Network Working Group
Request for Comments: 3662
Category: Informational

R. Bless
Univ. of Karlsruhe
K. Nichols
Consultant
K. Wehrle
Univ. of Tuebingen/ICSI

December 2003

A Lower Effort Per-Domain Behavior (PDB) for Differentiated Services

Status of this Memo

Copyright Notice

Abstract

This document proposes a differentiated services per-domain behavior (PDB) whose traffic may be "starved" (although starvation is not strictly required) in a properly functioning network. This is in contrast to the Internet's "best-effort" or "normal Internet traffic" model, where prolonged starvation indicates network problems. In this sense, the proposed PDB's traffic is forwarded with a "lower" priority than the normal "best-effort" Internet traffic, thus the PDB is called "Lower Effort" (LE). Use of this PDB permits a network operator to strictly limit the effect of its traffic on "best-effort"/"normal" or all other Internet traffic. This document gives some example uses, but does not propose constraining the PDB's use to any particular type of traffic.

1. Description of the Lower Effort PDB

This document proposes a differentiated services per-domain behavior [RFC3086] called "Lower Effort" (LE) which is intended for traffic of sufficiently low value (where "value" may be interpreted in any useful way by the network operator), in which all other traffic takes precedence over LE traffic in consumption of network link bandwidth. One possible interpretation of "low value" traffic is its low priority in time, which does not necessarily imply that it is generally of minor importance. From this viewpoint, it can be

considered as a network equivalent to a background priority for processes in an operating system. There may or may not be memory (buffer) resources allocated for this type of traffic.

Some networks carry traffic for which delivery is considered optional; that is, packets of this type of traffic ought to consume network resources only when no other traffic is present. Alternatively, the effect of this type of traffic on all other network traffic is strictly limited. This is distinct from "best-effort" (BE) traffic since the network makes no commitment to deliver LE packets. In contrast, BE traffic receives an implied "good faith" commitment of at least some available network resources. This document proposes a Lower Effort Differentiated Services per-domain behavior (LE PDB) [RFC3086] for handling this "optional" traffic in a differentiated services domain.

There is no intrinsic reason to limit the applicability of the LE PDB to any particular application or type of traffic. It is intended as an additional tool for administrators in engineering networks.

Note: where not otherwise defined, terminology used in this document is defined as in [RFC2474].

2. Applicability

A Lower Effort (LE) PDB is for sending extremely non-critical traffic across a DS domain or DS region. There should be an expectation that packets of the LE PDB may be delayed or dropped when other traffic is present. Use of the LE PDB might assist a network operator in moving certain kinds of traffic or users to off-peak times. Alternatively, or in addition, packets can be designated for the LE PDB when the goal is to protect all other packet traffic from competition with the LE aggregate, while not completely banning LE traffic from the network. An LE PDB should not be used for a customer’s "normal internet" traffic, nor should packets be "downgraded" to the LE PDB for use as a substitute for dropping packets that ought to simply be dropped as unauthorized. The LE PDB is expected to be applicable to networks that have some unused capacity at some times of day.

This is a PDB that allows networks to protect themselves from selected types of traffic rather than giving a selected traffic aggregate preferential treatment. Moreover, it may also exploit all unused resources from other PDBs.

3. Technical Specification

3.1. Classification and Traffic Conditioning

There are no required traffic profiles governing the rate and bursts of packets beyond the limits imposed by the ingress link. It is not necessary to limit the LE aggregate using edge techniques since its PHB is configured such that packets of the aggregate will be dropped in the network if no forwarding resources are available. The differentiated services architecture [RFC2475] allows packets to be marked upstream of the DS domain or at the DS domain's edge. When packets arrive pre-marked with the DSCP used by the LE PDB, it should not be necessary for the DS domain boundary to police that marking; further (MF) classification for such packets would only be required if there was some reason for the packets to be marked with a different DSCP.

If there is not an agreement on a DSCP marking with the upstream domain for a DS domain using the LE PDB, the boundary must include a classifier that selects the appropriate LE target group of packets out of all arriving packets and steers them to a marker that sets the appropriate DSCP. No other traffic conditioning is required.

3.2. PHB configuration

Either a Class Selector (CS) PHB [RFC2474], an Experimental/Local Use (EXP/LU) PHB [RFC2474], or an Assured Forwarding (AF) PHB [RFC2597] may be used as the PHB for the LE traffic aggregate. This document does not specify the exact DSCP to use inside a domain, but instead specifies the necessary properties of the PHB selected by the DSCP. If a CS PHB is used, Class Selector 1 (DSCP=001000) is suggested.

The PHB used by the LE aggregate inside a DS domain should be configured so that its packets are forwarded onto the node output link when the link would otherwise be idle; conceptually, this is the behavior of a weighted round-robin scheduler with a weight of zero.

An operator might choose to configure a very small link share for the LE aggregate and still achieve the desired goals. That is, if the output link scheduler permits, a small fixed rate might be assigned to the PHB, but the behavior beyond that configured rate should be that packets are forwarded only when the link would otherwise be idle. This behavior could be obtained, for example, by using a CBQ [CBQ] scheduler with a small share and with borrowing permitted. A PHB that allows packets of the LE aggregate to send more than the configured rate when packets of other traffic aggregates are waiting for the link is not recommended.

If a CS PHB is used, note that this configuration will violate the "SHOULD" of section 4.2.2.2 of RFC 2474 [RFC2474] since CS1 will have a less timely forwarding than CS0. An operator's goal of providing an LE PDB is sufficient cause for violating the SHOULD. If an AF PHB is used, it must be configured and a DSCP assigned such that it does not violate the "MUST" of paragraph three of section 2 of RFC 2597 [RFC2597] which provides for a "minimum amount of forwarding resources".

4. Attributes

The ingress and egress flow of the LE aggregate can be measured but there are no absolute or statistical attributes that arise from the PDB definition. A particular network operator may configure the DS domain in such a way that a statistical metric can be associated with that DS domain. When the DS domain is known to be heavily congested with traffic of other PDBs, a network operator should expect to see no (or very few) packets of the LE PDB egress from the domain. When there is no other traffic present, the proportion of the LE aggregate that successfully crosses the domain should be limited only by the capacity of the network relative to the ingress LE traffic aggregate.

5. Parameters

None required.

6. Assumptions

A properly functioning network.

7. Example uses

- Multimedia applications [this example edited from Yoram Bernet]:

  Many network managers want to protect their networks from certain applications, in particular, from multimedia applications that typically use such non-adaptive protocols as UDP.

  Most of the focus in quality-of-service is on achieving attributes that are better than Best Effort. These approaches can provide network managers with the ability to control the amount of multimedia traffic that is given this improved performance with excess relegated to Best Effort. This excess traffic can wreak havoc with network resources even when it is relegated to Best Effort because it is non-adaptive and because it can be significant in volume and duration. These characteristics permit it to seize network resources, thereby compromising the performance of other, more important applications that are

included in the Best Effort traffic aggregate but that use adaptive protocols (e.g., TCP). As a result, network managers often simply refuse to allow multimedia applications to be deployed in resource constrained parts of their network.

The LE PDB enables a network manager to allow the deployment of multimedia applications without losing control of network resources. A limited amount of multimedia traffic may (or may not) be assigned to PDBs with attributes that are better than Best Effort. Excess multimedia traffic can be prevented from wreaking havoc with network resources by forcing it to the LE PDB.

- For Netnews and other "bulk mail" of the Internet.
- For "downgraded" traffic from some other PDB when this does not violate the operational objectives of the other PDB or the overall network. As noted in section 2, LE should not be used for the general case of downgraded traffic, but may be used by design, e.g., when multicast is used with a value-added DS-service and consequently the Neglected Reservation Subtree problem [NRS] arises.
- For content distribution, peer-to-peer file sharing traffic, and the like.
- For traffic caused by world-wide web search engines while they gather information from web servers.

8. Experiences

The authors solicit further experiences for this section. Results from simulations are presented and discussed in Appendix A.

9. Security Considerations for LE PDB

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

10. History of the LE PDB

The previous name of this PDB, "bulk handling", was loosely based on the United States' Postal Service term for very low priority mail, sent at a reduced rate: it denotes a lower-cost delivery where the items are not handled with the same care or delivered with the same timeliness as items with first-class postage. Finally, the name was changed to "lower effort", because the authors and other DiffServ Working Group members believe that the name should be more generic in order to not imply constraints on the PDB's use to a particular type

of traffic (namely that of bulk data).

The notion of having something "lower than Best Effort" was raised in the Diffserv Working Group, most notably by Roland Bless and Klaus Wehrle in their Internet Drafts [LBE] and [LE] and by Yoram Bernet for enterprise multimedia applications. One of its first applications was to re-mark packets within multicast groups [NRS]. Therefore, previous discussions centered on the creation of a new PHB. However, the original authors (Brian Carpenter and Kathleen Nichols) believe this is not required and this document was written to specifically explain how to get less than Best Effort without a new PHB.

11. Acknowledgments

Yoram Bernet contributed significant amounts of text for the "Examples" section of this document and provided other useful comments that helped in editing. Other Diffserv WG members suggested that the LE PDB is needed for Napster traffic, particularly at universities. Special thanks go to Milena Neumann for her extensive efforts in performing the simulations that are described in Appendix A.

Appendix A. Experiences from a Simulation Model

The intention of this appendix is to show that a Lower Effort PDB with a behavior as described in this document can be realized with different implementations and PHBs respectively. Overall, each of these variants show the desired behavior but also show minor differences in certain traffic load situations. This comparison could make the choice of a realization variant interesting for a network operator.

A.1. Simulation Environment

The small DiffServ domain shown in Figure 1 was used to simulate the LE PDB. There are three main sources of traffic (S1-S3) depicted on the left side of the figure. Source S1 sends five aggregated TCP flows (A1-A5) to the receivers R1-R5 respectively. Each aggregated flow Ax consists of 20 TCP connections, where each aggregate experiences a different round trip time between 10ms and 250ms. There are two sources of bulk traffic. B1 consists of 100 TCP connections sending as much data as possible to R6 and B2 is a single UDP flow also sending as much as possible to R7.

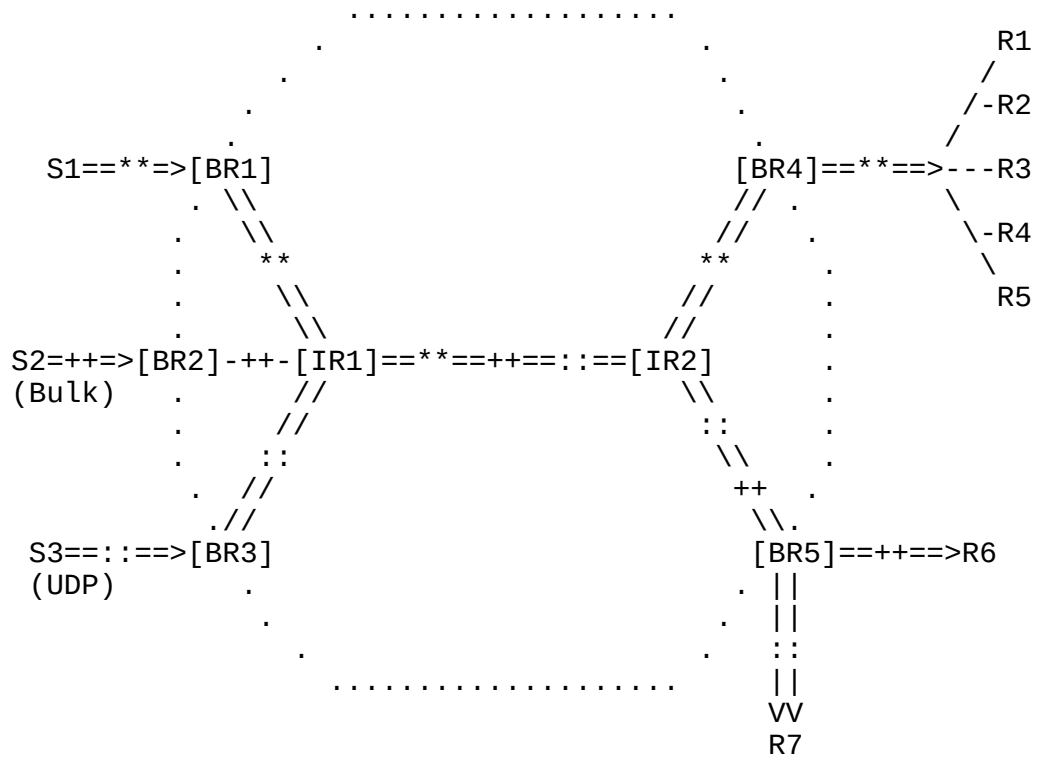

Figure 1: A DiffServ domain with different flows

In order to show the benefit of using the LE PDB instead of the normal Best Effort (BE) PDB [RFC3086], different scenarios are used:

A) B1 and B2 are not present, i.e., the "normal" situation without bulk data present. A1-A5 use the BE PDB.

B) B1 and B2 use the BE PDB for their traffic, too.

C) B1 and B2 use LE PDB for their traffic with different PHB implementations:

   1) PHB with Priority Queueing (PQ)
   2) PHB with Weighted Fair Queueing (WFQ)
   3) PHB with Weighted RED (WRED)
   4) PHB with WFQ and RED

C1) represents the case where there are no allocated resources for the LE PDB, i.e., LE traffic is only forwarded if there are unused resources. In scenarios C2)-C4), a bandwidth share of 10% has been allocated for the LE PDB. RED parameters were set to w_q=0.1 and max_p=0.2. In scenario C2), two tail drop queues were used for BE and LE and WFQ scheduling was set up with a weight of 9:1 for the ratio of BE:LE. In scenario C3), a total queue length of 200000 bytes was used with the following thresholds: min_th_BE=19000, max_th_BE=63333, min_th_LE=2346, max_th=7037. WRED allows to mark packets with BE or LE within the same microflow (e.g., letting applications pre-mark packets according to their importance) without causing a reordering of packets within the microflow. In scenario C4), each queue had a length of 50000 bytes with the same thresholds of min_th=18000 and max_th=48000 bytes. WFQ parameters were the same as in C2).

The link bandwidth between IR1 and IR2 is limited to 1200 kbit/s, thus creating the bottleneck in the network for the following situations. In all situations, the 20 TCP connections within each aggregated flow Ax (flowing from S1 to Rx) used the Best Effort PDB. Sender S2 transmitted bulk flow B1 (consisting of 100 TCP connections to R6) with an aggregated rate of 550 kbit/s, whereas the UDP sender S3 transmitted with a rate of 50 kbit/s.

The following four different situations with varying traffic load for the Ax flows (at application level) were simulated.

Situation	I	II	III	IV
Sender Rate S1 [kbit/s]	1200	1080	1800	800
Sender Rate S2 [kbit/s]	550	550	550	550
Sender Rate S3 [kbit/s]	50	50	50	50
Bandwidth IR1 -> IR2	1200	1200	1200	1200
Best Effort Load (S1)	100%	90%	150%	67%
Total load for link IR1->IR2	150%	140%	200%	117%

In situation I, there are no unused resources left for the B1 and B2 flows. In situation II, there is a residual bandwidth of 10% of the bottleneck link between IR1 and IR2. In situation III, the traffic load of A1-A5 is 50% higher than the bottleneck link capacity. In situation IV, A1-A5 consume only 2/3 of the bottleneck link capacity. B1 and B2 require together 50% of the bottleneck link capacity.

The simulations were performed with the freely available discrete event simulation tool OMNeT++ and a suitable set of QoS mechanisms [SimKIDS]. Results from the different simulation scenarios are discussed in the next section.

A.2. Simulation Results

QoS parameters listed in the following tables are averaged over the first 160s of the transmission. Results of situation I are shown in Figure 2. When the BE PDB is used for transmission of bulk flows B1 and B2 in case B), one can see that flows A1-A5 throttle their sending rate to allow transmission of bulk flows B1 and B2. In case C1), not a single packet is transmitted to the receiver because all packets get dropped within IR1, thereby protecting Ax flows from Bx flows. In case C2), B1 and B2 consume all resources up to the configured limit of 10% of the link bandwidth, but not more. C3) also limits the share of B1 and B2 flows, but not as precisely as with WFQ. C4) shows slightly higher packet losses for Ax flows due to the active queue management.

QoS Parameter	Flows	A) No bulk transfer	Bulk Transfer with PDB: B) Best Effort	C) Lower Effort 1) PQ	2) WFQ	3) WRED	4) RED&WFQ
Throughput [kbit/s]	A1	240	71	240	214	225	219
	A2	240	137	240	216	223	218
	A3	240	209	240	224	220	217
	A4	239	182	239	222	215	215
	A5	238	70	238	202	201	208
	B1	-	491	0	82	85	84
	B2	-	40	0	39	31	38
Total Throughput [kbit/s]	normal	1197	669	1197	1078	1084	1078
	bulk	-	531	0	122	116	122
Paket Loss [%]	A1	0	19.3	0	6.3	5.7	8.6
	A2	0	17.5	0	6.0	5.9	8.9
	A3	0	10.2	0	3.2	6.2	9.1
	A4	0	12.5	0	4.5	6.6	9.3
	A5	0	22.0	0	6.0	5.9	9.0
	B1	-	10.5	100	33.6	38.4	33.0
	B2	-	19.6	100	19.9	37.7	22.2
Total Packet Loss Rate [%]	normal	0	14.9	0	5.2	6.1	9.0
	bulk	0	11.4	100	29.5	38.2	29.7
Transmitted Data [MByte]	normal	21.9	12.6	21.9	19.6	20.3	20.3

Figure 2: Situation I - Best Effort traffic uses 100% of the available bandwidth

Results of situation II are shown in Figure 3. In case C1), LE traffic gets exactly the 10% residual bandwidth that is not used by the Ax flows. Cases C2) and C4) show similar results compared to C1), whereas case C3) also drops packets from flows A1-A5 due to active queue management.

QoS Parameter	Flows	A) No bulk transfer	Bulk Transfer with PDB: B) Best Effort	C) Lower Effort 1) PQ	2) WFQ	3) WRED	4) RED&WFQ
Throughput [kbit/s]	A1	216	193	216	216	211	216
	A2	216	171	216	216	211	216
	A3	216	86	216	216	210	216
	A4	215	121	215	215	211	215
	A5	215	101	215	215	210	215
	B1	-	488	83	83	114	84
	B2	-	39	39	39	33	38
Total Throughput [kbit/s]	normal	1078	672	1077	1077	1053	1077
	bulk	-	528	122	122	147	122
Paket Loss [%]	A1	0	9.4	0	0	1.8	0
	A2	0	14.6	0	0	2.0	0
	A3	0	22.4	0	0	2.1	0
	A4	0	15.5	0	0	1.8	0
	A5	0	17.4	0	0	1.9	0
	B1	-	11.0	32.4	32.9	35.7	33.1
	B2	-	21.1	20.3	20.7	34.0	22.2
Total Packet Loss Rate [%]	normal	0	14.9	0	0	1.9	0
	bulk	-	12.0	28.7	29.1	35.3	29.8
Transmitted Data [MByte]	normal	19.8	12.8	19.8	19.8	19.5	19.8

Figure 3: Situation II - Best Effort traffic uses 90% of the available bandwidth

Results of simulations for situation III are depicted in Figure 4. Due to overload caused by flows A1-A5, packets get dropped in all cases. Bulk flows B1 and B2 nearly get their maximum throughput in case B). As one would expect, in case C1) all packets from B1 and B2 are dropped, in cases C2) and C4) resource consumption of bulk data is limited to the configured share of 10%. Again the WRED implementation in C3) is not as accurate as the WFQ variants and lets more BE traffic pass through IR1.

QoS Parameter	Flows	A) No bulk transfer	B) Best Effort	Bulk Transfer with PDB: C) Lower Effort 1) PQ	2) WFQ	3) WRED	4) RED&WFQ
Throughput [kbit/s]	A1	303	136	241	298	244	276
	A2	316	234	286	299	240	219
	A3	251	140	287	259	236	225
	A4	168	84	252	123	209	219
	A5	159	82	132	101	166	141
	B1	-	483	0	83	73	83
	B2	-	41	0	38	31	38
Total Throughput [kbit/s]	normal	1199	676	1199	1079	1096	1079
	bulk	-	524	0	121	104	121
Paket Loss [%]	A1	9.6	17.6	12.1	9.3	8.6	12.8
	A2	8.5	13.6	8.4	9.8	8.1	14.5
	A3	8.8	18.7	7.7	11.6	7.8	13.6
	A4	14.9	22.3	11.2	18.9	8.2	12.4
	A5	12.8	19.0	15.6	19.7	8.3	14.3
	B1	-	11.9	100	32.1	39.5	33.0
	B2	-	17.3	100	22.5	37.7	22.8
Total Packet Loss Rate [%]	normal	10.4	17.3	10.3	12.2	8.2	13.4
	bulk	-	12.4	100	29.1	39.0	29.9
Transmitted Data [MByte]	normal	22.0	12.6	22.0	20.2	20.6	20.3

Figure 4: Situation III - Best Effort traffic load is 150%

In situation IV, 33% or 400 kbit/s are not used by Ax flows and the results are listed in Figure 5. In case B) where bulk data flows B1 and B2 use the BE PDB, packets of Ax flows are dropped, whereas in cases C1)-C4) flows Ax are protected from bulk flows B1 and B2. Therefore, by using the LE PDB for Bx flows, the latter get only the residual bandwidth of 400 kbit/s but not more. Packets of Ax flows are not affected by Bx traffic in these cases.

QoS Parameter	Flows	A) No bulk transfer	B) Best Effort	Bulk Transfer with PDB: C) Lower Effort 1) PQ	2) WFQ	3) WRED	4) RED&WFQ
Throughput [kbit/s]	A1	160	140	160	160	160	160
	A2	160	124	160	160	160	160
	A3	160	112	160	160	160	160
	A4	160	137	160	160	159	160
	A5	159	135	159	159	159	159
	B1	-	509	361	362	364	362
	B2	-	43	40	39	38	40
Total Throughput [kbit/s]	normal	798	648	798	798	797	798
	bulk	-	551	401	401	402	401
Paket Loss [%]	A1	0	9.2	0	0	0	0
	A2	0	12.2	0	0	0	0
	A3	0	14.0	0	0	0	0
	A4	0	9.3	0	0	0	0
	A5	0	6.6	0	0	0	0
	B1	-	7.3	21.2	21.8	25.0	21.3
	B2	-	14.3	19.4	20.7	24.5	20.7
Total Packet Loss Rate [%]	normal	0	10.2	0	0	0	0
	bulk	-	8.0	21.0	21.7	25.0	21.2
Transmitted Data [MByte]	normal	14.8	12.1	14.8	14.8	14.7	14.7

Figure 5: Situation IV - Best Effort traffic load is 67%

In summary, all the different scenarios show that the "normal" BE traffic can be protected from traffic in the LE PDB effectively. Either no packets get through if no residual bandwidth is left (LE traffic is starved), or traffic of the LE PDB can only consume resources up to a configurable limit.

Furthermore, the results substantiate that mass data transfer can adversely affect "normal" BE traffic (e.g., 14.9% packet loss in situations I and II, even 10.2% in situation IV) in situations without using the LE PDB.

Thus, while all presented variants of realizing the LE PDB meet the desired behavior of protecting BE traffic, they also show small differences in detail. A network operator has the opportunity to choose a realization method to fit the desired behavior (showing this is - after the proof of LE's efficacy - the second designation of this appendix). For instance, if operators want to starve LE traffic completely in times of congestion, they could choose PQ. This causes LE traffic to be completely starved and not a single packet would get through in case of full load or overload.

On the other hand, for network operators who want to permit some small amount of throughput in the LE PDB, one of the other variants would be a better choice.

Referring to this, the WFQ implementation showed a slightly more robust behavior with PQ, but had problems with synchronized TCP flows. WRED behavior is highly dependent on the actual traffic characteristics and packet loss rates are often higher compared to other implementations, while the fairness between TCP connections is better. The combined solution of WFQ with RED showed the overall best behavior, when an operator's intent is to keep a small but noticeable throughput in the LE PDB.

Normative References

[RFC3086] Nichols, K. and B. Carpenter, "Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification", RFC 3086, April 2001.

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

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

Informative References

[RFC2597] Heinanen, J., Baker, F., Weiss, W. and J. Wroclawski, "Assured Forwarding PHB Group", RFC 2597, June 1999.

[CBQ] Floyd, S. and V. Jacobson, "Link-sharing and Resource Management Models for Packet Networks", IEEE/ACM Transactions on Networking, Vol. 3, No. 4, pp. 365-386, August 1995.

[LBE] Bless, R. and K. Wehrle, "A Lower Than Best-Effort Per-Hop Behavior", Work in Progress, September 1999.

[LE] Bless, R. and K. Wehrle, "A Limited Effort Per-Hop Behavior", Work in Progress, February 2001.

[SimKIDS] Wehrle, K., Reber, J. and V. Kahmann, "A simulation suite for Internet nodes with the ability to integrate arbitrary Quality of Service behavior", in Proceedings of Communication Networks And Distributed Systems Modeling And Simulation Conference (CNDS 2001), Phoenix (AZ), USA, pp. 115-122, January 2001.

[NRS] Bless, R. and K. Wehrle, "Group Communication in Differentiated Services Networks", in Proceedings of IEEE International Workshop on "Internet QoS", Brisbane, Australia, IEEE Press, pp. 618-625, May 2001.

Authors' Addresses

Roland Bless
Institute of Telematics, Universitaet Karlsruhe (TH)
Zirkel 2
76128 Karlsruhe
Germany

EMail: bless@tm.uka.de
URI: http://www.tm.uka.de/~bless/

Kathleen Nichols
325M Sharon Park Drive #214
Menlo Park, CA 94025

EMail: knichols@ieee.org

Klaus Wehrle
University of Tuebingen, Computer Networks and Internet
Morgenstelle 10c, 72076 Tuebingen, Germany &
International Computer Science Institute (ICSI)
1947 Center Street, Berkeley, CA, 94704, USA

EMail: Klaus.Wehrle@uni-tuebingen.de
URI: http://net.informatik.uni-tuebingen.de/~wehrle/

Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.