6.1 Introduction

In the preceding four chapters we have discussed single-server queueing systems, though in Chapter 2 a multi-server queueing system has been analyzed. Multichannel queues constitute an important class of queueing processes and have wide applicability in real-life situations. Examples of such systems may be found in banks, telephone exchanges, hospital admission systems, seaports, toll booths, military tactics, courts of law (wherein servers are judges and customers correspond to cases) and information transmission systems (wherein messages containing a small random number of characters or a batch of characters arrive according to a Poisson process and must be transmitted to some destination).

Some situations arise in which the number of channels is so large (infinite for all practical purposes) that waiting times are negligible and only the level of server occupancy is of analytical or practical interest. An intensive care unit in a hospital, where delay in service may result in loss of a life, may provide an example of such a situation. Many service systems, where the number of service facilities arranged in parallel is large in comparison with the demand made on the system from time to time,
behave like infinite-server systems. A typical example is a self-service system. Other potential applications are large parking lots, theaters, or auditoriums.

Infinite server service systems can be categorized into two classes viz. homogeneous and non-homogeneous ones. Non-homogeneity may be introduced in the queueing system by several ways such as allowing the arrival rate or service rate or both to be time dependent, assigning to each customer a service time distribution (STD) that depend on the time at which the customer arrives etc. In this chapter, we shall discuss such a non-homogeneous infinite server system.

Non-homogeneous service systems with infinitely many servers have been studied at great length by many authors. Collings and Stoneman (1976) discussed an $M/M/\infty$ service system with time dependent arrival rate $\lambda(t)$ and time dependent service rate $\mu(t)$. For the infinite server queue Saaty (1961) shows that if the departure rate is constant, the resulting queue size distribution is always Poisson. Collings and Stoneman showed that the same result holds for time dependent departure rate if the initial queue size distribution is Poisson and used this model to describe the daily fluctuations in emergency health services. Kambo and Bhalaik (1982) studied the $M/M/\infty$ queueing model of Collings and Stoneman. They derived the generating function of the joint probability distribution of $N_1(t)$ and $N_2(t)$ where $N_1(t)$ is the number of units in service at time $t$ and $N_2(t)$ is the number of departures.
upto time $t$ and, from this, expressions for mean, variance and covariance of $N_1(t)$ and $N_2(t)$ were obtained. They proved that $N_1(t)$ and $N_2(t)$ are independent Poisson random variables provided at time $t = 0$, the system is empty and has no departures. Ramakrishnan (1980) proved the same result obtained by Kambo and Bhalaik for the M/D/∞ queue by a simple argument based on properties of the Poisson process. By generalizing Ramakrishnan's argument, Foley (1982) extended the result to the non-homogeneous M/G/∞ queue. In addition, he showed that the departure process in the interval $(0, t]$ is a Poisson process which is independent of the queue length at time $t$ and obtained the time-dependent queue-length distribution. These results have also been extended to tandem infinite server queueing systems in which the arrival process is a non-homogeneous Poisson process and the service time of a customer at the $n$th queue may depend on service time of all the previous customers as well as the time of his arrival to the system. Brown and Ross (1969) obtained some of the results of Foley by using different arguments. Reynolds (1968) analyzed a bulk arrival infinite server homogeneous Poisson queue where he showed that the limiting distribution of queue size is a Compound Poisson distribution. Later Holman et al. (1983) discussed the homogeneous service system $M^X/G/∞$ characterized by an infinite number of servers and a general STD. Using the standard properties of the Poisson process and elementary probability arguments they developed the PGF of the transient and steady-state
number of busy servers and several other connected properties. Expressions have been obtained for the means, variances of \( N_1(t) \) and \( N_2(t) \) and it has been shown that \( N_1(t) \) and \( N_2(t) \) are uncorrelated if and only if either the arrivals are in singlet's or else the STD is deterministic. Shanbhag (1965) studied the system \( M(X)_t) / G/\infty \) where the arrival rate is time-dependent. He obtained expressions for the joint Laplace Transform of several measures of the system, including the PGF of the number of busy servers in the transient state and deduced the steady state values therefrom. He derived the expected duration of busy period assuming constant arrival rate. The system \( M(X) / t)/ M/\infty \) with the input rate \( \lambda(t) \) has been investigated by Abolnikov (1968) to obtain transient behaviour of \( N_1(t) \). Kashyap and Chaudhry (1983) considered the system \( M(X) / (t) / M(t)/ \infty \) where they have shown that \( N_1(t) \) and \( N_2(t) \) are not independent, they turn out to be independent only in the case \( M(X) / D/\infty \). Mehata and Selvam (1982) considered the queueing system studied by Foley, in which units arrive in a non-homogeneous Poisson process and a service time distribution is assigned to each arriving unit at time \( t \). They presented a detailed account on the covariance structure of this queueing model. This type of study is quite relevant in optimal parameter estimation of the model.

In this chapter an infinite server queueing system is studied in which units arrive in batches of variable size following a non-homogeneous compound Poisson process and each arriving unit
has its own arbitrary STD. Using the standard properties of a Poisson process and elementary probability arguments, we derive the PGF of (i) the number of units in the system at time \( t \), (ii) the number of service completions up to time \( t \) and (iii) the number of arrivals up to time \( t \). Covariance functions between any two of these three r.v.'s are obtained separately. Finally, some steady state results are obtained under certain conditions. This study extends the work contained in Holman et al. (1983), Kashyap and Chaudhry (1983) and Mehata and Selvam (1982).

6.2 Description of the Model

Let units arrive in batches at the infinite server system \( M^X/G/\infty \) according to a non-homogeneous Poisson process with parameter

\[
\Lambda(t) = \int_0^t \Lambda(a) \, da,
\]

where \( \Lambda(t) \) is the mean rate (function) of arrival of a batch at time \( t \). The batch size, \( X \), is a r.v. with probability mass function (PMF)

\[
P(X = m) = c_m \quad (m = 1, 2, \ldots)
\]

Let the PGF of \( X \) be \( C(z) \) so that

\[
C(z) = E(z^X) = \sum_{m=1}^{\infty} c_m z^m
\]
\[ C^{(1)}(z) = \sum_{m=1}^{\infty} m c_m = \bar{c}, \text{ the mean batch size} \]

and

\[ C^{(k)}(z)/k! = \sum_{m=k}^{\infty} \binom{m}{k} c_m z^{m-k} \]

where \( C^{(k)}(z) \) denotes the \( k \)th derivative of \( C(z) \). Mean arrival rate of units in the system is \( \lambda(t)\bar{c} \). Units enter service immediately on arrival. Let \( G_t(x) \) be the STD of a unit which arrives at time \( t \) and \( g_t(x) \) denote the corresponding density function. We assume that at time \( t = 0 \) the system is empty. Let us define the following r.v.'s:

\[ \begin{align*}
N(t) &= \text{number of units in the system at time } t \\
D(t) &= \text{number of units that complete service during } (0, t] \\
R(t) &= \text{number of units that arrive during } (0, t] \\
Y_m(t) &= \text{number in the system at time } t \text{ from } n \text{ batches of size } m \text{ that arrive during } (0, t] \quad n = 0, 1, 2, \ldots
\end{align*} \]

Then the PMF of \( R(t) \) is given by

\[ \text{Prob}[R(t) = k] = \sum_{n=0}^{\infty} \exp\left[ -\Lambda(t) \right] \frac{\left[ \Lambda(t) \right]_n}{n!} \{c_k\}^{n^*} \quad k = 0, 1, 2, \ldots \quad (6.1) \]

where \( \{c_k\}^{n^*} \) is defined in (4.2).

The PDF of \( R(t) \) on using (6.1) is

\[ P_{R(t), z, t} = \sum_{k=0}^{\infty} \text{Prob}[R(t) = k] z^k, \quad |z| < 1 \quad (6.2 \text{ contd.}) \]
We have

\[ N(t) = \sum_{n=0}^{\infty} \sum_{m=1}^{n} Y_m(t) \]

= number in the system at time \( t \) - out of the units that arrive during \((0, t] \)

Clearly, \( \text{Prob}\left[ Y_m^0(t) = 0 \right] = 1 \) \hspace{1cm} (6.4)

6.3 **Transient Distributions of Various Functions Involved in the System**

We observe that

\[ Y_m^n(t) = \text{sum of } n \text{ i.i.d. r.v.'s} \]

each distributed as \( Y_m^1(t) \).

Now for the conditional distribution of \( Y_m^1(t) \) we note that it takes the values 0, 1, 2, ..., \( m \) with a binomial probability distribution. Specifically,

\[
\text{Prob}\left[ Y_m^1(t) = k \mid \text{there is one arrival epoch 'a' in } (0, t] \right] = \binom{m}{k} \{\alpha_a(t)\}^k \{\beta_a(t)\}^{m-k} \quad (6.5)
\]

where
\( a_a(t) = \text{Prob}[\text{Service time } T > t-a] \)

\[ = \int_{x=t-a}^{\infty} g_a(x) \, dx \quad (6.6) \]

and

\( \beta_a(t) = \text{Prob}[\text{Service time } T \leq t-a] \)

\[ = \int_{x=0}^{t-a} g_a(x) \, dx \quad (6.7) \]

It can be noted that if \( G_t(.) \) is a true distribution, then

\( \beta_a(t) = 1 - a(a) \) in (6.7). From (6.5) PGF of conditional distribution of \( Y_m(t) \) is given by:

\[ E\left[ z^{Y_m(t)} \mid \text{there is one arrival epoch 'a' in } (0,t] \right] = \left[ z a_a(t) + \beta_a(t) \right]^m \quad (6.8) \]

Now, arrivals of batches of size \( m \) form a non-homogeneous Poisson process with mean arrival rate \( \lambda(t) c_m \). Thus the probability that the arrival time (of a batch) lies in \((a,a+da)\) conditioned on the batch size given that there is an arrival epoch 'a' in \((0,t]\) is

\[ \frac{\lambda(a) c_m \, da}{t} = \frac{\lambda(a)}{\Lambda(t)} \, da \]

Hence unconditioning on 'a' we have from (6.8)
\[ E \left[ z^m(t) \right| \text{there is one arrival epoch in } (o, t] \right] = \int_o^t \frac{\lambda(s)}{\Lambda(t)} \left[ z \alpha_a(t) + \beta_a(t) \right]^m \, da , \]

whence we have

\[ E \left[ z^n_m(t) \right| \text{n. arrival epochs in } (o, t] \right] = \left[ \int_o^t \frac{\lambda(s)}{\Lambda(t)} \left[ z \alpha_a(t) + \beta_a(t) \right]^m \, da \right]^n \quad (6.9) \]

Now from (6.9) unconditioning on \( n \) and using (6.3) and (6.4) we have the PGF of \( N(t) \),

\[ P_N(t) = E \left[ z^{N(t)} \right|, \quad |z| \leq 1 \]

\[ = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left[ \int_o^t \frac{\lambda(s)}{\Lambda(t)} \left[ z \alpha_a(t) + \beta_a(t) \right]^m \, da \right]^n \]

\[ \times \exp \left[ -\Lambda(t) \frac{(\Lambda(t)c_m^n)}{n} \right] \]

\[ = \exp \left[ t \lambda(s) \left[ C(z \alpha_a(t) + \beta_a(t)) - 1 \right] \, da \right] \quad (6.10) \]

Following the same procedure to obtain (6.10), the PGF of \( D(t) \) is given by

\[ P_D(t) = E \left[ z^{D(t)} \right|, \quad |z| \leq 1 \quad (6.11 \text{ contd.}) \]
In this section we assume that $G(x)$ is a true distribution. From (6.10) expected number of busy servers at time $t$, $L_1(t)$ is given by

$$L_1(t) = \frac{2}{2} \sum_{n=0}^{\infty} \frac{1}{\Lambda(t)} \left[ \sum_{n=0}^{\infty} \frac{\Lambda(t) \lambda^n}{n!} \right] da$$

and variance of the number of busy servers at time $t$ is

$$\text{Var}[N(t)] = \frac{2}{2} \sum_{n=0}^{\infty} \frac{1}{\Lambda(t)} \left[ \sum_{n=0}^{\infty} \frac{\Lambda(t) \lambda^n}{n!} \right]^2 da$$

Similarly, from (6.11), we obtain the mean and variance of $D(t)$, which are respectively given by
\[ L_p(t) = \frac{c}{A(a)} \int_o^t \lambda(a) \beta_a(t) \, da \]  
\[ \text{and} \]
\[ \text{Var} \left[ D(t) \right] = \frac{c}{A(a)} \int_o^t \lambda(a) \beta_a(t) \, da + C(2)(1) \int_o^t \lambda(a) \left[ \beta_a(t) \right]^2 \, da \]  
\hspace{1cm} (6.14) \hspace{1cm} (6.15)

From (6.2) variance of \( R(t) \) is obtained as
\[ \text{Var} \left[ R(t) \right] = \Lambda(t) \left[ \frac{c}{A(a)} \int_o^t \lambda(a) \beta_a(t) \, da \right] + C(2)(1) \]  
\[ \hspace{1cm} (6.16) \]

To evaluate the covariance functions between \( N(t) \) and \( D(t) \), we note that
\[ R(t) = N(t) + D(t) \]  
\[ \text{Var} \left[ R(t) \right] = \text{Var} \left[ N(t) \right] + \text{Var} \left[ D(t) \right] + 2 \text{Cov} \left[ N(t), D(t) \right] \]  
\[ \hspace{1cm} (6.17) \hspace{1cm} (6.18) \]

where \( \text{Cov} \left[ N(t), D(t) \right] \) is the covariance function of \( N(t) \) and \( D(t) \). Using \( \Lambda(t) = \int_o^t \lambda(a) \, da \) and the results (6.13), (6.15), (6.16) and (6.18) we have
\[ \text{Cov} \left[ N(t), D(t) \right] = \frac{1}{2} \left[ \text{Var} \left[ R(t) \right] - \text{Var} \left[ N(t) \right] - \text{Var} \left[ D(t) \right] \right] \]
\[ = \frac{1}{2} \left[ \frac{c}{A(a)} \int_o^t \lambda(a) \left[ 1 - (\alpha_a(t) + \beta_a(t)) \right] \, da \right] + C(2)(1) \int_o^t \lambda(a) \left[ 1 - \left( \alpha_a(t) \right)^2 + \left( \beta_a(t) \right)^2 \right] \, da \]
\[ = C(2)(1) \int_o^t \lambda(a) \alpha_a(t) \left[ 1 - \alpha(t) \right] \, da \]  
\hspace{1cm} (since \( \beta_a(t) = 1 - \alpha_a(t) \))
\[= C(2) \int_0^t \lambda(a) G_a(l-a) \left[1 - G_a(t-a)\right] \, da \quad (6.19)\]

From (6.17)

\[\text{Cov} \left[R(t), N(t)\right] = \text{Var} \left[N(t)\right] + \text{Cov} \left[N(t), D(t)\right] \]

\[= \left[\bar{c} + C(2)(1)\right] \int_0^t \lambda(a) \alpha_a(t) \, da \quad (6.20a)\]

\[= \left[\bar{c} + C(2)(1)\right] \int_0^t \lambda(a) \left[1 - G_a(t-a)\right] \, da \quad (6.20b)\]

\[\text{Cov} \left[R(t), D(t)\right] = \text{Var} \left[D(t)\right] \cdot \text{Cov} \left[N(t), D(t)\right] \]

\[= \left[\bar{c} + C(2)(l)\right] \int_0^t \lambda(a) \beta_a(t) \, da \quad (6.21a)\]

\[= \left[\bar{c} + C(2)(1)\right] \int_0^t \lambda(a) \cdot G_a(t-a) \, da \quad (6.21b)\]

Since \(\lambda(a) \neq 0\) for at least one epoch 'a' in \((0,t)\)

\[\text{Cov} \left[N(t), D(t)\right] = 0 \text{ implies and is implied by} \]

either \(C(2)(1) = 0\) i.e. arrivals are one by one or,

\[G_a(t-a) \left[1 - G_a(t-a)\right] = 0\]

The second alternative i.e.

either \(G_a(t-a) = 0\) or \(1 - G_a(t-a) = 0\)

for \(a \in (0,t)\) corresponds to a degenerate distribution with a
constant service time $T$, say. Thus it is observed that $N(t)$, $D(t)$ are uncorrelated if and only if either arrivals occur one at a time or the STD is degenerate.

6.5 Some Steady State Results

Foley (1982) has shown by setting a counter example that steady state distributions for a non-homogeneous $M/G/\infty$ system do not always exist even if $\lambda(t) \to \lambda$ as $t \to \infty$. He obtained under certain conditions the steady state result for the number of units in the system. It can be easily seen that his conditions hold equally good for the bulk arrival system. Following the same conditions on the arrival process and STD viz.

(i) $\Lambda(t) = \int_0^t \lambda(s) \, ds < \infty$ for every $t > 0$

(ii) $\lim_{t \to \infty} \lambda(t) = \lambda$

(iii) $\lim_{t \to \infty} \int_0^t |G_a(x) - G(x)| \, dx = 0$

where $G(x)$ is assumed to be a distribution function with finite mean $1/\mu$, we obtain easily the steady state values of means, variances and covariance functions derived in section (6.4).

Expected values of $N(t)$ and $D(t)$ in the steady state on using (6.12), (6.14) and (6.22) are given by
\[ L_1 = \lim_{t \to \infty} L_1(t) = \lambda \bar{c} E(S) \]  
(6.23)

\[ L_2 = \lim_{t \to \infty} L_2(t) = \lambda c \int_0^\infty G(x) \, dx \]  
(6.24)

where \( E(S) = \int_0^\infty [1 - G(x)] \, dx \) is the expected service time of a unit.

Use of (6.13), (6.15), (6.16), (6.19)-(6.21) in the limit results in

\[ \text{Var} N = \lim_{t \to \infty} \text{Var}[N(t)] = \lambda \left[ \bar{c} E(S) + C(2)(1) \int_0^\infty [1 - G(x)]^2 \, dx \right] \]  
(6.25)

\[ \text{Var} D = \lim_{t \to \infty} \text{Var}[D(t)] = \lambda \left[ \bar{c} \int_0^\infty G(x) \, dx + C(2)(1) \int_0^\infty [G(x)]^2 \, dx \right] \]  
(6.26)

\[ \text{Cov}[N,D] = \lim_{t \to \infty} \text{Cov}[N(t),D(t)] = C(2)(1) \int_0^\infty G(x) [1 - G(x)] \, dx \]  
(6.27)

\[ \text{Cov}[R,N] = \lim_{t \to \infty} \text{Cov}[R(t),N(t)] = \lambda \left[ \bar{c} + C(2)(1) \right] E(S) \]  
(6.28)
Cov \[ R, D \] = \lim_{t \to \infty} \text{Cov} [R(t), D(t)]

= \lambda \left[ c + C^{(2)}(1) \right] \int_0^\infty G(x) \, dx \quad (6.29)

6.6 Some Particular Cases

When STD is degenerate we have from (6.20b), the following covariance function of \( R(t) \) and \( N(t) \):

**Case 1.** \( t < T \quad \Rightarrow \quad t-a < T \Rightarrow G_a(t-a) = 0 \)

Hence \( \int_0^t \xi(a) [1 - G_a(t-a)] \, da = \Lambda(t) \)

**Case 2.** \( t > T \). Then \( \int_0^t \xi(a) [1 - G_a(t-a)] \, da \)

\[
= \int_0^{t-T} \xi(a) [1 - G_a(t-a)] \, da + \int_{t-T}^t \xi(a) [1 - G_a(t-a)] \, da
\]

Now \( a < t-T \Rightarrow t-a > T \Rightarrow G_a(t-a) = 1 \)

and \( a > t-T \Rightarrow t-a < T \Rightarrow G_a(t-a) = 0 \)

whence we have

\[
\int_0^t \xi(a) [1 - G_a(t-a)] \, da = \int_{t-T}^t \xi(a) \, da = \Lambda(t) - \Lambda(t-T)
\]

Combining these two cases, we have

\[
\text{Cov} [R(t), N(t)] = \left[ c + C^{(2)}(1) \right] \{\Lambda(t) - \Lambda(t-T)H(t-T)\} \quad (6.30)
\]
By similar arguments we have from (6.21b)

\[ \text{Cov} \left[ R(t), D(t) \right] = \left[ \bar{c} + C(2)(1) \right] \Delta(t-T) H(t-T) \] (6.31)

where \( H(.) \) denotes Heaviside function.

When \( t < T \) we have from (6.12)

\[ \text{L}_1(t) = \bar{c} \left( t \right) \]

and when \( t > T \), \( \text{L}_1(t) = \bar{c} \left[ \Delta(t) - \Delta(t-T) \right] \)

and hence \( \text{L}_1(t) = \bar{c} \left[ \Delta(t) - \Delta(t-T) H(t-T) \right] \) (6.32)

Similarly, from (6.14) we have

\[ \text{L}_2(t) = \bar{c} \Delta(t-T) H(t-T) \] (6.33)

\[ \frac{M(X)}{M(t)/\infty} \]

For the non-homogeneous \( \frac{M(X)}{M/\infty} \) queue with

\[ G_t(x) = 1 - \exp \left[ \int_{t}^{t+x} \mu(y) \, dy \right] \]

We have from (6.6) and (6.7)

\[ a_a(t) = \int_{a}^{\infty} \mu(a+x) \exp \left[ \int_{a}^{a+x} \mu(y) \, dy \right] \, dx \]

\[ = \exp \left[ \int_{a}^{t} \mu(y) \, dy \right] - \exp \left[ \int_{a}^{\infty} \mu(y) \, dy \right] \] (6.34)

and
\[
\beta_a(t) = \int_{x=0}^{t-a} \mu(a+x) \exp \left[ - \int_{a}^{t} \mu(y) \, dy \right] \, dx
\]
\[
= 1 - \exp \left[ - \int_{a}^{t} \mu(y) \, dy \right]
\]

(6.35)

Results (6.20a) and (6.21a) for \( \alpha_a(t) \), \( \beta_a(t) \) given by (6.34) and (6.35) are just the same as those obtained by Kashyap and Chaudhry (1983).

If \( \mu(t) \) be such that \( G(t) = 1 \) then \( \alpha_a(t) + \beta_a(t) = 1 \).

In this case we have from (6.20b) and (6.21b)

\[
\text{Cov} \left[ R(t), N(t) \right] = \left[ c + C(2) (1) \right] \int_{0}^{t} \lambda(a) \exp \left[ - \int_{a}^{t} \mu(y) \, dy \right] \, da 
\]

(6.36)

and

\[
\text{Cov} \left[ R(t), D(t) \right] = \left[ c + C(2) (1) \right] \int_{0}^{t} \lambda(a) \left[ 1 - \exp \left[ - \int_{a}^{t} \mu(y) \, dy \right] \right] \, da
\]

(6.37)

Using (6.34), (6.35) we have from (6.12) and (6.14)

\[
L_1(t) = \int_{0}^{t} \lambda(a) \exp \left[ - \int_{a}^{t} \mu(y) \, dy \right] \, da
\]

(6.38)

\[
L_2(t) = \int_{0}^{t} \lambda(a) \left[ 1 - \exp \left[ - \int_{a}^{t} \mu(y) \, dy \right] \right] \, da
\]

(6.39)