Delivering Diverse Delay/Dropping QoS Requirements in a TDMA Environment*

Abstract

In *Wireless ATM networks* transmission resources are shared among geographically disperse applications with diverse Quality of Service (QoS) requirements and traffic characteristics. To provide QoS guarantees and use the bandwidth efficiently, call admission and scheduling functions are necessary. These functions should ensure the delivery of the target QoS to the supported applications while achieving statistical multiplexing gains, without explicit and continuous exchange of information between sources and scheduler.

In this paper the problem of sharing resources (slots of an up-link TDMA frame) among heterogeneous Variable Bit Rate (VBR) applications with diverse QoS requirements is addressed. The QoS requirements for each application is defined in terms of a maximum tolerable packet delay and dropping probability; a packet is dropped if it experiences excess delay. The region of achievable QoS vectors is established for policies that are *work-conserving* and satisfy the *earliest due date* (EDD) service criterion (WC-EDD policies); such policies are known to optimize the overall system performance. In addition to the determination of the region of achievable QoS vectors, this study leads also to the construction of scheduling policies which deliver any performance in the region established for WC-EDD policies. Finally, an upper bound on the region of QoS vectors that can be achieved under *any* policy (not limited to the WC-EDD policies) is determined.

1 Introduction

Recently, system architectures have been proposed to enable "Wireless ATM" (WATM) [1, 2, 3]. These systems employ a Data Link Control (DLC) layer, to combat the unreliability of the wireless link, and a medium access control (MAC) protocol, to organize the sharing of the multi-access channel (wireless up-link). MACs for WATM have been examined in [1, 4, 5, 6, 7] and all employ TDMA with on demand assignment of the transmission resources. By allocating resources on demand for Variable Bit Rate (VBR) sources (while a call is in progress) statistical multiplexing gains are achieved which results in efficient use of the available bandwidth.

Due to the lack of resources in a wireless network, complete continuous information of the status of the distributed source packets may not be known. The information needed to assign transmissions may only be updated periodically or at discrete moments, as in [1, 4, 5, 6, 7]. These systems use a combination of control channels, piggy-backing on information bearing packets and polling procedures to communicate service request and transmission assignments between the scheduler and source. In this environment a MAC protocol should determine a transmission scheduling policy in a way that the probability of violating the maximum tolerable packet delay - which leads to a packet dropping rate because of discarding packets with excess delay - does not exceed the required maximum packet

*Research supported in part by the Advanced Research Project Agency under Grant F49620-93-1-0564 monitored by the Air Force Office of Scientific Research (AFOSR) and the National Science Foundation under Grant NCR 9628116.

dropping probability. At the same time a call admission function should ensure that the requesting service from all of the supported applications is possible to deliver. The maximum packet delay and packet dropping probability are the critical QoS parameters for real-time oriented services. The call admission and scheduling functions ensure the delivery of target QoS to the supported applications while achieving statistical multiplexing gains.

In this paper, call admission and transmission scheduling policies are considered for a TDMA system, where service requests are processed at frame boundaries. This resource structure has been widely considered in both cellular systems [8, 9] and wireless LANs [10], as well as in recent work toward the development of wireless ATM networks [1]. Transmission scheduling policies for this environment were developed in [11] to accommodate heterogeneous Variable Bit Rate (VBR) applications with diverse packet dropping requirements. In that work, packets which are not serviced in the frame following their arrival are considered to have excess delay and are dropped. That is, all packets assume a common maximum tolerable delay of one frame. In the present work, the packets are considered to have diverse maximum delay as well as dropping probability requirements. As it becomes apparent in this paper, the added diversity regarding the maximum delay creates a number of issues. Some of these issues are addressed by considering a specific family of policies, as discussed in the main body of this paper.

In the next section, the system model considered in this work is described. In section 3, the region of achievable QoS vectors is established for policies that are *work-conserving* and satisfy the *earliest due date* (EDD) service criterion (WC-EDD policies). In this work, Determining this region is central to the call admission control problem as well as the design of effective transmission scheduling algorithms. In section 4, the probability distribution of the residual traffic process is derived for arrival processes modeled in terms of a sequence of i.i.d. random variables. Section 5 examines a class of policies that can deliver any achievable WC-EDD performance. An upper bound on the region of QoS vectors that can be achieved under *any* policy (not limited to the WC-EDD policies) is determined in section 6. Numerical examples are presented in section 7 and the conclusions of this work are contained in section 8.

2 System Model

In this paper, the problem of sharing T resources by N heterogeneous Variable Bit Rate (VBR) source applications is considered. The source packet arrival process is described in terms of a general arrival process embedded at the boundaries of fixed length intervals called service cycles (or frames). No additional assumptions for the packet arrival process are necessary at this point. Up to T packets may be transmitted (served) during each service cycle. Depending on the Quality of Service (QoS) requirements, packets which cannot be transmitted over the service cycle following their arrival may be dropped (due to delay violation) or may be delayed to compete for service in the next frame. A TDMA system in which arrivals are considered at frame boundaries may be modeled in terms of a discrete time system in which packet delays are measured in frames (1 Frame = L time units). In the system considered in this work, the N VBR sources are partitioned into two classes, $S_1 = \{1, 2, \ldots, K\}$ and $S_2 = \{K+1, K+2, \ldots, N\}$. Packets generated from sources in S_1 have a common maximum delay tolerance of L time units (1 frame) and packets generated from sources in S_2 have a common maximum delay tolerance of $2L$ time units (2 frames).

Let $\lambda_i(n),\ i \in \{S_1 \cup S_2\}$ denote the number of newly generated packets requesting service from sources in S_1 and S_2 at the n^{th} frame boundary. The aggregate traffic from sources in S_1 is given by $\lambda_{S_1}(n) = \sum_{i \in S_1} \lambda_i(n)$, and has a delay tolerance of L time units, thus, must be either serviced or dropped over frame n. Newly generated requests from sources in S_2, $\lambda_{S_2}(n) = \sum_{i \in S_2} \lambda_i(n)$, may be either serviced or delayed to the next frame to compete for service. Packets from S_2 that have been delayed (and must be serviced or dropped in the current frame) are denoted as, $\lambda_i^{r,f}(n)$; the superscript r is used to indicate residual traffic (those packets from $\lambda_i(n-1)$, $i \in S_2$ not serviced) and the superscript f indicates its dependency on the service policy f. Thus, the total residual traffic from sources in class S_2 requiring service in frame n is given by, $\lambda_{S_2}^{r,f}(n) = \sum_{i \in S_2} \lambda_i^{r,f}(n)$.

The number of packets from source i that are dropped under policy f during frame n are given by,

$$d_i^f(n) = \lambda_i(n) - a_i^f(n), \qquad i \in S_1 = \{1, 2, \ldots, K\}, \tag{1}$$
$$d_i^f(n) = \lambda_i^{r,f}(n) - a_{i,1}^f(n), \qquad i \in S_2 = \{K+1, K+2, \ldots, N\}; \tag{2}$$

$a_i^f(n)$ denotes the number of packets from source i serviced during frame n under policy f; $a_{i,1}^f(n)$, $i \in S_2$, denotes the amount of service (in slots) provided to the residual traffic associated with source i during frame n. Since in any frame two types of packet from a source in S_2 may be present (new arrivals and residual traffic), the total amount of resources devoted to source $i \in S_2$ in frame n, is given by,

$$a_i^f(n) = a_{i,1}^f(n) + a_{i,2}^f(n), \qquad i \in S_2, \tag{3}$$

where $a_{i,2}^f(n)$ is the amount of available resources allocated to the new arrivals from S_2 in frame n under some policy f. Since the residual traffic in frame n consist of the new arrivals from sources in S_2 which did not receive service in frame $(n-1)$, the residual traffic in frame n under policy f is given by,

$$\lambda_i^{r,f}(n) = \lambda_i(n-1) - a_{i,2}^f(n-1) \qquad i \in S_2. \tag{4}$$

Let $d_i^f = E\left[d_i^f(n)\right]$, $a_i^f = E\left[a_i^f(n)\right]$ and $\lambda_i = E\left[\lambda_i(n)\right]$ be the (assumed time invariant) expected values of the associated quantities. The residual traffic is illustrated in the realization depicted in Fig. 1.

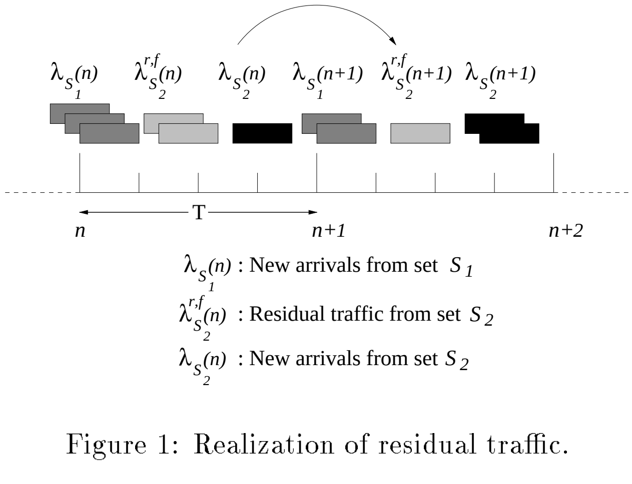

Figure 1: Realization of residual traffic.

If a Work Conserving Earliest Due Date (WC-EDD) policy is employed, then four out of the five packets with a service deadline (or due date) in the current frame will be served, the remaining one will be dropped, and the new arrival from class S_2 will form the residual traffic in the next frame (frame

$(n+1)$). According to the WC-EDD policy, all three packets with service deadline in frame $(n+1)$ will be served during the frame, as well as one of the two packets with service deadline in frame $(n+2)$.

From the above example it is evident that the employed EDD policy imposes restrictions on the level of QoS diversification that could be achieved otherwise. For instance, new arrivals from class S_2 cannot be serviced in the presence of packets from class S_1, imposing a limit on the minimum dropping rate for sources in class S_2. This limit is higher than the dropping rate achieved if, for instance, all packets (new and residual) from sources in class S_2 had service priority over those in S_1.

In most of this paper, the class of WC-EDD polices is considered for the following reasons. First, the WC-EDD policies are known to minimize the *system* packet dropping (delay violation) probability [12], resulting in throughput maximization. Unlike a more general case in which the service deadlines or due dates would form a continuum or may be drawn from a large collection of values, only two service deadlines are considered in the TDMA environment in this work. As a consequence, a potentially large number of packets from different sources will have identical service deadlines or due dates (one of two values) and, thus, significant room for dropping rate diversification may be possible without departing from the WC-EDD policies. In addition, it is possible to determine the region of achievable QoS vectors under *any* WC-EDD policy, as well as scheduling policies delivering any QoS vector in this region.

If the QoS vector is not in the region of achievable QoS vectors under the WC-EDD policies, it can be concluded that such level of QoS diversification *may* be achieved only at the expense of system throughput[13]. This may suggest that the sharing of the resources by such diverse applications may need to be restricted by allowing for resource sharing by less diverse applications. In any case, by deriving an upper bound of the region of achievable QoS vectors under *any* WC policy, it can be determined whether a given QoS vector is not achievable.

In this work the QoS requirements of application i is described in terms of a maximum tolerable delay and a maximum dropping probability p_i; this is the probability that a packet from source i experiences a delay greater than its maximum tolerable delay and, thus, is dropped. The corresponding packet dropping rate or delay violation rate, d_i (measured in expected number of dropped packets per frame) is easily determined by, $d_i = p_i \lambda_i, \ 0 \leq i \leq N$. In the rest of this paper the QoS vector associated with the supported applications will be described in terms of the dropping rates, with the understanding that these rates are induced due to violation of diverse delay tolerances. The QoS vector associated with the supported applications can be defined in terms of the (performance) packet dropping rate vector $\mathbf{d}$,

$$\mathbf{d} = (d_1, d_2, \ldots, K, K+1, \ldots, d_N). \tag{5}$$

The first question addressed in the sequel (section 3) is whether a given QoS vectors $\mathbf{d}$ is achievable under any WC-EDD policy f. Necessary and sufficient conditions are derived in order for the QoS vector to be achievable under these policies, leading to the precise determination of the region of achievable QoS vectors $\mathbf{d}$. The second question, addressed in section 5, is concerned with the design of scheduling policies that deliver any achievable QoS vector $\mathbf{d}$. Also, an upper bound of the region of QoS vectors achieved under *any* WC policy (not necessarily a WC-EDD policy) is determined in section 6.

3 Region of Achievable QoS Vectors Under WC-EDD Policies

3.1 Conservation Law and Inequality Constraints

A formal definition of a work-conserving scheduling policy for the system described in the previous section is given first.

Definition 3.1 *A scheduling policy f, is work-conserving iff it satisfies the following conditions:*

$$\sum_{i=1}^{K} d_i^f(n) + \sum_{i=K+1}^{N} \lambda_i^{r,f}(n+1) \;=\; 0 \quad \textit{if} \;\; \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) + \lambda_{S_2}(n) \leq T \tag{6}$$

$$\sum_{i=1}^{N} a_i^f(n) \;=\; T \quad \textit{if} \;\; \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) + \lambda_{S_2}(n) > T \;\; . \;\; \square \tag{7}$$

A work-conserving policy does not waste resources (slots) as long as there is work to perform (packets to transmit). Let $S = \{S_1 \cup S_2\}$ be the set of all sources and $d_{\{S\}}^f$ denote the average system packet dropping rate under scheduling policy f, given by,

$$d_{\{S\}}^f \triangleq E\left[\sum_{i=1}^{N} d_i^f(n)\right] = \sum_{i=1}^{N} E\left[d_i^f(n)\right] = \sum_{i=1}^{N} d_i^f. \tag{8}$$

Definition 3.2 *Let $\mathcal{F}$ be the family of <u>Work-Conserving Earliest Due Date (WC-EDD)</u> policies.* $\square$

Definition 3.3 *Frame n is said to be underloaded when $\lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) \leq T$ and overloaded when $\lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) > T$.* $\square$

Notice that during an overloaded frame, packets will be dropped and none of the $\lambda_{S_2}(n)$ packets will receive service under any policy $f \in \mathcal{F}$. Moreover, for any $f \in \mathcal{F}$ the following hold:

$$d_i^f(n) = \begin{cases} = 0 & \text{if } \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) \leq T \\ \geq 0 & \text{if } \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) > T \end{cases}, \; \forall i \in S, \tag{9}$$

and,

$$\lambda_i^{r,f}(n+1) = \begin{cases} \lambda_i(n) + \lambda_i^{r,f}(n) - a_i^f(n) & \text{if } \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) \leq T \\ \lambda_i(n) & \text{if } \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) > T \end{cases}, \; \forall i \in S_2. \tag{10}$$

Theorem 3.1 *The system dropping rate, $d_{\{S\}}^f$, is conserved under any $f \in \mathcal{F}$ and is a lower bound on the system dropping rate induced under any policy. More specifically,*

$$d_{\{S\}}^f \triangleq \left\{E\left[\lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) \;\middle|\; \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) > T\right] - T\right\} P\left(\lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) > T\right) \tag{11}$$

$$\triangleq b_{\{S\}} \quad \forall f \in \mathcal{F}. \;\; \square$$

Proof. First, the first equality in (11) is proved. Summing (1) and (2) over all $i \in S$, the following is obtained:

$$\sum_{i \in S} d_i^f(n) = \sum_{i \in S_1} \lambda_i(n) + \sum_{i \in S_2} \lambda_i^{r,f}(n) - \sum_{i \in S_1} a_i^f(n) - \sum_{i \in S_2} a_{i,1}^f(n). \tag{12}$$

Since f is a WC-EDD policy it satisfies (6) and (7) of Definition 3.1 and (9) and (10) of Definition 3.3. Thus, the above expression becomes,

$$\sum_{i \in S} d_i^f(n) = \begin{cases} 0 & \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) \leq T \\ \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) - T & \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) > T. \end{cases} \tag{13}$$

By applying the expectation operator in (13), the first equality is obtained. To prove the second equality in (11) - and thus, show that d_S^f is conserved - it suffices to show that $\lambda_{S_2}^{r,f}(n)$ is independent of the policy $f \in \mathcal{F}$. For any $f \in \mathcal{F}$,

$$\lambda_{S_2}^{r,f}(n+1) = \begin{cases} \max\{0, \lambda_{S_2}(n) + \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) - T\} & \text{if } \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) \leq T \\ \lambda_{S_2}(n) & \text{if } \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) > T. \end{cases} \tag{14}$$

Therefore, $\lambda_{S_2}^{r,f}(n)$ depends on $\lambda_{S_2}^{r,f}(n-1)$ which may in-turn be dependent on f and is otherwise independent from f. By induction $\lambda_{S_2}^{r,f}(n)$ is only dependent on the initial conditions $\lambda_{S_2}^{r}(0)$ which is independent of the policy and equal to zero. Thus, the *total* residual traffic in frame n is independent of the policy $f \in \mathcal{F}$ and therefore can be denoted as $\lambda_{S_2}^{r}(n)$. That is,

$$\lambda_{S_2}^{r,f}(n) = \sum_{i \in S_2} \lambda_i^{r,f}(n) \triangleq \lambda_{S_2}^{r}(n), \quad \forall n, \quad \forall f \in \mathcal{F}, \tag{15}$$

and therefore proving the second equality. Finally, it is a well known result that $d_{\{S\}}^f$ is the minimum dropping rate since it is induced by a WC-EDD policy[12]. □

Since the total residual traffic in frame n, $\lambda_{S_2}^{r}(n)$, is independent of the policy $f \in \mathcal{F}$, the following corollary is self-evident in view of the previous theorem.

Corollary 3.1 *The number of service opportunities for the new arrivals from S_2 in frame n is independent of the selected $f \in \mathcal{F}$. Therefore, the residual traffic from any source i in frame $(n+1)$, $\lambda_i^{r,f}(n+1)$, is only dependent on the policy f chosen in the present frame, n, and not in past frames.* □

Corollary 3.1 is employed in the proof of Theorem 3.5 and in the development of a class of policies in section 5.

Let $\lambda_g(n) = \sum_{i \in g} \lambda_i(n)$ be the aggregate arrivals from sources in subset g in frame n. Let d_g^f denote the aggregate packet dropping rate associated with sources in group g only, under policy f; all N sources in S are assumed to be present and served under the policy f. d_g^f is defined by,

$$d_{\{g\}}^f = E\left[\sum_{i \in g} d_i^f(n)\right] = \sum_{i \in g} E\left[d_i^f(n)\right] = \sum_{i \in g} d_i^f, \quad g \subseteq S. \tag{16}$$

The following lemma will be used in the proof of the theorem that follows.

Lemma 3.1 *Let $g \subseteq S_2$ and let $\lambda_g^{r,f}(n) = \sum_{i \in g} \lambda_i^{r,f}(n)$ be the aggregate residual traffic from subset g under some policy f. Then,*

$$\lambda_g^{r,f}(n) \geq \lambda_g^{r}(n), \quad \forall n \quad \text{and} \quad \forall f \in \mathcal{F}, \tag{17}$$

where $\lambda_g^{r}(n)$ is given by,

$$\lambda_g^{r}(n+1) = \begin{cases} \max\{0, \lambda_g(n) + \lambda_{S_1}(n) + \lambda_{S_2}^{r}(n) - T\} & \text{if } \lambda_{S_1}(n) + \lambda_{S_2}^{r}(n) \leq T \\ \lambda_g(n) & \text{if } \lambda_{S_1}(n) + \lambda_{S_2}^{r,f}(n) > T. \end{cases} \tag{18}$$

$\lambda_g^{r}(n)$ is called the minimum residual traffic process for sources in $g \in S_2$, for a system that is served under a WC-EDD policy. □

Proof. The above inequality is self-evident since $\lambda_g^r(n)$ corresponds to the residual traffic where new arrivals from sources in g are given all excess resources[1]. $\Box$.

Theorem 3.2 *Let $b_{\{g\}}$ denote the lower bound for the aggregate packet dropping rate for sources in set g, $g \subset S$, under any policy $f \in \mathcal{F}$. Then this bound is given by,*

$$b_{\{g\}} = \left\{ E\left[\lambda_x(n) + \lambda_y^r(n) \;\middle|\; \lambda_x(n) + \lambda_y^r(n) > T \right] - T \right\} P\left(\lambda_x(n) + \lambda_y^r(n) > T \right), \qquad (19)$$
$$g =\subseteq S,$$
$$x = \{g \cap S_1\} \neq \emptyset \text{ and } y = \{g \cap S_2\} \neq \emptyset. \quad \Box$$

Proof. If $g \subseteq S_2$, the bound $b_{\{g\}}$ is achieved when $\lambda_g^r(n)$ packets do not see any interference from packets in S_1 and $\{S_2 - g\}$, or equivalently, all T resources are available to service the $\lambda_g^r(n)$. By employing $\lambda_g^r(n)$ in (11), (19) is obtained for when $g \subseteq S_2$. The probability distribution of the minimum residual traffic, $\lambda_g^r(n)$, is calculated from (18) as in section 4. If $g \subseteq S_1$, the proof can be derived as in [11]. If $g \subset S$, where $x = \{g \cap S_1\}$ and $y = \{g \cap S_2\}$ are non-empty, (generalizing for the above cases) $b_{\{g\}}$ is achieved when the traffic $\lambda_x(n) + \lambda_y^r(n)$ does not see interference from traffic from sources in $\{S - g\}$. By generalizing (19), the proof follows. $\Box$

The following class of policies is needed for the establishment of the region of achievable QoS vectors under the WC-EDD policies, as well as for the development of policies which deliver any achievable QoS vector for this region.

Definition 3.4 *A Deadline-Sensitive Ordered Head of Line (DSO-HoL) priority service policy is defined to be the policy which first separates packets into two sets: packets having a service deadline in the current frame and packets having a service deadline in the next frame. Packets from sources having a service deadline in the current frame are serviced according to a priority service policy, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$; $\pi_i \in \{1, 2, \ldots, N\}$, $\pi_i \neq \pi_j$, $1 \leq i, j \leq N$. The index of π_i indicates the order of the priority given to the π_i source to service packets having a service deadline in the current frame. None of the π_j sources, $j > i$, may be served as long as packets with current service deadline from sources π_k, $k \leq i$, are present. After servicing the packets having a current deadline, the same service policy, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$, is followed for packets from sources that are present which do not have a current deadline.* $\Box$

Theorem 3.3 *The following constraints associated with the induced packet dropping rate vector $\mathbf{d}^f = (d_1^f, d_2^f, \ldots, d_N^f)$ are satisfied by any scheduling policy $f \in \mathcal{F}$,*

$$d_{\{g\}}^f \geq b_{\{g\}}, \quad \forall g \subset S, \quad \forall f \in \mathcal{F} \qquad (20)$$

where $d_{\{g\}}^f$ and $b_{\{g\}}$ are given in (16) and (19). $\Box$

Proof. Let $g \subset S$ and $x = g \cap S_1$ and $y = g \cap S_2$. Suppose that the packets from sources in g are served under a scheduling policy $f_o \in \mathcal{F}$ according to which they are given Deadline Sensitive Head of the Line (DS-HoL) priority over the packets from sources in $\{S - g\}$. That is, during overload conditions no packet from sources in $\{S - g\}$ is served unless no packets with a current service deadline from

[1] As state in corollary 3.1, the number of service opportunities (or excess resources) available to the new arrivals, $\lambda_i(n)$, $i \in S_2$, is independent of the policy $f \in \mathcal{F}$.

sources in g are present. During underload conditions, no packets from sources in $\{S_2 - y\}$ are serviced while packets from sources in $y = \{g \cap S_2\}$ are present, and therefore the minimum residual traffic is achieved for set y. Thus,

$$\sum_{i \in g} d_i^{f_o}(n) = \begin{cases} 0 & \text{if } \lambda_x(n) + \lambda_y^r(n) \leq T \\ \lambda_x(n) + \lambda_y^r(n) - T & \text{if } \lambda_x(n) + \lambda_y^r(n) > T. \end{cases} \tag{21}$$

From (19) and (21) it is clear that,

$$d_{\{g\}}^{f_o} = b_{\{g\}}, \tag{22}$$

for all policies $f_o \in \mathcal{F}$ that provide DS-HoL priority to the packets from sources in g. Since no other policy $f \in \mathcal{F}$ can provide better service (that is, lower aggregate packet dropping rate) to sources in g than f_o, it is evident that,

$$d_{\{g\}}^{f} \geq b_{\{g\}}, \quad \forall f \in \mathcal{F}, \quad g \subset S. \ \Box \tag{23}$$

3.2 Region of Achievable QoS Vectors

The main result of this section is the determination of the region $\mathcal{D}$ of the achievable QoS vectors $\mathbf{d}$ under WC-EDD policies. The following corollary provides a set of necessary conditions in order for a QoS vector $\mathbf{d}$ to be achievable, followed by a corollary regarding an upper bound on the achievable region for QoS vectors $\mathbf{d}$ under WC-EDD policies. Their proofs are self-evident in view of Theorems 3.1 and 3.3.

Corollary 3.2 *A necessary condition in order for a QoS vector $\mathbf{d} = (d_1, d_2, \ldots, d_N)$ to be achieved by a policy $f \in \mathcal{F}$ is that its components satisfy the following constraints,*

$$d_{\{g\}} \geq b_{\{g\}} \quad \forall g \subset S \tag{24}$$

$$d_{\{S\}} = b_{\{S\}}. \ \Box \tag{25}$$

Corollary 3.3 *Let $\mathcal{D}^u$ denote the collection of all vectors $\mathbf{d}$ satisfying (24) and (25). Then $\mathcal{D}^u$ is an upper bound on the region $\mathcal{D}$ of achievable QoS vectors $\mathbf{d}$. That is,*

$$\mathcal{D} \subseteq \mathcal{D}^u. \ \Box \tag{26}$$

The following theorem describes the vectors contained in $\mathcal{D}^u$.

Theorem 3.4 *Any vector in the set $\mathcal{D}^u$ can be expressed as a convex combination of extreme points (vertices) of $\mathcal{D}^u$; that is, $\mathcal{D}^u$ may be expressed as the convex hull of its extreme points, $\mathcal{D}^u$=conv[ext($\mathcal{D}^u$)].* $\Box$

Proof. The proof follows from the fact that $\mathcal{D}^u$ is a bounded set defined by a finite intersection of closed half spaces, (see (24) and (25)). Then by definition, $\mathcal{D}^u$ is a polytope [14] and Theorem 3.4 follows directly from properties of polytopes [14]. $\Box$

The following theorem establishes a relationship between scheduling policies and the vertices of $\mathcal{D}^u$.

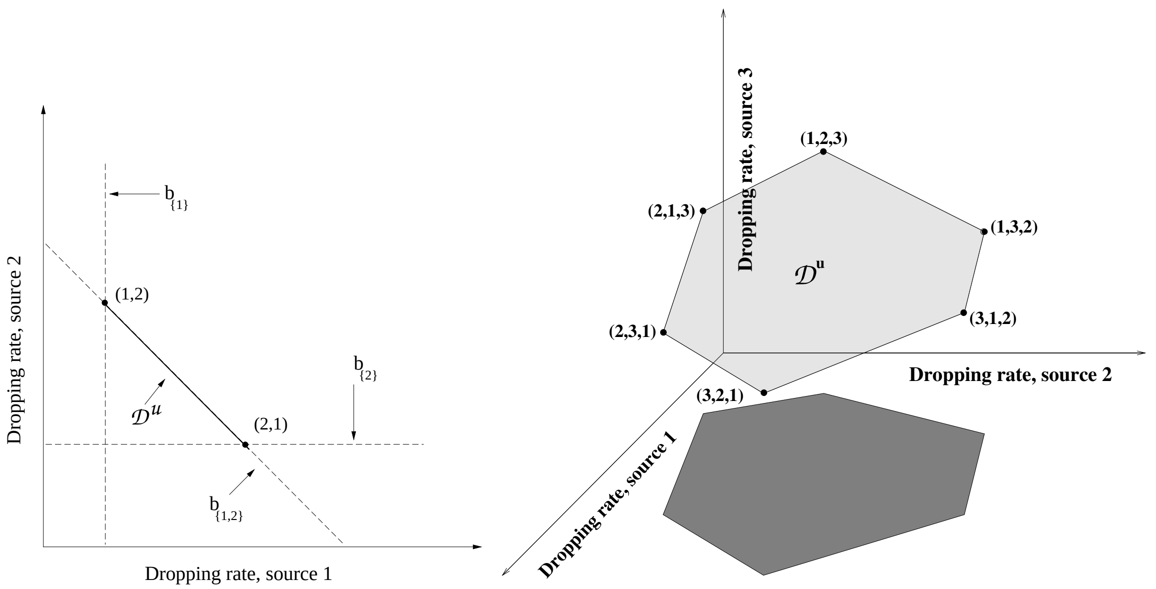

Figure 2: The region (polytope) $\mathcal{D}^u$ for a system with two sources.

Theorem 3.5 $\mathbf{d}^*$ *is a vertex (extreme point) of the set* $\mathcal{D}^u$ *iff* $\mathbf{d}^*$ *is a dropping rate vector resulting from a Deadline-Sensitive Ordered HoL (DSO-HoL) priority service policy,* $\boldsymbol{\pi}= (\pi_1, \pi_2, \ldots, \pi_N)$. $\Box$

The proof of Theorem 3.5 can be found in [15], and follows from the super-modularity property of the set function $b_{\{g\}}$.

Fig. 2 provides a graphical illustration of the region $\mathcal{D}^u$ for the case of $N = 2$ and $N = 3$ sources. The extreme points $\mathbf{d}_{ext-i}$'s correspond to QoS vectors induced by the $(N!)$ DSO-HoL priority policies $\boldsymbol{\pi}^i= (\pi_1, \pi_2, \ldots, \pi_N)$, $1 \leq i \leq N!$, as shown in Theorem 3.5. Referring to the first figure in Fig. 2, it may be observed that the policy $(\pi_1, \pi_2) = (1, 2)$ corresponds to the intersection of the line for the lower bound on the packet dropping rate line for source 1, $b_{\{1\}}$, with the system dropping rate line $b_{\{1,2\}}$. Similarly, the second extreme point induced by the policy $(\pi_1, \pi_2) = (2, 1)$ is the intersection of the lower bound on the packet dropping rate for source 2, $b_{\{2\}}$ and $b_{\{1,2\}}$. Similar observations can be made for the region $\mathcal{D}^u$ for a system of $N = 3$ sources shown in Fig. 2.

Let $\mathcal{D}$, $\mathcal{D} \subseteq \mathcal{D}^u$, be the achievable region of QoS vectors $\mathbf{d}$. The following theorem establishes its convexity.

Theorem 3.6 *Let* $\mathbf{d}_1, \mathbf{d}_2 \in \mathcal{D}$, *then* $\mathbf{d}_3$, *where*

$$\mathbf{d}_3 = \alpha \mathbf{d}_1 + (1 - \alpha)\mathbf{d}_2, \qquad \alpha \geq 0, \quad \alpha \leq 1, \tag{27}$$

is also in $\mathcal{D}$. *That is* $\mathcal{D}$ *is convex.* $\Box$

Proof. Let $\mathbf{d}_1$ ($\mathbf{d}_2$) be the dropping rate vector induced by the policy f_1 (f_2) $\in \mathcal{F}$; that is, $\mathbf{d}_1, \mathbf{d}_2 \in \mathcal{D}$. Consider a scheduling policy f_3 that at each underloaded frame decides to follow the scheduling rule of policy f_1 with probability α and policy f_2 with probability $(1 - \alpha)$; the decisions over consecutive underloaded frames are independent. Notice that no dropping occurs during underloaded frames, and thus, any policy $f \in \mathcal{F}$ could be selected. In fact any policy $f \in \mathcal{F}$ can be selected during an underloaded frame, except for the frame before an overloaded one, since shown in corollary 3.1, the residual traffic for source i in frame n, $\lambda_i^{r,f_1(f_2)}(n)$ is only dependent on the policy $f_1(f_2)$ selected in frame $(n-1)$ and not in earlier frames. Decisions over the first overloaded frame following an underloaded one, follow the policy selected for the previous frame, ensuring that the $\lambda_i^{r,f_1(f_2)}(n)$ packets are serviced under policy $f_1(f_2)$[2]. Over overloaded frames other than the first of an overloaded period, policy $f_1(f_2)$ is selected with probability α $(1-\alpha)$ independently from the selection in the previous frame. Note that the selected policy during an overloaded frame does *not* affect the residual traffic in the next frame, refer to (18). It should also be noted that the aggregate traffic, $\lambda_{S_1}(n) + \lambda_{S_2}^r(n)$, is independent of the policy $f \in \mathcal{F}$ (Theorem 3.1). This implies that the occurrence of overloaded periods is independent of the policy and the length of consecutive overload periods are independent and identically distributed random variables, regardless of the selected policy. In view of the above, it is easy to establish that the dropping rate performance of policy $f_1(f_2)$ is induced with probability α $(1-\alpha)$. Thus, the packet dropping rate induced by policy S_3 is given by,

$$\mathbf{d}_3 = \alpha\mathbf{d}_1 + (1-\alpha)\mathbf{d}_2. \tag{28}$$

Since $\mathbf{d}_3$ is achieved by a policy in $\mathcal{F}$, $\mathbf{d}_3 \in \mathcal{D}$, establishing the convexity of $\mathcal{D}$. $\square$

The next theorem establishes the region $\mathcal{D}$ of achievable QoS vectors under WC-EDD policies.

Theorem 3.7 $\mathcal{D} \equiv \mathcal{D}^u$. $\square$

Proof. Since $\mathcal{D} \subseteq \mathcal{D}^u$ (Corollary 3.3), it suffices to establish that $\mathcal{D}^u \subseteq \mathcal{D}$ to complete the proof. Notice that $\mathcal{D}^u{=}conv[ext(\mathcal{D}^u)]$ (Theorem 3.4) and that $\mathcal{D}$ is convex (Theorem 3.6) and, thus, if $\{ext(\mathcal{D}^u)\}$ $\subseteq \mathcal{D}$ then $\mathcal{D}^u \subseteq \mathcal{D}$. The latter holds true since the extreme points of $\mathcal{D}^u$ are induced by the DSO-HoL priority policies (Theorem 3.5) and, thus, these points are in the region $\mathcal{D}$ of achievable QoS vectors. $\square$

4 Analysis of Residual Traffic

In this section the probability distribution of the minimum residual traffic process, $\lambda_g^r(n)$, is derived when the arrival processes $\lambda_i(n), i \in S$, are independent and each modeled by a sequence of independent identically distributed random variables. As seen in (14), future evolution of the total residual traffic process, $\lambda_{S_2}^r(n+1)$, depends only on the present values of $\lambda_{S_2}^r(n)$, $\lambda_{S_1}(n)$ and $\lambda_{S_2}(n)$. Therefore, under the i.i.d. assumption for the arrival processes, $\lambda_{S_2}^r(n)$ is Markovian with transition probabilities,

$$P(\lambda_{S_2}^r(n+1) = i \big| \lambda_{S_2}^r(n) = j) = \begin{cases} P(\lambda_{S_2}(n) + \min(j + \lambda_{S_1}(n), T) - T = i) & i > 0, \\ P(\lambda_{S_2}(n) + \min(j + \lambda_{S_1}(n), T) \leq T) & i = 0. \end{cases} \tag{29}$$

Assuming the stationary distribution of $\lambda_{S_2}^r(n)$ exists, the distribution vector, $\boldsymbol{\lambda}_{S_2}^r$, is easily computed as,

$$\boldsymbol{\lambda}_{S_2}^r P = \boldsymbol{\lambda}_{S_2}^r, \tag{30}$$

where P is the probability transition matrix with elements $p_{j,i}$ given in (29).

[2]Otherwise, packets dropped in the overloaded frame under policy f_1 would be associated with the residual traffic generated in the previous (underloaded) frame under policy f_2; the resulting performance would be that of neither f_1 nor f_2.

The distribution of the minimum residual traffic process for subset g in (18) is found by conditioning on $\lambda^r_{S_2}(n)$. As shown in Theorem 3.1, no matter what the subset g under consideration, the quantity $\lambda^r_{S_2}(n)$ is conserved. The stationary distribution for $\lambda^r_g(n)$ is found as,

$$P(\lambda^r_g(n)=i)=\sum_j P(\lambda^r_g(n)=i \,\big|\, \lambda^r_{S_2}(n-1)=j)P(\lambda^r_{S_2}(n-1)=j). \tag{31}$$

From the i.i.d. assumption,

$$P(\lambda^r_g(n+1)=i \,\big|\, \lambda^r_{S_2}(n)=j)=\begin{cases} P(\lambda_g(n)+\min(j+\lambda_{S_1}(n),T)-T=i) & i>0, \\ P(\lambda_g(n)+\min(j+\lambda_{S_1}(n),T)\le T) & i=0. \end{cases} \tag{32}$$

5 A Class of Policies

Let $C_{DSO-HoL}$ denote the class of Delay-Sensitive Ordered Head of Line (DSO-HoL) service policies $\boldsymbol{\pi}^i$ introduced in section 3.1.

Definition 5.1 *In each underloaded frame, a Mixing DSO-HoL Policy f_m decides to follow the DSO-HoL policy $\boldsymbol{\pi}^i$ with probability α_i, $\alpha_i \ge 0$, $1 \le i \le N!$, $\sum_{i=1}^{N!}\alpha_i = 1$; decisions over consecutive underloaded or overloaded frames are independent. The DSO-HoL policy from the previous underloaded frame is chosen in the first overloaded frame. Clearly, f_m is completely determined by the $N!$ dimensional vector $\boldsymbol{\alpha}$, $\boldsymbol{\alpha} \ge \mathbf{0}$, $\mathbf{1}\cdot\boldsymbol{\alpha} = 1$. Let $M_{DSO-HoL}$ denote the class of such policies.* □

The proof of the following theorem follows directly from the proof of Theorem 3.6.

Theorem 5.1 *The dropping rate vector induced by a Mixing DSO-HoL policy is given by,*

$$\mathbf{d}^{f_m}=\sum_{i=1}^{N!}\mathbf{d}^{f_i}\alpha_i. \quad \square \tag{33}$$

The following theorem establishes the main results of this section.

Theorem 5.2 *For each packet dropping rate vector $\mathbf{d}\in\mathcal{D}$ there exist a policy $f_m \in M_{DSO-HoL}$ that induces $\mathbf{d}$.* □

Proof. Let $\mathbf{d}\in\mathcal{D}$, then $\mathbf{d}=\sum_{i=1}^{N!}\alpha_i\mathbf{d}_{ext-i}$ for some $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\ldots,\alpha_{N!})$ where $\alpha_i \ge 0$, $1\le i\le N!$, $\sum_{i=1}^{N!}\alpha_i=1$, since any point in $\mathcal{D}$ can be written as a convex combination of the extreme points (vertices) $\mathbf{d}_{ext-i}$ of $\mathcal{D}(=\mathcal{D}^u)$; each $\mathbf{d}_{ext-i}$ is induced by some policy in $C_{DSO-HoL}$ (Theorem 3.5).

Let f_m be the mixing policy which selects the HoL priority $\boldsymbol{\pi}^i$ (that induces $\mathbf{d}_{ext-i}$) with probability α_i. The packet dropping rate vector $\mathbf{d}^{f_m}$ induced by f_m is given by,

$$\mathbf{d}^{f_m}=\sum_{i=1}^{N!}\alpha_i\mathbf{d}_{ext-i} \tag{34}$$

and thus f_m induces $\mathbf{d}$. □

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

Corollary 5.1 *Let* $\mathbf{d} \in \mathcal{D}$ *be a target packet dropping rate vector. The mixing policy* $f_m \equiv \boldsymbol{\alpha}$ *induces* $\mathbf{d}$*, where* $\boldsymbol{\alpha}$ *is such that,*

$$E_{\boldsymbol{\alpha}}\left[\mathbf{d}_{ext}\right] \triangleq \sum_{i=1}^{N!} \alpha_i \mathbf{d}_{ext-i} = \mathbf{d}, \tag{35}$$

$$\boldsymbol{\alpha} \geq \mathbf{0}, \tag{36}$$

$$\mathbf{1} \cdot \boldsymbol{\alpha} = 1, \tag{37}$$

where $E_{\boldsymbol{\alpha}}[.]$ *is weighted average of the set of extreme points,* $\mathbf{d}_{ext}$*, of* $\mathcal{D}$ *with respect to the Probability Mass Function* $\boldsymbol{\alpha}$*.* $\Box$

6 Upper Bound on the Region of Achievable QoS Vectors Under Any Service Policy

In the previous section the region of achievable QoS vectors induced by WC-EDD policies (denoted in this section by $\mathcal{D}^{EDD}$) was determined and described in terms of conditions (24) and (25), restated (using superscript EDD) as,

$$\begin{aligned} d_{\{g\}} &\geq b_{\{g\}}^{EDD} \quad \forall g \subset S, \\ d_{\{S\}} &= b_{\{S\}}^{EDD}; \end{aligned} \tag{38}$$

$b_{\{S\}}^{EDD}$ and $b_{\{g\}}^{EDD}$ are lower bounds on the performance of WC-EDD policies as given in (11) and (19), respectively. It is well known that the WC-EDD polices optimize the system (S) performance by minimizing the system dropping rate. Therefore, any policy that attempts to improve (decrease) the dropping rate for a subset of sources g beyond the lower bound shown in (18)-(19) by relaxing the EDD condition will result in increased system (S) system dropping rate. That is, the lower bounds on the dropping rates achieved by the WC-EDD policies and the class of policies which do not necessarily satisfy the EDD condition (denoted here as an Unconstrained (UC) policy) satisfy the following conditions.

$$\begin{aligned} b_{\{g\}}^{EDD} &\geq b_{\{g\}}^{UC} \quad \forall g \subset S, \\ b_{\{S\}}^{UC} &\geq b_{\{S\}}^{EDD}. \end{aligned} \tag{39}$$

$b_{\{g\}}^{UC}$ is the *unconstrained* lower bound on the dropping rate for the sources in set g. This bound is achieved by considering that packets from sources in set g *only* are present and serviced under a work-conserving earliest due date policy; sources in $\{S - g\}$ are considered to be absent. $b_{\{g\}}^{UC}$ is calculated by applying (11) and (14) to the set g only. That is, replacing S_1 and S_2 with $x = \{g \cap S_1\}$ and $y = \{g \cap S_2\}$ in (11) and (14), respectively.

Since no policy can do better for sources in set g than $b_{\{g\}}^{UC}$ when all sources in S are present, the following necessary conditions must be satisfied by any QoS vector $\mathbf{d}$ which is achieved under some policy.

$$\begin{aligned} d_{\{g\}} &\geq b_{\{g\}}^{UC} \quad \forall g \subset S, \\ d_{\{S\}} &\geq b_{\{S\}}^{EDD}. \end{aligned} \tag{40}$$

The following proposition is self-evident in view of the above discussion.

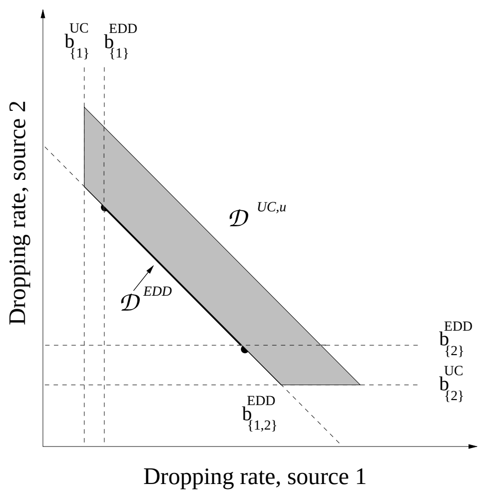

Figure 3: Necessary performance bounds under any policy and the sufficient bounds for the WC-EDD family of policies.

Proposition 6.1 *An upper bound on the region of QoS vectors achieved under* <u>*any*</u> *policy,* $\mathcal{D}^{UC,u}$*, is given by (40).* $\square$

Fig. 3 depicts the region $\mathcal{D}^{EDD}$ and $\mathcal{D}^{UC,u}$ for the case of two sources.

7 Numerical Example and Verification through Simulation

In this section two examples are presented for a system with two and three sources, respectively, competing for T slots in a TDMA frame. The source packet arrival processes are assumed to be mutually independent. Each process is described in terms of a sequences of independent and identically distributed random variables embedded at the frame boundaries. Let $E^k = \{0, 1, \ldots, M^k - 1\}$ denote the state space of the arrival process associated with source k.

Consider a system with 5 ($T = 5$) slots per frame and 2 sources, $k = 2$. In this example $M^1 = 7$ and $M^2 = 7$ and when in state i, a source generates i packets in the current frame. Source 1 has a maximum arrival rate of 6 packets per frame with an average per frame arrival rate of 3.6. Source 2 has an average arrival rate of 3.2 packets per frame with a maximum arrival rate of 6 packets per frame. Sources 1 and 2 have state probability distributions of (0.0, 0.1, 0.1, 0.3, 0.2, 0.2, 0.1) and (0.1, 0.0, 0.4, 0.0, 0.3, 0.0, 0.2), respectively. Sources 1 has a maximum delay tolerance of 1 frame while sources 2 has a maximum delay tolerance of 2 frames.

The region of achievable QoS vectors under WC-EDD policies for this system is calculated from (11), (14), (18) and (19); the bounds on the dropping rates are given in Table 1. The unconstrained lower bounds were evaluated by applying (11) and (14) to each source k only. Notice that the unconstrained lower bound for source 1, b_1^{UC}, cannot be decreased beyond the WC-EDD lower bound, b_1^{EDD}, since every packet from source 1 has a maximum delay tolerance of 1 frame. The WC-EDD bound for source 2 can be decreased by relaxing the EDD condition on the system; that is, if the policy allows service to new arrivals from source 2 before packets from source 1, a lower dropping rate will be delivered to source 2. The decrease in the dropping rate beyond the WC-EDD bound for source 2 is achieved at the expense of increasing the system dropping rate. This result is verified through simulations and displayed in Fig. 4.

b_1^{EDD}	b_2^{EDD}	$b_{\{1,2\}}^{EDD}$	b_1^{UC}	b_2^{UC}
0.100	0.160	1.856	0.100	0.001

Table 1: Lower bounds on dropping rates under WC-EDD policies and under any policy.

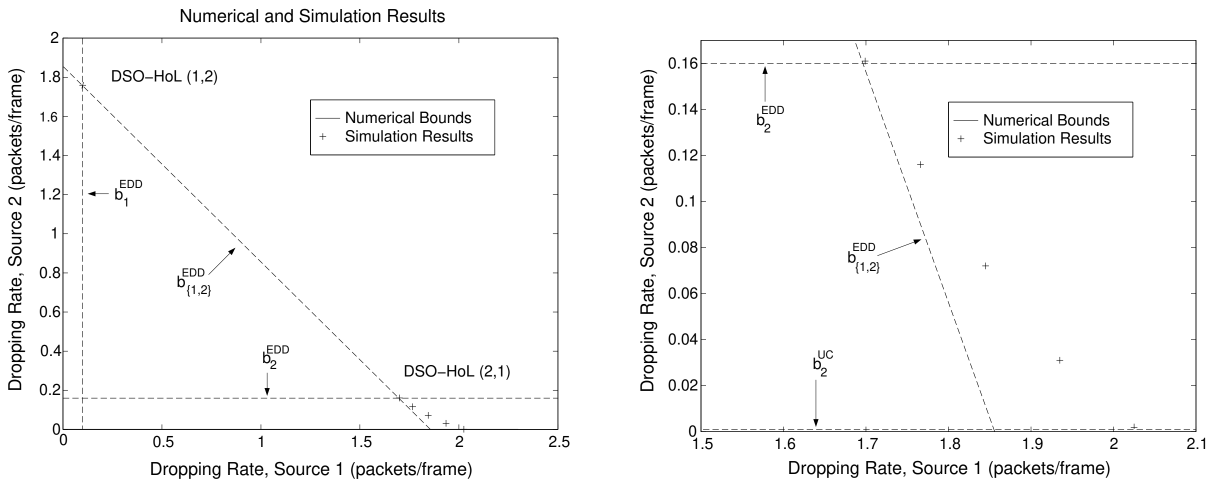

Figure 4: Evaluation of a numerical example with supporting simulation results.

In the simulation results, policies[3] were generated that ranged from a policy that gives *all* packets from source 2 service priority over packets from source 1, to the DSO-HoL service policy $\boldsymbol{\pi}$=(2,1) in which only the packets with a current service deadline from source 2 were given priority over packet from source 1. Under the former policy, the UC lower bound ($b_2 = 0.001$) is achieved for source 2, but the system dropping rate is increased from its minimum of 1.859 to 2.025 packets per frame. The three policies which induce the other points shown in Fig. 4 were obtained by varying the frequency at which source 2 was allowed to violated the EDD condition. As is clearly observed in Fig. 4, for each of the four policies that violate the EDD condition, the resulting system dropping rate is increased compared to the conserved system dropping rate induced by WC-EDD policies.

To illustrate the design of WC-EDD scheduling polices, consider the following example consisting of three VBR sources competing for 7 ($T = 7$) slots of a TDMA frame. In this example $M^1 = 3$, $M^2 = 5$ and $M^3 = 7$ and when in state i, a source generates i packets in the current frame. Each source is modeled by a sequence of i.i.d. random variables with a maximum arrival rate of 2, 4 and 6 packets per frame, respectively. The probability distributions for each source are (0.200, 0, 0.400, 0, 0.295, 0, 0.200), (0.400, 0, 0.600) and (0.36, 0.16, 0, 0.48) for source 1,2 and 3 respectively.

Packets from source 1 have a maximum delay tolerance of 1 frame while packets from sources 2 and 3 can tolerate delays of up to 2 frames. The maximum dropping probabilities (resulting from delay violations) acceptable for sources 1, 2 and 3 are: $p_1 = 0.02$, $p_2 = 0.01$ and $p_3 = 0.02$. The QoS vector in this case is: $\mathbf{d}= (0.064, 0.008, 0.032)$.

By evaluation of (24) and (25), it can be observed that $d_{\{g\}} \geq b_{\{g\}}^{EDD}$ and $d_{\{S\}} = b_{\{S\}}^{EDD}$[4]. Since $\mathbf{d} \in \mathcal{D}^{EDD}$, (as shown in section 5) there exists a mixing DSO-HoL policy $f_m \equiv \boldsymbol{\alpha}$, delivering it. Any

[3]The performance of the DSO-HoL policy $\pi = (1, 2)$ was also verified through simulation and is displayed in Fig. 4.

[4]If (25) had been satisfied with inequality, it can be shown that there exits an $f \in \mathcal{F}$ delivering "better" service to the supported applications[15]

$\boldsymbol{\alpha}$ satisfying (35)-(37) of Corollary 5.1 may be chosen. For this example, the following $\boldsymbol{\alpha}_o$ was chosen by employing linear programming techniques,

$$\boldsymbol{\alpha}_o = (0, 0, 0.3367, 0.5386, 0, 0.1248) \tag{41}$$

As it is expected from the formulation of the constraints, more than one solution may be found. This allows for the incorporation of additional constraints representing other desirable qualities of the policies. Functions of interest may be minimized subject to the constraints presented in this paper to guarantee the achievability of the resulting policies. For instance, among all mixing policies inducing $\mathbf{d}$, the one which minimizes the variance in the service provided to certain sources may be identified.

8 Conclusion

In this work, the region of achievable QoS vectors was precisely determined for a system of heterogeneous VBR sources serviced under policies that are *work-conserving* and satisfy the *earliest due date* service criterion (WC-EDD). The QoS requirements for each application were defined in terms of a maximum tolerable packet delay and dropping probability. In addition to determining the region of achievable QoS vectors, a class of scheduling policies was developed which delivers any achievable performance in the region established for WC-EDD policies. The effectiveness of these policies in delivering the QoS vectors was verified through simulation. Also, an upper bound on the region of QoS vectors that can be achieved under any policy was determined. Numerical examples were presented with simulation results verifying the theoretical results presented in this paper.

References

[1] D. Raychaudhuri and N. Wilson. ATM-Based Transport Architecture for Multiservices Wireless Personal Communication Networks. *IEEE Journal of Selected Areas Communications*, 12(8):1401–1414, October 1994.

[2] K.Y. Eng, M. Karol, and M. Veeraraghavan. A Wireless Broadband Ad-Hoc ATM Local Area Network. *ACM Wireless Networks Journal*, 1(2), December 1995.

[3] P. Mermelstien, A. Jalali, and H. Leib. Integrating Services on Wireless Multiple Access Networks. In *Proceedings of ICC*, pages 863–867, March 1993.

[4] D. Petras et al. MAC Protocol for Wireless ATM: Contention Free versus Contention Based Transmission reservation Requests. In *Proceedings of PIMRC*, Taipei, Taiwan, October 1996.

[5] M. Karol, Z. Liu, and K. Eng. An Efficient Demand-Assignment Multiple Access Protocol for Wireless Packet (ATM) Networks. *Wireless Networks*, 1(4):267–279, December 1995.

[6] A. Mahmoud, D. Falconer, and S. Mahmoud. A Multiple Access Scheme for Wireless Access to a Broadband ATM LAN Based on Polling and Sectored Antennas. *IEEE Journal on Selected Areas in Communications*, 14(4), May 1996.

[7] C. Chang, K. Chen, and M. You. Guaranteed Quality of Service Wireless Access to ATM Networks. In *ICC*, Dallas, TX, June 1996.

[8] N.D. Newman, R. Ganesh, K. Joseph, and D. Raychaudhuri. Packet CDMA Versus Dynamic TDMA for Multiple Access in an Integrated Voice/Data PCN. *IEEE Journal on Selected Areas in Communications*, 11(6):870–884, August 1993.

[9] D.J. Goodman, R.A. Valenzeula, et al. Packet Reservation Multiple Access for Local Wireless Communication. *IEEE Transactions on Communications*, 37(8):885–890, August 1989.

[10] D. Bantz and F. Baucho. Wireless LAN Design Alternatives. *IEEE Network Magazine*, 8(2):43–45, March-April 1994.

[11] *reference omitted to preserve anonymity*

[12] S.S. Panwar, D. Towsley, and J.K. Wolf. Optimal Scheduling Policies for a Class of Queues with Customer Deadlines to the Beginning of Service. *Journal of ACM*, pages 832–844, 1988.

[13] G. Chen and I. Stavrakakis. ATM Traffic Management with Diversified Loss and Delay Requirements. In *Proceedings of IEEE INFOCOM*, San Francisco, CA, March 1996.

[14] Arne Brøndsted. *An Introduction to Convex Polytopes*. Springer-Verlag, 1983.

[15] *reference omitted to preserve anonymity*

[16] E.G. Coffman and I. Mitrani. A Characterization of Waiting Time Performance Realizable by Single-Server Queues. *Operations Research*, 28(3):810–82, May-June 1980.

[17] W.W Sharkey. Cooperative Games with Large Cores. *International Journal of Game Theory*, 11:175–182, 1982.

[18] D.J.A. Welsh. *Matroid Theory*. Academic Press, 1976.