通信网络理论基础

Information Science Institute,
Xidian University

M/G/1型排队系统

M/G/1型排队系统

用户数不具有无后效性

- 服务过程
  - 假定第*i*个用户的服务时间为***Xi***，***Xi***是独立同分布的，
  - 与到达间隔相互独立。
  - 令 $X=\{X_1,X_2,\cdots\}$，则服务时间的均值和二阶矩为

平均服务时间 = $\overline{X}$ = $E\{X\}=\dfrac{1}{\mu}$

服务时间的二阶矩 = $\overline{X^2}$ = $E\{X^2\}$

P-K公式

- 下面我们将证明，M/G/1排队系统的平均等待时间为

$$\boxed{W=\frac{\lambda\overline{X^2}}{2(1-\rho)}}$$

- 其中， $$\rho={\lambda}/{\mu}=\lambda\overline{X}$$

- 该式称为P-K（Pollaczek-Khinchin）公式。

P-K公式

- 根据上述P-K公式，应用Little定理,可以得该系统的平均时延为

$$T=\overline{X}+W=\overline{X}+\frac{\lambda\overline{X^2}}{2(1-\rho)}$$

- 平均队列长度为

$$N_Q=\lambda W=\frac{\lambda^2\overline{X^2}}{2(1-\rho)}$$

- 系统中的平均用户数为

$$N=\lambda T=\lambda\overline{X}+\frac{\lambda^2\overline{X^2}}{2(1-\rho)}$$

P-K公式

- 如果G＝M，即服务时间服从指数分布，M/M/1，

$$\overline{X^2} = 2/\mu^2$$

$$\boxed{W = \frac{\lambda \overline{X^2}}{2(1-\rho)} = \lambda \frac{2}{\mu^2}\frac{1}{2(1-\rho)} = \frac{\rho}{\mu(1-\rho)}}$$

- 如果服务时间是常量，即M/D/1

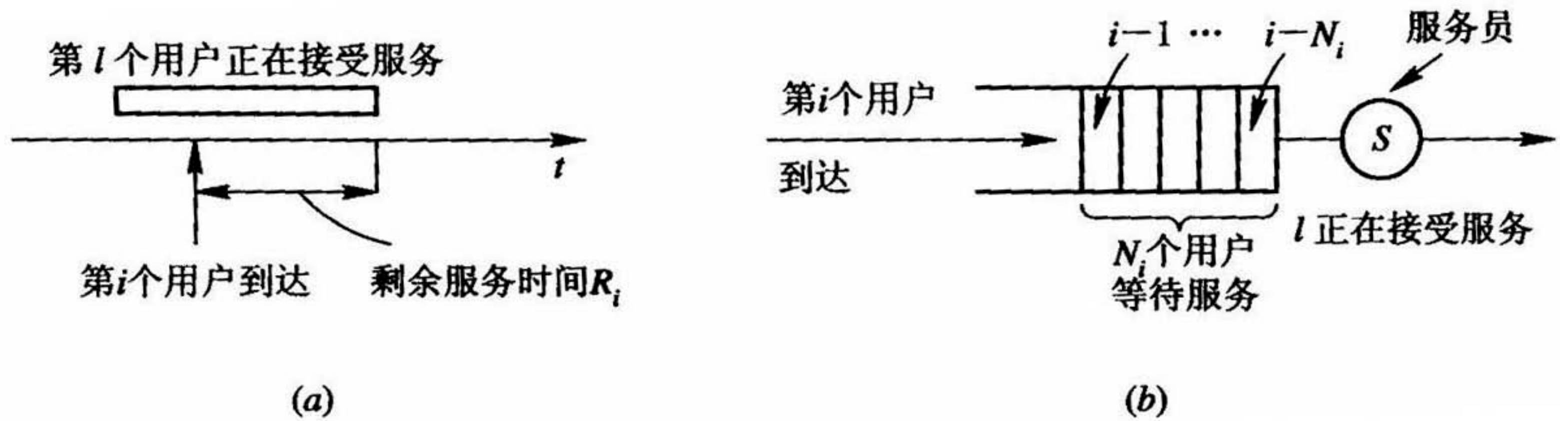
等待时间是**M/M/1**系统的一半

$$\overline{X} = 1/\mu \qquad \overline{X^2} = 1/\mu^2$$

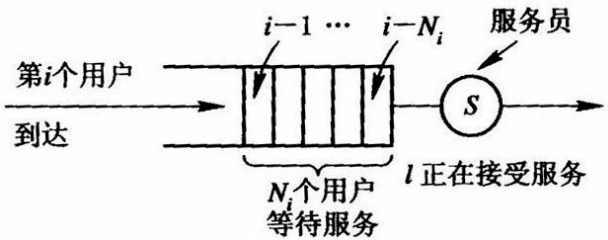
$$\boxed{W = \frac{\lambda \overline{X^2}}{2(1-\rho)} = \lambda \frac{1}{\mu^2}\frac{1}{2(1-\rho)} = \frac{\rho}{2\mu(1-\rho)}}$$

P-K公式

- P-K公式证明的思路是基于对平均剩余服务时间（Mean Residual Service Time）的求解。设第i个用户到达系统时，第l个用户正在接受服务，其剩余服务时间为R_i，此时等待队列中有N_i个用户，



P-K公式



- 设第***k***个用户的服务时间为X_k，则由图可知，用户***i***的等待时间为

$$W_i = R_i + N_i\text{个用户的服务时间} = R_i + \sum_{k=i-N_i}^{i-1} X_k$$

- 对上式求平均得

$$\overline{W_i} = E\{R_i\} + E\left\{\sum_{k=i-N_i}^{i-1} X_k\right\} = E\{R_i\} + \overline{X} \bullet E\{N_i\}$$

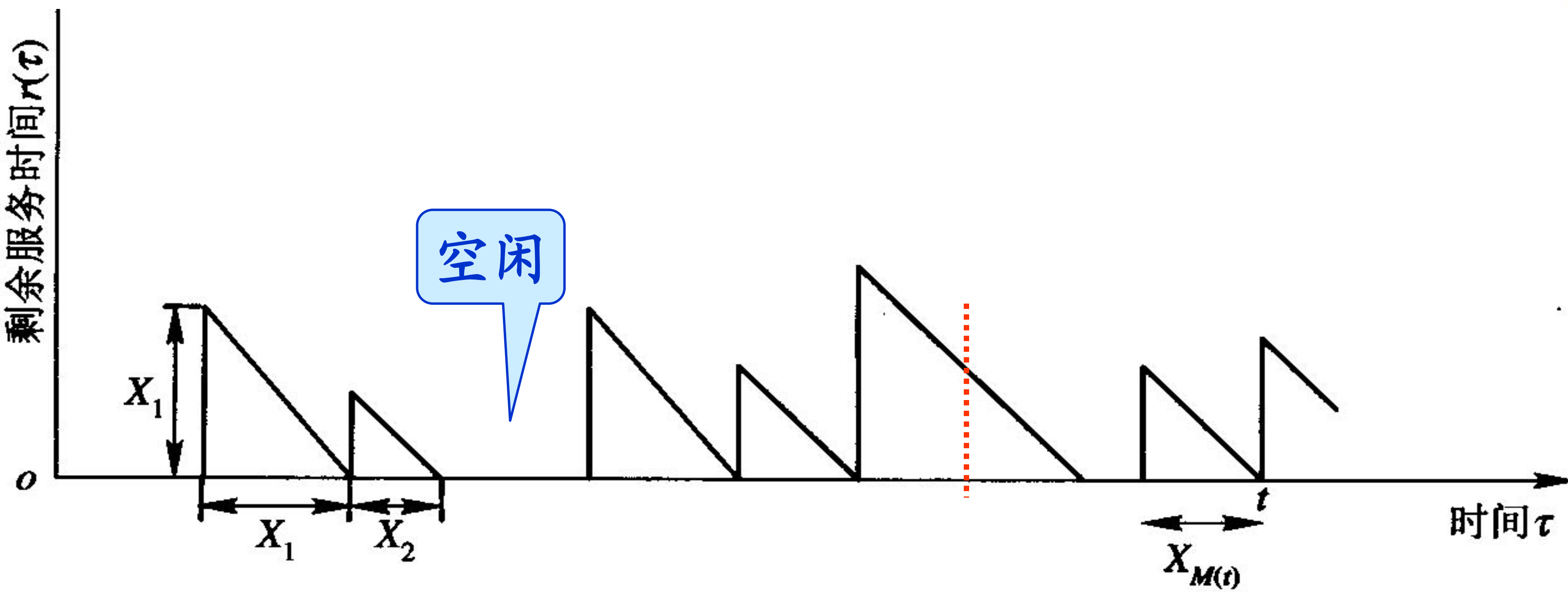
$$W = \lim_{i\to\infty} \overline{W_i} \qquad R = \lim_{i\to\infty} E\{R_i\} \qquad N_Q = E\{N_i\}$$

$$W = R + \overline{X} N_Q = R + \frac{1}{\mu} N_Q = R + \frac{1}{\mu}\lambda W = R + \rho W$$

$$W = \frac{R}{1-\rho}$$

P-K公式

- 假定系统有稳态解，且具有各态历经性，则剩余服务时间可用下图表示



P-K公式

- 为了方便起见，取 t 为 $r(t)=0$ 的时刻，则 [0, t] 区间的平均剩余服务时间为

时间平均=集合平均

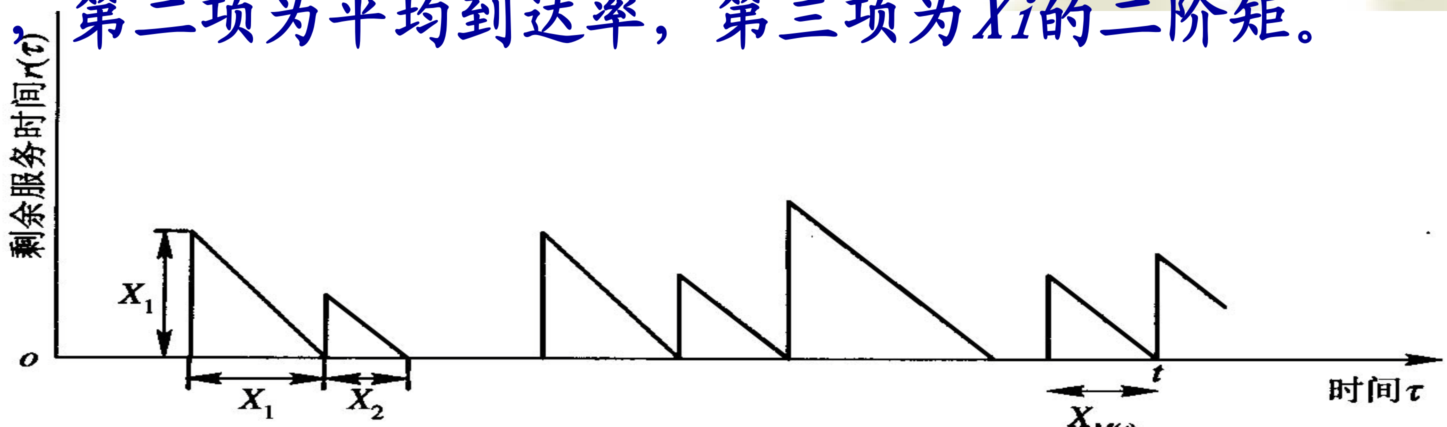
$$R_t=\frac{1}{t}\int_0^t r(\tau)d\tau=\frac{1}{t}\sum_{i=1}^{M(t)}\frac{1}{2}X_i^2$$

- 式中，$M(t)$ 表示 [0, t] 区间内已服务的用户数。 上式可以写成

系统稳态时，到达率=退去率

$$R_t=\frac{1}{2}\bullet\frac{M(t)}{t}\bullet\frac{\sum_{i=1}^{M(t)}X_i^2}{M(t)}$$

- 式中，第二项为平均到达率，第三项为 Xi 的二阶矩。



P-K公式

- 令 $t \to \infty$ ，得 $R = \frac{1}{2}\lambda \overline{X^2}$

$$W = \frac{R}{1-\rho} = \frac{\lambda \overline{X^2}}{2(1-\rho)}$$

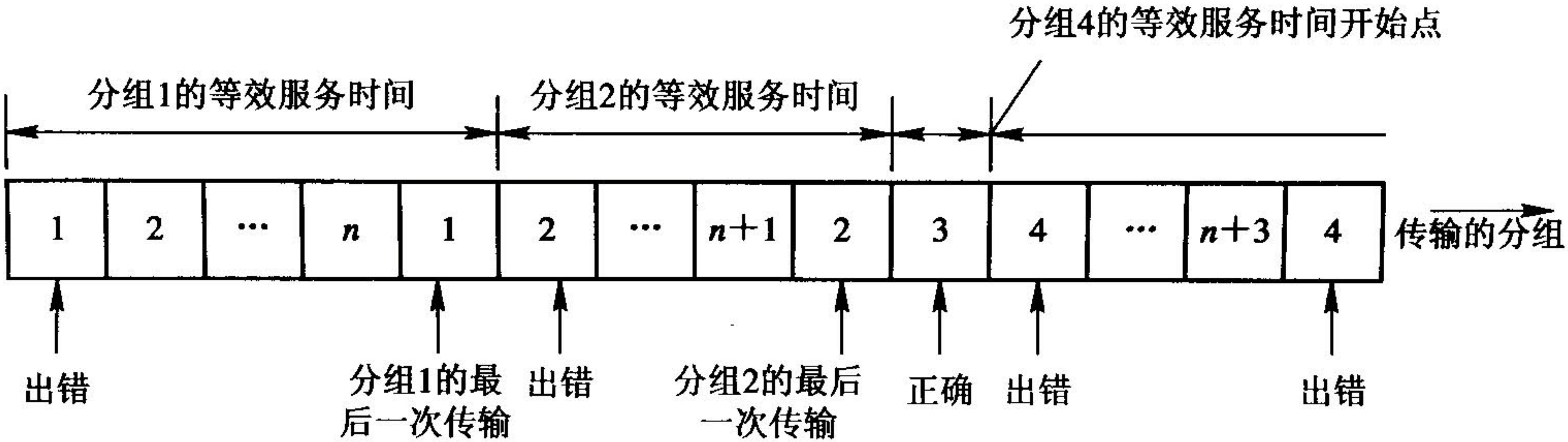
$$\boxed{W \propto \overline{X^2}}$$

- 即使 $\rho < 1$ ，如 果 $\overline{X^2} \to \infty$ ，则 $W \to \infty$

- 它说明，如果有少量用户有非常长的服务时间，一旦这些用户被服务，将导致队列非常长。

P-K公式

- 例 返回$n-$ARQ系统的时延性能分析
  - 设返回$n-$ARQ系统中的分组到达过程是速率为λ的Poisson过程，
  - 分组（帧）长度相同，为一个单位，
  - 等待应答的最长时间（重传间隔）为 $n-1$ 个分组长度。因此，任一分组传输出错后，将在n-1个分组后重传。



1. A given packet transmitted in frame i might be rejected at the receiver due to errors, in which case the transmitter will transmit packets in frames $i+1, i+2, \ldots, i+n-1$, (if any are available), and then go back to retransmit the given packet in frame $i+n$.

2. A packet transmitted in frame i might be accepted at the receiver, but the corresponding acknowledgment (in the form of the receive number) might not have arrived at the transmitter by the time the transmission of packet $i+n-1$ is completed. This can happen due to errors in the return channel, large propagation delays, long return frames relative to the size of the goback number n, or a combination thereof.

P-K公式

- 在该系统中，一个分组（帧）的服务时间，不仅仅是一个分组的一次传输时间，而且还应当包括分组的重传时间。因此，这里的服务时间为一个等效服务时间，它等于从一个分组的第一次传输开始，到该分组的最后一次传输结束时刻之间的时间长度，它服从一般性分布，因而该系统可用M/G/1队列模型来描述。

P-K公式

- 假定分组传输错误的概率为p。如果分组重传次数为k，则等效服务时间为 $1+kn$ ，其概率为（一次正确传输，k次传输错误，共k+1次传输），即等效服务时间X_k的概率分布为 $(1-p)p^k$
- 其一阶矩和二阶矩为

$$P(X_k = 1+kn) = (1-p)p^k$$

$$\overline{X} = \sum_{k=0}^{\infty}(1+kn)(1-p)p^k = 1+\frac{np}{1-p}$$

$$\overline{X^2} = \sum_{k=0}^{\infty}(1+kn)^2(1-p)p^k = 1+\frac{2np}{1-p}+\frac{n^2(p+p^2)}{(1-p)^2}$$

P-K公式

- 该公式的推导过程中，利用了下列结果：

$$\sum_{k=0}^{\infty} p^k = \frac{1}{1-p} \qquad \sum_{k=0}^{\infty} kp^k = \frac{p}{(1-p)^2} \qquad \sum_{k=0}^{\infty} k^2 p^k = \frac{p+p^2}{(1-p)^3}$$

- 利用P-K公式可得，分组的平均等待时延和分组的平均时延分别为

$$\boxed{W = \frac{\lambda \overline{X^2}}{2(1-\rho)}} \qquad W = \frac{\lambda \overline{X^2}}{2(1-\lambda \overline{X})} \qquad T = \overline{X} + W$$

M/G/1队列应用举例

假定p=10^{-3}，n=8，则有， $\overline{X}=1.008$ ， $\overline{X^2}=1.08$ ，

当λ=0.5、0.8、0.9和0.95分组/单位分组时间，则有w≈0.54，2.23，5.24和11.29个单位分组时间。

由此可以看出，当分组的到达率接近系统容量时，分组的平均等待时间急剧上升。因此，为了保证分组的最大等待时间低于某一门限，应当限制分组的到达率。

$$\overline{X}=\sum_{k=0}^{\infty}(1+kn)(1-p)p^k=1+\frac{np}{1-p}$$

$$W=\frac{\lambda\overline{X^2}}{2\left(1-\lambda\overline{X}\right)}$$

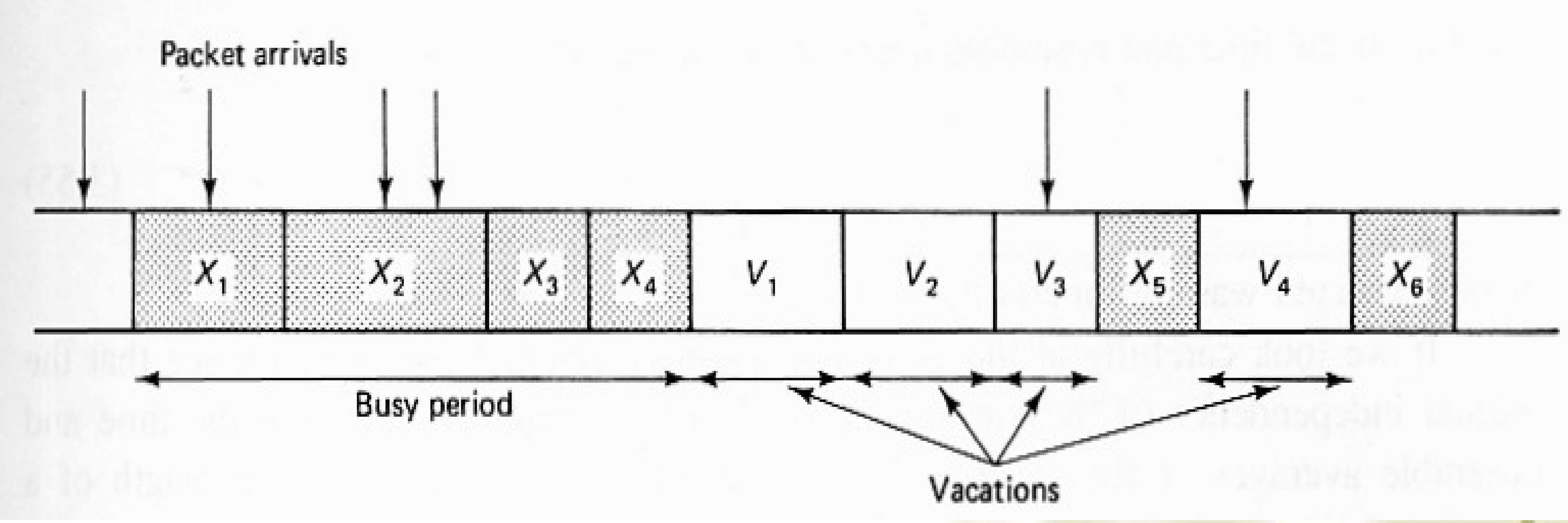
$$\overline{X^2}=\sum_{k=0}^{\infty}(1+kn)^2(1-p)p^k=1+\frac{2np}{1-p}+\frac{n^2\left(p+p^2\right)}{(1-p)^2}$$

3.3.2 服务员有休假的M/G/1排队模型

服务员有休假的M/G/1排队模型

- 服务员有休假的M/G/1（M/G/1 Queues with Vacations）排队系统是指在每一个忙周期后（分组传输结束后），服务员需要休假（休假对应于服务员（通信节点）要进行其它处理，如存储数据、信令交换等），在服务员休假期内到达的用户，要等待服务员休假结束后，才能被服务。如服务员休假期满后，没有用户到达，服务员进入另一个休假期。

服务员有休假的M/G/1排队模型



服务员有休假的M/G/1排队系统的分组传输过程

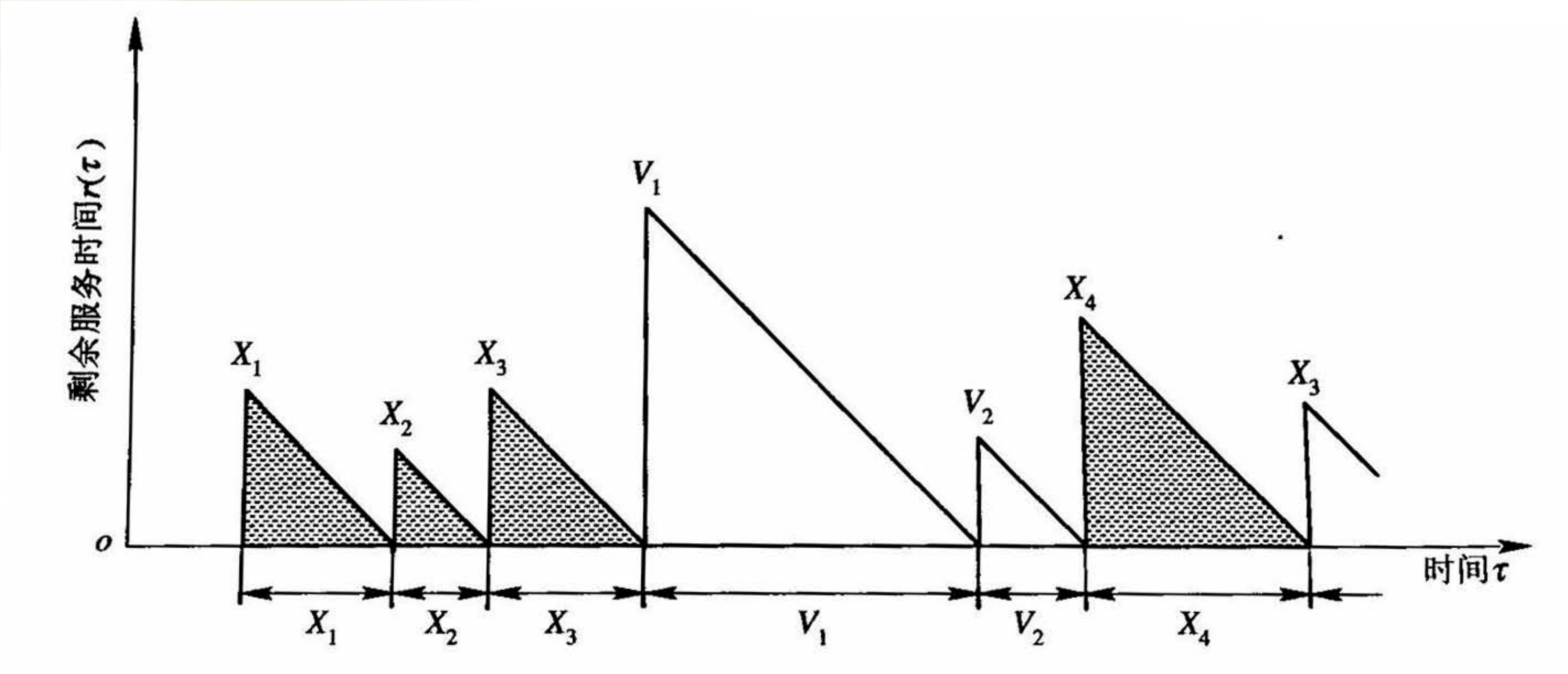
服务员有休假的M/G/1排队模型

- 服务员的休假期用 V_1，V_2，…表示，V_i是独立同分布的随机变量，并且与到达间隔相互独立，其均值为 $\overline{V}$ 。
- 用户（分组）的到达是速率为λ的Poisson过程，每个用户的服务时间也是独立同分布的随机变量，其平均服务时间为 $1/\mu$

服务员有休假的M/G/1排队模型

- 一个新用户到达系统时，它可能会遇到两种情况
  - 当前有一个分组正在接受服务
  - 服务员正在处于休假期
- 前一种情况的剩余服务时间与标准的M/G/1队列相同，后一种情况的剩余服务时间为服务员休假的剩余服务时间。

服务员有休假的M/G/1排队模型



图中阴影部分表示分组传输的剩余服务时间X_i，其余部分是服务员休假期的剩余服务时间V_i。

服务员有休假的M/G/1排队模型

- 令$M(t)$和$L(t)$分别为 $[0,t]$ 区间内到达的用户数和服务员休假的次数。平均剩余服务时间为

$$\frac{1}{t}\int_0^t r(\tau)d\tau = \frac{1}{t}\sum_{i=1}^{M(t)}\frac{1}{2}X_i^2 + \frac{1}{t}\sum_{i=1}^{L(t)}\frac{1}{2}V_i^2$$

$$= \frac{1}{2}\frac{M(t)}{t}\frac{\sum_{i=1}^{M(t)}X_i^2}{M(t)} + \frac{1}{2}\frac{L(t)}{t}\frac{\sum_{i=1}^{L(t)}V_i^2}{L(t)}$$

$$\frac{1}{t}\int_0^t r(\tau)d\tau = \frac{1}{2}\frac{M(t)}{t}\frac{\sum_{i=1}^{M(t)} X_i^2}{M(t)} + \frac{1}{2}\frac{L(t)}{t}\frac{\sum_{i=1}^{L(t)} V_i^2}{L(t)}$$

- 由于分组传输和休假期的交替到达占满整个时间轴，则在单位时间，分组所占的比例为 $\lambda/\mu=\rho$ ，
- 休假期占的比例为 $1-\rho$ 。
- 因而有休假到达率为 $t(1-\rho)/L(t) \to \overline{V}$ $(1-\rho)/\overline{V}$
- 当 $t \to \infty$ ，得平均剩余服务时间为

$$R = \frac{1}{2}\lambda\overline{X^2} + \frac{1}{2}\frac{1-\rho}{\overline{V}}\overline{V^2}$$

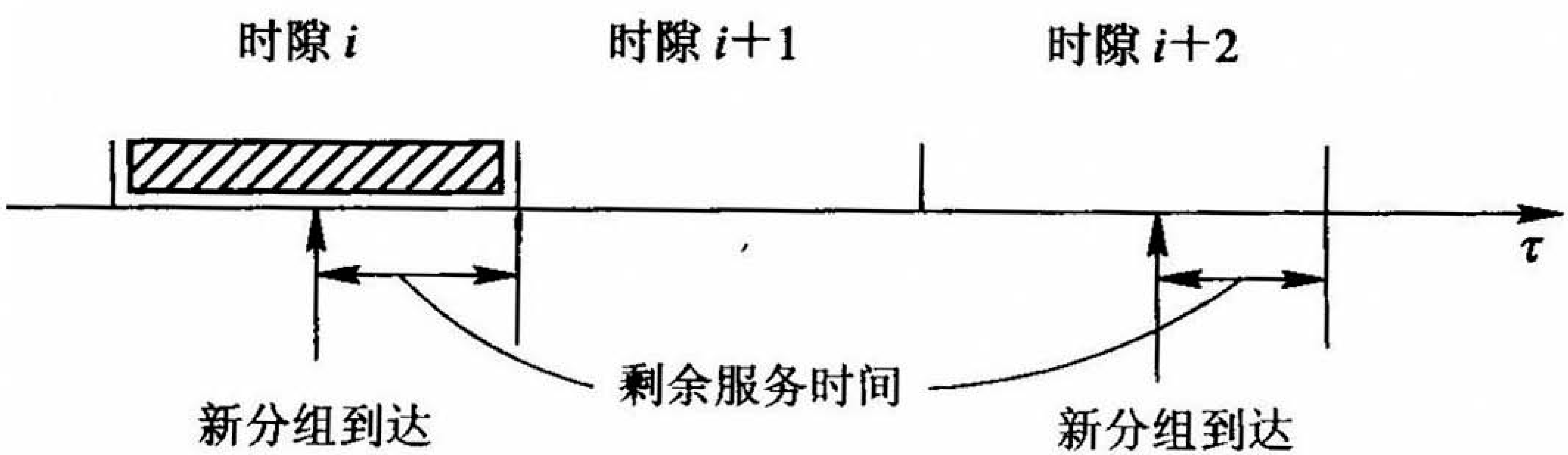
$$\boxed{W = \frac{R}{1-\rho} = \frac{\lambda\overline{X^2}}{2(1-\rho)} + \frac{1}{2}\frac{\overline{V^2}}{\overline{V}}}$$

- 例：时隙FDM和TDM系统的性能比较
- 设基本的FDM（Frequency Division Multiplexing）系统
  - 有m个信道，每个信道的分组到达率为 λ/m ，
  - 每个分组的传输时间为m个单位时间，即服务时间 $1/\mu = m$ 。
  - 在该系统中，只要分组到达时信道空闲，该分组就会立即得到服务。显然每个信道是标准的M/D/1排队系统，
- 其等待时间为

$$\rho = \frac{\left(\lambda/m\right)}{\mu} = (\lambda/m)/(1/m) = \lambda$$

$$W_{FDM} = \frac{\rho}{2\mu(1-\rho)} = \frac{\lambda m}{2(1-\lambda)}$$

- 假设m个信道变成以时隙为基础，时隙的宽度为m个单位，所有的分组都在时隙的开始点进行传输。如果在时隙开始时，没有分组到达，信道将空闲一个时隙（对应于服务员休假一次）。该系统被称为SFDM系统。



(a)

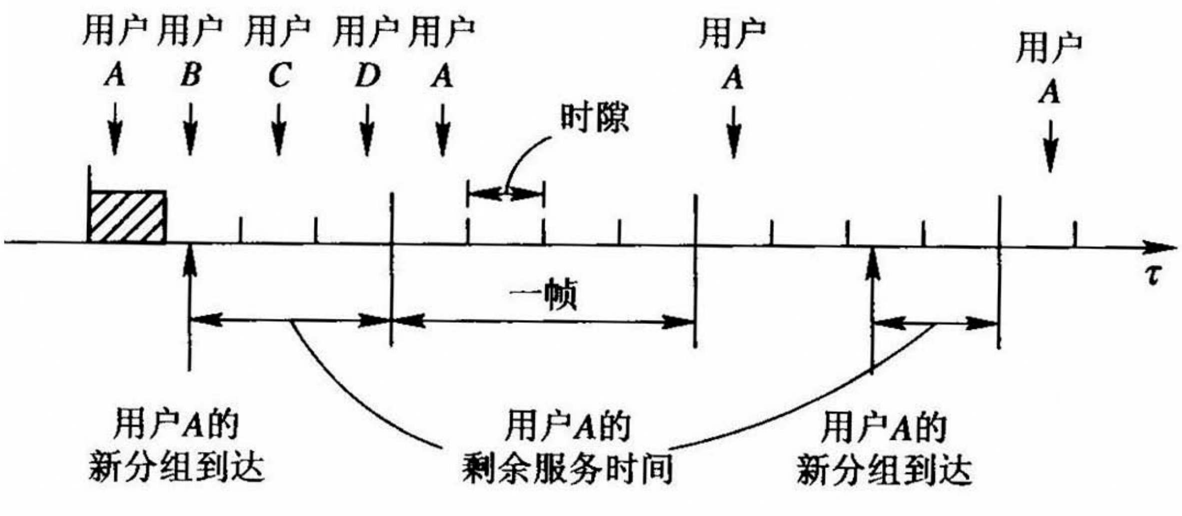
- SFDM系统的每个信道是服务员有休假的M/D/1系统，

$$\rho = \lambda \qquad \overline{X^2} = m^2 \qquad \overline{V} = m \qquad \overline{V^2} = m^2$$

- 该公式与我们直观上的理解是一致的，即在SFDM系统中，分组到达后平均要比FDM系统多等半个时隙。

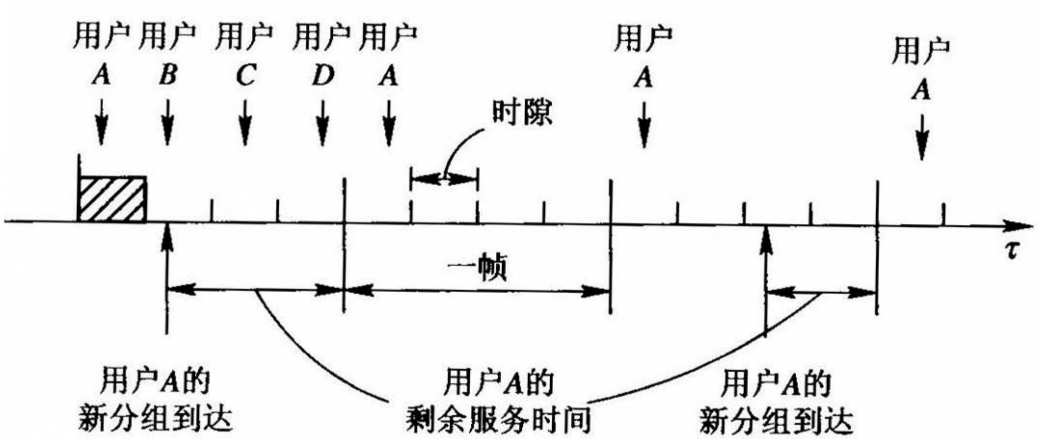
$$W_{SFDM} = \frac{\lambda m}{2(1-\lambda)} + \frac{1}{2}m = W_{FDM} + \frac{1}{2}m = \frac{m}{2(1-\lambda)}$$

- TDM（Time Division Multiplexing）系统，
  - M个信道合成一个信道，信道速率提高了m倍，故分组的传输时间降低为原传输时间的 $1/m$
  - 一帧由m个时隙组成，每个分组的传输时间为一个时隙，时隙的宽度为一个单位时间。



(b)

- TDM（Time Division Multiplexing）系统，
  - 由于每个用户在一帧中仅占一个时隙，如果在时隙开始时刻该用户无分组到达，则必须等到下一帧中相同的时隙才可能开始传输分组。
  - 因此，对于每个用户来说，信道将暂停（休假）一帧（*m*个时隙）。在TDM系统中，剩余服务时间与SFDM系统的完全相同，即当该用户在某一帧有分组传送时，他等待的剩余服务时间不是一个时隙的剩余时间，而是一帧中的剩余时间，



(b)

- 因此，**TDM**系统中每个用户的等待时间为

$$W_{TDM} = W_{SFDM} = \frac{m}{2(1-\lambda)}$$

- 上述三种系统的分组总的时延分别为

$$T_{FDM} = \boxed{m} + \frac{\lambda m}{2(1-\lambda)}$$

$$T_{SFDM} = T_{FDM} + \frac{m}{2}$$

$$T_{TDM} = \boxed{1} + \frac{m}{2(1-\lambda)} = T_{FDM} - (\frac{m}{2} - 1)$$

总结

- 排队系统模型及表示方法
- Little定理
- 各种排队模型的性质和特点
  - M/M/1型排队系统
  - M/M/m型排队系统
  - M/G/1型排队系统

作业

3.13

3.14