Rate Control for UWB Mobile Radios

Gianluca Reali

Abstract: **This paper shows a technique for regulating the rate emitted by UWB mobile terminals. This technique is designed to adapt the emission rate to the effects of channel capacity variations. We assume that individual flows are shaped at edge routers by means of dual leaky buckets. The proposal is based on the evaluation of certain parameters, which reflect the channel status, sent to the application layer. This information is generated by the scheduler designed to guarantee appropriate channel sharing for the various traffic components.**

We describe some results which highlight how performance depends on the specific operating conditions, as well as on the traffic descriptors of the flows, their reciprocal difference, and performance requirements.

Index Terms: **Additional delay, DLB, rate control, scheduling, UWB.**

I. INTRODUCTION

Ultra-wideband (UWB) technology consists of impulse radio transmission, which is expected to allow users in operation to access information infrastructure at data rates of tens of megabits per second. Since UWB signals consist of pulses (monocycles) with a duration of less than one nanosecond, the relevant bandwidth is very much higher than in traditional telecommunication systems. It is specified that the minimum bandwidth of UWB signals is 25% of center frequency. Since the relative position of pulses may be associated with the information content, this information proves very difficult to detect. UWB signals are virtually unaffected by short term fading due to the short pulse duration. In addition, the use of multi-branch receivers, such as the *rake* receiver, enables a number of isolated replicas of the pulses to be collected and combined in a constructive way, thus enhancing their resistance to short term fading [1].

Despite the potentials of this technique, some technological impairments [2], the presence of UWB, and non-UWB variable interference [3], [4], together with the variability of radio propagation conditions associated with user mobility, as well as shadowing from the surrounding environment, may cause some problems for the required radio link quality to be achieved and for network resources to be allocated efficiently.

At various protocol layers, researchers have already defined and evaluated a number of solutions to address propagation problems and optimize the exploitation of the channel capacity [3]–[5].

In this paper, we deal with the variability of the path gain due to terminal mobility and shadowing effects, which could induce a similar variability on the signal to noise ratio. In turn, this implies that the link capacity is also time-variable. In this situation, it could be desirable for applications to adapt to the current link capacity by means of rate control algorithms and protocols. They should be able to react promptly to channel variations, for example, by dropping high-resolution sub-streams of a hierarchically coded information flow to reduce the data rate. This paper puts forward a proposal to manage variable channel capacity. Some concepts have been outlined concisely in [6]. In addition to a more detailed description, the value added by this manuscript to the contents of [6] consists of different control functions, the analysis of some scheduling effects which were not shown in [6] due to limited space, and different numerical results, which have been oriented to the UWB transmission.

The proposal is based on the evaluation of certain parameters, which reflect the channel status, sent to the application layer. This information is generated by the scheduler designed to guarantee appropriate channel sharing to the various traffic components. Via this information, applications can dynamically adapt the flow emission rate, thus avoiding a large number of packets which cannot be transmitted successfully from being pushed into the radio interface and causing possible problems to MAC protocols (e.g., increasing packet retransmission rate and link saturation). This interaction of the link layer with the application layer can be used to develop a number of strategies to manage channel degradation.

The basic assumption of our work is that each traffic component is regulated by means of a dual leaky bucket (DLB). Our proposal consists of adapting the parameters of the traffic shapers by using the channel status of the relevant traffic component. The definition of the adaptation rules aims at minimizing service degradation.

The structure of this paper is as follows. Section II contains some background on source shaping and resource management. Section III shows the UWB terminal architecture and the rate control problem. Section IV illustrates the proposed solution. A numerical example is given in Section V. Some final comments conclude the paper.

II. BACKGROUND CONCEPTS: SOURCE RATE REGULATION

The dual leaky bucket (DLB) has been standardized for both ATM and IP networks. A DLB jointly implements two generic cell rate algorithms [7], [8]. Our DLB model is very simple and based on three traffic descriptors only: P_s , r_s, and B_{TS}. This approach is widely used in literature; e.g., in [9], the authors represented DLB as shown in Fig. 1. Using this model, each information unit (IU) needs a token to be admitted. Tokens arrive at the token buffer at the sustainable rate r_s. When tokens are available inside the token buffer, they can be picked up by the incoming IUs. Therefore, the output rate process $U(t)$ is the minimum between the input rate $I(t)$ and r_s. When the token

Manuscript received August 1, 2003; approved for publication by M. Gabriella Di Benedetto, Guest Editor, November 17, 2003.

G. Reali is with the Dipartimento di Ingegneria Elettronica e dell' Informazione(DIEI), University of Perugia, Italy, email: reali@diei.unipg.it.

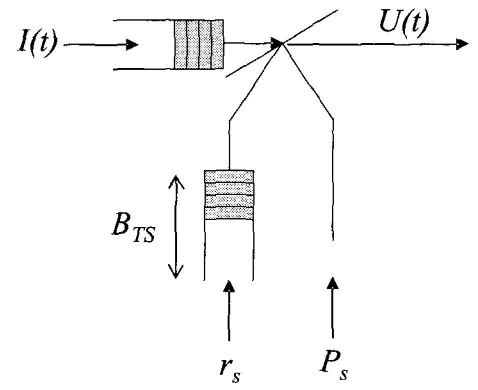


Fig. 1. DLB equivalent model.

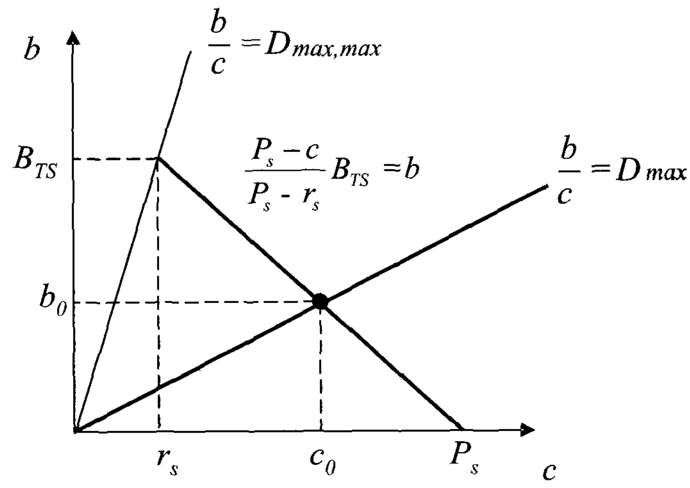


Fig. 2. Determination of the equivalent bandwidth (c_0) and the equivalent buffer (b_0) for loss-free multiplexing.

buffer is empty, the output rate cannot be higher than the token arrival rate r_s.

In this paper, the DLB parameters are used to determine the *equivalent buffer* and the *equivalent bandwidth* of the shaped flows. A number of models have been proposed in literature (see [10] for a survey). The models shown in [9] and [11], together with the one shown in [12], provide analogous results, but it is useful to consider both of them, as they highlight different aspects which help to define the rules of allocation. It is worth noting that in [9] and [11] the authors have used the fluid traffic model. Clearly, this model cannot be used to represent the packet-based operation of a scheduler. Nevertheless, since packet queuing delays are virtually unaffected in the heavy traffic condition, for the sake of simplicity we prefer to use the same traffic descriptors as in [9] and [11].

The basic idea is as follows. Let us consider a buffer of unlimited size being loaded by a shaped flow and serviced at the output trunk capacity c. If c is above the average flow rate, the maximum buffer occupancy $b(c)$, which is a function of the bandwidth variable c, is the equivalent buffer to be associated with the flow. Using a buffer of limited size, the equivalent buffer is the memory space to be allocated to the flow in order to avoid losses. Clearly, the possible combinations of (b, c) pairs are infinite. To select any one pair in particular, another equation may be used. For example, if we consider an upper bound to the maximum queueing delay, we can select the pair (b_0, c_0) which satisfies this requirement. Thus, the (b_0, c_0) pair represents the equivalent buffer and the equivalent capacity of the regulated flow in the loss-free case. In [9], the authors have shown that the maximum buffer occupancy relevant to a transmission capacity c is given by

$$\frac{P_s - c}{P_s - r_s} B_{TS} = b, \tag{1}$$

where the (b, c) pair belongs to an infinite set of eligible values of equivalent buffer and capacity, respectively, for each shaped flow. In addition, we impose the maximum queueing delay of an IU at D_{max}. Therefore, the following additional equation applies:

$$\frac{b}{c} = D_{max}. \tag{2}$$

This equation provides another infinite set of pairs compliant with the tolerable delay D_{max}.

The intersection of (1) and (2) determines the association of the pair (b_0, c_0) with the regulated flow, as show in Fig. 2. The result is:

$$c_0 = \frac{P_s B_{TS}}{D_{max}(P_s - r_s) + B_{TS}}, \tag{3}$$

$$b_0 = c_0 D_{max}. \tag{4}$$

Analogous results may be obtained by making use of some mathematical tools developed in the framework of the *min-plus algebra.* Due to limited space we will only recall some basic concepts in this paper. For an exhaustive description the reader should refer to [13].

The most important analytical models used in this paper are the *arrival curve* and the *service curve* [12]. Let $R(t)$ be the number of data units of a flow crossing a given interface in the interval $[0, t]$. An arrival curve α of that flow is a wide-sense increasing (or non-decreasing) function, so that for all $s \leq t$ $R(t) - R(s) \leq \alpha(t - s)$. The function α is a meaningful constraint only if it is sub-additive, that is $\alpha(t + s) \leq \alpha(t) + \alpha(s)$, for all $s, t \geq 0$.

To define the service curve, we consider a system S giving a service to an incoming flow with an arrival curve $R(t)$. Let $R^*(t)$ be the corresponding output function. A service curve β of the system S is a function which enables a time $t_0 < t$ to be determined for all $t \geq 0$, so that $R^*(t + T) - R(t_0) \geq \beta(t - t_0)$.

A number of useful equations using the arrival curve and the service curve have been demonstrated (see [12] and its referenced papers). Some important results are concerned with the delay experienced within a system. The so-called *virtual delay* $d(t)$ is $d(t) = \inf\{T : T \geq 0, R(t) \leq R^*(t + T)\}$, which represents the delay experienced by an IU entered at time t into a system, which provides a service according to an FCFS scheduling. The virtual delay is bounded by:

$$d(t) \leq \sup_{s \geq 0}\{\inf[T : T > 0, \alpha(s) \leq \beta(s + t)]\} \triangleq h(\alpha, \beta), \tag{5}$$

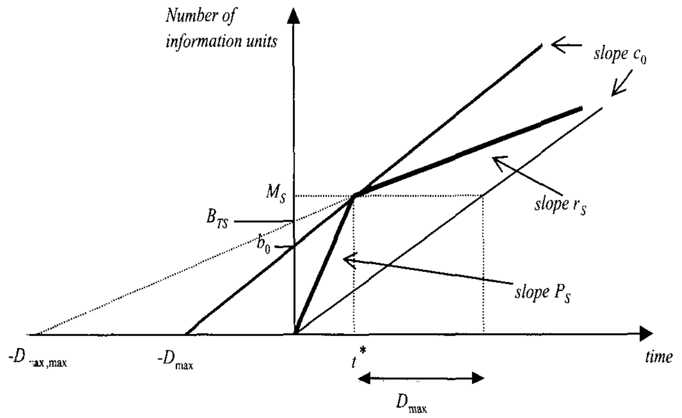


Fig. 3. Derivation of the equivalent capacity, buffer from the arrival curve of the shaped flow and the service curve of the channel.

where $h(\alpha,\beta)$ is defined as the horizontal deviation between the functions α and β.

All these concepts are applicable to shapers, that bound the arrival curve of the output flow by a given curve σ. It is shown that if σ is a sub-additive function, with $\sigma(0) = 0$, then the shaper offers σ as a service curve.

To determine the equivalent bandwidth we make use of an (5), which is the horizontal deviation between the arrival curve of a shaped flow and the service curve of the output link. The former is the service curve of the shaper, whereas the latter is the service curve of a link with capacity c, that is

$$\sigma_s(t) = ct\,, \quad t \geq 0. \tag{6}$$

The DLB service curve is

$$\sigma_{DLB}(t) = \inf_{t\geq 0}\left\{P_s t, B_{TS} + r_s t\right\}. \tag{7}$$

The minimum rate so that the horizontal deviation $h(\sigma_{DLB}, \sigma_s)$ is not greater than D_{max} is

$$c_0 = \sup_{s\geq 0}\frac{\sigma_{DLB}(s)}{s + D_{max}}. \tag{8}$$

c_0 is the slope of the tangent to the arrival curve, that intersects the time axis at $t = -D_{max}$ (see Fig. 3). The intersection of the tangent with the ordinate axis gives the equivalent buffer size. Note that the approach can be used for any type of shaper.

In Fig. 3, we indicate the quantity M_S, that is the maximum burst size that a DLB may emit at the peak rate, that is

$$M_S = B_{TS}\frac{P_s}{P_s - r_s}. \tag{9}$$

This burst ends at time t^*, which is the time when the DLB token buffer is emptied.

Thus, the pair (b_0, c_0), corresponding to the delay D_{max}, is determined. It is easy to prove that this pair corresponds to the analogous pair determined using (1) and (2), and Figs. 2 and 3 represent the same situation by emphasizing different aspects.

The performance measure we use is the maximum transmission delay, including the queueing delay. For the sake of brevity we will only consider hard delay bounds. Since each flow is

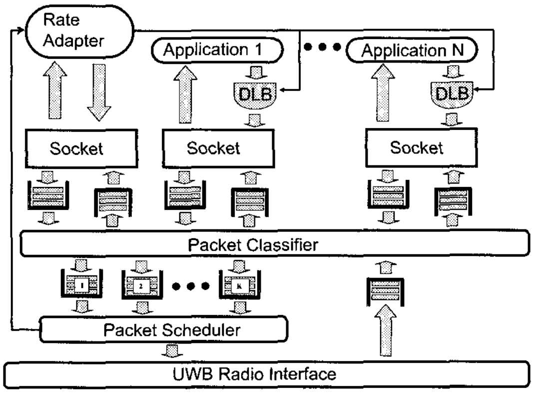


Fig. 4. Functional block representation of the UWB terminal.

shaped, the amount of capacity and buffer space needed for the delay bound D_{max} to be respected is given by the (3) and (4).

III. THE RATE CONTROL PROBLEM

Fig. 4 shows the reference terminal architecture. The scheme does not include all the network functions which must be implemented in the terminal, but only the subset necessary for our analysis. A number of network applications share the same UWB radio interface. The traffic flows emitted by the applications are shaped through DLBs, as described in Section II. The socket application programming interface permits applications using network protocols. Traffic is buffered and sent to the radio interface. The capacity of the link is managed in order to support all services by means of an appropriate scheduling technique. This technique is based on associating each flow with a capacity-sharing factor and provides strong guarantees under average use of the bandwidth portion reserved for each one. In addition, the excess bandwidth is shared among the other flows in the same proportion as their capacity sharing weights. This means that the server is always busy if there are packets waiting in the system (work conserving), that are scheduled using the relevant value of the weights. As regards the receive chain, dual functions are implemented.

Packet transmission through the radio interface undergoes a number of varying effects, essentially due to movements of both mutually interfering UWB users and a changing environment, including UWB and non-UWB interfering radio transmissions. Despite the countermeasures taken to face these variations, the signal to noise ratio (SNR) at receiver site is variable, and may occasionally drop to very low values. In [4], the authors have shown that, considering N UWB simultaneous links, the SNR at the receiver of the ith link is

$$\mathrm{SNR}_i = \frac{P_i g_{ii}}{R_i\left(\eta_i + T_f\sigma^2 \sum\limits_{k=1,k\neq i}^{N} P_k g_{ki}\right)}, \; i = 1, \cdots, N, \tag{10}$$

where T_f is the frame period consisting of periodic time intervals, where the terminals may transmit *position-modulated* monocycles, R_i is the bit rate of the ith link, P_i is the power emitted by the transmitter of the ith link, g_{kj} is the path gain from the transmitter of the kth link to receiver of the jth link, η_i is background noise which includes the non-UWB transmissions, and σ^2 is an adimensional parameter, the value of which depends on the shape of the monocycle.

When the path gain coefficients g_{kj} change due to a temporary variation in the environment, not only could the SNR decrease, but the related channel capacity could also drop and applications would not be served in accordance with the traffic descriptors. The effects on resulting performance may be either minimal or extremely critical. From (10), it is clear that the situation could in some way be fixed in a short time by acting on the transmission rate R_i, which, in this respect, could also be considered as a tuning knob to regulate the SNR value. Note that this control could also be implemented without making use of any additional signaling protocols distributed over all mutual interfering links. For example, it is possible to modify the channel coding rate over a single link, thus making transmission more immune to errors to give an appreciable result at a relatively low cost.

Capacity-dropping effects on provided services essentially depend on the DLB output rate and on the fade period. Qualitatively speaking, if the emission rate is low, a decrease in the channel capacity value does not significantly affect performance. If the rate increases, a brief fade duration may be absorbed by the statistical multiplexing of the various flows. In contrast, brief fades could be critical at high rates. For this reason, not only when action needs to be taken, but also what information is needed to take such a decision, together with what kind of counteraction is appropriate, all has to be determined. The diagram given in Fig. 4 shows a feedback connection from the packet scheduler to DLBs. Our proposal consists of using this connection to send an alarm message. The basic concept is that the alarm message is generated when the *effects* of link capacity variation have an impact on perceived performance. The action generated by the alarm message is a variation of the traffic descriptors used to shape traffic. Note that in the diagram shown in Fig. 4 we include the possibility that the rate adapter can obtain information on the network status from the outside through a dedicated socket, and consequently can take adequate decisions. Limited space prevents us from following up this idea in this paper.

The concepts shown in Section II can be used to extend the analysis to a situation in which flows are served according to service disciplines differing from the FCFS discipline. The interested reader can find appropriate background information on this subject in [14].

The buffers between the packet classifier and the scheduler, as shown in Fig. 4, are selected for transmission through a scheduling discipline. In the case of circular and periodic scanning of the buffers, this discipline is the well known *time division multiplexing* (TDM) scheduling. Fig. 5 shows the service curves for a generic buffer, for a TDM discipline, and for a generic scheduling discipline. Both curves are clearly upper bounded by the "ideal" service curve with a slope equal to the equivalent bandwidth, as determined in Section II. Nevertheless, apart from the queueing time necessary for an IU to get the head of the queue, once this position is reached, an additional delay for receiving

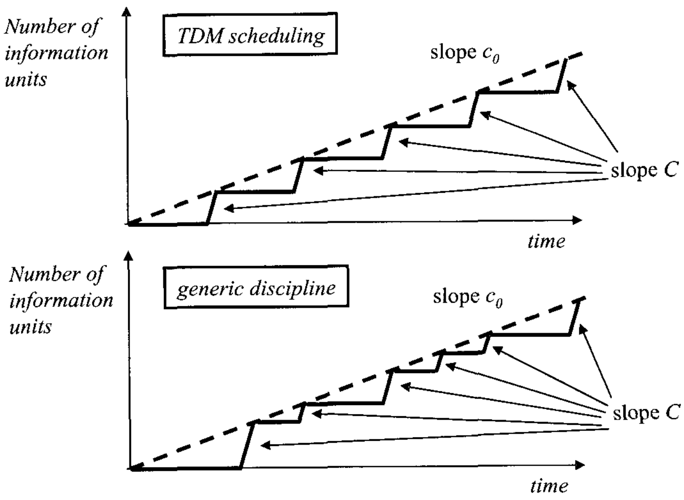


Fig. 5. Number of transmitted IUs vs. time for a generic buffer in the case of a round robin and generic scheduling discipline.

service could be suffered. In the case of a round-robin, this additional delay is upper bounded by $(N-1)T$, where N is the number of buffers and T is the periodic transmission time-slot allocated to each buffer. For a generic scheduling discipline this additional delay may be modeled as a random variable, with its maximum denoted as ΔD_{max}.

If we want the flows to suffer no more than the pre-determined delay D_{max}, ΔD_{max} has to be a portion of D_{max}. This implies that the allocation rules shown in the previous section are applied by considering a maximum acceptable delay equal to $D_{max} - \Delta D_{max}$. The consequence is that the equivalent bandwidth of the flows increases, hence the admissible number of flows decreases. On the contrary, if we ease the hard delay requirement, the possible approach may differ. In fact, if the additional delay is found to be acceptable, it may be added to D_{max}. In this way the same allocation results shown in the previous paragraph are maintained, to the detriment of an increased delay. In the remainder of the paper, we will support these qualitative concepts by a quantitative analysis.

As expected, the additional delay may also have an impact on the equivalent buffer. In fact, the maximum vertical deviation between the DLB arrival curve and the service curve could increase up to a maximum value of $\Delta b_{max} = c_0 \Delta D_{max}$. Moreover, Δb_{max} may be considered either as a portion of the equivalent buffer or as an addition to it.

Now we consider the additional delay ΔD_{max} and evaluate its effects on the equivalent bandwidth when it is considered as part of it. In other words, we determine the necessary *increment* of the equivalent bandwidth in order to *absorb* the additional delay ΔD_{max}. Fig. 6 shows the worst case in the delay perspective. In this figure, a small portion of the service curve has been drawn. This portion corresponds to the maximum value ΔD_{max} of the additional delay, which, in the worst case, adds to the maximum horizontal deviation between the DLB arrival curve and the straight line with slope c_0. In order to include this additional delay in D_{max} we fold the segment of ΔD_{max} backward over the deviation segment D_{max} (see Fig. 6). The straight line with slope c_0' represents a service curve relevant to

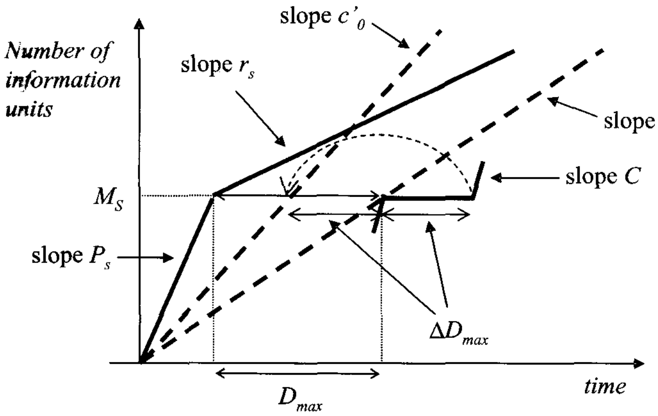


Fig. 6. Graphical representation of the increment of the equivalent buffer due to the additional delay.

an equivalent bandwidth c_0', which has a maximum horizontal deviation $D_{max} - \Delta D_{max}$ from the arrival curve. Thus, by using the increased equivalent bandwidth c_0' the increased delay never violates the hard delay specification D_{max}.

Fig. 6 shows that

$$\frac{M_S}{c_0} - \frac{M_S}{c_0'} = \Delta D_{max}, \tag{11}$$

hence

$$c_0' = \frac{c_0}{1 - c_0 \frac{\Delta D_{max}}{M_S}}. \tag{12}$$

Clearly (12) is not valid for all possible ΔD_{max} values. The most limiting bound is that $c_0' \leq P_S$. (12) proves that

$$\Delta D_{max} \leq \left(\frac{M_S}{c_0}\right)\left(\frac{P_s}{1 - P_s}\right). \tag{13}$$

As expected, (13) shows us that when c_0 is small, that is to say when the tolerable maximum delay and the corresponding multiplexing gain are large, it is also possible to compensate large additional delays. Nevertheless, (12) shows that this situation corresponds not only to a marked increase in the equivalent bandwidth, but also to a corresponding decrease in the number of flows that the UWB terminal can handle and to a pejorative effect on the multiplexing gain.

Even though this result refers to the time of the maximum horizontal deviation between the arrival curve and the service curve with slope c_0, we wonder whether the equivalent bandwidth c_0' enables the additional delay to be included at all times. In order to prove this, let t_0 and $\beta(t_0)$ be a generic time and the corresponding value of the arrival curve, respectively. The straight lines with slope c_0' reaches this value at time $\beta(t_0)/c_0'$. The time deviation from t_0 is $\beta(t_0)c_0't_0$, which, when added to the maximum additional delay ΔD_{max} must not be lower than D_{max}, as follows:

$$\frac{\beta(t_0)}{c_0'} - t_0 + \Delta D_{max} \leq D_{max}, \tag{14}$$

which leads to the following condition:

$$c_0' \geq \frac{\beta(t_0)}{D_{max} - \Delta D_{max} + t_0}. \tag{15}$$

The right-hand side of the inequality (15) reaches its maximum at the time of the maximum vertical deviation between the arrival curve $\beta(t)$ and service curve with slope c_0, which in turn corresponds to the starting time of the maximum horizontal deviation. This proves that (12) gives an equivalent bandwidth which enables the maximum delay D_{max} to be respected at any time.

A similar analysis may be performed to find the increase of the equivalent bandwidth in order to avoid increasing the necessary buffer space beyond the value computed in (4). Let t_0 and $\beta(t_0)$ be a generic time and the corresponding value of the arrival curve, respectively. Without considering any additional delay, the following equation stands:

$$b(t_0) = \beta(t_0) - c_0 t_0 \leq b_0, \tag{16}$$

where $b(t)$ is the buffer occupancy at time t.

If we consider an additional delay ΔD_{max}, in order to avoid any additional buffer space an increased equivalent bandwidth equal to the minimum c_0'' value must be used so that:

$$b(t_0) = \max(0, \beta(t_0) - c_0'' \max(0, (t_0 - \Delta D_{max}))) \leq b_0, \tag{17}$$

hence

$$c_0''(t_0 - \Delta D_{max}) \geq c_0 t_0, \quad t_0 \geq \Delta D_{max}, \tag{18}$$

with the result that

$$c_0'' \geq c_0 \frac{1}{1 - \frac{\Delta D_{max}}{t_0}}, \quad t_0 \geq \Delta D_{max}. \tag{19}$$

Equation (19) implies that c_0'' could also diverge. Hence, there is a possibility that $c_0'' > c_0'$. In such a situation we suggest the following choice. Even though the buffer space is a very important resource, an efficient bandwidth utilization could be preferable. Therefore our suggestion is to use the value c_0'. The reason for this choice is that a small increase in the implemented buffer space would not be dramatic, whereas an efficient use of an assigned bandwidth is a basic requirement that has greatly stimulated research, including research into UWB. Clearly, different considerations could apply in some instances, nevertheless limited space prevents us from outlining all of them. In the following we will refer to c_0' only, and we will assume that sufficient buffer space is available. Under this assumption, the maximum buffer occupancy is:

$$b_{max} = \begin{cases} \beta(\Delta D_{max}), \ \Delta D_{max} \geq t^* \\ \beta(t^*) - c_o'(t^* - \Delta D_{max}), \ \Delta D_{max} < t^*, \end{cases} \tag{20}$$

where t^*, shown in Fig. 3, is the time at which the DLB token buffer is emptied.

It is now necessary to determine the value of additional delay ΔD_{max} in order to make use of the results shown above. ΔD_{max} clearly depends on the scheduling discipline. This means that the following analysis should focus on the queueung delay, which should be weighted according to the delay specification in order to schedule the IU waiting for transmission. Clearly, when the delay suffered by some IUs approaches the delay tolerable for them, their weight should overcome that of the others so as to "force" the scheduler to select them. There are many potential approaches to obtain this result. In the next section, we illustrate our proposal

IV. PROPOSED SOLUTION

Among the possible solutions proposed in literature, we believe that the packetized implementation of GPS based on the virtual time [15] is the most applicable to our problem. For this implementation, the authors define the so-called *virtual finish time*. Since this parameter cannot provide any upper bound to the queueing delay if sources are generic, authors have included the LB source shaping. The LB model used in [15] is equivalent to the DLB used in this paper. This results in a bound on the maximum length of the busy period [15, *Lemma* 3]. This solution is valid, but we wondered whether different delay bounds can be found through a different approach. In addition, since the situation analyzed in this paper includes a variable channel capacity, we prefer a solution which does not require the knowledge of the current value of the channel capacity, thus saving the computational burden of a specific estimation. We have found that these requirements can be met by an intuitive weight definition, which consists of the queueing time normalized to the maximum tolerable queuing delay. In principle, each queued IU is associated with a weight. The ith IU in the kth buffer has an associated weight equal to:

$$P_{i,k}(t) = \frac{t - t_{a,i,k}}{D_{max,k}}, \; k = 1, \cdots, N_b, \tag{21}$$

where $t_{a,i,k}$ is the arrival time of the IU and $D_{max,k}$ the tolerable delay for the IUs loading the kth buffer.

This weight definition could lead the reader to the conclusion that the maximum additional delay for the generic kth flow is $D_{max,k}$. This would be true in the case of unregulated sources and constant channel capacity. In this paper we assume that the emission rate of the information sources is shaped, so that we will show that the maximum additional delay $\Delta D_{max,k}$ is strictly lower than $D_{max,k}$.

The maximum weight for the kth buffer can be found by inverting (20):

$$P_{i,k}(t) \leq \begin{cases} \frac{b_{max,k}}{P_{s,k} D_{max,k}}, \; \Delta D_{max,k} < t^*, k = 1, \cdots, K \\ \frac{1}{D_{max,k}} \left(t^* + \frac{b_{max,k} - P_{s,k} t^*}{r_{s,k}} \right), \; \Delta D_{max,k} \geq t^*, \\ k = 1, \cdots, K, \end{cases} \tag{22}$$

where K is the number of buffers, as shown in Fig. 4. Since the worst case is clearly $\Delta D_{max,k} \geq t^*$, we can write that in general

$$\begin{aligned} P_{i,k}(t) &\leq \frac{b_{max,k} - (P_{s,k} - r_{s,k}) t^*}{D_{max,k} r_{s,k}} \\ &= \frac{\beta(\Delta D_{max,k}) - (P_{s,k} - r_{s,k}) t^*}{D_{max,k} r_{s,k}}, k = 1, \cdots, K. \end{aligned} \tag{23}$$

Since $\Delta D_{max,k} \geq t^*$,

$$\begin{aligned} &\beta(\Delta D_{max,k}) = \\ &P_{s,k} t^* + r_{s,k} (\Delta D_{max,k} - t^*), \; k = 1, \cdots, K. \end{aligned} \tag{24}$$

As a result of (23) and (24):

$$P_{i,k}(t) \leq \frac{\Delta D_{max,k}}{D_{max,k}}, \; k = 1, \cdots, K. \tag{25}$$

This is an important result. Equation (25) means that the *maximum* value of the weights is only due to the *additional* delay. The physical interpretation is the following: The IU at the head of the kth queue can suffer the maximum additional delay only if the available capacity has previously been used to transmit the IUs preceding it in the same queue, and the queues in the other buffers have grown.

For the sake of simplicity in the following we consider two buffers only ($K = 2$), referred to as B1 and B2, with delay requirements $D_{max,1}$ and $D_{max,2}$, respectively, and $D_{max,1} < D_{max,2}$. The two buffers are served by a link capacity equal at least to the sum of the equivalent bandwidths of the relevant flows.

Let us assume that buffers are initially empty, and then they are loaded by IUs emitted by greedy sources. Since the weight of the IUs in B1 increases faster that in B2, the IUs are at first picked up from B1 at the maximum rate C, until, at a given transmission time, the weight at the head of B2 is larger than $1/CD_{max,1}$. When this occurs, the IUs accumulate in B1, until the next transmission time, and so on. If after a transmission from B1 the buffer is not emptied, the content in B1 increases to an asymptotic value. To determine this asymptotic value we follow a simple, slightly approximated model. We observe that for the two weights at the head of B1 and B2 to reach almost the same value, a number n of *time-slices* $1/C$ is necessary, so that:

$$\frac{n}{D_{max,2}} = \frac{1}{D_{max,1}}, \tag{26}$$

thus

$$\Delta D_{max,1} \cong \left\lceil \frac{D_{max,2}}{D_{max,1}} \right\rceil \frac{1}{C}. \tag{27}$$

As the initial value of the weights in the buffer is equal to zero, the transmission time, when the weight at the head of B2 is larger than the one in B1, can be found by solving the following inequality:

$$n \frac{1}{C} \frac{1}{D_{max,2}} \geq \frac{1}{D_{max,1}} \left(\Delta D_{max,1} - \frac{1}{C} \right), \tag{28}$$

from which it follows that

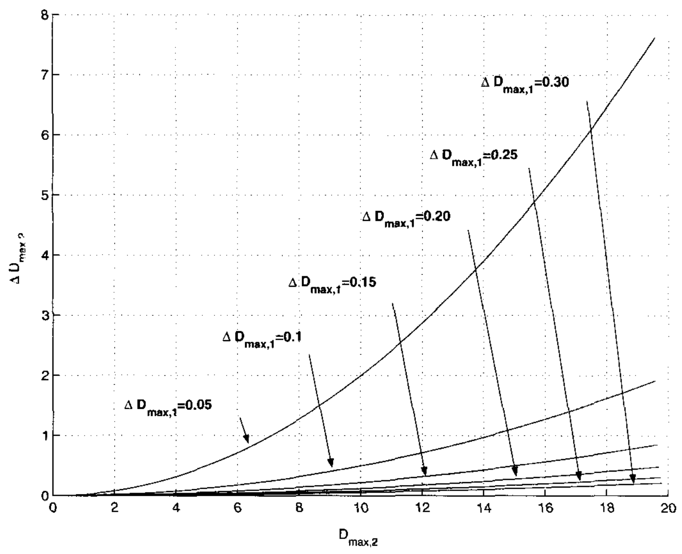


Fig. 7. Maximum additional delay $\Delta D_{max,2}$ vs. $D_{max,2}$.

$$n_0 = \left\lceil \frac{D_{max,2}}{D_{max,1}} \left(C\Delta D_{max,1} - 1 \right) \right\rceil, \tag{29}$$

thus

$$\Delta D_{max,2} = \frac{1}{C} \left\lceil \frac{D_{max,2}}{D_{max,1}} \left(C\Delta D_{max,1} - 1 \right) \right\rceil. \tag{30}$$

It is worth noting that the additional delay may or may not be significant. This mainly depends on the value of the traffic descriptors and on the delay requirements of the flows. A numerical example is shown in Fig. 7. This figure shows the maximum additional delay of IUs in B2, vs. their delay requirement. Each curve is relevant to a specific value of the maximum delay requirement in B1, $D_{max,1}$. The values of traffic descriptors of the flows, selected for this example are: $r_{s,1} = r_{s,2} = 10^4$ IU/s, $P_{s,1} = 2\cdot 10^4$ IU/s, $P_{s,2} = 10^5$ IU/s, $B_{TS,1} = r_{s,1}D_{max,1}$, and $B_{TS,2} = r_{s,2}D_{max,2}$. In this figure, we can appreciate the effects of $D_{max,1}$ and $D_{max,2}$ on $\Delta D_{max,2}$. In particular, the former has a greater impact since, due to the weight definition (21), it gives a major contribution to the probability that the scheduler selects B1. On the other hand, this effect may be marked only if $D_{max,2}$ is large, as expected, otherwise the effects of small values of both D_{max1} and $D_{max,2}$ tend to counterbalance each other.

The results obtained above refer to a fixed channel capacity C. Unfortunately, this condition very seldom occurs in radio transmissions, especially in the case of UWB transmissions, and mobile terminals may suffer some fluctuations in link capacity.

In general, fluctuations happen for all radio transmission techniques. Anyway, different transmission techniques might suffer different attenuation patterns, with different statistics. For example the narrowband transmission is affected, beyond other things, by the fast fading phenomenon due to the destructive interference of the symbol replicas. Differently, some measurements of the UWB channel reveal a noise-like shape of the impulse-response within the delay spread of some nanoseconds, together with increased dissipative losses. Our target is to control the emission rate of broadband terminals, as a countermeasure of varying channel conditions. Even if this problem is not exclusive of UWB, it characterizes the UWB deeply. In addition, UWB shows the capabilities of implementing these countermeasures due to its extreme flexibility in fixing the transmission parameters, as shown in (10) and related comments. Thus, when the link capacity drops below its nominal value, the UWB flexibility allows implementing a number of countermeasures. Basically the type of action depends on the effects of the capacity decrease over the performance of the transmitted flows. In order to capture these effects, we propose to maintain an updated global parameter which indicates the ability of the terminal to guarantee the agreed delay performance.

Given the value of the traffic descriptors and the delay requirements, it is possible to determine the maximum value of the weights. This means that if the link capacity does not drop below its nominal value, the weights cannot be larger than these mathematically determined values. Hence, the current value of the weights may be regarded as an indicator of the terminal ability to respect the predetermined delay performance, and can be used as the control parameter to regulate the DLB traffic descriptors.

(25), (27), and (30) give the following maximum value of the weights for a link capacity value C:

$$P_{max,1} = \left\lceil \frac{D_{max,2}}{D_{max,1}} \right\rceil \frac{1}{D_{max,1}C}, \tag{31}$$

and

$$P_{max,2} = \frac{1}{CD_{\max,2}} \left\lceil \frac{D_{max,2}}{D_{max,1}} \left(C\Delta D_{max,1} - 1 \right) \right\rceil. \tag{32}$$

We consider $P_{max,1}$ and $P_{max,2}$ as thresholds. The decrease of the channel capacity causes weight increase. If variations are brief and sporadic, and the terminal load is low, it is likely that the value of the weights will remain below the thresholds. This means that the terminal is still able to respect the transmission-delay bound. On the contrary, if variations are large and durable, the value of the weights may rise above their theoretical maximum value.

When this occurs, different actions may be taken. According to the diagram in Fig. 4, an alarm message is generated and DLB traffic descriptors may be updated so as to make traffic statistics compliant with the available channel capacity.

From the implementation point of view, it is sufficient to stamp the arrival time on the IUs and to calculate the weights for those at the head of the queues only. In addition, it is worth noting that if the link capacity is constant, the next transmission can be scheduled during the time of the service in progress. Differently, if the channel capacity decreases, and this can be argued from the feedback information, the value of the weights may be computed by the scheduler at the transmission times only. Anyway, the specific control may either be blind to the current value of the channel capacity (e.g., pre-determined adaptation of the DLB parameters), or be able to adapt to the estimated capacity value. The former is very simple, and the performance of a specific implementation is illustrated below. On the contrary, if the preferred control, not shown in this paper, is made according to the current value of the channel capacity, this value has to be estimated. This estimation may be made using (31) and (32). At a first glance it appears that (31) is sufficient to estimate the current capacity value, which is:

$$C_{est,1} = \frac{\left\lceil \frac{D_{max,2}}{D_{max,1}} \right\rceil}{\hat{P}_{max,1} D_{max,1}}, \tag{33}$$

where $\hat{P}_{max,1}$ is the observed maximum value of the weight at the head of B1, in the interval where the capacity value is assumed to be stationary. Nevertheless, the above estimation could be noisy, due to the assumption that all the sources are greedy and that the capacity in the observed interval is stationary. The estimation could be improved by considering (32), from which it is very easy to find that

$$C \geq \frac{D_{max,2} - D_{max,1}}{D_{max,2}\left(\Delta D_{max,1} - P_{max,2} D_{max,1}\right)}. \tag{34}$$

As a result of (25):

$$C_{est,2} = \frac{D_{max,2} - D_{max,1}}{D_{max,2} D_{max,1}\left(\hat{P}_{max,1} - \hat{P}_{max,2}\right)}, \tag{35}$$

where, in this case also, $\hat{P}_{max,2}$ is the observed maximum value of the weight at the head of B2, in the interval in which the capacity value is assumed to be stationary.

In conclusion, using only two buffers the most reasonable estimate of the current available capacity is

$$C_{est} = \max\left\{\min\left\{C_{est,1}, C_{est,2}\right\}, 0\right\}. \tag{36}$$

It is worth noting that the computational cost of the channel capacity estimation is extremely low. This result is another important strength of this approach. Thus, the radio interface can be monitored in real time, and counteractions can be taken before occasional losses could trigger congestion control or recovery protocols.

V. NUMERICAL RESULTS

Using the approach described above it is possible to design a variety of rate control algorithms. Some of these may simply be based on simple step-wise functions (e.g., a sudden decrease in the source rate when a weight value rises above a given threshold). Others could start from (10) and adapt the rate emission parameters (e.g., DLB parameters) by using the on-going value of the weights (e.g., by estimating the available current link capacity and by taking the appropriate actions). A comprehensive analysis of all possibilities is beyond the scope of this paper and the limited space will not permit it. Nevertheless, in this section we present an interesting case study. We have considered a terminal transmitting two types of flows. The former includes MPEG-like flows shaped by the following traffic descriptors: $r_{s,1} = 4.5 \cdot 10^6$ bit/s, $P_{s,1} = 9 \cdot 10^6$ bit/s,

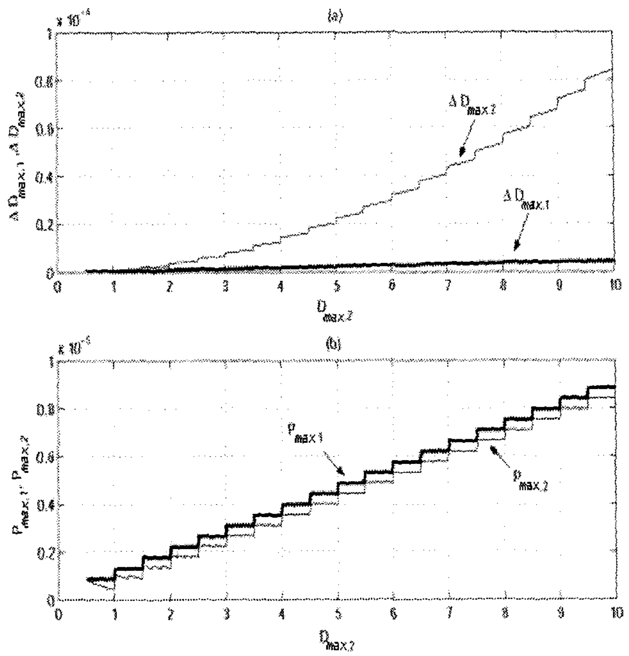


Fig. 8. Results of the case study: (a) Additional delays $\Delta D_{max,1}$ and $\Delta D_{max,2}$ vs. $D_{max,2}$, (b) maximum allowed values for the scheduling weights $P_{max,1}$ and $P_{max,2}$ vs. $D_{max,2}$.

$D_{max,1} = 0.5$ s, and $B_{TS,1} = r_{s,1} D_{max,1}$. The latter may be considered a (slightly) better than best effort traffic, shaped by the following traffic descriptors: $r_{s,2} = 10^4$ bit/s, $P_{s,2} = 10^5$ bit/s, $B_{TS,2} = 2r_{s,2} D_{max,2}$. Fig. 8 (a) shows $\Delta D_{max,1}$ and $\Delta D_{max,2}$ vs. $D_{max,2}$. For this particular case, we can see that if $D_{max,2}$ is limited to few seconds, the additional delays can be regarded as negligible, since their values are a few orders of magnitude lower than the delay specifications. Thus, performance is virtually unaffected by them. Fig. 8 (b) shows $P_{max,1}$ and $P_{max,2}$ vs. $D_{max,2}$. This calculation is necessary to design a rate control technique in the case of a link capacity decrease. We have selected the case $D_{max,2} = 10$ s. We have used the following control functions:

$$r_{s1,c} = \frac{r_{s,1}}{1 + \frac{P_1(t) - P_{max,1}}{P_{max,1}}}, \tag{37}$$

$$r_{s2,c} = \frac{r_{s,2}}{1 + \frac{P_1(t) - P_{max,2}}{P_{max,2}}}, \tag{38}$$

where $r_{s1,c}$ and $r_{s2,c}$ are the controlled values of the sustainable rates. The values of $P_{max,1}$ and $P_{max,2}$ are plotted in Fig. 8 (b) for $D_{max,2} = 10$ s. We stress that these control functions are only instances of the numerous possibilities.

The value of the link capacity used is equal to the sum of the equivalent bandwidth of the flows, which is $C = 4.52$ Mbps. In order to clearly outline the effects of the rate control, we first show in Fig. 9 the weight evolution for B1 and B2 when the capacity does not fluctuate. We can see that the weights converge to an asymptotic maximum value in accordance with the values shown in Fig. 8 (b). Fig. 10 shows the same situation with a varying channel capacity. We have modeled the varying capacity as an unsuccessful transmission probability of 10^{-2}.

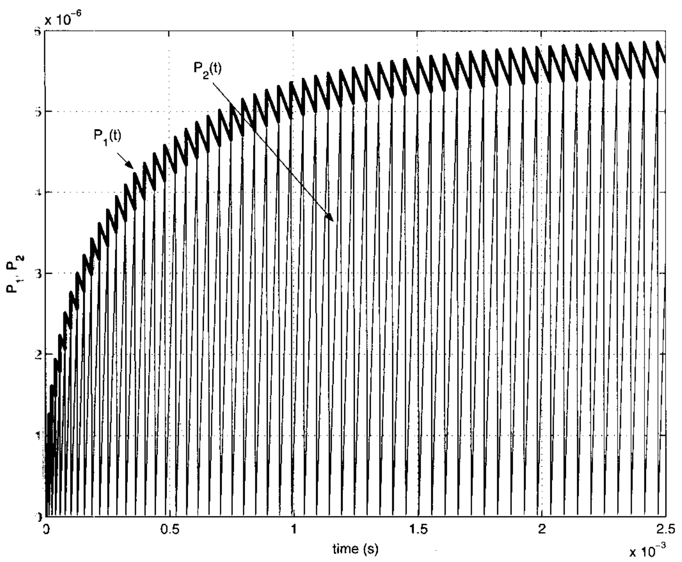


Fig. 9. Results of the case study: Weight evolution for B1 and B2 when the capacity does not fluctuate.

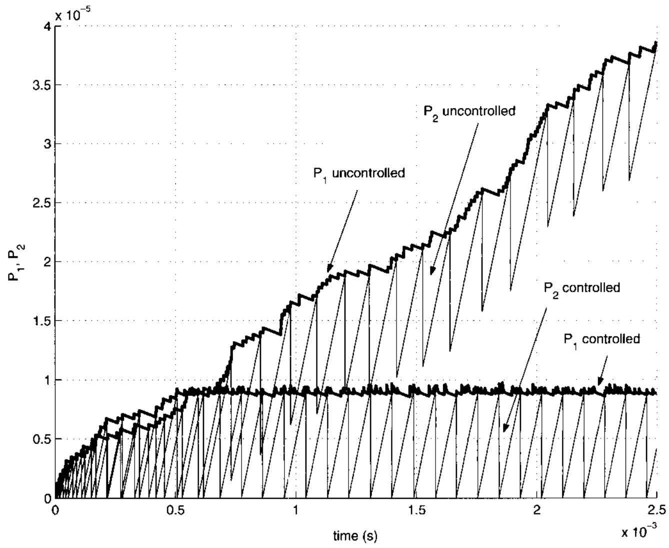


Fig. 10. Results of the case study: The controlled and uncontrolled weight evolution for an unsuccessful transmission probability of 10^{-2}.

Fig. 10 shows both the controlled and the uncontrolled weight evolution. We can see that the rate control is effective in keeping weights below the desired threshold, thus denoting its effectiveness.

VI. Conclusions

This paper deals with the emission rate of UWB radios. The starting assumption is that sources are shaped by means of DLBs. By using the requirements on the maximum tolerable transmission delay, it is possible to associate each flow with a relevant equivalent buffer and bandwidth. In order to define the transmission scheduling algorithm for the buffered IUs, we have highlighted some side effects that could have a significant negative impact on the expected performance. After this, we have calculated the performance obtainable through a specific rate control technique for variable channel capacity, applied to a terminal that has to manage two traffic types with a different set of traffic descriptors and performance specifications in terms of transmission delay. The analyses shown in this paper may be extended in different directions, such as rate control techniques for different traffic types, and scheduling algorithms with reduced transmission time.

References

[1] M. Z. Win and R. A. Scholtz, "Characterization of ultra-wide bandwidth wireless indoor channels: A communication-theoretic view," *IEEE J. Select. Areas Commun.*, vol. 20, no. 9, Dec. 2002.

[2] R. A. Scholtz and J-Y Lee, "Problems in modeling UWB channels," Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers 2002, vol. 1, Nov. 3–6, 2002.

[3] P. Baldi, L. De Nardis, and M. G. Di Benedetto, "Modeling and optimization of UWB communication networks through a flexible cost function," *IEEE J. Select. Areas Commun.*, vol. 20, no. 9, Dec. 2002.

[4] F. Cuomo *et al.*, "Radio resource sharing for ad hoc networking with UWB," *IEEE J. Select. Areas Commun.*, vol. 20, no. 9, Dec. 2002.

[5] D. Di Sorte *et al.*, "Network service provisioning in UWB open mobile access networks," *IEEE J. Select. Areas Commun.*, vol. 20, no. 9, Dec. 2002.

[6] G. Reali, "Rate control for wideband radios," *IEEE Commun. Lett.*, vol. 6, no. 1, Jan. 2002.

[7] ATM Forum, "Traffic management specification," v. 4.0, Apr. 1996.

[8] ITU-T Recommendation I.371, "Traffic control and congestion control in B-ISDN," (08/96).

[9] A. Elwalid and D. Mitra, "Traffic shaping at a network node: Theory, optimum design, admission control," in *Proc. IEEE INFOCOM'97*, Kobe, Japan, Apr. 1997, pp. 445–455.

[10] E. W. Knightly and N. B. Shroff, "Admission control for statistical QoS: Theory and practice," *IEEE Networks*, Mar. 1999.

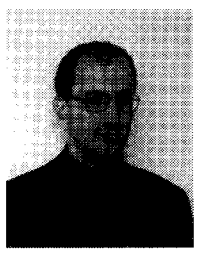
[11] A. Elwalid, D. Mitra, and R. H. Wentworth, "A new approach for allocating buffers and bandwidth to heterogeneous, regulated traffic in an ATM node," *IEEE J. Select. Areas Commun.*, vol. 13, no. 9, Aug. 1995.

[12] J. Y. Le Boudec, "Application of network calculus to guaranteed service networks," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1087–1096, May 1998.

[13] F. Baccelli *et al.*, *Synchronization and Linearity, an Algebra for Discrete Event Systems*, New York, Wiley, 1992.

[14] L. Kleinrock, *Queueing Systems, Vol. II: Computer Applications*, Wiley, 1976.

[15] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single-node case," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, June 1993.



Gianluca Reali has been a researcher of the Department of Information and Electronic Engineering of the University of Perugia since 1997. He received the "Laurea" degree in Electronic Engineering from the University of Perugia in 1991. Then he collaborated with Alenia Spazio S.p.A. working on VSAT networks. He has also acted as Consulting Engineer for the Institute of Electronics of the University of Perugia and CRA in Rome. He received his Ph.D. degree in Telecommunications from the University of Perugia in 1997, working on Spread Spectrum techniques and CDMA. His present research activity is in IP QoS techniques, particularly in switching, transport, and resource management protocols. He has been involved in the European ACTS projects CABSINET and ASSET, and in the European IST projects SUITED and WHYLESS.COM.