Internet Engineering Task Force
INTERNET-DRAFT
draft-heinanen-pbe-svc-01.txt

Integrated Services WG
Juha Heinanen
Telecom Finland Inc.
February 17, 1996

Protected Best Effort Service

Status of this Memo

This draft has been submitted to the Integrated Services Working Group of the Internet Engineering Task Force. Comments are solicited and should be addressed to the working group's mailing list at int-serv@isi.edu and/or the author.

Abstract

This document specifies the Protected Best Effort (PBE) service that provides to the application an isolated best effort flow with a possible minimum bandwidth.

Introduction

This document defines the requirements for network elements that support the Protected Best Effort (PBE) service. The end-to-end behavior provided to an application by a series of network elements conforming to this specification is

- fair share of the bandwidth available for PBE traffic along the path of the application's data flow or at least a minimum bandwidth whichever is larger,

- fair transit delay among the applications whose PBE data flows share the same network elements, and

- ordered delivery of packets as long as the network path of the application's PBE data flow doesn't change.

To ensure that the above behaviour can be met, an application requesting PBE service provides the network elements with a Minimum Requested Bandwidth (MRB). A network element that correctly implements the PBE service, doesn't need to support MRB values > 0.

The application may not assume succesful delivery of packets sent at a rate which is higher than the MRB, since there may not be more bandwidth available on the path for PBE traffic. Further, the application may not assume any given upper bound for transit delay or jitter, since the number of PBE data flows that share the same network path can change dynamically.

Motivation

The PBE service is intented to support any non-realtime applications that are developed for use in today's internets. In particular, the PBE service is suitable for applications that don't require a given maximum delay or delay variation and that can't characterize their traffic requirements in terms of average data rates and maximum burst lengths. The PBE service differs from the conventional (= unprotected) best effort service in that it guarantees bandwidth and delay fairness by isolating the PBE data flows from each other. If all the network elements along the path of the application's data flow support a greater than zero MRB, then the PBE service can also be used to provide an upper bound for the time it takes to transfer the information across the network.

Network Element Data Handling Requirements

Each network element accepting a request for a PBE service must ensure that adequate network element and link resources are available to handle the requested level of traffic as given by the MRB in the requestor's TSpec. This must be accomplished through active admission control. A network element may employ statistical methods to decide whether adequate resources are available to accept the service request.

Links are not permitted to fragment packets which receive the PBE service. Packets larger than the MTU of the link must be treated as nonconformant to the TSpec. This implies that they will be policed according to the rules described in the Policing section below.

The PBE service is invoked by specifying the data flow's requested traffic parameters (TSpec) to the network element. Requests placed for a new flow will be accepted if the network element has the capacity to forward the flow's packets as described above. Requests to change the TSpec for an existing flow should be treated as a new invocation in the sense that admission control must be reapplied to the flow. Requests that reduce the TSpec for an existing flow (in the sense that the new TSpec is strictly smaller than the old TSpec according to the ordering rules given below) should never be denied service.

The TSpec consists of a rate (MRB), a minimum policed unit (m), and a maximum packet size (M). MRB is measured in bytes of IP datagrams per second. Values of this parameter may range from 1 byte per second to 40 terabytes per second. Network elements must return an error for requests containing values outside this range. Network elements must return an error for any request containing a value within this range which cannot be supported by the element. In practice, only the first few digits of the MRB parameter are significant, so the use of floating point representations, accurate to at least 0.1% is encouraged.

The range of values allowed for these parameters is intentionally large to allow for future network technologies. Any given network element is not expected to support the full range of values.

The minimum policed unit, m, is an integer measured in bytes. All IP datagrams less than size m will be counted against MRB as being of size m. The maximum packet size, M, is the biggest packet that will conform to the traffic specification; it is also measured in bytes. Network elements must reject a service request if the requested maximum packet size is larger than the MTU of the link. Both m and M must be positive, and m must be less then or equal to M.

The preferred concrete representation for the TSpec is one floating point number in single-precision IEEE floating point format followed by two 32-bit integers in network byte order. The first value is the rate (MRB), the second value is the minimum policed unit (m), and the third is the maximum packet size (M).

Exported Information

The PBE service is assigned service_name X.

The PBE service has no required characterization parameters. Specific implementations may export appropriate measurement and monitoring information.

Policing

PBE traffic must be policed for conformance at each network element. All PBE traffic that exceeds either the flow's fair share of the network element's best effort resources or MRB*T over all time periods T (whichever is larger) is considered nonconformant. For the purposes of this accounting, network elements must count packets that are smaller than the minimal policing unit to be of size m.

The fair share of the flow is determined according to the max-min criteria [1-2]. In determining the number of best effort flows competing on the network element's resource, all unprotected best effort traffic using that resource must be counted as one single flow. The MRB of this flow is network element specific, i.e., the manager of the network element may choose to allocate some amount of bandwidth for unprotected best effort traffic.

A network element may discard nonconforming PBE packets, but should not do so as long as it has enough buffering resources available to protect conforming PBE flows, i.e., flows that are currently using less than their share of the network element's resources. Conversely, the network element must discard nonconforming packets if buffering resources would not otherwise be adequate to accommodate new packets belonging to conforming PBE flows.

At all network elements, packets bigger than the outgoing link MTU are considered nonconformant and must be discarded.

(I would like to add to the TSpec Peak Rate parameter that tells the maximum bandwidth ever possibly available for PBE traffic on the current path and then police all PBE traffic also according to this parameter. It doesn't namely make sense to allow traffic to the network that possibly can't get through at later network elements. The value of the Peak Rate parameter is determined using the path message.)

Ordering and Merging

The PBE service TSpec is ordered according to the following rule: TSpec A is a substitute for ("as good or better than") TSpec B if and only if

(1) MRB for TSpec A is greater than or equal to those of TSpec B,

(2) the minimum policed unit m is at least as small for TSpec A as it is for TSpec B, and

(3) the maximum packet size M is at least as large for TSpec A as it

is for TSpec B.

A merged TSpec may be calculated over a set of TSpecs by taking the largest MRB, smallest minimal policed unit, and largest maximum packet size across all members of the set. This use of the word "merging" is similar to that in the RSVP protocol; a merged TSpec is one that is adequate to describe the traffic from any one of a number of flows.

The sum of n PBE service TSpecs is used when computing the TSpec for a shared reservation of n flows. It is computed by taking:

- The minimum across all TSpecs of the minimum policed unit parameter m.

- The maximum across all TSpecs of the maximum packet size parameter M.

- The sum across all TSpecs of the parameter MRB.

The perfect minimum of two TSpecs is defined as a TSpec which would view as compliant any traffic flow that complied with both of the original TSpecs, but would reject any flow that was non-compliant with at least one of the original TSpecs. This perfect minimum can be computed only when the two original TSpecs are ordered, in the sense described above.

A definition for computing the minimum of two unordered TSpecs is:

- The minimum of the minimum policed units m.

- The maximum of the maximum packet sizes M.

- The minimum of the rates MRB.

Guidelines for Implementors

Network elements providing the PBE service are permitted to oversubscribe the available resources to some extent, in the sense that the MRB requirements indicated by summing the TSpec MRB values of all PBE flows may exceed the maximum capabilities of the network element. However, this oversubscription may only be done in cases where the element is quite sure that actual utilization is far less than the sum of the MRBs would suggest. The most conservative approach, rejection of new flows whenever the addition of their traffic would cause the sum of the MRBs to exceed the capacity of the network element, may be appropriate in other circumstances.

Evaluation Criteria

The quality of a PBE implementation can be evaluated by measuring the degree of bandwidth and delay fairness it provides to PBE flows sharing the same network path. In addition, it should be measured that the bandwidth available to a PBE flow is always at least MRB and that unprotected best effort traffic can't get more resources than a single PBE flow can get.

A bandwidth fairness is commonly measured using the so-called parking lot configuration. The service is bandwidth fair if each flow receives an equal amount of bandwidth no matter how many network elements it has traversed.

Delay fairness can be evaluated by measuring that the delay of a packets only depends on the number of the simultaneously active flows on the path. The delay should increase linearly when new flows become active, but should not depend on the bandwidth of each flow.

(A more detailed description of the evaluation criteria can be provided later.)

Examples of Implementation

One possible implementation of the PBE service uses a fair queuing mechanism where each PBE flow and all unprotected best efforts flows together have their own queues that are serviced in a round robin fashion or at least so often that the flow can be guaranteed its MRB.

Security Considerations

Security considerations are not discussed in this memo.

Acknowledgements

This specification was created using the Controlled-Load Network Element Service specification [3] as the template.

References

[1] Bertsekas, D., and R. Galager, Data Networks, Prentice Hall, 1987.

[2] Jaffe, J., Bottleneck Flow Control, IEEE Trans. Communications, July 1981.

[3] Wroclawski, J., Specification of the Controlled-Load Network Element Service. Internet draft, November, 1995.

Authors' Addresses

Juha Heinanen
Telecom Finland Inc.
PO Box 228
FI-33101 Tampere
Finland

Phone: +358 400 500 958
Email: jh@lohi.dat.tele.fi