M/G/n/n排队系统的平稳分布

1 多服务窗损失制排队模型

M/M/n/n

- 求平稳分布

$$\begin{cases} p_k = \frac{\lambda^k}{k! \mu^k} p_0 = \frac{1}{k!} \rho_1^k p_0 & \rho_1 = \frac{\lambda}{\mu} \\ \sum_{k=0}^n p_k = 1 \end{cases}$$

$$\Rightarrow p_0 = \left[\sum_{k=0}^n \frac{\rho_1^k}{k!} \right]^{-1}$$

$$p_k = \frac{\rho_1^k}{k! \sum_{k=0}^n \frac{\rho_1^k}{k!}}$$

流入话务量强度

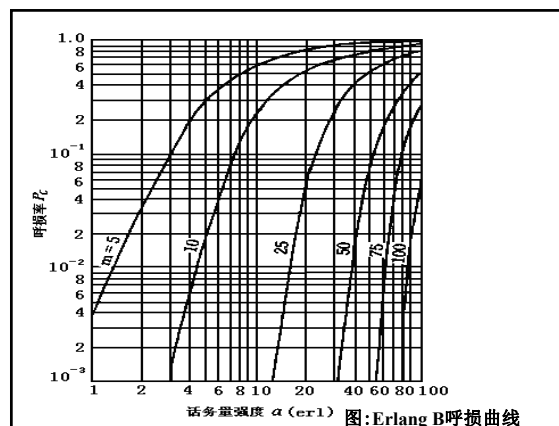


图: Erlang B呼损曲线

1 多服务窗损失制排队模型

M/M/n/n

$$\lambda_e = \lambda(1 - p_n)$$

— 服务窗占用的均值：

$$L_s = L_{\text{服}} = \sum_{k=0}^n k p_k = \dots = \frac{\lambda_e}{\mu}$$

$$L_s = W_{\text{服}} \cdot \lambda_e = \frac{\lambda_e}{\mu}$$

$$L_q = 0$$

— 服务窗的效率（劳动强度）

$$\eta = \frac{L_{\text{服}}}{n} = \frac{\lambda_e}{n\mu}$$

(3) 呼叫损失的概率为

$$P_{\text{损}} = P_5 = \frac{1}{5!} \left(\frac{\lambda}{\mu} \right)^5 P_0 = \frac{1}{5!} 3.75^5 \times 0.029 = 0.179 = 17.9\%$$

148

例5.6 某电话总机系统有5条中继线。设电话呼叫按泊松流发生，平均每分钟呼叫1.5次，且通话时间服从指数分布，平均每次通话时间为2.5分钟，试求：

- (1) 系统空闲的概率。
- (2) 一条线路被占用的概率。
- (3) 呼叫损失的概率。

146

5.3.2 通信业务量分析

例5.7 在一区域内有一无线蜂窝系统，系统中有400个蜂窝小区，每个小区有25个信道，呼叫阻塞概率为2%，每个用户每小时平均拨打1个电话，每个电话平均通话时间为1.5分钟。求该系统所能支持的用户数。

解：已知小区内的信道数 $C=25$ 个，

$$\lambda = 1 \text{ 个} / h = \frac{1}{60} \text{ 个} / \text{min}, 1 / \mu = 1.5 \text{ min}, P_c = 2\% = 0.02。$$

149

5.3.2 通信业务量分析

解：根据题意可知，此系一M/M/5(5)损失制系统，

$$\lambda = 1.5 \text{ 次} / \text{min}, \mu = \frac{1}{2.5} = 0.4 \text{ 次} / \text{min}, \alpha = \frac{1.5}{0.4} = 3.75$$

(1) 系统空闲的概率为

$$P_0 = \left[\sum_{k=0}^5 \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k \right]^{-1} = \left[1 + \frac{3.75}{1!} + \frac{3.75^2}{2!} + \frac{3.75^3}{3!} + \frac{3.75^4}{4!} + \frac{3.75^5}{5!} \right]^{-1}$$

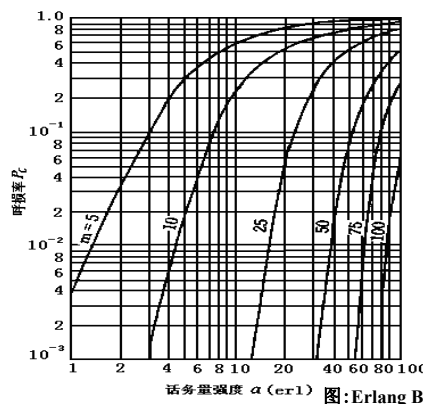
$$= [1 + 3.75 + 7.03 + 8.79 + 8.24 + 6.18]^{-1} = [34.99]^{-1}$$

$$= 0.029 = 2.9\%$$

(2) 系统有一条线路被占用的概率为

$$P_1 = \frac{1}{1!} \left(\frac{\lambda}{\mu} \right) P_0 = 1 \times 3.75 \times 0.029 = 0.109 = 10.9\%$$

147



图：Erlang B呼损曲线

5.3.2 通信业务量分析

则每个用户的话务量强度:

$$a_0 = \lambda / \mu = \frac{1}{60} \times 1.5 = 0.025 \text{ (erl)}$$

查Erlang B曲线图, 可得该系统所承载的总话务量强度为

$$a = 14 \text{ (erl)}$$

因此, 该小区所能支持的用户数为

$$N_0 = a / a_0 = 14 / 0.025 = 560 \text{ (个)}$$

该系统所能支持的用户数:

$$N = 560 \times 400 = 224\,000 \text{ (个)}$$

151

2 多服务窗等待制排队模型

M/M/n

- 求平稳分布

$$\text{令 } \rho_1 = \frac{\lambda}{\mu}, \rho = \frac{\lambda}{n\mu}, \rho < 1$$

$$p_k = \begin{cases} \frac{\rho_1^k}{k!} p_0 = \frac{n^k}{k!} \rho^k \cdot p_0 & 0 \leq k < n \\ \frac{\rho_1^k}{n! n^{k-n}} p_0 = \frac{n^n}{n!} \rho^k \cdot p_0 & k \geq n \end{cases}$$

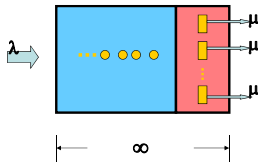
$$\sum_{k=0}^{\infty} p_k = 1$$

$$\Rightarrow p_0 = \left[\sum_{k=0}^{n-1} \frac{\rho_1^k}{k!} + \frac{\rho_1^n}{n!} \cdot \frac{1}{1-\rho} \right]^{-1}$$

2 多服务窗等待制排队模型

M/M/n

- 顾客到达的间隔时间——负指数分布, 参数为 λ
- 顾客接受服务的时间——负指数分布, 参数为 μ
- 系统有 n 个服务窗
- 系统最多容纳顾客 ∞ 个



2 多服务窗等待制排队模型

M/M/n

- 目标参量

$$- P_{\text{损}} = 0$$

$$- A = \lambda$$

$$- Q = I - P_{\text{损}} = I$$

$$- \text{平均等待队长}$$

$$L_q = \sum_{k=n}^{\infty} (k-n) p_k = \dots = \frac{\rho_1^{n+1}}{(n-1)!(n-\rho_1)^2} p_0$$

$$- \text{平均服务队长}$$

$$L_{\text{服}} = \sum_{k=0}^{n-1} k p_k + n \sum_{k=n}^{\infty} p_k = \dots = n \rho = \rho_1 = \frac{\lambda}{\mu}$$

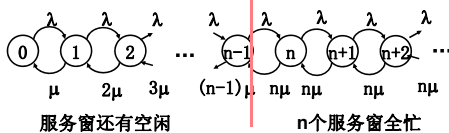
2 多服务窗等待制排队模型

M/M/n

$$E = \{0, 1, 2, \dots\}$$

$$\lambda_k = \lambda$$

$$\mu_k = \begin{cases} k\mu & k < n \\ n\mu & k \geq n \end{cases}$$



2 多服务窗等待制排队模型 M/M/n

$$- \text{平均系统队长}$$

$$L_s = L_q + L_{\text{服}} = \frac{\rho_1^{n+1}}{(n-1)!(n-\rho_1)^2} p_0 + \rho_1$$

$$- \text{平均等待时间}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\rho_1^n p_0}{\mu(n-1)!(n-\rho_1)^2} = \frac{\rho_1^n p_0}{\mu n \cdot n!(1-\rho)^2}$$

$$- \text{平均系统内逗留时间}$$

$$W_s = \frac{L_s}{\lambda}$$

$$W_s = W_q + W_{\text{服}} = W_q + \frac{1}{\mu}$$

2 多服务窗等待制排队模型M/M/n

— 来到系统的顾客必须排队等待的概率

$$C(n, \frac{\lambda}{\mu}) = C(n, \rho_1) = \sum_{k=n}^{\infty} p_k = \sum_{k=n}^{\infty} \frac{n^n}{n!} \rho^k p_0$$

$$= \dots$$

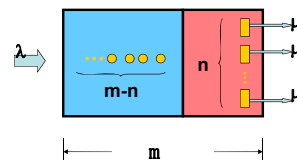
$$= \frac{np_n}{n - \rho_1}$$

— 称为爱尔兰等待公式，又称爱尔兰C公式，欧洲人称为爱尔兰第二公式

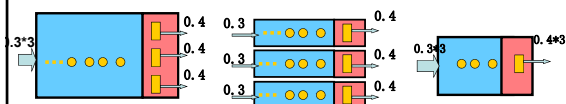
3 多服务窗混合制排队模型

M/M/n/m

- 顾客到达间隔时间——负指数分布，参数为 λ
- 顾客接受服务的时间——负指数分布，参数为 μ



等待制排队模型比较



	M/M/3	3个M/M/1	1个M/M/1
p_0	0.0748	0.25 (每个子系统)	0.25
顾客等待概率	0.57	0.75	0.75
L_q	1.7	6.75 (整个系统)	2.25
L_s	3.95	9.00 (整个系统)	3.00
W_s	4.39	10 (整个系统)	3.33
W_q	1.89	7.5 (整个系统)	2.5

为什么只有一队排队队列呢？

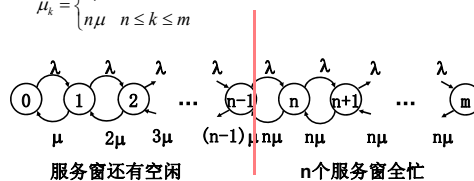
3 多服务窗混合制排队模型

M/M/n/m

$$E = \{0, 1, 2, \dots, m\}$$

$$\lambda_k = \lambda \quad 0 \leq k \leq m-1$$

$$\mu_k = \begin{cases} k\mu & 0 < k < n \\ n\mu & n \leq k \leq m \end{cases}$$



服务窗还有空闲

n个服务窗全忙

• 例2.1

3 多服务窗混合制排队模型

M/M/n/m

- 目标参量
 - $P_{\text{损}} = P_m$
 - $Q = I - P_{\text{损}}$
 - 单位时间内平均损失的顾客数，平均进入系统的顾客数

$$\lambda_L = \lambda P_{\text{损}}$$

$$\lambda_e = \lambda(1 - P_{\text{损}}) = \lambda Q$$

— 平均等待队长

$$L_q = \sum_{k=n}^m (k-n)p_k = \dots = \frac{n^n \rho^{n+1} p_0}{(n)!(1-\rho)^2} [I - (m-n+1)\rho^{m-n} + (m-n)\rho^{m-n+1}] (\rho \neq 1)$$

$$L_q = \frac{n^n}{2n!} (m-n)(m-n+1)p_0 \quad (\rho = 1)$$

— 平均服务队长 $L_{\text{服}} = \sum_{k=0}^{n-1} k p_k + n \sum_{k=n}^m p_k = \dots = n\rho(1 - p_m) = \rho_1(1 - p_m) = \frac{\lambda e}{\mu}$

3 多服务窗混合制排队模型

– 平均系统队长 $M/M/n/m$

$$\begin{cases} L_s = L_q + L_{\text{服}} = L_q + n - \sum_{k=0}^{n-1} (n-k)p_k \\ L_s = L_q + L_{\text{服}} = L_q + \frac{\lambda e}{\mu} \end{cases} \quad \lambda e = \mu \left[n - \sum_{k=0}^{n-1} (n-k)p_k \right]$$

– 平均等待时间

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

– 平均系统内逗留时间

$$W_s = \frac{L_s}{\lambda e}$$

$$W_s = W_q + W_{\text{服}} = W_q + \frac{1}{\mu}$$

> 转到 P.224

4 窗口能力不等的多服务窗排队模型

- 1. 八股方法
 - 状态图 → 方程组
- 2. 正则性求 P_0
- 3. 流量平衡求 P_0
 - 系统输入流平均强度 = 输出流平均强度

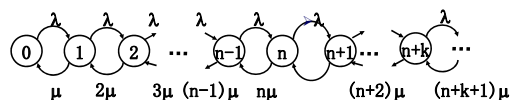
- LS
- Lq
- Ws
- Wq

- 例3.1

5. 无限多个服务窗排队模型

- 当服务窗口足够多，使得顾客随到随服务

$$\begin{cases} \lambda_i = \lambda \\ \mu_i = k\mu \end{cases}$$



4 窗口能力不等的多服务窗排队模型

n 个服务窗口

$$\text{总服务率 } \mu = \sum_{k=1}^n \mu_k$$

选 $n=2$ 考虑

假设顾客选窗口1和窗口2的服务概率分别为 φ_1 和 φ_2

假定 φ_1 为 φ , $\varphi_2 = 1 - \varphi$

$$\varphi = 0, \varphi = \frac{1}{2}, \varphi = 1$$

0状态, 01状态, 10状态, 2, 3, ..., $n+1$ 状态

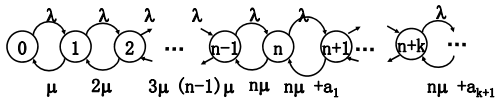
6. 具有不耐烦顾客的 $M/M/n$ 排队模型

- n 个服务窗，容量不限
- 泊松流到达系统
- 不耐烦顾客离去
 - 1) 排队队列长
 - 2) 服务效率低

6 具有不耐烦顾客的M/M/n排队模型

- 假如离开的顾客流泊松流，强度与系统内排队等候的顾客数 k 有关，标记为 a_k ，则系统内顾客数变化是生灭过程

$$\begin{cases} \lambda_i = \lambda \\ \mu_i = i\mu & (0 \leq i \leq n) \\ \mu_i = n\mu + a_{i-n} & (i > n) \end{cases}$$



6 具有不耐烦顾客的M/M/n排队模型

$$\text{平均排队等候队长: } L_q = \sum_{k=0}^{\infty} k p_{n+k} < \sum_{k=0}^{\infty} \frac{k n^n \rho^{k+n}}{n! k! b^k} p_0 = \frac{n^n \rho^{n+1}}{b n!} p_0 e^{\rho/b}$$

$$L_{\text{服}} = \bar{k} = \sum_{k=0}^{n-1} k p_k + \sum_{k=n}^{\infty} n p_k = n + n(\rho - 1) \sum_{k=0}^{n-2} p_k - n p_{n-1}$$

$$L_s = L_{\text{服}} + L_q$$

因此， L_q 中不耐烦顾客的均值 L 为 δL_q

(单位时间内平均离开队列的不耐烦的顾客数)

单位时间内被服务的顾客平均数 $\Lambda = \lambda - \delta L_q$

$$Q = \frac{A}{\lambda} = 1 - \frac{\delta}{\lambda} L_q$$

$$\text{占用服务窗的均值} \bar{k} = \frac{A}{\mu} = \rho_1 - \beta L_q$$

$$L_q = \frac{\rho_1}{\beta} - \frac{\bar{k}}{\beta}$$

6 具有不耐烦顾客的M/M/n排队模型

假定: $\alpha_k = k\delta$ ($\delta > 0$)

$$\text{状态0: } \lambda p_0 = \mu p_1, \quad p_1 = \frac{\lambda}{\mu} p_0 = \rho_1 p_0$$

$$\text{状态1: } \lambda p_1 = (2\mu) p_2, \quad p_2 = \frac{1}{2} \rho_1 p_1 = \frac{\rho_1^2}{2!} p_0$$

$$\alpha_1 = \delta \quad \beta \triangleq \frac{\delta}{\mu}$$

$$\text{状态2: } \lambda p_2 = (\mu + 2\delta) p_3, \quad p_3 = \frac{\rho_1^3}{(1 + \beta)(1 + 2\beta)} p_0$$

$$\text{状态n-1: } \lambda p_{n-1} = (n\mu) p_n, \quad p_n = \frac{\rho_1^n}{n!} p_0$$

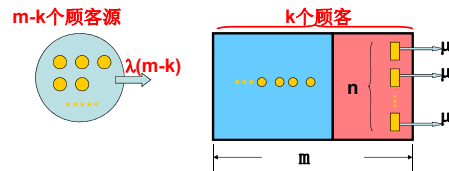
$$\text{状态n: } \lambda p_n = (n\mu + \delta) p_{n+1} \quad \text{状态n+1: } \lambda p_{n+1} = (n\mu + 2\delta) p_{n+2}$$

$$p_{n+1} = \frac{\lambda}{n\mu + \delta} p_n = \frac{\rho_1^{n+1}}{n!(n + \beta)} p_0 \quad p_{n+2} = \frac{\rho_1^{n+2}}{n!(n + \beta)(n + 2\beta)} p_0$$

7 多服务窗闭合式排队模型M/M/n/m/m

- 顾客源有限—— m

- 顾客源 m =系统最大顾客数 m ，∴任何的需求都可以得到满足， $P_{\text{损}}=0$
- 闭合式排队系统：排队系统内顾客与顾客源中顾客总数是固定的



6 具有不耐烦顾客的M/M/n排队模型

$$p_{n+k} = \frac{\rho_1^{n+k}}{n!(n + \beta)(n + 2\beta) \dots (n + k\beta)} p_0$$

正则性条件 $\sum_{k=0}^{\infty} p_k = 1$ ，容易得到：

$$p_0 = \left\{ \sum_{k=0}^{n-1} \frac{\rho_1^k}{k!} + \frac{\rho_1^n}{n!} \left[1 + \frac{\rho_1}{(n + \beta)} + \frac{\rho_1^2}{(n + \beta)(n + 2\beta)} + \dots + \frac{\rho_1^k}{(n + \beta)(n + 2\beta) \dots (n + k\beta)} + \dots \right] \right\}^{-1}$$

令 $\rho = \rho_1/n$, $b = \beta/n$

$$p_k = \begin{cases} \frac{(n\rho)^k}{k!} p_0 & (0 \leq k \leq n) \\ \frac{n^n \rho^k}{n!(k-n)b \dots [1 + (k-n)b]} p_0 & (k > n) \end{cases}$$

当 k 足够大时， $p_k \approx \frac{(nb)^{k-n} \rho^n}{n!(k-n)! b^{k-n}} p_0$

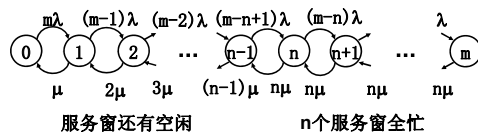
$$\text{因此，顾客排队等待概率 } P_q = \sum_{k=n}^{\infty} p_k \approx \sum_{k=n}^{\infty} \frac{n^n \rho^k}{n!(k-n)! b^{k-n}} p_0 = \frac{(n\rho)^n}{n!} p_0 e^{\rho/b}$$

7 多服务窗闭合式排队模型M/M/n/m/m

$$E = \{0, 1, 2, \dots, m\}$$

$$\lambda_k = (m-k)\lambda \quad 0 \leq k \leq m-1$$

$$\mu_k = \begin{cases} k\mu & 0 < k < n \\ n\mu & n \leq k \leq m \end{cases}$$



7 多服务窗闭合式排队模型M/M/n/m/m

- 求平稳分布

$$p_k = \frac{\lambda_0 \lambda_1 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_k} p_0 \quad (\rho_1 = \frac{\lambda}{\mu})$$

$$= \begin{cases} \frac{m(m-1)\dots(m-k+1)}{1 \cdot 2 \cdot 3 \dots k} \rho_1^k p_0 = C_m^k \rho_1^k p_0 & 0 \leq k \leq n-1 \\ \frac{m(m-1)\dots(m-k+1)}{n! n^{k-n}} \rho_1^k p_0 = \frac{C_m^k k!}{n! n^{k-n}} \rho_1^k p_0 & k \geq n \end{cases}$$

$$\sum_{k=0}^m p_k = 1$$

$$\Rightarrow p_0 = \left[\sum_{k=0}^{n-1} C_m^k \rho_1^k + \sum_{k=n}^m \frac{C_m^k k!}{n! n^{k-n}} \rho_1^k \right]^{-1}$$

7 多服务窗闭合式排队模型M/M/n/m/m

- 内线占用外线的概率

$$1 - p = \frac{L_s}{m}$$

- 内线占用、不占用的循环周期

$$W_{cycle} = \bar{t} + W_s$$

- 内线处于等待状态概率

$$r(m) = \frac{L_q}{m}$$

7 多服务窗闭合式排队模型M/M/n/m/m

- 目标参量

$$L_q = \sum_{k=n}^m (k-n) p_k$$

$$L_s = \sum_{k=0}^m k p_k$$

$$L_{\text{服}} = L_s - L_q$$

$$A = L_{\text{服}} \mu, A = E[(m - l_s) \lambda] = (m - L_q) \lambda = \lambda_e$$

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

$$W_s = \frac{L_s}{\lambda_e}$$

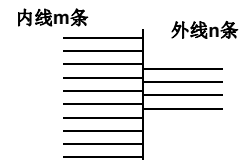
7 多服务窗闭合式排队模型M/M/n/m/m

- 外线利用率

$$\eta = \frac{L_{\text{服}}}{n}$$

- 外线损失系数，（空闲、浪费系数）

$$q(n) = 1 - \frac{L_{\text{服}}}{n} = 1 - \eta$$



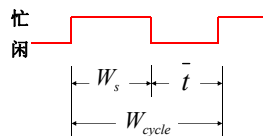
7 多服务窗闭合式排队模型M/M/n/m/m

- 例题（120页）
- 有m条内线 and n条外线，采用BCD（Blocked Call Delayed）排队规则
 - 内线平均空闲时间（不占用外线的平均时间）：

$$\bar{t} = \frac{1}{\lambda}$$

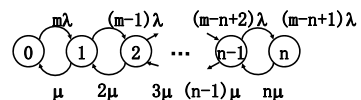
- 内线平均空闲概率

$$p = \frac{\bar{t}}{\bar{t} + W_s} = \frac{m - L_s}{m}$$



8.多服务窗损失制排队模型M/M/n/n/m

- 假设m部机器正常运转，配有n个维修工（m>=n）；
- 当n个维修工全部在忙时，再出故障的机器送到其它单位；或者全体整修
- 每台机器出故障服从泊松分布
- 工人修复故障的机器所用时间服从负指数分布



8. 多服务窗损失制排队模型M/M/n/n/m

- 0状态: m台机器正常运行, 无故障出现
- 1状态: m-1台机器正常运行, 1台故障
- K状态: k个故障, k个工人在忙 ($k \leq n$)

$$P_k = C_m^k \rho_1^k p_0 \quad (0 \leq k \leq n) \quad \rho_1 \triangleq \frac{\lambda}{\mu}$$

利用正则性, 推出: $p_0 = (\sum_{l=0}^n C_m^l \rho_1^l)^{-1}$

$$P_k = P_k(m) = \frac{C_m^k \rho_1^k}{\sum_{l=0}^n C_m^l \rho_1^l} \quad (0 \leq k \leq n) \text{ (engest: 恩格斯特分布)}$$

$$m > n \text{ 时, } P_s(m) = \frac{C_m^n \rho_1^n}{\sum_{l=0}^n C_m^l \rho_1^l}, \text{ 恩格斯特损失公式, 系统损失概率}$$

9. 多服务窗有备用品排队模型M/M/n/m+N/m

- 0状态: m台仪器正常运行, N台备用
- 1状态: m台仪器正常运行, N-1台备用
- N状态: m台仪器运行, 无备用
- N+1状态: m-1台仪器运行, 无备用
-
- N+m-1状态, 一台仪器运行, 无备用

8. 多服务窗损失制排队模型M/M/n/n/n

- m=n时, M/M/n/n/n排队模型
- 此时, 系统没有损失概率

举例: 当m=n时, 便成为M/M/n/n/n排队模型

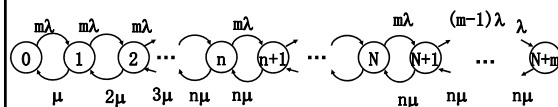
$$P_k(n) = \frac{C_n^k \rho_1^k}{(1 + \rho_1)^n} \quad (0 \leq k \leq n)$$

$$L_s = \sum_{k=0}^n k P_k = \frac{n \rho_1}{(1 + \rho_1)^n} \sum_{l=0}^{n-1} C_{n-1}^l \rho_1^l = \frac{n \rho_1}{(1 + \rho_1)}$$

- 例子8.1

9. 多服务窗有备用品排队模型M/M/n/m+N/m

- 1. $n \leq N$ 时



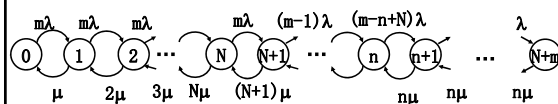
- 状态K, $0 \leq k \leq N$, $m \lambda$
- 状态k+N, $1 \leq k \leq m-1$, $(m-k) \lambda$
- $K < n$: $k \mu$
- $K > n$: $n \mu$

9. 多服务窗有备用品排队模型M/M/n/m+N/m

- m部仪器正常运行, N部备用, 且配有n个维修工
- 发生故障后, 由备用品替换, 维修好的做为备用品
- 若出故障的机器大于N时, 暂时缺额生产
- 仪器出故障服从负指数分布; 修复仪器所需时间服从负指数分布
- 状态空间E (0, 1, 2.....m+N)

9. 多服务窗有备用品排队模型M/M/n/m+N/m

- 2. $n > N$ 时



- 状态k: k个故障仪器
- $k \leq N (< n), m \lambda$
- $k > N, [m - (k - N)] \lambda$

排队网络M/M/n

10.1 多服务窗口串联排队模型

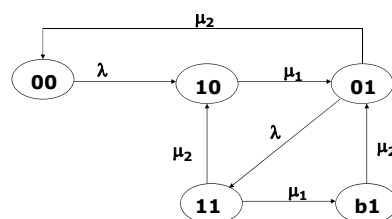
- 当 $n=2$ ，顾客服务的次序是先窗口I，然后再窗口II服务；并且每个服务窗口前不允许排队，并且假定服务窗口I和II的服务强度分别为 μ_1 和 μ_2 ，记 λ 为单位时间到达的顾客平均数；
- 以系统的顾客数为系统的状态：
 - 00状态，系统中无顾客；
 - 10状态，系统中有一个顾客，在窗口I接受服务，窗口II为空闲状态；
 - 01状态，系统中有一个顾客，在窗口II接受服务，窗口I为空闲状态；
 - 11状态，系统中有两个顾客，分别在窗口I和窗口II接受服务；
 - b1状态，系统中有一个在窗口II接受服务，在窗口I的顾客已经服务完成在等待进入窗口II。

内容

- 10、排队网络M/M/n
- 10.1 多服务窗口串联排队模型
- 10.2 成批到达的M^k/M/1排队模型

10.1 多服务窗口串联排队模型

- 状态流图



10.1 多服务窗口串联排队模型

- 顾客在系统中通过一系列服务阶段才能完成服务的模型称为**服务窗口串联排队模型**。
- 典型的系统
 - 医院看病的服务过程；
 - 通信过程的中转过程；
 - 新生到校/离校的注册手续；
- 多服务窗口串联系统的问题一般比较复杂，本节只考虑当窗口数为2时的情况。

10.1 多服务窗口串联排队模型

- 由状态流图，写出平衡条件下的K氏方程：

$$\begin{cases} \lambda p_{00} = \mu_2 p_{01} & (4.11-1) \\ \lambda p_{00} + \mu_2 p_{11} = \mu_1 p_{10} & (4.11-2) \\ \mu_1 p_{10} + \mu_2 p_{01} = (\lambda + \mu_2) p_{01} & (4.11-3) \\ \lambda p_{01} = (\mu_1 + \mu_2) p_{11} & (4.11-4) \\ \mu_1 p_{11} = \mu_2 p_{b1} & (4.11-5) \end{cases}$$

10.1 多服务窗口串联排队模型

• 解析:

- 由 (4.11-1) 得:

$$p_{01} = \frac{\lambda}{\mu_2} p_{00}$$

- 将其代入 (4.11-4) 得:

$$p_{11} = \frac{\lambda^2}{\mu_2(\mu_1 + \mu_2)} p_{00} \quad (4.11-6)$$

- 由式 (4.11-5) 得:

$$p_{b1} = \frac{\mu_1}{\mu_2} p_{11} = \frac{\mu_1 \lambda^2}{\mu_2^2 (\mu_1 + \mu_2)} p_{00} \quad (4.11-7)$$

- 由式 (4.11-2) 得:

$$p_{00} = \left(\frac{\lambda}{\mu_1} + \frac{\lambda^2}{\mu_1(\mu_1 + \mu_2)} \right) p_{00} \quad (4.11-8)$$

10.1 多服务窗口串联排队模型

• 目标参量计算:

- (1) 系统平均顾客数:

$$L_s = 1 \times (p_{01} + p_{10}) + 2 \times (p_{11} + p_{b1}) \quad (4.11-13)$$

$$= \frac{5\lambda^2 + 4\lambda\mu}{3\lambda^2 + 4\lambda\mu + 2\mu^2}$$

- (2)

- 平均忙的服务窗口数量:

$$L_{\text{服}} = 1 \times (p_{01} + p_{10} + p_{b1}) + 2 \times p_{11} \quad (4.11-14)$$

$$= \frac{4\lambda^2 + 4\lambda\mu}{3\lambda^2 + 4\lambda\mu + 2\mu^2}$$

- (3) 系统的损失概率:

$$p_{\text{损}} = p_{10} + p_{b1} + p_{11} \quad (4.11-15)$$

$$= \frac{3\lambda^2 + 2\lambda\mu}{3\lambda^2 + 4\lambda\mu + 2\mu^2}$$

10.1 多服务窗口串联排队模型

• 根据正则性条件:

$$p_{00} + p_{01} + p_{10} + p_{11} + p_{b1} = 1$$

• 可得:

$$p_{00} = \left[1 + \frac{\lambda}{\mu_2} + \frac{\lambda}{\mu_2} + \frac{\lambda^2}{\mu_1(\mu_1 + \mu_2)} + \frac{\lambda^2}{\mu_2(\mu_1 + \mu_2)} + \frac{\mu_1 \lambda}{\mu_2^2(\mu_1 + \mu_2)} \right]^{-1}$$

$$= \frac{\mu_1 \mu_2^2 (\mu_1 + \mu_2)}{\mu_2 (\mu_1 + \mu_2) (\mu_1 \mu_2 + \lambda \mu_1 + \lambda \mu_2) + \lambda^2 (\mu_1^2 + \mu_2^2 + \mu_1 \mu_2)}$$

10.1 多服务窗口串联排队模型

- 当: $\rho_1 = \frac{\lambda}{\mu} \rightarrow \infty$

$$\bar{k} = L_{\text{服}} \rightarrow \frac{4}{3}$$

- 最大利用率为: $\frac{4}{3} \times \frac{1}{2} = \frac{2}{3}$

10.1 多服务窗口串联排队模型

• 当令 $\mu_1 = \mu_2 = \mu$ 时, 则有:

$$p_{00} = \frac{2\mu^2}{3\lambda^2 + 4\lambda\mu + 2\mu^2} \quad (4.11-10)$$

$$p_{01} = \frac{\lambda}{\mu} p_{00} \quad p_{10} = \left(\frac{\lambda}{\mu} + \frac{\lambda^2}{2\mu^2} \right) p_{00} \quad (4.11-11)$$

$$p_{11} = \frac{\lambda^2}{2\mu^2} p_{00} \quad p_{b1} = \frac{\lambda^2}{2\mu^2} p_{00} \quad (4.11-12)$$

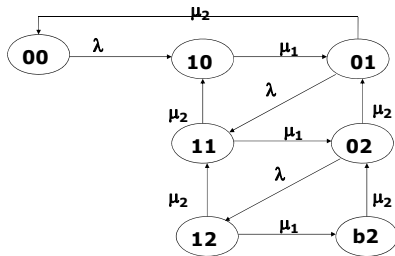
10.1 多服务窗口串联排队模型-

1buf

- 为了减少系统的损失, 假定在窗口II前增设一个排队等候位置, 其余条件不变, 那么系统的可能状态为:
- 00状态, 系统内无顾客;
- 01状态, 系统有一个顾客, 窗口I空闲, 窗口II正在为顾客服务;
- 02状态, 系统有两个顾客, 窗口I空闲, 窗口II正在为顾客服务, 而且有一位顾客排队等候;
- 10状态, 系统有一位顾客, 窗口I在为顾客服务, 窗口II空闲;
- 11状态, 系统中有两位顾客, 窗口I和窗口II均忙, 无排队等待服务的顾客;
- 12状态, 系统中有三位顾客, 窗口I和窗口II均忙, 而且窗口II前面有一个顾客排队等待;
- b2状态, 系统中有3个顾客, 窗口II忙, 而且窗口II前面有一个顾客排队等待, 窗口I的顾客服务完成, 但顾客不能进入窗口II。

10.1 多服务窗口串联排队模型

- 系统的状态流图



10.1 多服务窗口串联排队模型

- 联立以上两式，可以解得：

$$p_{11} = \frac{\lambda^3 + \lambda^2(\mu_1 + \mu_2)}{\lambda\mu_2^2 + \mu_2(\mu_1 + \mu_2)^2} p_{00} \quad (4.11-27)$$

$$p_{12} = \frac{\mu_1\lambda^3}{\mu_2^2[\lambda\mu_2 + (\mu_1 + \mu_2)^2]} p_{00} \quad (4.11-28)$$

- 将其代入式 (4.11-17) -- (4.11-22)，便有：

$$p_{02} = \frac{\mu_1\lambda^2(\mu_1 + \mu_2)}{\mu_2^2[\lambda\mu_2 + (\mu_1 + \mu_2)^2]} p_{00} \quad (4.11-29)$$

$$p_{10} = \frac{\lambda\mu_2\mu_1^2 + 2\lambda\mu_1\mu_2^2 + \lambda^2\mu_1\mu_2 + 2\lambda^2\mu_2^2 + \lambda^3\mu_2 + \lambda\mu_2^3}{\mu_1\mu_2[\lambda\mu_2 + (\mu_1 + \mu_2)^2]} p_{00} \quad (4.11-30)$$

$$p_{b2} = \frac{\mu_1}{\mu_2} p_{12} = \frac{\mu_1\lambda^3}{\mu_2^2[\lambda\mu_2 + (\mu_1 + \mu_2)^2]} p_{00} \quad (4.11-31)$$

10.1 多服务窗口串联排队模型

- 平衡条件下的K氏方程：

$$\lambda p_{00} = \mu_2 p_{01} \quad (4.11-16)$$

$$\mu_1 p_{10} = \lambda p_{00} + \mu_2 p_{11} \quad (4.11-17)$$

$$(\lambda + \mu_2) p_{01} = \mu_1 p_{10} + \mu_2 p_{02} \quad (4.11-18)$$

$$(\mu_1 + \mu_2) p_{11} = \lambda p_{01} + \mu p_{12} \quad (4.11-19)$$

$$(\lambda + \mu_2) p_{02} = \mu_1 p_{11} + \mu_2 p_{b2} \quad (4.11-20)$$

$$(\mu_1 + \mu_2) p_{12} = \lambda p_{02} \quad (4.11-21)$$

$$\mu_2 p_{b2} = \mu_1 p_{12} \quad (4.11-22)$$

10.1 多服务窗口串联排队模型

- 最后由正则性条件：

- 可得： $p_{00} + p_{01} + p_{10} + p_{02} + p_{11} + p_{12} + p_{b2} = 1$

$$p_{00} = \frac{\mu_1\mu_2^2[\lambda\mu_2 + (\mu_1 + \mu_2)^2]}{\mu_1\mu_2^2 + 2\mu_1^2\mu_2^2 + \mu_1^3\mu_2^2 + 4\lambda\mu_1\mu_2^2 + 3\lambda^2\mu_1^2\mu_2^2 + \lambda\mu_1^3\mu_2^2 + \lambda\mu_2^5 + 2\lambda^2\mu_2^4 + 3\lambda^2\mu_1\mu_2^3 + 2\lambda^2\mu_1^2\mu_2^2 + \lambda^2\mu_1^3\mu_2 + \lambda^3\mu_2^3 + \lambda^3\mu_1\mu_2^2 + \lambda^3\mu_1^2\mu_2 + \lambda^3\mu_1^3}$$

- 记 p_{00} 中的分母为 $G(\lambda, \mu_1, \mu_2)$ ，并且记

$$H(\lambda, \mu_1, \mu_2) = \lambda\mu_2 + (\mu_1 + \mu_2)^2$$

10.1 多服务窗口串联排队模型

- 由式 (4.11-16) 得：

$$p_{01} = \frac{\lambda}{\mu_2} p_{00} \quad (4.11-23)$$

- 由式 (4.11-21) 得：

$$p_{02} = \frac{\mu_1 + \mu_2}{\lambda} p_{12} \quad (4.11-24)$$

- 将式 (4.11-23)、(4.11-24) 代入式 (4.11-18)、(4.11-19) 中，并且利用式 (4.11-17) 整理得：

$$\lambda\mu_2 p_{11} + (\mu_1 + \mu_2)\mu_2 p_{12} = \frac{\lambda^3}{\mu_2} p_{00} \quad (4.11-25)$$

$$(\mu_1 + \mu_2)p_{11} - \mu_2 p_{12} = \frac{\lambda^2}{\mu_2} p_{00} \quad (4.11-26)$$

10.1 多服务窗口串联排队模型

$$p_{00} = \frac{\mu_1\mu_2^2 H(\lambda, \mu_1, \mu_2)}{G(\lambda, \mu_1, \mu_2)} \quad (4.11-32)$$

$$p_{01} = \frac{\lambda\mu_1\mu_2^2 H(\lambda, \mu_1, \mu_2)}{G(\lambda, \mu_1, \mu_2)} \quad (4.11-33)$$

$$p_{11} = \frac{\lambda^2\mu_1\mu_2^2(\lambda + \mu_1 + \mu_2)}{G(\lambda, \mu_1, \mu_2)} \quad (4.11-34)$$

$$p_{12} = \frac{\lambda^3\mu_1^2\mu_2}{G(\lambda, \mu_1, \mu_2)} \quad (4.11-35)$$

$$p_{02} = \frac{\lambda^2\mu_1^2\mu_2(\mu_1 + \mu_2)}{G(\lambda, \mu_1, \mu_2)} \quad (4.11-36)$$

$$p_{10} = \frac{\mu_2^2(\lambda\mu_2\mu_1^2 + 2\lambda\mu_1\mu_2^2 + \lambda^2\mu_1\mu_2 + 2\lambda^2\mu_2^2 + \lambda^3\mu_2 + \lambda\mu_2^3)}{G(\lambda, \mu_1, \mu_2)} \quad (4.11-37)$$

$$p_{b2} = \frac{\lambda^3\mu_1^3}{G(\lambda, \mu_1, \mu_2)} \quad (4.11-38)$$

10.1 多服务窗口串联排队模型

令 $\mu_1 = \mu_2 \triangleq \mu$ ，并记 $F(\lambda, \mu) = 4\lambda^3 + 8\lambda^2\mu + 9\lambda\mu^2 + 4\mu^3$

$$p_{00} = \frac{\lambda\mu^2 + 4\mu^3}{F(\lambda, \mu)} \quad (4.11-39)$$

$$p_{01} = \frac{\lambda^2\mu + 4\lambda\mu^2}{F(\lambda, \mu)} \quad (4.11-40)$$

$$p_{02} = \frac{2\lambda^2\mu}{F(\lambda, \mu)} \quad (4.11-41)$$

$$p_{10} = \frac{\lambda^3 + 3\lambda^2\mu + 4\lambda\mu^2}{F(\lambda, \mu)} \quad (4.11-42)$$

$$p_{11} = \frac{\lambda^3 + 2\lambda^2\mu}{F(\lambda, \mu)} \quad (4.11-43)$$

$$p_{12} = p_{22} = \frac{\lambda^3}{F(\lambda, \mu)} \quad (4.11-44)$$

10.1 多服务窗口串联排队模型-

2buf

- 当在窗口I之前也增设一个排队等待的位置（即在窗口I和窗口II之前各有一个等待位置时），其余条件不变，并令：
 $\mu_1 = \mu_2 \triangleq \mu$
- 系统的状态为：00、01、02、10、11、12如同仅在窗口II前有一个等候位置时一样。
- 其他状态有：
 - 20状态，窗口II空闲，窗口I忙，并且有一位顾客在排队等待；
 - 21状态，窗口I和窗口II均忙，并且窗口I前有一个顾客在排队等待；
 - 22状态，窗口I和窗口II均忙，并且两窗口前各有一位顾客排队等待；

目标参量计算：

- 1、系统平均长度：

$$L_s = 1 \times (p_{01} + p_{10}) + 2 \times (p_{02} + p_{11}) + 3 \times (p_{12} + p_{22}) \\ = \frac{9\lambda^3 + 12\lambda^2\mu + 8\lambda\mu^2}{F(\lambda, \mu)} \quad (4.11-45)$$

- 2、平均忙的服务窗口数：

$$\bar{k} = (p_{01} + p_{10} + p_{02} + p_{11} + p_{12} + p_{22}) + 2 \times (p_{11} + p_{12}) \\ = \frac{6\lambda^3 + 10\lambda^2\mu + 8\lambda\mu^2}{F(\lambda, \mu)} \quad (4.11-46)$$

10.1 多服务窗口串联排队模型

- b_02 状态，两个窗口都被占用，且有一位顾客在窗口II前排队等待，一位顾客在窗口I服务完成，等待进入窗口II，窗口I前没有顾客排队等待；
- b_12 状态，两个窗口均忙，两个窗口前各有一个顾客排队等待，而窗口I前有一位顾客已经服务完成，等待进入窗口II服务。

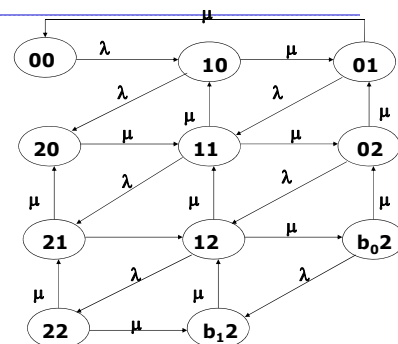
- 3、系统的损失概率：

$$P_{\text{损}} = p_{10} + p_{11} + p_{12} + p_{22} \\ = \frac{4\lambda^3 + 5\lambda^2\mu + 4\lambda\mu^2}{F(\lambda, \mu)} \quad (4.11-47)$$

- 4、窗口I和窗口II忙得概率：

$$p_{I*} = p_{10} + p_{11} + p_{12} = \frac{\bar{k}}{2} \\ p_{**} = p_{01} + p_{02} + p_{11} + p_{12} + p_{22} = \frac{\bar{k}}{2}$$

当 $\rho_1 = \frac{\lambda}{\mu} \rightarrow \infty$ 时， $\bar{k} \rightarrow \frac{3}{2}$ ，系统地最大利用率为 $\frac{3}{2} \times \frac{1}{2} = \frac{3}{4}$
显然，在窗口II前面增设排队等待位置比不设排队等待位置系统效率提高，
 $\frac{3}{4} - \frac{2}{3} = \frac{1}{12}$ 。



10.1 多服务窗口串联排队模型

— 平衡条件下的K氏方程:

$$\begin{aligned}
 \lambda p_{00} &= \mu p_{01} & (4.11-48) \\
 (\mu + \lambda) p_{01} &= \lambda p_{00} + \mu p_{11} & (4.11-49) \\
 (\mu + \lambda) p_{01} &= \mu (p_{10} + p_{02}) & (4.11-50) \\
 \mu p_{20} &= \lambda p_{10} + \mu p_{21} & (4.11-51) \\
 (2\mu + \lambda) p_{11} &= \lambda p_{01} + \mu p_{20} + \mu p_{12} & (4.11-52) \\
 (\mu + \lambda) p_{02} &= \mu p_{11} + \mu p_{02} & (4.11-53) \\
 2\mu p_{21} &= \lambda p_{11} + \mu p_{22} & (4.11-54) \\
 (2\mu + \lambda) p_{12} &= \lambda p_{02} + \mu p_{21} + \mu p_{02} & (4.11-55) \\
 (\mu + \lambda) p_{02} &= \mu p_{12} & (4.11-56) \\
 2\mu p_{22} &= \lambda p_{12} & (4.11-57) \\
 \mu p_{02} &= \mu p_{02} + \mu p_{22} & (4.11-58)
 \end{aligned}$$

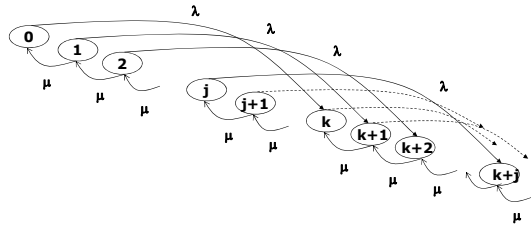
10.2 成批到达的M^k/M/1排队模型

• 有K氏方程得:

$$\begin{aligned}
 p_1 &= \rho p_0 \\
 p_2 &= (\rho + 1) p_1 = \rho(\rho + 1) p_0 \\
 &\dots\dots \\
 p_k &= \rho(\rho + 1)^{k-1} p_0 \\
 p_{k+1} &= (\rho + 1) p_k - \rho p_0 = \rho p_0 [(\rho + 1)^k - 1] \triangleq \alpha \quad (3.9-2) \\
 &\dots\dots \\
 p_{k+l+1} &= (1 + \rho)^l \alpha - l(1 + \rho)^{l-1} \rho^2 p_0 \quad 0 \leq l \leq k-1 \quad (3.9-3) \\
 p_{2k+l+1} &= (1 + \rho)^{k+1} \alpha - (k + l)(1 + \rho)^{k+l-1} \rho^2 p_0 \\
 &\quad + l(1 + \rho)^{l-1} \rho^2 p_0 + \frac{l(l-1)}{2} (1 + \rho)^3 \rho^3 p_0 \quad 0 \leq l \leq k-1 \quad (3.9-4) \\
 &\dots\dots
 \end{aligned}$$

10.2 成批到达的M^k/M/1排队模型

• 系统的状态流图:



10.2 成批到达的M^k/M/1排队模型

$$\begin{aligned}
 p_{3k+1} &= (1 + \rho) p_{2k} \\
 &= (1 + \rho)^{2k} \alpha - 2k(1 + \rho)^{2k-1} \rho^2 p_0 + k\rho^2(1 + \rho)^{k-1} p_0 \\
 &\quad + \frac{k(k+1)}{2} (1 + \rho)^{k-2} \rho^3 p_0 \\
 p_{3k+2} &= (1 + \rho) p_{3k+1} - \rho p_{2k+1} \\
 &= (1 + \rho)^{2k+1} \alpha - (2k + 1)(1 + \rho)^{2k} \rho^2 p_0 \\
 &\quad + (k + 1)(1 + \rho)^k \rho^2 p_0 + \frac{k(k+1)}{2} (1 + \rho)^{k-1} \rho^3 p_0 \\
 p_{3k+3} &= (1 + \rho)^{2k+2} \alpha - (2k + 2)(1 + \rho)^{2k+1} \rho^2 p_0 \\
 &\quad + (k + 2)(1 + \rho)^{k+1} \rho^2 p_0 + \frac{(k+1)(k+2)}{2} (1 + \rho)^k \rho^3 p_0 \\
 &\quad - \rho^3 p_0 \\
 p_{3k+4} &= (1 + \rho)^{2k+3} \alpha - (2k + 3)(1 + \rho)^{2k+2} \rho^2 p_0 \\
 &\quad + (k + 3)(1 + \rho)^{k+2} \rho^2 p_0 + \frac{(k+2)(k+3)}{2} (1 + \rho)^{k+1} \rho^3 p_0 \\
 &\quad - 3(1 + \rho) \rho^3 p_0 - \rho^4 p_0 \\
 &\dots\dots
 \end{aligned}$$

10.2 成批到达的M^k/M/1排队模型

• 根据状态流图, 写出平衡条件下K氏方程:

$$\begin{aligned}
 \lambda p_0 &= \mu p_1 \\
 (\lambda + \mu) p_j &= \mu p_{j+1} \quad 0 \leq j < k \\
 (\lambda + \mu) p_k &= \mu p_{k+1} + \lambda p_0 \\
 (\lambda + \mu) p_{k+l} &= \mu p_{k+l+1} + \lambda p_l \quad 1 \leq l \leq k-1
 \end{aligned}$$

10.2 成批到达的M^k/M/1排队模型

$$\begin{aligned}
 p_{3k+l+1} &= (1 + \rho)^{2k+l} \alpha - (2k + l)(1 + \rho)^{2k+l-1} \rho^2 p_0 \\
 &\quad + (k + l)(1 + \rho)^{k+l-1} \rho^2 p_0 + \frac{(k+l)(k+l-1)}{2} (1 + \rho)^{k+l-2} \rho^3 p_0 \\
 &\quad - \frac{l(l-1)}{2} (1 + \rho)^{l-2} \rho^3 p_0 \\
 &\quad - \frac{l(l-1)(l-2)}{6} (1 + \rho)^{l-3} \rho^4 p_0 \quad 0 \leq l \leq k-1
 \end{aligned}$$

10.2 成批到达的M^k/M/1排队模型

- 若 $0 \leq l \leq k-1$ ，则有：

$$\begin{aligned}
 p_{4k+l+1} = & (1+\rho)^{3k+l} p_0 [(1+\rho)^4 - 1] \\
 & - (3k+l)(1+\rho)^{3k+l-1} \rho^2 p_0 \\
 & + (2k+l)(1+\rho)^{3k+l-2} \rho^2 p_0 \\
 & + \frac{(2k+l)(2k+l-1)}{2} (1+\rho)^{3k+l-2} \rho^3 p_0 \\
 & - \frac{(k+l)(k+l-1)}{2} (1+\rho)^{3k+l-2} \rho^3 p_0 \\
 & - \frac{(k+l)(k+l-1)(k+l-2)}{6} (1+\rho)^{3k+l-3} \rho^4 p_0 \\
 & - \frac{l(l-1)(l-2)}{6} (1+\rho)^{l-3} \rho^4 p_0 \\
 & + \frac{1}{6} l(l-1)(l-2)(l-3)(l-4) \rho^5 p_0
 \end{aligned} \quad (3.9-6)$$

10.2 成批到达的M^k/M/1排队模型

- 若 k 是一个随机变量，求解如下：

– 记 W 为整批顾客逗留时间的总和， l_s 为批量到达系统时的系统长度，那么有：

$$W_s = \frac{EW}{Ek} \quad (3.9-11)$$

– 先求 EW ，为此先假定 l_s 和 k 是已知的；那么在 l_s 和 k 的条件下，整批顾客平均滞留时间总和为： $E(W|l_s, k)$ 由两个部分组成：一部分为当队长为 l_s 时，批量为 k 个顾客的 **全部等待时间的总和** kl_s/μ ，另一部分为服务窗口对批量为 k 的顾客提供服务的 **平均时间的总和**。

10.2 成批到达的M^k/M/1排队模型

- 利用正则性条件，计算 $\sum_{i=0}^{\infty} p_i$ = 求解 p_0 非常复杂，由定义去解析 L_s 、 W_s 的目标参量就更加复杂。
- 下面使用 **较为简单的方法** 求解 L_s 、 W_s
- 假定一个批量为 k 的顾客同时到达系统，系统内已经有了 n 个顾客，所以就需要等待这 n 个顾客服务完成后，服务窗口才能够开始为这批顾客服务。
- 服务窗口为对该批量顾客服务过程中，每位顾客所花费的时间为：

$$\begin{aligned}
 \sum_{i=1}^k \frac{1}{\mu} P_i & \quad (\text{该顾客在 } k \text{ 个当中排在第 } i \text{ 位}) \\
 = \sum_{i=1}^k \frac{i}{\mu} \times \frac{1}{k} & = \frac{k+1}{2\mu} \quad (3.9-7)
 \end{aligned}$$

10.2 成批到达的M^k/M/1排队模型

- 于是有：

$$E(W|l_s, k) = \frac{kl_s}{\mu} + \frac{k(k+1)}{2\mu} \quad (3.9-12)$$

- 在将 k 看作一个随机变量，则有：

$$\begin{aligned}
 E(W|l_s) & = E(E(W|l_s, k)) \\
 & = \frac{l_s Ek}{\mu} + \frac{Ek^2 + Ek}{2\mu} \quad (3.9-13)
 \end{aligned}$$

- 最后将 l_s 的条件由已知改为随机的，则有：

$$\begin{aligned}
 EW & = E(E(W|l_s, k)) \\
 & = \frac{l_s Ek}{\mu} + \frac{Ek^2 + Ek}{2\mu} \quad (3.9-14)
 \end{aligned}$$

10.2 成批到达的M^k/M/1排队模型

- 顾客在系统的滞留时间：

$$\begin{aligned}
 W_s & = \sum_{n=0}^{\infty} E(\text{滞留时间} | \text{顾客到达系统时队长为 } n) p_n \\
 & = \sum_{n=0}^{\infty} \left(\frac{n}{\mu} + \frac{k+1}{2\mu} \right) p_n = \frac{1}{\mu} \sum_{n=0}^{\infty} n p_n + \frac{k+1}{2\mu} \sum_{n=0}^{\infty} p_n \\
 & = \frac{L_s}{\mu} + \frac{k+1}{2\mu} \\
 & = \frac{k\lambda W_s}{\mu} + \frac{k+1}{2\mu} \quad (3.9-8)
 \end{aligned}$$

- 由此解得：

$$W_s = \frac{k+1}{2(\mu - k\lambda)} \quad \mu > k\lambda \quad L_s = \frac{k(k+1)\lambda}{2(\mu - k\lambda)} \quad \mu > k\lambda$$

10.2 成批到达的M^k/M/1排队模型

- 由式 (3.9-11)、(3.9-13)，并且利用 Little 公式： $L_s = (Ek)\lambda W_s$ ，可以解得：

$$W_s = \frac{Ek^2 + Ek}{2(\mu - \lambda Ek)Ek} \quad (3.9-15)$$

5.1 排队论基础

5.2 M/M/m/(n)排队系统

5.3 通信业务量分析

5.4 随机接入系统业务分析

223

5.3.1 通信业务量基本理论

1. 呼叫的发生过程

- ✓ 纯随机呼叫的情况
- ✓ 准随机呼叫的情况
- ✓ 呼叫合成发生的情况

(1) 纯随机呼叫的情况

满足以下条件的呼叫，称为纯随机呼叫：

- 呼叫源**无限多**，即能够发生呼叫的用户数很大；
- 处于占线状态（占用信道）的呼叫源数目相对少可不考虑；
- 用户（呼叫）之间相互独立；
- 呼叫的发生和交换网（或信道）的阻塞状态可分别考虑。

若同时满足**最简单流条件（平稳性、无后效性、普通性）**，即可表示为M/M/m(n)排队系统模型。

226

5.3 通信业务量分析

5.3.1 通信业务量基本理论

5.3.2 通信业务量分析

224

5.3.1 通信业务量基本理论

(2) 准随机呼叫的情况

满足以下条件的呼叫，称为准随机呼叫：

- 呼叫源有限；
- 用户之间仍相互独立。

若同时满足**最简单流条件**，即可表示为M/M/m(n, N)排队系统模型。

实际通信网中的顾客（用户）数总是有限的，所以不存在严格的纯随机呼叫，而多属于**准随机呼叫**。当N很大时(N >> k)，或用户数非常多时，准随机呼叫可近似当做纯随机呼叫处理。N越大，这种近似越合理。

227

5.3.1 通信业务量基本理论

设计和建设一个通信网（及所配置的设备）是以全网**业务量**为主要依据的。

呼叫

- 进入通信网送到通信设备和线路上进行传输的语音、数据等**输入信息**，统称为通信呼叫，简称呼叫。

通信业务量（流量）

- 在网中传送的信息量，称为通信业务量，也称为流量。

与排队论中的对应关系

- 呼叫→顾客
- 呼叫长度（呼叫持续时间）→服务时间。网中的呼叫源即是网内的所有用户。
- 信道数C或线路容量→窗口数m
- 不同类型的呼叫事件→不同的典型的排队模型

225

5.3.1 通信业务量基本理论

(3) 呼叫合成发生的情况

设有两个相互独立的呼叫源，各自按呼叫发生率 λ_1 、 λ_2 呈泊松分布，其呼叫发生概率分别为

$$P_k^1(t) = \frac{(\lambda_1 t)^k}{k!} e^{-\lambda_1 t}$$

$$P_k^2(t) = \frac{(\lambda_2 t)^k}{k!} e^{-\lambda_2 t}$$

则合成呼叫发生数为k的概率为

$$P_k(t) = \frac{[(\lambda_1 + \lambda_2)t]^k}{k!} e^{-(\lambda_1 + \lambda_2)t} \quad (5.59)$$

228

5.3.1 通信业务量基本理论

结论：两个分别按 λ_1 、 λ_2 的泊松分布的合成等于呼叫发生率为 $(\lambda_1 + \lambda_2)$ 的泊松分布。

推广：

若有 m 个各自任意速率为 $\lambda_1, \lambda_2, \dots, \lambda_m$ 的独立泊松流，则复合流本身也为泊松过程，其速率参数为

$$\lambda = \sum_{i=1}^m \lambda_i \quad (5.60)$$

2. 业务量和呼叫量

(1) 流量的特性

通信网中的流量受到各种因素的影响，信息流量处在经常的变化之中。以话务量变化为例说明。

229

5.3.1 通信业务量基本理论

Q 具有两方面的含义：

- 反映了信息源所发生的**用户需求业务量**，这时， Q 可能大于 mT ；
- 反映了通过 m 条信道的**实际的通信业务量**，此时， Q 一定不大于 mT 。

② 呼叫量

定义：业务量的强度通常称为呼叫量。它可定义为线路（或信道）可能占用的时间与观察时间之比，即呼叫量为

$$a = \frac{\text{业务量}}{\text{观察时间}} = \frac{Q}{T} (\text{erl}) \quad (5.62)$$

a 是没有量纲的，通常使用“小时呼”或“爱尔兰（erl）”表示它的单位。通常取 T 为一小时，

232

5.3.1 通信业务量基本理论

特点：

- 话务量的**波动性**：话务量随时间不断地变化。
- 话务量波动的**随机性**：每一次观测的结果都不同。
- 话务量波动的**周期性**：每一次观测的结果都遵从某种相同的规律。

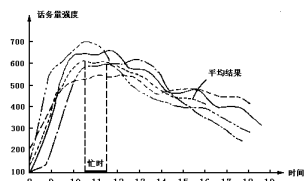


图 5.10 话务量的波动曲线

230

5.3.1 通信业务量基本理论

- ✓ 一个erl表示一小时一个完全被占用的信道的呼叫量（即单位小时或单位分钟的呼叫时长）。
- ✓ 根据定义，呼叫量也可表示为

$$a = \lambda \bar{\tau} \quad (5.63)$$

呼叫量的含义：

- 反映了信息源产生的**用户需求呼叫量**，它可以大于 m ；
- 反映了**实际通过 m 条信道的呼叫量**，它一定不大于 m 。当用户需求呼叫量大于 m 时，多余的呼叫将被拒绝。
- ✓ 作为通信网设计的原始数据，一般采用**用户需求呼叫量**。在测定呼叫量时，应认为信道数 m 足够大。

233

5.3.1 通信业务量基本理论

(2) 业务量和呼叫量的定义

① 业务量

定义：

是在指定观察时间内各个线路（或信道）可能被占用的时间之和**即占用的总时间**。这些时间可以是重叠的或不重叠的。

- 若某线路有 m 条信道，第 i 条信道被占用 Q_i 秒，则 m 条信道或该线路上的业务量 Q 为

$$Q = \sum_{i=1}^m Q_i \quad (5.61)$$

业务量的**量纲是时间**。若一个信道代表一个电话话路，则业务量的单位是秒·话路。

231

5.3.1 通信业务量基本理论

用户需求呼叫量分为两种：

- ✓ **日呼叫量**：一天中最忙的一小时内的呼叫量称为日呼叫量，也就是一天中最大的小时呼叫量。
- ✓ **年呼叫量**：一年内取三十天，取这些天的呼叫量的平均值称为年呼叫量，亦称基准呼叫量。

网设计依据的话务量：

- ✓ 日呼叫量：用于小网
- ✓ 年呼叫量：用于大网

234

5.3.1 通信业务量基本理论

(3) 话务量的定义与计算

- **话务量**：电话网中的业务量称为话务量。
用来反映电话用户的通话频繁程度和通话时间的长短。用下式表示：

$$Y = \lambda S T \quad (5.64)$$

- λ ：单位时间内的呼叫次数，即呼叫强度（次/h）；对应排队论中的系统到达率 λ 。
- S ：一次呼叫的平均占用时长（h/次）；对应平均服务时间 $\bar{\tau}$ 。
- T ：计算话务量的时间范围（h）。

235

5.3.1 通信业务量基本理论

例5.4 设有呼叫强度 $\lambda = 1800$ 次/h，平均占用时长

$$S = \frac{1}{30} \text{ h/次}，求话源话务量强度。$$

解：根据式(5.65)，有

$$\begin{aligned} a &= \lambda S \\ &= 1800 \text{ 次/h} \times \frac{1}{30} \text{ h/次} \\ &= 60 \text{ (erl)} \end{aligned}$$

238

5.3.1 通信业务量基本理论

话务量含义：

- 反映了占用设备的程度；
- 同时也反映了用户对电话网设备的需求。

话务量强度：

单位时间内的话务量（即电话网中的呼叫量），即

$$a = \lambda S \rightarrow \rho_1 \quad (5.65)$$

单位：“小时呼”或“erl”。

通常所说的话务量一般指**话务量强度**。

忙时话务量：指系统在24小时中最繁忙的一个小时内的平均话务量。工程设计中使用。

236

5.3.1 通信业务量基本理论

例5.5 假设在60条线的中继线群上，平均占用次数为1200次/h，平均占用时长为

$$\frac{1}{30} \text{ h/次}，求该中继线群的完成话务量强度。$$

解：已知： $\lambda = 1200$ 次/h， $S = \frac{1}{30}$ h/次

由式(5.66)得

$$\begin{aligned} a &= \lambda_e S \\ &= 1200 \text{ 次/h} \times \frac{1}{30} \text{ h/次} \\ &= 40 \text{ (erl)} \end{aligned}$$

239

5.3.1 通信业务量基本理论

话务量强度分为：

- ✓ 流入话务量强度：反映了用户对通话设备的**需求**程度。
- ✓ 完成话务量强度：反映了用户**占用**通话设备的程度。
- **流入话务量强度**
等于在平均占用时长内话源发生的平均呼叫次数。
又称作话源话务量强度。 $a = \lambda S$

此处，话源发生的平均呼叫次数就是系统到达率。

➤ 完成话务量强度

一组设备的完成话务量强度等于该组设备在平均占用时长内发生的平均占用次数。

这里平均占用次数就是系统的有效到达率，计算公式为 $a = \lambda_e S$

237

5.3.1 通信业务量基本理论

(4) 通过量 \bar{a}_c 和信道利用率 η

通过量：在单位时间内通过网络的呼叫量，即

$$\bar{a}_c = a(1 - P_c) \quad (\text{erl}) \quad (5.67)$$

$$\text{或} \quad \bar{a}_c' = \lambda(1 - P_c) \quad (\text{次/s}) \quad (5.68)$$

\bar{a}_c \rightarrow 完成话务量，或平均占用窗口数（参考5.56）

\bar{a}_c' \rightarrow λ_e

信道数 C 或线路容量相当于窗口数 m ，则信道利用率为

$$\eta = \frac{\bar{a}_c}{C} = \frac{a(1 - P_c)}{m} \quad (5.69)$$

240

5.3.1 通信业务量基本理论

(5) 时延

时延：指消息进入网内后直到被利用完毕所需的时间，包括

- ✓ 等待时间
- ✓ 服务时间
- ✓ 处理时间
- ✓ 传输时延
- 传输时延一般是较小的；
- 处理时间与消息内容有关，一般可从技术上缩短，所占的份额不一定太大，而且往往是恒定的。
- 时延的主要部分是系统时间，即等待时间和服务时间。

241

5.3.1 通信业务量基本理论

阻塞率：

是指拒绝状态占全部状态的百分比。阻塞率可有两种定义：时间阻塞率和呼叫阻塞率。

✓ 时间阻塞率

$$P_B = \frac{\text{阻塞时间}}{\text{总观察时间}} \quad (5.71)$$

P_B 就是本章5.1节中介绍的截止队长为 n 时的拒绝概率 P_n

✓ 呼叫阻塞率

$$P_c = \frac{\text{被拒绝的呼叫次数}}{\text{总呼叫次数}} \quad (5.72)$$

P_c ：通常称为呼损，即为Erlang B呼损公式。

244

5.3.1 通信业务量基本理论

3. 服务等级及服务系统

业务量理论：

利用Erlang B公式或C公式，即业务量、中继线（或信道）数量和阻塞概率（或呼叫等待概率）之间的关系式，在一定的服务等级上，在已知业务量预测值的条件下，确定中继电路数、长途电路数，或求移动网中核心网的电路数、无线网的信道配置等。

目的：使固定数量的中继线路或信道可为一个数量更大的、随机的用户群体服务。

242

5.3.1 通信业务量基本理论

一般来说，由于阻塞期间内可能没有顾客到达，则

$$P_c \leq P_n$$

在纯随机呼叫情况下，相当于顾客以泊松流到达，则

$$P_c = P_n$$

证明过程：见P146页（利用呼叫阻塞率的定义）

(2) 服务系统

按处理阻塞呼叫（未接续的呼叫）的方式不同，通信网中通常用到两种服务系统：

- 阻塞呼叫清除系统
- 阻塞呼叫延迟系统

245

5.3.1 通信业务量基本理论

(1) 服务等级 (GoS: Grade of Service)

服务等级 (GoS)：表示阻塞的量。

定义为

- ✓ 呼叫阻塞概率（也称呼叫阻塞率）。
- ✓ 或呼叫延迟时间大于某一特定排队时间的概率。

在实际的通信网中，多为截止型的排队系统。当系统处于拒绝状态时，系统是阻塞的，将出现呼损。

243

5.3.1 通信业务量基本理论

(2) 服务系统

按处理阻塞呼叫（未接续的呼叫）的方式不同，通信网中通常用到两种服务系统：

- 阻塞呼叫清除系统
- 阻塞呼叫延迟系统

• 阻塞呼叫清除系统

该系统不对阻塞呼叫请求进行排队，即放弃阻塞呼叫的接续。这种系统又叫做阻塞系统或损失制系统、立接制系统。

- 系统模型：M/M/m (m) 或 M/M/m (m, N)
- 基本的阻塞呼叫清除系统：M/M/m (m)

246

5.3.1 通信业务量基本理论

假设前提条件为：

- ✓ 用户数量为无限大，呼叫服从泊松分布。
- ✓ 呼叫请求的到达无记忆性，意味着所有的用户，包括阻塞的用户，都可以在任何时刻要求分配一个信道。
- ✓ 用户占用信道时间服从指数分布，那么根据指数分布，长时间的通话发生的可能性就很小。
- ✓ 可用的信道数目有限。此处用 C 表示信道数。通常，5个或以上的信道数才可认为是足够大量的信道。

该类系统的GoS: P_c

$$P_c = \frac{\frac{a^C}{C!}}{\sum_{k=0}^C \frac{a^k}{k!}}$$

247

5.3 通信业务量分析

5.3.1 通信业务量基本理论

5.3.2 通信业务量分析

250

5.3.1 通信业务量基本理论

- 当用户为有限时，阻塞呼叫清除系统的模型为 $M/M/m(m, N)$ ，此时通常会产生更小的阻塞概率。
- 阻塞呼叫延迟系统
 - ✓ 其前提条件同上。
 - ✓ 不同的只是该服务系统用一个队列缓冲器来保存阻塞呼叫，以等待接续。
 - ✓ 如果不能立即获得一个信道，呼叫请求就一直延迟一定时间后，直到有空闲信道为止或被拒绝接续。也可称为等待制系统、缓接制系统。
- 系统模型: $M/M/m(n)$ 、 $M/M/m$
- 基本的阻塞呼叫延迟系统: $M/M/m(n)$

248

5.3.2 通信业务量分析

1. 阻塞呼叫清除系统的业务分析

(1) 足够多用户的情况

这是一种纯随机呼叫发生的情况。

系统的排队模型: $M/M/m(m)$ ，是基本的阻塞呼叫清除系统。

由爱尔兰B公式→该系统的GoS:

$$P_c = \frac{a^m / m!}{\sum_{k=0}^m a^k / k!}$$

251

5.3.1 通信业务量基本理论

该系统的GoS: 呼叫在队列中需等待 t 秒以上的概率:

$$P_W[\text{延迟} > t] = P_W[\text{延迟} > 0] P_W[\text{延迟} > t | \text{延迟} > 0] \\ = P_W[\text{延迟} > 0] \exp(-(C-a)t / \bar{\tau}) \quad (5.76)$$

$\bar{\tau}$: 为呼叫的平均持续时间。

C : 信道数

到达系统的呼叫没有立即获得信道的概率(呼叫等待概率):

$$P_W[\text{延迟} > 0] = \frac{a^C}{a^C + C! \left(1 - \frac{a}{C}\right) \sum_{k=0}^{C-1} \frac{a^k}{k!}} \quad (5.75)$$

Erlang C公式

249

5.3.2 通信业务量分析

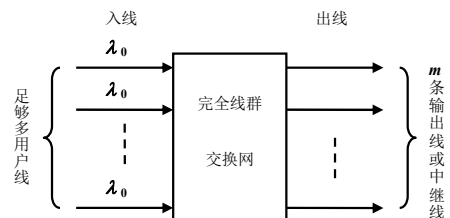


图5.11 固话网中的基本的阻塞呼叫清除系统 $M/M/m(m)$ 模型

252

5.3.2 通信业务量分析

(2) 有限用户的情况

这是准随机呼叫发生的情况。

系统模型: $M/M/m(m, N)$

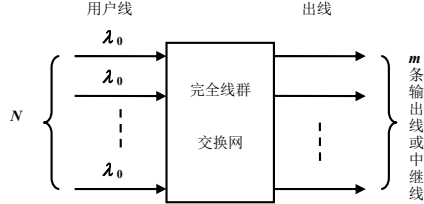


图5.12 固话网中的阻塞呼叫清除系统 $M/M/m(m, N)$ 模型

5.3.2 通信业务量分析

$$P_c = \frac{\binom{N-1}{m} \left(\frac{\lambda_0}{\mu}\right)^m}{\sum_{k=0}^m \binom{N-1}{k} \left(\frac{\lambda_0}{\mu}\right)^k} \quad (5.82)$$

$$\eta = \sum_{k=0}^m \frac{k}{m} P_k = \frac{N}{m} \frac{\lambda_0}{\mu} \frac{\sum_{k=0}^{m-1} \binom{N-1}{k} \left(\frac{\lambda_0}{\mu}\right)^k}{\sum_{k=0}^m \binom{N-1}{k} \left(\frac{\lambda_0}{\mu}\right)^k} \quad (5.83)$$

当 $N \rightarrow \infty$ 时, 得

$$P_c = \lim_{N \rightarrow \infty} \frac{\binom{N-1}{m} \left(\frac{\lambda_0}{\mu}\right)^m}{\sum_{k=0}^m \binom{N-1}{k} \left(\frac{\lambda_0}{\mu}\right)^k} = \frac{a^m / m!}{\sum_{k=0}^m a^k / k!} \quad \eta = \lim_{N \rightarrow \infty} \frac{N \lambda_0}{m \mu} \frac{\sum_{k=0}^{m-1} \binom{N-1}{k} \left(\frac{\lambda_0}{\mu}\right)^k}{\sum_{k=0}^m \binom{N-1}{k} \left(\frac{\lambda_0}{\mu}\right)^k} = \frac{a(1-P_m)}{m}$$

256

5.3.2 通信业务量分析

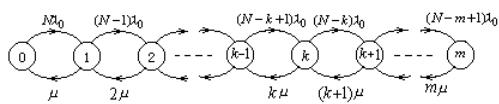


图5.13 电话交换即时拒绝系统 $M/M/m(m, N)$ 的状态转移图

$$\lambda_k = \begin{cases} (N-k)\lambda_0 & 0 \leq k < m-1 \\ 0 & k \geq m \end{cases} \quad (5.77)$$

$$\mu_k = \begin{cases} k\mu & 0 < k \leq m \\ 0 & k > m \end{cases} \quad (5.78)$$

254

5.3.2 通信业务量分析

结论:

当 $N \rightarrow \infty$ 时, 准随机呼叫变为纯随机呼叫, 所得结果与 $M/M/m(m)$ 排队系统的结果相同。

在实际交换系统中, 当 $N \gg m$ 时, 即可将用户的呼叫看做是纯随机呼叫过程。

257

5.3.2 通信业务量分析

$$P_0 = \left[\sum_{k=0}^m \binom{N}{k} \left(\frac{\lambda_0}{\mu}\right)^k \right]^{-1} \quad (5.79)$$

$$P_k = \binom{N}{k} \left(\frac{\lambda_0}{\mu}\right)^k P_0 \quad (5.80)$$

$$P_n = P_m = \frac{\binom{N}{m} \left(\frac{\lambda_0}{\mu}\right)^m}{\sum_{k=0}^m \binom{N}{k} \left(\frac{\lambda_0}{\mu}\right)^k} \quad (5.81)$$

255

5.3.2 通信业务量分析

例5.8 当电路数 $m=8$ 、服务质量等级 $P_c=0.01$ 时, 利用查表法求该系统最大能承担的话务量。

解: 查表5.1, 当 $m=8$ 、 $P_c=0.01$ 时, 可查得该系统最大能承担的话务量

$$a = 3.128 \text{ (erl)}.$$

258

5.3.2 通信业务量分析

表5.1

Erlang B呼损表

a	P_c	0.005	0.01	0.05	0.1	0.2	0.3
m							
1		0.005	0.010	0.053	0.111	0.250	0.429
2		0.105	0.153	0.381	0.595	1.000	1.499
3		0.349	0.455	0.899	1.271	1.930	2.633
4		0.701	0.869	1.525	2.045	2.945	3.891
5		1.132	1.361	2.218	2.881	4.010	5.189
6		1.622	1.909	2.960	3.758	5.109	6.514
7		2.157	2.501	3.738	4.666	6.230	7.857
8		2.730	3.128	4.543	5.597	7.369	9.213
9		3.333	3.783	5.370	6.546	8.522	10.579
10		3.961	4.461	6.216	7.511	9.685	11.953

259

5.3.2 通信业务量分析

(1) Erlang C公式——呼叫等待概率 P_w [延迟 > 0]

$$P_w [\text{延迟} > 0] = P[m \text{ 个信道忙}] = \sum_{k=m}^{n-1} P_k = \sum_{k=m}^{n-1} \frac{m^m}{m!} \left(\frac{a}{m}\right)^k P_0$$

$$= \frac{m^m}{m!} P_0 \sum_{k=m}^{n-1} \left(\frac{a}{m}\right)^k = \frac{a^m}{m!} \frac{1 - \left(\frac{a}{m}\right)^{n-m}}{1 - \frac{a}{m}} P_0 \quad (5.84)$$

$$P_n = \frac{m^m}{m!} \left(\frac{a}{m}\right)^n P_0 \quad \text{稳定} \quad (5.85)$$

若缓冲器容量无限大，即 $n \rightarrow \infty$ ，且 $a < m$ ，则有

$$P_0 = \left[\sum_{k=0}^{m-1} \frac{a^k}{k!} + \frac{a^m}{m!} \frac{m}{m-a} \right]^{-1} \quad (5.86)$$

262

5.3.2 通信业务量分析

2. 阻塞呼叫延迟系统的业务分析

- 该类系统主要应用于以数据终端为呼叫源的非实时性的数据通信中。GSM网中的切换呼叫处理，亦可采用阻塞呼叫延迟 (M/M/m (n)) 方式。

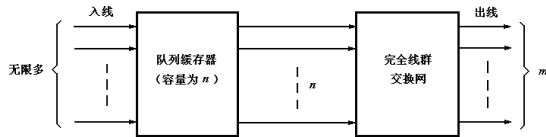


图5.14 基本的阻塞呼叫延迟系统M/M/m (n) 模型

260

5.3.2 通信业务量分析

$$P_k = \begin{cases} \frac{a^k}{k!} P_0 & 0 \leq k < m \\ \frac{m^m}{m!} \left(\frac{a}{m}\right)^k P_0 & k \geq m \end{cases} \quad (5.87)$$

$$P_n = 0$$

$$P_w [\text{延迟} > 0] = \frac{a^m}{a^m + m! \left(1 - \frac{a}{m}\right) \sum_{k=0}^{m-1} \frac{a^k}{k!}} \quad (5.89)$$

Erlang C公式

263

5.3.2 通信业务量分析

任意时刻k个用户同时接续的概率为

$$P_k = \begin{cases} \frac{a^k}{k!} P_0 & 0 \leq k \leq m \\ \frac{m^m}{m!} \left(\frac{a}{m}\right)^k P_0 & m \leq k \leq n \\ 0 & k > n \end{cases}$$

261

5.3.2 通信业务量分析

等待概率 P_w [延迟 > 0] 与所有呼叫的平均等待时间 W_q 的关系：(用 P_w 表示 P_w [延迟 > 0])

$$W_q = \frac{1}{m \mu} \left[\frac{m}{m-a} - \frac{(n-m)a^{n-m}}{m^{n-m} - a^{n-m}} \right] P_w \quad (5.90)$$

当 $n \rightarrow \infty$ 时，则有

$$W_q = \frac{1}{\mu(m-a)} P_w \quad (5.91)$$

排队呼叫的平均延迟D为

$$D = \frac{1}{\mu(m-a)} \quad (5.92)$$

264

5.3.2 通信业务量分析

$$\eta = \frac{a(1-P_c)}{m} = \frac{a(1-P_n)}{m} \xrightarrow{0} 1$$

在 $\eta \rightarrow 1$ 处, $W_q \rightarrow \infty$, $P_W = \frac{a^m}{m!} \frac{m}{m-a} P_0$ 收敛为 1。

(2) 等待延迟时间大于 t 秒的条件概率 $P_W[\text{延迟} > t | \text{延迟} > 0]$ 排队呼叫的平均延迟为 D , 则排队呼叫的平均离去率为 $1/D$ 。

$$P_W[\text{延迟} > t | \text{延迟} > 0] = \frac{(t/D)^k}{k!} \exp(-t/D) \Big|_{k=0} = \exp(-t/D)$$

$$= \exp(-\mu(m-a)t) = \exp(-(m-a)t/\bar{\tau})$$

$$P_W[\text{延迟} > t | \text{延迟} > 0] = \exp(-(m-a)t/\bar{\tau}) \quad (5.93)$$

$\bar{\tau} = 1/\mu$ 为一个呼叫的平均持续时间。

265

- (2) 一个被延迟的呼叫等待 10 秒以上的概率?
- (3) 一个呼叫被延迟 10 秒以上的概率?

268

5.3.2 通信业务量分析

(3) 该系统的 GoS: 呼叫需等待 t 秒以上的概率 $P_W[\text{延迟} > t]$

$$P_W[\text{延迟} > t] = P_W[\text{延迟} > 0] P_W[\text{延迟} > t | \text{延迟} > 0] \quad (5.94)$$

$$P_W[\text{延迟} > t] = P_W[\text{延迟} > 0] P_W[\text{延迟} > t | \text{延迟} > 0] \\ = P_W[\text{延迟} > 0] \exp(-(m-a)t/\bar{\tau}) \quad (5.95)$$

爱尔兰 C 公式

266

5.3.2 通信业务量分析

(2) 已知: $\lambda = 2$ 次/h, 每个用户的话务量强度为 $a_0 = 0.05$ erl

所以, 呼叫的平均持续时间

$$\bar{\tau} = a_0 / \lambda = 0.05 / 2 = 0.025 \text{ (h)} = 90 \text{ (s)}$$

一个被延迟的呼叫等待 10 秒以上的概率为

$$P_W[\text{延迟} > 10 | \text{延迟} > 0] = \exp(-(C-a)t/\bar{\tau}) = \exp(-(20-13)10/90) = 46\%$$

(3) $P_W = 5\% = 0.05$

一个呼叫被延迟 10 秒以上的概率为

$$P_W[\text{延迟} > 10] = P_W[\text{延迟} > 0] P_W[\text{延迟} > 10 | \text{延迟} > 0] = 0.05 \times 46\% = 2.3\%$$

269

5.3.2 通信业务量分析

例 5.9 在一个无线蜂窝系统内, 有 60 个小区, 每个小区有 20 个信道, 每个用户的话务量强度为 0.05 erl, 平均每小时呼叫两次。该系统为呼叫等待概率为 5% 的 Erlang C 系统。求:

(1) 该系统可支持多少用户数?

解: 已知 $C = 20$ 个, 呼叫等待概率 $P_W = 5\% = 0.05$,

每个用户的话务量强度为 $a_0 = 0.05$ erl。

(1) 利用 Erlang C 公式可得: 一个小区内的承载话务量强度 = 13 (erl)

所以, 一个小区内可支持的用户数为: $N_0 = 13 / 0.05 = 260$ (个)

则该系统可支持的总的用户数为: $N = 260 \times 60 = 15600$ (个)

267

• Thanks

270

5.3.2 通信业务量分析

3. 蜂窝移动网中的呼叫处理排队方案

在GSM网络呼叫切换处理中常用的三个典型的越区切换方案：

- 无优先级方案
- 切换呼叫排队方案
- 信道预留方案

(1) 无优先级方案

- 小区中所有的 C 个信道均被新呼叫和越区切换呼叫所共享。
- 基站处理以上两种呼叫的方法完全相同。
- 任意一种呼叫，如在其到达的时刻基站内没有空闲信道，那么到达的呼叫都将被系统阻塞。

系统模型：处理新呼叫和切换呼叫的排队模型为 $M/M/C(C)$ 。

271

5.3.2 通信业务量分析

结论：

- 预留信道方式是降低切换失败概率的一种有效方法，但它影响了信道利用率。
- 切换失败概率的降低都是以新呼叫阻塞概率的升高为代价的。

4. 提高网效率的一些措施

提高信道利用率和降低呼损，可采取四个措施：

- ✓ 大群化效应
- ✓ 延迟效应
- ✓ 综合效应
- ✓ 迂回效应

274

5.3.2 通信业务量分析

(2) 切换呼叫排队方案

- 小区中所有的 C 个信道同样被新呼叫和越区切换呼叫所共享。
- 当上述两种呼叫同时到达，并且小区中的信道全被占用时，将对切换呼叫进行排队，并阻塞新呼叫。
- 如果在最大排队时间内无空闲信道可用就将阻塞切换呼叫。

系统模型：

- ✓ 处理新呼叫的排队模型为 $M/M/C(C)$ ；
- ✓ 处理切换呼叫的排队模型为 $M/M/C(N)$ 。

272

5.3.2 通信业务量分析

(1) 大群化效应

大群化效应：在保证一定通信质量指标的前提下，变分散利用信道为集中利用信道，可增加信道的利用率，传送更多的业务量，从而提高网络效率，这种规律就是大群化效应。即采用集中业务量进行大容量的信道传输，系统效率可得到很大提高，而且节省了网络资源。

特点：

- 业务量越大，这种效应就越明显。
- 信道越集中，故障的影响面也越大，将使网络的可靠性下降。
- 过高的效率，当业务量变化时，会使信道对负荷的适应能力下降。

组建通信网应在兼顾业务量、呼损、时延、可靠性和自适应性等因素的情况下，尽量利用大群化效应。

275

5.3.2 通信业务量分析

(3) 信道预留方案

- 系统专门为越区切换呼叫预留了部分信道。预留信道数为 $C-S_c$ 。
- 剩余信道 S_c 则由新呼叫和越区切换呼叫所共享。
- 当新呼叫到达小区时，如果基站中剩余的闲置信道数小于或等于预留信道数，就阻塞该呼叫请求。
- 当切换呼叫请求到达小区时，如果基站中没有空闲信道，就将其阻塞。

系统模型：

- ✓ 处理新呼叫的排队模型为 $M/M/S_c(S_c)$ ；
- ✓ 处理切换呼叫的排队模型为 $M/M/C(C)$ 。

273

5.3.2 通信业务量分析

(2) 延迟效应

- 延迟效应：采用延迟拒绝系统，由于延时的存在，可提高系统效率和降低呼损。

应用：

- 通信网内的非实时性业务如数据、电邮等。
- 对于实时性业务如电话，也可采取呼叫信令的排队等待方式。这样就可以取得降低呼损的效果。

276

5.3.2 通信业务量分析

(3) 综合效应

综合效应：包括业务的综合和技术的综合。一般指将不同性质的业务综合起来在同一网内或同一条线路上传输，例如把宽带与窄带、或实时的与非实时的业务等综合起来一起传输，以实现大容量信道的大群化效应。

277

5.4 随机接入系统业务分析

5.4.1 概述

5.4.2 基本的随机接入系统业务分析

5.4.3 基于监听的随机接入系统业务分析

5.4.4 基于预约机制的随机接入系统业务分析

5.4.5 应用举例

280

5.3.2 通信业务量分析

(4) 迂回效应

迂回路由的使用：

- ✓ 在首选路由发生故障时被采用；
- ✓ 用来转接首选路由中的溢出业务。

溢出业务：

由于业务流的随机性或其他因素，在某些时间内所要求的业务量可能超出首选路由的信道数，超出部分就是溢出业务。

迂回效应：采用迂回路由方式，来传送超负荷随机业务流的溢出部分，可提高网络的接通率，降低呼损，减小时延。

特点：

- 迂回路由占用的网络资源比直达路由多。
- 当网络内业务流不过载的情况下，采用迂回路由比较有效。
- 在网络业务流过载的情况下，应对自动迂回路由的选择加以一定的控制，而对直达路由的业务流给以优先。

278

5.4.1 概述

1. 多址通信与随机接入技术

多址通信

指网络中多个节点（或多个用户终端）共享公共通信资源（信道）、实现连接访问。多址通信采用多址接入技术实现用户间的通信。

多址接入技术

- 是网络传输技术的一部分。它的实现机制将直接影响到网络的吞吐量、时延、业务能力、用户数、信道利用率等多方面性能。
- 多址接入技术的实现主要是通过数据链路层（MAC子层）的协议来完成的。多址接入协议的选择主要受到三方面的影响：业务要求，网络资源，信道环境。

281

第5章 排队论及其应用

5.1 排队论基础

5.2 M/M/m/(n)排队系统

5.3 通信业务量分析

5.4 随机接入系统业务分析

279

5.4.1 概述

典型的多址通信信道环境包括：

地面无线信道环境、局域网环境和卫星信道环境。

- 卫星通信环境：最重要的影响参数是传输延时，典型值是0.25s，这通常比传输一个分组的时间长度还长。
- 地面无线信道环境：传输延时通常比分组传输时间小，但要考虑信号功率的覆盖范围，而且地面无线信道中存在多径的影响。
- 局域网信道环境：信号传输距离短，要求传输速率高。

282

5.4.1 概述

多址接入方式包括：

- 固定分配信道接入
- 动态分配信道接入
- 在动态分配接入策略的动态FDMA、TDMA、CDMA方式中，**动态TDMA成为动态分配方式的首选。**
- 动态TDMA方式又分为三种分配调度算法：
 - ✓ **基本的随机接入方式：**即基于ALOHA的接入方式，包括P-ALOHA和S-ALOHA两种接入方式；
 - ✓ **基于监听的随机接入方式：**即基于CSMA的接入方式，包括了CSMA、CSMA/CD及CSMA/CA。
 - ✓ **基于预约机制的随机接入方式**

283

5.4.1 概述

在稳定状态下，平均通过量 \bar{a}_c 等于网络负荷（全部呼叫量） a 与分组成功发送概率 $P_{成功}$ 的乘积，即

$$\bar{a}_c = a \cdot P_{成功} \quad (5.96)$$

$$P_{成功} = 1 - P_c$$

在不发生碰撞的情况下，即当 $P_{成功} = 1$ 或 $P_c = 0$ 时，有 $\bar{a}_c = a$ 简称平均通过量为通过量。

286

5.4.1 概述

- 随机接入系统可以看做是M/M/1系统。设计网络时，利用基于排队论的概率模型去**分析、预测时隙的可用度**是解决冲突的有效途径。

随机接入方式的发展：

- 纯ALOHA (P-ALOHA) 技术；
- 时隙ALOHA (S-ALOHA) 技术；
- 载波监听多址访问 (CSMA——Carrier Sense Multiple Access) 技术；
- 带有冲突检测的载波监听多址访问 (CSMA/CD——CSMA with Collision Detection) 技术及带有冲突避免的载波监听多址访问 (CSMA/CA——CSMA with Collision Avoidance) 技术。

284

5.4.1 概述

两个归一化参数：吞吐量 S 和网络负载 G 。

- **吞吐量 S ：**又称为吞吐率，等于在帧的发送时间 T_0 内成功发送的**平均帧数**。显然

$$0 \leq S \leq 1$$

在**稳定情况**下，在时间 T_0 内到达且能够进入系统的平均帧数（即**输入负载**）**应等于吞吐量**。

- **网络负载 G ：**从网络角度来看，等于在 T_0 内总共发送的平均帧数（平均分组数），包括发送成功的帧和因冲突而重发的帧及发送不成功的帧。即

$$G = \lambda T_0 \quad (5.97)$$

287

5.4.1 概述

- **改进的ALOHA协议的一个思路是限制用户使用信道的随机性，从而减小碰撞，提高信道的通过率。**

2. 性能标准

多址通信系统的主要性能参数是：平均通过量、平均分组时延和稳定性。

(1) 平均通过量 \bar{a}_c

多址通信的平均通过量：

在每个发送周期 T_0 时间内，成功发送的平均分组数（即数据帧数）。

或**更准确地定义**为：在很长时间间隔内，成功发送的分组数与信道上连续传输的最大发送分组数之比。

$$\bar{a}_c = a \cdot P_{成功}$$

帧的长度不是用比特，而是用发送该帧所需的时间来表示

285

5.4.1 概述

$$G = a \quad (5.98)$$

在稳定状态下，有

$$S = G \cdot P_{成功} \quad (5.99)$$

通过量（有时也称通过率） \bar{a}_c 定义为

$$\bar{a}_c = \frac{\text{平均成功发送的数据帧所占的时间}}{\text{观察时间}} = \frac{T_0 \text{内成功发送的数据帧数} \cdot T_0}{T_0} = T_0 \text{内成功发送的数据帧数} = S \quad (5.100)$$

(2) 平均分组时延

分组时延时间或响应时间：为一个分组从信源发出时刻开始，到达信道，直至最后成功地被接收时刻为止的这段时间。

平均分组时延：为一个很长时间间隔内分组总延迟与间隔内分组数之比。

288

5.4.1 概述

(3) 稳定性

系统工作的稳定性：在较长时间段内，其**通过量**和**延迟特性**基本保持不变，即其性能参数基本保持不变。

289

5.4.2 基本的随机接入系统业务分析

(2) 数学模型

假设

- 设有无限多个用户共用一个信道，这些用户的总呼叫到达率为 λ （包括新发的数据帧和重发的数据帧），为泊松流到达。
- 设信息包（或数据帧）都采用纠错编码，其长度为定长，即帧长固定，用 P 表示帧长；
- T_0 为服务时间，即发送一帧占用信道的时间，用信息包长度来代表发送这个信息包的时间。则有 $P = T_0$ ；
- a 为呼叫量， $a = \lambda T_0$

292

5.4 随机接入系统业务分析

5.4.1 概述

5.4.2 基本的随机接入系统业务分析

5.4.3 基于监听的随机接入系统业务分析

5.4.4 基于预约机制的随机接入系统业务分析

5.4.5 应用举例

290

5.4.2 基本的随机接入系统业务分析

t 内有 k 个呼叫或信息包发送到信道的概率为

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad (5.101)$$

t 内无包发送的概率为

$$P_0(t) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t} \quad (5.102)$$

293

5.4.2 基本的随机接入系统业务分析

1. 纯ALOHA (P-ALOHA) 系统

P-ALOHA工作环境：无线信道或总线型网络中。设有无限个用户共用一个信道，这些用户的总呼叫到达率为 λ ，为泊松流到达。

(1) P-ALOHA的工作原理

- **纯随机方式抢占信道**：某数据站（用户）有信息要发送时，立即发送。若在规定的时间内收到确认信号ACK，表示发送成功；若收到否认信号NAK，表示发生了碰撞。若未收到ACK或NAK信号，表示未能成功发送，则该站重发此信号。
- **发生碰撞或发送失败后的重发**：当冲突现象发生后或未能成功发送，则数据站隔一段随机时间重发该信息。

291

5.4.2 基本的随机接入系统业务分析

(3) 性能分析

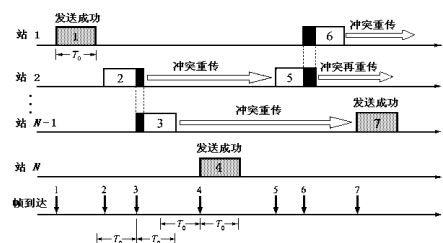


图5.16 P-ALOHA系统的工作原理

294

5.4.2 基本的随机接入系统业务分析

(3) 性能分析

① 发送成功的概率（无碰撞的概率）

➤ 一个数据帧发送成功的概率为

$$P[\text{成功概率}] = P[\text{连续两个到达时间间隔} > 2T_0] = (P[\text{到达时间间隔} > T_0])^2$$

T ：两个帧的到达时间间隔，则有

$$P_{\text{成功}} = P(T > 2T_0) = [P(T > T_0)]^2 = P_0^2(T_0) = (e^{-\lambda T_0})^2 = e^{-2\lambda T_0} = e^{-2G} \quad (5.103)$$

➤ 发生碰撞的概率为

$$P = 1 - e^{-2\lambda T_0} = 1 - e^{-2G} \quad (5.104)$$

G ：网络负载

295

5.4.2 基本的随机接入系统业务分析

结论：

➤ 设备简单，效率低。系统最多只能有18.4%的时间能成功地发送信息而不发生碰撞，实现正常通信。

➤ 当 a 较小时， $\bar{a}_c \approx a$ 基本上可以顺利通信。

➤ 为使系统能够稳定工作，网络负载 a

（或网络负载 G ）一定不能大于0.5。

(4) 帧的时延和重发

重发帧采用间隔一个随机时间再发送的策略。这个随机时间是由采用的**随机退避延时算法**决定的，不同的算法对系统的性能有不同的影响。

298

5.4.2 基本的随机接入系统业务分析

② 通过量 \bar{a}_c （或吞吐量 S ）

$$\bar{a}_c = S = G \cdot P_{\text{成功}} = G \cdot e^{-2G} = a e^{-2a}$$

最大通过量 $\bar{a}_{c\max}$

$G = a = 1/2$ 时，

$$\bar{a}_c = \bar{a}_{c\max} = \frac{1}{2} \cdot e^{-1} = 0.184 = 18.4\%$$

296

5.4.2 基本的随机接入系统业务分析

2. 时隙ALOHA (S-ALOHA) 系统

S-ALOHA系统特点：网内所有的用户都与主时钟同步，有通信要求的用户信息到达信道的时刻只能是各时隙的起始时刻。主时钟的同步信息要向所有用户广播。

(1) S-ALOHA的工作原理

• S-ALOHA系统把时间分成一段段等长的时隙 (Slot)，记为 T_0 ，并规定不论帧何时产生，每个用户只能在每个时隙的前沿发送信息。

T_0 长度的确定：每个帧正好在一个时隙内发送完毕。

• 当在一个时隙内有两个或两个以上的帧到达时，则在下一个时隙将产生冲突，两组数据完全重叠。

• 冲突后重发的策略与P-ALOHA相似，不同的是应该在一个随机整数倍的时隙时延后重发。

299

5.4.2 基本的随机接入系统业务分析

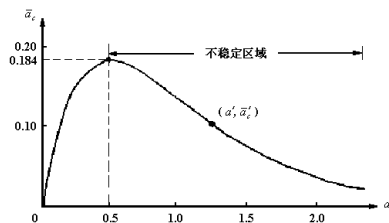


图5.17 P-ALOHA系统的通过量与网络负载 a 的关系曲线

297

5.4.2 基本的随机接入系统业务分析

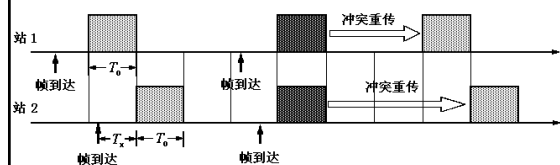


图5.19 S-ALOHA系统的工作原理

300

5.4.2 基本的随机接入系统业务分析

(2) 性能分析

假设条件与P-ALOHA系统相同。

① 发送成功的概率

发送成功的概率为

$$P_{\text{成功}} = P(T > T_0) = P_0(T_0) = \frac{(\lambda T)^k}{k!} e^{-\lambda T} \Big|_{k=0, T=T_0} = e^{-\lambda T_0} = e^{-G}$$

发生碰撞的概率为 $1 - e^{-G}$ (5.107)

② 通过量 \bar{a}_c (或吞吐量S)

$$\bar{a}_c = S = G \cdot P_{\text{成功}} = G \cdot e^{-G} = ae^{-a} \quad (5.108)$$

$$\bar{a}_{c \max} = G \cdot e^{-G} \Big|_{G=1} = a \cdot e^{-a} \Big|_{a=1} = 0.368 = 36.8\%$$

301

5.4.2 基本的随机接入系统业务分析

(2) 采用S-ALOHA方式, 取 $\bar{a}_c = 0.368$,

信道的总数据率为:

$$0.368 \times 4800 \text{ b/s} = 1766.4 \text{ (b/s)},$$

此时系统一共可容纳交互式用户数为:

$$(1766.4 \text{ b/s}) / (5 \text{ b/s}) = 353 \text{ (个)}$$

304

5.4.2 基本的随机接入系统业务分析

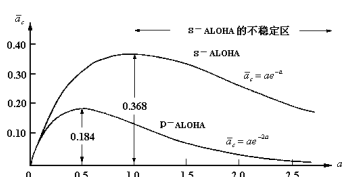


图5.20 S-ALOHA与P-ALOHA系统的通过量曲线

> 与P-ALOHA系统相比, 通过率提高了一倍

302

5.4 随机接入系统业务分析

5.4.1 概述

5.4.2 基本的随机接入系统业务分析

5.4.3 基于监听的随机接入系统业务分析

5.4.4 基于预约机制的随机接入系统业务分析

5.4.5 应用举例

305

5.4.2 基本的随机接入系统业务分析

例5.10 假定许多异步终端通过多点线路连到主机, 线路的数据率为4800b/s, 设每份报文有60个字符, 而用户用键盘输入一份报文需2分钟。每个字符用10比特进行编码, 则每个终端的平均数据率仅有5b/s [(60×10bit) / (2×60s) = 5b/s]。

试问: (1) 采用P-ALOHA方式, 取 $\bar{a}_c = 0.1$,

该系统一共可容纳多少个交互式用户?

(2) 采用S-ALOHA方式呢?

解: (1) 根据题意, 采用P-ALOHA方式, 取 $\bar{a}_c = 0.1$

, 即仅利用信道容量的10%, 则信道的总数据率为:

$$0.1 \times 4800 \text{ b/s} = 480 \text{ (b/s)},$$

此时该系统一共可容纳交互式用户数为:

$$(480 \text{ b/s}) / (5 \text{ b/s}) = 96 \text{ (个)}$$

303

5.4.3 基于监听的随机接入系统业务分析

载波监听多址访问(接入)方式, 又称为载波侦听多点访问, 简称CSMA。这是ALOHA系统的一种改进形式, 适用于延时较小的网络。

1. CSMA的工作原理及监听方式

(1) CSMA的工作原理

CSMA与ALOHA技术的主要区别:

- 每个用户终端(数据站)多了一个附加的硬件装置, 称为**载波监听装置**。
- 功能: 每个数据站在发送数据前先监听信道, 以接收到信息与否来判断信道上的忙闲状态, 各用户只能在信道空闲时启动发送数据。

306

5.4.3 基于监听的随机接入系统业务分析

① **监听发送过程**：用户（数据站）有数据发送，则在发送数据前，先监听信道是否空闲。如果空闲，则该站可以发送数据；若忙，则该站就暂时不发送数据，而是按某种算法，隔一段时间后再尝试发送。

② **重发过程**：数据发送以后，该站点在规定的时间内若收到对方的确认信号ACK，则发送成功，可以进行下一个数据的发送尝试，否则重发该数据。

(2) CSMA的监听方式

发生碰撞的原因有两个：

- 两个站或多个站同时监听到信道空闲，同时发送数据造成冲突。
- 由于传输时延 τ 的存在，使得在 τ 时间内，监听装置可能检测不到信道上信息包的存在而发送数据。

307

5.4.3 基于监听的随机接入系统业务分析

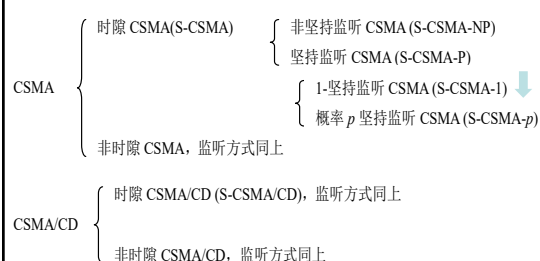


图5.21 CSMA技术分类

310

5.4.3 基于监听的随机接入系统业务分析

监听方式分为：非坚持监听CSMA和坚持监听CSMA两类。

• **非坚持监听CSMA (CSMA-NP)**：用户监听到信道“忙”状态信号（即发现有其他站在发送数据）或发送失败信号后，停止监听。在重新发送分组之前，根据协议的算法延迟一个随机时间后再重新监听，直到听到信道“闲”状态信号后再发送数据。

• **坚持监听CSMA (CSMA-P)**：用户一直连续在监听，一旦发现信道空闲就以不同方式发送数据。

该方式又分为**1-坚持监听**和**概率 p 坚持监听**两种方式：

308

5.4.3 基于监听的随机接入系统业务分析

CSMA技术的优点：

- 通过采用**先听后发**技术减少了冲突，提高了整个系统的吞吐量。CSMA-NP的最大通过率可达80%以上，

缺点：

- 由于退避的原因，时延性能较差；
- 由于传输时延 τ 的存在，冲突不可避免；限制了网络的覆盖范围，只在小时延（即短距离）情况下有效。

2. CSMA/CD多址接入系统

CSMA/CD系统（带有冲突检测的载波监听多址接入（访问））CSMA可采用冲突检测，就是**边发送边监听**。一旦监听到发生冲突（碰撞），则冲突的双方或多方都必须停止发送，以使信道不致无效地继续被占用。这就是**冲突检测功能**。

优点：

- 提高了通过量和信道利用率。在时延增加后也将逐渐消失。

311

5.4.3 基于监听的随机接入系统业务分析

✓ **1-坚持监听 (CSMA-1)**：用户（站）有数据要发送，则先一直连续监听信道并等待发送，当监听到信道忙时，仍坚持监听，一旦监听到信道空闲，就以概率1立即发送数据；

✓ **概率 p 坚持监听 (CSMA-p)**：当监听到信道空闲时，就以概率 p 在第一有效时隙内发送数据，或以概率 $(1-p)$ 延迟一段时间 τ ，重新监听信道，开始下一次的发送尝试。任意节点以 $(1-p)$ 的概率主动退避，放弃发送分组的机会，即信道空闲也可能不发送数据，因而可以进一步减小数据的碰撞的概率。可用于分时隙的信道。

确定 p ：

- ✓ p 事先给定，根据信道上的通信量的多少设定。
- ✓ 若信道空闲，在 $[0, 1]$ 区间选择一个随机数 I ，若 $I \leq p$ ，则发送数据；否则，延时 τ 后再重新监听信道。

309

5.4.3 基于监听的随机接入系统业务分析

CSMA/CD系统的工作原理如下：

(1) 用户监听到信道空闲就发送数据帧。

(2) 在发送数据帧后**继续监听**，进行**冲突检测**。

一种可行的**冲突检测方法**是：在发送帧的时候也同时进行接收，将从信道上接收到的数据与本站发送的数据逐比特进行比较。若不一致，则说明发生了冲突。

(3) 如果两个或多个终端同时开始发送数据，那么就会检测到碰撞。一旦**检测到冲突**，就**立即停止发送数据**，而不管这个数据发完；并发送一简短的**阻塞信号**，使所有的站点都知道已发生了冲突，停止发送数据，转入闲期。

(4) 在发送阻塞信号后，该站延迟一段随机时间，然后采用CSMA方式再次尝试发送。

阻塞信号可以通过发送若干比特的人为干扰信号（jamming signal）而产生。

312

5.4.3 基于监听的随机接入系统业务分析

3. 随机退避延时算法

不同的随机退避算法有不同的传输效率。

常用的随机退避延时算法主要有：

- 均匀随机数退避法
- 二进制指数退避法
- 线性增量退避法
- 顺序退避法

减少冲突的可能性的措施：

- 采用概率 p 发送数据帧的方式；
- 采用优先制发送数据帧的方式，即某些节点主动退避；
- 采用分时隙的方式。

完全消除碰撞——采用非竞争的集中控制机制

313

5.4.4 基于预约机制的随机接入系统业务分析

基于预约机制的随机接入方式的优点：

- 可以最大程度地减小或者消除随机因素对控制过程的影响；
- 能够进一步减少或消除冲突，提高传输质量。

缺点：

- 预约信息需要占用信道资源，同时也增加了系统的复杂度。

分类：

- 显式预约
- 隐式预约

不论是哪一类，都有S-ALOHA和TDMA两类接入方式。

316

5.4.3 基于监听的随机接入系统业务分析

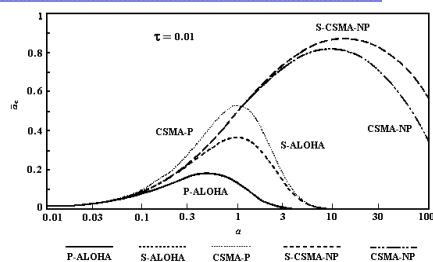


图5.22 $\tau=0.01$ 时的各种随机接入方式的通过量 \bar{s} 与网络负荷 α 的关系曲线

314

5.4.4 基于预约机制的随机接入系统业务分析

- 显式预约：要求系统固定地或动态地划出一小部分信道资源用于预约信息的开销，故在一次预约有效期间不会发生分组碰撞。
- 隐式预约：不要求系统固定地划出一小部分信道资源用于预约信息的开销，预约的信息是通过分组碰撞来预示的，由分组碰撞造成的信道资源损失就是等效的预约开销。

基于预约机制的多址接入协议有：

- 预约ALOHA
- 分组预约多址（PRMA）等。

317

5.4 随机接入系统业务分析

5.4.1 概述

5.4.2 基本的随机接入系统业务分析

5.4.3 基于监听的随机接入系统业务分析

5.4.4 基于预约机制的随机接入系统业务分析

5.4.5 应用举例

315

5.4.4 基于预约机制的随机接入系统业务分析

1. 预约ALOHA

- 是以时分多址技术（TDMA）为基础的一个分组接入模式。在此协议中，某个分组时隙被赋予优先级，并且能够为用户预约发送分组的时隙。时隙能够被永久预约或者按要求预约。

2. 分组预约多址（PRMA）

- 20世纪80年代末，出现的分组预约多址（PRMA：Packet Reservation Multiple Access）是S-ALOHA协议和TDMA技术相结合的动态多址接入协议。
- PRMA是用一种类似于预约ALOHA的离散分组时间技术，可以支持在无线信道中传输可变的混合语音分组和数据分组，实现综合的语音和数据接入。它以每一个TDMA时隙传送语音和数据，并且优先传输语音，同时利用了TDMA的周期帧结构。

316

5.4 随机接入系统业务分析
5.4.1 概述
5.4.2 基本的随机接入系统业务分析
5.4.3 基于监听的随机接入系统业务分析
5.4.4 基于预约机制的随机接入系统业务分析
5.4.5 应用举例
319

5.4.5 应用举例

表5.2 多址接入技术的应用

业务类型	多址技术
突发，短信息	竞争协议 (ALOHA, S-ALOHA, CSMA, CSMA/CD, CSMA/CA)
突发，长信息，大量用户	预约协议
突发，长信息，少量用户	显式 TDMA 预约信道的预约协议
流数据或确定性信息（如语音）	FDMA, TDMA, CDMA

322

5.4.5 应用举例
1. 基本的随机接入方式的应用
特点:
• 实现简单，但吞吐量低，适合于具有突发性、负荷较低的数据业务。
• 适用于网络覆盖较大的情况，如在广域网中的应用。
应用: 现多用于无线分组网及卫星通信系统中。
举例: 随机接入方式通常用于链路的建立阶段。例如，在GSM和IS-95系统中，上行链路的控制信道中的随机接入信道 (RACH: Random Access Channel) 的接入大多采用S-ALOHA协议。在呼叫建立期间为用户分配一个无冲突的信道，即一个TDMA子信道作为通话期间的固定话路。
320

5.4.5 应用举例
2. 基于监听的随机接入方式的应用
特点:
• 采用监听方式，减少了冲突，改进了系统的吞吐性能和时延性能，从而提高了信道利用率。
• 冲突没有完全消除。传输延时越大，影响越大。
应用:
适用于网络覆盖较小的情况，如计算机局域网的接入。现主要应用于短距离数据业务的多址接入。
➢ IEEE802.3以太网 (Ethernet) 采用了CSMA/CD协议
➢ IEEE802.11无线局域网 (WLAN) 采用了CSMA/CA协议机制。
321