CAPACITY ALLOCATION UNDER RANDOM SLOT ASSIGNMENT POLICY

Ioannis Stavrakakis

Department of Electrical Engineering and Computer Science
University of Vermont
Burlington, Vermont 05405.

Abstract

In this paper, two non-communicating stations share a time slotted communication channel on a random slot assignment basis. The optimal assignment policy is derived and a simple strategy, based on a threshold test, is developed for the implementation of the optimal policy in a dynamically changing environment. Although the random slot assignment policy is inferior to the optimal periodic, fixed slot assignment policy, it is shown than it achieves the performance of that policy as the variance of the packet arrival process increases; the optimal periodic, fixed slot assignment policy is not, in general, implementable and it is very difficult to become adaptive to the conditions of a dynamically changing environment.

I. Introduction

In this paper a common channel is shared by two distributed stations on a time slot assignment basis. Information is packetized and the packet size is assumed to be equal to the time slot; the time slot is defined to be equal to the transmission time of a single packet. Each station has no knowledge regarding the packet arrival process at the other station. Information regarding the common channel activity is assumed to be available to the stations, if an adaptive channel allocation scheme is to be implemented. In this case, the channel activity information is utilized for the estimation of the traffic at each station. The stations are assumed to be synchronized with the slot boundaries and they are allowed to transmit only at the slot boundaries. The slot boundaries form the clocking times of two identical and synchronized random number generators. Each station is equipped with one such random number generator for the implementation of the random slot assignment policy. If the value of the random number is less than β the slot is assigned to station 2; otherwise, it is assigned to station 1. Thus, $100(1-\beta)\%$ of the channel capacity is assigned to station 1 and $100\beta\%$ of it is assigned to station 2. This random slot assignment scheme will be defined as policy $\boldsymbol{R}(\beta)$ for station 1 and policy $\boldsymbol{R}(1-\beta)$ for station 2. The standard TDM policy which assigns every other slot to a station will be defined as the fixed slot assignment policy $\boldsymbol{F}$, [1], [2], [3].

The random, conflict-free, slot assignment policy has been briefly considered in [8] for the purpose of demonstrating the merit of the fixed slot assignment policy compared with the random one. It has been shown in [8] that the optimal periodic, fixed slot assignment policy is superior to the random slot assignment one, under independent packet arrival processes; the optimal policy is defined to be the policy which induces the minimum mean packet delay. Then effort was concentrated on the derivation of an implementable periodic, fixed slot assignment policy which would achieve the optimal capacity allocation. The golden ratio policy was proposed in [8] as a policy which could result in the optimal channel allocation, at least under certain traffic conditions. The difficulty in achieving the theoretical optimal allocation is discussed in [8], especially under certain traffic situations. In addition to this problem it is not clear how an implementable near-optimal, periodic, fixed slot assignment policy would become adaptive to a dynamically changing packet traffic environment and what the resulting performance would be.

In view of the above comments, the superior-

Research partially supported by the National Science Foundation under Grant NCR-9011962.

ity of an implementable periodic, fixed slot assignment policy over the random one becomes questionable, particularly when a simple policy which is adaptive to packet traffic changes is desirable. Furthermore, it is shown in the next section (Corollary 1) that the (easily and accurately implementable) random slot assignment policy achieves the performance of the theoretical periodic, fixed slot assignment policy asymptotically, as the variance of the traffic approaches infinity. As a consequence, when the variance of the traffic is large, the deviation of the random slot assignment policy from the optimal periodic, fixed slot assignment one could be smaller than that of an implementable near-optimal periodic, fixed slot assignment policy.

Under identical packet traffic conditions in both stations, policy $\boldsymbol{F}$ (defined above) is the optimal periodic fixed slot assignment policy for the system of the two stations. Before policies $\boldsymbol{F}$ and $\boldsymbol{R}(\beta)$ are applied to the system of the two stations (section III), their performance in terms of the induced mean packet delay to a single station is investigated in section II. Although the expressions for the mean packet delay induced by policies $\boldsymbol{F}$ and $\boldsymbol{R}(\beta)$ may be found in [8] or elsewhere for policy $\boldsymbol{F}$, [2], [3], they are derived in this paper (Theorems 1 and 2) by following a new, simple and unified approach for both policies. This approach is directly applicable to the case when the system consists of an arbitrary number of users, as well. Furthermore, although the closed form expression for the induced mean packet delay is obtained for independent per station packet arrival process, the derivation approach is applicable to the case of Markov modulated packet arrival processes to the stations, which are dependent processes. The resulting queueing models under policies $\boldsymbol{F}$ and $\boldsymbol{R}(\beta)$ have not been analyzed in the past, under the dependent, Markov modulated packet arrival processes. The optimal policy for a single station in $\boldsymbol{P}=\{\boldsymbol{F}, \boldsymbol{R}(\beta)$ for $0\leq\beta\leq1\}$ is derived and the excess capacity 1/2-β required for policy $\boldsymbol{R}(\beta)$ to achieve the optimal performance of policy $\boldsymbol{F}$ is established. Numerical results (Fig.4 and 5) show that the excess capacity is insignificant for large variance, which is a manifestation of the result (Corollary 1) that policy $\boldsymbol{R}(\beta)$ becomes optimal as the variance of the packet arrival process approaches infinity.

Notice that, unlike policy $\boldsymbol{F}$, policy $\boldsymbol{R}(\beta)$ assigns a variable portion of the channel capacity to a station, according to the value of $1-\beta$. In the system of the two stations, a larger capacity assignment to one station will reduce the available capacity to the other station and, thus, it will increase the intensity of the queueing problems in the latter station. The performance of the system of the two stations under policy $\bar{\boldsymbol{R}}(\beta)$ is investigated in section III; policy $\bar{\boldsymbol{R}}(\beta)$ is defined as the system policy which applies policy $\boldsymbol{R}(\beta)$ to station 1 and policy $\boldsymbol{R}(1-\beta)$ to station 2. The optimal policy $\bar{\boldsymbol{R}}(\beta_0)$ (defined as the one which minimizes the induced mean delay of a random packet) is derived and it is compared with the fixed policy $\boldsymbol{F}$. From the latter comparison, the optimal policy in $\bar{\boldsymbol{P}}=\{\bar{\boldsymbol{F}}, \bar{\boldsymbol{R}}(\beta)$ for $0\leq\beta\leq1\}$ is established. Again, although policy $\boldsymbol{F}$ is optimal under completely symmetric traffic conditions, the performance of the optimal policy $\boldsymbol{R}(\beta)$ is practically the same for large variance of the packet arrival process. Thus, when the variance of the packet arrival process is significant, both (easily implementable) policies perform similarly. It is under asymmetric traffic load that policy $\boldsymbol{F}$ looses its optimality and the optimal such policy is in general, only approximately and non-easily implementable. Such a near-optimal policy becomes more complicated if, in addition to the traffic asymmetry, the parameters of the traffic vary in time; in this case, the slot assignment policy needs to become adaptive. While an adaptive policy $\boldsymbol{R}(\beta)$ is easily implementable through the adaptation of the probability β (and a strategy is developed for this purpose, Corollary 4) an adaptive optimal or near-optimal implementable, periodic, fixed slot assignment policy would be very complicated and such a policy, to our knowledge, has not been proposed anywhere. For this reason, only the easily implementable policy $\boldsymbol{F}$ is consider here for comparison to the optimal policy $\boldsymbol{R}(\beta)$; under asymmetric and/or time varying traffic conditions the optimal policy $\boldsymbol{R}(\beta)$ is shown to outperform policy $\boldsymbol{F}$ in most cases. It should be pointed out that the objective in this paper is not to show that the random slot assignment policy is better than the periodic, fixed slot assignment one, but to study a simple, implementable and potentially adaptive slot assignment policy which is also comparable in performance to the theoretically optimal fixed slot assignment policy whose practical implementation is not always possible (especially under time varying traffic conditions) or it may be only approximately implementable at the expense of increased complexity and a possibly large deviation from the optimal (theoretical) performance.

In section IV numerical results are presented and useful conclusions regarding the relative performance of the policies are drawn. The effect of the asymmetry (with respect to the rate and/or the structure) of the traffic on the system performance is also illustrated.

II. Optimal policy in P for a single station

In this section, the performance of the policies in $P=\{F, R(\beta)$ for $0\leq\beta\leq 1\}$ is investigated and the optimal policy in P is determined. No constraint on the available capacity $1-\beta$ under policy $R(\beta)$ is imposed, as opposed to policy F under which the available capacity is always .5. This way, the capabilities of policy $R(\beta)$, $0\leq\beta\leq 1$, are fully investigated. An optimal policy in P determines the best slot allocation policy for a particular station without taking into consideration its possible negative effect on the resulting policy which is applied to the other station of the system. The policy in $\tilde{P}=\{\tilde{F}, \tilde{R}(\beta)$ for $0\leq\beta\leq 1\}$, under which optimality is achieved for the 2-station system, is investigated in section III.

Let μ_j, σ_j^2 be the mean and the variance of the (independent per slot) packet arrival process of the j^{th} station of the system, $j=1,2$. The following two theorems provide for the induced mean packet delay under policies F and $R(\beta)$, $0\leq\beta\leq 1$.

Theorem 1: The mean packet delay induced by policy F is given by

$$D_j^F=\frac{1}{2}+\frac{\sigma_j^2}{\mu_j(1-2\mu_j)} \quad , \quad j=1,2 \qquad (1)$$

for $\mu_j<\frac{1}{2}$, $j=1,2$ (stability condition).

Proof: The queueing problem in the buffer of station j operating under policy F is the one depicted in Fig. 1. The server (which is capable of serving one packet per slot) is assumed to be unavailable every other slot, giving rise to a discrete time queueing system with periodic service interruptions. This queueing system can be studied by developing the equivalent statistical multiplexer shown in Fig. 2 and applying the analysis presented in [4]. Let $\{a_i^j\}_{i\geq 0}$ denote the packet arrival process to station j. Let $\{\tilde{a}_i^j\}_{i\geq 0}$ denote a packet arrival process which delivers one packet every other slot; let $\tilde{\mu}_j=.5$ denote the resulting packet arrival rate. The packets delivered by $\{\tilde{a}_i^j\}_{i\geq 0}$ are assumed to have priority over those delivered by $\{a_i^j\}_{i\geq 0}$. Thus, the delay of the packets from $\{\tilde{a}_i^j\}_{i\geq 0}$ is equal to one. The two queueing systems are equivalent with respect to the induced delay for the packets delivered by $\{a_i^j\}_{i\geq 0}$. Whenever the server is unavailable in the queueing system of Fig. 1, the server serves packets from $\{\tilde{a}_i^j\}_{i\geq 0}$ in the queueing system of Fig. 2. The packet arrival process $\{\tilde{a}_i^j\}_{i\geq 0}$ together with the adopted priority policy in the queueing system of Fig. 2 completely represent the interruption policy in the queueing system of Fig. 1. Let D_{12} denote the mean packet delay in the queue of Fig. 2. The work conservation law [5], [6], implies that

$$D_{12}^F=\frac{\tilde{\mu}_j+D_j^F\mu_j}{\tilde{\mu}_j+\mu_j}$$

The queueing system of Fig. 2 has been analyzed in [4] under Markov modulated generalized Bernoulli packet arrival processes. According to this process, the state of an underlying Markov chain determines the distribution of the number of packet arrivals over a slot. It is easy to see that the packet arrival process $\{\tilde{a}_i^j\}_{i\geq 0}$ can be described in terms of an underlying Markov chain with state space $\tilde{S}_j=\{0,1\}$ and steady state and transition probabilities given by

$$\tilde{\pi}(0)=\tilde{\pi}(1)=\frac{1}{2} \quad , \quad \tilde{p}(0,0)=0 \quad , \quad \tilde{p}(1,1)=0$$

and distribution, $\tilde{\phi}(x,k)$ of the number of packet arrivals, k, given a certain state, x, given by

$$\tilde{\phi}(0,0)=1 \quad , \quad \tilde{\phi}(1,1)=1$$

When the number of stations is N, then the corresponding Markov chain would have N states, N-1 of which would deliver one packet, describing

Figure 1
The queueing system with service interruptions.

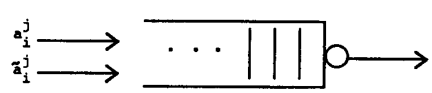


Figure 2
The equivalent queueing system.

the visit of the server to the other stations in the original queueing system with service interruptions.

The Markov modulated generalized Bernoulli model which describes the process $\{a_i^j\}_{i\geq 0}$ (assuming here an independent per slot packet arrival process) is simply a generalized Bernoulli model which can be seen as being based on a single state of an underlying Markov chain with state space $S_j=\{s\}$ and parameters

$$\pi(s)=1 \ , \ p(s,s)=1 \ , \ \phi(s,k)=g(k)$$

where g(k) is the distribution of the number of packets delivered by the independent process $\{a_i^j\}_{i\geq 0}$ over a slot. By applying the analysis presented in [4] the following expression is derived for the mean packet delay D_{12}^F.

$$D_{12}^F=\frac{1}{(\mu_j+\bar{\mu}_j)}\left\{\frac{1+\mu_j}{2}+\frac{\sigma_j^2}{1-2\mu_j}\right\}$$

Finally, (1) can be obtained from the above. A detailed proof may be found in [9]. □

Theorem 2: The mean packet delay induced by policy $\boldsymbol{R}(\beta)$, is given by

$$D_j^R(\beta)=\frac{1}{2}+\frac{\sigma_j^2+\mu_j\beta}{2\mu_j(1-\mu_j-\beta)} \quad , \ j=1,2 \ , \ 0\leq\beta\leq 1 \quad (2)$$

for $\mu_j<1-\beta$ (stability condition).

Proof: The queueing problem in the buffer of station j operating under policy $\boldsymbol{R}(\beta)$ is the one depicted in Fig. 1, where the service interruptions are now random. By following the approach used in the proof of Theorem 1, the equivalent queueing system shown in Fig. 2 can be derived. Under policy $\boldsymbol{R}(\beta)$, the packet arrival process $\{\tilde{a}_i^j\}_{i\geq 0}$, describing the service interruption policy, is a Bernoulli process with rate β. The server is absent in a slot with probability β (Fig. 1) or a priority packet is delivered by $\{\tilde{a}_i^j\}_{i\geq 0}$ with probability β (Fig. 2). By following the approach used in the proof of Theorem 1 and assuming that $\{a_i^j\}_{i\geq 0}$ is an independent process expression (2) is obtained. □

It is of interest to see how policy $\boldsymbol{F}$ compares with a policy $\boldsymbol{R}(\beta)$ for $0\leq\beta\leq 1$. Let $\boldsymbol{P}(\beta)=\{\boldsymbol{F}, \boldsymbol{R}(\beta)\}$. The following theorem provides for the optimal policy in $\boldsymbol{P}(1/2)$. By setting $\beta=1/2$ in (2), the following can be shown.

Theorem 3: Policy $\boldsymbol{F}$ is optimal in $\boldsymbol{P}(1/2)$ for $\mu_j<1/2$. That is, it is the optimal among those policies in $\boldsymbol{P}$ which assign half of the available capacity to the station under consideration. □

Corollary 1: For fixed packet arrival rate μ_j ($\mu_j<1/2$), the normalized deviation $\boldsymbol{d}$ of policy $\boldsymbol{R}(1/2)$ from the optimal policy in $\boldsymbol{P}(1/2)$ decreases monotonically as σ_j^2 increases. Policy $\boldsymbol{R}(1/2)$ is asymptotically optimal for $\sigma_j^2\to\infty$.

Proof:

$$\boldsymbol{d}=\frac{D^R(1/2)-D^F}{D^F}=\frac{1}{2(\mu_j-2\mu_j^2+\sigma_j^2)}$$

and thus $\lim_{\sigma_j\to\infty}\boldsymbol{d}=0$. □

Corollary 1 implies that, for sufficiently large σ_j^2, policy $\boldsymbol{R}(1/2)$ can be arbitrarily close to the optimal policy $\boldsymbol{F}$. Notice that as β increases, the capacity 1−β assigned to the station decreases and, thus, an increase in the induced mean packet delay is expected. This observation (or by inspection from (2)) establishes the following.

Corollary 2: $D^R(\beta)$ is a monotonically increasing function of β. □

Theorem 1 implies that policy $\boldsymbol{F}$ is optimal in $\boldsymbol{P}(1/2)$. The latter, in view of Corollary 2, implies that $\boldsymbol{F}$ is optimal in $\boldsymbol{P}_+(1/2)$, where $\boldsymbol{P}_+(1/2)=\{\boldsymbol{F}, \boldsymbol{R}(\beta)$ for $\beta\geq 1/2\}$. The following theorem provides for the set of policies $\boldsymbol{P}_-(\beta_0)=\{\boldsymbol{F}, \boldsymbol{R}(\beta)$ for $0\leq\beta\leq\beta_0\}$ in which policy $\boldsymbol{R}(\beta)$ is optimal.

Theorem 4: Policy $\boldsymbol{R}(\beta)$ is optimal in $\boldsymbol{P}_-(\beta_0)$. Policy $\boldsymbol{F}$ is optimal in $P_+(\beta_0)$ where β_0 is the optimality threshold given by

$$0<\beta_0=\frac{\sigma_j^2}{(1-2\mu_j)\mu_j+2\sigma_j^2}\leq\frac{1}{2} \quad (3)$$

Proof: The existence of a threshold β_0 as above is guaranteed in view of the monotonicity of $D^R(\beta)$ (Corollary 2) and the fact that $\boldsymbol{R}(0)$ is optimal in $\boldsymbol{P}(0)$. The latter is true since a policy which never makes the server unavailable (β=0) induces smaller delay than that under a policy which makes the server unavailable every other slot. Let β_0 be the value of β which makes both policies in $\boldsymbol{P}(\beta_0)$ optimal, that is $D^R(\beta_0)=D^F$. By using (1) and (2) and solving the previous equation with respect to β_0 we obtained the result. □

From (3) it can be seen that $\beta_0\to 1/2$ as $\sigma_j^2\to\infty$. That is, both policies become optimal in $\boldsymbol{P}(1/2)$ as $\sigma_j^2\to\infty$, which was shown before (Corollary 1). The following corollary is evident in view of the previous theorem.

Corollary 3: The excess capacity c required in order for policy $R(\beta)$ to become optimal in P is given by

$$c = \frac{1}{2} - \beta_0 \quad , \quad \mu_j < \frac{1}{2}.$$

Notice that $c \to 0$ as $\sigma_j^2 \to \infty$, which implies the additional capacity is insignificant for large variance of the traffic. □

III. Optimal policy in $\bar{P}$ for the system.

The developments of the previous section imply that the random slot allocation policy $R(\beta)$ can be optimal in P at the expense of an additional capacity c compared to that under policy F (Corollary 3). In a 2-station communication system, the latter implies that an optimal policy $R(\beta)$ for one station might cause increased queueing problems to the other station or even instability if its packet arrival rate is larger than the assigned capacity.

In the case of an asymmetric system, the reduction in the mean packet delay due to the adoption of an optimal policy $R(\beta)$ by one station might more than compensate for the increased mean packet delay of the other station. Policy F cannot be adjusted to asymmetric packet load conditions. Under such conditions, the optimal periodic, fixed slot assignment policy is not policy F and it can be, in general, only approximately implemented, [8]. Even if the traffic parameters were such that an implementation as suggested in [8] is possible and well performing, its complexity is significantly larger than that of policies $R(\beta)$ and F and its adaptation to a dynamically changing traffic environment probably more so. For these reasons only policies $R(\beta)$ and F are considered as candidate policies which are easily implementable. Furthermore, policy $R(\beta)$ is easily become adaptive through the appropriate selection of the parameter β. A strategy for the identification of the optimal capacity allocation under policy $R(\beta)$ is developed. By properly adjusting β, policy $R(\beta)$ is capable of handling temporary severe queueing problems or temporary queueing instabilities and outperforming the (easily implementable) policy F. These issues are investigated in this section.

Let $\bar{F}$ and $\bar{R}(\beta)$ denote the policies F and $R(\beta)$, respectively, applied to a 2-station communication system. Under policy $\bar{F}$, policy F is applied to each of the stations. Under policy $\bar{R}(\beta)$, policy $R(\beta)$ is applied to one station and policy $R(1-\beta)$ is applied to the other. Let $\bar{P}=\{\bar{F}, \bar{R}(\beta)$ for $0\le\beta\le1\}$ be the class of all policies considered here. An optimal policy in $\bar{P}$ is defined to be the policy which minimizes the mean delay of a random packet (coming from any of the stations). Let μ_j, σ_j^2 be the mean and the variance, respectively, of the packet arrival process in station j, j=1,2. The next theorem identifies the optimal policy in $\{\bar{R}(\beta)$ for $0\le\beta\le1\}$. The following lemma is useful for the proof of that theorem.

Lemma 1: Let $\mu=\mu_1+\mu_2<1$ and $\mu_2<\beta<1-\mu_1$. Then,
(a) $f_1(\beta)$ is strictly increasing for $0\le\beta<1-\mu_1$ and convex U
(b) $f_2(\beta)$ is strictly decreasing for $\mu_2<\beta\le1$ and convex U
(c) $G(\beta)$ is convex U for $\mu_2<\beta<1-\mu_1$, where

$$f_1(\beta)=\frac{\mu_1\beta+\sigma_1^2}{2(1-\mu_1-\beta)} \quad , \quad f_2(\beta)=\frac{\mu_2(1-\beta)+\sigma_2^2}{2(\beta-\mu_2)} \tag{4}$$

$$G(\beta)=f_1(\beta)+f_2(\beta) \tag{5}$$

Proof: It is easy to show that the first and the second derivatives of $f_1(\beta)$ with respect to β are strictly positive for $0\le\beta\le1-\mu_1$ and, thus, $f_1(\beta)$ is strictly increasing and convex U. Similarly, the first and second derivatives of $f_2(\beta)$ are strictly negative and strictly positive, respectively, for $\mu_2<\beta\le1$ and, thus, $f_2(\beta)$ is strictly decreasing and convex U. Part (c) is true in view of the fact that both functions $f_1(\beta)$ and $f_2(\beta)$ are convex U, [7]. □

Theorem 5: Let $\mu=\mu_1+\mu_2<1$. The optimal policy in $\{\bar{R}(\beta)$ for $0\le\beta\le1\}$ is policy $\bar{R}(\beta_0)$, where

$$\beta_0=\arg\{\min_{0\le\beta\le1} G(\beta)\}$$

and $G(\beta)$ is given in Lemma 1. It turns out that β_0 is the root in [0,1] of the second order polynomial

$$(c_1-c_2)\beta^2+\Big(2(1-\mu_1)c_2-2\mu_2c_1\Big)\beta+\mu_2^2c_1-(1-\mu_2)^2c_2 \tag{6}$$

where

$$c_1=\mu_1(1-\mu_1)+\sigma_1^2 \quad , \quad c_2=\mu_2(1-\mu_2)+\sigma_2^2$$

Proof: The optimal policy in $\{\bar{R}(\beta), 0\le\beta\le1\}$ is the policy $\bar{R}(\beta)$ which minimizes the induced mean packet delay

$$\bar{D}^R(\beta)=\frac{\mu_1}{\mu_1+\mu_2}D_1^R(\beta)+\frac{\mu_2}{\mu_1+\mu_2}D_2^R(1-\beta)$$

By substituting (2) and manipulating the resulting expression we obtain

$$\tilde{D}^R(\beta)=\frac{1}{2}+\frac{1}{\mu_1+\mu_2}[f_1(\beta)+f_2(\beta)]=\frac{1}{2}+\frac{1}{\mu_1+\mu_2}G(\beta)$$

Thus, minimizing $\tilde{D}^R(\beta)$ with respect to β is equivalent to minimizing $G(\beta)$ with respect to β. The existence of β_0 is guaranteed since $G(\beta)$ is a convex U function (Lemma 1). By setting the first derivative of $G(\beta)$ equal to zero and manipulating the resulting equation the proof of the theorem is completed. □

The above theorem provides for the optimal policy $\bar{R}(\beta_0)$ in $\{\bar{R}(\beta)$ for $0\leq\beta\leq1\}$ by identifying the optimal value β_0. In a real, dynamically changing environment, it is of interest to develop a simple mechanism capable of testing whether a certain current policy $\bar{R}(\beta)$ is optimal or not and, more important, to develop a strategy which brings the system close to the currently optimal point of operation. The following theorem sets the ground for the development of such a strategy.

<u>Theorem 6:</u> Let β be the operation point (adopted policy $\bar{R}(\beta)$) of a system. The optimal point β_0 (policy $\bar{R}(\beta_0)$) is such that

$$\beta_0<\beta \text{ if } h(\beta)>-1$$

$$\beta_0>\beta \text{ if } h(\beta)<-1$$

$$\beta_0=\beta \text{ if } h(\beta)=-1$$

where

$$h(\beta)=\frac{f_2'(\beta)}{f_1'(\beta)} \quad , \quad \mu_2<\beta<1-\mu_1$$

and $f_1'(\beta)$, $f_2'(\beta)$ are the first derivatives of $f_1(\beta)$, $f_2(\beta)$ given by (4). A typical function $h(\beta)$ is shown in Fig. 3.

<u>Proof:</u> It is easy to show that $f_1'(\beta)>0$, $f_1''(\beta)>0$, $f_2'(\beta)<0$ and $f_2''(\beta)>0$ (see proof of Lemma 1). Then,

$$h'(\beta)=\frac{f_2''(\beta)f_1'(\beta)-f_2'(\beta)f_1''(\beta)}{[f_1'(\beta)]^2}>0 \quad , \quad \mu_2<\beta<1-\mu_1$$

Thus, $h(\beta)$ is a strictly increasing function of β for $\mu_2<\beta<1-\mu_1$. Theorem 5 implies that $G'(\beta_0)=f_1'(\beta_0)+f_2'(\beta_0)=0$; therefore,

$$\frac{f_2'(\beta_0)}{f_1'(\beta_0)}=-1$$

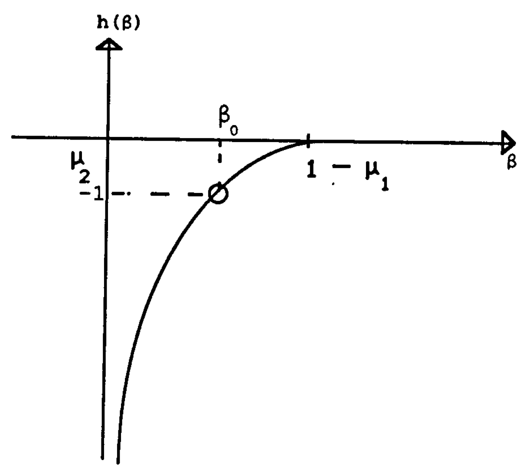
The above equation, together with the monotonicity property of $h(\beta)$ complete the proof. □

The following Corollary is obvious in view of the above theorem.

<u>Corollary 4:</u> Let $R(\beta)$ be the currently adopted policy. The optimal point β_0 (policy $R(\beta_0)$) can be reached by the following strategy : S = { increase β if $h(\beta)<-1$, decrease β if $h(\beta)>-1$, maintain β if $h(\beta)=-1$ }. □

The above strategy generates a sequence of policies $\{R(\beta_j)\}_j$ which converges to the optimal policy in $\{\bar{R}(\beta)$ for $0\leq\beta\leq1\}$, $R(\beta_0)$. Strategy S can be used for the adaptation of policy $R(\beta)$ to the varying optimal policy $R(\beta_0)$, in a dynamically changing environment. For instance, if the rates μ_1 and μ_2 change, strategy S is capable of adjusting the operation of the system so that optimality can be achieved, provided that some estimates of the traffic parameters be available. The mechanism for the generation of such estimates, its goodness and the detailed implementation of strategy S are beyond the scope of this paper. The common channel is assumed to be capable of providing the information necessary for the derivation of the estimates and the identical update of the random number generators (β). The following Corollary provides some intuitively expected results. The proof and some discussion on the implications of this corollary can be found in [9].

<u>Corollary 5:</u> Let $\mu=\mu_1+\mu_2<1$ and $\theta=(1-\mu_1-\mu_2)(\mu_1-\mu_2)$. (a) If $\mu_1>\mu_2$, then $\beta_0\leq\frac{1}{2}-\frac{\mu_1-\mu_2}{2}$ if and



<u>Figure 3</u>
A typical function h(β).

only if $\sigma_1^2 \geq \sigma_2^2 - \theta$. If $\mu_2 > \mu_1$, then $\beta_0 \geq \frac{1}{2} - \frac{\mu_1 - \mu_2}{2}$ if and only if $\sigma_2^2 \geq \sigma_1^2 + \theta$. Equalities hold when the corresponding conditions hold with equality as well. (b) If $\mu_1 = \mu_2$ then $\beta_0 = \frac{1}{2}$ or $\beta_0 < \frac{1}{2}$ or $\beta_0 > \frac{1}{2}$ depending on whether $\sigma_1^2 = \sigma_2^2$ or $\sigma_1^2 > \sigma_2^2$ or $\sigma_1^2 < \sigma_2^2$, respectively. □

So far the optimal policy in $\{\tilde{R}(\beta)$ for $0 \leq \beta \leq 1\}$ has been studied and the optimal value β_0 has been derived. The optimal policy in $\tilde{P} = \{\tilde{F}, \tilde{R}(\beta)$ for $0 \leq \beta \leq 1\}$ is given by the next theorem.

Theorem 7: Let $\mu_1 + \mu_2 < 1$. The optimal policy in $\tilde{P}$ is the optimal policy in $\{\tilde{R}(\beta), 0 \leq \beta \leq 1\}$, $\tilde{R}(\beta_0)$, given by Theorem 5, if and only if one of the following conditions is satisfied:

(a) $\mu_1 > \frac{1}{2}$ or $\mu_2 > \frac{1}{2}$

(b) $\frac{\sigma_1^2}{1-2\mu_1} + \frac{\sigma_2^2}{1-2\mu_2} > G(\beta_0)$, where $G(\beta)$ is given in Theorem 5.

Proof: (a) If either $\mu_1 > \frac{1}{2}$ or $\mu_2 > \frac{1}{2}$ is satisfied, then one of the two queues would be unstable under policy F and, thus, the optimal policy in $\tilde{P}$ will be $\tilde{R}(\beta_0)$, since $\tilde{F}$ would induce infinite mean packet delay to at least one of the stations. (b) The mean packet delay under policy $\tilde{F}$ is given by

$$\tilde{D}^F = \frac{\mu_1}{\mu_1+\mu_2} D_1^F + \frac{\mu_2}{\mu_1+\mu_2} D_2^F$$

where D_j^F is given by (1), j = 1,2. The mean packet delay under the optimal policy $\tilde{R}(\beta_0)$ in $\{\tilde{R}(\beta), 0 \leq \beta \leq 1\}$ is given by

$$\tilde{D}^R(\beta_0) = \frac{\mu_1}{\mu_1+\mu_2} D_1^R(\beta_0) + \frac{\mu_2}{\mu_1+\mu_2} D_2^R(1-\beta_0)$$

where $D_j^R(\beta_0)$ is given by (2), j = 1,2. $R(\beta_0)$ is optimal in P if and only if

$$\tilde{D}^R(\beta_0) < \tilde{D}^F$$

By substituting for $\tilde{D}^R(\beta_0)$ and $\tilde{D}^F$ and manipulating the resulting expression the desired result is obtained. □

Corollary 6: Policy $\tilde{F}$ is optimal in $\tilde{P}$ if $\mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2$ and $\mu_1 + \mu_2 < 1$.

Proof: It is easy to show by direct substitution that the necessary condition (b) of Theorem 7 is not satisfied. The proof can be also obtain by invoking Corollary 5. Corollary 5-(b) implies that the optimal policy in $\{\tilde{R}(\beta), 0 \leq \beta \leq 1\}$ for $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$ is policy $\tilde{R}(1/2)$. Under policy $\tilde{R}(1/2)$, each station of the system operates under policy $R(1/2)$. The latter policy is inferior to policy F, as proven in Theorem 3. Thus, the optimal policy in $\{\tilde{R}(\beta), 0 \leq \beta \leq 1\}$, $\tilde{R}(1/2)$, is inferior to policy $\tilde{F}$. □

IV. Numerical results.

In Fig. 4 the mean packet delay induced under policies $R(\beta)$, $0 \leq \beta \leq 1$ and F is plotted for mean packet arrival rate $\mu = .25$ packets/slot and variance $\sigma^2 = 2\mu$ and $\sigma^2 = 10\mu$. Notice that $D^R(1/2)$ is always greater than D^F, as shown in Theorem 3. Notice also that a small increase in the allocated capacity is sufficient for policy $R(\beta)$ to become optimal (.056 for $\sigma^2 = 2\mu$). As σ^2 increases both $D^F(\beta)$ and D^R increase as expected (see (1) and (2)). Notice that as σ^2 increases the additional capacity required for policy $R(\beta)$ to become optimal decreases, as implied by Corollary 1 (.013 for $\sigma^2 = 10\mu$). Also, notice that the optimality threshold β_0 is as computed by (3) and it is always less than .5, which

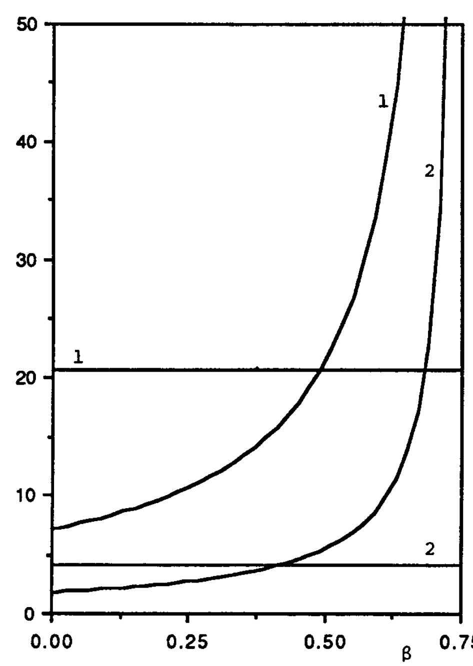


Figure 4

Mean packet delay (in slots) for a single station under policies $R(\beta)$ and F. (D^F: constant; D^R: convex U; $\mu = .25$; $1 \rightarrow \sigma^2 = 2.5$, $2 \rightarrow \sigma^2 = .5$).

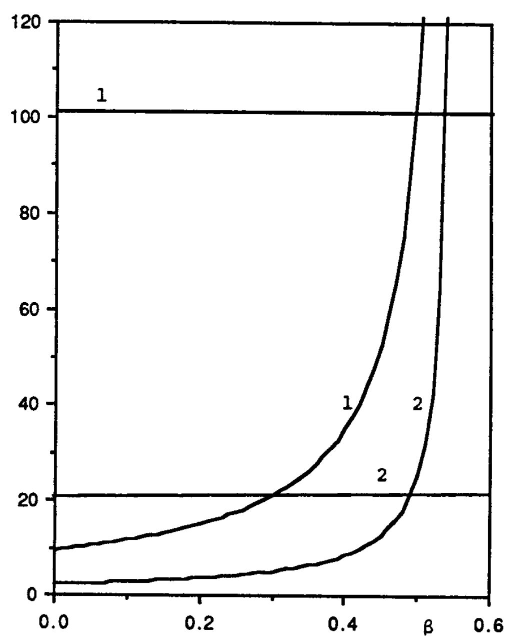
3C.4.7.

illustrates the optimality of policy $\boldsymbol{F}$ for $\beta=.5$; for $\beta \leq \beta_0$ policy $\boldsymbol{R}(\beta)$ is optimal as implied by Theorem 4.

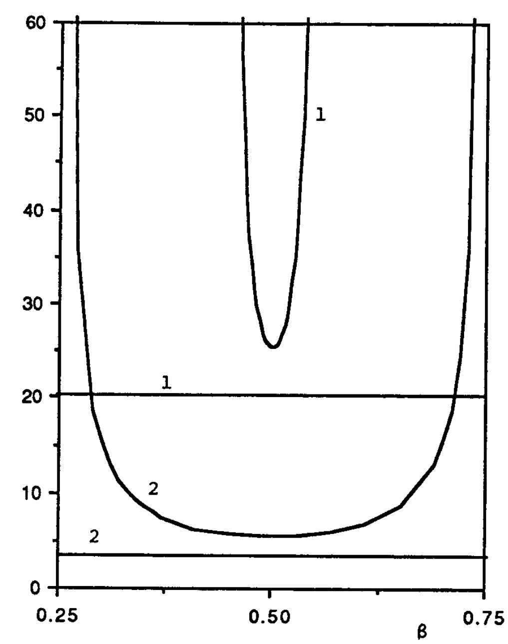
In Fig. 5 similar results are shown for heavier traffic load ($\mu=.45$). Notice that the additional capacity required for policy $\boldsymbol{R}(\beta)$ to be optimal is even smaller (.0025) since $\sigma^2=4.5$ is large. Practically, both policies $\boldsymbol{R}(1/2)$ and $\boldsymbol{F}$ are optimal for large variance. Notice also that if 55% (for $\sigma^2=4.5$) of the channel capacity (as opposed to 50% under policy $\tilde{\boldsymbol{F}}$) can be allocated to the station, then the induced mean packet delay under $\boldsymbol{R}(.45)$ is about half the one induced under policy $\tilde{\boldsymbol{F}}$ (52 versus 100 slots). The latter observation implies that the theoretical advantage (optimality) of policy $\tilde{\boldsymbol{F}}$ over policy $\tilde{\boldsymbol{R}}(1/2)$ may disappear in practical cases, if some additional capacity is offered to the station. The performance improvement may be tremendous at the cost of the utilization of slightly larger capacity. This cost may be insignificant when the rest of the capacity is under-utilized (for instance, under asymmetric traffic situation). The latter issue is discussed later (Fig. 7 and 8).

In Fig. 6, the mean packet delay in a 2-station communication system under policies $\tilde{\boldsymbol{R}}(\beta)$, $0 \leq \beta \leq 1$ and $\tilde{\boldsymbol{F}}$ is plotted, for the case of symmetric traffic load. Notice that the optimal policy in $\{\tilde{\boldsymbol{R}}(\beta), 0 \leq \beta \leq 1\}$ is policy $\tilde{\boldsymbol{R}}(1/2)$, as implied by Corollary 5-(b). The optimal policy in $\tilde{\boldsymbol{P}}$ is policy $\tilde{\boldsymbol{F}}$, as implied by Corollary 6.

In Fig. 7 similar results under asymmetric traffic load are shown. For $\mu_1=.3$, $\sigma_1^2=.6$ and $\mu_2=.4$, $\sigma_2^2=.8$, the asymmetry in the rate and the structure of the packet arrival processes is not strong enough to render policy $\tilde{\boldsymbol{R}}(\beta_0)$ optimal. For $\sigma_1^2=3$ and $\sigma_2^2=4$, policy $\tilde{\boldsymbol{R}}(\beta_0)$ has clearly become the optimal policy. This is due mostly to the structure (variance) of the packet arrival processes rather than the difference in the rates. The favoring effect of the structure on policy $\tilde{\boldsymbol{R}}(\beta_0)$ is due,



<u>Figure 5</u>
Mean packet delay (in slots) for a single station under policies $\boldsymbol{R}(\beta)$ and $\boldsymbol{F}$. (D^F: constant; D^R: convex U; $\mu=.45$; $1 \rightarrow \sigma^2=4.5$, $2 \rightarrow \sigma^2=.9$).



<u>Figure 6</u>
Mean packet delay (in slots) for the two symmetric stations under policies $\tilde{\boldsymbol{R}}(\beta)$ and $\tilde{\boldsymbol{F}}$. ($\tilde{D}^F$: constant; $\tilde{D}^R$: convex U; $1 \rightarrow \mu=.45$, $\sigma^2=.9$; $2 \rightarrow \mu=.25$, $\sigma^2=.5$).

first, to the larger variance of the traffics, which under symmetry would bring policy $\bar{R}(\beta_0)$ very close to the optimal policy F (Corollary 1) and, second, to the larger difference in the variances which further emphasizes the asymmetry of the traffic (since $\sigma_2^2>\sigma_1^2$ and $\mu_2>\mu_1$) and certainly results in the inequality/condition (b) of Theorem 7. For $\sigma_1^2=3$ and $\sigma_2^2=.4$, the asymmetry in the packet arrival processes, due to the asymmetry in the packet arrival rates, is balanced out by the non-coherent (i.e., if $\mu_1<\mu_2$ then $\sigma_1^2>\sigma_2^2$) asymmetry in their structure (variance). As a result, the two packet arrival processes behave as being almost symmetric, with respect to the intensity of the resulting queueing problems, and policy $\bar{F}$ becomes optimal. For $\sigma_1^2=10$ and $\sigma_2^2=.1$, the non-coherent asymmetry in the structure of the processes is strong and overwhelms the asymmetry in the packet arrival rates. Thus, the packet arrival processes behave as being asymmetric in a direction opposite to that implied by the packet arrival rates. As a result policy $\bar{R}(\beta_0)$ becomes optimal again due to this strong asymmetry. Notice that $\beta_0=.44$ which implies that more capacity (.56) is assigned to station 1 despite the significantly smaller packet arrival rate (.3 versus .4).

In Fig. 8 similar results are presented. The packet arrival rates are assumed to be asymmetric with $\mu_1=.1$ and $\mu_2=.45$. The large asymmetry in the rates together with the coherent asymmetry (i.e. if $\mu_1<\mu_2$ then $\sigma_1^2<\sigma_2^2$) in the structure of the packet arrival processes ($\sigma_1^2=1$, $\sigma_2^2=4.5$) render policy $\bar{R}(\beta_0)$ optimal. The asymmetry in the rates is capable of rendering policy $\bar{R}(\beta_0)$ optimal even if the structure of the two processes is symmetric ($\sigma_1^2=\sigma_2^2=1$). When the asymmetry in the structure

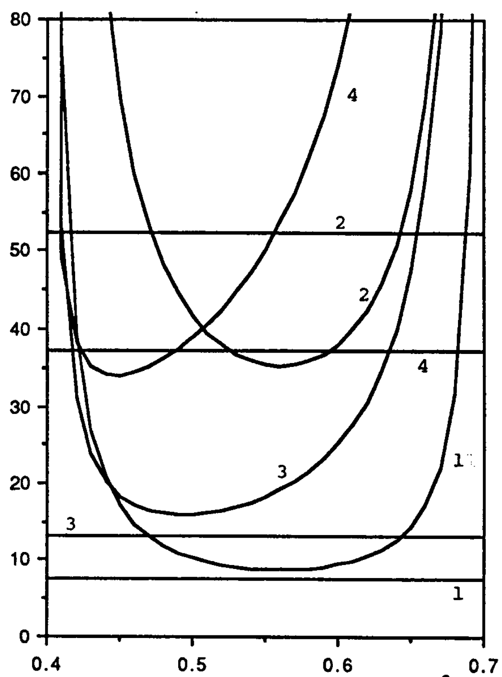


Figure 7
Mean packet delay (in slots) for the two stations under policies $\bar{R}(\beta)$ and $\bar{F}$. ($\bar{D}^F$: constant; $\bar{D}^R$: convex U; $\mu_1=.3$, $\mu_2=.4$; $1 \rightarrow \sigma_1^2=.6$, $\sigma_2^2=.8$; $2 \rightarrow \sigma_1^2=3$, $\sigma_2^2=4$; $3 \rightarrow \sigma_1^2=3$, $\sigma_2^2=.4$; $4 \rightarrow \sigma_1^2=10$, $\sigma_2^2=.1$).

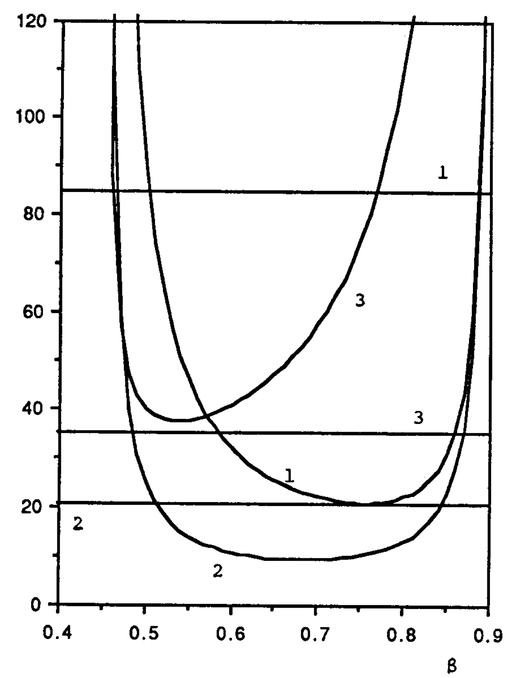


Figure 8
Mean packet delay (in slots) for the two stations under policies $\bar{R}(\beta)$ and $\bar{F}$. ($\bar{D}^F$: constant; $\bar{D}^R$: convex U; $\mu_1=.1$, $\mu_2=.45$; $1 \rightarrow \sigma_1^2=1$, $\sigma_2^2=4.5$; $2 \rightarrow \sigma_1^2=1$, $\sigma_2^2=1$; $3 \rightarrow \sigma_1^2=12$, $\sigma_2^2=.45$).

is non-coherent (i.e., if $\mu_1 < \mu_2$ then $\sigma_1^2 > \sigma_2^2$) and sufficiently large then its counter-effect will have a balancing effect on the two queues which will show a symmetric behavior and render policy $\tilde{F}$ optimal ($\sigma_1^2 = 12$, $\sigma_2^2 = .45$).

References

[1] D. Bertsekas, R. Gallager, "Data Networks", Prentice Hall, 1987.

[2] I. Rubin, "Message Delays in FDMA and TDMA Communication Channels", IEEE Transactions on Communications, Vol. 27, No. 5, May 1979.

[3] S. Lam, "Delay Analysis of a Time Division Multiple Access (TDMA) Channel", IEEE Transactions on Communications, Vol. 25, No. 12, Dec. 1977.

[4] I. Stavrakakis, "Analysis of a Statistical Multiplexer under a General Input Traffic Model", INFOCOM'90 conference, June 5-7, 1990, San Francisco.

[5] G. Barberis, "A Useful Tool in Priority Queueing", IEEE Transactions on Communications, Vol. 28, No. 9, Sept. 1980.

[6] D. Heyman, M. Sobel, "Stochastic Models in Operations Research", Vol.1, McGraw-Hill, 1982.

[7] A. Roberts, D. Varberg, "Convex Functions", Academic Press, New York, 1973.

[8] M. Hofri, Z. Rosberg,"IEEE Transactions on Information Theory", Vol. IT-33, No. 3, May 1987.

[9] I. Stavrakakis, "Channel Assignment under a Conflict-Free Allocation Policy", submitted for publication.