QUEUING WITH STRICT AND LAG PRIORITY MIXTURES

Files d'Attente de Priorité Stricte et Non-Stricte

LEONARD KLEINROCK

University of California, Los Angeles
United States of America

1. INTRODUCTION

Many priority queuing disciplines have been studied recently. Saaty's [1] summary of much of the progress up to 1961 includes the work of Cobham [2] and White and Christie [3] on strict head-of-the-line priority systems. Kleinrock [4] has analyzed a lag (or delay-dependent) queuing discipline. The characteristic of the strict system is that it can operate in a saturated condition while still giving finite waiting time to the higher priority groups; this system, however, allows no freedom in adjusting the relative waiting time among the groups. The lag system introduces a set of parameters into the model which allows manipulation of the relative waiting times; but the previous analysis of the system allowed operation only in a stable mode (i.e., if any group experienced an infinite average wait, then so did all the groups).

We recognize that the lag system has the advantage of adjustability of the relative waiting times; as developed in [4], it has the unfortunate disadvantage of causing all groups to suffer an unbounded waiting time whenever the system is saturated. On the other hand, the strict priority system allows finite average waits for high-priority groups beyond saturation but permits no adjustment of the relative delays. In this article we adjust the parameters in the lag system to allow finite waiting times for the higher priority groups under saturated conditions. In so doing, we develop an interesting priority queuing discipline that combines the desirable features (i.e., operation beyond saturation and adjustment of relative waiting times) from both of the earlier systems. The

This work was carried out under the author's consulting contract with T.R.W. Space Technology Laboratiories.

result is what we call a strict and lag priority mixture (SLPM). The SLPM system produces a number of strict priority groups, each of which consists of a number of lag priority subgroups. An example is given and the family of average waiting time curves is presented.

2. THE MODEL

We assume that we have P separate groups of users of which the pth group ($p = 2, 3, 4, \ldots, P$) is to be given preferential treatment over the $p-1$st group. At any time the priority of a particular unit is calculated according to a set of parameters assigned to that unit; the higher the value obtained by this function, the higher the priority. The sequence of arrivals forms a Poisson process in which the average number of arrivals per second from the pth group is λ_p. We assume that the service times are exponentially distributed with an average of $1/\mu_p$ sec for units from group p.

We define a pre-emptive priority system as one that removes a unit from the service facility as soon as another unit of higher priority appears in the queue. When the unit that was removed is returned to the service facility, it picks up from the point at which it was interrupted. A nonpre-emptive priority system is one that always allows a unit to complete its service, once that unit has begun service.

We define W_p = average time spent in the queue for a unit from group p.

$$\lambda = \sum_{p=1}^{P} \lambda_p, \tag{1}$$

$$\frac{1}{\mu} = \sum_{p=1}^{P} \frac{\lambda_p}{\lambda} \frac{1}{\mu_p}, \tag{2}$$

$$\rho_p = \frac{\lambda_p}{\mu_p}, \tag{3}$$

$$\sigma_p = \sum_{i=p}^{P} \rho_i, \tag{4}$$

$$\rho = \sigma_1 = \sum_{p=1}^{P} \rho_p = \frac{\lambda}{\mu}, \tag{5}$$

$$W_0 = \sum_{p=1}^{P} \frac{\rho_p}{\mu_p}. \tag{6}$$

3. STRICT PRIORITY SYSTEM

A common priority system (which we refer to as a *strict* priority system) is one in which each unit from group p is assigned a fixed value of priority equal to p. Within the pth group, a first-come-first-served ordering is used. Thus a member of the pth group will always be taken into service before a member of group p', where $p' < p$. The behavior of this system has been studied by Cobham in the nonpre-emptive case and by White and Christie in the pre-emptive case.

Theorem 1 (*Cobham*). *In the nonpre-emptive strict priority system we have for* $0 \leq \rho < 1$

$$W_p = \frac{W_0}{(1-\sigma_{p+1})(1-\sigma_p)} \tag{7}$$

and for $\rho \geq 1$

$$W_p = \begin{cases} \dfrac{(1-\sigma_j)/\mu_{j-1} + \sum_{i=j}^{P} \rho_i/\mu_i}{(1-\sigma_{p+1})(1-\sigma_p)}, & p \geq j, \\ \infty, & p < j, \end{cases} \tag{8}$$

where j *is the smallest integer such that*

$$\sigma_j < 1.$$

An example of this family of curves is plotted in Figure 1.

Theorem 2 (*White and Christie*). *In the pre-emptive strict priority system we have for* $0 \leq \rho$

$$W_p = \begin{cases} \dfrac{\rho_p/\mu_p + \sum_{i=p+1}^{P} \rho_i(1/\mu_i) + (1/\mu_p)}{(1-\sigma_{p+1})(1-\sigma_p)}, & p \geq j, \\ \infty, & p < j, \end{cases} \tag{9}$$

where j *is defined as in Theorem 1.*

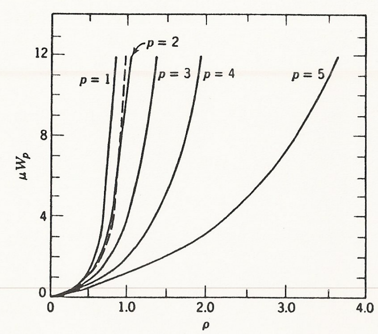


Figure 1. $\mu W_p(\rho)$ for the fixed-priority system with no preemption: $\lambda_p = \lambda/P$, $\mu_p = \mu$, $P = 5$.

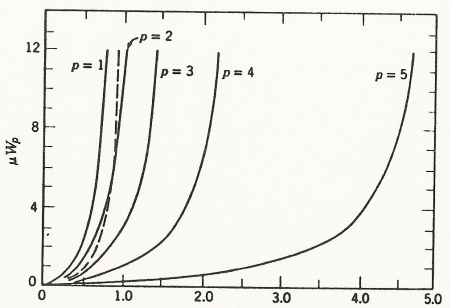


Figure 2. $\mu W_p(\rho)$ for the fixed-priority system with preemption: $\lambda_p = \lambda/P$, $\mu_p = \mu$, $P = 5$.

An example of this family of curves is plotted in Figure 2. From Figures 1 and 2 we see that the nonpre-emptive and pre-emptive systems are not significantly different in their behavior. Both families display the obvious discrimination that the system shows in favor of higher priority units; this is reflected in the uniformly shorter waiting times for the higher priority units. In these two figures (as well as in Figures 4 and 5) we show a dashed curve of the function $\rho/(1-\rho)$; this curve corresponds to the average

$$\frac{1}{W_0}\sum_{p=1}^{P}\rho_p W_p = \frac{\rho}{1-\rho}$$

which is invariant to a wide class of queue disciplines (see Kleinrock [5]).

An interesting behavior in the waiting time for $\rho > 1$ may be seen in these curves. Specifically, note that when the system is overloaded (i.e., $\rho > 1$), certain of the high-priority groups (namely, those for which $p \geq j$) experience only a finite delay in spite of the fact that there is an infinite queue. This is due to the fact that units from group p are sharing the service facility only† with units from groups $p' \geq p$. Thus the infinite queue is made up of units from groups for which $p < j$. The value of j increases to the value $j = p$ whenever the ρ_i increases such that

$$\rho = 1 + \sum_{i=1}^{p-1}\rho_i$$

(these points corresponding to $\sigma_p = 1$).

† In the pre-emptive case this is strictly true. In the nonpre-emptive case, an occasional unit from group $p' < p$ (where $p' \geq j - 1$) may be found in service, thus causing a slight coupling effect from lower priority groups also.

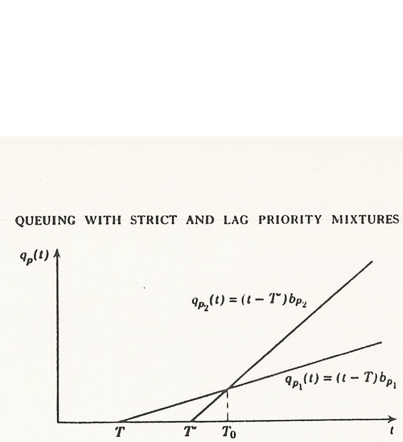


Figure 3. Interaction between priority functions for the delay-dependent priority system.

4. LAG (DELAY-DEPENDENT) PRIORITY SYSTEM

The strict priority system discussed suffers from a serious defect: the average waiting times W_p are completely determined once the arrival parameters (λ_p, μ_p) are specified. Consequently the system designer has no freedom in adjusting the relative waiting times among various priority groups. In this section we describe a lag (or delay-dependent) priority discipline that corrects this defect and enables the system designer to adjust the relative average waiting times of priority groups over a wide range.

In the lag system a unit from priority group p entering at time T is assigned a parameter b_p and its priority, $q_p(t)$ at time t, is calculated from

$$q_p(t) = (t - T)b_p \tag{10}$$

where

$$0 \leq b_1 \leq b_2 \leq \cdots \leq b_P. \tag{11}$$

As shown in Figure 3, all entering units begin with an initial priority of zero and gain priority linearly with time. A unit from priority group p_2 will be given preferential treatment over a unit from group $p_1 < p_2$ only after time T_0, as shown in the figure. Thus there is an interaction among all priority groups. This system has been studied by Kleinrock [4].

Theorem 3. In the nonpre-emptive lag priority system we have

$$W_p = \frac{W_0/(1-\rho) - \sum_{i=1}^{p-1}\rho_i W_i(1 - b_i/b_p)}{1 - \sum_{i=p+1}^{P}\rho_i(1 - b_p/b_i)}. \tag{12}$$

This family is plotted in Figure 4 for $\lambda_p = \lambda/P$, $\mu_p = \mu$, *and* $b_p = 2^{p-1}$.

Theorem 4. In the pre-emptive lag priority system we have

$$W_p = \frac{\dfrac{W_0}{1-\rho} + \displaystyle\sum_{i=p+1}^{P}\frac{\rho_i}{\mu_p}\left(1 - \frac{b_p}{b_i}\right) - \sum_{i=1}^{p-1}\frac{\rho_i}{\mu_i}\left(1 - \frac{b_i}{b_p}\right)\sum_{i=1}^{p-1}\rho_i W_i\left(1 - \frac{b_i}{b_p}\right)}{1 - \displaystyle\sum_{i=p+1}^{P}\rho_i\left(1 - \frac{b_p}{b_i}\right)}. \tag{13}$$

This family is plotted in Figure 5 for the same parameters as in Figure 4.

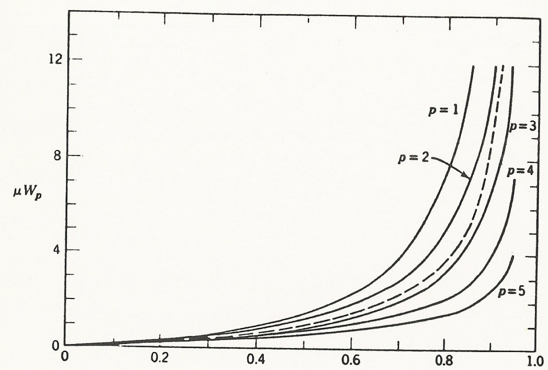
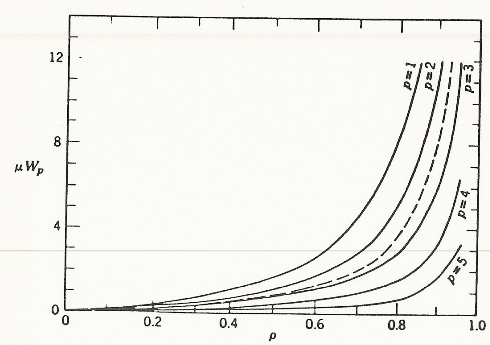


Figure 4. $\mu W_p(\rho)$ for the delay-dependent priority system with no preemption: $P = 5$.

The expressions for W_p in (12) and (13) are given recursively in terms of :he W_i for $i < p$. Thus each represents a set of P simultaneous linear equations n the W_p, where the sets are *triangular* and therefore trivial to solve. They ire expressed recursively only for simplicity of form.

In comparing Figures 4 and 5 we see again that the higher priority units iave a uniformly shorter waiting time than the lower priority units. Moreover, 'rom Theorems 3 and 4 we see that the values of the various W_p can be nodified by changing the free parameters b_p $(p = 1, 2, \ldots, P)$. With these



igure 5. $\mu W_p(\rho)$ for the delay-dependent priority system with preemption: $P = 5$.

degrees of freedom the designer can adjust the W_p to suit any relative waiting times that may be desired. We note that only the ratios b_i/b_j $(i < j)$ appear in the solutions. These curves, as well as the results, appear to imply that all W_p go to infinity at $\rho = 1$; this is an undesirable feature, for extremely high priority units must be able to get rapid service even under overload $(\rho > 1)$ conditions. A closer investigation of Theorems 3 and 4, however, shows that the behavior of finite delay for the higher priority groups (even under overload) can still be obtained by mixing the two priority systems discussed. For this purpose we introduce strict and lag priority mixtures.

5. STRICT AND LAG PRIORITY MIXTURES (SLPM)

Let us assume that we may wish to operate the system at some overloaded condition, say $\rho' \geq 1$. With this overload we insist that priority groups p_j, $p_{j+1}, \ldots, P$ experience a finite delay (and we are willing that groups $p_1, p_2, \ldots, p_{j-1}$ experience an infinite delay). Furthermore we wish to specify the relative waiting times of the various groups. In essence, then, we are asking for the desirable features of both the strict and lag priority systems (namely, the ability to operate beyond saturation and also to adjust the relative waiting times). We find that by judicious selection of the parameters b_p $(p = 1, 2, \ldots, P)$ we can achieve such a system, which we refer to as a strict and lag priority mixture (SLPM).

Specifically, we recognize that if $b_k/b_{k-1} \gg 1$ then all groups with $p \geq k$ will *always* be given preference over all groups with $p < k$ in the lag system, for the rate at which priority is attained is proportional to b_p [see (10)]. This system will look like a two-priority group, strict priority system, the lower of the two containing $k - 1$ subgroups which interact among themselves as a lag priority system and the higher of the two containing $P - k + 1$ subgroups which also interact among themselves as a lag priority system. If $b_k/b_{k-1} \gg 1$ for many (say, $M - 1$) k, we break the P priority groups into M strict priority groups, each of which is made up of subgroups that interact among themselves as lag priority groups. We define the parameters of an SLPM system, present the results for the average waiting times, and plot an example of this family of curves. Define

M = number of strict priority groups,
n_m = number of lag priority groups in the mth strict priority group $m = 1, 2, \ldots, M$,

$$N_m = \sum_{i=1}^{m} n_i, \tag{14}$$

$P = N_M$ = total number of priority groups,
λ_p = average arrival rate (Poisson) to pth group $p = 1, 2, \ldots, N_M$,

$\frac{1}{\mu_p}$ = average length of messages (exponential) in pth group $p = 1, 2, \ldots, N_M$,

$$\rho_p = \frac{\lambda_p}{\mu_p}, \tag{15}$$

$$\sigma_p = \sum_{i=p}^{N_M} \rho_p, \qquad [\sigma_1 \equiv \rho], \tag{16}$$

$$\beta_m = \sum_{i=N_{m-1}+1}^{N_m} \rho_i, \qquad [N_0 \equiv 0], \tag{17}$$

$$\gamma_m = \sum_{i=m}^{M} \beta_i = \sum_{i=N_{m-1}+1}^{N_M} \rho_i, \tag{18}$$

m^* = smallest positive integer such that $\gamma_{m^*} < 1$,

$$N^* = \sum_{i=1}^{m^*-1} n_i = N_{m^*-1}, \tag{19}$$

$$\bar{W}_m = \begin{cases} \dfrac{W_0}{(1-\gamma_m)(1-\gamma_{m+1})}, & \text{for } \rho < 1, \\ \dfrac{(1-\gamma_{m^*})/\mu_{N^*} + \sum_{i=N^*+1}^{N_M} \rho_i/\mu_i}{(1-\gamma_m)(1-\gamma_{m+1})}, & \text{for } \rho \geq 1. \end{cases} \tag{20}$$

Then, given a set of parameters b_p ($p = 1, 2, \ldots, N_M$) such that

$$0 \leq b_1 \leq b_2 \leq \cdots \leq b_{N_M} \tag{21}$$

and

$$\frac{b_{N_m+1}}{b_{N_m}} \gg 1, \qquad m = 1, 2, \ldots, M-1, \tag{22}$$

and given p, define $N_{m_{p-1}}$ and N_{m_p} as $N_{m_{p-1}} < p \leq N_{m_p}$ ($p = 1, 2, \ldots, N_M$).

Theorem 5†. *For an SLPM system with no pre-emption, the average waiting time W_p is*

$$W_p = \frac{1}{D_p}\left[C - \sum_{m=m^*}^{m_{p-1}} \bar{W}_m \beta_m - \sum_{i=N_{m_{p-1}}+1}^{p-1} \rho_i W_i \left(1 - \frac{b_i}{b_p}\right)\right], \tag{23}$$

where

$$D_p = 1 - \sigma_{N_{m_p}+1} - \sum_{i=p+1}^{N_{m_p}} \rho_i \left(1 - \frac{b_p}{b_i}\right) \tag{24}$$

and

$$C = W_0 + \frac{\sigma_{N^*+1}}{1-\sigma_{N^*+1}} \sum_{j=N^*+1}^{N_M} \frac{\rho_j}{\mu_j} + \frac{\sigma_{N^*+1}}{\mu_{N^*}}. \tag{25}$$

In Figure 6 we show this family of curves for $M = n_m = 5$, $\lambda_p = \lambda/25$, $1/\mu_p = 1/\mu$ ($p = 1, 2, \ldots, 25$; $m = 1, 2, \ldots, 5$), $b_{p+1}/b_p = 2$ for p and $p+1$ in the same strict priority group. This figure shows the characteristics we were looking for, namely, operation beyond saturation, and adjustability of the W_p. We note the strict priority groups, each composed of lag priority subgroups.

† See the appendix for proof of this theorem.

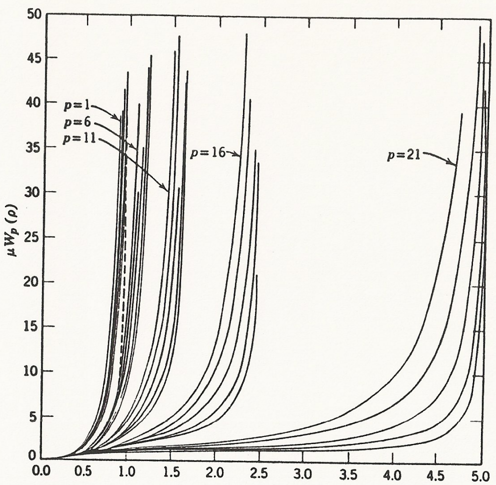


Figure 6. $\mu W_p(\rho)$ for the strict and lag priority mixtures (SLPM) with no preemption: $P = 25$, $M = 5$.

For the SLPM system with pre-emption we get the following:

Theorem 6†. *With pre-emption the SLPM system has an average waiting time W_p:*

$$W_p = \frac{1}{D_p}\left[C_1 + \frac{1}{\mu_p} - \sum_{m=m^*}^{m_{p-1}} \beta_m \bar{T}_m - \sum_{i=N_{m_{p-1}}+1}^{p-1} \rho_i \left(W_i + \frac{1}{\mu_i}\right)\left(1 - \frac{b_i}{b_p}\right)\right] - \frac{1}{\mu_p}, \tag{26}$$

where

$$\bar{T}_m = \frac{\overline{1/\mu_m}}{(1-\gamma_m)(1-\gamma_{m+1})}. \tag{27}$$

$$\overline{1/\mu_m} = \sum_{i=N_{m-1}+1}^{N_m} \frac{\lambda_i}{\lambda_m} \frac{1}{m_i}, \tag{28}$$

$$\lambda_m = \sum_{i=N_{m-1}+1}^{N_m} \lambda_i, \tag{29}$$

and

$$C_1 = \frac{1}{1-\sigma_{N^*+1}} \sum_{i=N^*+1}^{N_M} \rho_i \mu_i. \tag{30}$$

† See the appendix for *an outline of* the proof of this theorem.

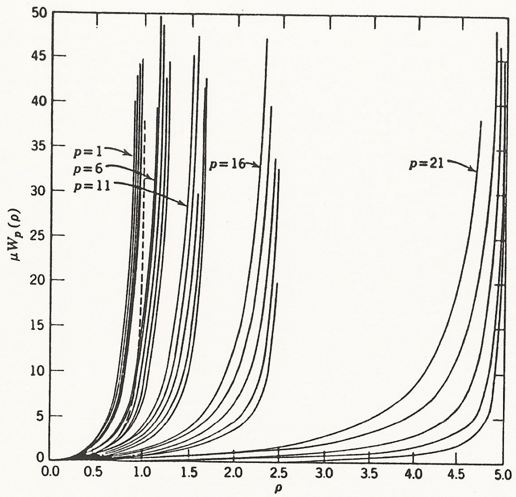


Figure 7. $\mu W_p(\rho)$ for the strict and lag priority mixtures (SLPM) with preemption: $P=25$, $M=5$.

This family of curves is plotted in Figure 7 for the same special parameters as in Figure 6. The difference between pre-emptive and nonpre-emptive SLPM systems is that in the latter the strict priority groups are more widely separated.

6. CONCLUSIONS

In this article we introduced the notion of a priority structure superimposed on the population of users. We then discussed two well-known priority systems, the strict priority system and the lag priority system (both with and without pre-emption). Curves of the typical performance of these systems are given in Figures 1, 2, 4, and 5. We observed that the strict priority system offered no degrees of freedom for varying the relative waiting times of the different priority groups and that the lag-priority system did not allow operation beyond saturation (i.e., heavy loads). Consequently we developed a new priority discipline, the strict and lag priority mixture (SLPM) which was shown to possess the desirable features of both of the other disciplines. Curves of the performance of the SLPM system are given in Figure 6 (for the nonpre-emptive system) and in Figure 7 (for the pre-emptive system). It is felt that the generality available in this new priority queuing discipline is great enough to satisfy the requirements of many desired priority structures.

7. REFERENCES

[1] T. L. Saaty, *Elements of Queuing Theory*, McGraw-Hill, New York, 1961.

[2] A. Cobham, "Priority Assignments in Waiting Line Problems," *Operations Res.*, 2, 70–76 (1954).

[3] H. White and L. S. Christie, "Queuing with Pre-emptive Properties or with Breakdown," *Operations Res.*, 6, 79–95 (1958).

[4] L. Kleinrock, "A Delay Dependent Queue Discipline," *Naval Res. Logistics Quart.*, 11, No. 4, 329–341 (December 1964).

[5] L. Kleinrock, "A Conservation Law for a Wide Class of Queuing Disciplines," *Naval Res. Logistics Quart.*, 12, No. 2 (June 1965).

[6] L. Kleinrock, *Communication Nets; Stochastic Message Flow and Delay*, McGraw-Hill, New York, 1964.

APPENDIX

In this appendix we give proofs for Theorems 5 and 6.

Proof (Theorem 5). Because we are dealing with a special case of the nonpre-emptive lag priority system, we may immediately write down the following equation which applies to such systems and which is developed in [6], p. 164:

$$W_p = \frac{1}{D'_p}\left(W_0 + \sum_{i=p}^{P} \rho_i W_i + \sum_{i=1}^{p-1} \rho_i W_i \frac{b_i}{b_p}\right), \tag{A.1}$$

where

$$D'_p = 1 - \sum_{i=p+1}^{P} \rho_i \left(1 - \frac{b_p}{b_i}\right). \tag{A.2}$$

Using the notation in Section 5 and recalling that $b_{N_m+1}/b_{N_m} \gg 1$ $(m = 1, 2, \ldots, M-1)$, we get for $N_{m_{p-1}} < p \le N_{m_p}$

$$W_p = \frac{1}{D'_p}\left(W_0 + \sum_{i=N_{m_{p-1}}+1}^{p-1} \rho_i W_i \frac{b_i}{b_p} + \sum_{i=p}^{P} \rho_i W_i\right) \tag{A.3}$$

and $D_{p'} = D_p$ [see (24)]. Applying Corollary 1 ([6], p. 155) and using the definition of C in (25), we have

$$\sum_{i=N^*+1}^{P} \rho_i W_i = \frac{\sigma_{N^*+1}}{1-\sigma_{N^*+1}} \sum_{i=N^*+1}^{P} \frac{\rho_i}{\mu_i} + \frac{\sigma_{N^*+1}}{\mu_{N^*}} = C - W_0. \tag{A.4}$$

Substitution of (A.4) in (A.3) gives us

$$W_p = \frac{1}{D_p}\left(C + \sum_{i=N_{m_{p-1}}+1}^{p-1} \rho_i W_i \frac{b_i}{b_p} - \sum_{i=N^*+1}^{p-1} \rho_i W_i\right). \tag{A.5}$$

The last sum in this equation involves summing $\rho_i W_i$ over the strict priority groups $m^*, m^*+1, \ldots, m_{p-1}$ (plus the sum from $N_{m_{p-1}}+1$ to $p-1$ within the m_pth strict priority group). We wish to show the following:

$$\sum_{i=N_{m-1}+1}^{N_m} \rho_i W_i = \bar{W}_m \beta_m \quad \text{for} \quad m = m^*, \quad m^*+1, \ldots, M, \tag{A.6}$$

where $\bar{W}_m$ is the average wait for the mth group of a strict priority system with M groups ($m = 1, 2, \ldots, M$) and where† $\lambda_m/\mu_m = \beta_m$; $\bar{W}_m$ is given by Theorem 1 and rewritten in terms of m in (20). From the conservation law [5] we have for a large class of queuing disciplines (for $0 \le \rho < 1$)

$$\sum_{p=1}^{P} \rho_p W_p = \frac{\rho}{1-\rho} W_0 . \tag{A.7}$$

This applies to our SLPM system and may be written as ($N_0 \equiv 0$):

$$\sum_{m=1}^{M} \sum_{i=N_{m-1}+1}^{N_m} \rho_i W_i = \frac{\rho}{1-\rho} W_0 . \tag{A.8}$$

We recognize further that an M-group strict priority system with $\lambda_m/\mu_m = \beta_m$ also obeys the conservation law; that is,

$$\sum_{m=1}^{M} \beta_m \bar{W}_m = \frac{\rho}{1-\rho} W_0 , \tag{A.9}$$

where we have chosen β_m, as in (17), so that (A.8) and (A.9) may be set equal (after normalizing‡ with respect to W_0),

$$\sum_{m=1}^{M} \left(\beta_m \frac{\bar{W}_m}{W_0} - \sum_{i=N_{m-1}+1}^{N_m} \frac{\rho_i W_i}{W_0} \right) = 0. \tag{A.10}$$

We may now change the values of ρ_i in this last equation and leave the sum on i (for any m) unchanged as long as we maintain constant the sum $\beta_m = \sum_{i=N_{m-1}+1}^{N_m} \rho_i$; this is because such a change can affect only those terms $\rho_i W_i$ for $N_{m-1} + 1 \le i \le N_m$ in (A.8), since the other priority groups see this set of (lag) priority groups as a single strict priority group with $\lambda_m/\mu_m = \beta_m$ (due to $b_{N_m+1}/b_{N_m} \gg 1$). Let us therefore choose $\rho_i = 0$ for $N_{m-1} + 1 \le i \le N_m$ and $\rho_{N_m} = \beta_m$ for $m = 2, 3, \ldots, M$. In this event it is clear that we have $W_{N_m} = \bar{W}_m$ for $m = 2, 3, \ldots, M$, which causes cancellation of these terms in (A.10) and gives

$$\beta_1 \bar{W}_1 = \sum_{i=1}^{N_1} \rho_i W_i . \tag{A.11}$$

By making use of (A.11) we may remove the $m = 1$ terms from (A.10) to obtain

$$\sum_{m=2}^{M} \left(\beta_m \frac{\bar{W}_m}{W_0} - \sum_{i=N_{m-1}+1}^{N_m} \rho_i \frac{W_i}{W_0} \right) = 0. \tag{A.12}$$

We now make the new selection $\rho_i = 0$ for $N_{m-1} + 1 \le i \le N_m$ and $\rho_{N_m} = \beta_m$ for $m = 3, 4, \ldots, M$. Again this gives $W_{N_m} = \bar{W}_m$ for $m = 3, 4, \ldots, M$. By cancelling them in (A.12) we get

$$\beta_2 \bar{W}_2 = \sum_{i=N_1+1}^{N_2} \rho_i W_i . \tag{A.13}$$

† λ_m is defined in (29); λ_m and μ_m are the arrival and service rates, respectively, for the mth group.

‡ Note that $\bar{W}_m$ and W_i contain the factor W_0 .

We continue this procedure to obtain (A.6). For $\rho \ge 1$ we use Corollary 1 to the conservation law from [5], which carries through to give us (A.6) over the range $m = m^*, \ldots, M$. We apply this to (A.5) to yield (23), thus proving Theorem 5.

Q.E.D.

Proof (*Theorem* 6). The proof here is almost identical to that given for Theorem 5. We begin with a general preemptive lag priority system that satisfies the following equation ([6], p. 167).

$$T_p = \frac{1}{\mu_p} + \sum_{i=p}^{P} \rho_i T_i + \sum_{i=1}^{p-1} \rho_i T_i \frac{b_i}{b_p} + \sum_{i=p+1}^{P} \rho_i T_p \left(1 - \frac{b_p}{b_i}\right), \tag{A.14}$$

where

$$T_p = W_p + \frac{1}{\mu_p} . \tag{A.15}$$

Thus

$$T_p = \frac{1}{D'_p} \left[\frac{1}{\mu_p} + \sum_{i=p}^{P} \rho_i T_i + \sum_{i=1}^{p-1} \rho_i T_i \frac{b_i}{b_p} \right], \tag{A.16}$$

where D'_p is given by (A.2). Using our SLPM notation, we get for $N_{m_{p-1}} < p < N_{m_p}$

$$T_p = \frac{1}{D_p} \left(\frac{1}{\mu_p} + \sum_{i=p}^{P} \rho_i T_i + \sum_{i=N_{m_{p-1}}+1}^{p-1} \rho_i T_i \frac{b_i}{b_p} \right). \tag{A.17}$$

By Corollary 2 ([6], p. 157) we get

$$\sum_{i=N^*+1}^{P} \rho_i W_i = \frac{\sigma_{N^*+1}}{1 - \sigma_{N^*+1}} \sum_{i=N^*+1}^{P} \frac{\rho_i}{\mu_i} . \tag{A.18}$$

Thus

$$\begin{aligned} \sum_{i=N^*+1}^{P} \rho_i T_i &= \sum_{i=N^*+1}^{P} \rho_i W_i + \sum_{i=N^*+1}^{P} \frac{\rho_i}{\mu_i} \\ &= \left(\frac{\sigma_{N^*+1}}{1 - \sigma_{N^*+1}} + 1 \right) \sum_{i=N^*+1}^{P} \frac{\rho_i}{\mu_i} \\ &= C_1 , \end{aligned}$$

where C_1 is defined in (30). Thus

$$\sum_{i=p}^{P} \rho_i T_i = C_1 - \sum_{i=N^*+1}^{p-1} \rho_i T_i . \tag{A.19}$$

Adding $\sum_{i=1}^{P} \rho_i/\mu_i$ to both sides of (A.7) we get

$$\sum_{i=1}^{P} \rho_i T_i = \frac{W_0}{1-\rho} . \tag{A.20}$$

We now wish to consider a pure strict priority system (with preemption) with M groups where $\beta_m = \lambda_m(\overline{1/\mu_m})$ $m = 1, 2, \ldots, M$ [see (17) and (28)]; $\overline{1/\mu_m}$ is the average of the mean service times of the groups $N_{m-1} + 1 \leq p \leq N_m$. Equation A.20, which applies to this strict priority system, gives

$$\sum_{m=1}^{M} \beta_m \tilde{T}_m = \frac{W_0}{1-\rho};$$

$\tilde{T}_m$ is just the average wait (in queue plus service) for the mth group (see [6]). We also apply (A.20) to the SLPM system to obtain

$$\sum_{m=1}^{M} \sum_{i=N_{m-1}+1}^{N_m} \rho_i T_i = \frac{W_0}{1-\rho}.$$

We now apply the arguments in the proof of Theorem 5 to get

$$\sum_{i=N_{m-1}+1}^{N_m} \rho_i T_i = \beta_m \tilde{T}_m \quad \text{for} \quad m = m^*, \quad m^* + 1, \ldots, M. \tag{A.21}$$

Use of (A.19) and (A.21) in (A.17) gives us

$$T_p = \frac{1}{D_p}\left[\frac{1}{\mu_p} + C_1 - \sum_{m=m^*}^{m_{p-1}} \beta_m \tilde{T}_m - \sum_{i=N_{m_{p-1}}+1}^{p-1} \rho_i T_i\left(1 - \frac{b_i}{b_p}\right)\right].$$

Substitution of (A.15) in the last equation establishes (26) and completes the proof.

Q.E.D.