ASTA IMPLIES AN M/G/1-LIKE LOAD DECOMPOSITION
FOR A SERVER WITH VACATIONS

by

Z. Rosberg and H.R. Gail

Technical Report #538
February 1989
ASTA IMPLIES AN $M/G/1$ - LIKE LOAD DECOMPOSITION FOR A SERVER WITH VACATIONS

Zvi Rosberg

H. Richard Gail

(October 1989)
Technical Report No. 538

Computer Science Dept.
Technion, IIT
Israel

Abstract

We consider a single server at which customers of several classes arrive according to a stochastic point process and the server may take vacations arbitrarily. We show that under stationary conditions, the ASTA property (Arrivals See Time Averages) implies an $M/G/1$-like load decomposition property. A network of quasi-reversible symmetric queues with server vacations is given as an example for non Poisson arrivals.

Keywords: ASTA, stochastic decomposition, point processes, vacation models, queues, quasi-reversibility.

*Computer Science Dept., Technion, IIT, Israel
†IBM T.J. Watson Research Center, Yorktown Heights, NY 10598
1 Introduction and preliminaries

The purpose of this note is to show that Poisson arrivals are not necessary for an $M/G/1$-like load decomposition property in a queueing model with server vacations. As will be demonstrated in section 3, an $M/G/1$-like decomposition property occurs in product-form queueing networks with server vacations.

Consider a single server at which customers arrive according to a stochastic point process $\mathcal{N} = \{N(t), t \geq 0\}$ on $[0, \infty)$. The random variable $N(t)$ is the number of arrivals during the interval $[0, t]$. We assume that the sample paths of $\mathcal{N}$ are right-continuous and that

$$0 < E[N(t)] < \infty, \text{ for every } t.$$

Every arrival brings into the system an amount of service that has to be processed by the server. Let $S_1, S_2, \ldots$ be the sequence of processing times that are brought by the arrivals. We assume that $(S_1, S_2, \ldots)$ is a sequence of i.i.d r.v.'s, and each $S_n$ is distributed according to the probability distribution $F_S(\cdot)$. The random variable $S_n$ represents the total processing time of the $n^{th}$ arrival, where the $n^{th}$ arrival may be a single customer or a batch. In the case where there are different classes of customers, each having its own service requirement distribution, $S$ is the appropriate weighted distribution of an arriving customer.

At any instant, the server may be either in service or on vacation. When the server is in service it serves customers at a constant unit rate. No particular service discipline is assumed, except of being work conserving; preemptive resume and switching are allowed at any moment of time. When the server is on vacation, no load is removed from the system. The server may go on vacation at any moment of time, and whenever the system is empty the server is regarded as on vacation.
Let $V = \{V(t), t \geq 0\}$ be a binary valued stochastic process that represents the activity of the server. That is, for every $t$, $V(t) = 1$ or $0$, depending upon whether or not the server is in service or on vacation, respectively. Also, let $X = \{X(t), t \geq 0\}$ be the stochastic process that represents the workload of the service system. We assume that the sample paths of $(X, V)$ are left-continuous. In particular, the load at an arrival instant does not include the processing time of the new arrival.

We assume that the process $\{(N(t), X(t), V(t)), t \geq 0\}$ has a non-degenerate limiting distribution with one ergodic set, and we consider its version under stationary conditions. Hence, we necessarily have,

$$\rho = \frac{E(S)}{E(T)} < 1,$$

where $E(T)$ is the expected interarrival time under stationary conditions.

For an $M/G/1$ queue with server vacations in which the rules that govern a beginning and a termination of a vacation do not anticipate future jumps of the Poisson arrival process, it was shown in [6] that the following decomposition property holds:

*Under stationary conditions, the number of customers present in the system at a random point in time is distributed as the sum of two independent r.v.'s. One is the number of customers present in the 'corresponding' $M/G/1$ queue without server vacations at a random point in time. The other is the number of customers in the system at a random point in time during a vacation.*

A similar decomposition result for the service load in the same model was established in [1] and [2]. Other vacation models for a $GI/G/1$ queue in which the server goes on vacation only when the queue is empty also have been studied. For an extensive survey see [5]. Generalizations of the stochastic decomposition for a single queue under a class of non-renewal arrival processes in which the server takes a vacation only when the
system becomes empty were recently obtained in [4] and [8].

In this study we show that for our general vacation model as above, we may obtain an \( M/G/1 \)-like load decomposition property as in [1], [2] and [6], even when the arrival process is not Poisson. This phenomenon is strongly coupled with the ASTA property (Arrivals See Time Averages). In section 3, we will give an example of a product-form queueing network where this happens.

This study is motivated by a recent result of Melamed and Whitt, [9], [10], where they derived necessary and sufficient conditions for ASTA to hold in a general stochastic model. In particular and in our model, their condition implies that the limiting distribution of \( \{(X(\tau_n-), V(\tau_n-)), n \geq 0\} \) is the same as the limiting distribution of \( \{(X(t), V(t)), t \geq 0\} \), where \( \tau_n \) is the \( n \)th arrival instant.

In this paper we prove that in our general vacation model, an \( M/G/1 \)-like decomposition property holds whenever ASTA holds. To show that, we first present the characterization of ASTA from [10], which is given in terms of the stochastic intensity of the point process \( N \). (For background see, e.g., [3] and [11].)

We assume that the point process \( N \) has a unique predictable stochastic intensity \( \{\nu(t), t \geq 0\} \). That is, \( \{\nu(t), t \geq 0\} \) is a non-negative \( \mathcal{F}_t \)-progressive process such that \( \int_0^\infty \nu(s)ds < \infty \) with probability 1; and for any non-negative \( \mathcal{F}_t \)-predictable process \( \{C(t), t \geq 0\} \),

\[
E\left[ \int_0^\infty C(s)dN(s) \right] = E\left[ \int_0^\infty C(s)d\nu(s) \right].
\]

It is further assumed that for every \( t \),

\[
E\left[ \int_0^t \nu(s)ds \right] = \int_0^t E[\nu(s)ds] < \infty.
\]
The stochastic intensity determines the probability law of the point process \( N' \), and it represents the conditional intensity of \( N' \) at time \( t \), given a history in \( \mathcal{F}_t \).

From [10] it follows: the ASTA property holds if and only if, for every bounded and continuous real function \( f \),

\[
\text{cov}[f(X(t), V(t)), \nu(t)] = 0,
\]

or equivalently,

\[
\text{cov}[f(X(t), V(t)), \lambda(t)] = 0,
\]

where \( \lambda(t) = E[\nu(t) | X(t), V(t)] \).

The condition in (1) is referred to as the Lack of Bias Assumption (LBA), and is shown to hold if and only if

\[
\lambda(t) \equiv \lambda \quad \text{with probability 1.}
\]

Remark 2.1: Observe that \( \lambda \) in (2) is the probability density that an arrival occurs at an arbitrary point in time under stationary conditions.

2 Stochastic decomposition

In this section, we show that an \( \hat{M}/G/1 \)-like load decomposition holds in our general vacation model, whenever LBA holds (i.e., ASTA holds).

Let \( X \) be the system load at an arbitrary instant, and \( Y \) be the system load at an arbitrary instant during a vacation. For every random variable \( Z \), let \( F_Z(\cdot) \) be its
probability distribution and \( \mathcal{L}_Z(s) \) be its Laplace transform. For a discrete random variable \( Z \) and a bounded real function \( c(x) \), we regard \( \int c(x)dF_Z(x) \) as the Riemann-Stieltjes integral. Let \( \tilde{X} \) be the workload process of the corresponding single server station without vacation. That is, the server is on vacation if and only if the system is empty.

**Theorem 2.1** In the vacation model above, if LBA holds, then the workload at an arbitrary instant is distributed as the sum of the workload in the corresponding system without vacation, \( \tilde{X} \), and the workload in the vacation system at an arbitrary instant in the vacation interval, i.e.,

\[
X = \tilde{X} + Y
\]

Furthermore, \( \tilde{X} \) and \( Y \) are independent.

**Proof:** Doshi's up-crossing, down-crossing argument as presented in [1] for an \( M/G/1 \) queue works for our model as well. For the sake of completeness we present it here.

For every load level \( x > 0 \), we equate the probability density function of having a down-crossing with that of having an up-crossing of level \( x \) at an arbitrary instant. Given that the system is at level \( z \), \( 0 \leq z \leq x \), at an arbitrary instant, then the LBA and Remark 2.1 imply that the probability density function of having an up-crossing of level \( x \) is given by \( \lambda(1 - F_S(x-z)) \). Thus, the probability density function of having an up-crossing of level \( x \) at an arbitrary instant is,

\[
\int_0^x \lambda(1 - F_S(x-z))dF_X(z).
\]

The probability density function of a load level \( x \) at an arbitrary instant during a vacation is \( (1 - \rho)dF_Y(x) \). Since service is provided at unit rate, a down-crossing of level \( x > 0 \) occurs whenever the server is in service and the load is \( x \). Therefore, the
probability density function of having a level \( x \) down-crossing is,

\[
dF_x(x) - (1 - \rho)dF_y(x).
\]

Under our ergodicity assumption, we have an equality between the probability density functions of up-crossing and down-crossing. Hence,

\[
dF_x(x) - (1 - \rho)dF_y(x) = \lambda \int_0^x (1 - F_S(x - z)) dF_x(z), \quad x > 0. \tag{3}
\]

Since the server is on vacation whenever the system is idle,

\[
dF_x(0) = (1 - \rho)dF_y(0) \tag{4}
\]

Taking the Laplace transform, we have from (3) and (4),

\[
\mathcal{L}_x(s) - (1 - \rho)\mathcal{L}_y(s) = \frac{\lambda(1 - \mathcal{L}_S(s))\mathcal{L}_x(s)}{s}.
\]

Therefore, we have

\[
\mathcal{L}_x(s) = \mathcal{L}_y(s) \frac{s(1 - \rho)}{s - \lambda + \lambda\mathcal{L}_S(s)} \tag{5}
\]

To complete the proof, we note that the second term on the righthand side is the Laplace transform of the workload for the \( M/G/1 \) system without vacation. Indeed, we recognize it as the corresponding transform for a non-vacation system with Poisson arrivals (or equivalently, \[10\], with ASTA arrivals). \( \square \)

In the next section, we will give an example with non Poisson arrivals.
3 A queueing network example

To generate our example, consider a symmetric queue as defined in section 3.3 in [7] (e.g., an $M/G/1$ queue with different classes of customers under a LIFO preemptive resume or a Processor Sharing regime.) Each customer of class $c$ requires a service $S_c$, which is generally distributed. In addition, after every service completion of a class $c$ customer (alternatively, before each service beginning), the server has to complete an over-head task that lasts $V_c$ units of time. We assume that all service requirement and over-head task durations are independent r.v.'s.

While working on over-head tasks, the server is regarded as being on vacation, and during that time the server obeys the same scheduling regime as for regular customers (e.g., if a LIFO regime is implemented, then an over-head task is preempted by new arrivals and resumes later according to the LIFO order).

From [7], symmetric queues without vacations are quasi-reversible under any service requirement distribution. To verify that symmetric queues with our defined vacations (before or after regular services) are also quasi-reversible, consider first the case where the over-head tasks are performed before regular services start. Clearly, over-head time can be regarded as part of the customer service, and therefore we have a quasi-reversible queue with services that take $S_c + V_c$ instead of $S_c$. For the case where vacations are taken after service completion, one observes that this queue is the time-reversed queue of the former one. Since a time-reversed quasi-reversible queue is also quasi-reversible, we have that property for both type of vacations.

Now, consider a network of quasi-reversible vacation queues as above, where external arrivals are Poisson, routing is Markovian and loops are permitted, [11]. It is well known that such networks admit product-form solutions, and the flow of customers along routes that contain loops, are not Poisson. Nevertheless, a customer in transit (who
just arrived from outside, or just-left a queue and is about to join another queue) sees
the state of the network (not including himself) according to the stationary distribution.
Thus, by considering a single queue in the network along a route that contains loops,
we do have the ASTA property, and arrivals at that queue are not Poisson. Therefore,
from Theorem 2.1 we have the decomposition property for that queue.

The LBA condition trivially follows from the independence of the queues under sta­
tionary conditions and condition (2). Also, the decomposition property can be verified
by direct computation of the stationary distribution of the queue state augmented with
binary attributes for each customer that has not yet completed his service or over-head
task.
References


