ANALYSIS OF INTEGRATED SERVICES TDM WITH CORRELATED TRAFFIC

IOANNIS STAVRAKAKIS and SOPHIA TSAKIRIDOU

University of Vermont
CS/EE Dept., Burlington, Vermont 05405, U.S.A.

ABSTRACT The problem of multiplexing Time Critical (*TC*) and Time Non Critical (*TNC*) packetized information in an Integrated Services Digital Network (ISDN) is considered in this paper. An appropriate Time Division Multiplexing (TDM) scheme is adopted, to accommodate the strict delay requirements for the *TC* traffic. The resulting system is different from those considered in the past in at least two aspects. First, the *TNC* traffic is assumed to exhibit correlation between consecutive slots. Second, the *TC* traffic is not necessarily accommodated in a contiguous subframe, but it may be spread over the whole frame according to any pattern. The non-gated service policy is adopted for the service of the *TNC* traffic, unlike most of the past work. A very flexible approach is developed for the exact analysis of the proposed TDM scheme. Numerical results are derived for the case of *TC* traffic generated by voice sources.

I. INTRODUCTION

The fundamental problem associated with the design of Integrated Services Digital Networks (ISDN's) is that of the development of protocols which guarantee certain predetermined quality of service for the diversified network users. The most severe requirements come from network users which generate Time Critical (*TC*) information. *TC* information must be delivered to its destination within a strict time limit; otherwise, it is useless and it is considered to be lost. This type of information is present in any real time application, such as voice and video transmission. Computer data, on the other hand, is an example of Time Non Critical (*TNC*) information, which is also present in an ISDN. No strict delivery time limitations are assumed for this type of information.

The information is assumed to be organized in packets of equal length. The time required for a packet transmission (slot) determines the time constant of the network. The *TC* sources are usually slow compared to the network time constant. That is, the minimum time separation between consecutive packets of information generated by a *TC* source, which determines the *TC* source time constant, T, is much larger than the slot. By dividing time into frames of length T (in slots) and assigning one slot per frame to each *TC* source, up to T *TC* sources can be served, suffering a delay of at most 2T time units (slots). This is the classical Time Division Multiplexing (TDM) protocol.

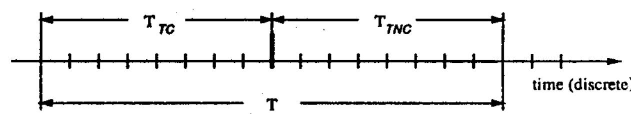

Figure 1
The *TC* and the *TNC* subframes of the TDM frame.

A TDM switching node of an ISDN is considered in this paper, supporting both *TC* and *TNC* sources of packetized information. The frame is divided into two parts: The *TC* subframe, of length T_{TC}, and the *TNC* subframe, of length T_{TNC} (Fig. 1), which are used for the transmission of *TC* and *TNC* information, respectively. The two subframes are separated by a conceptual boundary which may be either fixed or variable. The TDM switching node described above can also model a distributed system, where users share the network capacity on a reservation TDMA basis, [13], or when some distributed channel allocation function is used, [5], [8].

In this paper, the movable boundary TDM protocol is adopted for the capacity allocation in an ISDN. This protocol has been adopted for the voice / data integration and it has been analyzed in the past both in continuous [2], [3], [6], and discrete [1], [4], [5], [7], [9]-[11], time. The continuous time models are inherently approximate since they ignore the discrete slots of the TDM frames and assume exponentially

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

* Email: stavraka@uvm.edu , tsakirid@uvm.edu

distributed message lengths, when birth-death models are constructed, [9]. Fluid flow approximation approaches can be found in [3], [16]. Analysis of the protocol in discrete time has been based on the generating function approach [1], [4], [7], [9], [10], [11]. In [1] it is assumed that $\{R_i\}_{i\geq 0}$, that is, the number of active voice sources in the i^{th} frame, is a renewal process. In [4] the correlations in the voice traffic are taken into consideration and an exact solution is derived under a gated service policy for the data; that is, data packets cannot be transmitted before the beginning of the frame that follows their generation time instant. In [7], [11], the true model for $\{R_i\}_{i\geq 0}$ is considered under a gated service policy for the data. Although the models considered in [7], [11] are exact, only approximate, [7], or exact numerical results over a very limited range [11], are derived. The non-gated policy for data is considered in [9] and [10]. Because of the complexity of the solution (associated with the calculation of the characteristic roots within the unit disk of generating functions) bounds on the mean packet delay are actually computed.

The discrete-time TDM system considered in this paper is different from those considered in the past in at least two aspects. First, the *TNC* traffic is assumed to exhibit correlation between consecutive slots. Second, the *TC* traffic is not necessarily accommodated in a contiguous subframe, but it may be spread over the whole frame according to any pattern. The non-gated service policy for the TNC sources is assumed, unlike all previous work except [9] and [10].

II. SYSTEM MODEL

The TDM system considered here is similar to the one that has been adopted for the integration of packetized voice and data in the past work mentioned in the previous section. Information is organized in packets of fixed size. The (time) slot is defined to be equal to the packet transmission time. All *TC* sources are assumed to be identical. Their time constant is assumed to be equal to T (slots). As a result, each *TC* source is assumed to deliver one packet every T slots, when active. A *TC* source switches between the active (1) and the idle (0) states according to the first order Markov model; no packet is delivered when the *TC* source is idle. The *TC* source is completely described in terms of one of the steady state probabilities, π_0 or π_1, and the burstiness coefficient γ, where $\gamma_{TC}=p_{11}-p_{01}$; p_{ij} denotes the transition probability from state i to state j, $i,j \in \{0,1\}$. The *TC* packets must be transmitted within the time frame that follows their generation time instant, otherwise they are useless and they are dropped.

Unlike all related previous work, the *TNC* traffic is assumed to be correlated. This traffic is modeled as a Markov Modulated Generalized Bernoulli (MMGB) process. According to this process, *TNC* packet arrivals are governed by an underlying finite-state Markov chain $\{Z_j'\}_{j\geq 0}$. Transitions between the states of the chain occur at the slot boundaries. The steady state and the transition probabilities are denoted by $\pi'(i)$ and $p'(i,j)$, $i,j \in S'$; $S'=\{0,1,\cdots,N\}$ denotes the state space of the Markov chain. The *TNC* packet arrival process is determined in terms of the Markov chain and a probabilistic mapping $a'(\cdot)$: $S' \rightarrow \{0,1,\ldots,R'\}$. That is, k *TNC* packets are delivered to the *TNC* queue with probability $g'(i,k)=Pr\{a'(i)=k\}$, $0\leq k\leq R'$, when the Markov chain is in state i, $i \in S'$.

The characteristics and the time delivery requirements of the *TC* traffic support the adoption of a TDM scheme (frame size T) for the multiplexing of the *TC* and *TNC* traffic. The TDM system is modeled as a discrete-time queueing system with interruptions, as shown in Fig. 2. The buffer capacity of the *TNC* queue is assumed to be infinite; the buffer capacity of the *TC* queue is equal to $2T_{TC}$. To reflect the TDM policy, the server is assumed to switch to the TC queue at the beginning of a frame and serve all the *TC* packets found upon switching (at most T_{TC}); then, it switches to the *TNC* queue and serves the *TNC* traffic according to the FIFO service discipline. At the end of each frame, the *TC* packets which have been queued for at least T slots are dropped.

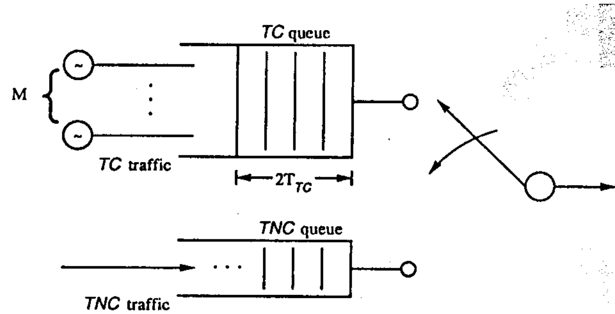

Figure 2
The discrete-time queueing model for the adopted TDM policy.

Let M denote the number of *TC* sources supported by the system and let T_{TC}, $T_{TC}\leq T$, denote the number of slots, in a frame, which are allocated to the *TC* packets. Let R_i denote the number of packets in the *TC* queue at the beginning of the i^{th} frame; R_i is equal to the number of *TC* sources which were active in the previous frame. If $R_i\leq T_{TC}$, then all *TC* packets will be transmitted in the i^{th} frame. To simplify the notation and the discussion in this paper, the movable boundary policy is adopted and the relation $M=T_{TC}\leq T$ is assumed to hold. The cases under the fixed boundary policy, as well as under various relations between M, T_{TC} and T, are trivially derived by modifying certain parameters of the unified model developed here, as discussed in section IV. For the same reasons, it is assumed that the *TC* packets are transmitted over the first R_i slots of the i^{th} frame. The case of the spreading of the *TC* packets over the T slots of the frame is easily derived, as discussed in section IV. Notice that the *TNC* packets do not necessarily have to wait for the frame following their generation instant before they consider transmission, as it is the case with the gated discipline. If a *TNC* packet reaches the head of the *TNC* queue and the server is available, then this packet is transmitted (non-gated discipline). The server is available when the corresponding slot is not used by the *TC* users.

The performance evaluation of the adopted TDM scheme for the integration of *TC* and *TNC* packet traffics is based on two measures. Assuming that the upper bounded by 2T delay of the *TC* packets is acceptable, the induced delay for these packets is not considered to be an issue. The most important performance measure for the *TC* traffic is the packet blocking (dropping) probability. This probability is zero when $M\leq T_{TC}\leq T$. When non-zero, it can be easily computed as discussed in section IV. The analytically challenging problem associated with the TDM system presented in this paper, is related to the calculation of the *TNC* packet delay induced by the transmission policy. This is considered in the next section.

III. ANALYSIS OF THE *TNC* QUEUE

In this section the *TNC* queue is studied under *TNC* packet arrivals described by a MMGB process (Section II). The behavior of this queue is affected by the activity in the TC queue. Fortunately, the TDM policy is such that the coupling between the two queues is loose. It is basically a one direction interference on the *TNC* queue coming from the *TC* queue; the behavior of the *TC* queue is, on the other hand, independent from the activity in the *TNC* queue. This observation allows for the modeling of the interference from the *TC* queue on the *TNC* queue as an independent - from anything associated with the *TNC* queue - process $\{\bar{R}_j\}_{j\geq 0}$, as it is explained in the next paragraph. The *TNC* queue is studied then, under the *TNC* packet traffic and the interfering process $\{\bar{R}_j\}_{j\geq 0}$, which captures the TDM policy. This approach is new and different from previous approaches developed for the study of integrated TDM systems. All policies and approximations which were adopted in the past can be analyzed in a unified way, by selecting the appropriate interfering process $\{\bar{R}_j\}_{j\geq 0}$. Furthermore, new policies, as well as correlations in the *TNC* traffic, can be considered for the first time. Unless stated otherwise, the *TC* packets will be considered to be transmitted over contiguous slots, starting from the beginning of the frame. No *TC* packet blocking will be assumed ($M=T_{TC}\leq T$).

The behavior of the *TNC* queue of the TDM system (Fig. 2) is identical to that of the *TNC* queue shown in Fig. 3. $\{\bar{R}_j\}_{j\geq 0}$ is a process which delivers at most one packet per slot according to the rule

$$\bar{R}_j = \begin{cases} 1 & \text{if } 0 \leq j\,mod\,T \leq R_i - 1 \quad, i = \lfloor j/T \rfloor \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

$\lfloor x \rfloor$ denotes the integer part of x. According to (1), a packet is delivered by $\{\bar{R}_j\}_{j\geq 0}$ at the j^{th} slot, if this slot is in the first R_i positions of the i^{th} frame; R_i is the number of active *TC* sources in that frame. A realization of $\{\bar{R}_j\}_{j\geq 0}$ is shown in Fig. 4. The Head of the Line (HoL) priority policy is assumed for the system in Fig. 3. The server switches to the *TNC* queue only if the buffer of the $\overline{TC}$ queue is empty; it switches back to the $\overline{TC}$ queue, as soon as its buffer becomes non-empty. Packet arrivals and service completions are declared at the end of the slots (slot boundaries). Let D_{TNC} and $D_{\overline{TC}}$ denote the mean packet delay of the *TNC* and the $\overline{TC}$ queues, respectively. Notice that $D_{\overline{TC}}=1$ due to the adopted service policy and the fact that $\{\bar{R}_j\}_{j\geq 0}$ may deliver at most one packet per slot.

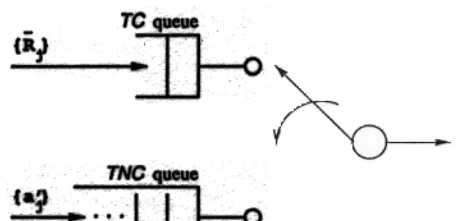

Figure 3
The equivalent queueing system associated with the *TNC* queue.

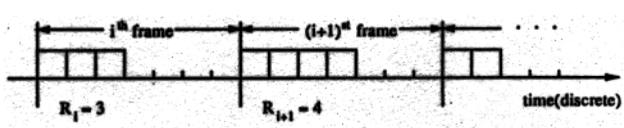

Figure 4
A realization of $\{\bar{R}_j\}_{j\geq 0}$ for $iT\leq j\leq (i+2)T+3$ and T=7.

Consider now a single queue, single server, First-In First-Out (FIFO) queueing system which is fed by the arrival processes of the system in Fig. 3. Let D_{FIFO} denote the mean packet delay induced by this system. D_{FIFO} is related to D_{TNC} and $D_{\overline{TC}}$ through the following expression (conservation law):

$$D_{FIFO} = \frac{\lambda_{TNC}D_{TNC}+\lambda_{\overline{TC}}D_{\overline{TC}}}{\lambda_{TNC}+\lambda_{\overline{TC}}} \tag{2}$$

where λ_{TNC} and $\lambda_{\overline{TC}}$ are the rates of the *TNC* traffic and the process $\{\overline{R}_j\}_{j\geq 0}$, respectively. Notice that the service policy of the queueing system is work-conserving and non-preemptive. It is clear that D_{TNC} can be computed from (2) if D_{FIFO} is known. The rest of the section is devoted to the computation of D_{FIFO}. At first, the interfering process $\{\overline{R}_j\}_{j\geq 0}$ is described in terms of an appropriate MMGB model.

Consider the process $\{R_i\}_{i\geq 0}$ defined at the beginning of the frames; R_i is equal to the number of active *TC* sources, that is, the *TC* sources which will be served over the i^{th} frame. From the Markovian structure of the *TC* sources it is easily shown that $\{R_i\}_{i\geq 0}$ is an (M+1)–state Markov chain with state space $S_R=\{0,1,\ldots,M\}$ and transition probabilities $p_R(i,j)$ given by

$$p_R(i,j) = Pr\{R_{n+1}=j / R_n=i\} = \sum_{k=\max\{0,i+j-M\}}^{\min\{i,j\}} \binom{i}{k} p_{11}^{k} p_{10}^{i-k} \binom{M-i}{j-k} p_{01}^{j-k} p_{00}^{M-i-(j-k)}, \quad i,j \in S_R, \tag{3}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Note that the one-step transition time of the Markov chain modeling the *TC* source is equal to T slots. The interfering process $\{\overline{R}_j\}_{j\geq 0}$ is defined in terms of the process $\{R_i\}_{i\geq 0}$ and their relation, as expressed in (1).

Consider the two-dimensional process $\{R_{\hat{i}},L_j\}_{j\geq 0}$ defined on the slot boundaries; $\hat{i} = \lfloor j/T \rfloor$ and $L_j=j\,mod\,T+1$. $R_{\hat{i}}$ is a Markov chain with transition probabilities given by (3) and state space S_R. L_j is also a Markov chain with state space $S_L = \{1,2,\ldots,T\}$ and transition probabilities given by

$$p_L(k,j) = \begin{cases} 1 & \text{if } j=k\,mod\,T+1\ ,\ 1\leq k\ ,\ j\leq T \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$\{L_j\}_{j\geq 0}$ is a periodic Markov chain whose state describes the position of the current slot within the current frame; this information is necessary for the determination of $\overline{R}_j$, as expressed in (1). As a result, $\{R_{\hat{i}},L_j\}_{j\geq 0}$ is a two-dimensional Markov chain with state space $S_R\times S_L=\{0,1,\ldots,M\}\times\{1,2,\ldots,T\}$. The state space is shown on Fig. 5. The transition probabilities are given by the following expression.

$$\overline{p}((i,j),(k,n)) = p_R(i,k)p_L(j,n) = \begin{cases} 1 & \text{if } k=i \text{ and } n=j+1 \text{ for } 0\leq i,k\leq M,\ i\leq j<T,\ 1<n\leq T \\ p_R(i,k) & \text{if } j=T,\ n=1 \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The interfering process $\{\overline{R}_j\}_{j\geq 0}$ determined by (1) can be easily described in terms of the Markov chain $\{R_{\hat{i}},L_j\}_{j\geq 0}$ and the following mapping

$$\overline{R}_j=\overline{a}\{(R_{\hat{i}},L_j)\}= \begin{cases} 1 & \text{with probability 1 , if } L_j\leq R_{\hat{i}} \\ 0 & \text{with probability 1, otherwise} \end{cases} \tag{6}$$

defined on $S_R\times S_L$. If $R_{\hat{i}}$ *TC* users are active in the $\hat{i}^{th}$ frame, then the above mapping will generate one packet over the first $R_{\hat{i}}$ slots of this frame. This is illustrated with the arrows originating from the appropriate states in Fig. 5. The packet generating states are contained in the triangle defined by the states (1,1), $(T_{TC},1)$ and (T_{TC},T_{TC}). Notice that $\{\overline{R}_j\}_{j\geq 0}$ has been described as a MMGB process (Section II) with underlying Markov chain $\{R_{\hat{i}},L_j\}_{j\geq 0}$ and probabilistic mapping given by (6). This mapping is alternatively described by the following probabilities

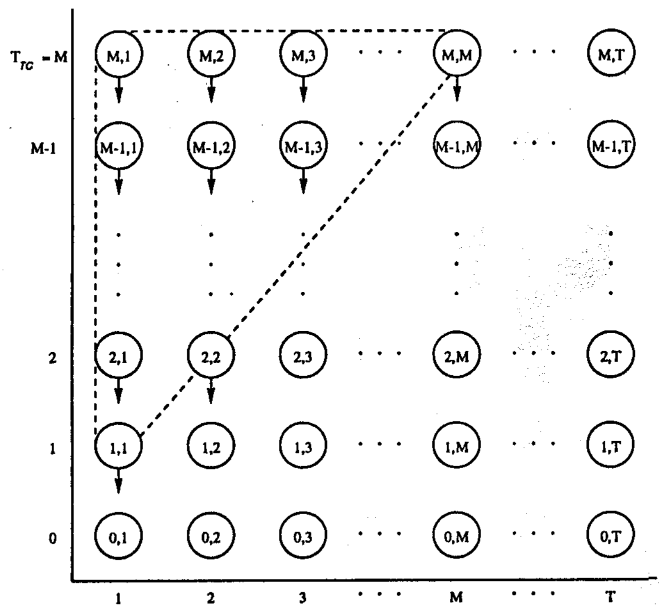

Figure 5

The state space $S_R\times S_L$;
the arrows denote a packet generation ($M=T_{TC}\leq T$).

$$\bar{g}((k,j),1) = Pr\{\bar{a}(k,j)=1\} = \begin{cases} 1 & \text{if } j \leq k \\ 0 & \text{otherwise} \end{cases} \quad (7a)$$

$$\bar{g}((k,j),0) = Pr\{\bar{a}(k,j)=0\} = 1-\bar{g}((k,j),1) \quad (7b)$$

for $(k,j) \in S_R \times S_L$.

At this point, the equivalent FIFO queueing system is considered. Notice that both packet arrival processes, $\{\bar{R}_j\}_{j\geq 0}$ and $\{a_j'\}_{j\geq 0}$, are MMGB processes. This queueing system has been analyzed in [14], [15]; the results may also be found in [17], where the first moment, Q_{FIFO}, and the variance, $var\{q_{FIFO}\}$, of the buffer occupancy process are calculated, as well as the mean packet delay, D_{FIFO}, by invoking Little's formula. Finally, D_{TNC} can be calculated by invoking (2); $\lambda_{\overline{TC}} = \pi_1 M/T$ $(M=T_{TC}\leq T)$.

The mean and the variance of the buffer occupancy in the *TNC* queue can also be calculated. The proof of the following expressions may be found in [17].

$$Q_{TNC} = E\{q_{TNC}\} = Q_{FIFO} - \lambda_{\overline{TC}} \quad (8)$$

$$var\{q_{TNC}\} = var\{q_{FIFO}\} + \lambda_{TNC}^2 - \lambda_{TNC} \quad (9)$$

IV. APPLICABILITY OF THE ANALYSIS APPROACH - SPECIAL CASES

In the previous section a general methodology was developed for the analysis of the behavior of the *TNC* queue. As it will become clear in this section, the generality of the approach is due to the flexibility provided through the adoption of the interfering process $\{\bar{R}_j\}_{j\geq 0}$ and its general description in terms of a MMGB model. The integrated services TDM that was considered in the previous sections has the following characteristics :

(a) It is non-blocking for the *TC* traffic $(M=T_{TC}\leq T)$.

(b) It is a movable boundary scheme.

(c) The *TC* sources are correlated, resulting in a Markovian process $\{R_i\}_{i\geq 0}$.

(d) The *TNC* traffic exhibits correlation.

(e) The *TC* sources generate one packet when active.

(f) The *TC* packets are transmitted over the first R_i slots of the i^{th} frame.

(g) The *TNC* packets can be served in the slot following their arrival instant (non-gated service).

In the sequel it is shown that characteristics (a)-(f) can be changed and the analysis approach be still applicable. The gated service policy for the *TNC* traffic cannot be analyzed through the developed approach. The gated policy would require that the *TNC* packets wait until the beginning of the next frame before they can be served. This policy results in the waste of the channel capacity. It has been considered in all related past work (except [9], [10]) for analytical convenience.

Modifying (a)

When the number of *TC* sources exceeds the maximum slot allocation, T_{TC} $(M>T_{TC})$ blocking of the *TC* packets may occur. The Markov chain $\{R_i\}_{i\geq 0}$, which describes the number of active *TC* sources at the frame boundaries, is as described in Section III (equation (3)). The interfering process $\{\bar{R}_j\}_{j\geq 0} \equiv \{\bar{a}(R_{\hat{i}},L_j)\}_{j\geq 0}$ is determined in terms of the underlying Markov chain $\{R_{\hat{i}},L_j\}_{j\geq 0}$ and the mapping in (6), which is now modified to

$$\bar{a}\{(R_{\hat{i}},L_j)\} = \begin{cases} 1 & \text{w.p. } 1 \quad \text{if } L_j \leq \min\{T_{TC},R_{\hat{i}}\} \\ 0 & \text{w.p. } 1 \quad \text{otherwise.} \end{cases} \quad (10)$$

The packet generating states are shown in Fig. 5 (arrows), where $T_{TC}<M$. The blocking probability can be computed from

$$p_B = \sum_{k=T_{TC}+1}^{M} \pi_R(k)(k-T_{TC})$$

where $\pi_R(k)$, $0\leq k\leq M$, is the steady state probability of $\{R_i\}_{i\geq 0}$.

Modifying (b)

When the fixed boundary policy is considered, the interfering process $\{\bar{R}_j\}_{j\geq 0}$ does not depend on the activity of the *TC* sources. The first T_{TC} slots are never available to the *TNC* packets. The process $\{R_i\}_{i\geq 0}$ becomes an i.i.d. process (single state Markov chain) which takes the value T_{TC} with probability one. The underlying Markov chain $\{R_{\hat{i}},L_j\}_{j\geq 0}$ reduces to the Markov chain $\{L_j\}_{j\geq 0}$. The interfering process $\{\bar{R}_j\}_{j\geq 0} \equiv \{\bar{a}\{L_j\}\}_{j\geq 0}$ is defined by the mapping

$$\bar{a}\{L_j\} = \begin{cases} 1 & \text{w.p. } 1 \quad \text{if } L_j \leq T_{TC} = M \\ 0 & \text{w.p. } 1 \quad \text{otherwise.} \end{cases} \quad (11)$$

The state space of $S_R \times S_L$ becomes a single row in Fig. 5 which has the arrow structure of row $\{(T_{TC},1), \ldots, (T_{TC},T)\}$.

It should be noted that the model for the fixed boundary policy is identical with that of the standard TDMA policy, where T_{TC} represents the slots that are not available to the tagged TDMA station.

Modifying (c)

When the *TC* sources are not correlated $\{R_i\}_{i\geq 0}$ becomes an i.i.d. process (single state Markov chain). In this case,

$$p_R(k,j) = p_R(j) = \binom{M}{j} \pi_1^j \, \pi_0^{M-j} \quad (12)$$

The interfering process, $\{\bar{R}_j\}_{j\geq 0}$, is, otherwise, identical to that described in Section III.

When the *TC* sources are always active, $R_i = M = T_{TC}$ for all $i \geq 0$. In this case, the behavior of the *TNC* queue is the same with that under the fixed boundary policy, which was discussed above.

Modifying (d)

Clearly, the case of the TNC traffic without correlation can be obtained as a special case of the MMGB model (adopted for the TNC traffic $\{a_j'\}_{j\geq 0}$), based on a single state underlying Markov chain.

Modifying (e)

According to the model for the *TC* sources, which was considered in Section III, a *TC* source generates a packet with probability one when active. When this probability is set to p^*, the interfering process $\{\bar{R}_j\}_{j\geq 0}$ is defined by the mapping

$$\bar{a}\{(R_{\dagger},L_j)\} = \begin{cases} 1 & \text{w.p. } p \quad \text{if } L_j \leq R_{\dagger} \\ 0 & \text{w.p. } 1-p^* \quad \text{if } L_j \leq R_{\dagger} \\ 0 & \text{w.p. } 1 \quad \text{otherwise.} \end{cases} \qquad (13)$$

Referring to Fig. 5, the arrows (reflecting packet generation) are contained within the same triangle but they exist with probability p^*. The case in which multiple packets may be generated by a *TC* source according to a given distribution can, in principle, be treated in a similar way.

Modifying (f)

In Section III it has been assumed that all *TC* packets are served over contiguous slots, starting from the beginning of a frame. To the best of our knowledge, this has been the policy in all relevant past work. This policy, on the other hand, is not necessary to meet the (real) time delivery requirement of the *TC* sources. What is important is that a *TC* source be allowed to transmit one packet in every frame, when active. The exact position of the transmission slot within the frame is not important. This is not the case regarding the *TNC* packets of the system.

The transmission of the *TC* packets over contiguous slots results in an interfering process $\{\bar{R}_j\}_{j\geq 0}$ which delivers blocks of high priority packets to the equivalent FIFO queue followed by idle periods over the rest of the frame. Clearly, this process generates more severe queueing problems than a process which spreads uniformly its packets over the whole frame. The more severe queueing problems will result in higher *TNC* packet delay, since the packets delivered by $\{\bar{R}_j\}_{j\geq 0}$ have a constant delay under the HoL priority policy (equal to one). Intuitively, it is expected that the *TNC* queue will be less occupied if the server is absent for many short intervals than for a large one (for the same total time), especially under light traffic and under the non-gated policy considered here.

An algorithm for the implementation of a uniform (or near-uniform) spreading of the *TC* packets within the frame has been developed and it may be found in [17]. Given a state of the underlying Markov chain $\{R_{\dagger},L_j\}_{j\geq 0}$, associated with the interfering process $\{\bar{R}_j\}_{j\geq 0}$, a packet will be delivered with probability one provided that this state is a legitimate one. The spreading algorithm operates on the rows of the state space of $\{R_{\dagger},L_j\}_{j\geq 0}$ (Fig. 5). For each row $\{(k,j)\}_{j=1}^{T}$, $1 \leq k \leq T_{TC} = M$, the algorithm distributes the k *TC* packets to the T slots in a uniform (or approximately uniform) pattern, determining the position of the k arrows of that row. Clearly, the arrows are not contained in the triangle indicated in Fig. 5, but they are spread throughout the entire state space in a deterministic manner.

The underlying Markov chain $\{R_{\dagger},L_j\}_{j\geq 0}$ remains unchanged. The interfering process $\{\bar{R}_j\}_{j\geq 0}$ is now determined by the mapping

$$\bar{a}\{(R_{\dagger},L_j)\} = \begin{cases} 1 & \text{w.p. } 1 \quad \text{if } L_j \in A(R_{\dagger}) \\ 0 & \text{w.p. } 1 \quad \text{otherwise,} \end{cases} \qquad (14)$$

where $A(R_{\dagger})$ denotes the set of the slot positions which are occupied by the $R_{\dagger}$ *TC* packets, as determined by the spreading algorithm.

V. NUMERICAL RESULTS

In this section, some numerical results are presented and the complexity of the methodology is discussed. The parameters of the *TC* sources are selected to be those of the packetized voice sources. That is, the steady state probabilitity that a *TC* source is active is equal to $\pi_1 = .35$ and the burstiness coefficient is equal to $\gamma = .93$. The *TNC* traffic is described in terms of a MMGB model based on an underlying Markov chain with state space $S' = \{0,1\}$. One *TNC* packet is generated from state 1 with probability one. Thus, the *TNC* packet arrival rate, λ_{TNC}, is equal to $\pi'(1)$. This process can model the traffic delivered by a high speed transmission line, where certain amount of correlation between consecutive slots is present. Let γ_{TNC} denote the burstiness coefficient for the *TNC* traffic.

The numerical results presented in this paper are derived under the assumption $M/T = 1/2$, where $T_{TC} = M$. That is, *TC* source blocking is not possible and at most half of the channel capacity can be allocated to the *TC* traffic. The movable boundary policy is considered. The results are plotted as a function of the *TC* source time constant T, which is equal to the frame size.

In Fig. 6, the mean *TNC* packet delay is plotted for $\lambda_{TNC} = .725$ and $\lambda_{\overline{TC}} = .35 \times .5$ in packets per slot. $\lambda_{\overline{TC}}$ is the cumulative *TC* packet rate, which is equal to the rate of the interfering process since no blocking is possible. The cumulative *TC* packet traffic is generated by the T/2 *TC* sources, each of which has a packet

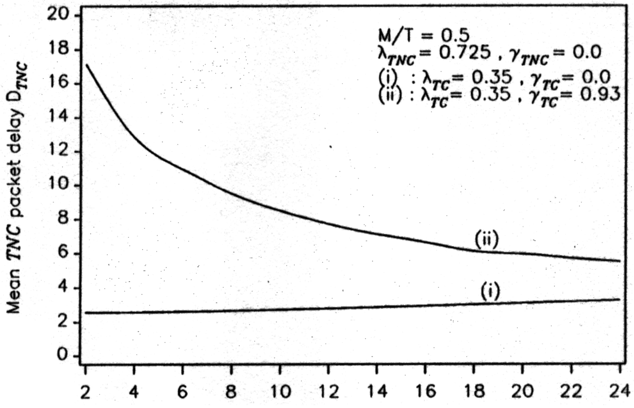

Figure 6

rate of .35 packets per T slots. The *TNC* traffic is assumed to be an i.i.d. process process (γ_{TNC}=0). Curve (i) corresponds to γ_{TC}=0; that is, the *TC* sources are assumed to be uncorrelated. In this case, $\{R_i\}_{i\geq 0}$ is an i.i.d. process. Curve (ii) corresponds to γ_{TC}=.93; that is, the *TC* sources behave like packetized voice sources. The results show that if the Markovian $\{R_i\}_{i\geq 0}$ (for γ_{TC}=.93) is approximated by the corresponding i.i.d. process (for γ_{TC}=.0), then the resulting error in the calculation of D_{TNC} is significant, particularly for small T. As T increases (and the number of *TC* sources increases, for a fixed ratio M/T = 1/2) the performance of the i.i.d. approximation is improved, as expected. The latter is due to the fact that the amount of dependence between consecutive random variables R_i's, as seen by the server, is reduced as T increases.

In principle, the delay results for the models assumed in Fig. 6, can be derived following the analysis in [9] or [10]. Although these analyses are exact (as is the one developed in this paper), the numerical complexity associated with the calculation of the poles of complex functions is enormous. As a result, bounds on the exact results are derived and they are actually computed for small range of (M,T). Exact results under the gated policy can be found in [11]. The numerical complexity of our approach - under the non-gated policy - seems to be very similar to that in [11]. The complexity increases in the order of T^2M^2 and is reflected in the increased required computer memory and computation time for the calculation of the boundary probabilities that the queue is empty and the input Markov chain in a certain state, [15], [17]. In view of the very limited range of system parameters (T=2, M=3) considered in [11], it seems that the structure of the matrices involved in the computation of the boundary elements of the present paper may result in faster convergence than that in [11]. It has also been observed that the computations are much faster when $\{R_i\}_{i\geq 0}$ is an i.i.d. process, compared to those under the Markovian $\{R_i\}_{i\geq 0}$.

The results shown in Fig. 7 are obtained under Markovian *TNC* traffic for γ_{TNC}=.5; the rest of the parameters are as in Fig. 6. No past work has considered a non-i.i.d. model for the *TNC* traffic.

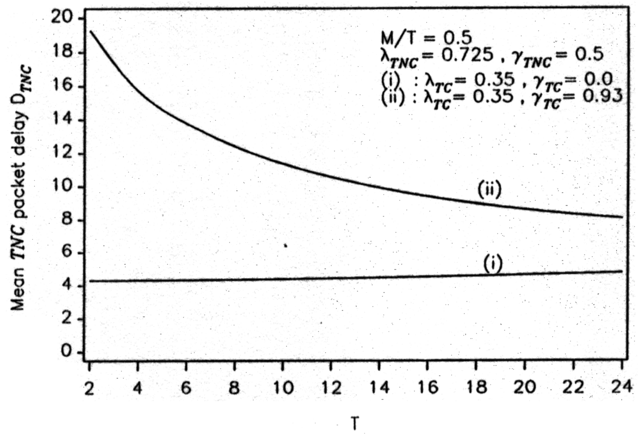

Figure 7

Fig. 8 shows the results presented in Fig. 6 and Fig. 7 for comparison purposes. Notice that the larger the burstiness coefficient for the *TNC* traffic, the larger the induced D_{TNC}, under both models for $\{R_i\}_{i\geq 0}$, as expected. To verify the expected limiting behavior of the queue, as the frame size increases, simulation results are derived beyond a certain range of T. These results are plotted in dotted lines. As it has been discussed earlier, the effect of the correlation in the *TC* traffic on the behavior of the *TNC* queue diminishes, as the frame size increases. The latter trend is clearly observed in Fig. 8, where the results under the true

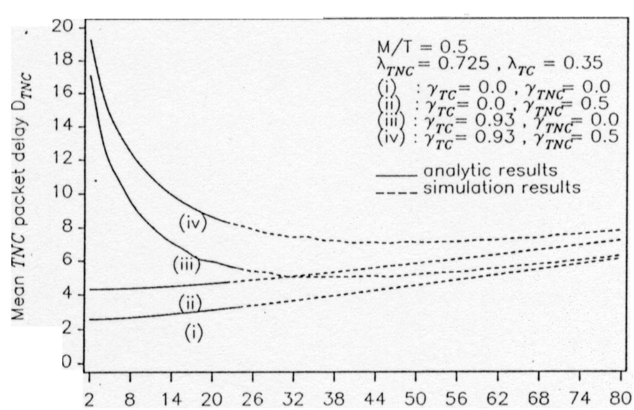

Figure 8

(markovian) model for $\{R_i\}_{i\geq 0}$ approach those under the i.i.d. model for $\{R_i\}_{i\geq 0}$, as T increases.

The case of the uniform spreading of the *TC* packets over the frame has not been considered in the past. Fig. 9 and Fig. 10 present the delay results under the (near-unform) spreading policy described in [17]; these results correspond to those in Fig. 6 and Fig. 7 (contiguous *TC* packet transmissions), respectively. For comparison purposes, the results from Fig. 6 and 9 are shown in Fig 11 (independent *TNC* traffic); similarly, the results from Fig. 7 and 10 are shown in Fig 12 (correlated *TNC* traffic). Notice that the (near-uniform) spreading of the *TC* packets over the frame has a positive effect on the induced delay for the *TNC* traffic. This effect increases as the frame size increases. The latter is expected since the larger the frame size, the more different the interference process under spreading is, compared to that without spreading.

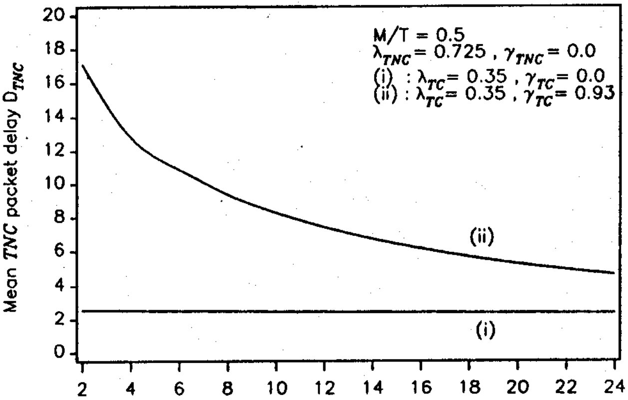

Figure 9

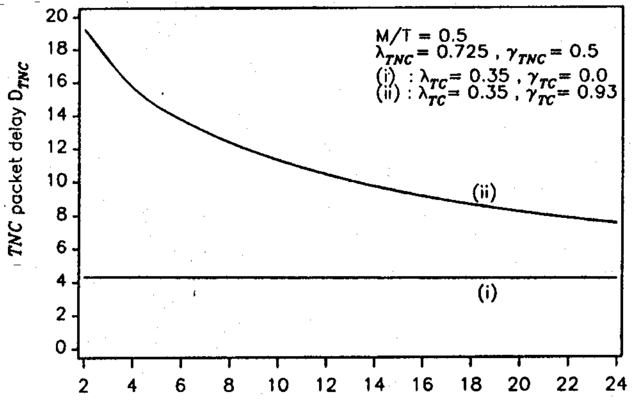

Figure 10

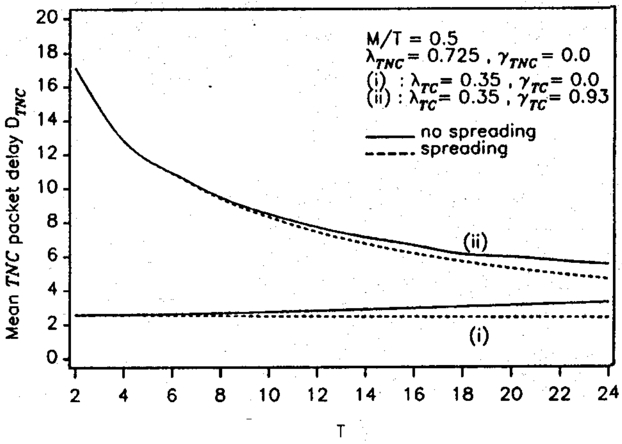

Figure 11

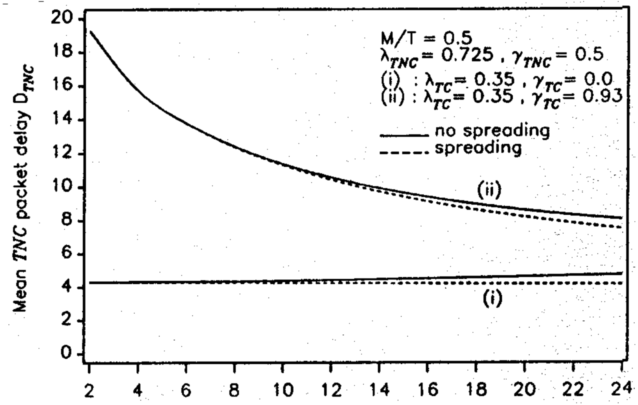

Figure 12

REFERENCES

[1] M.J. Fischer and T.C. Harris, "A model for evaluating the performance of an integrated circuit-and packet-switched multiplex structure," IEEE Transactions on Communications, vol. COM-24, pp. 198-202, Feb. 1976.

[2] C.J. Weinstein, M.L. Malpass, and M.J. Fischer, "Data traffic performance of an integrated circuit-and packet-switched multiplex structure," IEEE Transactions on Communications, vol COM-28, pp. 873-878, June 1980.

[3] A. Leon-Garcia, R.H. Kwong, and G.F. Williams, "Performance evaluation methods for an integrated voice/data link," IEEE Transactions on Communications, vol COM-30, pp. 1848-1858, Aug. 1982.

[4] K. Sriram, K.K. Varshney, and J.G. Shanthikumar, "Discrete-time analysis of integrated voice/data multiplexers with and without speech

activity detectors," IEEE Journal on Selected Areas in Communications, vol. SAC-1, pp. 1124-1132, Dec. 1983.

[5] A.G. Konheim and R.L. Pickholtz, "Analysis of integrated voice/data multiplexing," IEEE Transactions on Communications, vol. COM-32, pp. 140-147, Feb. 1984.

[6] G.F. Williams and A. Leon-Garcia, "Performance analysis of integrated voice and data hybrid-switched links," IEEE Transactions on Communications, vol. COM-32, pp. 695-706, June 1984.

[7] San-qi Li and J.W. Mark, "Performance of voice data integration on a TDM system," IEEE Transactions on Communications, vol. COM-33, pp. 1265-1273, Dec. 1985.

[8] B. Mukherjee and J.S. Meditch, "Integrating voice with the p_i - persistent protocol for unidirectional broadcast bus networks," IEEE Transactions on Communications, vol. COM-36, pp. 1287-1295, Dec. 1988.

[9] R.F. Chang and V.O.K. Li, "Analysis of packet delay for an integrated voice-data system," IEEE GLOBECOM'88 Conference, pp. 1095-1100, Nov. 28 - Dec. 1, 1988, Hollywood, Florida.

[10] Z. Zhang and I. Rubin, "Bounds on the mean system-size and delay for a movable-boundary integrated circuit and packet switched communications channel," IEEE INFOCOM '90 Conference, pp. 866-873, June 5-8, 1990, San Francisco.

[11] O.W. Yang and J.W. Mark, "Queueing analysis of an integrated services TDM system using a matrix-analytic method," IEEE Journal on Selected Areas in Communications, vol. SAC-9, pp. 88-94, Jan. 1991.

[12] R.R. Anderson, G.J. Foschini, and B. Gopinath, "A Queueing model for a hybrid data multiplexer," Bell System Technical Journal, vol.58, pp. 279-300, Feb. 1979.

[13] K. Apostolidis, X.H. Xing, and L. Merakos, "A protocol for voice and data integration in unidirectional bus networks," International Conference on Communications ICC'90, pp.279-283, April 16-19, 1990, Atlanta.

[14] I. Stavrakakis, "Analysis of a Statistical Multiplexer under a General Input Traffic Model," IEEE INFOCOM '90 Conference, pp. 1220-1225, June 5-8, 1990, San Francisco.

[15] I. Stavrakakis, "Efficient Modeling of Merging and Splitting Processes in Large Networking Structures," IEEE Journal on Selected Areas in Communications, Vol. 9, No. 8, Oct. 1991.

[16] M.L. Luhanger, "A fluid approximation model for an integrated packet voice and data multiplexer," IEEE INFOCOM '88 Conference, pp. 687-692, March 28-31, 1988, N. Orleans.

[17] I. Stavrakakis, S. Tsakiridou, "A Unified Analysis of Integrated Services TDM under Various Policies and Packet Arrival Processes", IEEE Transactions on Communications, submitted June 1991.