INTERNET DRAFT
Internet Engineering Task Force
Differentiated Services Working Group
Expires January, 2002

N. Seddigh
B. Nandy
Tropic Networks
J. Heinanen
Song Networks
July, 2001

An Assured Rate Per-Domain Behaviour for Differentiated Services

<draft-ietf-diffserv-pdb-ar-01.txt>

Status of this Memo

Abstract

This document describes a diffserv per-domain behaviour (PDB) called Assured Rate (AR). The AR PDB is suitable for carrying traffic aggregates that require rate assurance but do not require delay and jitter bounds. The traffic aggregate will also have the opportunity to obtain excess bandwidth beyond the assured rate. The PDB can be created using the diffserv AF PHB along with suitable policers at the domain ingress nodes.

1.0 Description of the Assured Rate PDB

This document defines a differentiated services per-domain behaviour (PDB) suitable for traffic aggregates that require rate assurance. This PDB ensures that traffic conforming to a committed information rate (CIR) will incur low drop probability. The aggregate will have the opportunity of obtaining excess bandwidth beyond the CIR but

there is no assurance. In addition to the CIR, the edge rules may also include other traffic parameters such as the peak information

rate (PIR) to place additional constraints for packets to which the assurance applies or to further differentiate packets which exceed the CIR.

This PDB is referred to as the Assured Rate (AR) PDB and is defined in accordance with the guidelines in [PDBDEF].

It may be possible to determine delay and jitter bounds for traffic aggregates using the AR PDB. However, such parameters are beyond the scope of this PDB definition and no attempt is made to characterize them. Development of a mathematical model to predict delay and jitter for the AR PDB is left as a subject of future research and investigation.

The PDB tries to avoid packet reordering within microflows. The PDB is applicable for a one-to-one, one-to-few as well as one-to-any types of service.

This document uses "one-to-one" to describe a traffic aggregate entering via a single ingress point of a domain and exiting from a single egress point for the domain. One-to-any refers to a traffic aggregate with single entry point and multiple (any) exit points in the domain. One-to-few refers to a traffic aggregate with single ingress point and fixed set of egress points within a domain.

The possibility of obtaining excess bandwidth allows development of various novel SLA models. For example, excess bandwidth is charged at a higher rate than assured bandwidth; excess bandwidth is cheaper than assured bandwidth; excess bandwith is charged proportionally etc. Development and discussion of such service and charging models are beyond the scope of this document.

2.0 Applicability

The Assured Rate PDB is intended to carry traffic aggregates that require assurance for a specific bandwidth level.

This document does not restrict the PDB to any particular application or traffic type. Regardless of the traffic mix, the CIR for the aggregate will be assured.

However, it is also possible to use this PDB to create a service for an aggregate consisting only of TCP microflows or non-responsive UDP microflows. The provider may wish to create a TCP-only service for a variety of reasons such as traffic isolation, better treatment of individual short microflows within an aggregate, greater fairness among TCP and UDP microflows access to the excess bandwidth allowed for the aggregate. Such service definitions are outside the scope of

this document. They are mentioned here simply to show that the PDB can be used to create diverse services.

The governing attributes of the PDB are only expressed in relation to the entire traffic aggregate. The PDB specification does not specify

any attributes for the individual microflows within an aggregate.

The grouping of microflows into the traffic aggregate can be done either at the customer site or by the provider's ingress router on behalf of the customer. The AR PDB definition can be used in either scenario. It is the responsibility of the service provider to specify which approach is adopted in the service level specification (SLS).

3.0 Technical Specification

The specification for this PDB consist of two parts:

1. A set of Edge rules that classifies packets arriving at the domain ingress into a traffic aggregate, performs metering/policing on the aggregate and associates a packet marking with the aggregate. Traffic shaping does not need to be performed on the aggregate as it enters the domain.

2. Per-node PHB treatment for the traffic aggregate as it weaves its way from the domain ingress to the domain egress.

3.1 Edge Rules

As packets enter the domain they will be classified into a traffic aggregate based on the specified filter at the domain ingress interface of the border router. The filter MUST be associated with a traffic profile that specifies committed information rate (CIR) AND a description on how it is to be measured. For example, the measurement may be based on a committed burst size (CBS) or an averaging time interval (T1).

The traffic profile MAY also include other traffic parameters. These parameters MAY place additional constraints on packets to which the assurance applies or MAY further differentiate traffic that exceeds the CIR.

Such parameters could include: peak information rate (PIR), peak burst size (PBS), excess burst size (EBS) or even a second averaging time interval (T2).

The policer causes each packet arriving into the domain to be marked with one of up to three levels of drop precedence, which we call (in the increasing order) green, yellow, red. The packets to which the assurance applies, MUST be marked green. The excess packets MUST be marked as either yellow or red. The details of packet colouring are

dependent on the specific policer utilized at the ingress router.

Red colour packets SHOULD be delivered with equal or lower probability than yellow colour packets. A special case of this is that all red colour packets are discarded by the ingress policer.

Yellow packets SHOULD not be dropped by the ingress policer. They MAY be dropped by the buffer management mechanisms of the ingress router but that will be due to PHB treatment.

The green, yellow and red packets MUST be marked with the DSCP for AFx1, AFx2 and AFx3 PHBs respectively, where x MUST be any one value from 1 to N. N is the number of AF classes supported by the routers in the domain.

The service provider may utilize any policer algorithm to colour the packets as long as it adheres to the general colouring principles outlined above. Examples of such policers include [SRTCM] [TRTCM] or [TSWTCM].

3.2 PHB Configuration

As described above, the AR traffic aggregate is to be treated using PHBs AFx1, AFx2 and AFx3 from a single AF class x. The resultant combination of the edge rules and PHB treatment within a single AF class, will ensure that:

"Within each AF class IP packets are marked (again by the customer or the provider DS domain) with one of three possible drop precedence values. In case of congestion, the drop precedence of a packet determines the relative importance of the packet within the AF class. A congested DS node tries to protect packets with a lower drop precedence value from being lost by preferably discarding packets with a higher drop precedence value." (taken from RFC 2597).

The requirement to achieve the PDB is as follows:

Nodes internal to the domain SHOULD not drop packets marked to receive treatment with AFx1. Under exceptional circumstances, network nodes MAY have to drop AFx1 packets for a short period. In such cases, they should only start dropping AFx1 packets after they have started dropping all AFx2 and AFx3 packets. See [TON98] for an example and justification of this approach. In the case where the AF class is lightly loaded, AFx2 and/or AFx3 packets MAY also be transmitted successfully through the node. This will allow the aggregate to obtain excess bandwidth beyond its assured rate.

As mentioned previously, any of the N AF classes may be selected to treat this PDB. This makes it possible to create in a single DS domain, multiple instances of the AR PDB, each with their own minimum forwarding resources. However, all aggregates using the same instance of the PDB in a single domain SHOULD utilize the same AF class PHB set.

4.0 Attributes of AR PDB

The attributes of this PDB include a rate that is assured and low drop probability for the traffic conformant to this rate.

5.0 Parameters

The AR PDB MUST have the following parameters:

- A committed information rate (CIR) that is assured with high probability. The AR PDB specification does not define "high" quantitatively, but an SLS MAY do so.

- Traffic parameters that are needed to measure CIR. The AR PDB specification does not define these parameters, since they depend on the policer used. Examples include a Committed Burst Size (CBS) and an averaging interval (T1).

- A maximum packet size for the aggregate - MAX_PACKET_SIZE.

In addition to the above, the AR PDB MAY have other, optional traffic parameters. These parameters MAY place further constraints on the packets to which the assurance applies or MAY further differentiate packets to which the assurance does not apply. The PDB does not define these parameters, since they depend on the policer used. Examples include a Peak Information Rate (PIR), a Peak Burst Size (PBS), an Excess Burst Size (EBS), and a second time averaging interval (T2).

6.0 Assumptions

Deployment of the AR PDB requires an assumption that the network is well-provisioned enough so that the likelyhood of green packets being dropped in case of congestion is very low. This draft does not dictate a particular method to achieve the above objective. Various traffic engineering methods may be used. As an example, the network operator monitors the level of green packets in the selected AF class on all links and takes appropriate action to limit the green packet loss.

The PDB also assumes that there is relatively stable routing within the domain.

7.0 Security

There are no specific security exposures for this PDB. See the general security considerations in the Diffserv Header RFC [RFC2474] and the Diffserv Architecture RFC [RFC2475].

All the security concerns expressed in [RFC 2597] apply for the AR PDB.

8.0 Example Uses

In this section, we provide only a few example services that are possible with this PDB - the list is not exhaustive. Example services that can be created out of this PDB include: (i) one-to-one or one-to-few VPN-like services and (ii) one-to-any general service.

In the case of VPN-like services, the PDB can be utilized to assure a rate for a traffic aggregate between an ingress and an egress within a domain or from one ingress to few different egress points in the domain.

In the case of one-to-any services, the PDB can be utilized to assure a rate for a traffic aggregate that originates from one ingress node but whose individual five-tuple flows may exit the domain at any of the egress nodes.

It is easier for a provider to demonstrate conformance with the SLS in the one-to-one service since all measurements can be performed at a single egress point. In the case of a one-to-any service, measurements need to be performed at all egress nodes where packets are sent from an ingress node during the measurement interval. These measurements then have to be correlated to determine the cumulative bandwidth of the aggregate as it exits the domain.

Most of the previous portion of this document discussed deployment of the AR PDB in the context of an IP routed network. It is also possible to deploy the AR PDB in other networks - for example, in an MPLS network which supports the Diffserv architecture and mechanisms [DIFFMPLS]. In a Diffserv-enabled MPLS network with dedicated, constraint based LSPs, the AR PDB could be deployed using either E-LSPs or L-LSPs. In such a network, traffic engineering can be used in conjunction with per-LSP admission control mechanisms to provide AR PDB-based services that are quantitative in nature.

In such a network, the AR PDB could be used as the basis for an SLS that includes additional parameters such as statistical assurance probabilities or packet drop assurance. As an example of the former, the SLS MAY specify a Y% assurance for the CIR (with CBS/T1) when sampled every X minutes. As an example of the latter, the SLS MAY specify a CIR (with CBS/T1) and a Y% drop ratio for AR PDB traffic for a particular customer.

9.0 Simulation Summary

There have been a number of simulation and emulation studies of involving PDBs that essentially provide similar behaviour to the AR PDB. In this section, we briefly summarize some of those studies. The primary aim is to show that with the deployment of the required mechanisms, in general, it is possible to provide rate assurance over a time period. Some of the studies also point to suitable provisioning or subscription levels for network links in order to support such a PDB.

In [TON98], the authors perform simulations that show how AR PDB-like

behaviour can be realized in a simple topology utilizing RIO in conjunction with a sliding window policer. This set of simulations showed that the rate for individual TCP or UDP flows can be assured.

In [YEOM], the authors perform simulation tests involving aggregates

of TCP flows with different RTTs. The topology in this case was a dumbell topology. The results indicated that for the utilized topology, the aggregate rate could be assured with a link subscription level of as high as 80%.

In [REZENDE], the authors perform simulations with more complicated topologies involving flow merge and split points, congestion in both directions of a link. In this set of experiments, individual flows had their assured rates. The simulations showed that for link provisioning levels of 40%, the network could assure the rates for the individual flows.

In [HPN2000], the authors perform simulations involving aggregates of TCP and UDP flows. It is observed that if UDP and TCP in-profile packets are marked green, the aggregates achieve their CIR. In another simulation, different flow aggregates with different CIRs, share the same bottleneck link with total allocated bandwidth constituting 40% and 80% of link capacity. It is observed that for the 40% case, the CIR for the different aggregates are always achieved. However, in the 80% case, the CIR for a few of the aggregates are not achieved.

10.0 Acknowledgements

The authors would like to thank the following individuals for their helpful comments and suggestions: Marcus Brunner, Brian Carpenter, Shahram Davari and Alper Demir.

10.0 References

[TON98] D.D. Clark, W. Fang, "Explicit Allocation of Best Effort Packet Delivery Service", IEEE/ACM Transactions on Networking, August 1998, Vol 6. No. 4, pp. 362-373.

[RFC2474] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", Internet RFC 2474, December 1998.

[RFC2475] D. Black, S. Blake, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services", Internet RFC 2475, December 1998.

[AFPHB] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999

[YEOM] Yeom I and Reddy N, "Impact of Marking Strategy for Aggregated Flows in a Differentiated Services Network", Proceedings of

IWQOS Workshop, June 1999

[SRTCM] J. Heinanen and R. Guerin, "A Single Rate Three Colour Marker", Internet RFC 2697, September 1999

[TRTCM] J. Heinanen and R. Guerin, "A Two Rate Three Colour Marker", Internet RFC 2698, September 1999

[REZENDE] J Rezende, "Assured Service Evaluation", Proceedings of GLOBECOM 99, Rio De Janiero, December 1999

[TSWTCM] W. Fang, N. Seddigh and B. Nandy, "A Time Sliding Window Three Colour Marker", Internet RFC 2859, June 2000

[HPN2000] Nandy B, Seddigh N, Pieda P and Ethridge J, "Intelligent Traffic Conditioners for Assured Forwarding Based Differentiated Services Networks", Proceedings of HPN 2000, Paris, June 2000.

[PDBDEF] Nichols K and Carpenter B, "Definition of Differentiated Services Per Domain Behaviours and Rules for their Specification", Internet Draft, October 2000

[DIFFMPLS] Le Faucher et al, "MPLS Support of Differentiated Services", Internet Draft, <draft-ietf-mpls-diff-ext-09.txt>, April 2001

9.0 Author Addresses

Nabil Seddigh
Tropic Networks,
135 Michael Cowpland Drive
Kanata, ON, K2M 2E9
Canada
Email: nseddigh@tropicnetworks.com

Biswajit Nandy
Tropic Networks,
135 Michael Cowpland Drive
Kanata, ON, K2M 2E9
Canada
Email: bnandy@tropicnetworks.com

Juha Heinanen
Song Networks, Inc.
Hallituskatu 16
33200 Tampere
Finland
Email: juha.heinanen@mail.com