Chapter 5

Elementary Performance Analysis
5.0 Queuing Systems

Queuing systems can be characterized by five components

1. The interarrival-time probability density function
2. The service-time probability density function
3. The number of servers
4. The queuing discipline
5. The amount of buffer space in the queues

The notation $A/B/M$

$A$ and $B$ can be $M$ (Markov), $D$ (Deterministic) or $G$ (General)
5.1 Poisson Process

Three basic statements are used to define the Poisson arrival process. Consider a small time $\Delta t (\Delta t \to 0)$

![Diagram showing time intervals $t$, $t+\Delta t$, and $\Delta t$.]

Then

1. The probability of one arrival in the interval $\Delta t$ is defined to be $\lambda \Delta t + o(t)$, $\lambda \Delta t \ll 1$, and $\lambda$ a specified proportionality constant.
2. The probability of zero in $\Delta t$ is $1 - \lambda \Delta t + o(t)$.
3. Arrivals are memoryless = An arrival (event) in one time interval of length $\Delta t$ is independent of events in previous or future intervals.

Ref: Mischa Schwartz “Telecommunication Networks” Addison-Wesley publishing company 1988
Consider a sequence of $m$ small intervals, each $\Delta t$ units long. Let the probability of one arrival in any interval $\Delta t$ be

$$p = \lambda \Delta t$$
$$q = 1 - \lambda \Delta t$$

Because it is memoryless, the probability of $k$ arrivals in the interval $T = m\Delta t$ is given

$$p(k) = \binom{m}{k} p^k q^{m-k}$$
\[
{m \choose k} = \frac{m!}{(m-k)!k!}
\]

Let \( \Delta t \to 0 \) with \( T = m \Delta t \) fixed (i.e. \( m \to \infty \))

\[
\lim_{m \to \infty} \left( \frac{m!}{m^k (m-k)!} \right) p^k q^{m-k} = \frac{(\lambda T)^k}{k!} \left( 1 - \frac{\lambda T}{m} \right)^m \left( 1 - \frac{\lambda T}{m} \right)^{-k}
\]

\[
= \frac{(\lambda T)^k}{k!} \lim_{m \to \infty} \left( 1 - \frac{\lambda T}{m} \right)^m
\]

\[
= \frac{(\lambda T)^k}{k!} e^{-\lambda T}
\]

\[
= \frac{\tau^k}{k!} e^{-\tau}
\]

\[
\tau = \lambda T
\]

\[
P(k) = \frac{\tau^k}{k!} e^{-\tau}
\]
Let $\tau$ be the random variable representing the time to the first arrival after some arbitrary time origin.

\[ p(\tau > x) = \text{prob(number of arrivals in (0,x)=0) } = e^{-\lambda x} \]

\[ p(\tau \leq x) = \text{The cumulative probability distribution } F_\tau(x) \text{ of the random variable } \tau \]

\[ = 1 - e^{-\lambda x} \]

The probability density function $f_\tau(\tau)$

\[ = \lambda e^{-\lambda \tau} \]

\[ E(\tau) = \int_0^\infty \tau f_\tau(\tau) d\tau = \frac{1}{\lambda} \quad \sigma^2_\tau = \frac{1}{\lambda^2} \]

\[ \int_{\lambda x}^\infty \lambda e^{-\lambda \tau} d\tau = \int_{\lambda x}^\infty \tau e^{-\lambda \tau} \frac{1}{\lambda} d(\lambda \tau) \]

\[ = \int_x^\infty \frac{x}{\lambda} e^{-x} dx = \frac{1}{\lambda} \]
5.2 The M/M/1 Queue

This is a queue of single-server type, with Poisson arrivals, exponential service-time statistics, and first-come-first-out service.

\[ \lambda : \text{ arrived rate} \]

\[ \mu : \text{ service rate} \]
Let $n$ denote the number of customers in the system at $\Delta t + t$, then
\[ p_n(t + \Delta t) = p_n(t)[(1 - \lambda \Delta t)(1 - \mu \Delta t) + \lambda \Delta t \mu \Delta t + o(\Delta t)] + p_{n-1}(t)[\lambda \Delta t(1 - \mu \Delta t) + o(\Delta t)] + p_{n+1}(t)[(1 - \lambda \Delta t) \mu \Delta t + o(\Delta t)] \]

\( \Delta t \) is small

\[ p_n(t + \Delta t) = p_n(t)[1 - (\lambda + \mu) \Delta t] + \lambda \Delta t p_{n-1}(t) + \mu \Delta t p_{n+1}(t) \]

\[ \frac{dp_n(t)}{dt} = -(\lambda + \mu) p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t) \]

At equilibrium, \( p_n(t) \) should approach a constant \( p_n \),
then \( \frac{dp_n(t)}{dt} = 0 \), so we have

\[
(\lambda + \mu) p_n = \lambda p_{n-1} + \mu p_{n+1} \quad n \geq 1
\]

State diagram of M/M/1 queue

\[\text{surface 1} \Rightarrow (\lambda + \mu)p_n = \lambda p_{n-1} + \mu p_{n+1} \quad n \geq 1\]

\[\text{surface 2} \Rightarrow \lambda p_n = \mu p_{n+1}\]
From the state diagram and the above equation, we have

\[ \lambda P_n = \mu P_{n+1} \quad n = 0, 1, 2, \ldots. \]

\[ P_{n+1} = \frac{\lambda}{\mu} P_n = \rho P_n \]

\[ P_n = \rho^n P_0 \]

\[ \sum P_n = 1 \quad \text{geometric distribution} \]

\[ P_0 = 1 - \rho \]

\[ E(n) = \sum_{n=0}^{\infty} nP_n = \rho/(1 - \rho) \]
$E(n) \quad \frac{\rho}{1-\rho}$
Delay is given by
\[ T = \frac{N}{\lambda} = \frac{\rho}{\lambda(1-\rho)} \]

The normalized throughput (input rate/output rate) (the probability that the system is nonempty) = \( \rho \)

The throughput (customer/see) = \( \lambda \) (absolute)
(nonsatulated \( \rho=\lambda \))
(satulated \( \rho=\mu \))
5.3 Finite M/M/1 queue (at most N customers)

\[ \sum_{n=0}^{N} P_n = 1 \]

\[ P_0 = \frac{1 - \rho}{1 - \rho^{N+1}} \]

\[ P_n = \frac{(1 - \rho)\rho^n}{1 - \rho^{N+1}} \]

The blocking probability = \( P_B = P_N = \frac{(1 - \rho)\rho^N}{1 - \rho^{N+1}} \)

The net arrival rate = \( \lambda (1 - P_N) = \mu (1 - P_0) = r \)

\[ \lambda \quad \text{finite M/M/1} \quad \text{throughput (customer/sec)} = \lambda (1 - P_B) \]

The throughput \( r/\mu = \frac{(1 - P_0)}{1 - \rho^{N+1}} = \frac{\rho(1 - \rho^N)}{(1 - \rho^{N+1})} \)
Figure 2–15 Blocking probability, finite M/M/1 queue, congested region

Figure 2–16 Throughput-load characteristic, finite M/M/1 queue
5.4 State-dependent Queues: Birth-death processes

\[ \lambda_n \xrightarrow{n \text{ customers: } p_n} \mu_n \]

arrival rate \hspace{2cm} departure rate

State-dependent queueing system

\[ \lambda_0 \xrightarrow{\mu_1} \lambda_1 \xrightarrow{\mu_2} \lambda_2 \xrightarrow{\ldots} \lambda_{n-1} \xrightarrow{\mu_n} \lambda_n \xrightarrow{\mu_{n+1}} \lambda_{n+1} \]

\[ \lambda_n p_n = \mu_{n+1} p_{n+1} \]

\[ p_n = p_0 \prod_{i=0}^{n-1} \lambda_i / \prod_{i=1}^{n} \mu_i \]
5.5 \( M/M/m : \) The m-server case

\[
\lambda P_{n-1} = n \mu P_n \quad n \leq m
\]

\[
\lambda P_{n-1} = m \mu P_n \quad n > m
\]

\[
P_n = \begin{cases} 
\frac{(m \rho)^n}{n!} & n \leq m \\
\frac{m^m \rho^n}{m!} & n > m
\end{cases}
\]

\[
\sum_{n=0}^{\infty} P_n = 1
\]

\[
P_0 = \left[ 1 + \sum_{n=1}^{m-1} \frac{(m \rho)^n}{n!} + \sum_{n=m}^{\infty} \frac{(m \rho)^n}{m!} \frac{1}{m^{n-m}} \right]^{-1}
\]

\[
\rho = \frac{\lambda}{m \mu}
\]

\[
\lambda_i = \lambda \quad \mu_i = \mu
\]
The probability that an arrival will find all servers busy

\[ P_Q = \sum_{n=m}^{\infty} P_n = \frac{P_0 (m \rho)^n}{m! (1 - \rho)} \]

Erlang C formula

The number of customers waiting in queue (not in service)

\[ N_Q = \sum_{n=0}^{\infty} n P_{n+m} = P_Q \frac{\rho}{1 - \rho} \]

\[ T = \frac{1}{\mu} + W_Q = \frac{1}{\mu} + \frac{P_Q}{m \mu - \lambda} \]

\[ W_Q = \frac{N_Q}{\lambda} = P_Q \frac{\rho}{(1 - \rho) \lambda} = \frac{P_Q}{\left(\frac{1}{\rho} - 1\right) \lambda} = \frac{P_Q}{m \mu - \lambda} \]

\[ \rho = \frac{\lambda}{m \mu} = \frac{1}{\rho} = \frac{m \mu}{\lambda} \]
Figure 2-27 Performance characteristic, M/M/2 queue

- Normalized delay
- Normalized load
Example 1  Statistical Multiplexing Compared with TDM and FDM

Assume \( m \) statistically iid Poisson packet streams each with an arrival rate of \( \lambda/m \) packets/sec.

The packet lengths for all streams are independent and exponentially distributed.

The average transmission time is \( 1/\mu \).

If the streams are merged into a single Poisson stream, with rate \( \lambda \), the average delay per packet is

\[
T = \frac{1}{\mu - \lambda}
\]

If, the transmission capacity is divided into \( m \) equal portions, as in TDM and FDM, each portion behaves like an M/M/1 queue with arrival rate \( \lambda/m \) and average service rate \( \mu/m \). Therefore, the average delay per packet is

\[
T = \frac{m}{\mu - \lambda}
\]
Example 2
Using One vs. Using Multiple Channels Statistical MUX(1)

A communication link serving \( m \) independent Poisson traffic streams with overall rate \( \lambda \).

Packet transmission times on each channel are exponentially distributed with mean \( 1/\mu \). The system can be modeled by the same Markov chain as the \( M/M/m \) queue. The average delay per packet is given by

\[
T = \frac{1}{\mu} + \frac{P_Q}{m\mu - \lambda}
\]

An \( M/M/1 \) system with the same arrival rate \( \lambda \) and service rate \( m\mu \) (statistical multiplexing with one channel having \( m \) times larger capacity), the average delay per packet is

\[
\hat{T} = \frac{1}{m\mu} + \frac{\hat{P}_Q}{m\mu - \lambda}
\]

\( P_Q \) and \( \hat{P}_Q \) denote the queueing probability.
When $\rho << 1$ (light load) $P_Q \approx 0$, $\hat{P}_Q \approx 0$, and

$$\frac{T}{\hat{T}} \approx m$$

At light load, statistical MUX with $m$ channels produces a delay almost $m$ times larger than the delay of statistical MUX with the $m$ channels combined in one.

When $\rho \approx 1$, $P_Q \approx 1$, $\hat{P}_Q \approx 1$, $1/\mu << 1/(m\mu - \lambda)$, and

$$\frac{T}{\hat{T}} \approx 1$$

At heavy load, the ratio of the two delays is close to 1.
5.6 The M/G/1 system

Assumptions:

a. existence of the steady-state average waiting time, residual time and no. of customers in queue

b. ergodic, long term average = ensemble average

\[ \lambda \rightarrow \text{general departure distribution} \]

Poisson arrival \hspace{1cm} (nonpreemptive, no priority)

Let \( X_i \) be the service time of the \( i \)th customer. The random variables \( (X_1,X_2\ldots) \) are identically distributed, mutually independent, and independent of the interarrival times

\[ \bar{X} = E[X] = \frac{1}{\mu} = \text{The average service time} \]

\[ \bar{X}^2 = E[X^2] = \text{The second moment} \]

\[ f(x) = \frac{1}{\pi} \frac{1}{1+x^2} \quad \bar{x^2} \text{ doesn't exist} \]
$W_i = \text{The waiting time in queue of the } i\text{th customer}$

$R_i = \text{The residual service time seen by the } i\text{th customer}$

If no customer is in service
( empty system when $i$ arrives )

$\Rightarrow R_i = 0$
\( N_i = \text{The no. of customer found waiting in queue by the } i^{th} \text{ customer upon arrival} \)

\[ \begin{aligned} i \text{ arrives} & \quad \downarrow \quad \cdots \cdots \quad \uparrow \quad j \text{ in service} \\
\downarrow \quad \text{customers in buffer} & \quad \downarrow \\
(i-1)^{th} \text{ customer} & \quad (i-N_i)^{th} \quad (i-N_i-1)^{th} \text{ customer} \\
\end{aligned} \]

The waiting time for the \( i^{th} \) customer is

\[
W_i = R_i + \sum_{j=i-N_i}^{i-1} X_j
\]

\( N_i, X_{i-1}, \ldots X_{i-N_i} \) are independent

\[
E[W_i] = E[R_i] + E\left\{ \sum_{j=i-N_i}^{i-1} E[X_j | N_j] \right\}
\]

w.r.t \( N_i \) \quad w.r.t \( X_j \)

Note that \( N_i \) is a random variable
Because $X_i$ and $N_i$ are independent

$$E\left\{ \sum_{j=i-N_i}^{i-1} E[X_j|N_j] \right\}$$

$$= E[N_i]E[X_j]$$

$$= E[N_i]\bar{X}$$

$$E[W_i] = E[R_i] + \bar{X}E[N_i]$$

By taking the limit as $t \to \infty$

We have the long term average quantities

$$W = R + \frac{1}{\mu} N_Q$$

Notice that all long term average quantities should be viewed as limits when time or customer index increase to $\infty$

$$\left( \begin{array}{c} W \\ R \\ N_Q \end{array} \right)$$

are limits of the average

waiting time

residual time

no. found

If $\lambda < \mu$ these limits exist
Because the arrival process is Poisson, 
The occupancy distribution upon arrival is 
“typical”
\[
\begin{align*}
\left\{ N_Q \right\} &= \left\{ \text{average no. in queue seen by an} \right\} \\
\Rightarrow \left\{ R \right\} &= \left\{ \text{mean residual time observed} \right\}
\end{align*}
\]

By Little’s Theorem, we have
\[
N_Q = \lambda W
\]
\[
\therefore W = R + \frac{1}{\mu} N_Q , \quad \mu(W - R) = N_Q
\]
\[
W - R = \frac{1}{\mu} W \quad W = R + \rho W
\]
\[
W = \frac{R}{1 - \rho}
\]

where R is not known yet

Denote \( r(\tau) \) is the residual service time

When a new service of duration X begins \( r(\tau) \)
starts at X and decreases linearly for X time units

From assumption (b)
\[
\lim_{t \to \infty} \frac{1}{t} \int_{t}^{t+1} r(\tau) d\tau = \text{The ensemble average of the residual time}
\]
Figure A. Derivation of the mean residual service time. During period \([0, t]\), the time average of the residual service time \(r(\tau)\) is

\[
\frac{1}{t} \int_{0}^{t} r(\tau) \, d\tau = \frac{1}{t} \sum_{i=1}^{M(t)} \frac{1}{2} X_i^2 = \frac{1}{2} \frac{M(t)}{t} \sum_{i=1}^{M(t)} \frac{X_i^2}{M(t)}
\]

where \(X_i\) is the service time of the \(i^{th}\) customer, and \(M(t)\) is the number of service completions in \([0, t]\). Taking the limit as \(t \to \infty\) and equating time and ensemble averages, we obtain the mean residual time \(R = (1/2)\lambda \overline{X^2}\).
consider in the interval \([0,t]\)

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

where \(M(t)\) = no. of service completions in \([0,t]\)

\[
\lim_{t \to \infty} \frac{1}{t} \int_0^t r(\tau) d\tau = \frac{1}{\lim_{t \to \infty} M(t)} \lim_{t \to \infty} \frac{\sum_{i=1}^{M(t)} X_i^2}{M(t)}
\]

- time average of the departure rate
- second moment of the service time
- arrival rate \(\overline{X^2}\)

\[
\Rightarrow R = \frac{1}{2} \lambda \overline{X^2}
\]

\[
W = \frac{R}{1 - \rho} = \frac{\lambda \overline{X^2}}{2(1 - \rho)} \quad \text{,} \quad \rho = \frac{\lambda}{\mu} = \frac{\lambda}{\overline{X}}
\]

It is called Pollaczek-Khinchin (P-K) formula.

The waiting time is function of arrival rate, second moment of service time and load

\[
\text{IF} \quad \overline{X^2} \to \infty \quad \Rightarrow \quad W \to \infty
\]

If service order depends on service time, then

\[
W \neq \frac{\lambda \overline{X^2}}{2(1 - \rho)}
\]
M/G/1 special cases

(a) M/M/1

\[ X^2 = \frac{2}{\mu^2} \]

\[ W_M = \frac{\lambda(2/\mu^2)}{2(1-\rho)} \]

\[ = \frac{\rho}{\mu(1-\rho)} \]

(b) M/D/1  D : deterministic

The service time is identical

\[ X^2 = \frac{1}{\mu^2} \]

\[ \sigma^2 = 0 \]

\[ W_D = \frac{\rho}{2\mu(1-\rho)} \]

Note that

\[ W_D = \frac{1}{2} W_M \]
5.7 Maximum Efficiency of ARQ Protocols

Stop-and-Wait Protocol

Assumptions:

a. unidirectional and the reverse channel is only used for ack
b. fixed length
c. full load
d. no errors

\[ t_{prop} : \text{propagation time} \]
\[ t_f : \text{frame transmitting time} \]
\[ t_{ack} : \text{ack transmitting time} \]
\[ t_{proc} : \text{processing time} \]
\[ t_0 = 2t_{prop} + 2t_{proc} + t_f + t_{ack} \]
\[ = 2t_{prop} + 2t_{proc} + \frac{n_f}{R} + \frac{n_a}{R} \]

where \( n_f \) : No. of bits in a frame
\( n_a \) : No. of bits in an ack

\( R \) : transmission rate (b/s)

The effective info transmission rate

\[ R^o_{eff} = \frac{n_f - n_o}{t_o} \]

\( n_o \) : No. of overhead bits

The transmission efficiency

\[ \eta_o = \frac{n_f - n_o}{t_o} = \frac{n_f - n_o}{R(2t_{prop} + 2t_{proc} + \frac{n_f}{R} + \frac{n_a}{R})} \]
\[ = \frac{1 - \frac{n_o}{n_f}}{1 + \frac{n_a}{n_f} + \frac{2(t_{prop} + t_{proc})}{n_f}} \]
(a) $n_{frame} = 8192$
$n_{overhead} = 64, n_{ack} = 64$

<table>
<thead>
<tr>
<th>$t_{prop} + t_{proc}$</th>
<th>sec</th>
<th>30 kbps</th>
<th>1.5 Mbps</th>
<th>45 Mbps</th>
<th>2.4 Gbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>0.95</td>
<td>0.35</td>
<td>1.77E-02</td>
<td>3.39E-04</td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>0.72</td>
<td>5.14E-02</td>
<td>1.80E-03</td>
<td>3.39E-05</td>
<td></td>
</tr>
<tr>
<td>0.500</td>
<td>0.21</td>
<td>5.39E-03</td>
<td>1.81E-04</td>
<td>3.39E-06</td>
<td></td>
</tr>
</tbody>
</table>

(b) $n_{frame} = 524,288$

<table>
<thead>
<tr>
<th>$t_{prop} + t_{proc}$</th>
<th>sec</th>
<th>30 kbps</th>
<th>1.5 Mbps</th>
<th>45 Mbps</th>
<th>2.4 Gbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>0.99</td>
<td>0.97</td>
<td>0.53</td>
<td>2.14E-02</td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>0.99</td>
<td>0.77</td>
<td>1.04E-01</td>
<td>2.18E-03</td>
<td></td>
</tr>
<tr>
<td>0.500</td>
<td>0.21</td>
<td>0.26</td>
<td>1.15E-02</td>
<td>2.18E-04</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 5.1** Efficiency of Stop-and-Wait ARQ in the absence of errors
consider $n_f = 1024 \text{bytes} = 8192 \text{bits}$, $n_0 = 8 \text{bytes} = 64 \text{bits} = n_{\text{ack}}$

for $t_{\text{prop}} + t_{\text{proc}} = 5 \text{msec} \ (1500 \text{km})$ 光速

<table>
<thead>
<tr>
<th>R</th>
<th>30kb/s</th>
<th>1.5kb/s</th>
<th>45mb/s</th>
<th>2.4Gb/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_0$</td>
<td>0.95</td>
<td>0.35</td>
<td>$1.77 \times 10^{-2}$</td>
<td>$3.39 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

For $t_{\text{prop}} + t_{\text{proc}} = 50 \text{msec} \ (15000 \text{km})$

| $\eta_0$ | 0.72  | $5.14 \times 10^{-2}$ | $1.80 \times 10^{-3}$ | $3.39 \times 10^{-5}$ |

If $n_f = 64000 \text{bytes} = 524288 \text{bits}$

<table>
<thead>
<tr>
<th>$t_{\text{prop}} + t_{\text{proc}} = 5 \text{msec}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
</tr>
<tr>
<td>$\eta_0$</td>
</tr>
</tbody>
</table>

If the channel is noisy, error may occur, let $P$ be the bit error rate and $P_f$ be the frame error probability

$$P_f = 1 - P \ [\text{no bit errors in a frame}]$$

$$= 1 - (1 - P)^{n_f}$$

$$= 1 - (1 - n_f p) \approx n_f p \quad \text{for } P << 1$$

For burst errors $\ p_f \approx \langle n_f \ p \rangle \quad 1/10 << 1/3$

ack 也会有 errors 但它比 frame 短非常多, 所以在计算时不列入

但在下图是考虑 either frame 或 ack errors 的情形计算 through pat
Let $N_t$ be the No. of transmissions to deliver a frame successfully

$$P[N_t = t] = (1 - P_f)P_f^{t-1} \quad \text{for } t = 1, 2, \ldots$$

$t_0 = 2t_{prop} + t_f + 2t_{proc} + t_{ack}$

$t_{out} = \text{time out interval}$

The average total time to transmit a frame

$$E[t_{tot}] = t_0 + \sum_{t=1}^{\infty} (t - 1)t_{out} P[N_t = t]$$

$$= t_0 + t_{out} (1 - P_f) \sum_{t=1}^{\infty} (t - 1)p_f^{t-1}$$

$$= t_0 + \frac{t_{out} p_f}{1 - p_f}$$

Assume $t_{out} = t_0$

$$E[t_{tot}] = \frac{t_0}{1 - p_f}$$
The effective transmission rate
\[ R_{\text{eff}} = \frac{n_f - n_o}{E[t_{\text{tot}}]} = (1 - p_f) \frac{n_f - n_o}{t_o} \]

\[ = (1 - p_f) R^o_{\text{eff}} \]

The maximum efficiency is given by
\[ \eta = \frac{n_f - n_o}{E[t_{\text{tot}}] R} \]

\[ = \frac{n_f - n_o}{t_o} (1 - p_f) \frac{1}{R} \]

\[ = (1 - p_f) \eta_o \]

Consider the Go-back-N ARQ with window size
Ws so the channel is busy always (Ws is large enough)

Assume \( t_{\text{out}} \) is equal to transmit Ws frames
Let $N_t$ be the no. of transmissions required to deliver a frame successfully.

$N_t = t$ there are $(t - 1)$ Ws frames to be retransmitted

$$E[t_{tot}] = t_f \{1 + Ws \sum_{i=1}^{\infty} (i - 1)P[N_t = i]\}$$

$$= t_f \{1 + Ws \frac{p_f}{1 - p_f}\}$$

$$= t_f \{\frac{1 + (Ws - 1) p_f}{1 - p_f}\} = t_f + \frac{Ws \ t_f \ p_f}{1 - p_f}$$

Go-back-N is a pipelined operation

Stop-and-Wait $E[t_{tot}] = t_o + t_{out} \frac{p_f}{1 - p_f}$

二者不相同 若要由Go-back-N導Stop-and-Wait

$t_f$ 要换成 $t_o$ $Ws \ t_f = t_{out}$

The efficiency is

$$\eta = \frac{n_f - n_o}{E[t_{tot}]}$$

$$= (1 - p_f) \frac{1 - \frac{n_o}{n_f}}{1 + (Ws - 1) p_f}$$
For Selective Repeat ARQ, when error occurs only the specific frame is retransmitted

\[ E[t_{\text{tot}}] = t_f \{1 + \sum_{i=1}^{\infty} (i-1)(1 - p_f)p_f^{i-1}\} \]

\[ = t_f \{1 + \frac{p_f}{1 - p_f}\} \]

\[ = t_f \frac{1}{1 - p_f} \]

\[ R_{\text{eff}} = (1 - p_f)(1 - \frac{n_o}{n_f}) R \]

The efficiency is

\[ \eta = (1 - p_f)(1 - \frac{n_o}{n_f}) \]

Consider \( R = 1.5 \text{Mb/s} \)  \( t_{\text{prop}} = 5 \text{m sec} \)  \( n_f = 1024 \text{bytes} \)

\( Ws = 4 \)  \( n_o = 4 \text{bytes} \)
\[ n_f = 1024 \text{ bytes} \quad n_o = n_a = 4 \text{ bytes} \]

\[ R = 1.5 \text{ Mb/s} \]

\[ t_{prop} = 5 \text{ ms} \]

\[ W_s = 4 \]

\[ \eta \]

\[ p \]
<table>
<thead>
<tr>
<th>ARQ technique</th>
<th>Efficiency</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop-and-Wait</td>
<td>$\eta = \frac{n_f - n_0}{E[t_{total}]} \cdot \frac{1 - n_0}{n_f} \cdot \frac{1}{1 + \frac{n_a}{n_f} + \frac{2(t_{prop} + t_{proc})}{n_f} R} = (1 - P_f)n_0$</td>
<td>Delay-bandwidth product is main factor</td>
</tr>
<tr>
<td>Go-Back-N</td>
<td>$\eta = \frac{n_f - n_0}{E[t_{total}]} = (1 - P_f) \cdot \frac{1 - n_0}{n_f}$</td>
<td>Average wasted time: $(W_S - 1)P_f$</td>
</tr>
<tr>
<td>Selective-Reject</td>
<td>$\eta = (1 - P_f) \left(1 - \frac{n_0}{n_f}\right)$</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 5.2 Summary of performance results**
Optimal frame size (given p)

Recall \( p_f \approx n_f p \)

Stop-and-Wait

\[
\eta = (1 - p_f) \frac{1 - \frac{n_o}{n_f}}{1 + \frac{n_a}{n_f} + \frac{2(t_{prop} + t_{proc})R}{n_f}}
\]

Go-back-N

\[
\eta = (1 - n_f p) \frac{n_f - n_o}{n_f + (Ws - 1)n_f^2 p}
\]

Selective Repeat

\[
\eta = (1 - n_f p)(1 - \frac{n_o}{n_f})
\]
\[ P = 10^4 \]

\[ R = 1.5 \text{Mb/s} \]

\[ n_o = n_a = 4 \text{bytes} \]
5.8 Example, Delay Analysis of go-back-n ARQ (special case)

Assumptions:
1. Packet length = 1 time unit
2. Maximum waiting time = $t_{out} = n-1$ units
3. Error may occur only on the forward channel
4. $P = \text{probability of packets in error}$
5. Poisson arrival with rate $\lambda$

Let $X$ be the no. of transmissions for a success packet

$\text{Prob } \{ X = k \text{ retransmissions and } 1 \text{ success} \} = \text{Prob } \{ X = 1 + nk \}$

\[ = (1 - P)P^k \quad k = 0,1,2... \]

It can be modelled as an $M/G/1$ System
Illustration of the effective service times of packets in the ARQ system of Example 3.15. For example, packet 2 has an effective service time of \( n + 1 \) because there was an error in the first attempt to transmit it following the last transmission of packet 1, but no error in the second attempt.
The mean no of transmissions

\[ \bar{X} = \sum_{k=0}^{\infty} (1 + kn)(1 - P)P^k \]

\[ = (1 - p)(\sum_{k=0}^{\infty} P^k + n\sum_{k=0}^{\infty} kP^k) \]

\[ = (1 - P)[\frac{1}{1 - P} + \frac{nP}{(1 - P)^2}] \]

\[ = 1 + \frac{nP}{1 - P} \]

The second moment

\[ \bar{X}^2 = \sum_{k=0}^{\infty} (1 + kn)^2 (1 - P)P^k \]

\[ = (1 - P)[\sum_{k=0}^{\infty} P^k + 2n\sum_{k=0}^{\infty} kP^k + n^2\sum_{k=0}^{\infty} k^2 P^k] \]

\[ = 1 + \frac{2nP}{1 - P} + \frac{n^2P(1 + P)}{(1 - P)^2} \]

From P-K Formula \[ W = \frac{\lambda \bar{X}^2}{2(1 - \rho)} \]

\[ W = \frac{\lambda \bar{X}^2}{2(1 - \lambda \bar{X})} = \frac{\lambda \bar{X}^2}{2(1 - \rho)} \]

\[ T = \bar{X} + W \]
5.9 Roll-call Polling

Stations are interrogated sequentially, one by one, by the central system, which asks if they have any messages to transmit.

\[ \begin{align*}
W_i & : \text{walk time} \\
T_i & : \text{frame transmission time}
\end{align*} \]
The scan or cycle time $t_c$ is given by

$$t_c = \sum_{i=1}^{N} \bar{w}_i + \sum_{i=1}^{N} \bar{t}_i$$

The average scan time $\bar{t}_c$

$$\bar{t}_c = \sum_{i=1}^{N} \bar{w}_i + \sum_{i=1}^{N} \bar{t}_i = L + \sum_{i=1}^{N} \bar{t}_i$$

$\bar{w}_i$, $\bar{t}_i$ are the ave. walk time and the ave. time to transmit pkt at station $i$. $L$ is the total walk time of the complete poling system.
For station $i$, let

- $\lambda_i$: the ave. pkt arrival rate
- $\overline{\ell}$: the ave. packet length
- $\ell'$: the number of overhead bits
- $C$: the channel capacity in bps
- $\overline{m}_i$: the ave. frame length in time $\overline{m}_i = (\overline{\ell} + \ell')/C$

The average number of packets waiting to be transmitted when station $i$ is polled is $\lambda_i \overline{t_c}$, the time required to transmit is

$$t_i = \lambda_i \overline{t_c} \overline{m}_i = \rho_i \overline{t_c}$$

With $\rho_i \equiv \lambda_i \overline{m}_i$ the traffic intensity, the average scan time $\overline{t_c}$ is given

$$\overline{t_c} = L / \left(1 - \sum_{i=1}^{N} \rho_i\right) = L/(1-\rho)$$

With $\rho = \sum_{i=1}^{N} \rho_i = \sum_{i=1}^{N} \lambda_i \overline{m}_i$ representing the total traffic intensity on the common channel.

$L = \sum_{i=1}^{N} \overline{w}_i$
For small $\rho$ the average access delay should be $\bar{t}_c / 2$.

Assume that each station has the same $\lambda$, same frame-length statistics, and the same $\bar{W}$.

The average access delay is

$$E(D) = \frac{\bar{t}_c}{2} \left( 1 - \frac{\rho}{N} \right) + \frac{N\lambda m^2}{2(1 - \rho)}$$

$$= \frac{L}{2} \frac{(1 - \rho/N)}{(1 - \rho)} + \frac{N\lambda m^2}{2(1 - \rho)}$$

$m^2$ is the second moment of the frame length, $\rho = N\lambda \bar{m}$.

The access delay is the average time a packet must wait at a station from the time it first arrives until the time transmission begins. Access delay is thus the average wait time in an M/G/1 queue.

5.10 Hub Polling

Control is passed sequentially from one station to another.

Let the polling message be a fixed value, $t_p$ sec in length. The time required per station to synchronize to a polling message is $t_s$ sec. The total propagation delay for the entire $N$-station system is $\tau'$ sec.
Hub Polling (Cont’)

- The total walk time for hub-call polling is \( L = N t_p + N t_s + \tau' \).

- Let the stations all be equally spaced, and the round-trip propagation delay between the controller and station \( N \) be \( \tau \) sec.

- The overall propagation delay \( \tau' \) is just

\[
\tau' = \frac{\tau}{2} (1 + N)
\]

- The analysis of the hub-polling strategy is identical to that of roll-call polling.

- The only difference is that the walk time \( L \) is reduced through the use of hub polling.

- For hub polling, \( L_{hub} = \tau + N t_s \).
5.11 Throughput Analysis of Nonpersistent CSMA.

5.12 Throughput Analysis of Slotted Nonpersistent CSMA

5.13 Stability of Slotted ALOHA
5.11 Performance Analysis of Nonpersistent CSMA

- Definitions
  - Total load = G
  - Vulnerable period = a
  - Packet transmission time = 1
  - Busy period = B
  - Idle period = I
  - Successful utilization period = U
  - The arrival time of the last packet collides with the first one = Y
  - a = Propagation delay packet transmission time
  - One cycle = B + I
  - The successful transmission period = 1 + a
  - The collision period = 1 + Y + a
FIGURE 4.5
Successful and unsuccessful transmission attempts for nonpersistent CSMA. Time is measured in units of the packet transmission time $t_p$. 
The throughput $s = \frac{E[u]}{E[B] + E[I]}$

Note that if no station transmits duration the first a time unit of the busy period, the transmission is successful.

$E[u] = e^{-aG}$

The idle period is just the time interval between the end of a busy period and the next arrival.

Because the packet arrivals follow a Poisson distribution

$P(k) = \frac{G^k e^{-G}}{k!}$
The average duration of an idle period

\[ E[I] = \frac{1}{G} \quad B = 1 + a + Y \]

Where \( y = 0 \) for successful transmissions

\[ E[B] = 1 + a + E[Y] \]

The probability density function of \( Y \) is the probability that no packet arrival occurs in an interval of length \( a-y \).

\[ f(y) = Ge^{-G(a-y)} \quad \text{for} \quad 0 \leq y \leq a \]

\[ E[Y] = \int_0^a y f(y) dy \]

\[ = \int_0^a y Ge^{-G(a-y)} dy \]

\[ = Ge^{-Ga} \int_0^a ye^{Gy} dy \]

\[ = a - \frac{1}{G} \left( 1 - e^{-aG} \right) \]

\[ E[B] = 1 + 2a - \frac{1}{G} \left( 1 - e^{-aG} \right) \]

\[ s = \frac{Ge^{-aG}}{G(1 + 2a) + e^{-aG}} \]
FIGURE 4.6
Throughput $S$ as a function of the offered load $G$ for various values of $a$ for nonslotted nonpersistent CSMA.
5.12 Throughput Analysis of Slotted Nonpersistent CSMA

- Slot time = $\tau$

All stations are synchronized and required to start transmission only at the beginning of a slot. All packets are assumed to be of a length $t_p$, which is an integral number of time slots.

When a packet arrives during a time slot, the station senses the state of the channel at the beginning of the next slot then either transmits if the channel is idle or defers to a later time slot if traffic is present.
Example: Packets station 1,2,n arrive during the first slot.
If the next slot is empty, three stations transmit.
If packets arrive in the busy period ($\alpha$, $\beta$, $\gamma$, $\delta$) the stations sense the channel to be busy.
These stations defer their packets for transmission at a later time.
The busy period $= t_p + \tau$

The normalized busy period $- \frac{(t_p + \tau)}{t_p} = (1 + a)$ where $a = \frac{\tau}{t_p}$

Define the average time during a cycle that a transmission is successful

$E[u] = P_s$

$= P[\text{one packet arrives is slot a | some arrival occurs}]$

$= \frac{P[\text{one packet arrives in slot a and some arrival occurs}]}{P[\text{some arrival occurs}]}$

$= \frac{P[\text{one packet arrives in slot a}]}{P[\text{some arrival occurs}]}$
Using Poisson arrival statistics, we have

\[ P[\text{one packet arrives in slot } a] = aGe^{-aG} \]

and \[ P[\text{some arrival occurs}] = 1 - e^{-aG} \]

\[ E[u] = \frac{aGe^{-aG}}{1-e^{-aG}} \]

The busy period is always 1+a

\[ E[B] = 1 + a \]

An idle period always consists of an integral number of time slots

\[ I \geq 0 \]
If a packet arrives during the last slot of a busy period, then the next slot immediately starts a new busy period, so that $I=0$.

When there are no arrivals during the last slot of a busy period then the next $I-1$ slots will be empty until there is an arrival in the final $I^{th}$ slot. This then marks the beginning of a new busy period.

Consider the case $I=0$.

The probability $p$ of this occurring is merely the probability of some packet arriving in the interval $a$:

$$P[I=0] = p = 1 - e^{-aG}$$

Next we look at the case $I=1$.

$P[I=1] = $ the joint probability that no arrival occurs in the last slot of the busy period and that some arrival occurs in the next time slot.

$$= (1-p)p$$
\[ P[I = i] = (1 - p)^i p \]

\[ E[I] = a \sum_{i=0}^{\infty} i(1 - p)^i p \]

\[ = a(1 - p) \frac{p}{1 - e^{-aG}} \]

\[ = \frac{ae^{-aG}}{1 - e^{-aG}} \]

\[ s = \frac{E[u]}{E[B] + E[I]} \]

\[ = \frac{aGe^{-aG}}{1 + a + \frac{ae^{-aG}}{1 - e^{-aG}}} \]

\[ = \frac{aGe^{-aG}}{(1 + a)(1 - e^{-aG}) + ae^{-aG}} \]

\[ = \frac{aGe^{-aG}}{1 - e^{-aG} + a} \]
5.13 Stability of Slotted ALOHA

Assumptions

a. There are N identical stations, each of which can hold at most one packet.

b. If a new packet arrives at a station with a full buffer, the newly arriving packet is lost. (backlogged)

c. The station generates new packets according to a Poisson distribution with an average rate $p$ packets/slot.

d. A previously collided packet is retransmitted with probability $\alpha$ in every succeeding time slot until a successful transmission occurs.
e. Let $k$ be the number of stations which are backlogged (Each of this stations decides to resend its backlogged packet with probability $\alpha$ or to skip the present slot with probability $1-\alpha$.

f. Let $r$ represent the number of retransmission attempts by the $k$ stations during the same slot.

g. Let $n$ be the number of new packets generated by the $N-k$ unblocked stations.

The probability $P_k$ of a successful transmission is

$$P_k = P[r = 1]P[n = 0] + P[r = 0]P[n = 1]$$

$$= [k\alpha(1-\alpha)^{k-1}][(1-P)^{N-k}] + [(1-\alpha)^k][(N-k)P(1-P)^{N-k-1}]$$

$$= S_{out,k}$$
Where \( P[n=0] \) is the probability that no new packets are generated, \( P[n=1] \) is the probability that a new packet is generated.

\( S_{\text{out},p} \) is the rate at which packets leave the system.

The rate at which packets enter the system is

\[ s = (N-k)p \]

For a large \( N \) and small \( p \)

\[ S_{\text{out},k} \approx k\alpha(1-\alpha)^{k-1}e^{-s} + (1-\alpha)^k se^{-s} \]

\[ e^{-s} = 1 - s - s^2/2! \approx 1 - s = 1 - (N-k)p \]

In equilibrium

\[ s = S_{\text{out},p} \]
FIGURE 4.15
Equilibrium throughput curve as a function of the number of backlogged stations in slotted ALOHA; at $k = 0$, $S_{out} = 0.164$. ($\alpha = 0.03$, $N = 200$, $p = 0.001$)
Equilibrium contours for various values of the retransmission probability $\alpha$ for slotted ALOHA. All curves go to 0.164 at $k = 0$. 

$N = 200$
$p = 0.001$
FIGURE 4.17
Example of a channel load line showing the input rate as a function of backlog.

\[ S = (N - k)p \]

\[ N = 200 \]
\[ p = 0.001 \]
Example of a globally stable, lightly loaded system.

\[ \alpha = 0.035 \]
\[ N = 150 \]
\[ p = 0.001 \]
FIGURE 4.19
Bistable system having two locally stable and one unstable equilibrium points.
FIGURE 4.20
Behavior of systems which are (a) stable but overloaded ($N$ and $p$ are large); (b) unstable with an infinite number of users.