ON ACCESS CONTROL DISCIPLINES FOR A TDMA SYSTEM WITH MULTIPLE-RATE REAL-TIME SOURCES*

MING LOUIE and IZHAK RUBIN
University of California
Los Angeles, California

Summary. We consider a communication medium, such as a satellite communication channel, which is shared by a number of sources on a time-division multiple-access (TDMA) basis. Sources require real-time transmission over the channel. A central controller stores the requests for transmission by the sources and assigns the freed slots within the time-frame to the appropriate sources, following a dynamic demand-assignment access control discipline. Sources are further assumed to require real-time transmission at different information rates, and thus require different number of slots per frame. Two access-control disciplines are studied, structurally optimized, and compared. A fixed-assignment discipline divides the frame slots among the various source classes on a predetermined basis. A priority-assignment discipline dynamically modifies the latter division by allowing one class of requests to utilize the freed slots of the other. The analysis demonstrates the extent to which the second scheme is preferrable, incorporating a message maximal waiting-time objective function and adaptability considerations.

Introduction. We are considering a communication channel which is shared by a number of information bearing sources on a time-division multiple-access (TDMA) basis. We assume the sources to require real-time transmission over the channel, once permission to transmit across the channel is granted. Since the number of sources which wish to transmit simultaneously over the channel is generally much larger than that allowed when considering the channel capacity, we need to incorporate a dynamic demand-assignment access control discipline in order that the channel will be efficiently utilized and the message response times appropriately minimized. For that purpose, we assume that a central controller has been established. The latter stores the requests for transmission by the various sources and accordingly assigns the freed channel slots to the appropriate sources. (The central controller communicates with the sources through either a separate established local control network, using for example telephone lines, or by utilizing a control channel as part of the major communication channel.)

*This work was supported by the Office of Naval Research under Contract N00014-75-C-0609 and by the National Science Foundation under Grant ENG 75-03224.

We furthermore assume that the sources require real-time transmission at different information rates. Thus, we can distinguish between classes of sources according to their information rates. For example, one class of sources can represent teletypewriters, while another class may represent vocoders which generate messages at an higher information rate. See [1][3] for examples of actual satellite-communication systems which operate as described above, integrating multiple-rate sources which require real-time transmission.

We assume here the sources to be divided into classes 1 and 2, so that a class i source, i = 1,2, generates messages according to information rate R_i , $R_1 > R_2$. Class 1 messages are further assumed to have longer average message lengths. We then consider two demand-assignment access control disciplines. The first discipline, the Fixed Assignment scheme, divides the time frame between the two source classes on a fixed basis, thus allowing the central controller to separately assign each class its own dedicated channel slots, on a demand assignment basis. The second discipline, called the Priority Assignment scheme, incorporates into the former scheme the possibility of allowing class 2 messages to utilize class 1 slots, when the latter are available (i.e.; not required for class 1 transmissions). Clearly, the second scheme makes the access control procedure more flexible and adaptive to fluctuations in the traffic characteristics of the sources.

In this paper, we study and compare the performance of the above-mentioned two schemes. We derive, for each scheme, results for the message queueing delays and obtain its optimal structure, in terms of a system maximal average delay measure (including message delays normalized by the corresponding average message lengths). The two schemes are then compared, and the extent of adaptivity (sensitivity to traffic fluctuations) of the Priority Assignment scheme is noted.

The System Model. Incorporating the given values of channel capacity, message information rates and the TDMA frame format we arrive at the following system model. A class 2 message is set to require a single slot each time-frame (to satisfy its requirement for real-time transmission at information rate R_2) . A class 1 message then requires M slots per frame, M > 1 . The total number of slots in each frame is K . Requests for transmission by class 1 and 2 sources arrive at the controller according to independent Poisson streams with rates λ_1 and λ_2 [requests/sec.], respectively. Each request by a class i source corresponds to a class i message, i - 1,2 . Class 1 and 2 messages are assumed to be statistically independent, with class i message lengths being i.i.d. and following an exponential distribution. We subsequently assume the transmission-times (holding times) required by class i messages (involving a single slot/frame for a class 2 message and the simultaneous utilization of M slots/frame for class 1 messages) to be i.i.d. exponentially distributed random-variables with mean μ_i^{-1} , i = 1,2 .

Under the Fixed Assignment scheme, the K slots of a frame are divided into two regions. Region 1 is assigned MK_1 slots, so that region 2 contains then $K_2 = K\text{-}MK_1$ slots. An arriving class i message (request) is then served by the controller, on a first-come first-served basis, which appropriately assigns it a server in region i (being composed of a single slot in region 2 and of M slots in region 1). The controller assigns available slots (servers) in region i to the corresponding class i messages, independent of assignments in region j to class j messages, $i \neq j$, $i = 1,2$, $j = 1,2$.

When the Priority Assignment scheme is employed, the frame is again divided into above regions 1 and 2 with appropriate number of slots MK_1 and K_2 , respectively, However, now class 2 messages are slowed to utilize any free class 1 slots if required. Thus, if an arriving class 2 message finds region 2 to be fully occupied, it is assigned a free slot in region 1, if available. Class 1 messages are however assigned only slots (servers, each containing M slots) within region 1, and are granted a higher priority status w.r.t. their assignment. Since class 1 messages are assumed to be longer than class 2 messages, only a small difference in performance analysis and design will occur if a preemptive priority status is given to class 1 messages with regard to region 1 slots, rather than a nonpreemptive priority status. Since the preemptive priority assumption results with a simplified queueing analysis, We will follow it.

Under both access control disciplines, we denote the (steady-state) average waiting-time of class 1 and 2 messages, under slot assignments K_1 , k_2 , and arrival rates λ_1 , λ_2 , by $W_1(K_1,\lambda_1)$ and $W_2(K_2,\lambda_2)$ respectively.

Steady-State Message Delay Analysis Under the Fixed Assignment and Priority Assignment Schemes

Fixed Assignment. Under a Fixed Assignment strategy, the system operates as two separate queueing systems. Thus, a class i message waiting-time distribution is equal to that of a customer in an $M/M/K_i$ queueing system i = 1,2 . (See [4], [5]). Therefore,

$$W_i(K_i,\lambda_i) = P_{x_i}(K_i)\ [K_i\mu_i(1-\frac{\lambda_i}{K_i\mu_i})^2]^{-1}$$

$$\text{for}\quad \lambda_i < K_i\mu_i\ ,\ i = 1,2, \tag{1}$$

where

$$P_{x_i}(K_i) = P_{x_i}(0)\,\frac{1}{K_i!}\,(\frac{\lambda_i}{\mu_i})^{K_i}\ ,\ i = 1,2\ , \tag{2}$$

$$P_{x_i}(0) = \left[1+\frac{\lambda_i}{\mu_i}+\ldots+\frac{1}{(K_i-1)!}\left(\frac{\lambda_i}{\mu_i}\right)^{K_i-1} + \frac{1}{\left(1-\frac{\lambda_i}{K_i\mu_i}\right)}\frac{1}{K_i!}\left(\frac{\lambda_i}{\mu_i}\right)^{K_i}\right]^{-1},$$

$$\text{for } \lambda_i < K_i\mu_i \ , \ i = 1,2 \ . \qquad (3)$$

Priority Assignment. Employing the Priority Assignment scheme, class 1 messages have waiting-time distribution determined by an $M/M/K_1$ queueing system. Hence,

$$W_1(K_1,\lambda_1) = P_{x_1}(K_1)\left[K_1\mu_1\left(1-\frac{\lambda_1}{K_1\mu_1}\right)^2\right]^{-1}, \quad \lambda_1 < K_1\mu_1 \ , \qquad (4)$$

where

$$P_{x_1}(K_1) = P_{x_1}(0)\frac{1}{K_1!}\left(\frac{\lambda_1}{\mu_1}\right)^{K_1}, \qquad (5)$$

and

$$P_{x_1}(0) = \left[1+\frac{\lambda_1}{\mu_1}+\ldots+\frac{1}{(K_1-1)!}\left(\frac{\lambda_1}{\mu_1}\right)^{K_1-1} + \frac{1}{\left(1-\frac{\lambda_1}{K_1\mu_1}\right)}\frac{1}{K_1!}\left(\frac{\lambda_1}{\mu_1}\right)^{K_1}\right]^{-1},$$

$$\lambda_1 < K_1\mu_1 \ . \qquad (6)$$

A queueing analysis for class 2 messages can be based on the Markovian characteristics of the joint queue size process of class 1 and 2 customers. A set of coupled linear equations can then be written for the state-state joint queue-size probabilities. This set of equations assumes however a complicated form so that no simple useful analytical expressions for the class 2 message waiting-time distributions result. Consequently, we develop an approximation to the latter variable by considering an appropriate single server M/G/1 queueing system. The latter queueing system is assumed to have customers arriving in a Poisson stream with intensity λ_2 The single server is assumed to process messages at rate $m\mu_2$ when m free slots are available for class 2 messages. Subsequently, we can consider class 2 customers to require i.i.d. service times with conditional exponential distribution of mean $(m\mu_2)^{-1}$, given m. Thus, unconditionally, the customer service distribution assumed is

$$\tilde{B}(x) = \sum_{j=0}^{K_1}\left(1-e^{-(K_2+jM)x}\right)\tilde{P}_m(K_2+jM) \ , \ K_2 > 0 \ , \ x \geq 0 \ , \qquad (7)$$

where

$\tilde{P}_m(i)$ = P(there are i free slots available for class 2 messages).

From the results of the M/M/K_1 queueing system, Eqs. (4)-(6), we obtain

$$\tilde{P}_m(K_2) = P_{x_1}(K_1)\left(1-\frac{\lambda_1}{K_1\mu_1}\right)^{-1}, \quad \lambda_1 < K_1\mu_1 , \tag{8}$$

$$\tilde{P}_m(K_2+jM) = P_{x_1}(0)\,\frac{1}{(K_1-j!)}\left(\frac{\lambda_1}{\mu_1}\right)^{K_1-j}, \quad 1 \le j \le K_1 . \tag{9}$$

By applying the Pollazeck-Khintchine formula, the average waiting time of a message in this M/G/1 system is given by (E(•) being w.r.t. $\tilde{P}_m$(•)),

$$\overline{W}_2(K_2,\lambda_2) = \lambda_2 E\left[\frac{1}{m^2}\right]\left[\mu_2^2\left(1-\frac{\lambda_2}{\mu_2}E\left[\frac{1}{m}\right]\right)\right]^{-1} . \tag{10}$$

We further modify the average waiting-time approximation as follows. The average number of slots available for class 2 messages is

$$E[m] = \sum_{j=0}^{K_1}(K_2+jM)\tilde{P}_m(K_2+jM) = K-M\lambda_1\mu_1^{-1} . \tag{11}$$

We note that E[m] is independent of K_1 . The system is assumed to be stable, yielding finite average class 2 waiting-time iff $\lambda_2 < E[m]\mu_2$. Therefore, to obtain a waiting-time approximation which yields instability at the latter point, we modify (10) to obtain the approximation

$$W_{2p}(K_2,\lambda_2) = \lambda_2 E\left[\frac{1}{m^2}\right]\left[\mu_2^2\left(1-\frac{\lambda_2}{E[m]\mu_2}\right)\right]^{-1}, \quad \lambda_2 < E[m]\mu_2 . \tag{12}$$

This approximation is compared with simulation results in two cases shown in Figure 1 and Figure 2. (In these figures we set $\overline{\gamma}_2 = W_{2p} + \mu_2^{-1}$) Around high and low traffic intensities, the approximation is seen to be excellent, whereas for moderate traffic intensities, the approximation is good and observed to be very satisfactory for design purposes.

Optimal Slot Allocations under Fixed and Priority Assignments. To determine the optimal slot allocation in each region for the Fixed and Priority assignment disciplines, we define a system delay performance measure W as the maximal average message normalized waiting time. Thus, we set

$$W(K_1,K_2,\lambda_1,\lambda_2) = \text{Max}[\tilde{W}_1(K_1,\lambda_1) \ , \ \tilde{W}_2(K_2,\lambda_2)] \ , \tag{13}$$

where

$$\tilde{W}_1(K_1,\lambda_1) = W_1(K_1,\lambda_1)/\mu_1^{-1} \ , \tag{14}$$

$$\tilde{W}_2(K_2,\lambda_2) = W_2(K_2,\lambda_2)/\mu_2^{-1} . \tag{15}$$

We thus denote by $\tilde{W}_i(K_i,\lambda_i)$ the ratio between the average waiting-time of a class i message and its required service time. The optimal slot assignments, denoted as K_1^* and $K_2^* = K-MK_1^*$, are those yielding the minimal value of $W(\cdot)$, denoted as $W^*(\cdot)$. We thus have,

$$W^*(K_1^*,K_2^*,\lambda_1,\lambda_2) = \min_{K_1} W(K_1,K-MK_1,\lambda_1,\lambda_2) \ . \tag{16}$$

In order to find K_1^* and K_2^*, given λ_1 and λ_2, we introduce the following two definitions. The region of arrival rates yielding a finite message average waiting time $\Omega(K_1)$, for each given K_1 is defined by $\Omega(K_1) = \{(\lambda_1,\lambda_2) \ , \ \lambda_1 \geq 0 \ , \ \lambda_2 \geq 0 : \tilde{W}_1(K_1,\lambda_1) < \infty \text{ and } \tilde{W}_2(K-MK_1,\lambda_2) < \infty\}$. The region of arrival rates for which $K_1^* = K_1$, denoted by $\psi(K_1)$, for a given K_1 value and yielding a finite average message waiting time, is defined by $\psi(K_1) = \{(\lambda_1,\lambda_2):(\lambda_1,\lambda_2)\in\Omega(K_1) \text{ and } K_1^* = K_1\}$. Let $[x]$ denote the largest integer less than x. We prove the following Theorem which yields a procedure for calculating K_1^* given (λ_1,λ_2).

<u>Theorem 1</u>. -Partition the region $\Omega = \bigcup_{i=1}^{[K/M]} \Omega(i)$ into the disjoint regions $A(K_1)$, $1 \leq K_1 \leq [K/M]$, $\Omega = \bigcup_{K_1=1}^{[K/M]} A(K_1)$, where $A(K_1)\subset\Omega(K_1)$ and is further contained between boundaries $B(K_1-1)$ and $B(K_1)$, where

$$B(K_1) = \{(\lambda_1,\lambda_2) : \tilde{W}_1(K_1,\lambda_1) = \tilde{W}_2(K-M(K_1+1) \ , \ \lambda_2) < \infty\} \ ,$$

$$1 \leq K_1 \leq [K/M]-1 \ ,$$

$$B(0) = \{(\lambda_1,\lambda_2) : \lambda_1 = 0\} \ ,$$

$$B([K/M]) = \{(\lambda_1,\lambda_2) : \lambda_2 = 0\} \ .$$

Then,

$$A(K_1) = \psi(K_1) \ . \tag{17}$$

<u>Proof</u>. -Let $C(K_1) = \{\Omega(K_1)\cap\Omega(K_1-1)\}\cup\{\Omega(K_1)\cap(K_1+1)\}$. Any point (λ_1,λ_2) that belongs to the region $\Omega(K_1)-C(K_1)$ satisfies the following equations, by construction.

$$\tilde{W}_1(K_1-1,\lambda_1) = \infty , \tag{18}$$

$$\tilde{W}_2(K-M(K_1+1) , \lambda_2) = \infty , \tag{19}$$

$$\tilde{W}_1(K_1,\lambda_1) < \infty , \tag{20}$$

$$\tilde{W}_2(K-MK_1,\lambda_2) < \infty . \tag{21}$$

Subsequently, in this region,

$$\tilde{W}_1(K_1-i,\lambda_1) > \tilde{W}_1(K_1-1,\lambda_1) = \infty , \quad \text{for} \quad 1 < i < K_1 , \tag{22}$$

and

$$W_2(K-M(K_1+i) , \lambda_2) > W_2(K-M(K_1+1) , \lambda_2) = \infty ,$$
$$\text{for} \quad 1 < i \leq [K/M]-K_1 . \tag{23}$$

Hence, $K_1{}^* = K_1$ for any point (λ_1,λ_2) in the region $\Omega(K_1)-C(K_1)$.

Now, we consider the region $C(K_1)$. We examine first the boundary $B(K_1-1)$. From the construction of $C(K_1)$, we obtain

$$\{(\lambda_1,\lambda_2):: 0 \leq \lambda_1 < (K_1-1)\mu_1 , 0 \leq \lambda_2 < (K-MK_1)\mu_2\}\subset C(K_1) .$$

We also observe the following properties.

<u>Property 1</u>. -$\tilde{W}_1(K_1-1,0) = 0$ and $\tilde{W}_1(K_1-1,\lambda_1)$ is a monotonic increasing function w.r.t. λ_1 given K_1-1 , $K_1-1 > 0$.

<u>Property 2</u>. -$\tilde{W}_2(K-MK_1,0) = 0$ and $\tilde{W}_2(K-MK_1,\lambda_2)$ is a monotonic increasing function w.r.t. λ_2 given $K-MK_1$, $K-MK_1 > 0$.

Thus $B(K_1-1)$ is well defined in $C(K_1)$ and for any point $(\lambda_1,\lambda_2)\in B(K_1-1)$, we have

$$\tilde{W}_2(K-MK_1,\lambda_2+\Delta) > \tilde{W}_2(K-MK_1,\lambda_2) , \text{ for any } \Delta > 0 , \tag{24}$$

and

$$\tilde{W}_2(K-MK_1,\lambda_2-\Delta) < \tilde{W}_2(K-MK_1,\lambda_2) , 0 < \Delta \leq \lambda_2 . \tag{25}$$

The same conclusions apply to $B(K_1)$ and therefore $B(K_1)$ is well defined in $C(K_1)$ and for any point $(\lambda_1,\lambda_2)\in B(K_1)$, we have

$$\tilde{W}_2(K-M(K_1+1) , \lambda_2+\Delta) > \tilde{W}_2(K-M(K_1+1) , \lambda_2) , \Delta > 0 , \tag{26}$$

and

$$\tilde{W}_2(K-M(K_1+1)\ ,\ \lambda_2-\Delta) < \tilde{W}_2(K-M(K_1+1)\ ,\ \lambda_2)\ ,\ \lambda_2 \geq \Delta > 0\ . \tag{27}$$

Consider any point $(\lambda_1,\lambda_2)\epsilon B(K_1-1)$. Since

$$\tilde{W}_1(K_1-1,\lambda_1) > \tilde{W}_1(K_1,\lambda_1)\ ,\ \lambda_1 \neq 0\ ,$$

we have, by definition of $B(K_1-1)$,

$$W(K_1,K-MK_1,\lambda_1,\lambda_2) = \text{Max}(\tilde{W}_1(K_1,\lambda_1)\ ,\ \tilde{W}_2(K-MK_1,\lambda_2)) = \tilde{W}_2(K-MK_1,\lambda_2). \tag{28}$$

Similarly, since

$$\tilde{W}_2(K-M(K_1-1)\ ,\ \lambda_2) < \tilde{W}_2(K-MK_1,\lambda_2)\ ,\ \lambda_2 \neq 0\ ,$$

we obtain

$$\begin{aligned} W(K_1-1,K-M(K_1-1)\ ,\ \lambda_1,\lambda_2) &= \text{Max}(\tilde{W}_1(K_1-1,\lambda_1)\ ,\ \tilde{W}_2(K-M(K_1-1)\ ,\ \lambda_2) \\ &= \tilde{W}_1(K_1-1,\lambda_1)\ . \end{aligned} \tag{29}$$

By similar arguments, for any point $(\lambda_1,\lambda_2)\epsilon B(K_1)$, we have

$$W(K_1+1,K-M(K_1+1)\ ,\ \lambda_1,\lambda_2) = \tilde{W}_2(K-M(K_1+1)\ ,\ \lambda_2)\ , \tag{30}$$

and

$$W(K_1,K-MK_1,\lambda_1,\lambda_2) = \tilde{W}_1(K_1,\lambda_1)\ . \tag{31}$$

Consider any point $(\lambda_1,\lambda_2)\epsilon B(K_1-1)$. Let $\Delta_1 > 0$. From (28) and (24) we have

$$W(K_1,K-MK_1,\lambda_1,\lambda_2 + \Delta_1) = \tilde{W}_2(K_1-MK_1,\lambda_2 + \Delta_1)\ .$$

Incorporating (24) and the definitions of $B(K_1-1)$ we obtain

$$\tilde{W}_2(K_1-MK_1,\lambda_2 + \Delta_1) > \tilde{W}_1(K_1-1,\lambda_1)\ .$$

Since

$$\tilde{W}_2(K_1-MK_1,\lambda_2 + \Delta_1) > \tilde{W}_2(K-M(K_1-1)\ ,\lambda_2 + \Delta_1)\ ,$$

we conclude that

$$W(K_1,K-MK_1,\lambda_1,\lambda_2 + \Delta_1) > W(K_1-1,K-M(K_1-1)\ ,\ \lambda_1,\lambda_2 + \Delta_1)\ .$$

Thus, $K_1^* \neq K_1$ at $(\lambda_1,\lambda_2 + \Delta_1)$ and so in the region bounded by $(0,\lambda_2)$ and $B(K_1-1)$ we have $K_1^* \neq K_1$.

Consider now any point $(\lambda_1,\lambda_2)\epsilon B(K_1-1)$ and let $0 < \Delta_2 \leq \lambda_2$. From the definition of $B(K_1-1)$ and by applying (25), we have

$$\tilde{W}_1(K_1-1,\lambda_1) > \tilde{W}_2(K-MK_1,\lambda_2-\Delta_2)\ .$$

We also have

$$\tilde{W}_1(K_1-i,\lambda_1) > \tilde{W}_1(K_1-1,\lambda_1) > \tilde{W}_1(K_1,\lambda_1) \;,\; K_1 > i > 1 \;.$$

Therefore,

$$W(K_1-i,K-M(K_1-i),\, \lambda_1,\lambda_2-\Delta_2) > W(K_1,K-MK_1,\lambda_1,\lambda_2-\Delta_2) \;.$$

Hence, in the subregion of $C(K_1)$ which is bounded by $B(K_1-1)$ and $(\lambda_1,0)$, we conclude that $K_1{}^* \neq K_1-i$, $i \geq 1$.

Now consider any point $(\lambda_1,\lambda_2)\epsilon B(K_1)$ and let $\Delta_3 > 0$. From the definition of $B(K_1)$ and by applying (26) we obtain

$$\tilde{W}_1(K_1,\lambda_1) < \tilde{W}_2(K-M(K_1+1)\;,\; \lambda_2 + \Delta_3) \;.$$

We also have

$$\tilde{W}_2(K-MK_1,\lambda_2 + \Delta_3) < \tilde{W}_2(K-M(K_1+1)\;,\; \lambda_2 + \Delta_3) \;.$$

Furthermore,

$$\tilde{W}_2(K-M(K_1+1)\;,\; \lambda_2 + \Delta_3) < \tilde{W}_2(K-M(K_1+i)\;,\; \lambda_2 + \Delta_3) \;,\; [K/M]-K_1 \geq i > 1 \;.$$

Thus,

$$W(K_1,K-MK_1,\lambda_1,\lambda_2 + \Delta_3) < W(K_1+i,K-M(K_1+i)\;,\; \lambda_1,\lambda_2 + \Delta_3) \;,$$
$$[K/M]-K_1 \geq i \geq 1 \;.$$

So in the subregion of $C(K_1)$ which is further bounded by $B(K_1)$ and $(0,\lambda_2)$, $K_1{}^* \neq K_1+i$, $1 \leq i \leq [K/M]-K_1$.

Consider any point $(\lambda_1,\lambda_2)\epsilon B(K_1)$ and let $0 < \Delta_4 \leq \lambda_2$. From (31) and (27) we have

$$W(K_1,K-MK_1,\lambda_1,\lambda_2-\Delta_4) = \tilde{W}_1(K_1,\lambda_1) \;.$$

By the definition of $B(K_1)$ and by applying (27),

$$\tilde{W}_1(K_1,\lambda_1) > \tilde{W}_2(K-M(K_1+1)\;,\; \lambda_2-\Delta_4) \;.$$

We also have

$$\tilde{W}_1(K_1,\lambda_1) > \tilde{W}_1(K_1+1)\;,\; \lambda_1) \;.$$

Therefore,

$$W(K_1,K-M(K_1+1)\;,\; \lambda_1,\lambda_2-\Delta_4) > W(K_1+1,K-M(K_1+1)\;,\; \lambda_1,\lambda_2-\Delta_4) \;.$$

So $K_1{}^* \neq K_1$ in the subregion of $C(K_1)$ which is further bounded by $B(K_1)$ and $(\lambda_1,0)$.

Since $\tilde{W}_1(K_1,\lambda_1) < \tilde{W}_1(K_1-1,\lambda_1)$, $B(K_1)$ and $B(K_1-1)$ have no common point except $(0,0)$. Therefore, $A(K_1) = \psi(K_1)$.

Q.E.D.

When the Priority Assignment scheme is applied to the system, we note that the associated approximation for the class 2 message average waiting time have the same properties as described in (24), (25), (26), (27). Thus the same algorithm for optimal slot allocation is applied, with $\tilde{W}_2(K_1,\lambda_2)$ appropriately replaced by $\tilde{W}_{2p}(K_1,\lambda_2)$ where

$$\tilde{W}_{2p}(K_1,\lambda_2) = W_{2p}(K_1,\lambda_2)/\mu_2^{-1} \;. \tag{32}$$

To optimally allocate the slots for a system with given K and M, we proceed as follows. Construct Ω according to its definition. Then draw the boundaries $B(i)$ (defined in Theorem 1), $0 \le i \le [K/M]$, in Ω. For any given pair of traffic rates λ_1 and λ_2, if $(\lambda_1,\lambda_2)\in\Omega$ and is between $B(i-1)$ and $B(i)$, set $K_1^* = i$, $1 \le i \le [K/M]$.

Comparison of the Two Schemes and Numerical Results. -When the Fixed Assignment scheme is applied, each class of messages have their own designated region in the frame. An arriving message has to wait if all the slots in its own region are occupied, even if there are free slots in the other region. To make use of these free slots, and thus increase the utilization of the channel we have introduced the Priority Assignment scheme. By applying the latter scheme, we allow the sharing of slots between the two classes of messages at the expense of requiring a more complex scheme and protocol and possible preemption of the class 2 messages. However, since we consider application of this scheme to cases where $\mu_1^{-1} >> \mu_2^{-1}$, the probability of preemption is small. The performance of the corresponding non-preemptive Priority Assignment scheme will thus be well approximated by that of the present scheme.

The two schemes are now compared through the example shown in Figure 3 and Figure 4. These Figures show the regions and boundaries associated with the optimal slot assignments. The slots are optimally allocated by the algorithm described by Theorem 1. Because of the sharing of the slots between the two classes of messages, the regions $\psi(K_1)$ expand (contract) under the Priority scheme for large (small) values of K_1. For $\lambda_1 = 0.2$, $K_1 = 3$, and $\tilde{W}_1 = \tilde{W}_2$, so that the message waiting-time $W(\cdot)$ is the same under both schemes, we observe from Figures 3-4 that the maximal class 2 arrival rate (λ_2) under the Fixed and Priority schemes, is equal to 2 and 4, respectively.

In Figure 5, the message response time $W = \mathrm{Max}(\tilde{W}_1,\tilde{W}_2)$ is plotted vs. λ_2 with $\lambda_1 = 0.14$. For $\lambda_2 = 2$, $K_1^* = 3$ for both schemes, W is reduced from 0.5 to 0.12 when the Priority scheme is applied. Increasing then λ_2 to 6, we obtain, using the Fixed Assignment scheme, $K_1^* = 2$, so that K_1 need to be decreased from 3 to 2, for otherwise the system will become unstable. Then, we find $W = \mathrm{Max}(\tilde{W}_1,\tilde{W}_2) = 0.95$. Applying the priority scheme for $\lambda_2 = 6$, we still obtain $K_1^* = 3$ and $W = \mathrm{Max}(\tilde{W}_1,\tilde{W}_2) = 0.4$. Thus the Priority scheme has been observed to yield lower message waiting-time values while simultaneously accomodating wider fluctuations in the message traffic characteristics.

Conclusions. We consider the sharing of a communication channel by a number of sources for real-time transmission on a TDMA basis. The messages from the sources are assumed to require two different information rates. They are classified according to their rates as class 1 and class 2 messages. Requests for transmission are stored by a central controller. Two dynamic demand-assignment access control disciplines are presented. The first discipline, a Fixed Assignment scheme, divides the slots of the frame into two fixed regions, one region for each class of messages. A separate assignment of slots within each region is employed by the controller. The second discipline, a Priority Assignment scheme, incorporates into the first scheme the possibility of utilization of the free class 1 slots by class 2 messages. We calculate, for each scheme, the message queueing delays, and optimally allocate the slots in terms of a maximum normalized average delay measure. The optimal structure of the frames for both schemes are compared. The advantage of the priority scheme in terms of response-time and adaptability measures is demonstrated.

References

1. R. J. Huff, "TDMA Space Communications Systems: Concepts and Practical Techniques", Rome Air Development Central Technical Report RADC-TR-71-255, November, 1971.
2. G. Frenkel, G. V. Kinal, D. G. Gan, and J. E. Doggett, "TDMA Network Control Study", Rome Air Development Center Technical Report RADC-TR-73-270, August, 1973.
3. J. E. Doggett, G. V. Kinal, S. H. Talbot, "Control of Channel Assignment in Tactical TDMA Satellite Communications". Proc. National Telecommunication Conference, Volume 1, 1973.
4. N. V. Prabhu, Queues and Inventories, J. Wiley, New York, New York, 1965.
5. T. L. Saaty, Elements of Queueing Theory, McGraw-Hill, New York, New York, 1961.

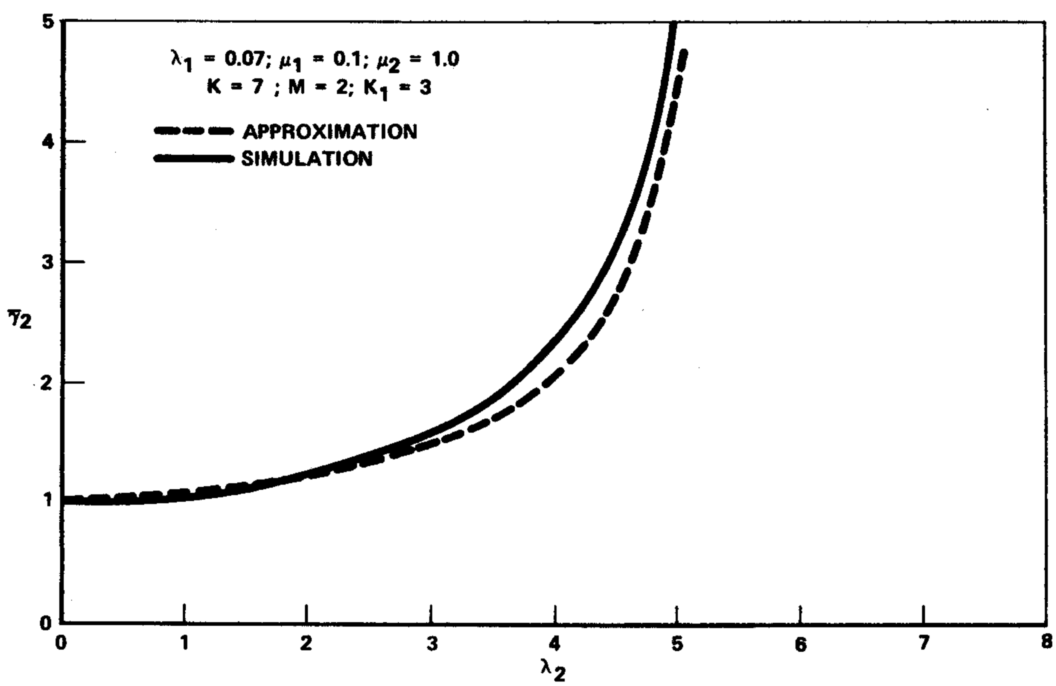

Figure 1. Average Delay of a Class 2 Message vs λ_2.

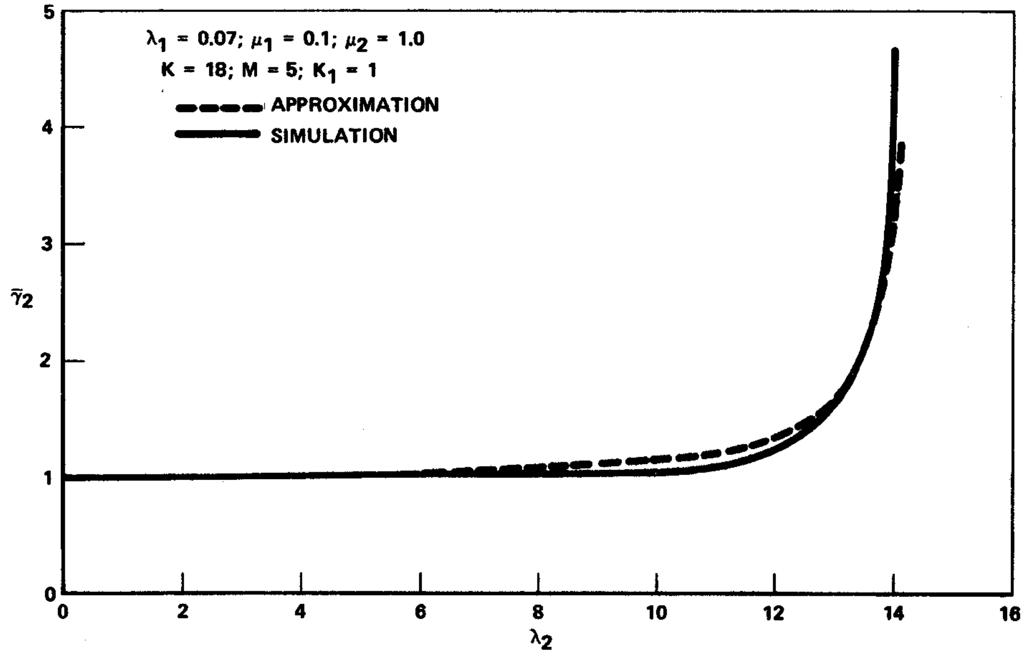

Figure 2. Average Delay of a Class 2 Message vs λ_2.

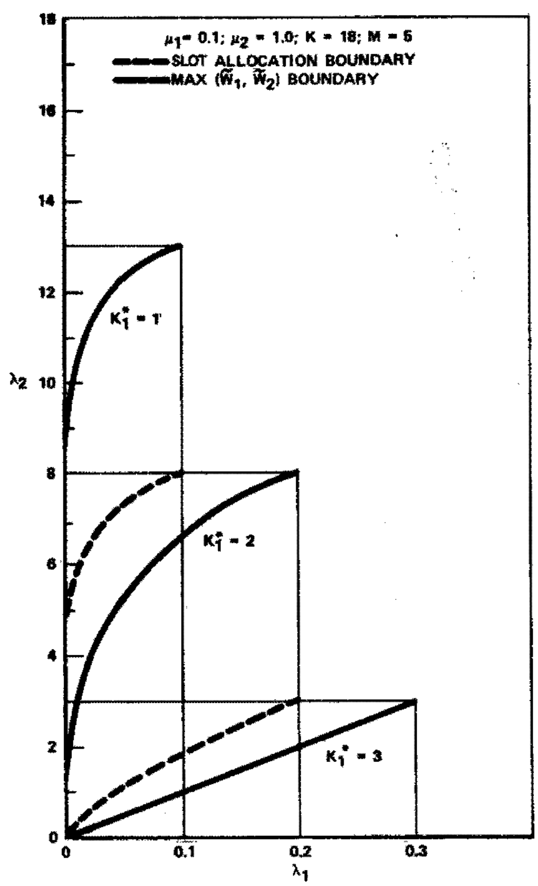

Figure 3. Slot Allocation Under Fixed Assignment.

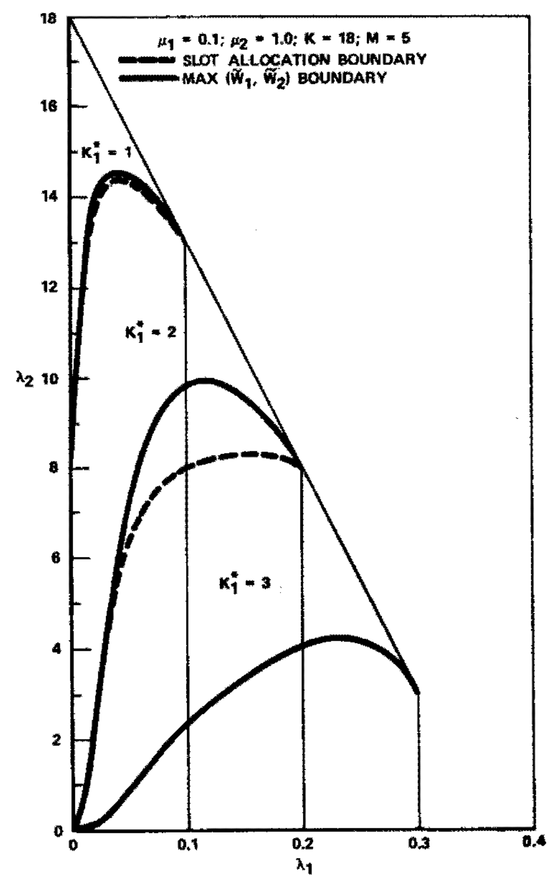

Figure 4. Slot Allocation under Priority Assignment.

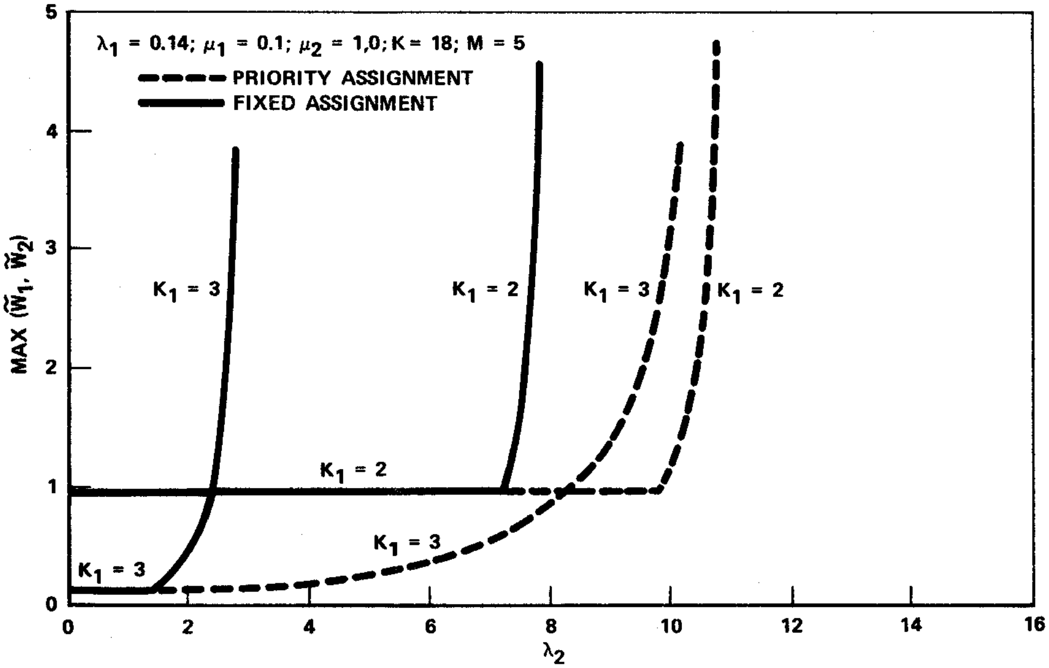

Figure 5. $\text{Max}(\tilde{W}_1, \tilde{W}_2)$ versus λ_2 Under Fixed Assignment and Priority-Assignment.