Channel Assignment Under a Conflict-Free Allocation Policy

Ioannis Stavrakakis, *Senior Member, IEEE*

Abstract— **In this paper, a random, conflict-free slot assignment policy is adopted for the allocation of a common channel between two (non-communicating) stations. Although this policy is inferior to the optimal periodic, fixed slot assignment policy, it is shown than it achieves the performance of that optimal policy as the variance of the packet arrival process increases. The main advantage of the random, conflict-free slot assignment policy is that it is simple and always feasible unlike the optimal, periodic, fixed slot assignment policy. Furthermore, the proposed policy is easily implemented in a dynamically changing environment; the optimal such policy is derived and a simple strategy based on a threshold test is developed for the identification of the optimal such policy, when estimates of the traffic parameters are available. No such strategy is known for the adaptation of the parameters of the optimal periodic, fixed slot assignment policy. The developed analysis approach can be applied, to a great extent, to a system with more than two stations.**

I. Introduction

The allocation of a common resource among a number of users is the central problem in a multi-user communication network. Depending on a number of factors, such as the user traffic characteristics, the packet delivery requirements and the topology, certain resource allocation schemes perform better than others. Random access schemes, such as those in ALOHA and Ethernet networks, generate packet collisions whose intensity increases as the number of users decreases and the per user traffic increases. As a result, random access schemes are efficient under light or moderate traffic generated by a large number of bursty users. On the other hand, when the number of users is small and the per user traffic significant and regular (as opposed to bursty), conflict-free resource allocation schemes, such as polling and time division multiplexing, have been shown to be superior. The literature regarding the previously mentioned schemes is large [1].

In this paper a common channel is shared by two distributed stations on a time slot assignment basis. The stations are assumed to have no knowledge of the other station's state. 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 in each station.

The stations are assumed to be synchronized and are allowed to transmit only at the slot boundaries. A random number x $(0 \leq x \leq 1)$ generated at every slot boundary is used for the determination of the channel allocation. If $x \leq \beta$, for some $0 \leq \beta \leq 1$, then the slot is assigned to station 2; otherwise, it is assigned to station 1. According to this policy, a slot is assigned to station 1 with probability $1-\beta$ and it is assigned to station 2 with probability β. The previously described random, conflict-free slot assignment policy will be denoted by $\mathbf{R}(\beta)$ for station 1 and by $\mathbf{R}(1-\beta)$ for station 2. The standard TDM policy which assigns every other slot to a station will be defined as the fixed slot assignment policy $\mathbf{F}$ [2], [3].

The implementation of the random, conflict-free slot assignment policy requires that the random number x be available to all the stations in every slot. This can be implemented in a number of ways, without establishing a special communication link between the stations. For instance, all stations could be equipped with identical and synchronized random number generators. The slot boundaries may serve as the clocking times for the number generation process. Maintaining the synchronization of the slot boundaries (and thus the random number generators) is required for the implementation of any distributed time division multiplexing scheme. Thus, provided that the number generators are simultaneously initialized, maintaining their synchronization is easily achieved. When the allocation parameter β is updated based on traffic estimates, all stations must determine the same value of β in every slot. This is possible if all stations can obtain the channel activity information, which is necessary for the determination of β, and the same algorithm is utilized. The random, conflict-free slot assignment policy can also be implemented by using one random number generator and one mechanism to update β located at the entity (one of the stations or a third party) that provides for the slot synchronization. This entity would generate x, update β and take a decision as to which station should be permitted to access the current slot. This permission could then be communicated to the stations by setting certain flag(s) at the beginning of each slot.

The random, conflict-free slot assignment policy has been briefly considered in [4] for the purpose of demonstrating the merit of the fixed slot assignment policy. It has been shown in [4] that the optimal periodic, fixed slot assign-

Paper approved by the Editor for Random Access and Distributed Systems of the IEEE Communications Society. Manuscript received December 14, 1990; revised June 10, 1991, September 30, 1991 and April 24, 1992. This work was supported in part by the National Science Foundation under Grant NCR-9011962. This paper was presented in part at the IEEE INFOCOM'91 Conference, April 7-11, 1991, Miami, FL.

The author is with the Department of Computer Science and Electrical Engineering of the University of Vermont, Burlington, Vermont 05405, USA.

IEEE Log Number 9400930.

ment policy is superior to the random, conflict-free 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 a feasible periodic, fixed slot assignment policy which would achieve the optimal capacity allocation. The golden ratio policy was proposed in [4] as a policy which could achieve the optimal channel allocation, at least under certain traffic conditions. The difficulty in achieving the theoretical optimal allocation is discussed in [4]. In addition to the latter problem, it is not clear how a feasible 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 superiority of a feasible periodic, fixed slot assignment policy over the random, conflict-free 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 implemented) random, conflict-free slot assignment policy achieves the performance of the optimal 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, conflict-free slot assignment policy from the optimal periodic, fixed one could be smaller than that of a feasible near-optimal periodic, fixed slot assignment policy!

Under identical packet traffic conditions at both stations, policy $\mathbf{F}$ (defined above) is the optimal periodic, fixed slot assignment policy for the system of two stations. Before policies $\mathbf{F}$ and $\mathbf{R}(\beta)$ are applied to the system of 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 $\mathbf{F}$ and $\mathbf{R}(\beta)$ may be found in [4] or elsewhere for policy $\mathbf{F}$, [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 an independent packet arrival process in each station, the derivation approach is also applicable under Markov modulated packet arrival processes. The resulting queueing models under policies $\mathbf{F}$ and $\mathbf{R}(\beta)$ have not been analyzed in the past under Markov modulated packet arrival processes. The optimal policy for a single station in $\mathbf{P} = \{\mathbf{F},\ \mathbf{R}(\beta) \text{ for } 0 \le \beta \le 1\}$ is derived and the excess capacity $1/2 - \beta$ required for policy $\mathbf{R}(\beta)$ to achieve the optimal performance of policy $\mathbf{F}$ is established. Numerical results (Figs. 4 and 5) show that the excess capacity is insignificant for large variance, which is a manifestation of the result (Corollary 1) that policy $\mathbf{R}(\beta)$ becomes optimal as the variance of the packet arrival process approaches infinity.

Notice that, unlike policy $\mathbf{F}$, policy $\mathbf{R}(\beta)$ assigns a variable portion of the channel capacity to a station, according to the value of $1-\beta$. In the system of two stations, a larger capacity assignment to one station will result in smaller available capacity to the other station and, thus, the intensity of the queueing problems in the latter station will be increased. The performance of the system of two stations under policy $\tilde{\mathbf{R}}(\beta)$ is investigated in section III; policy $\tilde{\mathbf{R}}(\beta)$ is defined as the system policy which applies policy $\mathbf{R}(\beta)$ to station 1 and policy $\mathbf{R}(1-\beta)$ to station 2. The optimal policy $\tilde{\mathbf{R}}(\beta_0)$ (defined as the one which minimizes the induced mean delay of a random packet) is derived and it is compared with the fixed policy $\tilde{\mathbf{F}}$; under policy $\tilde{\mathbf{F}}$, policy $\mathbf{F}$ is applied to each of the two stations. From this comparison, the optimal policy in $\tilde{\mathbf{P}} = \{\tilde{\mathbf{F}},\ \tilde{\mathbf{R}}(\beta) \text{ for } 0 \le \beta \le 1\}$ is established.

Although policy $\tilde{\mathbf{F}}$ is optimal under completely symmetric traffic conditions, the performance of the optimal policy $\tilde{\mathbf{R}}(\beta_0)$ 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 implemented) policies perform similarly. It is under asymmetric traffic load that policy $\tilde{\mathbf{F}}$ looses its optimality and the optimal periodic, fixed slot assignment policy is, in general, only approximately and non-easily implemented. 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 $\tilde{\mathbf{R}}(\beta)$ is easily implemented through the adaptation of the probability β (and a strategy is developed for this purpose, Corollary 4), an adaptive optimal or near-optimal feasible, 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 implemented policy $\tilde{\mathbf{F}}$ is consider here for comparison to the optimal policy $\tilde{\mathbf{R}}(\beta_0)$; under asymmetric and/or time varying traffic conditions the optimal policy $\tilde{\mathbf{R}}(\beta_0)$ is shown to outperform policy $\tilde{\mathbf{F}}$ in most cases. It should be pointed out that the objective in this paper is not to show that the random, conflict-free slot assignment policy is better than the periodic, fixed one, but to study a simple, feasible 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 which may be only approximately implemented 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 of the traffic with respect to both the rate and the structure is also illustrated. Finally, the conclusions of this work and some extensions are presented in section V.

II. Optimal Policy in P for a Single Station

In this section, the performance of the policies in $\mathbf{P} = \{\mathbf{F},\ \mathbf{R}(\beta) \text{ for } 0 \le \beta \le 1\}$ is investigated and the optimal policy in $\mathbf{P}$ is determined. No constraint on the available

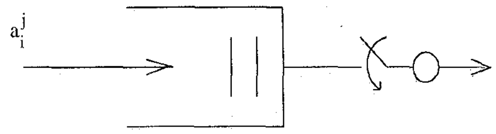

Fig. 1. The queueing system with service interruptions.

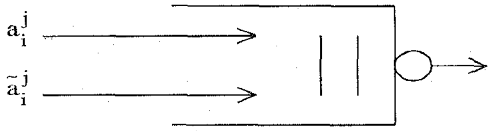

Fig. 2. The equivalent queueing system.

capacity $1-\beta$ under policy $\mathbf{R}(\beta)$ is imposed; note that under policy $\mathbf{F}$ the available capacity is always equal to .5. This way, the capabilities of policy $\mathbf{R}(\beta)$, $0 \leq \beta \leq 1$, are fully investigated. An optimal policy in $\mathbf{P}$ 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 $\tilde{\mathbf{P}} = \{\tilde{\mathbf{F}}, \tilde{\mathbf{R}}(\beta)$ for $0 \leq \beta \leq 1\}$, under which optimality is achieved for the 2-station system, is investigated in section III.

Each station is assumed to be equipped with a buffer of infinite capacity. The performance of a policy is evaluated in terms of the behavior of the buffer associated with a station operating under that policy. The packet arrival process, $\{a_i^j\}_{i\geq 0}$, associated with station j, $j=1,2$, is assumed to be an idependent and identically distributed discrete-time process, defined at the slot boundaries. This process may deliver k packets at any discrete-time instant with some probability $g_j(k)$, $0 \leq k \leq K$ for $K \geq 1$ (generalized Bernoulli process). Let μ_j, σ_j^2 denote the mean and the variance of this probability distribution. When the buffer is non-empty, a packet leaves the buffer (is transmitted) at the end of a slot assigned to the station by the policy. Thus, the buffer behavior can be studied by considering a discrete-time queueing system with interruptions. The following two theorems provide for the induced mean packet delay under policies $\mathbf{F}$ and $\mathbf{R}(\beta)$, $0 \leq \beta \leq 1$.

Theorem 1: Under stability ($\mu_j < \frac{1}{2}$, $j=1,2$), the mean packet delay induced by policy $\mathbf{F}$ is given by

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

Proof: The queueing system formulated in the buffer of station j operating under policy $\mathbf{F}$ is 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 considering the equivalent statistical multiplexer shown in Fig. 2 and applying the analysis presented in [5]. 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_{FIFO} denote the mean packet delay in the queueing system of Fig. 2 operating under the FIFO (First-In First-Out) service policy. The work conservation law [6], [7], [8] implies that

$$D_{\text{FIFO}}^F = \frac{\tilde{\mu}_j + D_j^F \mu_j}{\tilde{\mu}_j + \mu_j} \tag{2}$$

A simple proof of (2) is presented in the Appendix. The FIFO queueing system of Fig. 2 has been analyzed in [5] 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 $\tilde{\pi}(0) = \tilde{\pi}(1) = 1/2$, $\tilde{p}(0,0) = 0$, $\tilde{p}(1,1) = 0$ and distribution $\tilde{\phi}(x,k)$ of the number of packet arrivals, k, given a certain state, x, given by $\tilde{\phi}(0,0) = 1$, $\tilde{\phi}(1,1) = 1$. When the number of stations is N, then the corresponding Markov chain would have N states, $N-1$ of which would deliver one packet, representing the visits of the server to the other stations in the original queueing system with service interruptions.

The Markov modulated generalized Bernoulli model describing the independent per slot packet arrival process $\{a_i^j\}_{i\geq 0}$ (generalized Bernoulli) is based on a single state underlying Markov chain with state space $S_j = \{s\}$ and parameters $\pi_j(s) = 1$, $p_j(s,s) = 1$ and $\phi_j(s,k) = g_j(k)$. By applying the analysis presented in [5] the mean buffer occupancy, Q_{FIFO}^F, can be computed from the following expression

$$Q_{\text{FIFO}}^F = w(0) + w(1) \tag{3}$$

where $w(0)$ and $w(1)$ are the two solutions of the following linear equations

$$\begin{aligned} w(0) = {} & w(0)p_j(0,0) + w(1)p_j(1,0) \\ & +(\mu_j + \tilde{\mu}(0) - 1)p_j(0,0)\pi_j(0) \\ & +(\mu + \tilde{\mu}(1) - 1)p_j(1,0)\pi_j(1) \\ & +(1-\mu)p_j(0,0)p_j(0,0) \\ & +(1-\mu)p_j(0,1)p_j(1,0) \end{aligned} \tag{4}$$

and

$$0 = 2(\mu_j + \tilde{\mu}(0) - 1)w(0)$$

$$\begin{aligned}&+2(\mu_j+\tilde{\mu}(0)-1)(1-\mu)p_j(0,0)\\&+\pi_j(0)(2+s(0)-3(\mu_j+\tilde{\mu}(0))\\&+2(\mu_j+\tilde{\mu}(1)-1)w(1)\\&+2(\mu_j+\tilde{\mu}(1)-1)(1-\mu)p_j(0,1)\\&+\pi_j(1)(2+s(1)-3(\mu_j+\tilde{\mu}(1))\end{aligned} \tag{5}$$

where $\tilde{\mu}(k)$ is the mean packet arrivals delivered by $\{\tilde{a}_i^j\}_{i\geq 1}$ under state k, $k=0,1$; $s(k)$ is the second moment of the cumulative packet arrival process under state k, $k=0,1$; μ is the mean value of the cumulative packet arrival process. They are given by

$$\begin{aligned}&\tilde{\mu}(0)=0,\quad \tilde{\mu}(1)=1,\quad \mu=\mu_j+\tilde{\mu}_j,\\&s(0)=\sigma_j^2+\mu_j^2,\quad s(1)=\sigma_j^2+(\mu_j+1)^2\end{aligned} \tag{6}$$

By using (3)–(6) it can be found that

$$Q^F_{\text{FIFO}}=\frac{1+\mu_j}{2}+\frac{\sigma_j^2}{1-2\mu_j} \tag{7}$$

Finally, (1) can be obtained from (7) by using (2) and applying Little's theorem ($D^F_{\text{FIFO}}=Q^F_{\text{FIFO}}/(\mu_j+\tilde{\mu}_j)$).

Theorem 2: Under stability ($\mu_j<1-\beta$), the mean packet delay induced by policy $\mathbf{R}(\beta)$, is given by

$$D_j^R(\beta)=\frac{1}{2}+\frac{\sigma_j^2+\mu_j\beta}{2\mu_j(1-\mu_j-\beta)},\quad j=1,2,\quad 0\leq\beta\leq 1. \tag{8}$$

Proof: The queueing system formulated in the buffer of station j operating under policy $\mathbf{R}(\beta)$ is 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 $\mathbf{R}(\beta)$, 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). Assuming that $\{a_i^j\}_{i\geq 0}$ is an independent process, both input processes to the queueing system in Fig. 2 are independent processes. The mean buffer occupancy, Q_{FIFO}, is easily found to be equal to, [5],

$$Q^R_{\text{FIFO}}=\mu+\frac{s-\mu}{2(1-\mu)} \tag{9}$$

where μ and s are the first and the second moments at the cumulative packet arrival process given by $\mu=\mu_j+\tilde{\mu}_j=\mu_j+\beta$, $s=\beta(1-\beta)+\sigma_j^2+(\mu_j+\beta)^2$. Finally, (8) can be obtained by applying Little's theorem to (9).

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

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

Proof: By setting $\beta=1/2$ in (8), it is easy to show that $D^R(1/2)=D^F+\frac{1}{2(1-2\mu_j)}$ and thus $D^R(1/2)>D^F$.

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

Proof: From Theorem 3,

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

and thus $\lim_{\sigma_j\to\infty}\mathbf{D}=0$.

Corollary 1 implies that, for sufficiently large σ_j^2, policy $\mathbf{R}(1/2)$ can be arbitrarily close to the optimal policy $\mathbf{F}$. The latter is intuitively expected since for very large σ_j^2 and $\mu_j<\frac{1}{2}$, packet arrivals will occur mostly in bursts of very large length l. The delay of the packets in the burst will be approximately the same under both policies since the law of large number implies that, on the average, there will be needed approximately $2l$ slots to transmit the l packets under policy $\mathbf{R}(1/2)$ which is equal to the time required by policy $\mathbf{F}$.

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

Proof: The above statement is easily justified by inspection (Theorem 2). Note that as β increases, the capacity $1-\beta$ assigned to the station decreases and, thus, an increase in the induced mean packet delay is expected.

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

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

$$0<\beta_0^*=\frac{\sigma_j^2}{(1-2\mu_j)\mu_j+2\sigma_j^2}\leq\frac{1}{2} \tag{10}$$

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 $\mathbf{R}(0)$ is optimal in $\mathbf{P}(0)$. The latter is true since a policy which never makes the server unavailable ($\beta=0$) 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 $\mathbf{P}(\beta_0^*)$ optimal, that is $D^R(\beta_0^*)=D^F$. By using (1) and (8) and solving the previous equation with respect to β_0^* we obtain the result.

From (10) it can be seen that $\beta_0^*\to 1/2$ as $\sigma_j^2\to\infty$. That is, both policies become optimal in $\mathbf{P}(1/2)$ 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 minimum excess capacity $\mathbf{c}$ required in order for policy $\mathbf{R}(\beta)$ to become optimal in $\mathbf{P}$ is given by

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

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

III. Optimal policy in $\tilde{\mathbf{P}}$ for the communication system.

The developments of the previous section imply that the random, conflict-free slot assignment policy $\mathbf{R}(\beta)$ can be optimal in $\mathbf{P}$ at the expense of an additional capacity $\mathbf{c}$ compared to that under policy $\mathbf{F}$ (Corollary 3). In a 2-station communication system, the latter implies that an optimal policy $\mathbf{R}(\beta)$ for one station could 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 $\mathbf{R}(\beta)$ by one station may compensate for the increased mean packet delay of the other station. Policy $\mathbf{F}$ cannot be adjusted to asymmetric packet load conditions. Under such conditions, the optimal periodic, fixed slot assignment policy is not policy $\mathbf{F}$ and it can be, in general, only approximately implemented, [4]. Even if the traffic parameters were such that an implementation as suggested in [4] is possible and well performing, its complexity is significantly larger than that of policies $\mathbf{R}(\beta)$ and $\mathbf{F}$ and its adaptation to a dynamically changing traffic environment probably more so. For these reasons only policies $\mathbf{R}(\beta)$ and $\mathbf{F}$ are considered as candidate policies which are easily implemented. Furthermore, policy $\mathbf{R}(\beta)$ can easily become adaptive through the appropriate selection of the parameter β. A strategy for the identification of the optimal capacity allocation under policy $\mathbf{R}(\beta)$ is developed. By properly adjusting β, policy $\mathbf{R}(\beta)$ is capable of handling temporary severe queueing problems or temporary queueing instabilities and outperforming the (easily implemented) policy $\mathbf{F}$. These issues are investigated in this section.

Let $\tilde{\mathbf{P}} = \{\tilde{\mathbf{F}}, \tilde{\mathbf{R}}(\beta) \text{ for } 0 \le \beta \le 1\}$, where $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{R}}(\beta)$ are as defined in section I. An optimal policy in $\tilde{\mathbf{P}}$ is defined to be the policy which minimizes the mean delay of a random packet. The next theorem identifies the optimal policy in $\{\tilde{\mathbf{R}}(\beta) \text{ for } 0 \le \beta \le 1\}$. 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 $\mu_2 < \beta \le 1$ and convex U and

(c) $G(\beta)$ is convex U for $\mu_2 < \beta < 1 - \mu_1$

where

$$f_1(\beta) = \frac{\mu_1\beta + \sigma_1^2}{2(1-\mu_1-\beta)}, \quad f_2(\beta) = \frac{\mu_2(1-\beta) + \sigma_2^2}{2(\beta-\mu_2)} \tag{11}$$

and

$$G(\beta) = f_1(\beta) + f_2(\beta) \tag{12}$$

Proof: It is easy to show that the first and second derivatives of $f_1(\beta)$ with respect to β are strictly positive for $0 \le \beta \le 1 - \mu_1$ 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 $\mu_2 < \beta \le 1$ 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, [9].

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

$$(c_1-c_2)\beta^2+(2(1-\mu_1)c_2 - 2\mu_2c_1)\beta+\mu_2^2c_1-(1-\mu_1)^2c_2 = 0 \tag{13}$$

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

Proof: The optimal policy in $\{\tilde{\mathbf{R}}(\beta),\ 0 \le \beta \le 1\}$ is the policy $\tilde{\mathbf{R}}(\beta)$ which minimizes the induced mean packet delay

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

By substituting (8) 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, equation (13) is obtained.

The above theorem provides for the optimal policy $\tilde{\mathbf{R}}(\beta_0)$ in $\{\tilde{\mathbf{R}}(\beta) \text{ for } 0 \le \beta \le 1\}$ 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 $\tilde{\mathbf{R}}(\beta)$ 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.

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

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

where

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

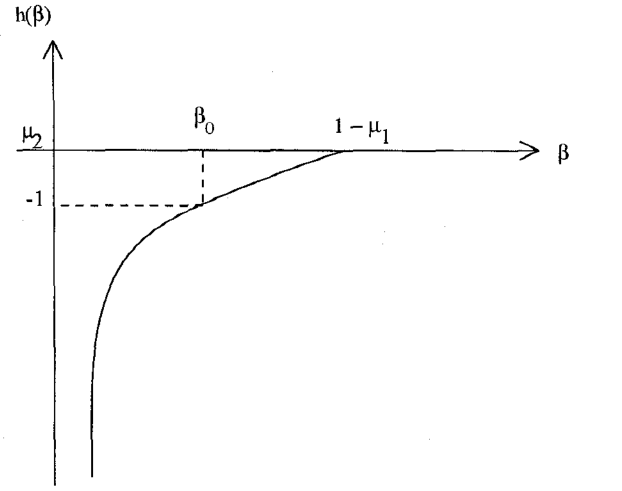

Fig. 3. A typical function $h(\beta)$.

and $f_1'(\beta)$, $f_2'(\beta)$ are the first derivatives of $f_1(\beta)$, $f_2(\beta)$ given by (11). A typical function $h(\beta)$ is shown in Fig. 3.

Proof: 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,

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

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, $f_2'(\beta_0)/f_1'(\beta_0) = -1$. The latter equation together with the monotonicity property of $h(\beta)$ complete the proof.

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

Corollary 4: Let $\tilde{\mathbf{R}}(\beta)$ be the currently adopted policy. The optimal point β_0 (policy $\tilde{\mathbf{R}}(\beta_0)$) can be reached by the following strategy: $\mathbf{S} = \{$ 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 $\{\tilde{\mathbf{R}}(\beta_j)\}_j$ which converges to the optimal policy in $\{\tilde{\mathbf{R}}(\beta)$ for $0 \leq \beta \leq 1\}$, $\tilde{\mathbf{R}}(\beta_0)$. Strategy $\mathbf{S}$ can be used for the adaptation of policy $\tilde{\mathbf{R}}(\beta)$ to the varying optimal policy $\tilde{\mathbf{R}}(\beta_0)$, in a dynamically changing environment. For instance, if the rates μ_1 and μ_2 change, strategy $\mathbf{S}$ is capable of adjusting the operation of the system so that optimality can be achieved, provided that some estimates of μ_1 and μ_2 be available. The mechanism for the generation of such estimates, its goodness and the detailed implementation of strategy $\mathbf{S}$ 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 for some intuitively expected results.

Corollary 5: 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 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.

Proof:

(a) The function $h(\beta)$, defined in Theorem 6, takes the following form for $\beta = \frac{1}{2} - \frac{\mu_1 - \mu_2}{2}$.

$$h\left(\frac{1}{2} - \frac{\mu_1 - \mu_2}{2}\right) = \frac{A + \sigma_2^2 - \mu_2^2}{A + \sigma_1^2 - \mu_1^2 + \mu_1 - \mu_2}$$

where $A = \frac{1}{4}(1 - \mu_1 + \mu_2)^2 + \mu_1\mu_2$. Theorem 5 implies that $\beta_0 < \beta = \frac{1}{2} - \frac{\mu_1 - \mu_2}{2}$ if $h(\frac{1}{2} - \frac{\mu_1 - \mu_2}{2}) > -1$. A necessary and sufficient condition for the latter is that $\sigma_2^2 - \mu_2^2 \leq \sigma_1^2 - \mu_1^2 + \mu_1 - \mu_2$, or $\sigma_1^2 \geq \sigma_2^2 - \theta$ where $\theta = (\mu_1 - \mu_2)(1 - \mu_1 - \mu_2)$. Clearly, if $\sigma_1^2 = \sigma_2^2 - \theta$ then $h(\frac{1}{2} - \frac{\mu_1 - \mu_2}{2}) = -1$ and thus $\beta_0 = \frac{1}{2} - \frac{\mu_1 - \mu_2}{2}$. The case for $\mu_2 > \mu_1$ can be studied in a similar way.

(b) If $\mu_1 = \mu_2$ then $\theta = 0$ and the result follows by applying part (a). This result can also be derived by noting that $\beta = 1/2$ satisfies the second order equation in (13), or by observing that, under these conditions, the convex function $G(\beta)$ in (12) is symmetric about $1/2$ since $G(\beta) = G(1 - \beta)$ for $\mu_1 < \beta < 1 - \mu_1$.

Corollary 5-(b) implies that under symmetric load ($\mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2$) the optimal policy $\tilde{\mathbf{R}}(\beta_0)$ assigns the same amount of channel capacity to each of the stations, which is intuitively expected. Corollary 5-(a) implies that for $\mu_1 = \mu_2$ the optimal policy $\tilde{\mathbf{R}}(\beta_0)$ assigns more channel capacity to the station with the largest variance. This is also intuitively expected since a larger variance results in more intense queueing problems (see (8)). Under equal packet arrival rates, the optimal policy will try to equate the intensity of the queueing problems (mean packet delay) in the two stations by assigning more capacity to the station with the largest variance. When $\mu_1 > \mu_2$ and $\sigma_1^2 = \sigma_2^2$, at least half of the remaining capacity $1 - \mu_1 - \mu_2$ (after the capacity necessary for the stability of each of the queues has been allocated) will be assigned to the heavier station 1 by the optimal policy, resulting in a capacity equal to $1 - \beta \geq \mu_1 + \frac{1 - \mu_1 - \mu_2}{2}$ or $\beta \leq \frac{1}{2} - \frac{\mu_1 - \mu_2}{2}$ for station 1, as indicated by Corollary 5-(a). Notice that even if $\sigma_1^2 = \sigma_2^2$, equality in the condition of Corollary 5-(a) is not achieved since $\theta > 0$ (for $\mu_1 > \mu_2$) and thus $\beta_0 < \frac{1}{2} - \frac{\mu_1 - \mu_2}{2}$ (strict inequality). That is, even if the arrival processes have similar structure as far as the mean queueing delay is concerned $(\sigma_1^2 = \sigma_2^2)$ and the stability of the associated queueing systems has been guaranteed, the remaining capacity is not equally assigned since the latter would result in unequal offered load to each queue, as determined by

$$\zeta_1 = \frac{\mu_1}{\mu_1 + \frac{1 - \mu_1 - \mu_2}{2}} > \frac{\mu_2}{\mu_2 + \frac{1 - \mu_1 - \mu_2}{2}} = \zeta_2$$

Given the identical structure of the two packet processes, $\zeta_1 > \zeta_2$ implies that most of the packets (since $\mu_1 > \mu_2$)

will undergo larger mean delay (since $\zeta_1 > \zeta_2$) and clearl an optimal policy should assign more capacity to statio 1. The more intense queueing problems associated wit station 1 under equal allocation of the remaining capacit $1-\mu_1-\mu_2$ (since $\zeta_1 > \zeta_2$) can be reduced by a better be having structure of the packet arrival process of station 1 regarding the induced mean packet delay, compared to tha of station 2 ($\sigma_1^2 < \sigma_2^2$). Under the latter conditions, th difference between the intensity of the queueing problem in the two stations will be decreased. When $\sigma_1^2 = \sigma_2^2 - \theta$ the structure of the packet arrival process of station 1 ha become sufficiently better than that of station 2 to be pos sible to balance the increase in the intensity of the queue ing problems of station 1 due to the larger offered load $\zeta_1 > \zeta_2$). Under the latter condition, the optimal polic assigns equally the remaining capacity ($\beta_0 = \frac{1}{2} - \frac{\mu_1-\mu_2}{2}$). If the structure of the packet arrival process of station 1 improves further ($\sigma_1^2 < \sigma_2^2 - \theta$), then more of the remaining capacity will be assigned to station 2 despite the fact that the corresponding offered load (ζ_2) is smaller than that of station 1 ($\beta_0 > \frac{1}{2} - \frac{\mu_1-\mu_2}{2}$). Under the latter conditions the structure of the packet arrival processes, rather than their intensities, is the dominating factor for the optimal allocation of the remaining capacity.

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

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

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

(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 $\mu_1 > 1/2$ or $\mu_2 > 1/2$ is satisfied, then one of the two queues would be unstable under policy $\mathbf{F}$ and, thus, the optimal policy in $\tilde{\mathbf{P}}$ will be $\tilde{\mathbf{R}}(\beta_0)$, since $\tilde{\mathbf{F}}$ would induce infinite mean packet delay to at least one of the stations.

(b) The mean packet delay under policy $\tilde{\mathbf{F}}$ 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 $\tilde{\mathbf{R}}(\beta_0)$ in $\{\tilde{\mathbf{R}}(\beta), 0 \leq \beta \leq 1\}$ 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 (8), $j = 1, 2$. $\tilde{\mathbf{R}}(\beta_0)$ is optimal in $\tilde{\mathbf{F}}$ if and only if $\tilde{D}^R(\beta_0) < \tilde{D}^F$. By substituting

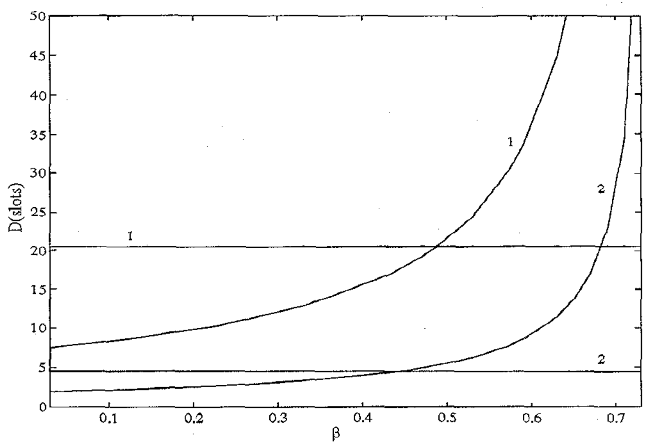

Fig. 4. Mean packet delay for a single station under policies $\mathbf{R}(\beta)$ and $\mathbf{F}$ for $\mu = 0.25$; 1:$\sigma^2 = 2.5$; 2:$\sigma^2 = 0.5$; D^F:constant; $D^R(\beta)$:convex U.

for $\tilde{D}^R(\beta_0)$ and $\tilde{D}^F$ and manipulating the resulting expression we obtain the desired result.

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

Proof: The above statement can be proven in two ways.

(a) It is easy to show by direct substitution that the necessary condition (b) of Theorem 7 is not satisfied.

(b) Corollary 5-(b) implies that the optimal policy in $\{\tilde{\mathbf{R}}(\beta),\ 0 \leq \beta \leq 1\}$ for $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$ is policy $\tilde{\mathbf{R}}(1/2)$. Under policy $\tilde{\mathbf{R}}(1/2)$, each station of the system operates under policy $\mathbf{R}(1/2)$. The latter policy is inferior to policy $\mathbf{F}$, as proven in Theorem 3. Thus, the optimal policy in $\{\tilde{\mathbf{R}}(\beta),\ 0 \leq \beta \leq 1\}$, $\tilde{\mathbf{R}}(1/2)$, is inferior to policy $\tilde{\mathbf{F}}$.

IV. Numerical results

In this section, the theory developed before is applied and some numerical results are obtained. These results illustrate the relative performance of the fixed and the random slot assignment policies and their dependence on the traffic characteristics. In Fig. 4 the mean packet delay induced under policies $\mathbf{R}(\beta)$, $0 \leq \beta \leq 1$ and $\mathbf{F}$ 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 $\mathbf{R}(\beta)$ to become optimal (.056 for $\sigma^2 = 2\mu$). As σ^2 increases both $D^R(\beta)$ and D^F increase as expected (see (1) and (8)). Notice that as σ^2 increases the additional capacity required for policy $\mathbf{R}(\beta)$ 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 (10) and it is always less than .5, which illustrates the optimality of policy $\mathbf{F}$ for $\beta = .5$; for $\beta \leq \beta_0^*$ policy $\mathbf{R}(\beta)$ 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 $\mathbf{R}(\beta)$ to be optimal is even smaller (.0025) since

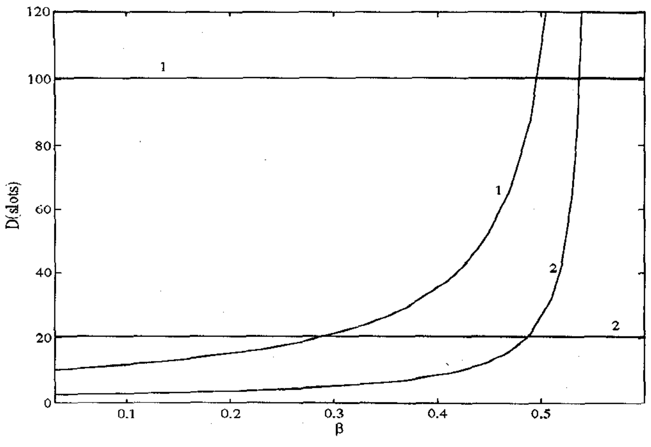

Fig. 5. Mean packet delay for a single station under policies $\mathbf{R}(\beta)$ and $\mathbf{F}$ for $\mu = 0.45$; $1{:}\sigma^2 = 4.5$; $2{:}\sigma^2 = 0.9$; D^F:constant; $D^R(\beta)$:convex U.

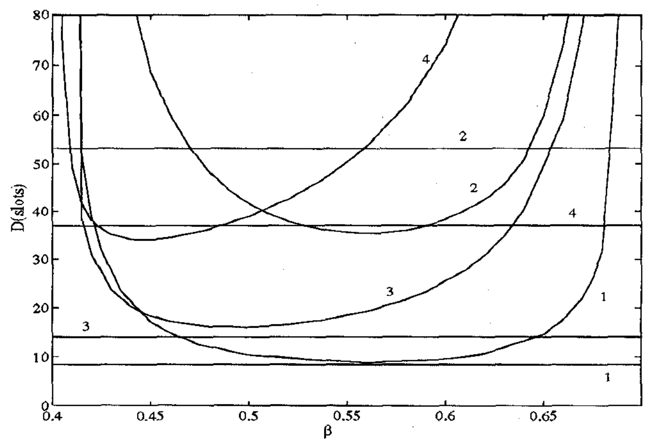

Fig. 7. Mean packet delay for the two stations under policies $\tilde{\mathbf{R}}(\beta)$ and $\tilde{\mathbf{F}}$ for $\mu_1 = .3$, $\mu_2 = .4$; $1{:}\sigma_1^2 = .6$, $\sigma_2^2 = .8$; $2{:}\sigma_1^2 = 3$, $\sigma_2^2 = 4$; $3{:}\sigma_1^2 = 3$, $\sigma_2^2 = .4$; $4{:}\sigma_1^2 = 10$, $\sigma_2^2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

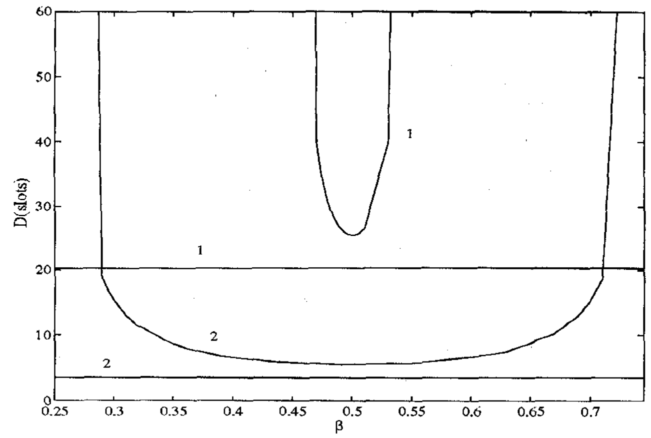

Fig. 6. Mean packet delay for the two symmetric stations under policies $\tilde{\mathbf{R}}(\beta)$ and $\tilde{\mathbf{F}}$ $1{:}\mu = 0.45$, $\sigma^2 = 0.9$; $2{:}\mu = 0.25$, $\sigma^2 = 0.5$; D^F:constant; $D^R(\beta)$:convex U.

$\sigma^2 = 4.5$ is large. Practically, both policies $\mathbf{R}(1/2)$ and $\mathbf{F}$ 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 $\mathbf{F}$) can be allocated to the station, then the induced mean packet delay under $\mathbf{R}(.45)$ is about half the one induced under policy $\mathbf{F}$ (52 versus 100 slots). The latter observation implies that the theoretical advantage (optimality) of policy $\mathbf{F}$ over policy $\mathbf{R}(1/2)$ may disappear in practical cases, if some additional capacity is offered to the station. The performance improvement may be tremendous at the cost of utilizing 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 $\tilde{\mathbf{R}}(\beta)$, $0 \leq \beta \leq 1$ and $\tilde{\mathbf{F}}$ is plotted, for the case of symmetric traffic load. Notice that the optimal policy in $\{\tilde{\mathbf{R}}(\beta),\ 0 \leq \beta \leq 1\}$ is policy $\tilde{\mathbf{R}}(1/2)$, as implied by Corollary 5-(b). The optimal policy in $\tilde{\mathbf{P}}$ is policy $\tilde{\mathbf{F}}$, 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 structure of the packet arrival processes is not strong enough to render policy $\tilde{\mathbf{R}}(\beta_0)$ optimal. For $\sigma_1^2 = 3$ and $\sigma_2^2 = 4$, policy $\tilde{\mathbf{R}}(\beta_0)$ 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 $\tilde{\mathbf{R}}(\beta_0)$ is due, first, to the larger variance of the traffics, which under symmetry would bring policy $\tilde{\mathbf{R}}(\beta_0)$ very close to the optimal policy $\mathbf{F}$ (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 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 $\tilde{\mathbf{F}}$ 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 dominates 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 $\tilde{\mathbf{R}}(\beta_0)$ 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 $\tilde{\mathbf{R}}(\beta_0)$ optimal. The asymmetry in the rates

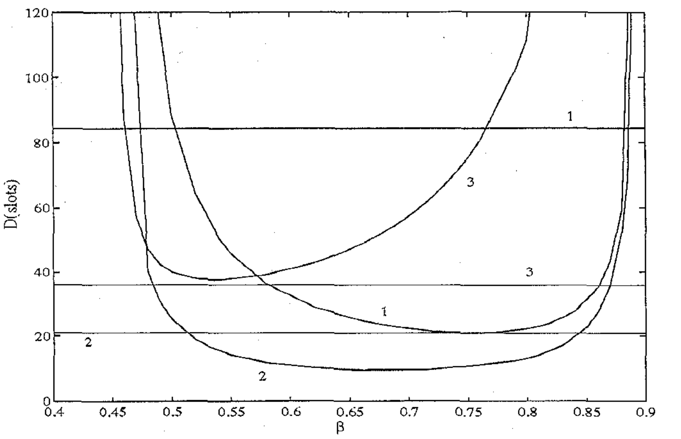

Fig. 8. Mean packet delay for the two stations under policies $\tilde{\mathbf{R}}(\beta)$ and $\tilde{\mathbf{F}}$ for $\mu_1 = .1$, $\mu_2 = .45$; 1:$\sigma_1^2 = 1$, $\sigma_2^2 = 4.5$; 2:$\sigma_1^2 = 1$, $\sigma_2^2 = 1$; 3:$\sigma_1^2 = 12$, $\sigma_2^2 = .45$; D^F:constant; $D^R(\beta)$:convex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

(a)

H	D^F
2	24936.0
4	6.2
6	69.1
8	1064.0
10	6.4
20	5.7
30	6.5
40	6.7
50	7.6
100	10.8
150	14.3

(b)

H	D^F
4	836.5
6	984.0
8	10.6
10	16.8
20	19.7
30	31.0
40	41.3
50	39.5
100	103.0
150	124.9

Table I. Mean delay results under the periodic, fixed slot assignment policy (D^F) for various values of the desired implementation horizon H: (a) under fixed traffic conditions (the mean delay under the random policy is equal to 9.9); (b) under varying traffic conditions (the mean delay under the random policy is equal to 8.2).

is capable of rendering policy $\tilde{\mathbf{R}}(\beta_0)$ optimal even if the structure of the two processes is symmetric ($\sigma_1^2 = \sigma_2^2 = 1$). When the asymmetry in the structure 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 $\tilde{\mathbf{F}}$ optimal ($\sigma_1^2 = 12$, $\sigma_2^2 = .45$).

The following discussion[1] (along with the results shown in Table I) is presented to illustrate the complexity in the implementation of the periodic, fixed slot assignment policy, its dependence on the implementation horizon and its potential inefficiency when traffic conditions change. The optimal capacity allocation determined by the random, conflict-free slot assignment policy can be implemented within a horizon of one slot, by simply setting properly the value of β. This is not the case with the optimal periodic, fixed slot assignment policy, except from special cases as such as when $\tilde{\mathbf{F}}$ is the optimal policy. Suppose that r_1 ($0 \leq r_1 \leq 1$) of the total available capacity needs to be allocated to station 1, as determined by the optimal periodic, fixed slot assignment policy, [4]. One simple way to implement (approximately, in general) this allocation is by selecting a desired implementation horizon H (in slots) and allocate $n_1 = \text{Int}\{r_1 H\}$ slots to station 1 and $n_2 = \text{Int}\{(1 - r_1)H\}$ slots to station 2, where $\text{Int}\{x\}$ denotes the largest integer which is smaller than x. After $H' = n_1 + n_2$ slots the implementation of the next horizon is initiated. Thus, the actual implementation horizon H' is in general smaller than the desired one (H). The n_1 and n_2 slots are provided in an alternating fashion. In general, $r_1' = n_1/H' \neq r_1$ and $r_2' = n_2/H' \neq r_2$. That is, the actual allocated capacity within a finite actual implementation horizon H' is different from the optimal. In general, it is impossible to achieve the theoretical allocations under the periodic, fixed slot assignment policy [4]. Table I-(a) illustrates the impact of the desired implementation horizon H on the induced mean delay. The results are based on simulations run for 100,000 slots; results remained unchanged for larger simulation time. The packet arrival processes to the stations are assumed to be Bernoulli with rates $\mu_1 = .3$ and $\mu_2 = .6$. Notice that when H is sufficiently small, the large deviation of the actual capacity allocated to a station from the optimal, may result in an allocation which is below that required for the queue stability. In such cases, the mean delay results will become unbounded ($H = 2$ and $H = 8$ in Table I-(a)).

For sufficiently large H, the allocated capacity approaches the optimal one. The average delay increases, though, due to the fact that after the slot alternation period is over, the remaining of the actual implementation horizon is allocated to the heaviest station. As a result, packets of the lightest station may have to wait for an increasingly (as H' increases) large sequence of slots, before the slot alternation period is initiated in the horizon that follows and these packets be given a chance. The induced mean delay under the random slot assignment policy is equal to 9.9 slots. Notice that it is higher than the lowest ones shown in Table I-(a) (for $H = 20$ and $H = 4$), obtained under the periodic, fixed slot assignment policy. This difference decreases as the variance of the arrival process increases, as it has been proven.

In Table I-(b), similar results are presented when the rate of the arrival processes changes in the next slot with probability $p = .8$. When it changes, it takes the value .8 with probability q and the value .1 with probability 1-q; q is selected so that the long term packet arrival rate for stations 1 and 2 be equal to $\mu_1 = .3$ and $\mu_2 = .6$, same with those considered for the results shown in Table I-(a). To avoid introducing the rate estimation error, it is assumed that the new rates are known to both stations under both policies. Notice that the random slot assignment policy can implement the optimal capacity allocation in the next slot based on the current rates. This is not the case with the periodic, fixed slot assignment policy which requires an actual implementation horizon H'. As a result, the lat-

[1]Recommended during the reviewing process

ter policy does not seem to be capable of responding as efficiently to fast rate changes. According to the results shown in Table I-(b), the random slot assignment policy induces the lowest mean packet delay, which is equal to 8.2 slots. Furthermore, notice that for the desired implementation horizon H which induces the lowest mean delay under constant packet arrival rates ($H = 20$ and $H = 4$ in Table I-(a)), the resulting mean delay is far from the minimum under traffic changing conditions ($H = 8$ in Table I-(b)).

V. Conclusions

In this paper, a random, conflict-free slot assignment policy, $\tilde{\mathbf{R}}(\beta)$, has been considered for the allocation of a common resource (channel) between two distributed entities. This policy has been analyzed and its performance under both symmetric and asymmetric traffic conditions has been investigated. The effect of both the rate and structure (variance) of the packet arrival process on the performance of policy $\tilde{\mathbf{R}}(\beta)$ has been fully investigated. Although the standard TDM policy, $\tilde{\mathbf{F}}$, which would assign the slots to the stations in a deterministic and periodic fashion, performs better under symmetric traffic load, it is inferior to policy $\tilde{\mathbf{R}}(\beta)$ in most practical cases. The latter has been shown to be usually the case under asymmetric (regarding the rate or the structure) packet traffic load. Furthermore, even if the traffic load is considered to be symmetric, policy $\tilde{\mathbf{F}}$ fails to adjust the capacity assignment to temporary traffic fluctuations which are present in most practical cases. Policy $\tilde{\mathbf{R}}(\beta)$ can easily adjust the capacity allocation to the current traffic conditions. A very simple strategy has been developed for this purpose which achieves the optimal capacity allocation under policy $\tilde{\mathbf{R}}(\beta)$. When more than two stations share the common channel the performance of the corresponding policies $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{R}}(\beta)$ can be evaluated by following the same approach. The mean packet delay for a single station under the corresponding policy $\tilde{\mathbf{R}}(\beta)$ is given by Theorem 2, where the capacity $1-\beta$ assigned to the station is properly adjusted. Under policy $\tilde{\mathbf{F}}$ the mean packet delay can be obtained by following the approach shown in the proof of Theorem 1 (or can be found in [4]). Notice also that the proof of Theorems 1 and 2 provide for a method for the derivation of the mean packet delay under both policies when packets arrive to the stations according to a Markov modulated process.

To summarize, the major contributions of this paper are the following:

(a) A simple unified method for the calculation of the mean packet delay induced by the fixed (TDMA) and the random, conflict-free slot assignment policies has been developed which, unlike other methodologies, is applicable to similar system under dependent packet arrival processes described by a Markov Modulated model (see proofs of Theorems 1 and 2).

(b) Although policy $\mathbf{F}$ outperforms policy $\mathbf{R}(\beta)$ under symmetric traffic conditions, it has been shown that the difference in performance is insignificant in most practical cases when the variance of the traffic is significant; the two policies are asymptotically the same as the variance of the traffic approaches infinity. Thus, the superiority of the periodic, fixed slot assignment policy over the random, conflict-free slot assignment policy, shown in [4], could be only of theoretical interest, in view of the asymptotic optimality of the random, conflict-free slot assignment policy shown here and the fact that the optimal random, conflict-free slot assignment policy is always feasible (as opposed to the optimal periodic, fixed slot assignment one [4]), and it is easily become an adaptive one.

(c) A simple strategy, based on a threshold test, has been developed for the adjustment of the optimal random, conflict-free slot assignment policy to dynamically changing traffic conditions. Under such conditions, it has been illustrated that the random slot assignment policy may outperform the periodic, fixed slot assignment policy, due to its fast adaptability to the changing traffic conditions.

(d) The impact of the structure (variance) of the arrival processes on the performance of both policies has been investigated in detail. Numerical results have illustrated that the structure of the arrival process can be the major factor for the determination of the optimal capacity allocation, which may turn out to be against expectations based on the rates.

Appendix

A very simple proof of equation (2) can be obtained based on the observation that the packet service time (or the work associated with one packet) is deterministic and equal to 1 slot. Since the queueing system is work-conserving $E\{L^s\} = E\{L^s_{\mathrm{FIFO}}\}$ where $E\{L^s\}$ denotes the expected value of the work in the system and $E\{L^s_{\mathrm{FIFO}}\}$ denotes the corresponding quantity associated with the equivalent system operating under the FIFO service discipline. Since each unit of work in the system corresponds to one customer, the previous equation implies that $E\{Q^s\} = E\{Q^s_{\mathrm{FIFO}}\}$ where $E\{Q^s\}$ denotes the average number of customers in the system and Q^s_{FIFO} denotes the corresponding quantity associated with the equivalent system operating under the FIFO service discipline. From Little's theorem it is obtained that $E\{Q^s\} = \tilde{\mu}_j 1 + \mu_j D^F_j$, $E\{Q^s_{\mathrm{FIFO}}\} = (\tilde{\mu}_j + \mu_j)D_{\mathrm{FIFO}}$. Equation (2) is easily obtained from the above.

References

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

[2] I. Rubin, "Message delays in FDMA and TDMA communications channels," *IEEE Trans. Commun.*, vol. 27, May 1979.

[3] S. Lam, "Delay analysis of a time division multiple access (TDMA) channel," *IEEE Trans. Commun.*, vol. 25, Dec. 1977.

[4] M. Hofri and Z. Rosberg, "Packet delay under the golden ratio weighted tdm policy in a multiple-access

channel," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 3, 1987.

[5] I. Stavrakakis, "Analysis of a statistical multiplexer under a general input traffic model," in *INFOCOM'90 conference*, (San Francisco), 5–7 June 1990.

[6] G. Barberis, "A useful tool in priority queueing," *IEEE Trans. Commun.*, vol. 28, Sept. 1980.

[7] D. Heyman and M. Sobel, *Stochastic Models in Operations Research*, vol. 1. McGraw-Hill, 1982.

[8] L. Schrage, "An alternative proof of a conservation law for the G/G/1 queue," *Operations Research*, vol. 18, pp. 185–187, 1970.

[9] A. Roberts and D. Varberg, *Convex Functions*. New York: Academic Press, 1973.

Ioannis Stavrakakis (S'85-M'89-SM'93) received the Diploma in Electrical Engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, 1983, and the Ph.D. degree in Electrical Engineering from the University of Virginia, 1988.

Since 1988, he has been an Assistant Professor in the Department of Electrical Engineering and Computer Science, University of Vermont. His research interests are in stochastic system modeling, teletraffic analysis and discrete-time queueing theory, with primary focus on the design and performance evaluation of Broadband Integrated Services Digital Networks (B-ISDN).

Dr. Stavrakakis is a Senior Member of IEEE and a Member of the IEEE Communications Society, Technical Committee on Computer Communications. He has organized and chaired sessions, and has been a technical committee member, for conferences such as GLOBECOM, ICC and INFOCOM.