Enabling IP-Multicast in Differentiated Services Networks

Roland Bless and Klaus Wehrle

Institute of Telematics, University of Karlsruhe
Zirkel 2, D-76128 Karlsruhe, Germany
{bless, wehrle}@telematik.informatik.uni-karlsruhe.de

The Differentiated Services approach is currently intensively discussed in the Internet community. It will also bring benefits for multicast applications that need quality of service support. For instance, a highly reliable multicast can be provided based on Expedited Forwarding. However, current efforts mainly concentrate on unicast services, whereas multicast services have not been addressed in a very detailed manner yet.

This article illustrates some of the problems which will arise when IP Multicast is used in Differentiated Services networks without taking special provisions into account for supplying it. Those problems mainly lead to situations in which other service users are affected adversely. In order to retain the benefits of the DiffServ approach, a quite simple and scalable solution for those problems is required, not resulting in additional complexity or costs in a DiffServ domain. The proposed solution in this paper requires only an additional entry for the DiffServ Codepoint in the multicast routing table and some support by management mechanisms. The discussion of the related problems and presentation of the solution is illustrated and affirmed by some measurements with IP Multicast groups using Expedited Forwarding.

1 Introduction

Services in the Internet offering a better quality than the current best-effort service are increasingly required. Many advanced applications need certain assurances from the network layer, e.g., a maximum delay, a minimum packet loss rate or guaranteed transmission rate. The currently used IP mechanisms are not able to offer such guarantees, especially, if group communication is additionally required.

The IETF attempted to meet these trends in defining the Integrated Services (IntServ) architecture, which provided quality based services in the Internet. After some years it was recognized that the IntServ Architecture has some scalability problems. So a new IETF working group was founded with the goal to specify a highly scalable architecture for sup-

porting differentiated services in the Internet. In the Differentiated Services Architecture [1, 3] scalability is achieved by avoiding complexity and maintaining per-flow state information in core routers and pushing unavoidable complexity to the network edges. Therefore, individual flows belonging to the same service are aggregated, thereby eliminating the need for complex classification or managing state information per flow in interior routers.

On the other hand, the reduced complexity in routers makes it more complex to use such better services together with IP Multicast. Problems that emerge from this fact are described in section 3. However, it is important to integrate IP Multicast functionality right from the beginning into the architecture, and, to provide simple solutions for those problems not defeating the so far gained advantages.

2 Differentiated Services

In the Differentiated Services Architecture different services can be constructed from different *per-hop forwarding behaviors (PHB)* which packets may experience. A packet is usually classified and marked to receive a particular forwarding behavior in the first DiffServ-capable node along its path. Classification at this early stage can be based on multiple fields of a packet (Multi-Field Classifier), which identify an instance of an application-to-application "microflow" by comprising source address, source port, destination addresses, destination port and protocol identification number. Such a classification rule for packets is part of a *Traffic Conditioning Agreement (TCA)* which also comprises a corresponding traffic profile, i.e., a description of a traffic stream's properties. Thus, packets are marked according to their corresponding traffic profile that is selected by the classifier. A TCA is usually derived from the service contract between a customer and a service provider which specifies the forwarding service desired by this customer. This contract is often also denoted as *Service Level Agreement (SLA)*.

The forwarding behavior that a packet experiences is identified by a so-called *codepoint* in the IP packet header. Each codepoint is a specific value in a part of the *Differentiated*

Services Field (DS-field), that replaces the common *Type of Service* (ToS) field in IPv4 packets and the class field in IPv6 packets [1].

Packets on the same link in a particular direction carrying the same codepoint are denoted as *Behavior Aggregate (BA)*. After an initial setting of the codepoint, subsequent nodes on the path only operate on those aggregates, that means they do not distinguish between different flows within an aggregate. Consequently, this leads to a higher scalability by avoiding complexity and per-flow state information expecially in "core" routers that are located in the inner network. Therefore, interior routers only have to classify packets according to their specific codepoint and treat them with the corresponding forwarding mechanism.

A Differentiated Services Domain consists of DiffServ-capable nodes. The DiffServ Architecture distinguishes two types of nodes [3]: *boundary nodes* interconnecting a DS domain with other domains, and, *interior nodes* being all other nodes within the same DS domain and having interconnections to other interior nodes and boundary nodes of this DS domain (cf. fig. 1). DS domains which are directly connected to each other define a *DS region*. In a DS region typically two types of DS domains can be distinguished: a *stub DS domain* and a *transit DS domain*. Stub DS domains constitute the first or last access point for end-systems to the DS region. Thus, stub DS domains contain first-hop and last-hop routers to connect sender and receiver respectively. In contrast, transit domains only interconnect different DS domains to each other. Some stub DS domains will serve also concurrently as transit DS domains in some cases, whereas pure transit DS domains have no direct links to any end-systems. As described above, the first boundary node on a packet's path plays an important role, because it decides about the initial setting of the DS codepoint, and, it is usually the only node that distinguishes between microflows by using an MF classifier. Subsequent nodes cannot act on specific microflows within a behavior aggregate. Therefore, one can distinguish two different types of boundary nodes and describe three types of routers:

- A *First-Hop Router* is the first DS-capable router on the path from the sender to the receiver (in the DS Architecture traffic is only considered simplex – duplex flows consist of two separate DS flows). Its task is to classify and mark packets from each (micro-)flow according to a corresponding service profile. Usually, it also carries out complex traffic conditioning functions, such as policing and shaping. Thus, it protects the DS domain ingress from entrance of unallowed or malicious pre-marked traffic.
- *Border Routers* are placed on the network boundary of a DS domain and typically interconnect two Internet Service Providers or two administrative regions. Their task is to police and condition traffic, that is leaving a DS domain or flowing into it, so they are often acting as ingress and egress border router. Classification will be simpler compared to first-hop routers, because traffic has to be differentiated only in dependence on the behavior aggregate and incoming or outgoing link. Therefore, they are also highly scalable, because they do not have to keep and maintain per flow states and reservation information. Instead of this they will use profiles for different BA/link pairs.
- *Interior Routers* are usually the simplest type of DS-capable routers and interconnect interior nodes and boundary nodes in their domain. They classify incoming packets only according to their particular DS codepoint and forward them with the corresponding PHB. Therefore, interior routers do not need to maintain any traffic profiles or per flow state information in order to check for conformance of the incoming traffic. As a consequence, they scale very well even with high amounts of microflows.

It is important to recognize that the first-hop router mainly determines the service that an incoming packet will experience, whereas all subsequent routers cannot change codepoints for a specific microflow, but only for packets of the aggregate of a service. For example, a border router in a transit DS domain has no possibility to filter out incoming packets of a particular flow that causes problems at some egress point. This can be the case when the destination of this flow is altered without reserving the appropriate resources for the new

path in this domain, possibly causing packet dropping – even of other conforming flows – if the resource capacity is exhausted at the egress link. Similarly, the change of a codepoint for packets of a particular flow in the interior network requires per flow classification leading to the same scalability problems which the Integrated Services approach possesses.

2.1 Per-Hop Behaviors

The *Expedited Forwarding PHB (EF)* [6] provides a guaranteed bandwidth, low delay and low loss service that shows the same characteristics as a "virtual leased line". This is achieved by keeping EF queues very small or almost empty, which in turn can only be accomplished by guaranteeing that the maximum arrival rate of an aggregate is less than that aggregate's minimum departure rate. Thus, admission control, policing and shaping is needed with respect to the guaranteed rate as a required configuration parameter. Usually, all packets waiting in the EF queue for transmission are served before all packets waiting in queues of other services.

The *Assured Forwarding* PHB Group (AF) [2] permits a statistical guaranteed rate only. It allows senders to use additional available capacity while providing a minimum base rate. Packets that exceed the negotiated rate are marked for subsequent treatment with higher drop precedence. Consequently, bursts of packets belonging to an AF flow may transit successfully a DS domain if enough capacity is available.

The Expedited Forwarding shows very interesting properties within a multicast context. The very low packet loss characteristic makes it suitable as a basis for a highly (but not absolute) reliable multicast service [8]. Packet loss cannot be fully precluded, because of aggregation effects that may lead to packet loss. Nevertheless, in reality packet losses should occur so infrequently that many applications can tolerate these losses, or if this is not the case, that at least very simple retransmission schemes can be applied.

2.2 Management of Differentiated Services

As mentioned above, at least for Expedited Forwarding admission control and resource reservation is required. Furthermore, installation and updating of traffic profiles in boundary

nodes is necessary. Most network administrators will not accomplish this task manually, even for long term SLAs. Furthermore, offering *services on demand* requires some kind of signaling and automatic admission control procedures. Therefore, the concept of *Bandwidth Brokers* was already suggested by Van Jacobson at a very early stage of DiffServ research [7]. In this concept, the *Bandwidth Broker (BB)* is a dedicated node in each DS domain, keeping track of the amount of available and reserved bandwidth for services, and, processing admission control requests from customers or BBs of adjacent domains. Additionally, it installs or alters traffic profiles in boundary routers.

Protocols for signaling a reservation request to a Differentiated Services Domain are required. For accomplishing end-system signaling to DiffServ domains RSVP may be used with new DS specific reservation objects. RSVP is mainly designed for use in multicast scenarios and is already supported by many operating systems. However, when applying RSVP to a DiffServ network some problems will arise that are described in the next section.

3 Problems of IP Multicast in DS Domains

Although potential problems and the complexity of providing multicast with Differentiated Services are considered in a separate section of [3], both aspects have to be discussed in greater detail. The simplicity of the Differentiated Services Architecture and its router models is necessary to reach high scalability, but it causes also fundamental problems in conjunction with the use of IP Multicast in DS domains. The following subsections describe these problems for which a simple solution is proposed in section 4. This solution is scalable and needs no resource separation by using different codepoint values for unicast and multicast traffic.

Because Differentiated Services are unidirectional by definition, we also consider the point-to-multipoint communication being of unidirectional nature. In traditional IP Multicast any node can send packets spontaneously and asynchronously to a multicast group, respectively to its multicast group address (therefore, IP Multicast offers a multipoint-to-multipoint

service). How to partially retain this feature in a Differentiated Services context is discussed in section 3.4.

For subsequent considerations we assume, unless stated otherwise, at least a unidirectional point-to-multipoint communication scenario in which the sender generates packets which experience a "better" Per-Hop Behavior (PHB, see [1, 3] for its exact meaning) than the traditional default PHB, resulting in a service of better quality compared to the default best-effort service. In order to accomplish this, a traffic profile corresponding to the traffic conditioning specification has to be installed in the sender's first-hop router (the first boundary node of the first DS domain receiving those packets). Furthermore, it must be assured that the corresponding resources are available on the path from the sender to all the receivers, possibly requiring adaptation of traffic profiles at involved domain boundaries. The latter process may be also initiated on demand of a receiver.

3.1 Neglected Reservation Subtree Problem (NRS)

Typically, resources for Differentiated Services must be reserved before actually using them. But in a multicast scenario group membership is often highly dynamic, therefore limiting the use of a sender-initiated resource reservation in advance. Unfortunately, dynamic addition of new members of the multicast group using Differentiated Services can adversely affect existing other traffic, if resources were not explicitly reserved before use.

IP Multicast packet replication usually takes places when the packet is handled by the routing process. Thus, a DiffServ capable node would also copy the content of the DS field [1] into the IP packet header of every replicate. Consequently, replicated packets get exactly the same DS codepoint as the original packet, and, therefore experience the same forwarding treatment as the incoming packets of this multicast group. Normally, the replicating node cannot test whether a corresponding reservation exists for a particular flow of replicated packets on an output link (resp. its corresponding interface), because a flow-specific traffic profile is usually not available in boundary (except in first-hop nodes) and interior nodes.

When a new receiver joins an IP multicast group, the corresponding multicast routing protocol (e.g., DVMRP, PIM-DM or PIM-SM) accomplishes that the multicast tree is expanded by a new subtree which connects the new receiver to the already existing multicast tree. As a result of tree expansion and missing per-flow classification mechanisms (cf. section 2), the new receiver will implicitly use the value-added-service of better quality.

If the additional amount of resources which are consumed by the new part of the multicast tree are not taken into account by the domain management (cf. section 2.2), the currently provided level of quality of service of other receivers (with correct reservations) will be adversely affected or violated. This negative effect on existing traffic contracts by a neglected reservation – in the following designated as *Neglected Reservation Subtree Problem (NRS Problem)* – must be avoided under any circumstances.

One can distinguish two distinct major cases of the NRS problem. In order to compare their different effects a simple example of a share of bandwidth is illustrated in Fig. 1. Three types of services (respectively their corresponding behavior aggregates) share the bandwidth of the considered output link: Expedited Forwarding, Assured Forwarding and the traditional Best-Effort service. In this example we assume, in accordance to the implementation in KIDS [4], that routers perform simple priority queueing, where Expedited Forwarding has the highest and Best-Effort the lowest assigned priority. When Weighted Fair Queueing (WFQ) would be used, the described effects would also occur, only with minor differences.

The Neglected Reservation Subtree problem occurs in two different cases:

- *Case 1:* If the branching point of the new subtree and the previous multicast tree is an (egress) border router, as shown in Fig. 2, the additional multicast flow now increases the amount of used resources for the corresponding aggregate and will be greater than the originally reserved amount. Consequently, the policing component in the egress border router discards packets until the traffic aggregate is conforming to the traffic contract. But during discarding packets, the router can not identify the responsible flow (be-

cause of missing flow classification functionality), and, thus randomly discards packets, whether they belong to a correct reserved flow or not. As a result, there will be no longer any service guarantee for the reserved flows.

Fig. 3 shows the resulting share of bandwidth in cases when (a) Expedited Forwarding and (b) Assured Forwarding are used for the flow causing the NRS problem. Assuming that the additional traffic would use another 30% of link bandwidth, Fig. 3 (a) illustrates that the resulting aggregate of Expedited Forwarding (70% of the outgoing link bandwidth) is throttled down to its originally reserved 40%. In this case, the amount of dropped Expedited Forwarding bandwidth is equal to the amount of excess bandwidth. The marked parts in Fig. 3 indicate that each part of the Expedited Forwarding aggregate is affected by packet losses, too. The other services, e.g., Assured Forwarding or Best-Effort, are not disadvantaged.

Fig. 3 (b) shows the same situation for Assured Forwarding. The only difference is that now Assured Forwarding is affected by discards and the other services will get their guarantees.

In either case, packet losses are restricted to the misbehaving service class by the traffic meter and policing mechanisms in border routers. Moreover, the latter problem (case 1) occurs only in egress border routers, because they are responsible, that not more traffic leaves the Differentiated Services domain, than the following ingress border router will accept. Therefore, those violations of SLAs will be already detected and processed in egress border routers.

– *Case 2:* The Neglected Reservation Subtree problem can also occur, if the branching point between the previous multicast tree and the new subtree is located in an interior router (as shown in Fig. 4). Because the router is not equipped with metering or policing functions it will not recognize any excess amount of traffic and will forward the new multicast flow. If the latter belongs to a higher priority service, such as Expedited

Forwarding, bandwidth of the aggregate is higher than the aggregate's reservation and will steal bandwidth from lower priority services. The additional amount of Expedited Forwarding without a corresponding reservation is forwarded together with the aggregate that has a reservation. This results in no packets losses for Expedited Forwarding as long as the resulting aggregate is not higher than the output link bandwidth. Because of its higher priority, Expedited Forwarding gets as much bandwidth as needed and as is available (strictly speaking, it is implementation dependent whether interior routers have something like a maximum configured service rate).

The result is, that there is no restriction for Expedited Forwarding, but as Fig. 5 (a) shows, other services will be extremely disadvantaged by this use of non-reserved resources. Their bandwidth is stolen by the new additional flow. In this case, the additional 30% Expedited Forwarding traffic preempts resources from the Assured Forwarding traffic, which in turn preempts resources from the best-effort traffic, resulting in 10% packet losses for the Assured Forwarding aggregate and complete loss of best-effort traffic. The example in Fig. 5 (b) shows that this can also happen with lower priority services like Assured Forwarding. When a reservation for service flow with lower priority is neglected other services (with even lower priority) can be reduced in their quality (in this case the best-effort service). As shown in the example, the service's aggregate causing the problem can itself be affected by packet losses (10% of the Assured Forwarding aggregate is discarded).

Besides the described problems of case 2, case 1 will arise in the next border router which performs traffic metering and policing for flows of the service aggregate.

3.2 RSVP causes NRS Problem

Directly applying RSVP to Differentiated Services would also result in an NRS Problem, because a receiver has to join the IP multicast group before sending a resource reservation request (RESV message), in order to receive the sender's PATH messages at first. Thus,

the join for receiving PATH messages will already cause an NRS Problem if this situation is not handled in a special way (like the solution in 4.1).

3.3 ,,Snooping" into a multicast group

In the current architecture of IP multicast, it is possible to join each group and to get their data. When a value added service, for which a receiver has to pay, is used by a sender, each receiver first has to reserve the quality for the value added service. A simple ,,snooping" into the group is no more possible.

The solution of the NRS problem, which is presented in chapter 4.1 allows this ,,snooping" into a group by tagging packets with the best effort codepoint, when they are transmitted on a multicast subtree without reservation.

3.4 Dynamics of Arbitrary Sender Change

Basically, within an IP multicast group *any* participant (actually, it can be any host not even receiving packets of this multicast group) can act as a sender. This is an important feature that should also be available in case a specific service other than best-effort is used within the group. Differentiated Services possess conceptually a simplex character. Therefore, for every multicast tree implied by a sender resources must be reserved separately if *simultaneous sending* should be possible with a better service. This is even true if shared multicast delivery trees are used (e.g., with PIM-SM or Core Based Trees). If not enough resources are reserved for a service within a multicast tree not enough for simultaneous sending of more than one participant, the NRS problem will occur again.

4 Solutions for Enabling Multicast in Differentiated Services Networks

The problems described in the previous section are mainly caused by the simplicity of the Differentiated Services Architecture. Solutions have to be developed that do not introduce an additional amount of complexity that diminishes the scalability of this approach.

In this paper, a simple solution is suggested for most of the problems. In order to keep things simple, we restrict this first solution for supporting heterogeneous groups to the case in which only two different services within a multicast group can be used simultaneously.

4.1 Solution for the NRS Problem

A usage of resources which where not reserved before must be precluded. In our example, we want to consider the case when the join of a new receiver to a DS multicast group requires grafting of a whole new subtree to an already existing multicast delivering tree. The connecting node which joins both trees converts the codepoint (and therefore the Per-Hop Behavior) to a codepoint of a PHB which is similar to the default PHB (see ③ Fig. 6) in order to provide a best-effort-like service for the new subtree. More specifically, this particular PHB has to be different from the default PHB and should provide a service which is even worse than the best-effort service of the default PHB.

The conversion to this specific PHB is necessary in order to avoid unfairness being introduced otherwise within the best-effort service aggregate, and, which results from the higher amount of resource usage of the incoming traffic belonging to the multicast group. If the rate at which remarked packets are injected into the outgoing aggregate is not reduced, those remarked packets will probably cause discarding of other flow's packets in the outgoing aggregate if resources are scarce. Therefore, the remarked packets from this multicast group should be discarded more aggressively than other packets in this outgoing aggregate. This could be accomplished by using a new *Lower Than Best Effort* (LBE) PHB (and a related DSCP) for those packets [5]. Merely dropping packets more aggressively at the remarking node is not sufficient, because there may be enough resources in the outgoing BA to transmit every remarked packet and not requiring discarding any other packets within the same BA. However, resources in the next node may be short for this particular BA. Therefore, those "excess" packets must be identifiable at this node. Mechanisms to provide

a “fair” share within the LBE aggregate or between an LBE aggregate and a BE aggregate are out of scope of this document and are discussed in [5].

The better service will be only provided if a reservation request was processed by the management (e.g., Bandwidth Brokers). In case the admission test is successful, the remarking node will be instructed by the management entity to stop remarking and to use the original codepoint again ④. Because reservation requests can also be initiated by the sender, an incoming JOIN-Request of a new receiver subtree should be also forwarded by a boundary node to the management node (indicated by the JOIN_INDICATION message in step ② in Fig. 6), so that the remarking node can be instructed (via the SET_CODEPOINT message in step ④)to immediately use the same codepoint value for replicated packets belonging to this group as for incoming packets (Fig. 6).

The proposed solution does not require any additional classification of multicast groups within an aggregate. Because every multicast packet has to be handled by the multicast routing process (in this context, this notion signifies the multicast forwarding part and not the multicast route calculation and maintenance part), addition of an extra byte in each multicast routing table for containing the DS field, and, thus its DS codepoint value, per output link (resp. child virtual interface) results in nearly no additional cost. Packets will be replicated according to the multicast routing process, so this is also the right place for setting the correct DSCP values of the replicated packets. Their DSCP values are not copied from the incoming original packet, but from the additional DS field in the multicasting routing table entry for the corresponding output link (only the DSCP value must be copied, while the two remaining bits are ignored and are present for simplification reasons only). This field contains initially the codepoint of the LBE PHB if incoming packets for this specific group do not carry the codepoint of the default PHB. When a packet arrives with the default PHB, the outgoing replicates should also get the same codepoint in order to retain the behavior of todays common multicast groups using the default PHB. The SET_CODEPOINT message

changes the DSCP values in the multicast routing table and may also carry the new DSCP value which should be set in the replicated packets.

Furthermore, there must be a mechanism for boundary nodes to inform a management entity about the join request of a new subtree (something like the JOIN_INDICATION message). In order to keep the complexity of interior nodes low, this task should be preferably handled by boundary routers. Additionally, a mechanism must be supplied for instructing a router to change the DSCP value in the related multicast routing table entry (something like the SET_CODEPOINT message). This mechanism may be also incorporated into an existing multicast routing protocol as an extension.

In summary, only those receivers will obtain a better service within a DiffServ multicast group, which previously reserved the according resources in the new subtree with assistance of the management. Otherwise they get a quality that is even lower than best-effort.

4.2 Solution for „Snooping" and the RSVP problem

With the presented solution, receivers are allowed to obtain a best effort or Lower Than Best-Effort service without a reservation, so that at least two different service classes within a multicast group are possible ((L)BE and the value added service class). Therefore, it is possible that any receiver may participate in the multicast session without getting any quality of service. This is useful if a receiver just wants to see whether the content of the multicast group is of interest for him, before requesting a better service that must be paid for.

Additionally, applying the RSVP concept of listening for PATH messages before sending any RESV message is now possible again. Without using our proposed solution this would have caused an NRS problem.

4.3 Solution for Arbitrary Sender Change

Every participant would have to initiate an explicit reservation if a receiver wants to make sure that it is possible to send with a value-added-service to the group, regardless whether

other senders already use the same service class simultaneously. This would require a separate reservation for each sender rooted multicast tree.

However, in the specific case of best-effort service (the default PHB), it is nevertheless possible for participants to send packets anytime to the group without requiring any additional mechanisms. The cause for this is that the first-hop router will mark those packets with the DSCP for the default PHB because of a missing traffic profile for this sender. First hop routers should therefore always classify multicast packets in dependence of the sender's address and multicast group address.

5 Proof of the Neglected Reservation Subtree Problem

In the following sections, it is showed that the NRS problem actually exists and occurs in reality. Hence, we investigated the problem and its solution using our own Linux-based implementation KIDS, which is described detailed in [4].

5.1 Test Environment and Execution

In order to proof NRS problem case 1, as described in the above section, a testbed shown in Fig. 7 was built. It is a reduced version of the network shown in Fig. 4 and consists of two DS-capable routers, a first-hop router and an egress border router. The absence of interior routers does not have any effects to proof the described problem.

The testbed comprises two Personal Computers used as Differentiated Services routers (Pentium III at 450 Mhz, 128 MB Ram, 3 network cards Intel eepro100), as well as one sender and three receiver systems (also PCs). On the routers KIDS has been installed and an *mrouted* was used to perform multicast routing. The network was completely build of separate 10BaseT Ethernet segments in full-duplex mode. In [4] we evaluated the performance of the software routers and found out that even a PC at 200 Mhz had no problem to handle up to 10 Mbps DS traffic on each link. Therefore, in the presented measurements are no performance bottlenecks on part of the software router.

The sender generates two shaped UDP traffic flows with an amount of 500 kbps (packets of 1.000 byte constant size) and sends them to the multicast group 1 (233.1.1.1) and 2 (233.2.2.2). In both measurements receiver A has a reservation on the path to the sender for each flow, receiver B has reserved for flow 1 and C for flow 2. Therefore, two static profiles are installed in the first-hop router with 500 kbps Expedited Forwarding and a token bucket size of 10.000 byte for each flow. In the egress border router one profile has been installed for the output link to host B and one related for the output link to host C. Each of them permits 500 kbps Expedited Forwarding, but only the aggregate of Expedited Forwarding traffic carried on the outgoing link is considered.

In measurement 1 hosts join to the groups as shown in Fig. 2. Those joins are using a reservation for the group towards the sender. Only the join of host B to group 2 has no admitted reservation. As described in section 3.1 this will cause the NRS problem (case 1). Metering and policing mechanisms in the egress border router throttle down the Expedited Forwarding aggregate to the reserved 500 kbps, whether individual flows have reserved or not.

Fig. 8 shows the obtained results. Host A and C received their flows without any interference. But host B received data from group 1 only with half of the reserved bandwidth, so one half of the packets has been discarded. Fig. 8 also shows that receiver B got the total amount of bandwidth for group 1 and 2, that is exactly the reserved 500 kbps. Flow 2 got a Expedited Forwarding without actually having reserved any bandwidth and additionally violated the guarantee of group 1 on that link.

For measurement 2 the previously presented solution (cf. section 4) has been installed in the border router. Now it checks during duplicating the packets, whether the codepoint has to be changed to (Lower Than) Best-Effort or whether it can be just copied. In this measurement it changed the codepoint for group 2 on the link to Host B to best-effort.

Results of this measurement are presented in Fig. 9. Each host gets its flows with the reserved bandwidth and without any packet loss. Packets from group 2 are remarked in the border router so that they have been treated as best-effort traffic. In this case, they got the same bandwidth as the Expedited Forwarding flow, because there was not enough other traffic on the link present, so that there was no need to discard packets.

The above measurements confirm that the Neglected Reservation Subtree problem is to be taken serious and that the presented solution will solve it.

6 Summary and Future Work

Aspects of providing Differentiated Services within multicast groups were not addressed in very much detail so far. In this paper, three problems were identified: resource conflicts due to neglected reservations, problems concerning support for heterogenous groups, and, finally supporting different senders within a group. The proposed solution uses an additional entry in the multicast routing table within DS nodes that holds a DS field (merely one byte) for setting the DSCP value in replicated packets. In order to accomplish this, additional support from a domain management has to be provided.

There are still some open problems according to support Differentiated Services within multicast groups, especially when providing more than two different services within a group. Furthermore, the effects (fairness/unfairness) caused by packets that are remarked to the proposed Lower Than Best-Effort PHB have to be investigated in detail. Last but not least, an open solution for implementing the additional management functionality (e.g., a means for setting the DSCP value in the multicast routing table) has to be developed.

References

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

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

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

[4] R. Bless and K. Wehrle. Evaluation of Differentiated Services using an Implementation under Linux. In *Proceedings of the 7th IFIP Workshop on Quality of Service, London, June 1999*. IEEE, 1999.

[5] R. Bless and K. Wehrle. A Lower Than Best-Effort Per-Hop Behavior. Internet-Draft – draft-bless-diffserv-lbe-phb-00.txt, Sept. 1999.

[6] V. Jacobson, K. Nichols, and K. Poduri. An Expedited Forwarding PHB. RFC 2598, June 1999.

[7] V. Jacobson, K. Nichols, and L. Zhang. A Two-bit Differentiated Services Architecture for the Internet. RFC 2638, July 1999.

[8] C. Metz. RELIABLE MULTICAST: When Many Must Absolutely Positively Receive It. *IEEE Internet Computing*, pages 9–13, July/August 1999.

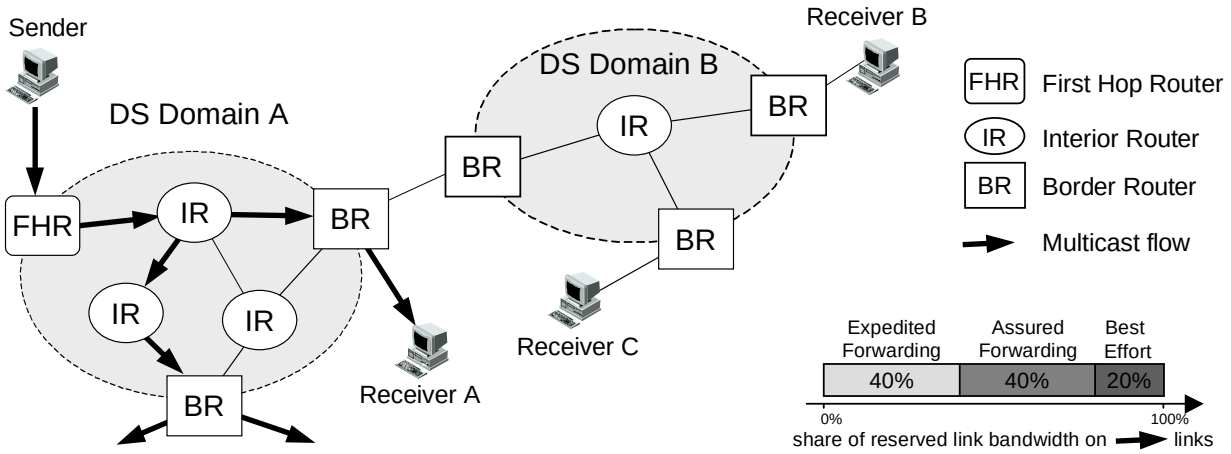


Fig. 1. Example of two Differentiated Services Domains using IP multicast with reserved bandwidth

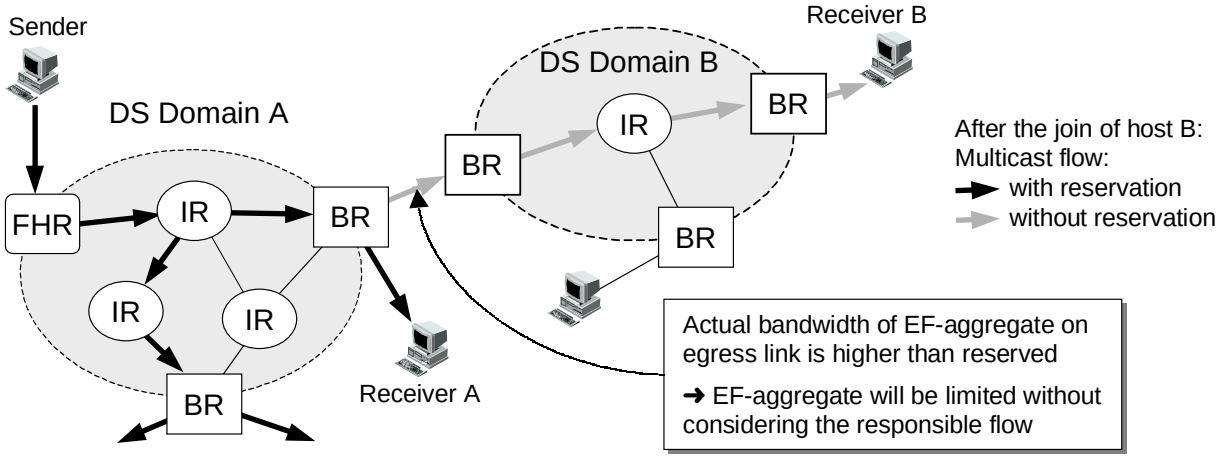


Fig. 2. Case 1 of NRS problem: After the join of Host B, the EF aggregate between the Border Routers is higher than reserved.

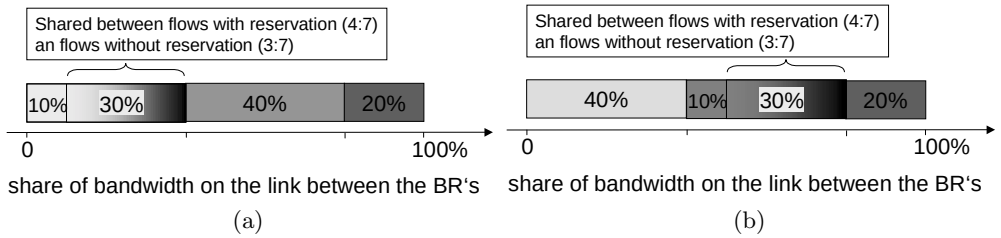


Fig. 3. Resulting share of bandwidth between the Border Routers with a neglected reservation of a (a) EF flow or (b) an AF flow. In the marked areas the traffic conditioning component could not distinguish between flows with and without reservation.

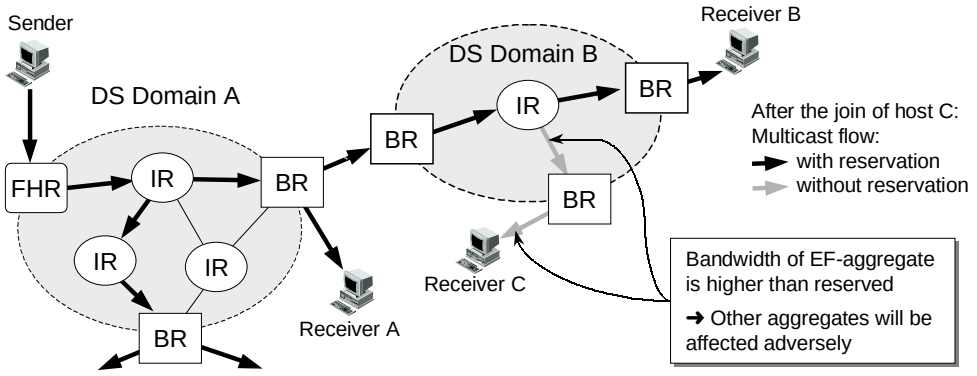
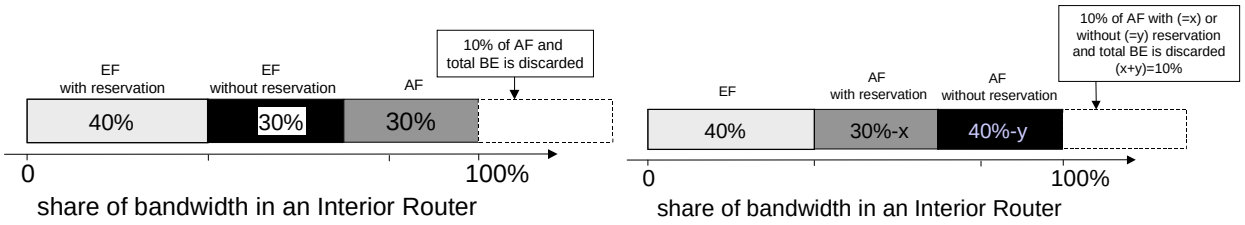


Fig. 4. Neglected Reservation Subtree problem case 2 after join of host C



(a) (b)

Fig. 5. Resulting share of bandwidth in an interior router with a neglected reservation of (a) a Expedited Forwarding flow or (b) an Assured Forwarding flow. Because of the missing traffic meter and reservation information lower or equal priority flows will be disadvantaged.

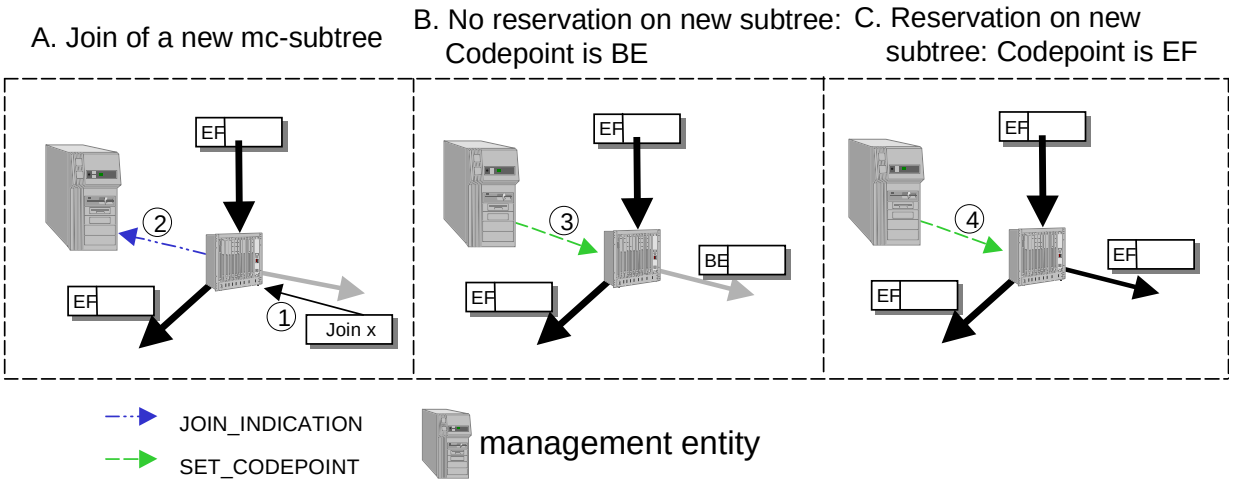


Fig. 6. Sequence of the proposed solution:
A. Join of a new multicast subtree.
B. If no reservation is made for the new subtree, all packets for this group will be remarked to (L)BE.
C. If a reservation has been made before (or later), the management entity tells the router not to remark the packets, and the new subtree gets the QoS.

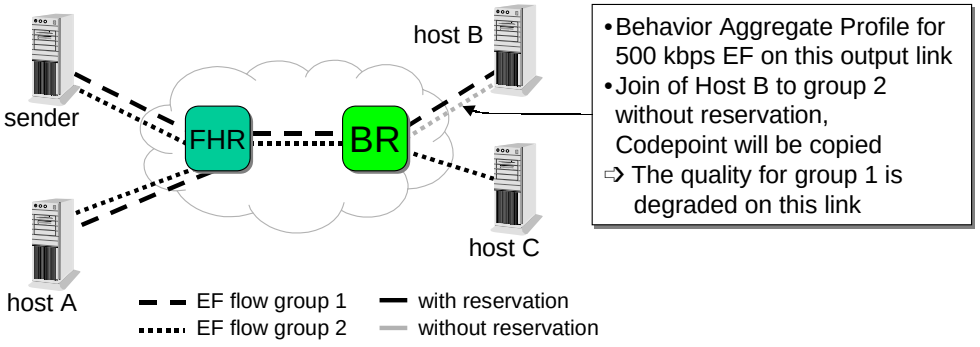
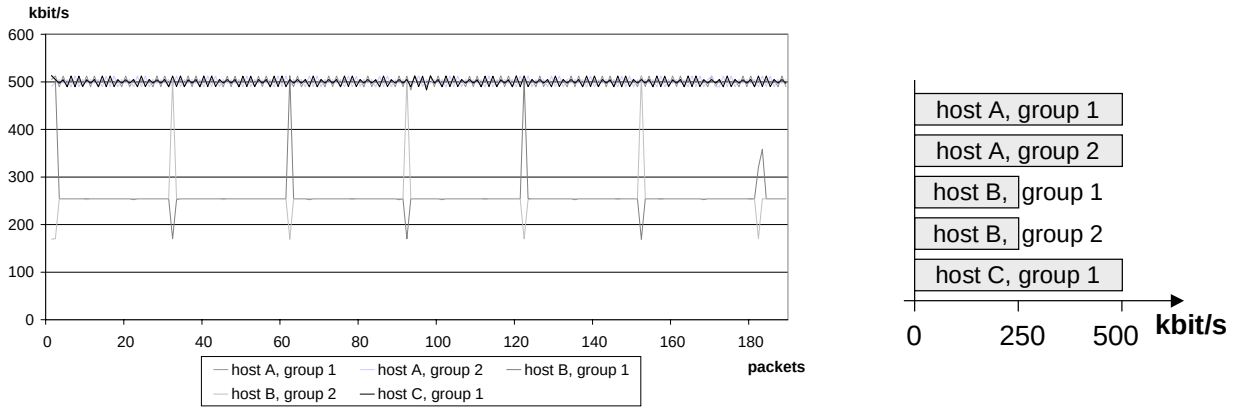
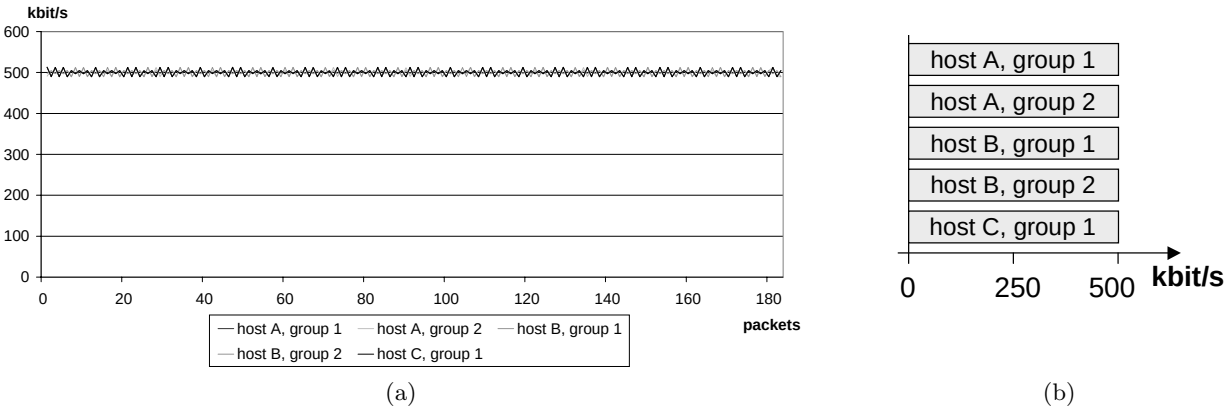


Fig. 7. Evaluation of the Neglected Reservation Subtree Problem from Fig. 4



(a) (b)

Fig. 8. Results of measurement 1: (a) measured bandwidth for a period of 190 packet times and (b) average bandwidth of these flows.



(a) (b)

Fig. 9. Results of measurement 2 (with proposed solution): (a) measured bandwidth for a period of 190 packet times and (b) average bandwidth of these flows.