Internet Engineering Task Force Flaminio Borgonovo
INTERNET-DRAFT Antonio Capone
Expires: February 1999 Luigi Fratta
Mario Marchese
Chiara Petrioli
Politecnico di Milano
29 July 1998

End-to-end QoS provisioning mechanism for Differentiated Services

<draft-borgonovo-qos-ds-00.txt>

Status of this Memo

Abstract

This document presents an end-to-end mechanism to guarantee bandwidth and delay into the Differentiated Services mechanism to constant rate traffic such as voice and video. The mechanism requires network routers to be able to serve packets according to three classes of priority. The needed call admission control is performed by an end-to-end signaling procedure that implicitly looks for the required bandwidth and seizes it, if available. Short delays are guaranteed by the regular structure of constant rate traffic. No entities other than source and destination are involved and multicast operation comes at no further cost, which makes the mechanism fully scalable and integrable into the existing Internet.

1. Introduction

The transport mechanism capable of guaranteeing demanding Quality of Service (QoS) requirements is under discussion in the INTERNET community. The Differentiated Services architecture [1] is the approach that has recently gained more credit. The goal is to offer an alternative to carry voice, video and multimedia with respect to classic Telephone/ISDN and ATM networks. The basic problem is how to guarantee bandwidth, delay and packet dropping probability, in a datagram network architecture where the only service is the Best Effort packet transmission.

In this contribution we consider only approaches that can be completely implemented at the IP level and do not assume any lower level QoS guarantee capability. In this context, an existing approach is represented by RSVP (Resource reSerVation Protocol) [2][3].

RSVP is a signaling mechanism among routers and hosts that includes support to ``flows'' of packets with different QoS and the ability to dedicate end-to-end capacity to real-time traffic by means of hop-by-hop resource reservation protocols. In practice, this solution changes the entire network architecture by relying on the virtual circuit connection mechanism, the paradigm of the telephony world, today extended to the B-ISDN. During the set-up phase, needed to install a virtual circuit, the network nodes cooperate, using complex protocols, to determine a path within the network and to reserve network resources such as bandwidth and buffers. The implementation of such a signaling and its related features over the layered structure of a pure datagram network will require large investments because of the heavy software and hardware modifications to be introduced in the already worldwide installed networks.

A much simpler alternative is represented by the Differentiated Services approach. The basic idea is to use the IPv4 header TOS bits or the IPv6 Traffic Class octet, the "DS byte" to designate the "per-hop behaviors" that packets are to receive. By carefully aggregating a multitude of QoS-enabled flows into a reasonable number of differentiated services offered by the network it is no longer necessary to recognize and store information about each individual flow in the core routers. Though some signaling mechanism is needed to manage the service assignment to individual flows, the network mechanism still remains purely datagram and scales well. However, since the control on packets is performed hop-by-hop, it is not easy to design a suitable call acceptance policy that is able to guarantee end-to-end QoS.

As the result of our research work in this field we have gained the belief that if one only wants to enforce QoS control over constant-rate or almost-constant-rate streams, simple procedures based on non-preemptive priority mechanisms and simple end-to-end signaling procedures are effective. One such procedure has been investigated in the framework of deflection networks, i.e., data networks with very small or no queuing at nodes [4].

The Bandwidth Guaranteed Service (BGS) mechanism, presented in this document, guarantees constant rate traffic bandwidth and delay requirements in datagram networks such as the Internet. BGS is based on three priority levels. It integrates and completes the DS approach and allows a QoS control as in RSVP, without adding explicit signaling protocols. No entities other than source and destination are involved and multicast operation is obtained at no further cost, which makes the mechanism fully scalable.

The document is organized as follows. In Section 2 we present BGS, its call setup and call acceptance procedures, while Section 3 illustrates its behavior by means of some preliminary results obtained by simulation.

2. The BGS mechanism

In the following we assume connections with constant rate packet traffic. The jitter introduced by the network is eliminated at the destination user by storing the packets in a play-out buffer which is emptied at the constant nominal rate, starting with a delay D_b after the reception of the first packet of the connection. With this technique the total delivery delay experienced by packets is constant and equal to

$$W = D_b+d_1 \qquad (1)$$

where d_1 is the network delay suffered by the first packet. If a packet is delayed more than W, it is useless and is dropped. Therefore the QoS guaranteed traffic requires, besides the bandwidth B, also a maximum packet delay W and a maximum packet-dropping percentage gamma.

The BGS mechanism requires that each packet in the network belongs to one of the three following priority classes.

Class 0: (lowest priority) if the packet requests best effort service;

Class 1: (intermediate priority) if the packet is a scout packet, used in the set-up procedure as defined below;

Class 2: (highest priority) if the packet belongs to a constant ate flow that has been granted guaranteed service.

The priority information is carried in the TOS field and is used by the routers to serve all packets according to a non-preemptive head-of-the-line priority scheme. Class 1 and 2 packets have the same constant length.

The set-up procedure operates end-to-end and is activated when a new connection, characterized by the bandwidth B and the maximum allowed transfer delay W, is requested.

Upon reception of a connection request C_j addressed to NODE_B, NODE_A immediately starts transmitting scout packets at the rate r_j corresponding to the requested bandwidth.

The scout packets do not carry data traffic in the payload, since they are used to perform "bandwidth scouting and seizing" within the network. To this purpose they only include set-up information to signal to the receiving NODE_B the request for an incoming call and the related service quality parameters.

Upon receiving the first scout packet, NODE_B starts performing bandwidth measures and delay evaluations to verify whether QoS requirements are met. If the outcome is positive NODE_B sends a call acceptance packet to NODE_A. Otherwise, either it rejects the connection or starts a bandwidth negotiation procedure with NODE_A.

As far as NODE_A is concerned, it keeps on transmitting scout packets until either a time-out occurs or a response from NODE_B is received. If the time out expires or the response is negative the call is aborted, meaning that the bandwidth has not been found. If a positive response is received, the connection set-up phase ends and NODE_A replaces scout packets with data packets. The bandwidth is automatically released when packets transmission ends.

The priority scheme adopted guarantees that new connections can not steal bandwidth to already established ones. In fact, if some constant bandwidth connections have already been admitted, as soon as new connections attempt to steal bandwidth on a link, old connections are delayed, a queue builds up and the priority mechanism cuts out the newly arrived scout packets.

2.1 Call acceptance procedure

The call acceptance procedure is directly performed by the destination node and is very simple: a new guaranteed-bandwidth connection is accepted only if the connection parameters measured on the N scout packets satisfy the QoS constraints.

First, a test of significance on the Hypothesis H0 that the bandwidth requirement is satisfied is performed on the sample X_1, X_2, ..., X_(N-1), where X_i is the interarrival time between scout packet i and i+1. Under the hypothesis H0 we have E[X]=T, being $1/T$ the packet constant rate of each flow. So, the test can be exploited on the sample average

M_X = (X_1+X_2+...+X_(N-1))/(N-1) (2)

which is assumed to be normally distributed with average T and variance

S2_x =[(X_1-T)*(X_1-T)+...+(X_(N-1)-T)*(X_(N-1)-T)]/(N-1) (3)

So, for a given confidence level alpha, a threshold T_alpha exists such that if M_X < T_alpha the hypothesis H_0 is accepted. The threshold value can be obtained as

T_alpha = T + xi_alpha * sqrt(S2_x/(N-1))   (4) where xi_alpha is the standard normal deviate that is exceeded with probability 1 - alpha. The difference

Delta B = 1/T - 1/T_alfa (5)

represents the resolution power of the measure.

The precision of the estimate increases with the number of statistically independent samples. In a network that carries periodic packet streams, delay samples can be considered independent if taken at a distance larger than the transmission period T_max of the connection with the lowest admissible bandwidth. Thus, the set-up period is determined in time rather than by the number of signaling packets. By doing this we guarantee that the measure captures all the existing traffics, including the slowest.

For example, if we assume that the lowest bandwidth corresponds to 32 Kb/s with 640 bit packets, the packet transmission time is 20 ms so that a measure period of 1-2 seconds is needed to collect a reasonable number (50-100) of samples.

The measure indicated above may be critical since, due to the error (5), the connection could be accepted even if the required bandwidth slightly exceeds the available one. In this case, more traffic than the capacity is accepted and all connections would be affected. A simple way to avoid this unwanted phenomenon is to scout for more bandwidth than needed in the set-up phase, and to turn to the required bandwidth once the call is accepted. With this modification links can not be used up to their capacity, but the unused fraction can be kept very small.

2.2 The delay issue

Since packets are dropped when their delay overcame the required threshold, it is important, for the effectiveness of the scheme, to derive some conservative peak delay estimate to be used at the set up phase. To this purpose we use an nD/D/1 queuing model where the n sources generate service requests at a constant inter-arrival time equal to T [5][6].

To obtain a conservative estimate under any network load condition, we evaluate the delay suffered when the utilization factor is close to one.

Using the approach in [5], we have considered three cases in which the channel capacity is M = 25,50,100 connections of the same rate. The average waiting plus transmission delay suffered by the packets of a connection when all connections are active is shown in Table I. The delays are expressed in transmission time units (m), in interarrival time units (m'), in milliseconds (m'') assuming T=30 ms (which can model 32 Kb/s voice channels and 1000 bit packets), and in milliseconds (m''') assuming T=1 ms (which can model 1 Mb/s channels and 1000 bit packets).

Table I.

M	m(D)	s(D)	m'(T)	m"(ms) (T=30 ms)	m"'(ms) (T=1 ms)
25	3.51	1.52	0.140	4.21	0.140
50	4.88	2.09	0.0977	2.93	0.0977
100	6.85	2.79	0.0685	2.05	0.0685

Column two shows that, for any link bandwidth, absolute delays decrease when the size of connections increases. For the case in which sources have different, though constant, rates no general solution exists. However, it is expected that the replacing of some lower rate connections with an equivalent high rate connection will not impair performance since the high rate connection generates a more regular arrival pattern than the slower connections replaced. This conjecture is substantiated by our simulation results, therefore we assume that in an environment with mixed rate sources a conservative delay bound is attained by assuming that all connections are of the smallest allowed rate.

Column four shows that if delay is measured in interarrival times, the delay decreases as the number of connections increases. This means that if a lower bound to the capacity of the path is used for delay evaluations, an upper bound is guaranteed.

Finally, a conservative estimate on the connection end-to-end delay can be attained by assuming that delays add hop-by-hop as independent random variables. Also this assumption is conservative as the pipeline effect along the connection path reduces the queuing delay at nodes other than the first. Thus, the peak delay evaluation of a connection with k hops can be attained assuming a normal distribution for the total delay.

Table II shows three examples of peak delay evaluations using the values in Table I. The first two are based on a minimum allowed rate of 33 packet/sec (e.g. voice channels at 32 kb/s with 30 ms interarrival time), and assuming that at least either 100 or 25 connections can be accommodated along an 8 hop path.

The third is based on a minimum allowed rate of 1 Mb/s with 1 ms interarrival time, representing the case for voice trunk channels within a provider domain, and assuming that at least 25 connections can be accommodated along an 8 hop path.

The results shown refer only to class 2 traffic. Class 1 traffic (scout packets) has very little influence on the delay since it can delay class 2 packets for at most one transmission time, because of the non-preemptive priority. Class 0 packets may alter the analysis shown if they are allowed a size considerably larger than the constant rate packets.

Table II.

rate capacity	mean delay (ms)	Standard deviation (ms)	0.999 percentile (ms)
32 kb/s (100)	16.4	1.58	21.2
32 kb/s (25)	33.7	5.17	49.3
1 Mb/s (25)	1.12	0.172	1.64

The set-up procedure uses the values indicated in Table I to evaluate the 0.999 percentile, which depends on the number of hops traversed, and to determine whether the delay QoS is met.

Note, however, that due to the strong correlation among packets of the same connection, the fraction 0.001 does not represent the packet dropping probability suffered by any connection, but it represents the fraction of connections that suffer loss of packets.

3. Measures

Preliminary simulation results have been obtained by simulating an 8 node network, interconnected by a unidirectional and homogeneous ring structure. Equal rate traffic is generated at any nodes, and destination nodes are chosen either 1 or 8 hops apart in the ratio 12 to 1, so that at any node a consistent fraction of traffic (about 60%) is renewed. With this structure the complexity of the simulation environment is kept low while observing sufficiently long paths with limited pipeline effect, a most critical condition.

Table III.

connection rate	pck.dropping classes	delay thresholds (ms)						average delay
1 Mb/s		1	1.1	1.2	1.3	1.4	1.5	1.11
32 Kb/s		30	33	36	39	42	45	33.3
	0 - 0.001	0	0.084	0.167	0.417	0.75	1	
	0.001-0.01	0	0	0	0	0	0	
	0.01 - 0.1	0	0	0	0	0	0	

	0.1 - 1	1	0.916	0.833	0.583	0.25	0	

In our simulations the network has been loaded one connection at a time, until saturation is reached, using the call set-up mechanism described in the previous sections. In Table III we report, for a few delay thresholds, the percentage of 8 hop connections that experienced a packet dropping probability (i.e. a delay greater than the threshold) within the class indicated in the first column. The average delay is reported in the last column. The capacity of the links is assumed 25 times the bandwidth required by each connection.

The bimodal behavior of the distribution in Table III confirms that a connection is either good or bad. Furthermore, the delay threshold with no packet dropping is within the bound given in Table II.

4. References

[1] K. Nichols, V. Jacobson, L. Zhang, ``A Two-bit Differentiated Services Architecture for the Internet'', IETF Internet Draft, Nov, 1997.

[2] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, ``Resource ReSerVation Protocol (RSVP)-Version 1 Functional Specification''IETF Request For Comments 2205, Sep. 1997.

[3] P.P White, ``RSVP and Integrated Services in the Internet: A Tutorial'', IEEE Communication Magazine, vol. 35, no. 5, May 1997,pag. 100.

[4] F. Borgonovo and L. Fratta, `` Deflection Networks: Architectures for Metropolitan and Wide Area Networks'', Computer Networks and ISDN Systems, Vol. 24, No. 2, April 1992, pp.171-183.

[5] A. E. Eckberg, "The single server queue with periodic arrival process and deterministic service time", IEEE Trans. on Comm., Vol. COM-27, pp. 556-562, 1979.

[6] G. Ramamurthy, B. Sengupta, "Delay analysis of the packet voice multiplexer by the SD_i/D/1 queue", IEEE Trans. on Comm., Vol. COM-39, no. 7, July 1991.

5. Authors' Addresses

Flaminio Borgonovo
Dipartimento di Elettronica e Informazione
Politecnico di Milano
P.zza L. da Vinci 32,
20133 MILANO, Italy
Email: borgonov@elet.polimi.it
Fax: +39 02 2399 3413
http://www.elet.polimi.it/people/borgonov/

Luigi Fratta
Dipartimento di Elettronica e Informazione
Politecnico di Milano
P.zza L. da Vinci 32,
20133 MILANO, Italy
Email: fratta@elet.polimi.it
Fax: +39 02 2399 3413
http://www.elet.polimi.it/people/fratta/

Antonio Capone
Dipartimento di Elettronica e Informazione
Politecnico di Milano
P.zza L. da Vinci 32,
20133 MILANO, Italy
Email: capone@elet.polimi.it
Fax: +39 02 2399 3413
http://www.elet.polimi.it/people/capone/

Mario Marchese
Dipartimento di Elettronica e Informazione
Politecnico di Milano
P.zza L. da Vinci 32,
20133 MILANO, Italy
Email: mmarches@elet.polimi.it
Fax: +39 02 2399 3413

Chiara Petrioli
Dipartimento di Elettronica e Informazione
Politecnico di Milano
P.zza L. da Vinci 32,
20133 MILANO, Italy
Email: chiara@cerbero.elet.polimi.it
Fax: +39 02 2399 3413