Resource Allocation Rules to Provide QoS Guarantees to Traffic Aggregates in a DiffServ Environment

N. Blefari-Melazzi, D. Di Sorte, M. Femminella, G. Reali

D.I.E.I. Department - University of Perugia - Italy
{blefari,disorte,femminella,real}@diei.unipg.it

Abstract: In this paper, we present a resource allocation protocol and the relevant allocation rules to provide end-to-end QoS guarantees to traffic aggregates in a DiffServ environment. The work is realized within the SUITED project\(^1\), whose main goal is allowing mobile and portable broadband connectivity to end users with different degrees of QoS, to adequately support future Internet applications, such as voice and video over IP.

I Introduction

The SUITED project has the challenging objective of providing end to end Quality of Service (QoS) guarantees to users. The reference network architecture is depicted in Fig. 1: it consists of a wireless portion, based on four different access segments (the EuroSkyWay satellite network, the UMTS network, the GPRS network, a Wireless LAN networks, based on the 802.11 standard), and a fixed segment, represented by a portion of the Internet, suitably upgraded with QoS support features (and named Federated ISPs). The former is able to provide QoS guarantees by means of the Integrated Services/RSVP (IntServ/RSVP) model [BRCS94][BZBH97], which allocates network resources by means of a per-flow management. The user mobile terminal is equipped with a network entity, called InterWorking Unit, responsible for mobility management and call set-up functions. The fixed network is assumed to be divided in two sections: an access segment, which adopts the IntServ/RSVP approach, and a core portion, which is based on the Differentiated Services (DiffServ) model [BBCD98], in order to better face scalability issues.

However, the DiffServ model is unable to provide hard QoS guarantees, as the IntServ model and as required within the SUITED project, due to the lack of an admission control scheme. In order to solve this problem, we enhance the DiffServ paradigm with admission control functions by means of the so-called GRIP (Gauge&Gate Reservation with Independent Probing) solution [BBM01][BB01].

Following the DiffServ paradigm, the GRIP algorithm does not manage traffic on a per-flow basis in core routers; traffic conditioning (and shaping) functions are executed at edge routers only. Since traffic is not reshaped in core routers, its statistical properties, imposed by the shaping at edge routers, could be altered, due to buffering within routers; this means that flows could not be correctly described, in the inner network stages, by the parameters declared at the network access points. In general, the statistical flow behaviour tends to become smoother when crossing a cascade of routers. However, there are some cases, even if not very likely, in which flows crossing core routers can become more bursty.

We analyse this alteration and focus our attention on the issue of flows description in the core network, in order to suitably dimension the GRIP router parameters and to verify which guarantees can be provided to flows by operating over traffic aggregates only, without per-flow management. To this end we make use of the so-called Network Calculus [LEBO98].

The paper is organized as follows. In Section II, we describe the core network architecture. In Section III, we describe the GRIP solution. In Section IV, we introduce some basic concepts relevant to source shaping, Network Calculus and the equivalent bandwidth characterization of user flows. In Section V, we define allocation rules in edge and core routers. Finally, in Section VI, conclusive remarks are reported.

\(^1\) The background of this work is the IST project SUITED, sponsored by the European Union. The effort of the project’s partners is gratefully acknowledged. This paper, however, does not implicitly and necessarily represent the opinion of the other project partners.
II Fixed Portion Network Architecture

We consider a DiffServ domain; routers are classified into edge and core routers by their position in the flow path. Edge routers are connected to edge routers of other domains. Core routers may receive traffic from both edge and other core routers.

The basic approach of DiffServ consists in managing traffic in core routers by applying different Per-Hop Behaviors (PHBs) [BBCD98]. PHBs are specified by a code, named DiffServ codepoint (DSCP), placed in a dedicated field in the header of IP datagrams [NICH98]. This way, traffic flows can be classified and aggregated into a small number of PHBs, thus avoiding scalability problems. Each link transports a number of traffic aggregates, corresponding to different services, including the best-effort service, some qualitative better than best-effort services, called Assured Services [HBWW99] in the DiffServ architecture, and high quality services, called Premium (or Expedited) Services [JANP99] in the DiffServ architecture. Therefore, the capacity of each link has to be managed by means of appropriate link sharing and scheduling techniques applied to the different traffic aggregates (e.g. the Class Based Queueing [FLJA93] or the well-known Weighted Fair Queueing).

Following the DiffServ model, traffic management complexity is pushed to the boundary of the network. In particular, traffic conditioning function (metering, marking, shaping, policing) are executed at edge routers only. We stress that we avoid reshaping in core routers.

We assume that Dual Leaky Buckets (DLBs) shape and regulate the traffic at the input ports of edge routers only.

Our aim is determining the maximum traffic load passing through each router such that the delay suffered within the core network is upper bounded by a desired value. This bound allows using the DiffServ region also as the core of a larger network supporting IntServ end-to-end. We assume, as often suggested in the literature [BERN00][HUST00], the adoption of a resource allocation protocol and of a Call Admission Control (CAC) procedure on top of the DiffServ architecture, i.e. the GRIP algorithm. This algorithm has the aim of solving the problem of congestion control, avoiding a potentially harsh degradation of service due to overload, and assuring end-to-end service guarantees. For instance, this allocation protocol could support and enhance Premium Services for delivering real time services with performance guarantees.

Moreover, we suppose that our domain implements the Multi Protocol Label Switching (MPLS) forwarding scheme [ROSE01]. This guarantees that all packets of the same flow pass through the same path. MPLS is also a fast forwarding scheme and allows traffic engineering.

III GRIP Solution

The resource allocation protocol for the DiffServ network is based on the introduction of the GRIP admission control scheme [BB01][BBM01]. This solution requires that each router/access device
supports a PHB defined in terms of service priority between two classes of packets (served by distinct logical queues): higher service priority to Active packets and lower service priority to Probing packets. The latter are delivered only when no Active packets are stored into the router high priority queue. The goal of Probing packets is determining the network state. Each router also implements a module, which measures the aggregate Active traffic (i.e. traffic that has already passed an admission control test) that it is handling. This module implements a Decision Criterion (DC), which drives the router to switch continuously from an ACCEPT state to a REJECT state and conversely. This way, it controls the Probing packet queue server. When the router is in the ACCEPT state, the Probing queue is served (with lower service priority). Instead, when the router switches to the REJECT state, it discards all the Probing packets contained in the Probing queue and blocks all new Probing packets arriving. In other words, the router acts as a gate for the probing flow, where the gate is opened or closed on the basis of traffic estimations. If the probing packets succeed in reaching the destination, the receiver responds to the source notifying the connection acceptance. If no response is received in a suitable set-up timeout, the set-up attempt is aborted. It is clear that the key of GRIP is the definition of a DC being able to provide performance guarantee; to simplify the definition of the DC, we assume that:

- each traffic source emission is regulated by a DLB;
- the sources are homogeneous, that is they are characterized by the same parameters of the DLBs;
- the sources are greedy, that is they do not waste tokens, otherwise the regulator force them to emit null packets.

GRIP estimates the average amount of traffic offered at each router's output link by means of a sliding window. During a window, each router counts the number of Active packets emitted by all the flows that are using a given link. By exploiting the DLB characterization of the traffic sources, we can calculate the minimum and the maximum number of packets emitted during the measurement window. Since no conjecture is made on the statistical properties of the emission process for each source, beyond the presence of DLB regulators, the distribution of the number of the emitted packets between these two extremes remains unknown. To provide a conservative estimate, the CAC scheme estimates the number of allocated flows according to the minimum emission pattern of a DLB. A new flow is accepted if the estimated number of admitted sources is lower than a theoretical maximum value \( K \), chosen by the network operator according to the desired performance levels.

In this CAC scheme, we include also a transient management procedure (the use of a stack variable, [BB01][BBM01]) to avoid activation of a number of flows higher than \( K \). This way, we provide strict guarantees in any operational condition.

In the following Section we define suitable rules for choosing \( K \) both in edge routers and in core routers (where we take into account the alteration suffered by flows after crossing upstream routers).

### IV Shaping Background and the Equivalent Bandwidth Concept

The traffic entering the domain is regulated by means of DLBs. A DLB implements jointly two Generic Cell Rate Algorithms. The regulated traffic is characterized by a set of parameters which consists of a token bucket size \( BTS \), a token (sustainable) rate \( rS \), a peak rate \( PS \), a minimum policed unit \( m \), and a maximum datagram size \( M \). The presence of the parameters \( m \) and \( M \) is related to the packetized nature of the exchanged traffic. In this paper, we adopt the fluid traffic model, which neglects such packetized nature and assume that the traffic is a continuous quantity. As a consequence, our DLB model is based on three traffic descriptors only: \( PS \), \( rS \), and \( BTS \). This approach is widely used in literature (e.g. [ELMI97]).

With this model, each information unit needs a token to be transmitted. Tokens arrive at the token buffer at the sustainable rate \( rS \). When there are available tokens inside the token buffer, they can be picked up by the incoming information units. Therefore, the output rate process \( U(t) \) is the minimum between the input rate \( I(t) \) and \( PS \). When the token buffer is empty the output rate cannot be higher than the token arrival rate \( rS \) (see Fig. 2). In this paper, the DLB parameters are used to determine the so-called equivalent buffer and equivalent bandwidth of the shaped flows. To this end, a number of models have been proposed in the literature (e.g. [ELMI97][LEBO98]). We follow the approach proposed in [LEBO98] since it is helpful in analysing the statistical modifications suffered by a flow as it passes through a cascade of routers. The model in [LEBO98] is based on the so-called Network Calculus. The Network Calculus is a set of rules and models of network entities that allows determining the performance of packet data networks, such as the end-to-end delay. It is based on the
**min-plus algebra**, that is a dioid algebra in turn based on the operations $\inf(a,b)$ and $a+b$, for any real numbers $a$ and $b$, where the $\inf$ denotes the inferior extreme operator. In the following we recall some basic concepts of the Network Calculus (see [LEBO98] and its references for details).

\[ R(t) \quad U(t) \]

**Fig. 2 – DLB equivalent model.**

The **min-plus convolution** for two functions $\gamma_1$ and $\gamma_2$ is a commutative and associative operation that may be expressed as $(\gamma_1 \otimes \gamma_2)(t) = \inf_{u \in [a,b]} \{ \gamma_1(u) + \gamma_2(t-u) \}$.

The most important concepts are the **arrival curve** and the **service curve**. Let $R(t)$ be the number of data units of a flow crossing a given interface in the interval $[0,t]$. An arrival curve $\alpha$ of that flow is a wide-sense increasing (or not decreasing) function such that for all $s \leq t$ it results that $R(t)-R(s) \leq \alpha(t-s)$. This condition is equivalent to the following relation: $R \leq R \otimes \alpha$. To define the service curve, we consider a system $S$ giving service to an incoming flow with an arrival process $R(t)$. Let $R^*(t)$ be the corresponding output function. A service curve $\beta$ of the system $S$ is a function such that for all $t \geq 0$, it is possible to determine a time $t_0 \leq t$ such that $R^*(t)-R(t_0) \geq \beta(t-t_0)$. This relation can be expressed as follows: $R^* \geq R \otimes \beta$.

Now let us consider a buffer of unlimited size being loaded by a shaped flow and serviced by an output trunk capacity $c$, whose service curve is $\beta(t) = ct$. If $c$ is higher than the average flow rate, the maximum buffer occupancy $b$, which is a function of the bandwidth variable $c$, that is $b = b(c)$, is the buffer space to be allocated to the flow, in a limited size buffer, to avoid losses. Clearly, the number of pairs $(b,c)$ is infinite. To select a particular pair, an additional relation has to be used. For example, if an upper bound $D_{\max}$ on the maximum queueing delay is required, we can select the pair $(b_0,c_0)$ that satisfies the following requirement: $b/c = D_{\max}$. This way, the pair $(b_0,c_0)$ represents the equivalent buffer and the equivalent bandwidth of the regulated flow in the lossless case.

We want to determine this pair with reference to a DLB shaped flow. By using the fluid traffic model, the DLB service curve depends on three parameters only, and it is given by $\alpha_{DLB}(t) = \inf_{t \geq 0} \{P_S t, B_{TS} + r_S t\}$. It is shown in [LEBO98] that the minimum value of the trunk capacity $c$ such that the delay experienced by the packets of the flow be not greater than $D_{\max}$ is $c_0 = \sup_{s \geq 0} \frac{\alpha_{DLB}(s)}{s + D_{\max}}$, that is

\[
\frac{P_S B_{TS}}{D_{\max} (P_S - r_S) + B_{TS}}.
\]

As shown in Fig. 3, $c_0$ is the slope of the tangent to the arrival curve, which intersects the time axis at $t = D_{\max}$. The intersection of the tangent with the ordinate axis gives the equivalent buffer size, that is $b_0 = c_0 D_{\max}$. This way, the pair $(b_0,c_0)$, corresponding to the delay $D_{\max}$, is determined. In Fig. 3 we highlight also the quantity $M_S$, that is the maximum burst size that a DLB may emit at the peak rate, that is $M_S = P_S B_{TS} / (P_S - r_S)$.

**V Resource Allocation Rules**

In this Section, we focus our attention on the choice of the number of allocable flows $K$, in the framework of the GRIP algorithm, by means of the Network Calculus technique. Our goal is providing a lossless service, while guaranteeing a maximum delay $D_{\max}$ to each information unit in each node. For the sake of simplicity, we assume that the routers’ output port speeds are all equal to the same
quantity, $C$. Since flows are shaped at edge routers only, we need to address separately edge and core routers.

![Fig. 3 – Derivation of the equivalent bandwidth and equivalent buffer.](image)

**V.I Edge Routers**

We assume that a set of regulated sources, each described by the DLB parameters $(r_s, P_s, B_{TS})$, are sending traffic to an edge router. For the sake of simplicity we assume that the regulated sources are homogeneous, that is they have the same DLB parameters. The router performs, among other things, a statistical multiplexing of all input flows by means of a buffer of size $B$ served at the output link capacity $C$. Since we consider a single traffic class (homogeneous sources) the queueing discipline is FCFS. Since the router is supposed to guarantee a maximum delay $D_{\text{max}}$ to each information unit, the maximum number $K$ of allocable flows in edge routers is equal to $K_{\text{edge}} = \min\left(\frac{B}{b_0}, \left\lfloor \frac{C}{c_0} \right\rfloor \right)$. In the following, we will assume that a buffer space $B \geq CD_{\text{max}}$ is available, therefore $\frac{B}{b_0} \geq \frac{C}{c_0}$ and the allocation rules will be based on the available bandwidth only. This way

$$K_{\text{edge}} = \left\lfloor \frac{C}{c_0} \right\rfloor + \left\lfloor \frac{D_{\text{max}} C (P_s - r_s)}{P_s B_{TS}} \right\rfloor. \quad (2)$$

Note the increase of the number of the allocable flows with respect to the peak allocation strategy.

**V.II Core Routers**

Now we move our attention from edge routers to core routers. First of all, we have to consider that buffering at edge routers may cause statistical variations of the rate process generated by the DLBs. In most cases, the flow rate processes tend to become smoother at downstream routers. Nevertheless, since we want to guarantee the desired performance levels, we have to consider also the unlikely cases in which a flow passing through a router gets more bursty and then its statistical behaviour worsen. Our goal is determining the worst case flow descriptors in the core network.

When information units experience queueing delay, it may happen that a number of information units of a flow are buffered one after the other. In this case, all of them are transmitted as a single burst at rate $C$. This means that the equivalent bandwidth of the flow may increase. We are interested in the case in which the number of consecutive buffered information units of a reference flow is maximized. This is a very unlucky situation, since for it to happen, it is necessary that the rate processes of all the DLBs be ON/OFF shaped and time-aligned so that the buffer occupancy is maximized. The information units of the reference flow are buffered one after the other, if such source starts transmitting immediately after all the others stop.

It is shown in [Tech_R] that, if flows are ON/OFF shaped at edge routers and if the delay jitter is such that different bursts do not join themselves along the path, the worst case flow description at the input of core routers is given by the following DLB parameters: $(r_s, C, M_S (C - r_s)/C)$, instead of the original ones declared at network border. Note that the assumption that the output link speeds are identical preserves the homogeneity of the flows also in the core network. This way, the equivalent bandwidth of the flow $C_0^*$ in core routers is given by (1) with updated values of peak rate ($P_s \rightarrow C$)
and burst tolerance \( B_{TS} \rightarrow M_S (C - r_S)/C \). Therefore the maximum number \( K \) of allocable flows in core routers, under the same delay constraint of the edge ones, is equal to
\[
K_{\text{core}} = \left[ \frac{C}{c_0} \right] = 1 + \left[ \frac{D_{\max} C (P_S - r_S)}{P_S B_{TS}} \right].
\]

We define the \textit{pejorative factor of utilization} as the value \( \xi = K_{\text{edge}}/K_{\text{core}} \), that is the decrease of the maximum number of allocable flows in core routers with respect to edge ones. It follows that
\[
\xi = \frac{K_{\text{edge}}}{K_{\text{core}}} = \frac{c_0^*}{c_0} = 1 + \frac{B_{TS} (C - P_S)}{CD_{\max} (P_S - r_S) + P_S B_{TS}}.
\]

This is the price, in terms of resource utilization, that must be paid to avoid reshaping at core routers. For further details please refer to [Tech_R].

VI Conclusions

We have considered the hybrid IntServ-DiffServ QoS approach of the SUITED project, and we have focused our attention on the DiffServ portion of the network. We have described the adopted shaping model and the GRIP solution (able to provide end-to-end performance guarantees to traffic aggregates). Moreover, since the DiffServ model provides flows shaping at access points only, we have studied the flows behaviour in the core network, so as to define suitable allocation rules in core routers, implementing the GRIP algorithm. This way, our approach is able to provide hard guarantees to traffic flows both in the edge and core routers, without reshaping flows within the DiffServ domain; the price to pay is an under-utilization of network resources.

References