A COMPLETE CLASS OF M/G/n QUEUEING DISCIPLINES WITH THE MM PROPERTY

A. S. Noetzel

June 1975 TR-50

Technical Report TR-50

Department of Computer Sciences
University of Texas
Austin, Texas 78712

The work reported here was partially supported by
National Science Foundation Grant No. GJ-39658

Abstract

A symmetric queueing discipline is defined as a scheduling algorithm that assigns processing rates to the jobs in a queue only at arrival and departure times, and for each n, the processing rate distribution over the n customers in the queue, ordered by their relative arrival time, is constant. A simple and complete characterization of a particular subclass of symmetric disciplines, called the Batch Processor-Sharing (BPS) algorithms, is shown to be a generalization of the Last-Come-First-Served Preemptive, Processor Sharing, and the No Queueing disciplines. It is shown that all BPS disciplines have the departure independence or MM property, and furthermore, any symmetric algorithm with the MM property must be a BPS algorithm. Disciplines from this class can be chosen to form analyzable queueing networks with general service time distributions, priority queueing, and load-dependent processing rates.

1. Introduction

A queue has the MM (Markov implies Markov) property if, whenever the arrival process is Poisson, the departure process is also Poisson. The MM property is of importance in the analysis of queueing networks, since the steady-state behavior of a network of queues with the MM property is determined only by the average input and output rate of each queue. The queues are then independent of each other to the extent that the state of the network can be expressed as the product of the states of the individual queues.

It is well known that every queue for which the processing-time distribution is exponential has the MM property. (Queueing disciplines that schedule on the basis of future or accumulated processing times, or which allow the processor to remain idle when a customer is ready, are excluded.) The product form for queueing networks with exponential service-time distributions is introduced in the classical papers by Jackson (J1) and Gordon and Newell (G1).

If the processing-time distribution of the queue is general, the MM property is an attribute of the queueing discipline, that is, the rule by which processing power is allocated to the customers in the queue. Chandy (C1) defines the concept of local balance and shows several instances in which this property leads to the product form for networks of queues with general service-time distributions. In particular, it is shown that networks of queues with the Last-Come-First-Served Preemptive (LCFSP) discipline and the Processor Sharing (PS) discipline have the product form. Also, if an arriving customer can be guaranteed immediate service by an available processor, the No-Queueing (NQ) discipline (the $M/G/\infty$ queue) affords the product form.

It is shown by Muntz (M2) that if the steady-state balance equations of a queue satisfy the departure independence condition, the queue has the MM property. It is thereby shown that the LCFSP and PS disciplines have the MM property. The NQ discipline is shown to have the MM property in (M1).

The results in (C1) are generalized by Baskett, et al, (B1) to queues with different classes of customers. In (C2), a characterization of a general queue is presented, and the local balance condition for a single queue is defined to be equivalent to the departure independence property when the balance equation is satisfied. With these definitions, local balance, departure independence, the MM property and the product form are all equivalent.

Furthermore, through the concept of station balance, (C2) defines a sufficient condition for local balance. The PS, LCFSP, and NQ disciplines have these properties, but the complete class of disciplines that have these properties is unknown.

In this paper, the MM property is considered for a particular class of disciplines, called symmetric disciplines. A queueing discipline is symmetric if the allocation of processing power to customers in the queue is made only at arrival and departure times, and for a given number of jobs in the station, the processing-rate assignments made after an arrival match those made after a departure. PS, LCFSP, and NQ are included in the class of symmetric disciplines. A simple rule for processing power assignments in symmetric disciplines is presented, and it is shown that this rule is both a necessary and sufficient condition for the discipline to have the MM property.

2. Departure Independence and Local Balance

The concepts of departure independence and local balance for a queue are defined in this section.

Let S represent the state of a queue. Let S^+ be the set of states which have exactly one more customer than S, let S^- be the set of states that have exactly one less customer than S, and let U be the universe of all states except S. Let R(S,S') be the conditional rate at which the queue state changes from S to S', given that the queue state is S.

The balance equation for a queue equates the rate at which the queue leaves state S to the rate at which it enters state S. It can be written to show the various kinds of transitions as follows:

$$P(S)\left[\sum_{S'\varepsilon S^+} R(S,S') + \sum_{S'\varepsilon U-S^+} R(S,S')\right]$$

$$= \sum_{S'\varepsilon S^+} P(S')\ R(S',S) + \sum_{S'\varepsilon U-S^+} P(S')\ R(S',S). \qquad (1)$$

If a queue has the departure independence property, then for all S, the rate at which the queue leaves state S due to an arrival, equals the rate at which it enters state S due to a departure. Departure independence equates the terms represented by the leftmost summations on either side of (1).

If the input process to the queue is Poisson with average arrival rate λ, (an M/G/n queue), the conditional rates of transition out of a state due to an arrival, is state-independent, i.e. $\sum_{S'\varepsilon S^+} R(S,S') = \lambda$ for all S. Muntz (M2) has shown that if the departure independence equation, which in this case is expressed as

$$\lambda P(S) = \sum_{S'\varepsilon S^+} P(S')\ R(S',S) \qquad (2)$$

holds for all S, then the output of the queue is a Poisson process. Queues with the departure independence property are therefore also said to have the MM, or "Markov implies Markov" property.

If the departure independence equation is assumed to hold, the terms of the balance equation remaining after the departure independence terms are cancelled, will be called the local balance condition.[1]

If the processing-time distribution has a rational Laplace transform, it can be represented by an exponential network. Each customer in the queue is represented in the state by the stage of processing that he has reached in the exponential network, and in particular a newly arriving job always enters the queue in the first stage. For deterministic disciplines, then, only one state transition is possible when an arrival occurs in a particular state S, and the conditional rate of that transition is the input rate. If the input process is Poisson with parameter λ, let $S^* \subseteq S^-$ be the set of states such that $R(S',S) = \lambda$ for $S' \varepsilon S^*$. Then the local balance condition is expressed as follows:

$$P(S) \sum_{S' \varepsilon U-S^+} R(S,S') = \lambda \sum_{S' \varepsilon S^*} P(S') + \sum_{S' \varepsilon U-S^+-S^*} P(S')\ R(S',S) \qquad (3)$$

The system of balance equations describes the state probabilities of a queue. If these state probabilities satisfy either the departure independence or the local balance conditions, they will satisfy both conditions,

[1] Local Balance, as originally defined by Chandy (C1), equated the rate at which a network leaves a state due to a customer arriving at a particular queue, to the rate at which the network enters the state due to a customer departing from the queue. The local balance defined here is analogous. It states that the rate at which the queue leaves a particular state due to a customer changing stage (including departures) equals the rate at which the queue enters the state due to a customer changing stage (including arrivals). This definition is consistent with that of (C2).

and the queue (the combination of queueing discipline and processing-time distribution) will be said to have the MM-property. Conversely, any state probabilities that satisfy both the local balance and departure independence conditions are the state probabilities of the queue.

If the departure independence conditions are satisfied for any processing-time distribution then the queueing discipline will be said to have the MM-property.

3. Symmetric Disciplines

Of the many forms of possible queueing algorithms, those that schedule on the basis of accumulated or expected processing time are excluded. Processor reassignments are in principle possible at times other than at arrivals and departures. But since accumulated processing times are not considered, periodic reassignments are the only remaining possibility. Periodic reassignments of processing power by various forms of the RR-algorithm are commonly used to time-share the processor, or to make it appear as though a single processor can be fractionally allocated, so the need for considering periodic reassignments is obviated by the capability of non-integral processing rates. Last, the queueing algorithms considered are restricted to the simple but broad subclass of algorithms called symmetric algorithms. These are defined as follows.

> Definition: A symmetric algorithm is a rule for allocating processing power to the customers in a queue that has the following properties:
>
> a) Processing rates are assigned to the customers in the queue only when there is a change in the number of customers in the queue.
>
> b) For each $n>0$, the processing rate distribution over

the n customers in the queue, ordered by their relative arrival times, is constant.

In a symmetric discipline, if the last customer to arrive receives service and departs before any other arrivals or departures, the remaining customers continue processing at the same rates as they did before the last arrival. PS, LCFSP, and NQ are symmetric disciplines.

4. Sufficiency of the BPS-discipline for Departure Independence and Local Balance

A general rule for allocating processing power to the customers in a queue, called the Batch-Processor Sharing (BPS) rule, is defined. Disciplines that allocate processing power according to this rule are called BPS-disciplines. In this section, it is shown that BPS-disciplines have the local balance and departure independence properties.

Definition: A processing-rate sequence $\{g_0, g_1, g_2 \ldots\}$ is a sequence representing any mapping from the non-negative integers to the non-negative reals, with $g_0 = 1$ and $g_n \leq n$, $n > 0$. For $n > 0$, g_n is the total number of processors available when there are n customers in the queue.

Definition: An E-sequence $E = \{e_1, e_2 \ldots e_s \ldots\}$ is an arbitrary finite or infinite strictly increasing sequence of integers with $e_1 = 1$.

Definition: A BPS-discipline is a rule for allocating processing power to customers in a queue such that for processing-rate sequence $\{g_n\}$ and some E-sequence E, when there are n customers in the queue and $e_s = \max\{e \mid e \in E \;\&\; e \leq n\}$, each of the last $n - e_s + 1$ customers that arrived receive processing service at the rate $\frac{g_n}{n - e_s + 1}$, and the remainder do not receive service.

It is obvious that if $g_n = 1$ for $n \geq 0$, when $E = \{1\}$, the BPS-discipline described is PS, and when $E = \{1,2,3...\}$, the BPS-discipline described is LCFSP. If $g_n = n$ for $n > 0$, and $E = \{1\}$, the BPS-discipline described is NQ.

Theorem 1

Any BPS-discipline has the departure independence property for any processing-time distribution with a rational Laplace Transform.

Proof: It is well known that distributions with rational Laplace Transforms can be represented by an exponential network (B1, C1). A general exponential network has k stages; processing is exponentially distributed with parameter μ_i at each stage $1 \leq i \leq k$. Processing terminates after stage i with probability $1-a_i$ and continues in stage i+1 with probability a_i.

A BPS-discipline is characterized by a set of processor-sharing intervals described by $E = \{e_1, e_2 ... e_s ...\}$. If there are n customers in the queue, and

$$e_s = \max\{e \mid e \in E \;\&\; e \leq n\},$$

let $n_s = n - e_s + 1$ be the number of customers in the s^{th} interval. For $j < s$, let $n_j = e_{j+1} - e_j$ be the number of customers in the j^{th} interval. For $1 \leq j \leq s$, $1 \leq i \leq k$, define $n_{j,i}$ to be the number of customers of interval j that are in stage i of processing. Within a processor-sharing interval, there is no distinction between customers that are in the same stage of processing, since they all receive processing at the same rate. Therefore, the state of queue with a BPS-discipline, in which the processing-time distribution is a k-stage exponential, is the vector

$$(n_{s,1}, n_{s,2} \cdots n_{s,k}, n_{s-1,1} \cdots n_{s-1,k}, \cdots n_{1,1}, \cdots n_{1,k}).$$

The departure independence and local balance equations for the queue are the following:

Departure Independence

If $n+1 < e_{s+1}$ or $e_s = \max\{E\}$,

$$P(n_{s,1}\cdots n_{s,k},\cdots n_{1,1}\cdots n_{1,k})\ \lambda$$

$$= g_{n+1} \sum_{j=1}^{k} P(n_{s,1}\cdots n_{s,j}+1\cdots n_{s,k},\cdots n_{1,1}\cdots n_{1,k})\ \frac{n_{s,j}+1}{n_s+1}\ \mu_j(1-a_j). \qquad (4a)$$

And, for $n+1 = e_{s+1}$,

$$P(n_{s,1}\cdots n_{s,k},\cdots n_{1,1}\cdots n_{1,k})\ \lambda$$

$$= g_{n+1} \sum_{j=1}^{k} P(n_{s+1,1}\cdots n_{s+1,j}\cdots n_{s+1,k},n_{s,1}\cdots n_{1,1}\cdots n_{1,k})\ \mu_j(1-a_j) \qquad (4b)$$

where $n_{s+1,i} = \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j \end{cases}$.

Local Balance

$$P(n_{s,1}\cdots n_{s,k},\cdots n_{1,1},\cdots n_{1,k})\ g_n \sum_{i=1}^{k} \frac{n_{s,i}\mu_i}{n_s}$$

$$= \lambda P(n_{s,1}-1,n_{s,2}\cdots n_{s,k},\cdots n_{1,1}\cdots n_{1,k})\ \alpha(n_{s,1})\ \alpha(n_s-1)$$

$$+ \lambda P(n_{s-1,1},n_{s-1,2}\cdots n_{s-1,k},\cdots n_{1,1}\cdots n_{1,k})\ \alpha(n_{s,1})(1-\alpha(n_s-1))$$

$$+ g_n \sum_{j=2}^{k} P(n_{s,1},\cdots n_{s,j-1}+1,n_{s,j}-1,\cdots n_{s,k},\cdots n_{1,1},\cdots n_{1,k})\ \frac{n_{s,j-1}+1}{n_s}\ \mu_{j-1}\ a_{j-1}\ \alpha(n_{s,j}) \qquad (5)$$

where $\alpha(n) = \begin{cases} 0 \text{ if } n = 0 \\ 1 \text{ if } n > 0. \end{cases}$

The form for the state probabilities of a BPS-discipline is then

$$P(n_{s,1}\cdots n_{s,k},\cdots n_{1,1},\cdots n_{1,k})$$

$$= P_o\ \frac{\rho^n}{G(n)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{\mu A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] \qquad (6)$$

where $\rho = \frac{\lambda}{\mu}$, λ = the average arrival rate for the Poisson input process,

$$G(n) = \prod_{i=1}^{n} g_n, \quad P_o^{-1} = \sum_{i=0}^{\infty} \frac{\rho^n}{G(n)}, \quad A_i = \prod_{j=1}^{i-1} a_j, \text{ and } \mu^{-1} = \sum_{j=1}^{k} \frac{A_j}{\mu_j}.$$

The theorem is proven by showing that (6) satisfies the departure independence (4) and local balance (5) equations of the queue. Then (6) satisfies the balance equation, hence describes the state probabilities, and the queueing discipline has the MM property. The details are in Appendix A.

5. Necessity of the BPS Processor Allocation Rule for Departure Independence and Local Balance in Symmetric Disciplines

In this section, it is shown that the BPS processor allocation rule is the only rule that provides the departure independence and local balance properties in a symmetric discipline.

Theorem 2

Any symmetric algorithm for allocating processing power to the customers in a queue that affords the departure independence and local balance properties when the processing-time distribution has a rational Laplace Transform, must be a BPS-discipline.

Proof: Suppose the customers entering a general queue are numbered by the inverse of their relative time of arrival; the most recent arrival is numbered one. The state of the queue can then be described by the vector $(k_1 \ldots k_n)$ when there are n customers in the queue, where k_i, $1 \le i \le n$, is the processing stage of the i^{th} most recent customer to enter the queue. $P_{k_1 \ldots k_n}$ is the probability of this state.

Let $\{r_{n,i}\}$, $1 \le i \le n$, be the service rate of the i^{th} customer when n customers are in the station. Necessarily, $\sum_{i=1}^{n} r_{n,i} = g_n$.

It must be shown that for some processing-time distribution, the only values for $\{r_{n,i}\}$ for all n, that satisfy the departure independence and local balance equations, are those that are specified by a BPS-discipline. An exponential distribution will not serve, since it affords these properties under any discipline. Therefore, a two-stage exponential network is chosen. The processing rates of the two stages are μ_1 and μ_2. With probability 1-a, service terminates after the first stage; with probability a, it enters the second stage. The general balance equation for this distribution is

$$P_{k_1\ldots k_n}(\lambda + \sum_{i=1}^{n} r_{n,i}\mu_{k_i}) = \lambda P_{k_2\ldots k_n}\ \delta(k_1,1)$$

$$+ \sum_{i=1}^{n} P_{k_1\ldots k_{i-1}\,1\,k_{i+1}\ldots k_n}\ r_{n,i}\ \mu_1\ a\ \delta(k_i,2)$$

$$+ \sum_{i=1}^{n+1} (P_{k_1\ldots k_{i-1}\,1\,k_i\ldots k_n}\ \mu_1(1-a) + P_{k_i\ldots k_{i-1}\,2\,k_i\ldots k_n}\ \mu_2)\ r_{n+1,i} \qquad (7)$$

where $\delta(x,y) = 1$, if $x = y$, and $\delta(x,y) = 0$ otherwise.

To simplify the notation, a normalized probability $Q_{k_1\ldots k_n}$ is defined as follows:

$$Q_{k_1\ldots k_n} = \frac{G(n)P_{k_1\ldots k_n}}{a^j\left(\frac{\lambda}{\mu_1}\right)^n P_o} \qquad (8)$$

where $j = \sum_{i=1}^{n} (k_i - 1)$ is the number of customers in stage two, $G(n) = \prod_{i=1}^{n} g_i$, where g_n is the processing power when there are n customers, and P_o is the probability of no customers in the queue.

Also, the normalized total processing rate for state $(k_1 \ldots k_n)$ is defined as

$$R_{k_1 \ldots k_n} = \frac{1}{\mu_1} \sum_{i=1}^{n} r_{n,i}\, \mu_{k_i} . \tag{9}$$

Also, let $z = \dfrac{\mu_2}{\mu_1}$.

The general balance equation can now be rewritten

$$Q_{k_1 \ldots k_n} \left(\frac{\lambda}{\mu_1} + R_{k_1 \ldots k_n} \right) = g_n\, Q_{k_2 \ldots k_n}\, \delta(k_1, 1)$$

$$+ \sum_{i=1}^{n} Q_{k_1 \ldots k_{i-1}\, 1\, k_{i+1} \ldots k_n}\, r_{n,i}\, \delta(k_i, 2)$$

$$+ \frac{1}{g_{n+1}} \sum_{i=1}^{n+1} \left[Q_{k_1 \ldots k_{i-1}\, 1\, k_i \ldots k_n} \frac{\lambda}{\mu_1} (1-a) + Q_{k_1 \ldots k_{i-1}\, 2\, k_i \ldots k_n} \frac{\lambda}{\mu_1}\, az \right] r_{n+1,i} . \tag{10}$$

The departure independence equation is

$$Q_{k_1 \ldots k_n} = \frac{1}{g_{n+1}} \sum_{i=1}^{n+1} \left[Q_{k_1 \ldots k_{i-1}\, 1\, k_i \ldots k_n} (1-a) + Q_{k_1 \ldots k_{i-1}\, 2\, k_i \ldots k_n}\, az \right] r_{n+1,i} \tag{11}$$

and the local balance equation is

$$Q_{k_1 \ldots k_n} R_{k_1 \ldots k_n} = g_n\, Q_{k_2 \ldots k_n}\, \delta(k_1, 1) + \sum_{i=1}^{n} Q_{k_1 \ldots k_{i-1}\, 1\, k_{i+1} \ldots k_n}\, \delta(k_i, 2)\, r_{n,i} . \tag{12}$$

It should be noted that (12) is a set of recursive equations defining the normalized probabilities $Q_{k_1 \ldots k_n}$. Since $Q_o = 1$, and a does not appear in (12), the $Q_{k_1 \ldots k_n}$ are independent of a.

Let 1^n represent the state $(k_1 \dots k_n)$, where $k_i = 1$, $1 \le i \le n$, and let $1^{j-1}21^{n-j}$ represent the state $(k_1 \dots k_n)$ where $k_j = 2$ and $k_i = 1$, $i \ne j$.

Then, from (12), the following special cases are obtained:

$$\text{For } n \ge 1,\ Q_{1^n} R_{1^n} = g_n Q_{1^{n-1}}\,. \tag{13}$$

$$\text{For } n > 1,\ Q_{21^{n-1}} R_{21^{n-1}} = Q_{1^n} r_{n,1}\,. \tag{14}$$

And, for $2 \le k \le n$,

$$Q_{1^{k-1}21^{n-k}} R_{1^{k-1}21^{n-k}} = g_n Q_{1^{k-2}21^{n-k}} + Q_{1^n} r_{n,k}\,. \tag{15}$$

In fact, since $R_{1^n} = g_n$, $Q_{1^n} = 1$ for all n.

At this point, it is convenient to prove the following lemma.

<u>Lemma 1</u>:

Suppose processing rates are assigned according to a BPS-discipline with E-sequence E whenever there are n or fewer customers in the queue, and departure independence holds for the entire discipline.

If $j = n+1-\max\{e \mid e \in E \,\&\, e \le n\}$ is the number of customers receiving service, then for all $1 \le k \le j$,

$$Q_{1^{k-1}21^{n-k}} = \sum_{i=j+1-k}^{j} \frac{1}{i} \prod_{h=i}^{j} \frac{h}{h-1+z}\,. \tag{16}$$

Proof: The proof is by induction on j. If $j = 1$, $n = e_s \in E$, and by the BPS-discipline, $r_{n,1} = g_n$ and $R_{21^{n-1}} = z\,g_n$. Hence, by (14), $Q_{21^{n-1}} = \frac{1}{z}$.

For the inductive step, assume (16) holds when $j > 1$ is replaced by j-1. From the definition of j, n is replaced by n-1. Then (16) yields

$$Q_{1^{k-1}21^{n-k-1}} = \sum_{i=j-k}^{j-1} \frac{1}{i} \prod_{h=i}^{j-1} \frac{h}{h-1+z} \qquad \text{for } 1 \le k \le j-1\,, \tag{17}$$

which can be expressed

$$Q_{1^{k-2}21^{n-k}} = \sum_{i=j+1-k}^{j-1} \frac{1}{i} \prod_{h=i}^{j-1} \frac{h}{h-1+z} \qquad \text{for } 2 \le k \le j\,. \qquad (18)$$

The rule for allocating processing power when there are n customers and $j=n+1-e_s$ specifies

$$r_{n,k} = \frac{g_n}{j} \text{ and } R_{1^{k-1}21^{n-k}} = \frac{(j-1+z)\, g_n}{j} \qquad \text{for } 1 \le k \le j\,.$$

Substituting these values, and (18) into (15) yields

$$\begin{aligned} Q_{1^{k-1}21^{n-k}} &= \frac{j}{j-1+z}\left[\sum_{i=j+1-k}^{j-1} \frac{1}{i} \prod_{h=1}^{j-1} \frac{h}{h-1+z} + \frac{1}{j}\right] \\ &= \sum_{i=j+1-k}^{j-1} \frac{1}{i} \prod_{h=1}^{j} \frac{h}{h-1+z} + \frac{1}{j-1+z} \\ &= \sum_{i=j+1-k}^{j} \frac{1}{i} \prod_{h=1}^{j} \frac{h}{h-1+z}\,, \end{aligned}$$

which completes the proof of the lemma.

The departure independence equation (11) yields the following special case of a state with n-1 customers in the queue, and $1 \le k \le n-1$:

$$Q_{1^{k-1}21^{n-k-1}} = \frac{1}{g_n}\left(Q_{1^k21^{n-k-1}} \sum_{i=1}^{k} r_{n,i} + Q_{1^{k-1}21^{n-k}} \sum_{i=k+1}^{n} r_{n,i}\right)(1-a)$$

$$+ \frac{1}{g_n}\left(\sum_{i=0}^{k} Q_{1^i21^{k-i-1}21^{n-k-1}}\, r_{n,i} + \sum_{i=k+1}^{n} Q_{1^{k-1}21^{i-k-1}21^{n-i}}\, r_{n,i}\right) a z\,. \qquad (19)$$

But, since each of the Q's is independent of a, (19) can be satisfied only if the coefficients of each of a and 1-a equal the left side of the equation.

Equating the coefficient of 1-a in (19) to $Q_{1^{k-1}21^{n-k-1}}$, for the case $k = 1$, and applying (14) and (15),

$$Q_{21^{n-2}} = \frac{g_n Q_{21^{n-2}} + r_{n,2}}{g_n R_{121^{n-2}}} \, r_{n,1} + \frac{r_{n,1}}{g_n R_{21^{n-1}}} \sum_{i=2}^{n} r_{n,i} \tag{20}$$

which can be rewritten

$$Q_{21^{n-2}} = r_{n,1} \frac{Q_{21^{n-2}} + \frac{r_{n,2}}{g_n}}{g_n - r_{n,2} + r_{n,2} z} + (g_n - r_{n,1}) \frac{\frac{r_{n,1}}{g_n}}{g_n - r_{n,1} + r_{n,1} z} \; . \tag{21}$$

This has the obvious solution $r_{n,1} = g_n$ and $r_{n,2} = 0$. Other solutions depend on the form of the Q's.

Equating the coefficient of 1-a in (19) to $Q_{1^{k-1}21^{n-k-1}}$ for $k > 1$ and applying (15),

$$Q_{1^{k-1}21^{n-k-1}} = \frac{\sum_{i=1}^{k} r_{n,i}}{g_n R_{1^k 21^{n-k-1}}} \left(g_n Q_{1^{k-1}21^{n-k-1}} + Q_{1^n} r_{n,k+1}\right)$$
$$+ \frac{\sum_{i=k+1}^{n} r_{n,i}}{g_n R_{1^{k-1}21^{n-k}}} \left(g_n Q_{1^{k-2}21^{n-k}} + Q_{1^n} r_{n,k}\right) . \tag{22}$$

Let $S_{n,k} = \sum_{i=1}^{k} r_{n,i}$. Then, (22) can be rewritten

$$Q_{1^{k-1}21^{n-k-1}} = S_{n,k} \frac{Q_{1^{k-1}21^{n-k-1}} + \frac{r_{n,k+1}}{g_n}}{g_n - r_{n,k+1} + r_{n,k+1} z}$$
$$+ (g_n - S_{n,k}) \frac{Q_{1^{k-2}21^{n-k}} + \frac{r_{n,k}}{g_n}}{g_n - r_{n,k} + r_{n,k} z} \; , \tag{23}$$

which has an obvious solution $S_{n,k} = g_n$ and $r_{n,k+1} = 0$. Other solutions depend on the form of the Q's.

It is now shown by induction on n that the BPS-discipline is the only rule for allocating processing power that satisfies the departure independence equations (21) and (23). For $n=1$, this is trivial, because $r_{1,1}=g_1$ must be true for any discipline.

Let $E=\{e_1 \ldots e_s\}$ with $e_1=1$ and $e_s<n$ be the partial E-sequence of a BPS-discipline. For the inductive hypothesis, assume that for all $m<n$, the only processing rates that satisfy (19) are $r_{m,i}=\frac{g_m}{j}$, $1 \le i \le j$ and $r_{m,i}=0$, $j<i\le m$, where $j=m+1-\max\{e \mid e \in E \;\&\; e<m\}$,

In particular, the processing rates for n-1 customers are $r_{n-1,i}=\frac{g_{n-1}}{j-1}$, $1 \le i \le j-1$ and $r_{n-1,i}=0$, $j \le i \le n$, where $j=n+1-e_s$. Hence, $R_{21^{n-2}}=g_{n-1}(1-\frac{1}{j-1}+\frac{z}{j-1})$, and by (14) $Q_{21^{n-2}}=\frac{1}{j-2+z}$. Then (21) is reduced to the polynomial in z

$$g_n = r_{n,1}\,\frac{g_n+(j-2+z)\,r_{n,2}}{g_n-r_{n,2}+r_{n,2}\,z} + (g_n-r_{n,1})\,\frac{(j-2+z)\,r_{n,1}}{g_n-r_{n,1}+r_{n,1}\,z}\,, \tag{24}$$

which must be satisfied for all z. This equation is straightforwardly reduced to the form

$$r_{n,2}\,g_n\left[j\,r_{n,1}-g_n\right]z+(g_n-r_{n,1})\,g_n\left[(j-1)\,r_{n,1}+r_{n,2}-g_n\right]=0\,.$$

Setting the coefficient of z equal to zero yields the solutions $r_{n,2}=0$ or $r_{n,1}=\frac{g_n}{j}$. Setting the constant term equal to zero yields the solutions $r_{n,1}=g_n$ or $r_{n,2}=g_n-(j-1)\,r_{n,1}$. Since both conditions must be satisfied,

$$(r_{n,1},r_{n,2})=(g_n,0) \quad \text{or} \quad (\frac{g_n}{j},\frac{g_n}{j})$$

are the only solutions satisfying the departure independence equation (21) for n customers.

If the first solution is chosen, then $r_{n,i}=0$ for $2 \le i \le n$. This assignment of processing rates corresponds to a BPS-discipline with the partial E-sequence $E=\{e_1,\ldots e_s,e_{s+1}\}$ where $e_{s+1}=n$.

If the second solution is chosen, induction on k will show that $r_{n,k} = \frac{g_n}{j}$ for $1 \le k \le j$ are the only remaining rate solutions consistent with this choice. Choosing the second solution then corresponds to a BPS-discipline with $e_{s+1} \neq n$ in the e-sequence. The choice of $r_{n,1} = r_{n,2} = \frac{g_n}{j}$ forms the basis of the induction.

For the inductive hypothesis, assume

$$S_{n,k} = \frac{kg_n}{j} \quad \text{and} \quad r_{n,k} = \frac{g_n}{j} .$$

By the hypothesis of the induction on n, it has been assumed that a BPS-discipline is used for rate assignments for n-1 or fewer customers. Then, the state probabilities $Q_{1^{k-1}21^{n-k-1}}$ and $Q_{1^{k-2}2^{n-k}}$ are given by Lemma 1. Because these are states with n-1 customers, their forms are given explicitly in (17) and (18). Using these values and the above rates in the departure independence equation (23), results in a polynomial in z, which must be satisfied for all z. The possible solutions for $r_{n,k+1}$ may be obtained by letting $z = 1$. In this case, (17) and (18) simplify to

$$Q_{1^{k-1}21^{n-k-1}} = \sum_{i=j-k}^{j-1} \frac{1}{i}, \quad 1 \le k \le j-1$$

and

$$Q_{1^{k-2}21^{n-k}} = \sum_{i=j+1-k}^{j-1} \frac{1}{i}, \quad 2 \le k \le j .$$

Substituting these values into (23), with $z = 1$, yields, for $2 \le k < j$,

$$j \sum_{i=j-k}^{j-1} \frac{1}{i} = k \left[\sum_{i=j-k}^{j-1} \frac{1}{i} + \frac{r_{n,k+1}}{g_n} \right] + (j-k) \left[\sum_{i=j+1-k}^{j-1} \frac{1}{i} + \frac{1}{j} \right]. \tag{25}$$

This straightforwardly reduces to the unique solution $r_{n,k+1} = \frac{g_n}{j}$.

Since, if the processing rates are chosen with $e_{s+1} \neq n$, the solutions are $r_{n,i} = \frac{g_n}{j}$ for $1 \leq i \leq j$, then $r_{n,i} = 0$, $j < i \leq n$. Therefore, for departure independence, the processing rates with n customers are necessarily assigned according to a BPS-discipline and completes the proof of the theorem.

6. Generalizations and Conclusion

The class of Batch-Processor-Sharing algorithms is defined as a subset of the class of symmetric disciplines, all of which employ a static processor allocation rule. This rule states that for each n, the distribution of processing power to each of the last n customers to enter the queue, ordered by their arrival time, must be constant. However, relaxing the symmetric requirement of the BPS-discipline, to allow dynamic variation of the processor-allocation rule, will not necessarily eliminate the MM-property. It has been shown that for the MM-property in symmetric disciplines, every incoming customer must be assigned to a processor-sharing class, within which each customer receives an equal amount of service. The only difference between processor-sharing in one of the BPS classes, and that of the PS-discipline demonstrated in (C2), is that the BPS customers are ordered by their arrival times. But, because all the customers in a BPS class receive equal service, this ordering is irrelevant to any external, or input-output, view of processing in the class. The jobs of the class may be randomly permuted among the stations of the BPS class, without disturbing the input-output characteristics. In other words, the BPS classes are memoryless. As a consequence, if a BPS class is interrupted at some random time, as for example at the arrival of a new customer, the BPS class will continue to provide a Poisson output process when its processing is resumed after the interrupt.

As long as the discipline used during the period of the interrupt results in a Poisson output process, the length of the interrupt makes no difference. Therefore, assume BPS classes $\{C_1, \ldots C_n\}$ have been filled with customers, and the last customer of class C_{n+1} departs. Then, resuming processing with any one of the remaining n BPS classes will preserve the Poisson output process. The preemptive round-robin discipline, in which a new customer always preempts the customer on the processor, causing him to be placed at the end of the processor queue, is an example of such a discipline. It is a generalization of the symmetric BPS algorithm.

As another consequence of the memoryless property of the BPS classes, the size of any BPS class may vary dynamically. A BPS class may once be preempted in favor of initiating a new BPS class on the arrival of the n^{th} customer, but after processing of the class is resumed, it may be allowed to accept a total of n or more customers. The decision to either continue the current class or to initiate a new BPS class may be made by the priority of the new customer, as long as arrivals of the high priority customers occur in a Poisson process within the total Poisson input process. This allows the design of MM-disciplines in which a favored class of customers will have a smaller wait time than other customers.

Network analysis is greatly simplified when the queues of the network have the MM-property. Because algorithms in which the processing power of a queue is variable, and may retain the MM-property, it is possible to analyze networks that are designed to achieve maximum throughput through allocating greater processing power to the queues with the greatest number of customers. This reflects a practical design approach for computer networks, since the servers in the network may be general-purpose processors that are capable of workload tradeoffs.

Proof of Departure Independence

Substituting (6) into (4a) for the case $n+1 < e_{s+1}$ or $e_s = \max\{e \mid e \in E\}$,

$$\frac{P_o \rho^n \mu^n}{G(n)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] \lambda$$

$$= g_{n+1} \sum_{j=1}^{k} \frac{P_o \rho^{n+1} \mu^{n+1}}{G(n+1)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] (n_s+1) \frac{\left(\frac{A_j}{\mu_j}\right)}{n_{s,j}+1} \frac{n_{s,j}+1}{n_s+1} \mu_j (1-a_j) .$$

The terms $\frac{P_o \rho^{n+1} \mu^{n+1}}{G(n)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right]$ cancel, leaving

$1 = \sum_{j=1}^{k} A_j (1-a_j)$, which, since $A_1 = 1$ and $a_k = 0$, and $A_{j+1} = A_j a_j$, is an identity.

Substituting (6) into (4b) for the case $n+1 = e_{s+1}$,

$$\frac{P_o \rho^n \mu^n}{G(n)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] \lambda$$

$$= g_{n+1} \sum_{j=1}^{k} \frac{P_o \rho^{n+1} \mu^{n+1}}{G(n+1)} \frac{A_j}{\mu_j} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] \mu_j (1-a_j) .$$

Cancelling $\frac{P_o \rho^{n+1} \mu^{n+1}}{G(n)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right]$ leaves

$1 = \sum_{j=1}^{k} A_j (1-a_j)$, which is an identity. Q.E.D.

Proof of Local Balance

Substituting (6) into (5) yields

$$\frac{P_o \rho^n \mu^n}{G(n)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] g_n \sum_{j=1}^{k} \frac{n_{s,j}\mu_j}{n_s}$$

$$= \lambda \frac{P_o \rho^{n-1} \mu^{n-1}}{G(n-1)} (n_s - 1)! \left[\prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{s,i}}}{n_{s,i}!} \right] \frac{n_{s,1}}{\left(\frac{A_1}{\mu_1}\right)} \prod_{h=1}^{s-1} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] \alpha(n_{s,1}) \ \alpha(n_s - 1)$$

$$+ \lambda \frac{P_o \rho^{n-1} \mu^{n-1}}{G(n-1)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] \frac{1}{\left(\frac{A_1}{\mu_1}\right)} \alpha(n_{s,1}) \ [1 - \alpha(n_s - 1)]$$

$$+ g_n \sum_{j=2}^{k} \frac{P_o \rho^n \mu^n}{G(n)} n_s! \left[\prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{s,i}}}{n_{s,i}!} \right] \frac{\left(\frac{A_{j-1}}{\mu_{j-1}}\right)}{n_{s,j-1}+1} \frac{n_{s,j}}{\left(\frac{A_j}{\mu_j}\right)} \prod_{h=1}^{s-1} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right] \frac{n_{s,j-1}+1}{n_s} \mu_{j-1} a_{j-1} \alpha(n_{s,j})$$

The terms $\frac{P_o \rho^n \mu^n}{G(n-1)} \prod_{h=1}^{s} \left[n_h! \prod_{i=1}^{k} \frac{\left(\frac{A_i}{\mu_i}\right)^{n_{h,i}}}{n_{h,i}!} \right]$ can be cancelled, leaving

$$\sum_{j=1}^{k} \frac{n_{s,j}\mu_j}{n_s} = \frac{1}{n_s} \frac{n_{s,1}}{\left(\frac{A_1}{\mu_1}\right)} \alpha(n_{s,1}) \ \alpha(n_s - 1) + \frac{1}{\left(\frac{A_1}{\mu_1}\right)} \alpha(n_{s,1}) \ [1 - \alpha(n_s - 1)]$$

$$+ \sum_{j=2}^{k} \frac{A_{j-1}}{\mu_{j-1}} \ \frac{n_{s,j}}{n_{s,j-1}+1} \ \frac{\mu_j}{A_j} \ \frac{n_{s,j-1}+1}{n_s} \ \mu_{j-1} \ a_{j-1} \ \alpha(n_{s,j}) \ .$$

But $A_1 = 1$ and $A_{j-1}\, a_{j-1} = A_j$, hence

$$\sum_{j=1}^{k} \frac{n_{s,j}\,\mu_j}{n_s} = \left[\frac{n_{s,1}}{n_s}\, \mu_1\ \alpha(n_s - 1) + \mu_1 [1 - \alpha(n_s - 1)] \right] \alpha(n_{s,1})$$

$$+ \sum_{j=2}^{k} \frac{n_{s,j}}{n_s}\, \alpha(n_{s,j})\ \mu_j \,.$$

This yields the identity

$$\sum_{j=1}^{k} \frac{n_{s,j}\,\mu_j}{n_s} = \frac{n_{s,1}}{n_s}\, \mu_1 + \sum_{j=2}^{k} \frac{n_{s,j}}{n_s}\, \mu_j \qquad \text{Q.E.D.}$$

References

B1. Baskett, F., K. M. Chandy, R. R. Muntz and F. Palacios-Gomez, "Open, closed and mixed networks of queues with different classes of customers", to appear in JACM.

C1. Chandy, K. M., "The analysis and solutions for general queueing networks", Proc. Sixth Annual Princeton Conf. on Information Sciences and Systems, Princeton Univ., Princeton, N. J., (March 1972), pp. 219-224.

C2 Chandy, K. M., Howard, J. H., and Towsley, D. F.,"Product Form and local Balance in Queueing Networks," to appear in JACM.

C3 Cox, D. R., "A use of complex probabilities in the theory of stochastic processes", Proc. Camb. Phil. Soc., 51 (1955), pp. 313-319.

G1 Gordon, W. J. and G. F. Newell, "Closed queueing systems with exponential servers", Oper. Res. 15, 2 (1967), pp. 252-267.

J1 Jackson, J. R., "Job-shop like queueing systems", Man. Sci., 10, (1963), pp. 131-142.

M1 Mirasol, N. M., "The Output of an M/G/∞ Queue in Poisson", Oper. Res. 11, (1968), pp. 282-284.

M2 Muntz, R. R., "Poisson departure processes and queueing networks", IBM Research Report, RC-4145, Yorktown Heights, New York, 1972.