Differentiated Internet Services

Florian Baumgartner, Torsten Braun, Hans Joachim Einsiedler and Ibrahim Khalil

Institute of Computer Science and Applied Mathematics
University of Berne, CH-3012 Bern, Switzerland,
Tel +41 31 631 8681 / Fax +41 31 631 39 65
http://www.iam.unibe.ch/~rvs/

Abstract. With the grown popularity of the Internet and the increasing use of business and multimedia applications the users' demand for higher and more predictable quality of service has risen. A first improvement to offer better than best-effort services was made by the development of the integrated services architecture and the RSVP protocol. But this approach proved only suitable for smaller IP networks and not for Internet backbone networks. In order to solve this problem the concept of differentiated services has been discussed in the IETF, setting up a working group in 1997. The Differentiated Services Working Group of the IETF has developed a new concept which is better scalable than the RSVP-based approach. Differentiated Services are based on service level agreements (SLAs) that are negotiated between users and Internet service providers. With these SLAs users describe the packets which should be transferred over the Internet with higher priority than best-effort packets. The SLAs also define parameters such as the desired bandwidth for these higher priority packets. The implementation of this concept requires additional functionality such as classification, metering, marking, shaping, policing etc. within routers at the domain boundaries. This paper describes the Differentiated Service architecture currently being defined by the IETF DiffServ working group and the required components to implement the DiffServ architecture.

1 Introduction

The Internet, currently based on the best-effort model, delivers only one type of service. With this model and FIFO queuing deployed in the network, any non-adaptive sources can take advantage to grab high bandwidth while depriving others. One can always run multiple web browsers or start multiple FTP connections and grab substantial amount of bandwidth by exploiting the best effort model. The Internet is also unable to support real time applications like audio or video.

Incredible rapid growth of Internet has resulted in massive increases in demand for network bandwidth performance guarantees to support both existing and new applications. In order to meet these demands, new Quality of Service (QoS) functionalities need to be introduced to satisfy customer requirements including efficient handling of both mission critical and bandwidth hungry web applications. QoS, therefore, is needed for various reasons:

- Better control and efficient use of networks resources (e.g. bandwidth).
- Enable users to enjoy multiple levels of service differentiation.
- Special treatment to mission critical applications while letting others to get fair treatment without interfering with mission sensitive traffic.
- Business Communication.
- Virtual Private Networks (VPN) over IP.

1.1 A Pragmatic Approach to QoS

A pragmatic approach to achieve good quality of service (QoS) is an adaptive design of the applications to react to changes of the network characteristics (e.g. congestion). Immediately after detecting a congestion situation the transmission rate may be reduced by increasing the compression ratio or by modifying the A/V coding algorithm. For this purpose functions to monitor quality of service are needed. For example, such functions are provided by the Real-Time Transport Protocol (RTP) [SCFJ96] and the Real-Time Control Protocol (RTCP). A receiver measures the delay and the rate of the packets received. This information is transmitted to the sender via RTCP. With this information the sender can detect if there is congestion in the network and adjust the transmission rate accordingly. This may affect the coding of the audio or video data. If only a low data rate is achieved, a coding algorithm with lower quality has to be chosen. Without adaptation the packet loss would increase, making the transmission completely useless. However, rate adaptation is limited since many applications need a minimum rate to work reasonably.

1.2 Reservation-based Approach

To achieve the QoS objective as mentioned in the earlier section, basically two approaches can be offered in a heterogeneous network like the Internet :

Integrated Service Approach: The Integrated Services Architecture based on the Resource Reservation Setup Protocol (RSVP) is based on absolute network reservation for specific flows. This can be supported in small LANs, where routers can store a small number of flow states. In the backbone, however, it would be extremely difficult, if not impossible, to store millions of flow states even with very powerful processors. Moreover, for short-lived HTTP connections, it is probably not practical to reserve resources in advance.

Differentiated Service (DiffServ): To avoid the scaling problem of RSVP, a differentiated service is provided for an aggregated stream of packets by marking the packets and invoking some differentiation mechanism (e.g. forwarding treatment to treat packets differently) for each marked packet on the nodes along the stream's path. A very general approach of this mechanism is to define a service profile (a contract between

a user and the ISP) for each user (or group of users), and to design other mechanisms in the router that favors traffic conforming to those service profiles. These mechanisms might be classification, prioritization and resource allocation to allow the service provider to provision the network for each of the offered classes of service in order to meet the application (user) requirements.

2 DiffServ Basics and Terminology

The idea of differentiated services is based on the aggregation of flows, i.e. reservations have to be made for a set of related flows (e.g. for all flows between two subnets). Furthermore, these reservations are rather static since no dynamic reservations for a single connection are possible. Therefore, one reservation may exist for several, possibly consecutive connections.

IP packets are marked with different priorities by the user (either in an end system or at a router) or by the service provider. According to the different priority classes the routers reserve corresponding shares of resources, in particular bandwidth. This concept enables a service provider to offer different classes of QoS at different costs to his customers.

The differentiated services approach allows customers to set a fixed rate or a relative share of packets which have to be transmitted by the ISP with high priority. The probability of providing the requested quality of service depends essentially on the dimensions and configuration of the network and its links, i.e. whether individual links or routers can be overloaded by high priority data traffic. Though this concept cannot guarantee any QoS parameters as a rule it is more straightforward to be implemented than continuous resource reservations and it offers a better QoS than mere best-effort services.

2.1 Popular Services of the DiffServ Approach

At present, several proposals exist for the realization of differentiated services. Examples are:

Assured and Premium Services: The approach allowing the combination of different services like Premium and Assured Service seems to be very promising. In both approaches absolute bandwidth is allocated for aggregated flows. They are based on packet tagging indicating the service to be provided for a packet. Actually, assured service does not provide absolute bandwidth guarantee but offers soft guarantee with high probability that traffic marked with high priority tagging will be transmitted with high probability.

User Share Differentiation and Olympic Service: An alternative approach called User-Share Differentiation (USD) assigns bandwidth proportionally to aggregated flows in the routers (for example all flows from

or to an IP address or a set of addresses). A similar service is provided by the Olympic service. Here, three priority levels are distinguished assigning different fractions of bandwidth to the three priority levels gold, silver and bronze, for example 60% for gold, 30% for silver and 10% for bronze.

2.2 DS byte marking

In differentiated services networks where service differentiation is the main objective, the differentiation mechanisms are triggered by the so-called DS byte (or ToS byte) marking of the IP packet header. Various service differentiation mechanisms (queuing disciplines), as we will study them in section 3, can be invoked dependent on the DS byte marking. Therefore, marking is one of most vital DS boundary enabling component and all DS routers must implement this facility.

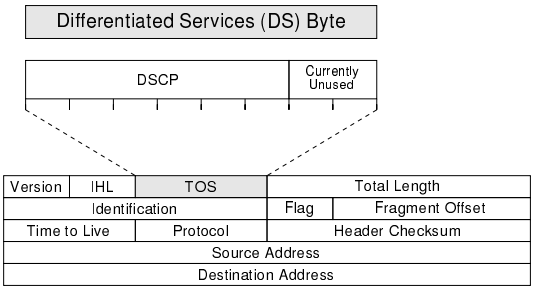

Fig. 1. DS byte in IPv4 [NBBB98]

In the latest proposal for packet marking the the first bit for IN or OUT-of-Profile traffic, the first 6 bits, called Differentiated Services Code point (DSCP), are used to invoke PHBs (see Figure 1). Router implementation should support recommended code point-to-PHB mappings. The default PHB, for example, is 000000. Since the DSCP field has 6 bits, the number of code points that can defined is $2^6 = 64$. This proposal will be the basis of future DiffServ development.

Many existing routers already use IP precedence field to invoke various PHB treatment similar to the fashion of DSCP. To remain compatible, routers can be configured to ignore bit 3,4 and 5. Code point 101000 and 101010 would, therefore, map to the same PHB. Router designers must consider the semantics described above in their implementation and do necessary and appropriate mapping in order to remain compatible with old systems.

2.3 Per Hop Behavior (PHB)

An introduction of PHB has already been given while discussing DS byte marking 2.2. Further [BW98] writes: *"Every PHB is the externally observable*

forwarding behavior applied at a DS capable node to a stream of packets that have a particular value in the bits of the DS field (DS code point). PHBs can also be grouped when it is necessary to describe the several forwarding behaviors simultaneously with respect to some common constraints."

However, there is no rigid assignments of PHBs to DSCP bit patterns. These has several reasons:

- There are (or will be) a lot of more PHBs defined, than DSCPs available, making a static mapping impossible.
- The understanding of good choices of PHBs is at the beginning.
- It is desirable to have complete flexibility in the correspondence of PHB values and behaviors.
- Every ISP shall be able to create/map PHBs in his DiffServ domain.

For these reasons there are no static mappings between DS code points and PHBs. The PHBs are enumerated as they become defined and can be mapped to every DSCP within a DiffServ domain. As long as the enumeration space contains a large number of values (2^{32}), there is no danger of running out of space to list the PHB values. This list can be made public for maximum interoperability. Because of this interoperability, mappings between PHBs and DSCPs are proposed, even when every ISP can choose other mappings for the PHBs in his DiffServ domain.

Until now, two PHBs and corresponding DSCPs have been defined.

Table 1. The 12 different AF code points

Drop Precedences	AF Code points			
	Class 1	Class 2	Class 3	Class 4
Low Drop Precedence	001010	010010	011010	100010
Medium Drop Precedence	001100	010100	011100	100100
High Drop Precedence	001110	010110	011110	100110

Assured Forwarding PHB: Based on the current Assured Forwarding PHB (AF) group [HBWW99], a provider can provide four independent AF classes where each class can have one of three drop precedence values. These classes are not aggregated in a DS node and Random Early Detection (RED) [FJ93] is considered to be the preferred discarding mechanism. This required altogether 12 different AF code points as given in table 1.

In a Differentiated Service (DS) Domain each AF class receives a certain amount of bandwidth and buffer space in each DS node. Drop precedence indicates relative importance of the packet within an AF class. During congestion, packets with higher drop precedence values are discarded first

to protect packets with lower drop precedence values. By having multiple classes and multiple drop precedences for each class, various levels of forwarding assurances can be offered. For example, Olympic Service can be achieved by mapping three AF classes to it's gold, silver and bronze classes. A low loss, low delay, low jitter service can also be achieved by using AF PHB group if packet arrival rate is known in advance. AF doesn't give any delay related service guarantees. However, it is still possible to say that packets in one AF class have smaller or larger probability of timely delivery than packets in another AF class. The Assured Service can be realized with AF PHBs.

Expedited Forwarding PHB: The forwarding treatment of the Expedited Forwarding (EF) PHB [JNP98] offers to provide higher or equal departure rate than the configurable rate for aggregated traffic. Services which need end-to-end assured bandwidth and low loss, low latency and low low jitter can use EF PHB to meet the desired requirements. One good example is premium service (or virtual leased line) which has such requirements. Various mechanisms like Priority Queuing, Weighted Fair Queuing (WFQ), Class Based Queuing (CBQ) are suggested to implement this PHB since they can preempt other traffic and the queue serving EF packets can be allocated bandwidth equal to the configured rate. The recommended code point for the EF PHB is 101110.

2.4 Service Profile

A service profile expresses an expectation of a service received by a user or group of users or behavior aggregate from an ISP. It is, therefore, a contract between a user and provider and also includes rules and regulations a user is supposed to obey. All these profile parameters are settled in an agreement called Service Level Agreement (SLA). It also contains Traffic Conditioning Agreement (TCA) as a subset, to perform traffic conditioning actions (described in the next subsection) and rules for traffic classification, traffic re-marking, shaping, policing etc. In general, a SLA might include performance parameters like peak rate, burst size, average rate, delay and jitter parameters, drop probability and other throughput characteristics. An Example is:

Service Profile 1: Code point: X, Peak rate= 2Mbps, Burst size=1200 bytes,
avg. rate = 1.8 Mbps

Only a static SLA, which usually changes weekly or monthly, is possible with today's router implementation. The profile parameters are set in the router manually to take appropriate action. Dynamic SLAs change frequently and need to be deployed by some automated tool which can renegotiate resources between any two nodes.

2.5 Traffic Conditioner

Traffic conditioners [BBC+98] are required to instantiate services in DS capable routers and to enforce service allocation policies. These conditioners are, in general, composed of one or more of the followings: classifiers, markers, meters, policers, and shapers. When a traffic stream at the input port of a router is classified, it then might have to travel through a meter (used where appropriate) to measure the traffic behavior against a traffic profile which is a subset of SLA. The meter classifies particular packets as IN or OUT-of-profile depending on SLA conformance or violation. Based on the state of the meter further marking, dropping, or shaping action is activated.

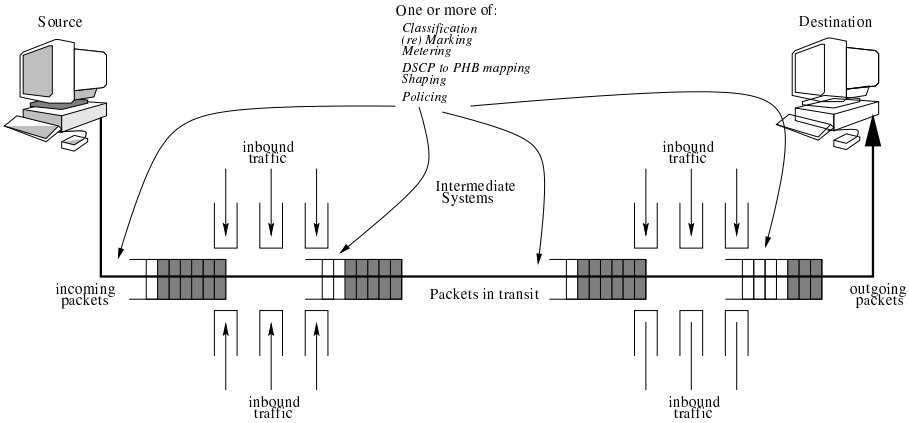

Fig. 2. DS Traffic Conditioning in Enterprise Network (as a set of queues)

Traffic Conditioners can be applied at any congested network node (Figure 2) when the total amount of inbound traffic exceeds the output capacity of the switch (or router). In Figure 2 routers between source and destination are modeled as queues in an enterprise network to show when and where traffic conditioners are needed. For example, routers may buffer traffic (i.e. shape them by delaying) or mark them to be discarded later during medium network congestion, but might require to discard packets (i.e. police traffic) during heavy network congestion when queue buffers fill up. As the number of routers grows in a network, congestion increases due to expanded volume of traffic and hence proper traffic conditioning becomes more important.

Traffic conditioners might not need all four elements. If no traffic profile exists then packets may only pass through a classifier and a marker.

Classifier: Classifiers categorize packets from a traffic stream based on the content of some portion of the packet header. It matches received packets to statically or dynamically allocated service profiles and pass those packets to an element of a traffic conditioner for further processing. Classifiers

must be configured by some management procedures in accordance with the appropriate TCA.
Two types of classifiers exist:
BA Classifier: classifies packets based on patterns of DS byte (DS code point) only.
MF classifier: classifies packets based on any combination of DS field, protocol ID, source address, destination address, source port, destination port or even application level protocol information.

Markers: Packet markers set the DS field of a packet to a particular code point, adding the marked packet to a particular DS behavior aggregate. The marker can (i) mark all packets which are mapped to a single code point, or (ii) mark a packet to one of a set of code points to select a PHB in a PHB group, according to the state of a meter.

Meters: After being classified at the input of the boundary router, traffic from each class is typically passed to a meter. The meter is used to measure the rate (temporal properties) at which traffic of each class is being submitted for transmission which is then compared against a traffic profile specified in TCA (negotiated between the DiffServ provider and the DiffServ customer). Based on the the comparison some particular packets are considered conforming to the negotiated profile (IN-profile) or non-conforming (OUT-of-profile). When a meter passes this state information to other conditioning functions, an appropriate action is triggered for each packet which is either IN or OUT-of-profile (see Table 1).

Shapers: Shapers delay some packets in a traffic stream using a token bucket in order to force the stream into compliance with a traffic profile. A shaper usually has a finite-size buffer and packets are discarded if there is not sufficient buffer space to hold the delayed packets. Shapers are generally placed after either type of classifier.For example, shaping for EF traffic at the interior nodes helps to improve end to end performance and also prevents the other classes from being starved by a big EF burst. Only either a policer or a shaper is supposed to appear in the same traffic conditioner.

Policer: When classified packets arrive at the policer it monitors the dynamic behavior of the packets and discard or re-mark some or all of the packets in order to force the stream into compliance (i.e. force them to comply with configured properties like rate and burst size) with a traffic profile. By setting the shaper buffer size to zero (or a few packets) a policer can be implemented as a special case of a shaper. Like shapers policers can also be placed after either type of classifier. Policers, in general, are considered suitable to police traffic between a site and a provider(edge router) and after BA classifiers (backbone router). However, most researchers agree that policing should not be done at the interior nodes since it unavoidably involves flow classification. Policers are usually present in ingress nodes and could be based on simple token bucket filters.

3 Realizing PHBs: The Queuing Components

Since differentiated service is a kind of service discrimination, some traffic need to be handled with priority, some of the traffic needs to be discarded earlier than other traffic, some traffic needs to be serviced faster, and in general, one type of traffic always needs to better than the other. In earlier sections we have discussed about service profile and PHBs. It was made clear that in order to conform to the contracted profile and implement the PHBs, queuing disciplines play a crucial role. The queuing mechanisms typically need to be deployed at the output port of a router.

Since we need different kinds of differentiation under specific situations, the right queuing component (i.e PHB) needs to be invoked by the use of a particular code point. In this section, therefore, we will describe some of the most promising mechanisms which have already been or deserve to be considered for implementation in varieties of DS routers.

3.1 Absolute Priority Queuing

In absolute priority queuing (Figure 3), the scheduler gives higher-priority queues absolute preferential treatment over lower priority queues. Therefore, the highest priority queue receives the fastest service, and the lowest priority queue experiences slowest service among the queues.

The basic working mechanism is as follows: the scheduler would always scan the priority queues from highest to lowest to find the highest priority packet and then transmit it. When that packet has been completely served, the scheduler would start scanning again. If any of the queues overflows, packets are dropped and an indication is sent to the sender.

While this queuing mechanism is useful for mission critical traffic (since this kind of traffic is very delay sensitive) this would definitely starve the lower priority packets of the needed bandwidth.

3.2 WFQ

WFQ [Kes91](Figure 4)is a discipline that assigns a queue for each flow. A weight can be assigned to each queue to give a different proportion of the network capacity. As a result, WFQ can provide protection against other flows.

WFQ can be configured to give low-volume traffic flows preferential treatment to reduce response time and fairly share the remaining bandwidth between high volume traffic flows. With this approach bandwidth hungry flows are prevented from consuming much of network resources while depriving other smaller flows.

WFQ does the job of dynamic configuration since it adapts automatically to the changing network conditions. TCP congestion control and slow-start

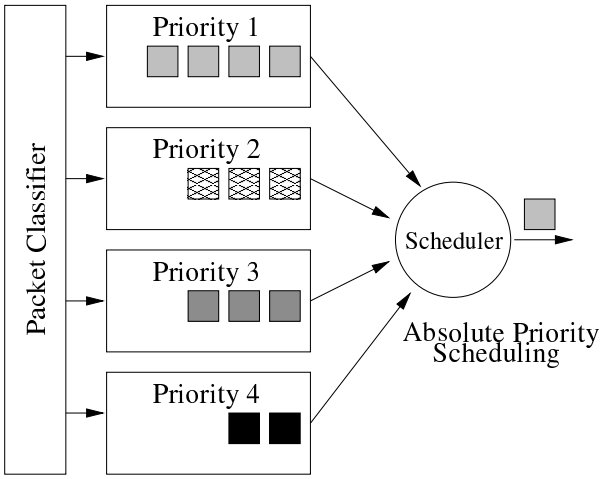

Fig. 3. Absolute Priority Queuing. The queue with the highest priority is served at first

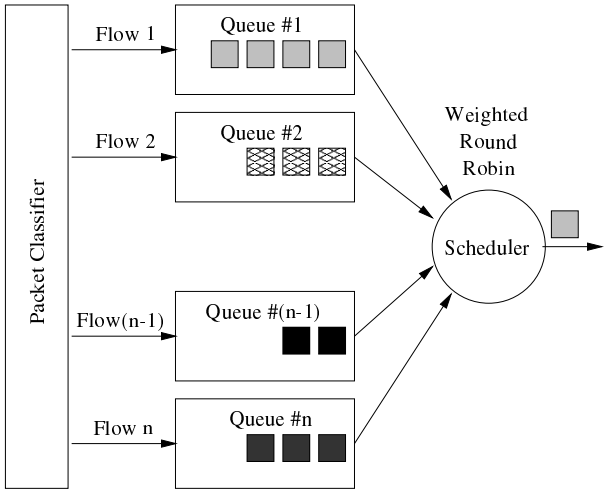

Fig. 4. Weighted Fair Queuing (WFQ)

features are also enhanced by WFQ, resulting in predictable throughput and response time for each active flow.

The weighted aspect can be related to values in the DS byte of the IP header. A flow can be allocated more access to queue resources if it has a higher precedence value.

3.3 Class Based Queuing (CBQ)

In an environment where bandwidth must be shared proportionally between users, CBQ [FJ95] (Figure 6) provides a very flexible and efficient approach to

first classifying user traffic and then assigning a specified amount of resources to each class of packets and serving those queues in a round robin fashion.

A class can be an individual flow or aggregation of flows representing different applications, users, departments, or servers. Each CBQ traffic class has a bandwidth allocation and a priority. In CBQ, a hierarchy of classes (Figure 5) is constructed for link sharing between organizations, protocol families, and traffic types. Different links in the network will have different link-sharing structures. The link sharing goals are:

- Each interior or leaf class should receive roughly its allocated link-sharing bandwidth over appropriate time intervals, given the sufficient demand.
- If all leaf and interior classes with sufficient demand have received at least their allocated link-sharing bandwidth, the distribution of any excess bandwidth should not be arbitrary, but should follow some set of reasonable guidelines.

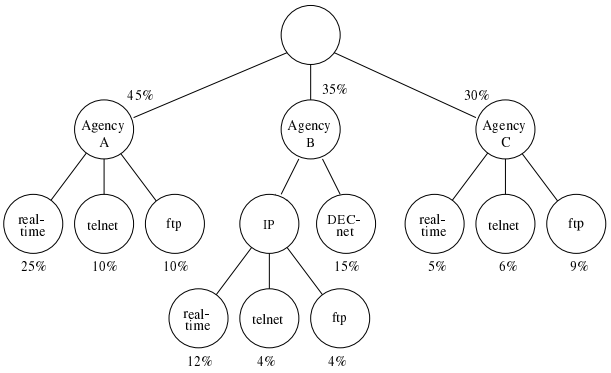

Fig. 5. Hierarchical Link-Sharing

The granular level of control in CBQ can be used to manage the allocation of IP access bandwidth across the departments of an enterprise, to provision bandwidth to the individual tenants of a multi-tenant facility.

Other than the classifier that assigns arriving packets to an appropriate class, there are three other main components that are needed in this CBQ mechanism: scheduler, rate-limiter (delayer) and estimator.

Scheduler: In a CBQ implementation, the packet scheduler can be implemented with either a packet-by-packet round robin (PRR) or weighted round robin (WRR) scheduler. By using priority scheduling the scheduler uses priorities, first scheduling packets from the highest priority level. Round-robin scheduling is used to arbitrate between traffic classes within the same priority level. In weighted round robin scheduling the scheduler uses weights proportional to a traffic class's bandwidth allocation. This weight finally allocates the number of bytes a traffic class is allowed to

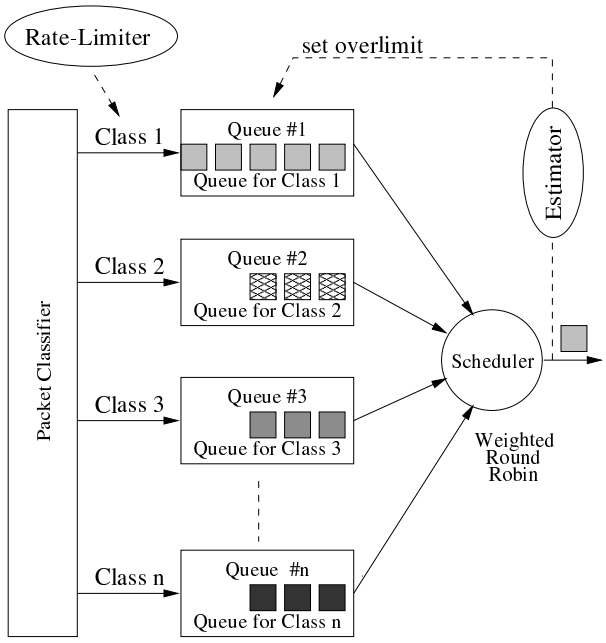

Fig. 6. Class Based Queuing: Main Components

send during a round of the scheduler. Each class at each round gets to send its weighted share in bytes, including finishing sending the current packet. That class's weighted share for the next round is decremented by the appropriate number of bytes. When a packet to be transmitted by a WRR traffic class is larger than the traffic class's weight but that class is underlimit[1], the packet is still sent, allowing the traffic class to borrow ahead from its weighted allotment for future rounds of the round-robin.

Rate-Limiter: If a traffic class is overlimit[2] and is unable to borrow from it's parent classes, the scheduler starts the overlimit action which might include simply dropping arriving packets for such a class or rate-limit overlimit classes to their allocated bandwidth. The rate-limiter computes the next time that an overlimit class is allowed to send traffic. Unless this future time has arrived, this class will not be allowed to send another packet until .

Estimator: The estimator estimates the bandwidth used by each traffic class over the appropriate time interval and determines whether each class is over or under its allocated bandwidth.

[1] If a class has used less than a specified fraction of its link sharing bandwidth (in bytes/sec, as averaged over a specified time interval)

[2] If a class has recently used more than its allocated link sharing bandwidth (in bytes/sec, as averaged over a specified time interval)

3.4 Random Early Detection (RED)

Random Early Detection (RED) [FJ93] is designed to avoid congestion by monitoring traffic load at points in the network and stochastically discarding packets when congestion starts increasing. By dropping some packets early rather than waiting until the buffer is full, RED keeps the average queue size low and avoids dropping large numbers of packets at once to minimize the chances of global synchronization.Thus, RED reduces the chances of tail drop and allows the transmission line to be used fully at all times. This approach has certain advantages:

- bursts can be handled better, as always a certain queue capacity can be reserved for incoming packets.
- by the lower average queue length real-time applications are better supported.

The working mechanism of RED is quite simple. It has two thresholds, minimum threshold *X1* and a maximum threshold *X2* for packet discarding or admission decision which is done by a dropper. Referring to Figure 7, when a packet arrives at the queue, the average queue (av_queue) is computed. If, $av\_queue < X1$, the packet is admitted to the queue; if $av\_queue \geq X2$, the packet is dropped. In the case, when the average queue size falls between the thresholds $X1 < av\_queue < X2$, the arriving packet is either dropped or queued, mathematically saying, it is dropped with linearly increasing probability.

When congestion occurs, the probability that the RED notifies a particular connection to reduce its window size is approximately proportional to that connection's share of the bandwidth. The RED congestion control mechanism monitors the average queue size for each output queue and using randomization choose connections to notify of that congestion.

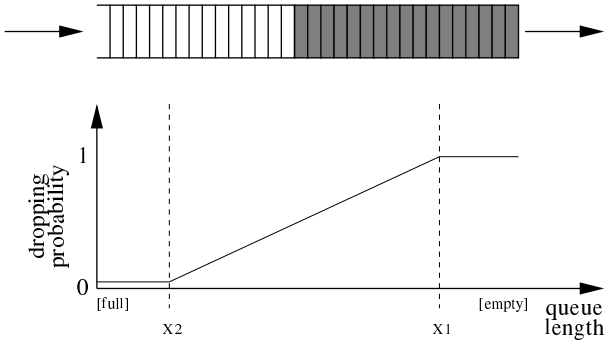

Fig. 7. Random Early Detection

It is very useful to the network since it has the ability to flexibly specify traffic handling policies to maximize throughput under congestion conditions.

RED is especially able to split bandwidth between TCP data flows in a fair way as lost packets automatically cause a reduction to a TCP data flow's packet rate. More problematic is the situation if non TCP conforming data flows (e.g. UDP based real-time or multicast applications) are involved. Flows not reacting to packet loss have to be handled by reducing their data rate specially to avoid an overloading of the network.

In general, RED statistically drops more packets from large users than from small ones. Therefore, traffic sources that generate the most traffic are more likely to be slowed down than traffic sources that generate little traffic.

3.5 RED with In and Out (RIO)

The queuing algorithm proposed for assured service RIO (RED with In and Out) [CW97] is an extension of the RED mechanism. This procedure shall make sure, that during overload primarily packets with high drop precedence (e.g. best-effort instead of assured service packets) are dropped. A data flow can consist of packets with various drop precedences, which can arrive at a common output queue. So changes to the packet order can be avoided affecting positively the TCP performance.

For in and out-of-profile packets a common queue using different dropping techniques for the different packet types is provided. The dropper for out of profile packets discards packets much earlier (e.g. a lower queue length) than the dropper for in profile packets. Further more the dropping probability for out of profile packets increases more than the probability for in packets. So, it shall be achieved that the probability for dropping in profile packets is kept very low. While the out-dropper used the number of all packets in the queue for the calculation of his probability, the in-dropper only uses the number of in profile packets (see figure 8). Using the same queue both types of packets will have the same delay. This might be a disadvantage of this concept. By dropping all out-of-profile packets at a quite small queue length this effect can be reduced but not eliminated.

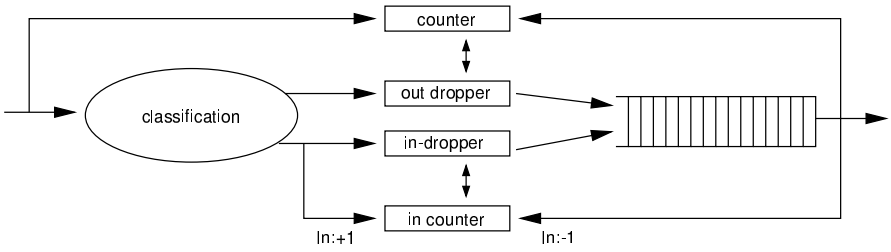

Fig. 8. RIO-Queuing

4 Differentiated Services in End-to-End Scenarios

4.1 Premium Service and Expedited Forwarding

With Premium Service the user negotiates with his ISP a maximum bandwidth for sending packets through the ISP network. Furthermore, the aggregated flow is described by the packets' source and destination addresses or address prefixes. In Figure 9 users and ISPs have agreed on a rate of three packets/s for traffic from A to B. The user configures the first-hop router in the individual subnet accordingly. In the example above a packet rate of two packets/s is allowed in every first-hop router as it can be expected that no two end systems will use the full bandwidth of two packets/s at the same time.

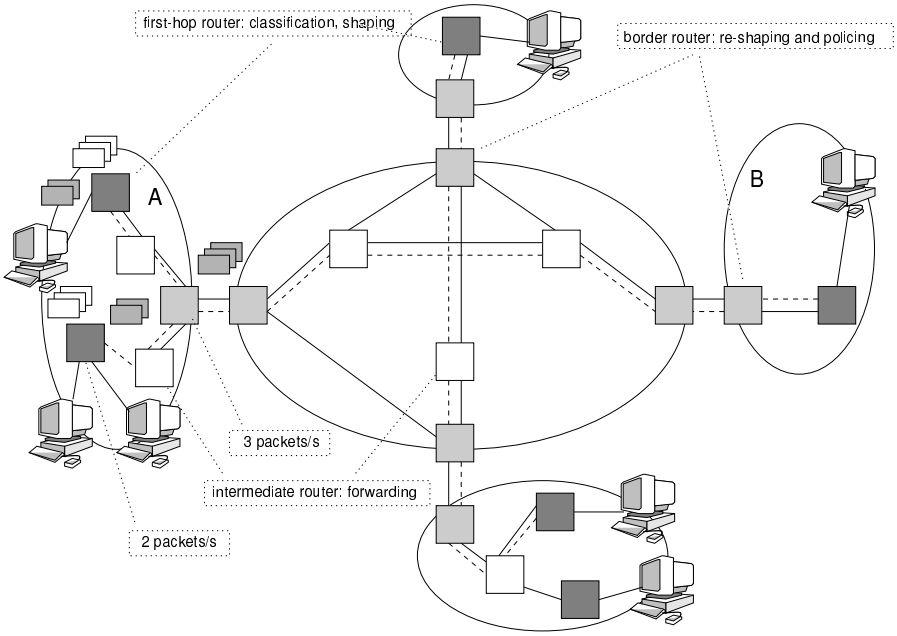

Fig. 9. Premium Service

First-hop routers have the task to classify the packets received from the end systems, i.e. to analyze if the Premium Service shall be provided to the packets or not. If yes, the packets are tagged as Premium Service and the data stream is shaped according to the maximum bandwidth. The user's border router re-shapes the stream (e.g. three packets per second) and transmits the packets to the ISP's border router, which performs policing functions, i.e. it checks whether the user's border router remains below the negotiated bandwidth of three packets/s. If each of the two first-hop routers allows two

packets/s, one packet per second will be dropped by shaping or policing at the border routers. All first-hop and border-routers own two queues, one for EF-packets and one for all other (see Figure 9). If the EF-queue contains packets these are transmitted prior to others. The implementation of two queues in every router of the network (ISP and user network) equals to the realization of a virtual network for Premium Service traffic.

Premium Service offers a service corresponding to a private leased line, with the advantage of making free network capacities available to other tasks, resulting in lower fees for the users.

4.2 Assured Service

A potential disadvantage of Premium Service is the weak support for bursts and the fact that a user has to pay even if he is not using the whole bandwidth. The Assured Service tries to offer a service which cannot guarantee bandwidth but provides a high probability that the ISP transfers high-priority-tagged packets reliably. The definition of concrete services has not yet happened, but it is obvious to offer services similar to the IntServ controlled load service. The probability for packets to be transported reliably depends on the network capacity. An ISP may choose the sum of all bandwidths for Assured Service to remain below the bandwidth of the weakest link. In this case, only a small portion of the available capacity may be allocated in the ISP network. An advantage of the Assured Service is that users do not have to establish a reservation for a relative long time. With ISDN or ATM, users might be unable to use the reserved bandwidth because of the burstiness of their traffic, whereas Assured Service allows the transmission of short time bursts.

With the Assured Service the user negotiates a service profile with his service provider, e.g. the maximum amount or rate of high priority, i.e. Assured Service, packets. The user may then tag his packets as high priority within the end system or the first-hop router, i.e. assign them a tag for assured forwarding (AF) (see Figure 10). To avoid modifications in the end systems the first-hop router may analyze the packets with respect to their IP addresses and UDP-/TCP-Port and then assign them the according priority, i.e. set the AF-DSCP for conforming Assured Service packets. The maximum rate of high-priority (AF-DSCP) packets must not be exceeded. This is done by (re-)classification in the first-hop routers and in the user's border routers at the border to the ISP network. Nevertheless, the service provider has to check if the user remains below the maximum rate for high priority packets and apply corrective actions such as policing if necessary.

For example, the border router at the network entrance will tag the non-conforming packet as low priority (out of service, out of profile). An alternative would be to charge higher fees for non-conforming packets by the ISP. The tagging of low and high priority packets is done by use of the DS byte.

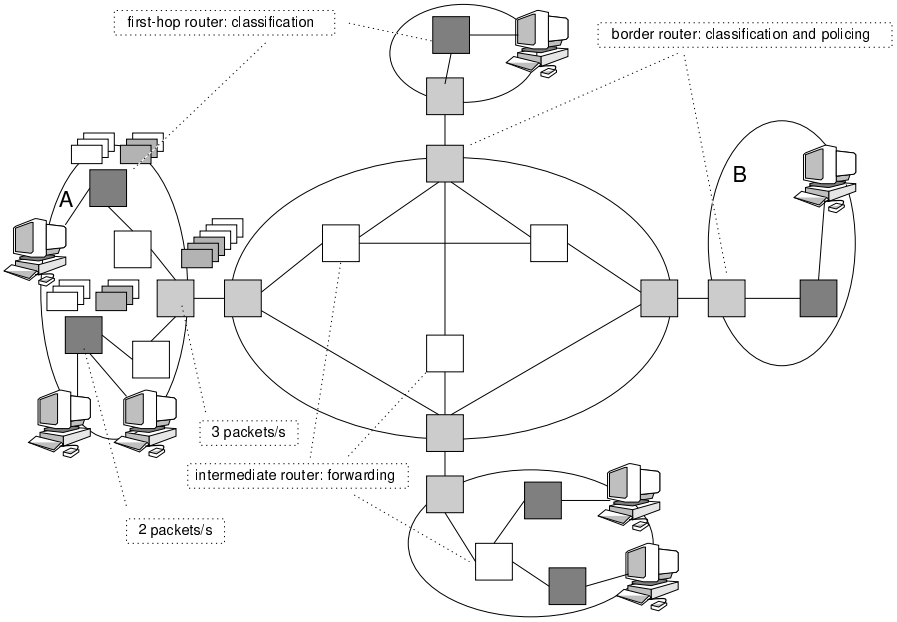

Fig. 10. Assured Service

Bursts are supported by making buffer capacity available for buffering bursty traffic. Inside the network, especially in backbone networks bursts can be expected to be compensated statistically.

4.3 Traffic Conditioning for Assured and Premium Service

The implementation of Assured and Premium Service requires several modifications of the routers. Mainly classification, shaping, and policing functions have to be performed to the router. These functions are necessary at the border between two networks, for example at the transition of the customer network to the ISP or between the ISPs. Service profiles have to be negotiated between the ISPs similar to the transition to the user.

First-hop router Figure 11 shows the first-hop router function for Premium and Assured Service. Received packets are classified and according to this the AF or EF-DSCP is set if the packet should be supported with Assured or Premium Service. As a parameter for the classification, source and destination addresses or information of higher protocols (e.g. port numbers) may be used. There are separate queues for each AF class, for EF and best effort traffic. So, a pure best-effort packet will be forwarded directly to the best-effort RED queue and the Assured Service packets get to their RED queues. The Assured Service packets are checked whether they conform to the service profile. The

drop precedence will only be kept unchanged if the Assured Service bucket contains a token. Otherwise the drop precedence will be increased. The RED-based queuing shall guarantee that AF packets with higher drop precedence are dropped prior to AF packets with lower drop precedence, if the capacity is exceeded.

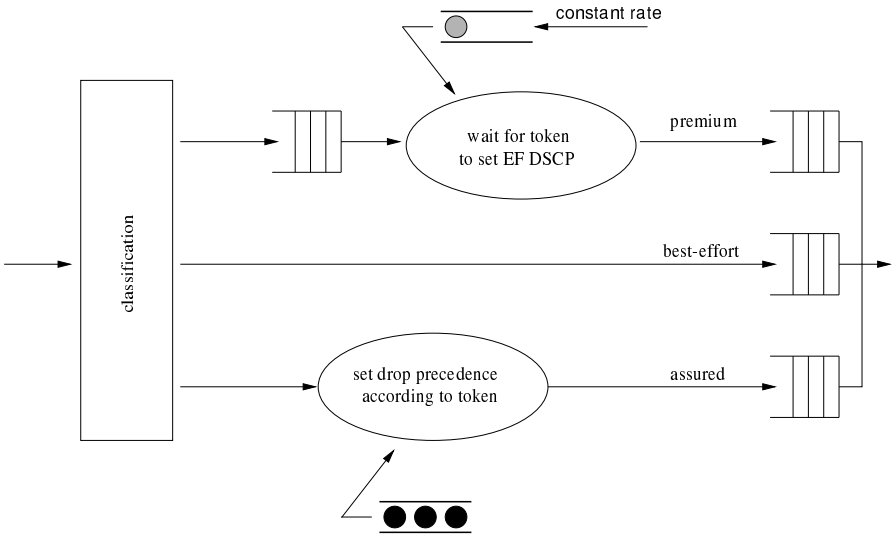

Fig. 11. First-hop router for Premium, Assured and best effort services

Border router Similar to the first-hop router an intermediate router has to perform shaping functions in order to guarantee that not more than the allowed packet rate is transmitted to the ISP. This is important since the ISP will check whether the user remains within the negotiated service profile. The border router in Figure 12 will therefore drop non conforming Premium service packets and increase the drop priority of non conforming Assured Service packets. Packets within an AF class but with different precedence values share the same queue since both types of packets may belong to the same source. A common queue avoids re-ordering of packets. This is especially important for TCP performance reasons.

First-Hop and Egress Border Routers Figure 13 shows the working principle of a first hop and an egress router for assured service. An egress border router is the border router, at which the packets are leaving the differentiated service domain. Received packets are classified and the AF DSCP is set, if assured service should be given to the packet. Source and destination addresses and information of higher protocols (e.g. port numbers) may

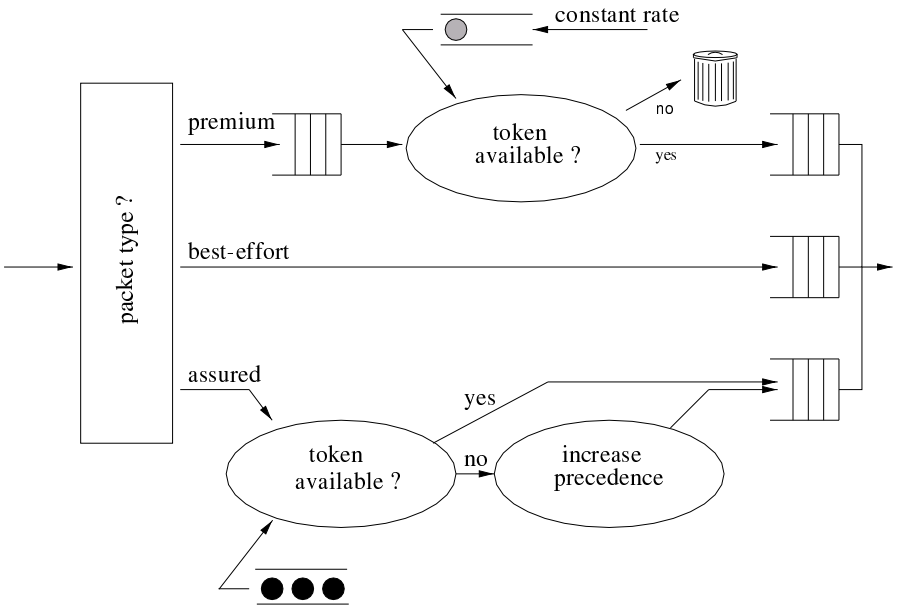

Fig. 12. Policing in a border router

be used as classification parameters.A pure best effort packet will directly be pushed to the output queue.

The AF-DSCP is set according to the availability of a token and then written to the AF output queue. Normal best effort traffic is directly pushed to the best effort queue.

The token buckets are configured according to the SLAs consisting of bit rates and the burst parameter. The bucket may be capable of keeping several tokens to support short time bursts. The bucket's depth depends on the arranged burst properties.

The difference between a first hop and an egress border router is the fact, that at the first hop router a packet is classified for the first time for this task information of higher protocols (TCP ports, type of the application) may be used, whereas the egress border router is capable of changing the drop precedence to meet the negotiated service profile.

Ingress Border Router The ISP has to ensure that the user meets the negotiated traffic characteristics. To achieve this, the ISP has to check in his ingress border router, which transmits the packets into his DS domain whether the user keeps the SLA. So the ingress border router of Figure 14 will change the drop precedence of non conforming packets.

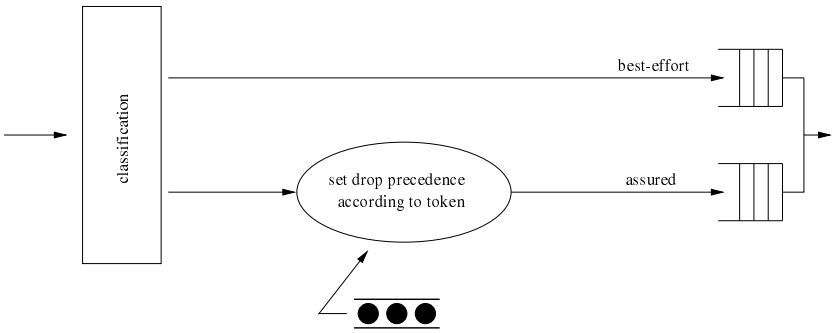

Fig. 13. First hop and egress border router for Assured Service

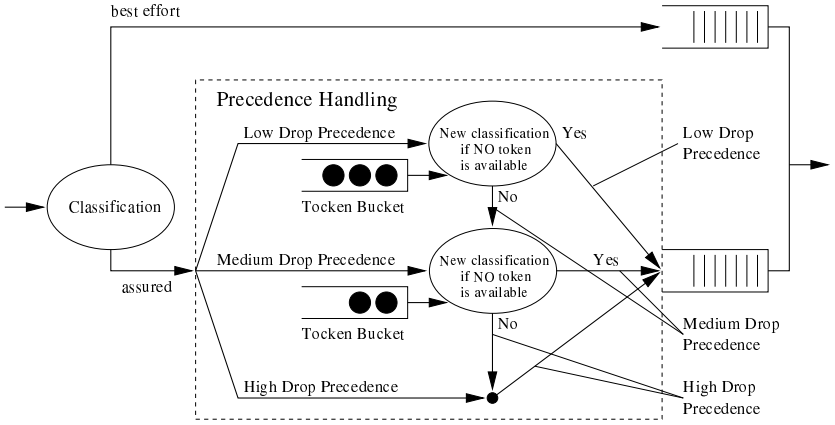

Fig. 14. Ingress border router with three drop precedences for Assured Service

4.4 User-Share Differentiation

Based upon packet tagging Premium and Assured Service models can fulfill the stipulated service parameters like bit rates with a high degree of probability only if the ISP network is dimensioned appropriately and non best-effort traffic is transmitted between certain known networks only.

If for instance two users have contracted a bit rate of 1 Mbps for Assured Service packets with an ISP and both wish to receive data simultaneously at a rate of 1 Mbps each from a WWW server which is connected to the network with a 1.5 Mbps link, the requested quality of service cannot be provided.

The User-Share Differentiation approach [Wan97] avoids this problem by contracting not absolute bandwidth parameters but relative bandwidth shares. A user will be guaranteed only a certain relative amount of the available bandwidth in an ISP network. In practice, the size of this share will be in direct relation to the charged costs.

In Figure 15, user A has allocated only half of the bandwidth of user B and one third of the bandwidth of user C. If A and B access the network on

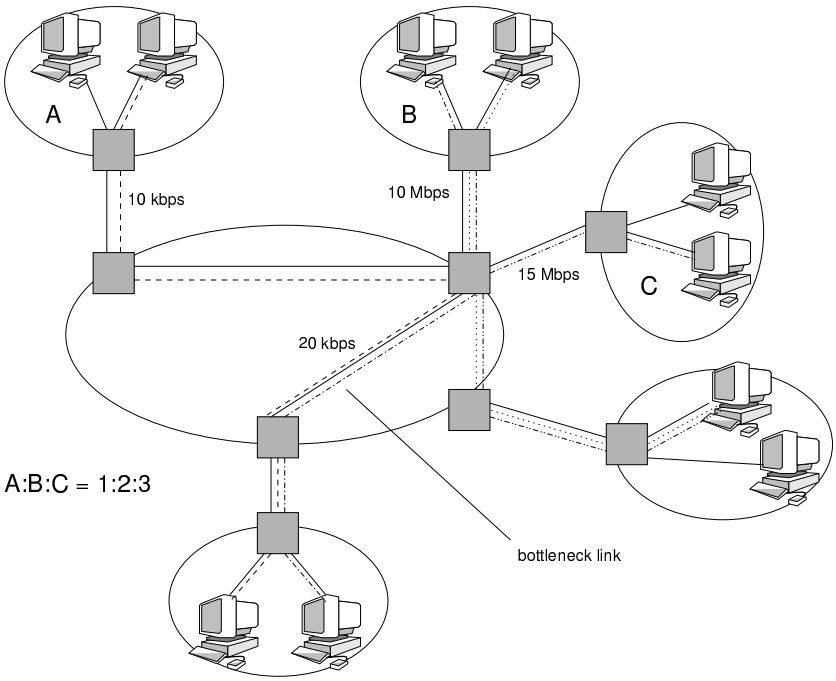

Fig. 15. User Share Differentiation (USD)

low bandwidth links with a capacity of 30 kbps at the same time, e.g. user B will receive a bandwidth of 20 kbps but user C will get merely 10 kbps. If B and C access the same or possibly a different network via a common high bandwidth link with a capacity of 25 Mbps, B will receive 10 Mbps and C only 15 Mbps.

Simpler router configuration is an important advantage of the USD approach. However, absolute bandwidth guarantees cannot supported. An additional drawback is that not only edge routers must be configured (as in the case of Premium or Assured Service) but also interior routers must be configured with the bandwidth shares.

5 Conclusion and Outlook

Standardization of Differentiated services is still under discussion. So far most of discussions have been centered around RED and Assured Service. Virtual Leased Line (or Premium Service) and it's implementations by EF PHB has been recently been discussed in [JNP98] which would require implementation of Priority Queuing, WFQ, CBQ etc. It is not clear where the policing and shaping should take place. Although, both AF and EF PHBs have been proposed, interaction between these two is a debatable issue.

RED and it's variants are complimentary to different scheduling algorithms, and fit very nicely with CBQ. RED is designed to keep queue sizes small (smaller than their maximum in a given implementation), and thus avoid tail drop and global TCP resynchronization. It is, therefore, expected that in router implementation all these service discipline need to coexist and some of those be complementary to each other. Nevertheless, new proposals for both AF and EF PHB strongly suggests that Class Based Queuing (CBQ), WFQ, and their variants will play stronger roles in the implementation of DiffServ.

Regarding interaction between the PHBs the EF draft says that other PHBs can coexist at the same DS node given that the requirements of AF classes are not violated. These requirements include timely forwarding which is at the heart of EF. On the other end, the AF PHB group distinguishes between the classes based on timely forwarding. The AF draft also says that "any other PHB groups may coexist with the AF group within the same DS domain provided that the other PHB groups do not preempt the resources allocated to the AF classes". The question here is: If they coexist should EF have more timely forwarding than the highest timely forwarded AF class by preempting any AF class as the EF document basically states?

What is needed here is EF must leave AF whatever has been allocated for AF.This would mean EF can actually preempt forwarding resources for AF. For example, one could take a 1.5 Mbps link and allow for 64 Kbps of it to be available to EF, with the remaining capacity available to AF. One could also state that EF has absolute priority over AF (up to the 64 Kbps allocated). In this case, EF would preempt AF (so long as it conforms to the 64 Kbps limit) and AF would always be assured that it has 1.5 Mbps - 64 Kbps of the link throughput.

There are lot more issues which are debatable and need attention for further research. However, we should always keep in mind that the whole point of DiffServ is to allow service providers to implement QoS pricing strategies in the first place.

References

[BBC+98] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weis. An architecture for differentiated services. Request for Comments 2475, December 1998.

[BW98] Marty Borden and Christoph White. Management of phbs. Internet Draft `draft-ietf-diffserv-phb-mgmt-00.txt`, August 1998. work in progress.

[CW97] D. Clark and J. Wroclawski. An approach to service allocation in the internet, work in progress. Internet Draft `draft-clark-diff-svc-alloc-00.txt`, Juli 1997. work in progress.

[FJ93] Sally Floyd and Van Jacobson. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, August 1993.

[FJ95] Sally Floyd and Van Jacobson. Link-sharing and resource management models for packet networks. *IEEE/ACM Transactions on Networking*, 3(4), August 1995.

[HBWW99] Juha Heinanen, Fred Baker, Walter Weiss, and John Wroclawski. Assured forwarding phb group. Internet Draft `draft-ietf-diffserv-af-06.txt`, February 1999. work in progress.

[JNP98] Van Jacobson, K. Nichols, and K. Poduri. An expedited forwarding phb. Internet Draft `draft-ietf-diffserv-af-02.txt`, October 1998. work in progress.

[Kes91] S. Keshav. *Congestion Control in Computer Networks*. PhD thesis, Berkeley, September 1991.

[NBBB98] K. Nichols, S. Blake, F. Baker, and D. Black. Definition of the differentiated services field (ds field) in the ipv4 and ipv6 headers. Request for Comments 2474, December 1998.

[SCFJ96] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. Rtp: A transport protocol for real-time applications. Request for Comments 1889, January 1996.

[Wan97] Z. Wang. User-share differentiation (usd) scalable bandwidth allocation for differentiated services. Internet Draft `draft-wang-diff-serv-usd-00.txt`, November 1997. work in progress.