THE INHOMOGENEOUS $BMAP/G/\infty$ QUEUE

Lothar Breuer
Dept. of Computer Science
University of Trier
54286 Trier, Germany
email: breuer@info04.uni-trier.de

In queueing theory, most models are based on time-homogeneous arrival processes and service time distributions. However, in communication networks arrival rates and/or the service capacity usually vary periodically in time. In order to reflect this property accurately, it is most natural to consider inhomogeneous arrival processes in queueing models. In the present paper, the inhomogeneous $BMAP/G/\infty$ queue with time–dependent service time distributions is examined. Exact transient distributions as well as approximation formulae are derived. For the special case of homogeneous $BMAP/G/\infty$ queues, stability conditions and asymptotic distributions are given. Finally, it is shown how to derive bounds for the (inhomogeneous) $BMAP/G/c/c$ loss system by means of analyzing the $BMAP/G/\infty$ (inhomogeneous) queue.

1 Introduction

In queueing theory, most models are based on time-homogeneous arrival processes and service time distributions. One of the most important features to be exploited is the Markov property which often appears after the construction of embedded Markov chains. The search for Markovian but versatile arrival processes has led to the concept of batch Markovian arrival processes (BMAPs, see Neuts [20] and Lucantoni [17]) which allow for a phase process controlling the arrival rates. This arrival process is often used for modelling communication networks.

A typical property of communication traffic is the dependence of its arrival rates on time. This aspect incites the use of time-inhomogeneous processes and queues for modelling communication networks. Typically, a periodic dependence of the arrival rates and/or the service time distribution can be assumed with period lengths of a day or a week.

While queues with periodic input naturally reflect the time-dependent amount of traffic that arrives in communication networks, the analysis of queues with inhomogeneous arrival rates is far less developed than the one for homogeneous queues. Some of the existing results in the literature are given in Asmussen and Thorisson [1], Bambos and Walrand [2], Falin [9], Harrison and Lemoine [11], Hasofer [12], Heyman and Whitt [13], Lemoine [16], [15], Massey [18], Rolski [21], [22], and Willie [23]. Although many types of stability conditions could be established, explicit formulae for asymptotic behaviour have not been derived yet, except for Markovian queues (see Breuer [4]).

The present paper gives exact transient distributions as well as approximation formulae for the inhomogeneous $BMAP/G/\infty$ queue. For the special case of homogeneous $BMAP/G/\infty$ queues, stability conditions and asymptotic distributions are given. Finally, it is shown how to derive bounds for the (inhomogeneous) $BMAP/G/c/c$ loss system by means of analyzing the $BMAP/G/\infty$ (inhomogeneous) queue.

The paper is organized as follows. In section 2, the concept of inhomogeneous BMAPs is introduced. For these processes, the transition probabilities as well as z–transforms and expectation formulae are derived. In the special case of periodic BMAPs, a useful recursion formula for the computation of the transition probabilities is given. Section 3 contains the analysis of the (inhomogeneous) $BMAP/G/\infty$ queue. This is examined using the same method of analysis as the classical one for the $M/G/\infty$ queue (cf. Kalashnikov [14], p.80–82) or the inhomogeneous $M_t/G/\infty$ queue (cf. Eick et al. [7]). Thus many results by Eick et al. [7] for the $M_t/G/\infty$ queue can be proven for the inhomogeneous $BMAP/G/\infty$ queue, especially the possibility of using time–dependent service time distributions. First, the homogeneous case will be considered in order to show the method of analysis. For this case, asymptotic behaviour can be examined easily. For the general (i.e. possibly inhomogeneous) case, the same method of analysis applies. Furthermore, an approximation can be given. The last section contains an application of the derived results for the planning procedure of a communication network. This application is of the same form as the approximation of multi–server loss systems by infinite server queues announced in Eick et al. [7, 8].

2 Inhomogeneous BMAPs

Like all Markov arrival processes in queueing theory, BMAPs are Markov jump processes. As main reference for the theory of Markov jump processes, the book by Gikhman, Skorokhod [10] shall be referred to. Analogous to the definition of a BMAP (see Lucantoni [17]), an inhomogeneous BMAP shall be defined by its time-dependent transition rate matrices $(D_n(t) : n \in I\!N_0, t \in I\!R_0^+)$, assuming that the continuity conditions for the transition rates of inhomogeneous Markov jump processes are satisfied (see Gikhman, Skorokhod [10], p.362). Let the number of phases, which is the dimension of the square matrices $D_n(t)$, be denoted by $m \in I\!N$. Then $D_n(t)(i,j)$ is the infinitesimal transition rate of observing n arrivals at time t while changing from phase i to phase j. A periodic BMAP with period T is an inhomogeneous BMAP with the property $D_n(s+T) = D_n(s)$ for all $n \in I\!N_0$ and $s \in [0, T[$. Inhomogeneous BMAPs keep the most important structural properties of BMAPs, such that the computations necessary for analyzing them and the respective queues are still tractable.

Since inhomogeneous BMAPs generalize the classical homogeneous BMAP concept, they often will be referred to as general BMAPs or simply BMAPs in this paper. In the remainder of this section, the most important properties of general BMAPs shall be derived. Formulae for the transition probabilities, which uniquely determine an arrival process in distribution, are given. The concept of z–transforms leads to an expression for the expectation kernel of an SMAP.

2.1 Transition Probabilities

In order to determine the transition probabilities of an inhomogeneous BMAP (N, J), the convolution calculus developed in Baum [3] proves very useful. The method introduced therein yields explicit formulae to compute those transition probabilities via convolutions of matrices.

Define the (time–dependent) generating sequence of an inhomogeneous BMAP by

$$\Delta(t) := (D_n(t) : n \in I\!N_0)$$

for all $t \in I\!R_0^+$. Furthermore, define $N_k(s,t)(i,j)$ as the probability of having $k \in I\!N_0$ arrivals and being in phase j at time $t > s$ under the condition of having 0 arrivals and being in phase i at time s. Further define $N_k(s,t)$ as the $m \times m$ matrix with entries $N_k(s,t)(i,j)$ and $N(s,t) = (N_k(s,t) : k \in I\!N_0)$ as the sequence with entries $N_k(s,t)$.

Solving Kolmogorov's forward equations via the iteration method by Picard and Lindelöf (cf. Gikhman, Skorokhod [10], p.317), the convolution formulae for the transition probabilities (see Baum [3]) assume the following form:

Definition 1 For all $s < t \in \mathbb{R}_0^+$ set

$$N_k^{(0)}(s,t)(i,j) := \delta_{k,0} \cdot \delta_{i,j} = \begin{cases} 1 & \text{for } k = 0,\ i = j \\ 0 & \text{otherwise} \end{cases}$$

for $k \in \mathbb{N}_0$ and $i, j \in \{1, \ldots m\}$. Further, define the sequence

$$N^{(1)}(s,t) := \int_s^t \Delta(u)du$$

and recursively for $n \in \mathbb{N}$

$$N_k^{(n)}(s,t) := \sum_{l=0}^{k} \int_s^t N_l^{(n-1)}(s,u) D_{k-l}(u)du$$

for all $k \in \mathbb{N}_0$.

Theorem 1 *For every $s < t \in \mathbb{R}_0^+$, the transition probability kernel $N(s,t)$ is given by*

$$N(s,t) = \sum_{n=0}^{\infty} N^{(n)}(s,t),$$

meaning that for $k \in \mathbb{N}$

$$N_k(s,t) = \sum_{n=0}^{\infty} N_k^{(n)}(s,t)$$

for all $k \in \mathbb{N}_0$.

Proof: Since inhomogeneous BMAPs are special Markov jump processes, the transition probabilities are given as the solution of the Kolmogorov forward equations (cf. Gikhman, Skorokhod [10], p.314–319). It is easy to verify that they assume the given convolution form.

For two sequences $A = (A_n : n \in \mathbb{N}_0)$ and $B = (B_n : n \in \mathbb{N}_0)$ of matrices, define the convolution sequence $C = (C_n : n \in \mathbb{N}_0) = A * B$ by

$$C_n := \sum_{i=0}^{n} A_i B_{n-i}$$

for all $n \in \mathbb{N}_0$. Then the sequences $N^{(k)}(s,t)$ can be expressed in the closed form

$$N^{(k)}(s,t) = \underbrace{\int_s^t \int_s^{u_k} \ldots \int_s^{u_2}}_{k\ integrals} \Delta(u_1) * \ldots * \Delta(u_k)du_1 \ldots du_k \tag{1}$$

for every $k \in I\!N_0$ and $s < t \in I\!R_0^+$. The sequence $N(s,t)$ of transition probabilities can be computed iteratively by starting with $I_0(s,t) := N^{(0)}(s,t)$ and iterating

$$I_{n+1}(s,t) := \int_s^t I_n(s,u) * \Delta(u)\ du + I_0(s,t)$$

for $n \in I\!N_0$. Then $N(s,t) = \lim_{n\to\infty} I_n(s,t)$ for all $s < t \in I\!R_0^+$.

2.2 Periodic BMAPs

A BMAP shall be called periodic with period $T > 0$ if its generating sequence has the property

$$\Delta(t+T) = \Delta(t)$$

for all $t \in I\!R_0^+$. In the periodic case, formula 1 can be simplified as follows. Define

$$N_k(n) := N^{(k)}(0,nT) = \underbrace{\int_0^{nT}\int_0^{u_k}\ldots\int_0^{u_2}}_{k\ integrals} \Delta(u_1) * \ldots * \Delta(u_k) du_1 \ldots du_k$$

for all $k, n \in I\!N_0$. Now we can prove

Theorem 2 *The following recursion formula holds for all $k, n \in I\!N_0$:*

$$N_k(n+1) = \sum_{j=0}^{k} N_{k-j}(n) \underbrace{\int_0^{T}\int_0^{u_j}\ldots\int_0^{u_2}}_{j\ integrals} \Delta(u_1) * \ldots * \Delta(u_j) du_1 \ldots du_j \qquad (2)$$

Proof: By definition, the partition

$$\begin{aligned}
N_k(n+1) &= N_k(n) + \int_{nT}^{(n+1)T}\int_0^{u_k}\ldots\int_0^{u_2} \Delta(u_1) * \ldots * \Delta(u_k)\ du_1 \ldots du_k \\
&= N_k(n) + \int_{nT}^{(n+1)T} N_{k-1}(n) * \Delta(u_k) du_k \\
&\quad + \int_{nT}^{(n+1)T}\int_{nT}^{u_k}\ldots\int_0^{u_2} \Delta(u_1) * \ldots * \Delta(u_k)\ du_1 \ldots du_k \\
&= \ldots \\
&= \sum_{j=0}^{k} \underbrace{\int_{nT}^{(n+1)T}\int_{nT}^{u_j}\ldots\int_{nT}^{u_2}}_{j\ integrals} N_{k-j}(n) * \Delta(u_1) * \ldots * \Delta(u_j)\ du_1 \ldots du_j
\end{aligned}$$

holds for all $k, n \in \mathbb{N}_0$, defining the case of zero integrals as $P_k(n)$. Exploiting the periodicity of the generator $G(t)$ leads to formula 2.
☺

The initial values for this recursion are $P_0(n) = Id$ for all $n \in \mathbb{N}_0$ and $P_k(0) = \delta_{k0} \cdot Id$ for all $k \in \mathbb{N}_0$. This results in first iterates

$$N_1(n) = \int_0^{nT} \Delta(u)du$$

and

$$N_k(1) = \int_0^T \int_0^{u_k} \dots \int_0^{u_2} \Delta(u_1) * \dots * \Delta(u_k)\, du_1 \dots du_k$$

for all $k, n \in \mathbb{N}$.

Define

$$\lfloor t/T \rfloor := \max\{n \in \mathbb{N}_0 : nT \leq t\}$$

as the number of period lengths that have passed until time $t \in \mathbb{R}^+$. Now the transition probability matrix from time $s \in \mathbb{R}^+$ until time $t > s$ is given by

$$N(s,t) = N(s, \lfloor s/T \rfloor + 1) \sum_{k=0}^{\infty} N_k(\lfloor t/T \rfloor - \lfloor s/T \rfloor)\, N(0, t - \lfloor t/T \rfloor)$$

using recursion formula 2 for computing the matrices $N_k(\lfloor t/T \rfloor - \lfloor s/T \rfloor)$. This expression allows a computation of the transition probability matrix without needing to integrate over ranges larger than the period T.

2.3 Z-Transforms and Expectations

In this section, a z-transform for inhomogeneous BMAPs is derived. This is used to determine the expectation of the process (N, J). Since the proofs for the results of this section are rather lengthy and technical, it is referred to Breuer [5], where they are taken from.

Definition 2 Let (N, J) be an inhomogeneous BMAP. The z-transform of (N, J) over the time interval $]s, t]$ is defined as the function $z \to N(s, t; z)$ with values being the matrices determined by

$$N(s,t;z)(i,j) := \sum_{n=0}^{\infty} P(N_t - N_s = n, J_t = j | J_s = i)\, z^n$$

for all $i, j \in \Phi$ and $z \in \mathbb{C}$ with $|z| \leq 1$.

Theorem 3 *The z-transform of (N, J) over the time interval $]s, t]$ has the form*

$$N(s,t;z) = \sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n\ integrals} \sum_{k=0}^{\infty} D_k(u_1) z^k \dots \sum_{k=0}^{\infty} D_k(u_n) z^k \, du_1 \dots du_n$$

for all $z \in \mathbb{C}$ with $|z| \leq 1$.

Proof: see Breuer [5]

Definition 3 Let (N, J) be an inhomogeneous Batch Markovian Arrival Process. The **expectation** $E(N_t - N_s)$ of the marginal process N during the time interval $]s, t]$ is defined as the matrix with entries

$$E(N_t - N_s)(i,j) := E((N_t - N_s) \cdot \delta_{J_t,j} | J_s = i)$$

Further denote the transition probability matrix of the phase process from time s to time $t \geq s$ by $P^{\Phi}(s,t)$. Then the expectation of an inhomogeneous BMAP is determined as follows:

Theorem 4 *The expectation matrix of the marginal process N over the time interval $]s, t]$ is given by*

$$E(N_t - N_s) = \int_s^t P^{\Phi}(s,u) \sum_{k=1}^{\infty} k \cdot D_k(u) \, P^{\Phi}(u,t) \, du$$

for all $s < t \in \mathbb{R}_0^+$.

Proof: see Breuer [5]

3 The $BMAP/G/\infty$ Queue

The $BMAP/G/\infty$ queue shall be examined in two stages of generality. Essentially, the same method of determining the queue process applies to both of them. In a first step (section 3.1), this method will be shown for the simplest model of time–homogeneous arrival rates. Section 3.3 treats non-homogeneous arrival rates and time–dependent service time distributions.

3.1 The Homogeneous Case

In this section, the $BMAP/G/\infty$ queue with homogeneous arrival rates is examined. Let Q denote a queue with an homogeneous BMAP arrival process (N, J) and infinitely many independent servers. Let Δ denote the generating sequence determining (N, J). Every incoming user is served immediately, i.e. there is no waiting time and the queue length is always zero. Further, let G denote the service time distribution function.

Define $Q_k(s,t)(i,j)$ as the probability of observing k users being served at time t and the system being in phase $j \in \Phi$ under the condition that at time $s < t$ there were no users being served in S and the system was in phase $i \in \Phi$. Let $Q_k(s,t)$ denote the matrix with entries $Q_k(s,t)(i,j)$. Further define $Q(s,t) := (Q_k(s,t) : k \in I\!N_0)$ and $Q(t) := Q(0,t)$.

Fix any time $t \in I\!R_0^+$ the queue shall be observed at. In any infinitesimal time interval $du = \lim_{h\to 0} \,]u, u+h]$ with $u \in [0,t[$, batch arrivals occur with rates $\Delta = (D_k : k \in I\!N_0)$ according to the BMAP arrival process. Given that an arrival of batch size $k \in I\!N$ occured during $]u, u+h]$, the probability of $n \in \{0, \ldots, k\}$ arrivals still being served at time t is distributed binomially by

$$\binom{k}{n} G^c((t-u)-)^n G((t-u)-)^{k-n},$$

denoting $G((t-u)-) := \lim_{h\to 0} G(t-(u+h))$ and $G^c(t) := 1 - G(t)$.

Conditioning upon the size of the batch arrivals, the total rate of n arrivals during a time interval du which are still in service at time t is

$$R_n(u,t) = \sum_{k=n}^{\infty} D_k \binom{k}{n} G^c((t-u)-)^n G((t-u)-)^{k-n} \tag{3}$$

for every $u \in [0,t[$ and $n \in I\!N_0$.

The sequence $R(u,t) := (R_n(u,t) : n \in I\!N_0)$ can be interpreted as the time-dependent generating sequence of an inhomogeneous BMAP $H_t = (H_t(u) : u \le t)$ which at time $u = t$ coincides in distribution with the infinite server queue that is to be examined. Note that for every time $t \in I\!R_0^+$ the queue is observed at, the sequence $R(u,t)$ and hence the BMAP H_t is different. The transient distribution of the queue process at time $t \in I\!R_0^+$ is determined by $Q(t) = H_t(t)$. Theorem 1 yields the following representation:

Theorem 5 *The transient distribution of the marginal queue process of Q at time $t \in I\!R_0^+$ is*

determined by

$$Q(t) = \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_k} \dots \int_0^{u_2}}_{k \text{ integrals}} R(u_1,t) * \dots * R(u_k,t)\, du_1 \dots du_k \tag{4}$$

defining the case of zero integrals as the sequence $(I, 0, 0, \dots)$ *with* I *denoting the identity matrix and* 0 *the zero matrix.*

Proof: This merely is formula 1 for the BMAP H_t.

A great disadvantage in the above formula (4) lies in the fact that the sequences $R(u,t)$ need to be determined separately for every time $t \in I\!R_0^+$ the queue shall be observed at. In the next theorem, a closed form for all times of observance is obtained:

Theorem 6 *The transient distribution of the marginal queue process of* Q *at time* $t \in I\!R_0^+$ *is determined by*

$$Q(t) = \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_1} \dots \int_0^{u_{k-1}}}_{k \text{ integrals}} \tilde{R}(u_1) * \dots * \tilde{R}(u_k)\, du_k \dots du_1 \tag{5}$$

with

$$\tilde{R}(u) := R(t-u,t)$$

being independent of $t > u$.

Proof: By equation (3), the rates

$$\tilde{R}_n(u) := R_n(t-u,t) = \sum_{k=n}^{\infty} D_k \binom{k}{n} G^c(u-)^n G(u-)^{k-n}$$

are independent of $t > u$. Starting from equation (4), we have

$$\begin{aligned}
Q(t) &= \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_k} \dots \int_0^{u_2}}_{\text{k integrals}} R(u_1,t) * \dots * R(u_k,t)\, du_1 \dots du_k \\
&= \sum_{k=0}^{\infty} \int_0^t \int_{u_1}^t \dots \int_{u_{k-1}}^t R(u_1,t) * \dots * R(u_k,t)\, du_k \dots du_1 \\
&= \sum_{k=0}^{\infty} \int_0^t \int_0^{u_1} \dots \int_0^{u_{k-1}} R(t-u_1,t) * \dots * R(t-u_k,t)\, du_k \dots du_1
\end{aligned}$$

which proves the statement.
☺

Remark 1 The representation (5) is from the point of view that one looks backward in time from time t until time 0. Thus, equation (5) gives the transition kernel of a process running backward in time with the phase process still running forward.

The next two theorems yield expressions for the expectation kernel of the marginal queue process N at any time of observance.

Theorem 7 *Assume that the arrival rate is finite for any phase $i \in \Phi$, i.e.*

$$\sum_{n=1}^{\infty}\sum_{j=1}^{m} nD_n(i,j) < \infty \tag{6}$$

for all $i \in \Phi$. Then the expectation kernel of the queue process $Q = (N, J)$ in the subset at time $t \in I\!R_0^+$ is given by

$$E(N_t) = \int_0^t P^{\Phi}(0,u) \sum_{n=1}^{\infty} nD_n P^{\Phi}(u,t)\; G^c((t-u)-)\; du \tag{7}$$

with P^{Φ} denoting the transition kernel of the phase process.

Proof: Fix a time $t \in I\!R_0^+$. Since $Q(t)$ equals the distribution of the BMAP H_t at time t, the expectation kernel at time t is given by theorem 4 as

$$\begin{aligned} E(N_t) &= \int_0^t P^{\Phi}(0,u) \sum_{n=1}^{\infty} nR_n(u,t) P^{\Phi}(u,t)\; du \\ &= \int_0^t P^{\Phi}(0,u) \sum_{n=1}^{\infty} n \sum_{k=n}^{\infty} D_k \binom{k}{n} G^c((t-u)-)^n G((t-u)-)^{k-n} P^{\Phi}(u,t)\; du \end{aligned}$$

Abbreviating $p(u) := G^c((t-u)-)$ and $q(u) := G((t-u)-)$, we have

$$\begin{aligned} E(N_t) &= \int_0^t P^{\Phi}(0,u) \sum_{n=1}^{\infty} n \sum_{k=n}^{\infty} D_k \binom{k}{n} p^n(u) q^{k-n}(u) P^{\Phi}(u,t)\; du \\ &= \int_0^t P^{\Phi}(0,u) \sum_{k=1}^{\infty} \sum_{n=1}^{n} D_k \frac{k!}{n! \cdot (k-n)!} p^n(u) q^{k-n}(u) P^{\Phi}(u,t)\; du \\ &= \int_0^t P^{\Phi}(0,u) \sum_{k=1}^{\infty} kD_k p(u) \sum_{n=1}^{k} \frac{(k-1)!}{(n-1)! \cdot (k-n)!} p^{n-1}(u) q^{k-n}(u) P^{\Phi}(u,t)\; du \\ &= \int_0^t P^{\Phi}(0,u) \sum_{k=1}^{\infty} kD_k G^c((t-u)-) P^{\Phi}(u,t)\; du \end{aligned}$$

since

$$\sum_{n=1}^{k} \frac{(k-1)!}{(n-1)! \cdot (k-n)!} p^{n-1}(u) q^{k-n}(u) = \sum_{n=0}^{k-1} \frac{(k-1)!}{n! \cdot (k-1-n)!} p^{n}(u) q^{k-1-n}(u)$$

sums up to 1.
☺

Theorem 8 *If condition (6) holds and the phase process has stationary distribution π, then the expectation of the additive process of Q at time $t \in I\!R_0^+$ starting in phase equilibrium is*

$$E_\pi(N_t, J_t \in \Phi) = \pi E(N_t) 1_m = \pi \sum_{n=1}^{\infty} n D_n 1_m \cdot \int_0^t G^c(u-)\, du \tag{8}$$

Proof: This follows from the above theorem 7. Since $\pi P^\Phi(0,u) = \pi$ for all $u \in I\!R_0^+$ and $P^\Phi(u,t)1_m = 1_m$ for all $u < t \in I\!R_0^+$, we have

$$\begin{aligned} E_\pi(N_t, J_t \in \Phi) &= \pi E(N_t) 1_m = \int_0^t \pi \sum_{n=1}^{\infty} n D_n 1_m G^c((t-u)-)\, du \\ &= \pi \sum_{n=1}^{\infty} n D_n 1_m \cdot \int_0^t G^c(v-)\, dv \end{aligned}$$

after substituting $v := t - u$.
☺

3.2 Stability and Asymptotic Distribution

The asymptotic behaviour of the homogeneous BMAP/G/∞ queue already is apparent in formulae (5) and (7) of the transient distributions and expectation matrices, respectively. The marginal expectation $E(N_t)$ remains finite for $t \to \infty$ if and only if the mean service time $E(G)$ as well as the mean arrival rate are finite. As will be shown in this section, these conditions are, as intuition would suggest, the main ingredients of a stability condition for the homogeneous SMAP/G/∞ queue.

After first defining stability for homogeneous infinite server queues, the asymptotic distribution and marginal expectation of users in the queue are given for stable BMAP/G/∞ queues. Finally, a necessary and sufficient stability condition is proven for the case of a countable phase space at the end of this section.

Definition 4 An homogeneous queue $Q = (Q_t : t \in I\!R^+)$ shall be called **stable** if the asymptotic distribution $q := \lim_{t\to\infty} Q_t$ does exist and the marginal asymptotic distribution of users in the queue has finite expectation.

Theorem 9 *If $Q = (N, J)$ has an asymptotic distribution, it is determined by the kernel*

$$Q = \sum_{k=0}^{\infty} \underbrace{\int_0^{\infty} \int_0^{u_1} \dots \int_0^{u_{k-1}}}_{\text{k integrals}} \tilde{R}(u_1) * \dots * \tilde{R}(u_k)\, du_k \dots du_1$$

Denote the asymptotic distribution of the phase process J by π. Then the asymptotic expectation of the marginal process N is given by

$$E(N, J \in \Phi) := \lim_{t\to\infty} E(N_t, J_t \in \Phi) = E(G) \cdot \pi \sum_{n=1}^{\infty} nD_n)1_m$$

independent of the initial distribution.

Proof: The first statement follows immediately from formula (5) and the existence of the limit $Q = \lim_{t\to\infty} Q(t)$. The second formula follows from equation (7) if one recognizes that with t growing to infinity, the term $G^c((t-u)-)$ tends to zero for the times u during which the phase process is not in equilibrium yet.
☺

In the rest of this section, a stability condition is derived.

Theorem 10 *Let $Q = (N, J)$ denote an homogeneous BMAP/G/∞ queue. Assume that the phase process J is irreducible, has no absorbing states and an asymptotic distribution π. Then Q is stable if and only if the stability condition*

$$\pi \sum_{n=1}^{\infty} nD_n 1_m \cdot E(G) < \infty \tag{9}$$

holds.

Proof: First we show necessity. Assume that the queue is stable. Then the asymptotic expectation of the marginal process N is given by

$$E(N, J \in \Phi) = E(G) \cdot \pi \sum_{n=1}^{\infty} nD_n 1_m < \infty$$

according to the above theorem 9. This yields the necessity of condition (9).

Sufficiency will be shown by comparison to the $M/G/\infty$ queue. Denote the work load process of Q by $W = (W_t : t \in I\!R_0^+)$. Define $\tilde{Q}$ as the following $M/G/\infty$ queue. The arrival process of $\tilde{Q}$ shall be a Poisson process with rate

$$\gamma_{\max} := -\min_{i\in\Phi} D_0(R)(i,i) = \max_{i\in\Phi} |D_0(R)(i,i)| < \infty$$

Let $i_{\max} \in \Phi$ denote a phase with the highest expectation of the arrival batch size, i.e.

$$\frac{1}{\gamma_{i_{\max}}} \sum_{n=1}^{\infty}\sum_{h=1}^{m} nD_n(i_{\max},h) \geq \frac{1}{\gamma_j} \sum_{n=1}^{\infty}\sum_{h=1}^{m} nD_n(j,h)$$

for all $j \in \Phi$. Since there are no absorbing states, the exit rate $\gamma_{i_{\max}}$ is a positive number. Define the service time distribution of $\tilde{Q}$ by

$$\tilde{G} := \frac{1}{\gamma_{i_{\max}}} \sum_{n=1}^{\infty}\sum_{h=1}^{m} D_n(i_{\max},h) G^{*n}$$

denoting the service time distribution of Q by G and the n–fold convolution of G by G^{*n}. The $M/G/\infty$ queue constructed in this way has an arrival rate which is an upper bound of the phase specific arrival rates of Q. Furthermore, the batch arrivals of $Q(R)$ are interpreted as single arrivals in $\tilde{Q}$ with only one server working off the accrued service requirement of all the users in the batch arrival. Thus the work load $\tilde{W} = (\tilde{W}_t : t \in I\!R_0^+)$ of $\tilde{Q}$ is certainly at least as high as the work load W of Q. Since by Wald's equation

$$E(\tilde{G}) = \frac{1}{\gamma_{i_{\max}}} \sum_{n=1}^{\infty}\sum_{h=1}^{m} D_n(i_{\max},h) \cdot nE(G) = \frac{1}{\gamma_{i_{\max}}} \cdot E(G) \cdot \sum_{n=1}^{\infty}\sum_{h=1}^{m} nD_n(i_{\max},h)$$

we have $E(\tilde{G}) < \infty$ by assumption of the stability condition. This implies that $\tilde{W}$ is positive recurrent. i.e. the expected duration between the time instants of $\tilde{W}$ reaching the state 0 is finite. Since

$$\tilde{W}_t \geq W_t \geq 0$$

certainly for all $t \in I\!R_0^+$, the work load process W of Q is positive recurrent, too. Hence Q has an asymptotic distribution. The finiteness of the asymptotic expectation of the marginal process N is immediate from theorem 9 and the stability condition.
☺

3.3 The General Case

In communication networks, the characteristics of the arrival stream typically change in time (e.g. over the course of the day or the week). Stochastic models reflect this best by inhomogeneous arrival processes. The concept of inhomogeneous BMAPs introduced in section 2 does

provide for arrival rates varying in time and hence can be used to model this phenomenon. An examination analogous to the homogeneous case yields a solution for the corresponding infinite server queue. In this section, the same method of analysis as in section 3.1 shall be applied to infinite server queues with general arrival rates.

The infinite server queue fed by an inhomogeneous BMAP can be analyzed analogously to the homogeneous case in section 3.1. Let $Q = (Q(t) : t \in I\!R_0^+)$ denote a spatial queue with general BMAP input (N, J), general service time distribution G and infinitely many servers. The arrival process (N, J) shall be defined by its time-dependent generating sequence $(\Delta(t) : t \in I\!R_0^+)$.

Be $t \in I\!R^+$ the time instant the queue is observed at. Now the arrival rates at times $s \in [0, t]$ are not constant anymore, but depend on the time instant s. The total rate of n arrivals during a time interval $du = \lim_{h\to 0}]u, u + h]$ is

$$R_n(u,t) = \sum_{k=n}^{\infty} D_k(u) \binom{k}{n} G^c((t-u)-)^n G((t-u)-)^{k-n}$$

for every $u \in [0, t[$ and $n \in I\!N_0$. As in the case of homogeneous arrival rates, the sequence $R(u,t) = (R_n(u,t) : n \in I\!N_0)$ can be interpreted as the time-dependent sequence of transition rates of an inhomogeneous SMAP $H_t = (H_t(u) : u \le t)$ which at time $u = t$ coincides in distribution with the infinite server queue that is to be examined.

The same arguments as in section 3.1 lead to the result

Theorem 11 *The transient distribution of the marginal queue process of Q at time $t \in I\!R_0^+$ is determined by*

$$Q(t) = \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_k} \dots \int_0^{u_2}}_{k \text{ integrals}} R(u_1,t) * \dots * R(u_k,t)\, du_1 \dots du_k \tag{10}$$

defining the case of zero integrals as the sequence $(I, 0, 0, \dots)$ with I denoting the identity matrix and 0 the zero matrix.

Remark 2 Unfortunately, a unification of the above formula analogous to formula (5) cannot be achieved in general. This is due to the fact that

$$R_n(t-u,t) = \sum_{k=n}^{\infty} D_k(t-u) \binom{k}{n} G^c(u-)^n G(u-)^{k-n}$$

still depends on t if the arrival rates are inhomogeneous.

On the other hand, it is possible to extend the theory towards time–dependent service time distribution functions. This would lead to rates of the form

$$R_n(u,t) = \sum_{k=n}^{\infty} D_k(u) \binom{k}{n} G_u^c((t-u)-)^n G_u((t-u)-)^{k-n}$$

for every $u \in [0,t[$ and $n \in I\!N_0$, with G_u denoting the service time distribution function which holds for users arriving during the infinitesimal time interval du.

3.4 An Approximation

This section shows an approximation method which can be used in order to determine the distribution of the queue process at an arbitrary time $t \in I\!R_0^+$ without needing to integrate over the whole range $[0,t[$. The approximation regards only arrivals up to a certain time distance into the past. All arrivals which have occured before are neglected. Thus, this method approximates the service time distribution by a distribution with cut tail.

Conditions for this approximation are the standard assumptions of finite mean service time and finite arrival rates. First, it is shown that the matrices of the generating sequence of H_t are uniformly bounded. Then, this result is used to estimate the approximation error.

Assume in this chapter that the mean service time

$$E(G) = \int_0^{\infty} G^c(u)\, du < \infty \tag{11}$$

is finite. Further assume that the arrival rate

$$\sum_{n=1}^{\infty} \sum_{j=1}^{m} n D_n(t)(i,j) < M < \infty$$

is bounded by a finite value $M \in I\!R^+$ for all $t \in I\!R^+$ and $i \in \Phi$. The next result gives the decisive bound for the approximation following.

Theorem 12 *For every $\varepsilon > 0$, there is a time distance $T(\varepsilon) \in I\!R^+$ such that the inequality*

$$\left\| \int_0^s R_n(u,t)\, du \right\| < M \cdot \varepsilon$$

holds for all $n \in I\!N_0$ and $s \le t - T(\varepsilon)$, with $\|K\| := \sup_{i\in\Phi} \sum_{j=1}^{m} K(i,j)$ denoting the row norm for matrices.

Proof: Choose any $\varepsilon > 0$. By assumption (11), there is a $T(\varepsilon) \in I\!R^+$ such that

$$\int_{T(\varepsilon)}^{\infty} G^c(u-)\ du < \varepsilon$$

Fix any $y \in \Phi$ and $s \leq t - T(\varepsilon)$. Since all elements of $R_n(u,t)$ are positive for all $u \in [0,t]$ and $n \in I\!N$, we have

$$\left\| \int_0^s \sum_{n=1}^{\infty} R_n(u,t)\ du \right\| \leq \left\| \int_0^s \sum_{n=1}^{\infty} nR_n(u,t)\ du \right\|$$

Analogously to the proof of theorem 7, one can show that

$$\int_0^s \sum_{n=1}^{\infty} nR_n(u,t)\ du = \int_0^s \sum_{n=1}^{\infty} nD_n(u)\ G^c((t-u)-)\ du$$

For all $k \in I\!N$, the estimation

$$\begin{aligned} \left\| \int_0^s R_k(u,t)\ du \right\| &\leq \left\| \int_0^s \sum_{n=1}^{\infty} R_n(u,t)\ du \right\| \leq \left\| \int_0^s \sum_{n=1}^{\infty} nD_n(u)\ G^c((t-u)-)\ du \right\| \\ &\leq \int_0^s \left\| \sum_{n=1}^{\infty} nD_n(u) \right\| G^c((t-u)-)\ du < M \cdot \int_0^s G^c((t-u)-)\ du \end{aligned}$$

holds. Substituting $v := t - u$ leads to

$$\begin{aligned} \left\| \int_0^s R_k(u,t)\ du \right\| &\leq \left\| \int_0^s \sum_{n=1}^{\infty} R_n(u,t)\ du \right\| < M \cdot \int_{t-s}^{t} G^c(v-)\ dv \\ &\leq M \cdot \int_{T(\varepsilon)}^{\infty} G^c(v-)\ dv < M \cdot \varepsilon \end{aligned}$$

Since

$$\int_0^s \sum_{n=1}^{\infty} \sum_{j=1}^{m} R_n(u,t)(i,j)du = -\int_0^s \sum_{j=1}^{m} R_0(u,t)(i,j)du$$

for all $s \in I\!R_0^+$, the proof is complete.
☺

Using this bound, we can prove the following approximation:

Theorem 13 *At any time $t \in I\!R^+$ and for every $i \in \Phi$, the distribution of the queue process can be approximated by*

$$\sum_{j=1}^{m} Q(t)(i,j) = \sum_{j=1}^{m} Q(0,t)(i,j) \approx \sum_{j=1}^{m} Q(t - T(\varepsilon), t)(i,j)$$

The approximation error is at most

$$\left| \sum_{j=1}^{m} Q_n(t)(i,j) - \sum_{j=1}^{m} Q_n(t - T(\varepsilon), t)(i,j) \right| < M \cdot \varepsilon$$

for all $n \in I\!N_0$ *and* $i \in \Phi$.

Proof: For every $k \in I\!N$, the difference between the respective k-fold integrals appearing in formula (10) is

$$\begin{aligned}
&\int_0^t \ldots \int_0^{u_{k-1}} R(u_k,t) * \ldots * R(u_1,t)\; du_k \ldots du_1 \\
&\quad - \int_{t-T(\varepsilon)}^{t} \ldots \int_{t-T(\varepsilon)}^{u_{k-1}} R(u_k,t) * \ldots * R(u_1,t)\; du_k \ldots du_1 \\
&= \int_0^t \ldots \int_0^{u_{k-2}} \int_0^{t-T(\varepsilon)} R(u_k,t) * \ldots * R(u_1,t)\; du_k \ldots du_1 \\
&= \int_0^{t-T(\varepsilon)} R(u_k,t)\; du_k * \int_0^t \ldots \int_0^{u_{k-2}} R(u_{k-1},t) * \ldots * R(u_1,t)\; du_{k-1} \ldots du_1
\end{aligned}$$

Summing up over all $k \in I\!N_0$ yields

$$\begin{aligned}
&Q(t) - Q(t - T(\varepsilon), t) = \\
&\quad = \int_0^{t-T(\varepsilon)} R(u,t)\; du \; * \sum_{k=0}^{\infty} \int_0^t \ldots \int_0^{u_{k-1}} R(u_k,t) * \ldots * R(u_1,t)\; du_k \ldots du_1 \\
&\quad = \int_0^{t-T(\varepsilon)} R(u,t)\; du * Q(t)
\end{aligned}$$

Hence, for any $i \in \Phi$ and $n \in I\!N_0$ we have

$$\begin{aligned}
&\left| \sum_{j=1}^{m} Q_n(t)(i,j) \; - \sum_{j=1}^{m} Q_n(t - T(\varepsilon), t)(i,j) \right| \\
&\qquad = \left| \sum_{k=0}^{n} \int_{u=0}^{t-T(\varepsilon)} \sum_{h,j=1}^{m} R_k(u)(i,h) du \; Q_{n-k}(t)(h,j) \right| \\
&\qquad \leq \max_{k=0,\ldots,n} \left| \int_0^{t-T(\varepsilon)} \sum_{h=1}^{m} R_k(u)(i,h)\; du \right|
\end{aligned}$$

since for every $h \in \Phi$ the probability $\sum_{k=0}^{n} \left| \sum_{j=1}^{m} Q_{n-k}(t)(h,j) \right|$ does not exceed 1. Because

$$\max_{k=0,\ldots,n} \left| \int_0^{t-T(\varepsilon)} \sum_{h=1}^{m} R_k(u)(i,h) du \right| < M \cdot \varepsilon$$

according to theorem 12, the proof is complete.
☺

4 Bounds for the $BMAP/G/c/c$ Loss System

Now the results which have been obtained shall be applied to support the planning procedure of a communication network. Assume that a network with a maximum loss (or outage) probability of ε is to be designed, which means that a user who wants to use the network shall find it busy with a probability of at most ε. Since a network with finite capacity can be modelled by a (possibly inhomogeneous) $BMAP/G/c/c$ queue, the search is for the minimal sufficient capacity such that the loss probability can be guaranteed to remain smaller than ε.

Then the analysis of infinite server queues can help to derive a capacity function $C_\varepsilon : I\!R^+ \to I\!N_0$ which yields for every time $t \in I\!R^+$ a number $C_\varepsilon(t)$ such that if the network has the capacity of serving $C_\varepsilon(t)$ users at time t, then the quality of service with respect to the given maximum outage probability ε is guaranteed. Hence, in order to build a network that can guarantee the maximum outage probability ε, it suffices to provide service capacity according to the function C_ε.

In order to determine C_ε, one can proceed as follows. Define a $BMAP/G/\infty$ queue with appropriate time–dependent arrival rates. An estimation procedure for the parameters of an homogeneous BMAP can be found in Breuer [5, 6]. Then either the stationary distribution (in the homogeneous case) or the distributions at a given time of day (in the case of periodic arrival rates with period length being a day) can be represented by a set of functions $P_t : I\!N_0 \to [0,1]$ which yields the probability $P_t(n)$ of n users being served at day time t. For the homogeneous case, we have a function P instead of a set $(P_t : t \in [0,24])$ of functions which depend on the day time.

These are the distributions for an infinite server queue, which means that the probabilities were derived under the assumption that every user can be served. Hence the work load for any real network with some positive outage probability would be lower. It thus suffices to determine the capacity function C_ε for the infinite server queue in order to find a sufficient capacity function for any real network. Since the outage probability ε usually will be chosen very small, the approximation by an infinite server queue seems reasonably sharp.

Now the capacity function for the infinite server queue can easily be determined as

$$C_\varepsilon(t) := \min\left\{ n \in I\!N_0 : \sum_{k=n+1}^{\infty} P_t(k) < \varepsilon \right\}$$

for all $t \in I\!R^+$. This means that the value $C_\varepsilon(t)$ of the capacity function at a time t is given by the lowest number n of users such that the probability of more than n users being served

would be smaller than ε for the respective infinite server queue. Since the work load of any real network is lower than the work load for the infinite server queue, the outage probability of a network satisfying this capacity function can be guaranteed to stay smaller than the given threshold ε.

References

[1] S. Asmussen, H. Thorisson (1987): "A Markov Chain Approach to Periodic Queues", J. Appl. Prob. 24, pp.215-225

[2] BAMBOS, N. AND WALRAND J. (1989) On Queues with periodic inputs. *J. Appl. Prob.* **26,** 381–389

[3] D. Baum (1996): "Ein Faltungskalkül für Matrizenfolgen und verallgemeinerte Poisson-Gruppenprozesse", Research Report No.96-36, Department of Mathematics and Computer Science, University of Trier

[4] L. Breuer (1999): "Markovian Spatial Queues with Periodic Arrival and Service Rates", Proceedings of the 10th MMB conference in Trier, Research Report No.99-17, Department of Mathematics and Computer Science, University of Trier

[5] L. Breuer (2000): "Spatial Queues", Ph.D. thesis, University of Trier, to appear

[6] L. Breuer (2000): "Parameter Estimation for a Class of BMAPs", in: G. Latouche, P. Taylor (ed.): "Advances in Algorithmic Methods for Stochastic Models - Proceedings of the 3rd International Conference on Matrix Analytic Methods", Notable Publications Inc. (New Jersey), to appear

[7] S. Eick, W. Massey, W. Whitt (1993): "The Physics of the $M_t/G/\infty$ Queue", Operations Research 41(4), pp.731–742

[8] S. Eick, W. Massey, W. Whitt (1993): "$M_t/G/\infty$ Queues with Sinusoidal Arrival Rates", Management Science 39(2), pp.241–252

[9] G.I. Falin (1989): "Periodic Queues in Heavy Traffic", Adv. Appl. Prob. 21, pp.485-487

[10] I. Gikhman, A. Skorokhod (1969): "Introduction to the Theory of Random Processes", W.B. Saunders

[11] HARRISON, J.M. AND LEMOINE, A.J. (1977) Limit Theorems for Periodic Queues. *J. Appl. Prob.* **14,** 566–576

[12] A.M. Hasofer (1964): "On the Single-Server Queue with Non-Homogeneous Poison Input and General Service Time", J. Appl. Prob. 1, pp.369-384

[13] HEYMAN, D.P. AND WHITT, W. (1984) The Asymptotic Behavior of Queues with Time-Varying Arrival Rates. *J. Appl. Prob.* **21,** 143–156

[14] V. Kalashnikov (1994): "Mathematical Methods in Queueing Theory", Kluwer Academic Publishers

[15] LEMOINE, A.J. (1981) On Queues with Periodic Poisson Input. *J. Appl. Prob.* **18,** 889–900

[16] LEMOINE, A.J. (1989) Waiting Time and Workload in Queues with Periodic Poisson Input. *J. Appl. Prob.* **26,** 390–397

[17] D. Lucantoni (1991): "New Results on the Single Server Queue with a Batch Markovian Arrival Process", Comm. Statist. - Stochastic Models, Bd. 7(1), pp.1-46

[18] W. Massey (1981): "Non–Stationary Queues", Ph.D. thesis, Stanford University

[19] W.A. Massey, W. Whitt (1993): "Networks of infinite server queues with non-stationary Poisson input", Queueing Systems 13, pp.183-250

[20] M. Neuts (1979): "A versatile Markovian point process", J. Appl. Prob. 16, pp.764–79

[21] ROLSKI, T. (1987) Approximation of Periodic Queues *Adv. Appl. Prob.* **19,** 691–707

[22] T. Rolski (1989): "Relationships between Characteristics in Periodic Poisson Queues", Queueing Systems 4, pp.17-26

[23] WILLIE, H. (1998) Periodic Steady State of Loss Systems with Periodic Inputs. *Adv. Appl. Prob.* **30,** 152–166