Achievable QoS in an Interference/Resource Limited Shared Wireless Channel

Jeffrey M. Capone, Member, IEEE, and Ioannis Stavrakakis, Senior Member, IEEE

Abstract --- In this work, the region of achievable quality-ofservice (QoS) is precisely described for a system of real-time heterogeneous variable bit rate (VBR) sources competing for slots (packet transmission times) of a time division multiple access (TDMA) frame. The QoS for each application is defined in terms of a maximum tolerable packet-dropping probability. Packets may be dropped due to delay violations and channel induced errors. The region of achievable QoS is precisely described for an interference/resource limited network by considering the underlying TDMA-multiple access control (TDMA-MAC) structure and the physical channel. A simple OoS-sensitive errorcontrol protocol that combats the effects of the wireless channel while satisfying the real-time requirements is proposed and its impact on the region of achievable QoS is evaluated. The results presented here clearly illustrate the negative impact of a poor channel and the positive impact of the employed error-control protocol on the achievable QoS. The region of achievable QoS vectors is central to the call admission problem, and in this work, it is used to identify a class of scheduling policies capable of delivering any achievable performance.

Index Terms—Achievable quality-of-service (QoS), integrated services, wireless resource.

I. INTRODUCTION

TN integrated services, wireless network transmission resources are shared among geographically dispersed applications with diverse traffic characteristics and quality-of-service (QoS) requirements. In this work, the shared transmission resources are defined to be the slots (packet transmission times) of a time division multiple access (TDMA) frame. This resource structure has been widely considered in personal communication networks (PCN) [1], [2], satellite networks [3], and wireless LAN's [4], as well as in recent work toward the development of wireless ATM networks [5].

In [5], channel access is based on a multiservice dynamic reservation (MDR)-TDMA frame format. At the beginning of each uplink MDR-TDMA frame, a slotted ALOHA control channel is employed for accommodating allocation requests, followed by a number of slots available to service the traffic. A typical uplink TDMA frame where slots are allocated to various classes of service [such as constant bit rate (CBR),

I. Stavrakakis is with the Communications and Digital Signal Processing (CDSP) Center, Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (ioannis@cdsp.neu.edu).

Publisher Item Identifier S 0733-8716(99)08244-X.

Request Channel C C V A V C V C V L C Assigned CBR slot V Slot available for VBR traffic A Slot available for ABR traffic

Fig. 1. A typical uplink TDMA frame structure supporting CBR, VBR, and ABR traffic classes.

variable bit rate (VBR), and available bit rate (ABR)] and service requests are processed at frame boundaries is illustrated in Fig. 1.

The objective is to enable sharing of the resources among diverse applications while delivering the required QoS. Typically, the services to be supported are designed to be compatible with the prevalent network architecture for integrated services over fiber/copper based channels, the asynchronous transfer mode (ATM). Some of the services offered by ATM to support the various traffic classes are CBR, VBR, and ABR. The focus of this work is on the support of real-time VBR service in a wireless network.

To support real-time VBR applications, the multiple access control (MAC) should employ a resource management scheme (transmission scheduling policy) that allocates transmission resources on demand, thereby effectively using the available bandwidth. Scheduling policies to support real-time VBR traffic have been proposed in [6]–[9]. The performance of these scheduling policies are evaluated for different combinations of heterogeneous VBR sources.

However, before a scheduling policy can be designed, it should be known whether or not the required performance can be delivered to the set of VBR sources through some scheduling policy. The main objective of this work is to determine the region of achievable QoS vectors for real-time heterogeneous VBR applications in a shared wireless environment. Knowledge of this region is central to the development of a call admission control mechanism. For example, if with the addition of the new source, the new multidimensional target QoS vector is in the region of achievable QoS vectors, then the call can be admitted. If the call cannot be admitted and more resources can be made available, a precisely defined region of achievable QoS can be used to determine the minimum additional resources required in order for the new call to be admitted. Furthermore, in this work, the established

Manuscript received January 6, 1999; revised June 14, 1999. This work was supported in part by the National Science Foundation under Grant NCR 9628116.

J. M. Capone is with the Telecommunications Research Center (TRC), Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287-7206 USA (e-mail: icapone@asu.edu).

region also leads to the identification of a class of scheduling policies capable of delivering any achievable performance.

In a wireless network, achievable QoS is shaped not only by the amount of available resources and the employed resources management scheme, but also by the channel quality (interference). Therefore, the region of achievable QoS is shaped collectively by the capabilities of the physical channel and the MAC protocol. For example, in an error-free channel, all transmitted packets are successfully received. In this case, the delivered QoS is shaped practically entirely by the packetdiscarding process at the transmitter (source) due to delay violations; the latter occurs when the demand exceeds the amount of available resources for a sufficiently long period. Thus, the performance is limited by the amount of available resources (resource limited). The region of achievable QoS vectors in an error-free channel environment has been investigated in [10]–[12]. Although the necessary resources may become available on time, packets may be corrupted due to channel errors and be dropped at the receiver. Under these conditions, the performance is limited by the interference introduced in the channel (interference limited). Such packet discarding may occur with a frequency comparable to that of the packet discards at the transmitter due to resource limitations. As a consequence, the region of achievable QoS vectors is shaped by the packet-discarding process at both the transmitter and the receiver due to resource and interference limitations, respectively.

In addition, in an integrated services wireless network, such as wireless ATM, systems may have error-controlling capabilities [13] to combat the unreliability of the wireless link. For the real-time VBR applications, traditional automatic repeat request (ARQ) strategies used to combat the effects of the wireless channel may not be possible due to the realtime constraints. In this work, a QoS-sensitive error-control protocol is designed to meet the real-time constraints of the supported applications, and its impact on the region of achievable is evaluated.

II. DESCRIPTION OF THE SYSTEM MODEL

Consider a system where N heterogeneous VBR sources compete for T slots of an uplink TDMA channel. At the beginning of each frame n, each source i requests a random number of slots denoted by $\lambda_i(n)$. If the aggregate demand in frame n, $\sum_{i=1}^{N} \lambda_i(n)$, exceeds the number of slots available to service the VBR traffic—referred to as an overloaded frame—then decisions must be made regarding the amount of service that will be provided to each source. The number of slots $a_i^f(n)$ under scheduling policy f allocated to source i may be less than what is required by that source $\lambda_i(n)$ due to resource limitation. Packets from a source which do not receive service over a frame following their arrival are considered to have excess delay and are dropped at the source.

The effects of the wireless channel are modeled as in [13], where a Gaussian noise channel with random bit erasure interference is considered. Bit erasures may be correlated due to a slow fading channel (or a jamming signal), but in this work, it is assumed that the time scale of the fade is smaller than the packet duration. As a result, packet erasures are considered to be statistically independent and occur when the interference in the channel is such that the packet is corrupted beyond correction. The corrupted transmitted packets are discarded (dropped) at the receiver.

The event that the packet is corrupted and therefore dropped is a function of the interference in the channel, the transmitted power, the modulation and coding scheme, and the packet length. Let Z be an indicator function of a packet erasure. That is

$$Z = \begin{cases} 1, & \text{if the packet is corrupted} \\ 0, & \text{otherwise.} \end{cases}$$
(1)

Therefore, the expected value of Z, that is $E[Z] = \beta$, is the probability that a packet is corrupted by the channel and dropped.

Considering the combined impact of the scheduling policy f and the physical channel, the number of packets from source i dropped in frame n due to the competition for the resources and the interference in the channel is given by

$$d_{i}^{f}(n) = \begin{cases} \sum_{m=1}^{\lambda_{i}(n)} Z_{m} & \text{if } \sum_{j=1}^{N} \lambda_{j}(n) \leq T\\ \lambda_{i}(n) - a_{i}^{f}(n) + \sum_{m=1}^{a_{i}^{f}(n)} Z_{m} & \text{if } \sum_{j=1}^{N} \lambda_{j}(n) > T. \end{cases}$$
(2)

 Z_m is an indicator function associated with the transmission of the *m*th packet from source *i*. Considering the effects of the wireless channel, it must be determined whether *under given channel conditions* a QoS vector is achievable under some policy *f*.

Suppose that the QoS requirement of application i is defined in terms of a maximum tolerable average per frame packetdropping rate $d_i, 1 \leq i \leq N$. The QoS vector associated with the application then can be defined in terms of the (performance) packet-dropping rate vector $\mathbf{d} = (d_1, d_2, \dots, d_N)$. When the QoS requirement of the application i is defined in terms of a maximum tolerable packet-dropping probability p_i , the corresponding packet-dropping rate d_i is easily determined by $d_i = \lambda_i p_i$.

The first question addressed in Section III-A is whether (under the given channel conditions) a given QoS vector d is achievable under some policy f. The results from Section III-A are modified to reflect the impact of the proposed QoS-sensitive error-control scheme and are presented in Section III-B. The second question, addressed in Section V, is concerned with the design of scheduling policies that deliver an achievable target QoS vector d.

III. DETERMINATION OF THE REGION OF ACHIEVABLE QoS VECTORS

The establishment of the region of achievable QoS vectors is based on a set of inequalities and an equality constraint derived by employing work-conserving arguments. The superscript fis used to denote the employed packet scheduling policy. It is assumed that the scheduling policies are work-conserving¹ (that is, nonidling) and induce a performance vector d^{f} .

A. Achievable QoS Provided by the Underlying MAC and Physical Channel

Let $S = \{1, 2, ..., N\}$ be the set of all sources and d_S^f denote the average system packet-dropping rate under scheduling policy f denoted by

$$d_{S}^{f} \stackrel{\Delta}{=} E\left[\sum_{i=1}^{N} d_{i}^{f}(n)\right] = \sum_{i=1}^{N} E[d_{i}^{f}(n)] = \sum_{i=1}^{N} d_{i}^{f}.$$
 (3)

Let $\lambda_g(n)$ denote the aggregate arrival rate from sources in set $g, g \subseteq S$, i.e., $\lambda_g(n) = \sum_{i \in g} \lambda_i(n)$.

Summing (2) over all sources $i \in S$ and by considering the expected value of the associated quantity, the average system packet-dropping rate under work-conserving scheduling policy f is derived and it is given by

$$d_{S}^{I} = \{E[\lambda_{S}(n)|\lambda_{S}(n) > T] - T(1 - \beta)\}P(\lambda_{S}(n) > T) + \beta\{E[\lambda_{S}(n)|\lambda_{S}(n) \le T]\}P(\lambda_{S}(n) \le T).$$
(4)

As it can be seen from (4), d_S^f is independent from the policy f; it only depends on the aggregate arrival process, the number of resources T, and the channel characteristics β . Therefore, the system dropping rate d_S^f is conserved under any work-conserving policies f and is denoted as b_S .

Let d_g^f denote the average subsystem g packet-dropping rate under policy f defined by $d_g^f \triangleq E[\sum_{i \in g} d_i^f(n)] =$ $\sum_{i \in g} E[d_i^f(n)] = \sum_{i \in g} d_i^f, g \subset S$. That is, d_g^f is equal to 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 policy f.

Let b_g denote the lower bound on the aggregate packetdropping rate for sources in g. This bound is equal to the packet-dropping rate of a system in which only sources in g are present and served under a work-conserving policy; sources in set $\{S - g\}$ are considered to be removed. It is given by

$$b_g = \{E[\lambda_g(n)|\lambda_g(n) > T] - T(1 - \beta)\}P(\lambda_g(n) > T) + \beta\{E[\lambda_g(n)|\lambda_g(n) \le T]\}P(\lambda_g(n) \le T).$$
(5)

It is apparent that no policy can deliver a lower dropping rate than b_g to sources in set g when all sources in S are present. It can be seen that this lower bound is attained by all policies f which give service priority to packets from sources in set g over those in the complement set $\{S - g\}$.

It has been shown in [10] and [11] that in an error-free channel, the following conditions

$$d_g \ge b_g \quad \forall g \subseteq S \tag{6}$$

$$d_S = b_S \tag{7}$$

are necessary and sufficient in order for a QoS vector $d = (d_1, d_2, \ldots, d_N)$ to be achieved by some scheduling policy f.

¹The policies are work-conserving assuming that the packets arrive on frame boundaries.

This result can be extended to account for the channel quality provided that b_g [given in (5)] is a supermodular set function; the proof is provided in Section VIII.

Let \mathcal{D} denote the collection of all vectors **d** satisfying (6) and (7); then, by definition, \mathcal{D} is a convex polytope [14]. Using results from convex polytopes [14], any vector in the set \mathcal{D} can be expressed as a convex combination of extreme points (vertices) of \mathcal{D} ; that is, \mathcal{D} may be expressed as the convex hull of its extreme points, $\mathcal{D} = \operatorname{conv}[\exp(\mathcal{D})]$.

In addition, from the polytope structure and the supermodularity property of the set function b_g , it can be shown (see [10]) that d^* is a vertex of the set \mathcal{D} iff d^* is a dropping rate vector resulting from an ordered head-of-line (O-HoL) priority service policy $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_N); \pi_i \in$ $\{1, 2, \dots, N\}, \pi_i \neq = \pi_j, 1 \leq i, j \leq N$. The index of π_i indicates the order of the priority given to the π_i source. None of the π_j sources, j > i, may be served as long as packets from sources $\pi_k, k \leq i$, are present.

Fig. 2 provides a graphical illustration of the region \mathcal{D} for the case of N = 2 and N = 3 sources. The extreme points correspond to QoS vectors **d** induced by the N! O-HoL priority policies $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_N)$.

The region of achievable QoS is shaped by both the amount of available resources and the level of interference in the wireless channel. To illustrate this, consider the following example of two sources competing for T slots in a TDMA frame. The source packet arrival processes are assumed to be mutually independent. Each arrival process is embedded at the frame boundaries. The number of packets generated (and requesting service) by a source in the current frame boundary is (probabilistically) determined by the present state of the underlying arrival process.

In this example, each VBR source is modeled by a sequence of independent and identically distributed (i.i.d.) random variables, embedded at frame boundaries, with mean rate of 3.6 and 3.2 packets/frame, and variance of 2.04 and 3.36 packets/frame, respectively. In Fig. 3, the conserved system packet-dropping probability $p_S = (b_S/E[\lambda_S(n)])$ is plotted as a function of available resources T (time slots) for an errorfree channel ($\beta = 0$) and an error-prone wireless channel with channel quality $\beta = 0.02$.

As it can be seen in this figure, there are three distinct regions of operations: 1) resource-limited; 2) interference/resource-limited; and 3) interference-limited. In the resource-limited region, the performance is primarily determined by the amount of available resources. This result is evident since the packet-dropping probability for the system with the error-free channel and the (nonideal) wireless channel are almost identical. In the interference-limited region, the dropping probability in the error-free channel is zero, while the performance in the wireless system is limited by the interference and given by $\beta = 0.02$. The performance in the interference/resource-limited region is determined by both, the available resources and the level of interference in the channel. In this example, the system packet-dropping probability in this region ranges from 10^{-1} to 10^{-2} , an operation region of interest for real-time applications. It is important to note that, in general, satisfying the system



Fig. 2. The region (polytope) \mathcal{D} for a system with two and three sources under channel conditions β .

packet-dropping rate (probability) is only necessary and not sufficient to guarantee that the target QoS vector is achievable.

B. Effects of a QoS-Sensitive Error-Control Scheme on the Region of Achievable QoS

Many error-control architectures have been proposed to enable wireless ATM [13], [15]. The primary focus of these works is to design an FEC code and an ATM-HEC producing a concatenated code which protects the ATM cell header against undetected errors; undetected errors in an ATM cell header can lead to misrouting and other impairments. In [13], the architecture also involves a traditional data link ARQ protocol applied only to traffic that is not sensitive to increased delays. Recently, an ARQ scheme has been developed to accommodate real-time service [16]. The result is achieved by scheduling the feedback (ACK and NAK) based on the relative urgency of the transmitted packets; in addition, if the due date of the packet has expired, the packet is discarded, and the receiver is notified. In this manner, the number of retransmissions of each packet is controlled.

In this work, the probability of correct reception of a packet is increased (or the probability of dropping is reduced) by transmitting multiple copies of certain packets (before the packet has expired) without any feedback. In this way, the probability of packet-dropping can be controlled by the scheduling policy which controls the number of copies transmitted from each source based on the QoS required by the applications. The focus of this section is to examine the impact of the wireless channel and a simple QoS-sensitive forward error-control scheme on the region of achievable QoS.

The error-control scheme considered in this work will generate multiple copies² of certain packets for transmission over the current frame. Copies are transmitted only during underloaded frames utilizing the remaining resources (T - T) $\lambda_{S}(n)$). Transmitting a copy from a set q during an overloaded frame would reduce the probability of packet-dropping at the receiver for the set g, but would force an original packet from the complement set $\{S - g\}$ to be dropped at the source. The effect would be that the aggregate dropping rate for the subset g of sources is reduced, but the aggregate dropping rate for the complement set $\{S - g\}$ is increased by an amount greater (in realistic systems) than the decrease attained for the subset q, causing the overall system dropping rate to increase. In view of the previous discussion, if the objective of the error-control protocol is to minimize the system packet-dropping probability (or equivalently packet-dropping rate) and therefore maximize system throughput, then multiple copies of packets can be sent only during underloaded frames-utilizing the remaining resources.

During underloaded frames, the number of copies generated by the sender is a function of the scheduling policy and the amount of remaining resources. Therefore, the number of packets dropped due to channel induced errors is also a function of the scheduling policy. For a given $\lambda_S(n) \leq T$, let $\mathbf{k}^f = (k_1^f, k_2^f, \dots, k_{\lambda_S(n)}^f), 1 \leq k_m^f \leq T, 1 \leq m \leq \lambda_S(n)$, be a vector which determines the number of transmissions of each packet $m \in \{1, 2, \dots, \lambda_S(n)\}$. Notice that $k_m^f \geq 1$, which guarantees that no original packet is dropped at the expense of transmitting a copy, and consequently, only the remaining resources can be used to transmit copied packets. Under this policy, f the expected number of packets arriving at the receiver in error, is equal to

$$\sum_{m=1}^{\lambda_S(n)} \beta^{k_m^f} \tag{8}$$

where

$$\sum_{m=1}^{\lambda_S(n)} k_m^f = T$$

$$1 \le k_m^f \le T, \quad \forall m \in \{1, 2, \dots, \lambda_S(n)\}.$$
(9)

²Corrupted copies are not used for error correction.



Fig. 3. System packet-dropping probability for two source system given in Section III-A, in an error-free channel and a (nonideal) error-prone wireless channel with channel conditions $\beta = 0.02$.

1) Unbiased Error Control: An unbiased error-control protocol, or equivalently, a scheme that employs a policy that attempts to "fairly" allocate the remaining resources among sources, will result in the minimum system dropping rate. This result is a direct consequence of the convexity of the function β^x in x. That is, $\gamma\beta^a + (1-\gamma)\beta^b \ge \beta^{\gamma a + (1-\gamma)b}$ for $0 \le \gamma \le 1$; therefore

$$\sum_{m=1}^{\lambda_S(n)} \beta^{k_m^f} \ge \lambda_S(n) \beta^{T/\lambda_S(n)}.$$
 (10)

Since (8) is minimum when $k_m^f = (T/\lambda_S(n) = k, (8))$ is minimized when each sources *i* receives $\lambda_i(n)k = (\lambda_i(n)/\lambda_S(n))T$ slots for transmission of the $\lambda_i(n)$ packets and its copies. In this sense the policy is "fair" or unbiased.

However, due to the granularity in the system (that is, resources can be allocated only in integer multiples), the following is considered. During underloaded frames, let each packet be transmitted $R \triangleq \lfloor T/\lambda_S(n) \rfloor$, $R \ge 1$, number of times, where $\lfloor \cdot \rfloor$ denotes the integer part, and let $X \triangleq (T - \lambda_S(n)R)$ be the number of packets that can be transmitted one additional time, (R + 1). The system dropping rate is conserved regardless of the set of sources to which the packets that are transmitted *one additional time* belong. Allocating the additional retransmission during underloaded frames according to a policy allows for further diversification of the resulting QoS vector delivered, while still satisfying the requirement of minimum system dropping rate. Therefore, under any fair work-conserving policy, the system dropping rate (given to follow) is minimum and also conserved.

The results from Section III-A can be modified to account for the impact of the QoS-sensitive error-control protocol. Considering the effects of the physical channel, the scheduling policy, and the error-control protocol, the number of packets from source i dropped in frame n is given by

$$d_{i}^{f}(n) = \begin{cases} \sum_{m=1}^{\lambda_{i}(n)} \prod_{q=1}^{R+1_{m}^{f}} Z_{m}^{q}, & \text{if } \lambda_{S}(n) \leq T \\ \lambda_{i}(n) - a_{i}^{f}(n) + \sum_{m=1}^{a_{i}^{f}(n)} Z_{m}, & \text{if } \lambda_{S}(n) > T \\ 0 \leq i \leq N. \end{cases}$$
(11)

 $1_m^f \in \{0,1\}$ and indicates the dependency of the additional transmission of a copy of packet m on the fair policy f, where $\sum_{m=1}^{\lambda_S(n)} 1_m^f = X, \forall f. Z_m^q$ is an indicator function associated with the qth transmission of the mth packet from source i.

The system dropping rate b_S is conserved and is found by summing (11) over all i in S

$$\sum_{i \in S} d_i^f(n) = \begin{cases} \sum_{m=1}^{\lambda_S(n)} \prod_{q=1}^{R+1_m^f} Z_m^q, & \text{if } \lambda_S(n) \le T \\ \lambda_S(n) - T + \sum_{m=1}^T Z_m, & \text{if } \lambda_S(n) > T. \end{cases}$$
(12)

Taking the expectation of (12) conditioned on $\lambda_S(n)$

$$E\left[\sum_{i\in S} d_i^f(n)|\lambda_S(n)\right]$$

=
$$\begin{cases} \beta^R(\beta^{1_1^f} + \beta^{1_2^f} + \cdots \beta^{1_{\lambda_S(n)}^f}) & \text{if } \lambda_S(n) \le T\\ \lambda_S(n) - T + T\beta & \text{if } \lambda_S(n) > T. \end{cases}$$
(13)



Fig. 4. The impact of the QoS-sensitive error-control protocol on the system packet-dropping probability in a wireless channel with channel conditions $\beta = 0.02$.

Since $1_m^f \in \{0,1\}$ such that $\sum_{m=1}^{\lambda_S(n)} 1_m^f = X, \forall f$, then (13) set g. b_g is derived and it is given by becomes

$$E\left[\sum_{i\in S} d_i^f(n)|\lambda_S(n)\right]$$

=
$$\begin{cases} \beta^R(X\beta + \{\lambda_S(n) - X\}), & \text{if } \lambda_S(n) \le T\\ \lambda_S(n) - T(1 - \beta), & \text{if } \lambda_S(n) > T \end{cases}$$
(14)

and taking the expectation of (14) over $\lambda_S(n)$, the system dropping rate is derived, and it is given by

$$b_{S} \stackrel{\Delta}{=} \{E[\lambda_{S}(n)|\lambda_{S}(n) > T] - T(1 - \beta)\}P(\lambda_{S}(n) > T) + E[X\beta^{R+1}|\lambda_{S}(n) \le T]P(\lambda_{S}(n) \le T) + E[(\lambda_{S}(n) - X)\beta^{R}|\lambda_{S}(n) \le T]P(\lambda_{S}(n) \le T).$$
(15)

As it is seen in (15), the system dropping rate is not dependent on the policy f, and thus, it is conserved for all fair workconserving policies f.

With the addition of the unbiased error-control scheme, the lower bound b_g for the aggregate packet-dropping rate for sources in g under any fair work-conserving policy fcan be derived. It is determined to be equal to the packetdropping rate of a system in which only sources in g are present and served under a work-conserving policy, and all additional retransmission privileges are provided to set g. That is, $1_m^f = 1$ (if $\lambda_g(n) \leq X$) for all packets m from sources in

$$b_{g} \stackrel{\Delta}{=} \{E[\lambda_{g}(n)|\lambda_{g}(n) > T] - T(1-\beta)\}P(\lambda_{g}(n) > T) \\ + E[\min[\lambda_{g}(n), X]\beta^{R+1}|\lambda_{g}(n) \leq T]P(\lambda_{g}(n) \leq T) \\ + E[\max[0, \lambda_{g}(n) - X]\beta^{R}|\lambda_{g}(n) \leq T]P(\lambda_{g}(n) \leq T).$$
(16)

The expected value in (16) is with respect to $\{\lambda_a(n), \lambda_S(n)\},\$ which is easily computed since $P(\lambda_g(n) = i, \lambda_S(n) = j) =$ $P(\lambda_g(n) = i)P(\lambda_{\{S=g\}}(n) = j - i)$. The expressions in (15) and (16) reduce to (4) and (5) respectively, for R fixed and equal to one. The extreme points in this case correspond to fair-ordered HoL (F-O-HoL) priority service policies. An F-O-HoL service policy is an O-HoL service policy π = $(\pi_1, \pi_2, \ldots, \pi_N)$ in which the additional retransmissions are also allocated according to π .

With the addition of the QoS-sensitive error-control protocol, the region of achievable QoS for an interference/resources or interference-limited system can be improved compared to the system not employing the error-control protocol. The impact that the error-control protocol has on the system packet-dropping probability for the example described in Section III-A is shown in Fig. 4.

In this figure, the performance of the two systems is compared to that in an error-free environment. As it can be seen in this figure, the system employing the error-control protocol induces a lower packet-dropping probability than the system without. The impact is most significant in the interference/resource-limited and interference-limited regions.



Fig. 5. The impact of the error-control protocol on D for the two source system given in Section III-A, in an error-free channel, in an error-prone wireless channel ($\beta = 0.02$), with and without the QoS-sensitive error control.

 TABLE I

 IMPACT OF CHANNEL QUALITY ON THE REGION OF ACHIEVABLE QOS

Nur	nerical R	esults	
	<i>b</i> {1}	$b_{\{2\}}$	$b_{\{1,2\}}$
Error-Free	0.0	0.0	0.08
Error-Prone	0.0720	0.0640	0.2144
Error Control	0.0303	0.0316	0.1606

TABLE II COMPARISON OF LOWER BOUNDS OBTAINED UNDER AN UNBIASED AND BIASED ERROR CONTROL

Numerical Results				
	$b_{\{1\}}$	$b_{\{2\}}$		
Unbiased	0.0303	0.0316		
Biased	0.0043	0.0086		

In these regions, the system with the error-control protocol takes advantage of the remaining resources and can reduce the packet-dropping probability.

The impact that the QoS-sensitive error control has on the region of achievable QoS \mathcal{D} is illustrated in Fig. 5 and Table I. In this figure, \mathcal{D} is derived for the system of sources given in Section III-A and with T = 10 slots. Thus, according to Fig. 3, the system is interference/resource-limited. As it can be seen, this simple forward error control (FEC) protocol moves the region of achievable QoS vectors toward lower dropping rates. This implies that a larger collection of QoS vectors can be accommodated, as explained in Section IV.

2) Biased Error Control: In Section III-B1, the region of achievable QoS was determined under the constraint that the system packet-dropping rate be minimized. This constraint results in throughput maximization, but imposes restrictions on the level of QoS diversification that could be achieved otherwise. If the QoS vector is not in the region of achievable QoS vectors under this restriction, it can be concluded that such level of QoS diversification may be achieved only at the expense of system throughput [17]. This result 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. This constraint also leads to the development of a fair or unbiased error-control protocol, as described in the previous section.

In this section, the effect of a biased error-control protocol is examined. A biased error-control protocol allocates the remaining resources, $(T - \lambda_S(n))$, to favor a certain subset q. As a consequence, (8) is no longer minimized, and the system dropping rate is increased. In addition, the system dropping rate would depend on which subset is favored, and thus, the system dropping is dependent on the policy and no longer conserved. Although the conservation is lost, the lower bounds for a subset q achievable under any biased error-control protocol can be evaluated. This lower bound b_g^{biased} under a biased scheme can be obtained by setting $k_m^f = 1$ in (8) for all packets m from sources in set $\{S - q\}$, guaranteeing that the remaining resources, $(T - \lambda_S(n))$, are used only for transmitting copies from set g. Under this constraint, the lower bound for the subset q is obtained when $R \stackrel{\Delta}{=} \lfloor T - \lambda_{\{S-g\}}(n) / \lambda_g(n) \rfloor \text{ and } X \stackrel{\Delta}{=} \{T - \lambda_g(n)R\}.$

For the previous example described in Section III-A, the lower bounds are evaluated and compared to the lower bounds obtained with the unbiased error-control scheme, and the results are displayed in Table II. Since the system dropping rate under any biased errorcontrol protocol is increased compared to the unbiased system dropping rate (denoted in this section as b_S^{unbiased}), the following conditions are necessary for a QoS vector d to be achieved under any (biased or unbiased) scheme

$$d_g \ge b_g^{\text{biased}} \quad \forall g \subseteq S$$
$$d_S \ge b_S^{\text{unbiased}}.$$
 (17)

The constraints presented in (17) are only necessary and not sufficient to guarantee that the target QoS vector d is achievable; thus, they determine only an upper bound on the region of achievable QoS vectors. However, this region can be used to determine if a QoS vector with such diversity is *not* achievable by any (biased or unbiased) error-control scheme.

IV. CALL ADMISSION REGION

The call admission region $\mathcal{A}(\mathcal{D})$ associated with the region \mathcal{D} of achievable QoS vectors is defined to be the region of vectors *d* satisfying

$$d_g \ge b_g \quad \forall g \subset S \tag{18}$$

$$d_S \ge b_S. \tag{19}$$

Thus, $\mathcal{A}(\mathcal{D})$ also expands with the addition of the QoS-sensitive error-control protocol.

It is show in [10] that if $d \in \mathcal{A}(\mathcal{D})$, then there exists a vector $d' \in \mathcal{D}$ which is such that $d'_i \leq d_i \quad \forall i \in S$. This result implies that if the target QoS vector d is in $\mathcal{A}(\mathcal{D})$, there exists a policy which can deliver d or better (that is, less than the dropping rate required by any source).

Finally, as stated in Section III-A, satisfying the condition on the system performance, given here by (19), is only necessary and not sufficient to guarantee that the target QoS vector is achievable.³ To illustrate this concept, consider the network given in Section III-A with T = 10 and $\beta = 0.02$ and let d = (0.1291, 0.0315) be the target QoS vector (source 1 and 2 can tolerate a packet-dropping probability of 0.035 and 0.010, respectively). According to Table I, d satisfies the condition on the system dropping rate for the system that employs the real-time unbiased scheme, that is $0.1606 = d_S \ge b_S^{\text{imbiased}} = 0.1606$, but $d_2 < b_2^{\text{imbiased}} =$ 0.0316. Thus, the target QoS vector cannot be achieved under any f with an unbiased error-control scheme. For this system, the best possible performance source 2 can achieve is if it is given service priority, resulting in dropping rates of 0.1290 and 0.0316 for sources 1 and 2, respectively. The overall system performance is satisfied, but source 2 is provided poorer than desired service, while source 1 is provided improved service. Furthermore, the QoS vector d =(0.1291, 0.0315) is not achievable under any biased scheme, since $b_S^{\text{biased}} > b_S^{\text{unbiased}} = 0.1606 = d_S$, and the system dropping rate can not be satisfied.

V. A CLASS OF POLICIES DELIVERY ANY ACHIEVABLE QoS VECTOR IN \mathcal{D}

Once a call is given admission into the system, management of the available transmission resources is necessary to ensure that the required QoS is delivered to each source application. In addition to determining the region of achievable QoS vectors, this study also leads to the development of scheduling policies that deliver any achievable performance for a given amount of available resources and channel quality.

A. Mixing Policies

Let C_{HoL} denote the class of F-O-HoL priority service policies π introduced in Section III-B1. A mixing F-O-HoL priority service policy f_m is defined to be one that at each frame decides to follow the F-O-HoL priority policy $\pi^i =$ $(\pi_1^i, \pi_2^i, \ldots, \pi_N^i)$ with probability $\alpha_i, \alpha_i \geq 0, 1 \leq i \leq$ $N!, \sum_{i=1}^{N!} \alpha_i = 1$; decisions over consecutive frames are independent. Clearly, f_m is completely determined by the N!dimensional vector $\alpha, \alpha \geq 0, 1 \cdot \alpha = 1$. Let M_{HoL} denote the class of such policies.

For each packet-dropping rate vector $\boldsymbol{d} \in \mathcal{D}$, there exists a policy $f_m \in M_{\text{HoL}}$ that induces \boldsymbol{d} . The proof follows from the fact that \mathcal{D} can be written as a convex combination of the extreme points (vertices) $\boldsymbol{d}_{\text{ext}-i}$ of \mathcal{D} ; that is, $\boldsymbol{d} =$ $\sum_{i=1}^{N!} \alpha_i \boldsymbol{d}_{\text{ext}-i}$ for some $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_{N!})$ where $\alpha_i \geq$ $0, 1 \leq i \leq N!, \sum_{i=1}^{N!} \alpha_i = 1$. Since each $\boldsymbol{d}_{\text{ext}-i}$ is induced by some policy in C_{HoL} , a mixing policy f_m that selects the F-O-HoL priority policy $\boldsymbol{\pi}^i$ (that induces $\boldsymbol{d}_{\text{ext}-i}$) with probability α_i such that $\alpha_i \geq 0, 1 \leq i \leq N!, \sum_{i=1}^{N!} \alpha_i = 1$ will have a packet-dropping rate vector \boldsymbol{d}^{f_m} given by

$$\boldsymbol{d}^{f_m} = \sum_{i=1}^{N!} \alpha_i \boldsymbol{d}_{\text{ext-}i}$$
(20)

and thus f_m induces **d**.

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

$$E_{\boldsymbol{\alpha}}[\boldsymbol{d}_{\text{ext}}] \stackrel{\Delta}{=} \sum_{i=1}^{N!} \alpha_i \boldsymbol{d}_{\text{ext}-i} = \boldsymbol{d}$$
(21)

$$\alpha \ge 0$$
 (22)

$$\mathbf{1} \cdot \boldsymbol{\alpha} = 1 \tag{23}$$

where $E_{\alpha}[\cdot]$ is weighted average of the set of extreme points d_{ext} of \mathcal{D} with respect to the probability mass function α .

B. Priority Indexing Policy

The priority indexing policy⁴ uses the past history of the system to form performance indexes that are uses for resource management. The performance index is a relative measure of the current performance to the desired performance. The sources are ordered according to the performance indexes and the corresponding F-O-HoL priority service policy π is selected. A priority indexing policy uses the system's past

 $^{^{3}}$ In the special case of a homogeneous system, such as a cellular voice system, satisfying the system constraint in (19) is sufficient to guarantee that the target QoS vector is achievable. This result has been established in [10].

⁴This policy is similar to the policy used in [18] to manage the buffer resources.

history to select an F-O-HoL priority policy. In every frame, each source is assigned an index as follows

$$I_i(n) = \frac{\hat{d}_i(n)}{d_i} \tag{24}$$

where

$$\widehat{d_i(n)} = \frac{1}{n} \sum_{k=1}^{n} d_i(k).$$
(25)

 $d_i(n)$ is an estimate for the mean dropping rate for source *i*. In each frame, service priority is given to source *j* over source *i*, if and only if

$$I_j(n) > I_i(n). \tag{26}$$

All priority indexes are ordered according to (26), and the appropriate priority policy is selected. A comparison in performance between the two classes of policies is examined in the following section.

VI. A NUMERICAL EXAMPLE

In this section, an example is presented to demonstrate the call admission procedure and the design of a scheduling policy to deliver the target QoS vector. Consider the network presented in Section III, where the source packet arrival processes are assumed to be mutually independent, and each arrival process is described in terms of a sequence of i.i.d. random variables embedded at the frame boundaries. It is assumed that there are T = 10 slots available to service the two VBR sources and the channel quality is given by $\beta = 0.02$.

Let p_k denote the target packet-dropping probability, for source k, k = 1, 2. In this example, it is assumed that $p_1 = 0.020$ and $p_2 = 0.028$. By multiplying these probabilities by the corresponding per frame arrival rate, the per frame packet-dropping rates $d_1 = 0.071$ and $d_2 = 0.0896$ are obtained. Thus, the target QoS vector d is given by d = $(d_1, d_2) = (0.071, 0.0896)$ and the associated system packetdropping rate is equal to $d_S = d_1 + d_2 = 0.1606$. Considering the inequalities in (18) and (19) (evaluated in Table I), it is determined that d is achievable.

Since $d \in \mathcal{A}(\mathcal{D})$ and particularly $d \in \mathcal{D}$, there exists a mixing F-O-HoL priority policy $f_m \equiv \alpha$ achieving exactly the QoS vector d. Any α satisfying the conditions (21)–(23) may be chosen. In this example, there exists only one solution to (21)–(23), since N = 2, and it is found to be

$$\boldsymbol{\alpha_o} = \begin{bmatrix} \alpha_1 = 0.5876\\ \alpha_2 = 0.4124 \end{bmatrix}.$$
 (27)

In systems with greater than two sources (N > 2), 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 delivery of the target QoS vector. For instance, among all mixing policies inducing **d**, the one which minimizes the variance of the service provided to certain sources may be identified.

A simulation was performed using the policy α_o derived for this example. The sources, as described in Section III-A, were

TABLE III Comparison of Theoretical and Time-Averaged Steady-State Mean Packet-Dropping Rate for Each Source Under Each Policy

Source, i	Theoretical, d_i	Mixing Policy	Indexing Policy
1	0.0710	0.0654	0.0740
2	0.0896	0.0880	0.0890

simulated and generated the service demand, and the packet erasures occurred with probability 0.02. Resources (slots) were allocated according to the prescribed policy, and the number of dropped packets/frame from each source was recorded. The simulation time was 3000 frames. Throughout the simulation in each frame n, the time-averaged mean dropping rate was calculated for each source i as

$$\widehat{d}_i(n) = \frac{1}{n} \sum_{k=1}^n d_i^f(k).$$
 (28)

The resulting time-averaged mean packet-dropping rate for each source under the policy was calculated and is displayed in Table III. Throughout the progression of the simulation the time-averaged mean packet-dropping rates were recorded and are displayed in Fig. 6. Under the policy, the objective of mean packet-dropping rate d_i for each source *i* is met (see Table III and Fig. 6).

Results from the priority indexing policy are also displayed in Table III and Fig. 7. As it can be seen in Fig. 7, the priority indexing policy attempts to control the variation in the dropping rates by using the history of the past performance.

Although both the mixing policy and the priority indexing policy deliver the desired dropping rates, the priority indexing policy has less variability in the delivered service, as is seen in Fig. 7.

VII. CONCLUSION

In this work, the region of achievable QoS has been precisely described for a system of real-time heterogeneous VBR sources competing for an unreliable wireless channel. The QoS has been defined in terms of a packet-dropping probability (or equivalently packet-dropping rate). Packets from sources in the system are dropped as a result of delay violations and channel induced errors. As a consequence, it has been shown that the region of achievable QoS is shaped by both the interference in the physical channel and the amount of available resources. In addition, a simple QoS-sensitive errorcontrol protocol has been proposed to combat the effects of the wireless channel while still satisfying the real-time service constraints of the associated applications. The results presented in this paper illustrate the positive impact of the employed simple error-control protocol on the region of achievable QoS.

APPENDIX

Lemma 8.1: b_A [as defined in (5)] is a supermodular set function; that is

$$b_A + b_B \le b_{\{A \cup B\}} + b_{\{A \cap B\}} \tag{29}$$

and equality hold only if $A \subset B$ or $B \subset A$.



Fig. 6. Simulation results obtained with the derived mixing policy α_o .



Fig. 7. Simulation results obtained with priority indexing policy.

Proof: Let λ_A (λ_B) be the generic random variables representing the number of packets generated by sources in

set A (B) over a frame $A, B, \subseteq S$. Clearly $\lambda_A + \lambda_B = \lambda_{\{A \cup B\}} + \lambda_{\{A \cap B\}} \quad \forall A, B \subseteq S.$ (30) As given in (5)

$$b_A = \beta \{ E[\lambda_A | \lambda_A \le T] \} P(\lambda_A \le T)$$

+ $\{ E[\lambda_A | \lambda_A > T] - T(1 - \beta) \} P(\lambda_A > T).$ (31)

Writing $E[\lambda_A|\lambda_A \leq T]P(\lambda_A \leq T)$ as $E[\lambda_A] - E[\lambda_A|\lambda_A > T]P(\lambda_A > T)$ and expanding the second term in (31), the following is obtained:

$$b_A = \beta(E[\lambda_A] - E[\lambda_A|\lambda_A > T]P(\lambda_A > T)) + \{E[\lambda_A|\lambda_A > T] - T\}P(\lambda_A > T).$$
(32)

Equation (32) can be written in terms of the error-free channel given in [10], and it is denoted in this section as b_A^{ef}

$$b_A = (E[\lambda_A] - b_A^{ef})\beta + b_A^{ef}.$$
(33)

Employing the result in (33) and the result from [10] that $b_A^{ef} + b_B^{ef} \le b_{A\cup B}^{ef} + b_{A\cap B}^{ef}$

$$b_{A} + b_{B} = \left(E[\lambda_{A}] - b_{A}^{ef}\right)\beta + b_{A}^{ef} + \left(E[\lambda_{B}] - b_{B}^{ef}\right)\beta + b_{B}^{ef}$$
$$= (E[\lambda_{A}] + E[\lambda_{B}])\beta + (1 - \beta)\left(b_{A}^{ef} + b_{B}^{ef}\right)$$
$$\leq (E[\lambda_{\{A\cup B\}}] + E[\lambda_{\{A\cap B\}}])\beta$$
$$+ (1 - \beta)\left(b_{\{A\cup B\}}^{ef} + b_{\{A\cap B\}}^{ef}\right)$$
$$= b_{\{A\cup B\}} + b_{\{A\cap B\}}$$
(34)

completing the proof of Lemma 8.1.

REFERENCES

- 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 J. Select. Areas Commun.*, vol. 11, pp. 870–884, Aug. 1993.
- [2] D. J. Goodman, R. A. Valenzeula *et al.*, "Packet reservation multiple access for local wireless communication," *IEEE Trans. Commun.*, vol. 37, pp. 885–890, Aug. 1989.
- [3] G. Falk et al., "Integration of voice and data in the wideband packet satellite network," *IEEE J. Select. Areas Commun.*, vol. SAC-1, pp. 1076–1083, Dec. 1983.
- [4] D. Bantz and F. Baucho, "Wireless LAN design alternatives," *IEEE Network Mag.*, vol. 8, pp. 43–34, Mar./Apr. 1994.
 [5] D. Raychaudhuri *et al.*, "WATMnet: A prototype wireless ATM sys-
- [5] D. Raychaudhuri *et al.*, "WATMnet: A prototype wireless ATM system for multimedia personal communications," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 83–95, Jan. 1997.
- [6] S. Biswas, D. Reininger, and D. Raychaudhuri, "User parameter control based bandwidth allocation for VBR video in wireless ATM links," in *Proc. IEEE ICC'97*, Phoenix, AZ, pp. 1073–1079.
- [7] B. Walke, D. Petras *et al.*, "Wireless ATM: Air interface and network protocols of the mobile broadband system," *IEEE Personal Commun. Mag.*, vol. 3, p. 50–56, Aug. 1996.
- [8] G. Anastasi, D. Grillo, and L. Lenzini, "An access protocol for speech/data/video integration in TDMA-based advanced mobile systems," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1498–1509, Oct. 1997.
- [9] C. Chang, K. Chen, M. You, and J. Chang, "Guaranteed qualityof-service wireless access to ATM networks," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 106–118, Jan. 1997.

- [10] J. Capone and I. Stavrakakis, "Achievable QoS and scheduling policies in integrated services wireless networks," *Performance Evaluation*, vols. 26 and 27, no. 1, pp. 347–365, Oct. 1996.
- [11] _____, "Delivering diverse delay/dropping QoS requirements in a TDMA environment," in *Proc. ACM MobiCom*, Budapest, Hungary, Sept. 1997, pp. 110–119.
- [12] _____, "Determining the call admission region for real-time heterogeneous applications in wireless TDMA networks," *IEEE Networks*, vol. 12, pp. 38–47, Mar./Apr. 1998.
- [13] J. B. Cain and D. McGregor, "A recommended error control architecture for ATM networks with wireless links," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 16–28, Jan. 1997.
- [14] A. Brødsted, An Introduction to Convex Polytopes. Berlin, Germany: Springer-Verlag, 1983.
- [15] Y. Nakayama and S. Aikawa, "Cell discard and TDMA synchronization using FEC in wireless ATM systems," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 29–34, Jan. 1997.
- [16] D. Petras and A. Hettich, "Performance evalution of a logical link control protocol for an ATM air interface," to appear in *Int. J. Wireless Information Networks*, 1997.
- [17] G. Chen and I. Stavrakakis, "ATM traffic management with diversified loss and delay requirements," in *Proc. IEEE INFOCOM*, San Francisco, CA, Mar. 1996.
- [18] Y. Jeon and I. Viniotis, "Achievability of combined GOS requirements in broadband networks," in *Proc. IEEE ICC*, 1993, pp. 192–196.



Jeffrey M. Capone (S'92–M'98) received the B.S.E.E. degree from the University of Vermont, Burlington, in 1992, and the M.S. and Ph.D. degrees from Northeastern University, Boston, MA, in 1995 and 1997, respectively.

In 1997, he joined the Faculty of Electrical Engineering, Arizona State University, Tempe. His primary research is in the design and analysis of controlling policies for bandwidth management in integrated services wireless communication networks, multihop packet radio networks, and

large-scale netwrok models.

Dr. Capone is a member of the IEEE Communications Society and the Technical Committee on Computer Communications (TCCC). He has served on the technical committee for GLOBECOM and the program committee for MobiCom and ICASSP, as well as organized and chaired sessions for ICC and ICUPC. He is also currently serving as the Conference Coordinator for the TCCC.

Ioannis Stavrakakis (S'84–M'88–SM'93) received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1983, and the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, in 1988.

In 1988, he joined the Faculty of Computer Science and Electrical Engineering at the University of Vermont, Burlington, as an Assistant, and later, Associate Professor. Since 1994, he has been as Associate Professor of electrical engineering at Northeastern University, Boston, MA. His research interests are in stochastic system modeling, teletraffic analysis, and discrete time queueing theory, with primary focus on the design and performance evaluation of broadband integrated services digital networks (B-ISDN).

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