

Differentiated Services –architecture

Mikko Vanhala, 44368D

Teknillinen korkeakoulu

Teletekniikan laboratorio

S-38.128 Teletekniikan erikoistyö

Ohjaaja: Mika Ilvesmäki

Abstract

This work is a literature study of Differentiated Services architecture that is currently being prepared by a workgroup in IETF. Differentiated Services is intended to provide different classes of service for Internet traffic with backward compatibility for the current best-effort service. DiffServ is mainly implemented in area border routers and no necessary modifications to hosts are required.

DiffServ is a scalable approach. It doesn't transit state information across Internet but works on aggregates. It functions on both IPv4 and IPv6 and it uses one header field in an IP-packet. A packet receives forwarding treatment from network devices based on the value on the field. Network devices map the packet to some behavioural aggregate in which the packets receive uniform treatment.

Different services are not standardised within DiffServ. Instead, the building blocks are standardised from which the services can be built.

Keywords: Differentiated Services, QoS, network services, network provisioning.

Index

1. ABBREVIATIONS	4
2. INTRODUCTION	5
2.1. QUALITY OF SERVICE IN THE CURRENT INTERNET	6
2.2. DEMANDS OF THE FUTURE.....	6
3. SERVICE PROVISIONING APPROACHES	7
3.1. RELATIVE PRIORITY MARKING.....	7
3.2. SERVICE MARKING	8
3.3. LABEL SWITCHING	8
3.4. INTEGRATED SERVICES/RSVP	9
3.5. PER-HOP CLASSIFICATION	9
4. OVERVIEW OF DIFFERENTIATED SERVICES	10
4.1. CONCEPTS IN DIFFSERV	11
4.2. DIFFSERV –NETWORK.....	12
4.2.1. <i>Traffic Example</i>	13
5. SERVICE REALISATIONS	15
5.1. USAGE OF IP-HEADER	16
5.2. TRAFFIC CONDITIONING.....	16
5.2.1. <i>Location of Traffic Conditioners</i>	18
5.2.2. <i>SLA and TCA Configurations</i>	18
5.2.3. <i>Network Provisioning</i>	20
5.3. PER-HOP-BEHAVIOURS.....	21
5.3.1. <i>Network Resource Usage</i>	22
5.3.2. <i>Proposed PHBs</i>	23
5.4. SERVICE EXAMPLES.....	26
5.4.1. <i>Better Than Best-Effort</i>	26
5.4.2. <i>Leased Line</i>	27
6. TRAFFIC PROVISIONING BETWEEN PROVIDERS AND CUSTOMERS	29
6.1. CUSTOMERS’ RESPONSIBILITIES.....	29
6.2. INTEROPERABILITY	29
7. REQUIREMENTS FOR NETWORK ELEMENTS	31
7.1. QUEUE MANAGEMENT.....	31
7.2. BOUNDARY ROUTERS	32
7.3. INTEROPERABILITY OF INTSERV/RSVP AND DIFFSERV	33
8. OTHER ASPECTS	35
8.1. SECURITY AND TUNNELLING CONSIDERATIONS	35
8.2. MULTICAST IN DIFFSERV	36
8.3. EFFECTS ON NETWORK MANAGEMENT	36
9. CONCLUSIONS	38
10. REFERENCES	39
A.1 DEFINITIONS	41

1. Abbreviations

AF	Assured Forwarding, a PHB group defined in [AF/DS].
ATM	Asynchronous Transfer Mode
BA	Behaviour Aggregate, separate traffic flows receiving same treatment in a router.
DS, DiffServ	Differentiated Services
DSCP	Differentiated Services Codepoint, value of a field in IP-packets in DS-capable domains based on which the forwarding behaviour is applied.
EF	Expedited Forwarding, a PHB defined in [EF/DS].
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPv4	Internet Protocol version 4, the current widely used IP version.
IPv6	Internet Protocol version 6, the developed future version of IP.
ISP	Internet Service Provider
MF	Multi-Field
MPLS	Multiprotocol Label Switching
NE	Network element, e.g. a router.
PHB	Per-Hop-Behaviour, forwarding behaviour applied to each IP-packet in routers in DS-capable domains.
QoS	Quality of Service
RFC	Request For Comments
RSVP	Resource Reservation Protocol
SLA	Service Level Agreement, an agreement of a service between a provider and a customer.
SNMP	Simple Network Management Protocol
TC	Traffic Conditioner
TCA	Traffic Conditioning Agreement, a part of SLA that describes in detail the agreed service from the technical point of view
TCP	Transmission Control Protocol
TOS	Type of service
VPN	Virtual Private Network

2. Introduction

DiffServ working group in the Internet Engineering Task Force (IETF) is currently preparing an architecture for providing different classes of service into Internet. This work describes the architecture, Differentiated Services (DiffServ, DS), based on the RFCs and drafts prepared by the working group.

Differentiated Services is not yet publicly in use due to its novelty. Instead, there are currently several lightly utilised service differentiation approaches in Internet, such as Integrated Services, static administratively set services and ATM's virtual circuit model. While being promising, the approaches either require modifications to existing applications, dedicated kinds of network technology is needed or the complexity incurs reasonably raised prices of network devices. Mainly due to these reasons, the universal service differentiation method has not yet been encountered. Whether Differentiated Services takes that role can not yet be decided.

Differentiated Services is described in this work starting from more general level and proceeding more into detail. Most concentration is put on services and how they are supposed to be implemented and taken into use. Few possible types of service implementations are presented. Services are discussed from both provider's and customer's point of view. IETF is not trying to describe possible implementations of DiffServ in detail but rather leaves the manufacturers and service providers with vast degree of freedom. Therefore an overview of the framework of required technology for implementing DiffServ is given while the details are left outside of the scope of this work.

As the work on defining and standardising Differentiated Services is still unfinished, the contents of this work must be considered somewhat preliminary. However, only minor changes to the basic architecture can be expected, based on the on-going discussion around the Differentiated Services media [WG/DS].

2.1. Quality of Service in the Current Internet

Traditionally, network service providers (both enterprise and traditional ISPs) provide all customers with the same level of performance (best-effort service). Most service differentiation has been in the pricing structure (individual vs. business rates) or the connectivity type (dial-up access vs. leased line, etc.).

The major difficulties in the current Internet that make providing services with high demands difficult or even impossible are random delay and discarded packets. These are the consequences of treating all packets equally. In other words, packets are dropped randomly in case of congestion because their queuing policy is similar.

2.2. Demands of the Future

More and more different kinds of applications are coming into existence that use Internet in a way that it wasn't originally designed to be used. IP telephone, real-time video and data transport with minimum bandwidth requirement, just to name few, are among these. This all has resulted from rapid growth in Internet and Intranet deployment and usage. The consequences are the massive increases in demand for network bandwidth, performance and flexibility to support both existing and emerging applications and services. [CISCO_QoS]

However, the above-mentioned demand has often left Internet Service Providers with insufficient network capabilities to fully leverage the opportunity. This is due to absence of widely used method for flexibly provisioning networks and thus inability to provide consistent end-to-end QoS. A rather easily employable and flexible architecture, such as DiffServ, could be the key to future.

3. Service Provisioning Approaches

What is actually meant with “a service”? First of all it can be thought of as an agreement between two parties, e.g. customer and provider, one of who receives some special treatment for which he has paid. On the other hand “a service” has been defined as “some significant characteristics of packet transmission in one direction across a set of one or more paths within a network” [ARCH/DS]. However, regardless of what definition is used only one “service” is extensively employed in the present Internet, that being the best-effort service.

Several methods have been proposed for the universal approach providing service differentiation on Internet. The methods can be divided into two categories; those utilised on layer-2 switched infrastructure and those on layer-3 routed infrastructure. The former class requires the support of the link-layer network technology in order services to function as intended. Into this class belongs for example ATM, utilising label switching (or virtual circuit) model. The latter class’s methods are built on the network layer and they are therefore more or less insensitive to the underlying network technology. Into this class belongs for example Integrated Services and Differentiated Services.

On the other hand, service differentiation approaches can be classified into five categories based on the way the services are accomplished in terms of packet handling. These categories are Relative Priority Marking, Service Marking, Label Switching, Integrated Services/RSVP and Per-hop Classification. The categories are presented below.

3.1. *Relative Priority Marking*

In this model the application, host, or proxy node selects a relative priority or “precedence” for a packet (e.g. delay or discard priority), and the network nodes along the transit path apply the appropriate priority forwarding behaviour corresponding to the priority value within the packet’s header. Examples of the relative priority-marking model include IPv4 Precedence marking as defined in [RFC791], 802.5 Token Ring

priority [TR], and the default interpretation of 802.1 traffic classes [802.1]. In IPv4-datagrams the precedence bits, i.e. bits 0-2, of the TOS-field allow senders to indicate the importance of each datagram. In practice, most host and router software ignore type of service. [Co95, ARCH/DS]

3.2. Service Marking

In service marking model the requested service type is associated with each individual packet as an input to route selection. An example of service marking is IPv4 TOS as defined in [RFC1349]. The 'DTRC'-bits in TOS-field mark requested forwarding behaviour, including "minimise delay", "maximise throughput", "maximise reliability", or "minimise cost". Due to generality of the behaviours and the limited codepoint space, only a small range of possible service are realisable. In addition, the "TOS→forwarding behaviour" association in each core network node involves configuration. In practice the utilisation of IPv4 TOS is close to none. [ARCH/DS]

3.3. Label Switching

In this model an end-to-end route is established before the actual transmission begins. Network's resources are tied to the connection until it is torn down. Examples of label switching (or virtual circuit) include ATM, Frame Relay and MPLS [ATM, FRELAY, MPLS]. As a datagram arrives at an ingress interface of a network node, the egress interface is looked for in the routing table. A record is selected using the forwarding label found in each datagram. This label has only local significance (i.e. hop-by-hop) and it is replaced with a new label found from the record that was got from the routing table.

This model permits finer resource allocation to traffic streams, since label values are not globally significant but are only significant on a single link. This allows resources to be reserved for aggregates of streams (i.e. paths) each marked with a particular label, and the switching semantics govern the next-hop selection. The establishment and maintenance of these label switched paths require extensive management and configuration. [ARCH/DS]

3.4. *Integrated Services/RSVP*

Integrated Services or IntServ in short, uses service differentiation for each microflow individually. As the reservation of network's resources is done on per-microflow basis, the number of concurrent reservations increases heavily, especially on high-speed links. Reservation is done using RSVP and application support for it is required. Differentiated Services can be used as the transport for IntServ.

3.5. *Per-hop Classification*

A variant of the Integrated Services/RSVP model eliminates the requirement for hop-by-hop signalling by utilising only "static" classification and forwarding policies, which are implemented in each node along a network path. Configuration is updated administratively and thus the prevailing state does not respond to the instantaneous mixture of microflows active in the network.

Also DiffServ belongs to this category but being more or less dynamic compared to its static counterpart.

4. Overview of Differentiated Services

In DiffServ, classification of traffic is achieved by assigning forwarding behaviour to aggregates instead of microflows. The DS fields (see sec. 5.1) in the headers of IP-datagrams in a traffic flow are marked with a DS codepoint value either by the sending host or a traffic conditioner (usually in a boundary node) according to the behaviour wished for. The forwarding behaviours are applied hop-by-hop, based on DS codepoints (DSCPs) and on per-packet basis to aggregates of traffic streams in DS-capable network's nodes. Ingress traffic is classified and conditioned at boundary nodes in order to make it comply with the service agreement made between a customer and a service provider. [ARCH/DS]

Taking DiffServ into use does not necessarily require changes in applications. Separate traffic streams from a single host are either identified by a node in the network (how this is done is outside the scope of DiffServ) or are marked with a DSCP by a traffic conditioner that specifies a service level the host is justified for. In addition to the well-known best-effort (default) service, a variety of services can be applied, a leased-line emulation for example. Different types of service will not be standardised within DiffServ, but rather the implementation behind them. Applications or operating systems may also request a forwarding behaviour for packets by marking the DS field. The security implications caused by this are discussed in sec. 8.1.

The cornerstones of composing services in DiffServ are the **Per-Hop-Behaviours** (PHBs). PHBs present externally observable forwarding behaviours applied at DS-compliant nodes to DS behaviour aggregates, i.e. to packets with the same DSCP crossing a link in a particular direction. DS-capable domains operate with a common set of PHB definitions. The PHBs intended for wide usage are to be standardised and few have already been proposed [e.g. AF/DS, EF/DS]. Fig. 1 illustrates how microflows are aggregated into PHBs and how the PHBs can be handled in routers.

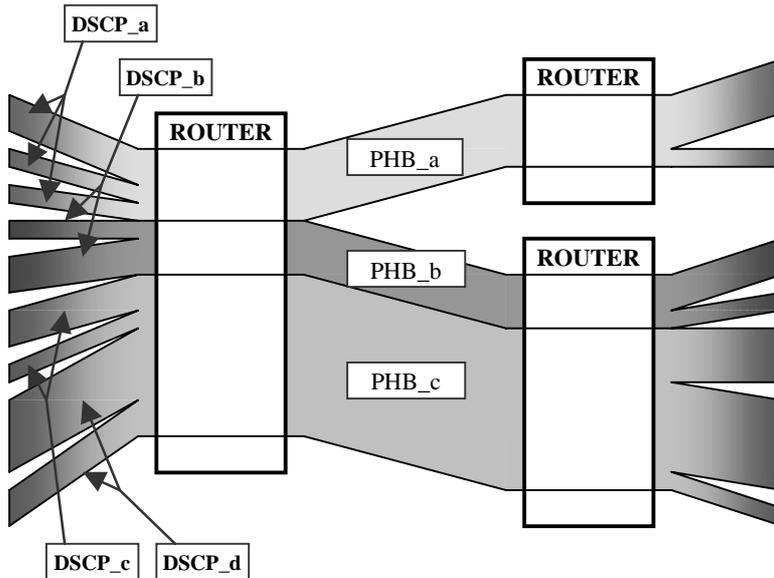


Fig. 1. Mapping of DSCP-marked streams to PHBs and aggregate routing.

A customer is entitled to services for which he has made an agreement with a service provider. Formed **Service Level Agreements (SLAs)** specify the forwarding service a customer's traffic should receive. A customer in this context refers to e.g. an organization or a DS-capable upstream domain. A sub-set of SLA is **Traffic Conditioning Agreement (TCA)**, which specifies in detail how customer's traffic is policed in traffic conditioners to comply with the SLA. TCA contains traffic classifier rules and any corresponding traffic profiles and metering, marking, discarding and/or shaping rules, which are applied to traffic streams selected by a traffic classifier.

In short, DiffServ is a scalable architecture in which per-microflow or per-customer state information is not carried along IP-datagrams' paths.

4.1. Concepts in DiffServ

Traffic conditioning in DS domains is carried out in boundary nodes, i.e. nodes that connect two domains. By conditioning only in boundary nodes, ingress and egress traffic is shaped to comply with free resources in the target network and scalability is obtained. By **traffic conditioner** is meant an entity which may contain meters, markers, droppers and shapers. These devices meter the ingress traffic and based on the result possibly re-mark packets with a new DSCP, drop excessive packets or delay packets in order to bring it into compliance with a traffic profile.

Forwarding behaviours (i.e. services) in DiffServ are determined by the codepoint value in the DS field in the IPv4 or IPv6 headers. The DS field in the IPv4 header is the TOS octet and in the IPv6 header the Traffic Class octet. Each DSCP maps to at least one PHB. PHBs are defined to permit a reasonably granular means of allocating buffer and bandwidth resources at each node among competing traffic streams. The PHBs don't solely define services but they are the building blocks.

A domain which implements DiffServ as defined in [ARCH/DS] is called DS-capable. A DS-capable domain has DS ingress and DS egress nodes. The former ones handle traffic entering the domain, the latter ones handle traffic leaving the domain. In practice, the division is only logical. DS ingress and egress nodes are boundary nodes that connect the domain to another DS domain or to a domain that is not DS-capable.

4.2. DiffServ –network

In the following is discussed how functions a network that supports DiffServ.

Fig. 2 shows three domains, which are interconnected with a backbone network. Two of the domains are DS-capable, i.e. the interior and boundary nodes support forwarding based on DSCP values. Introducing a non-DS-compliant node in a DS-capable domain may result in service degradation and the required service characteristics (e.g. low delay) may become unreachable [ARCH/DS]. Non-DS-capable domains do not employ differentiated services, but rather the best-effort service and possibly other service provisioning approaches.

Nodes in a DS-domain, e.g. domain 1 in the Fig. 2, are supposed to employ uniform sets of PHBs on which the services can be built. It is to be noted that the PHBs need not be the same among separate domains. Instead, mappings of PHBs are agreed between parties so that traffic from a domain that has certain PHB characteristics is remarked to an equivalent PHB in another DS-domain. Bilateral human agreement is needed at least in the early employment of DiffServ.

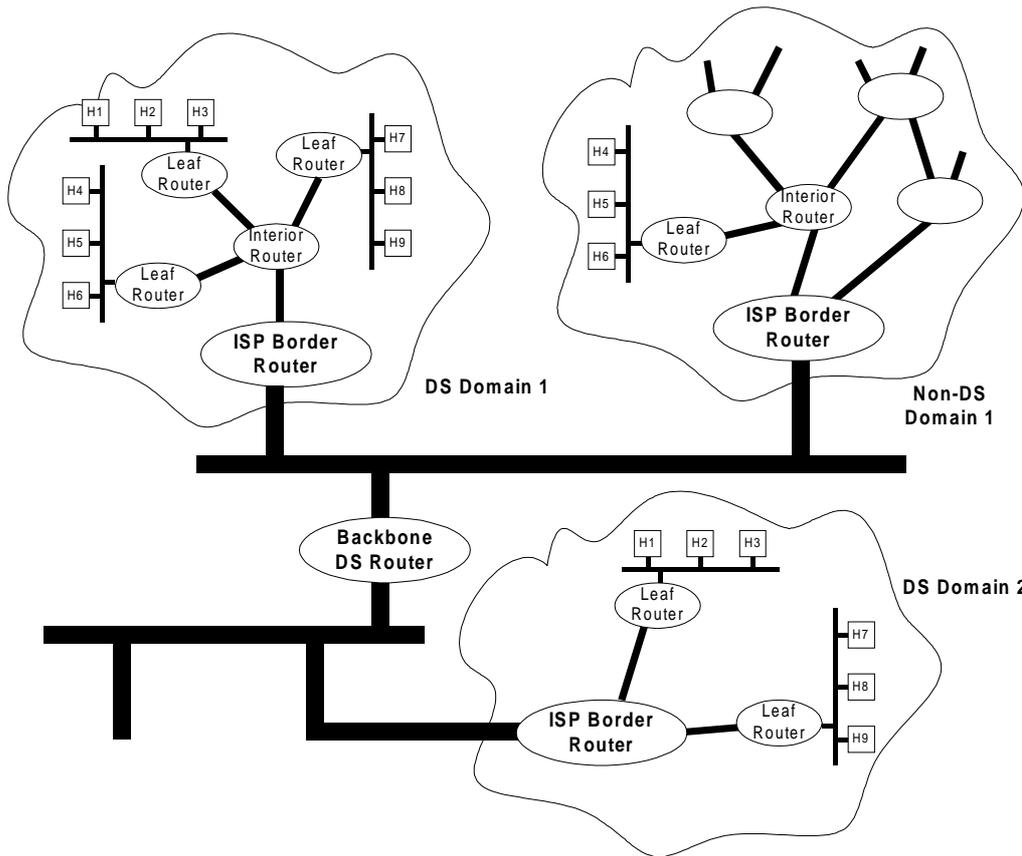


Fig. 2. A non-DS-capable and two DS-capable domains connected to backbone network.

4.2.1. Traffic Example

As some host, say H5 in DS-domain 1, in the Fig. 2 above starts generating traffic, there are certain implications in the functioning of the network that are different compared to if the network was non-DS-capable.

First, the near-by leaf router receives packets from H5. The router checks the DSCP value from the headers of the received packets and reads a PHB corresponding to the value. The router verifies that the host is entitled to the PHB and that the traffic is within the limits of the PHB. The router may re-mark the DSCP if it has been configured to do so for all traffic from H5 or it may re-mark packets in order to force the traffic stream to a justified level for H5.

The destination of the traffic from H5 is, say, host H2 in DS-domain 2. Based on its routing table, the leaf router forwards the traffic to domain's sole boundary router. The boundary router's traffic conditioner meters the traffic and possibly shapes it into

compliance with the agreement it has made with a boundary router in the backbone network. The presented backbone DS-router always performs policing to incoming traffic, as it can't trust upstream routers to have already done so. When packets are pre-marked and conditioned in the upstream domain, potentially fewer classification and traffic conditioning rules need to be supported in the downstream. [ARCH/DS]

The backbone router forwards the traffic to the boundary router of DS-domains 2, which in turn polices the traffic to comply with the service agreement with the backbone network. Next, the traffic is forwarded to the leaf router close to H2 and finally to H2. How successfully requested QoS is achieved depends greatly on how well the mappings of PHBs are performed between inter-domain boundaries and how similar the traffic conditioning rules are. When functioning correctly, the usage of DiffServ should result in sought-after provisioning of network's resources for each of the PHBs.

5. Service Realisations

Services are based on demands from the customers' side, they are feasible from the business point of view and they are technically realisable. When speaking of the future of data communication, it is essential that new services can be easily presented and that the already implemented ones are scalable. Differentiated Services tries to make all this possible. The DiffServ architecture can be thought of to be a framework within which service providers can offer their customers a range of network services that are differentiated in terms of performance and pricing.

Service realisations in DiffServ can be thought of to consist of three levels. On the first level are the DSCP values, on the second the PHBs specified by the DSCP values and on the third level are profiles. Customer and provider typically negotiate a profile at each service level for describing the rate at which traffic can be submitted (policing profile). Packets submitted in excess of this profile may not be allotted the service level requested [FRM/DS]. The value to the customer comes from the services that are obtained from the usage of these three levels.

PHBs are merely building blocks for services. Service providers combine PHB implementations with traffic conditioners, provisioning strategies and billing models, which enable them to offer services to their customers. Providers and customers negotiate agreements with respect to the services to be provided at each customer/provider boundary. These take the form of Service Level Agreements (SLAs).

End-to-end services can be constructed by the concatenation of domain services and their associated customer-provider SLAs for each of the domains, which the service traffic has to cross. However, difficulties easily arise as the providers can freely decide what are the services available in their networks, which in turn results in situations where the possible service level in some domain doesn't match with the service level in the originating host's domain. This will be a major issue in the development of DiffServ in the future.

5.1. Usage of IP-header

DiffServ uses a 6-bit header field, the DS-field. In IPv4 the TOS octet [RFC791] and in IPv6 the Traffic Class octet [IPv6] acts as the DS-field. Even though the usage of TOS and Traffic Class octets are different from the usage of DS-field, the differences are such that deployment of DiffServ doesn't cause serious impacts on existing applications.

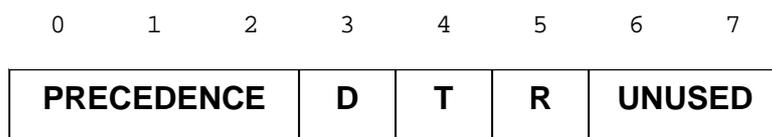


Fig. 3. Type Of Service field in IPv4.

Fig. 3 shows the coding of the TOS-field in IPv4 [RFC791]. The coding of the Traffic Class octet in IPv6 is not specified but it is intended to be used in a similar manner as the TOS-field in IPv4 [RFC2460]. The precedence bits 0-2 are used for marking the importance of an IPv4-datagram over another. Their usage and queuing requirements in IPv4 routers are specified in [RFC1812]. DiffServ supports with limitation the widely used values, or PHBs, for backward compatibility. The 'D', 'T' and 'R' bits have been specified to be marked in IPv4-datagrams that need special handling considering Delay, Throughput or Reliability. Backward compatibility with these bits will not be maintained in DiffServ. A node that interprets the precedence bits as they were originally supposed to is called a legacy node. The value of the unused bits is ignored when employing DiffServ.

The bits 0-5 in the TOS-field make up the DS-codepoint. Its value is in turn mapped to one or more PHBs, as configured by a network administrator.

5.2. Traffic Conditioning

What **traffic conditioning** basically means, is treating traffic differentially so that traffic entering a DS-domain conforms to the rules specified in a TCA, in accordance with the domain's service provisioning policy. In DiffServ conditioning happens mostly in boundary nodes (routers). Conditioning can also be employed in interior nodes but it is not required for a network to be DS-capable. When conditioning in an

interior node is required, only a subset of functioning of a boundary node may be needed. Naturally interior nodes need to support forwarding based on the DSCP-values.

Traffic conditioner [ARCH/DS] is a part of a network node that takes the node's ingress packets as its input and places the packets in its output in such an order that best satisfies the forwarding (i.e. service) requirements set for the packets and uses the network's resources in a best possible way. From the service point of view, traffic conditioners are the only elements needed for services to function. How they are maintained is merely a question of administration and network management. A traffic conditioner and its functional components are shown in Fig. 4.

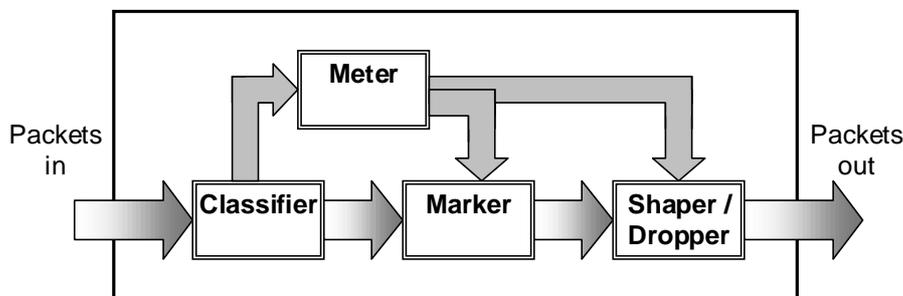


Fig. 4. A traffic conditioner and its functional components.

Packet *classifiers* select packets from a traffic stream based on the content of some portion of the packet header. The **BA classifier** (Behaviour Aggregate) classifies packets based on the DS codepoint only while the **MF classifier** (Multi-Field) selects packets based on the values on one or more header fields, such as source and destination addresses and port numbers, DS field and protocol ID. Classification is done for separating traffic aggregates from each other so that the amount of resources allotted to the aggregate is not exceeded. More on classifiers in Sec. 5.2.2.

Packet *meters* measure the temporal properties of the stream of packets selected by a classifier against a traffic profile specified in a TCA. A meter passes state information to a marker/shaper, which shapes the traffic if it is out of profile.

Packet *markers* receive the incoming packets from a classifier and the state information from a packet meter. If some portion of incoming packets is out of profile, a marker

can re-mark those packets by another codepoint, according to the state of a meter. A marker can also be configured to mark all incoming packets to a single codepoint.

Shapers delay some or all of the packets in a traffic stream in order to bring the stream into compliance with a traffic profile. A shaper usually has a finite-size buffer, and packets may be discarded if there is not sufficient buffer space to hold the delayed packets. *Droppers* police the incoming stream by dropping packets that are out of profile.

5.2.1. Location of Traffic Conditioners

Traffic is conditioned on either end of a boundary link. Boundary nodes may refer to the formed SLA if it is their responsibility to condition the ingress/egress traffic to conformance with the appropriate TCA or is it the responsibility of the node in the other end of a boundary link. However, ingress traffic can't be assumed to conform to the TCA but policing must be prepared to enforce the TCA.

In addition to conditioning in boundary nodes, it is recommended that traffic is conditioned as close to the sending host as possible. This way potentially fewer classification and traffic conditioning rules need to be supported in the downstream DS domain and the requested service level is more likely to be met. Therefore a limited number of conditioning rules (i.e. some, but not all of the TC components in Fig. 4) may be needed applied within the originating host's domain.

5.2.2. SLA and TCA Configurations

Traffic conditioning is applied at the ingress interfaces of boundary routers for traffic that either leaves or enters the domain. In cases where DSCP value based classification is adequate for the incoming packets (i.e. coming to the node), only BA-classifiers are needed. Instead, MF-classification is needed e.g. when per-customer based differentiation is required or if some service can only be offered between ingress boundary router and a specific egress point from the DiffServ network. Two example configurations are presented in Fig. 5. Conditioned packets are routed to the appropriate egress interface where they are applied to a PHB. [BROUT]

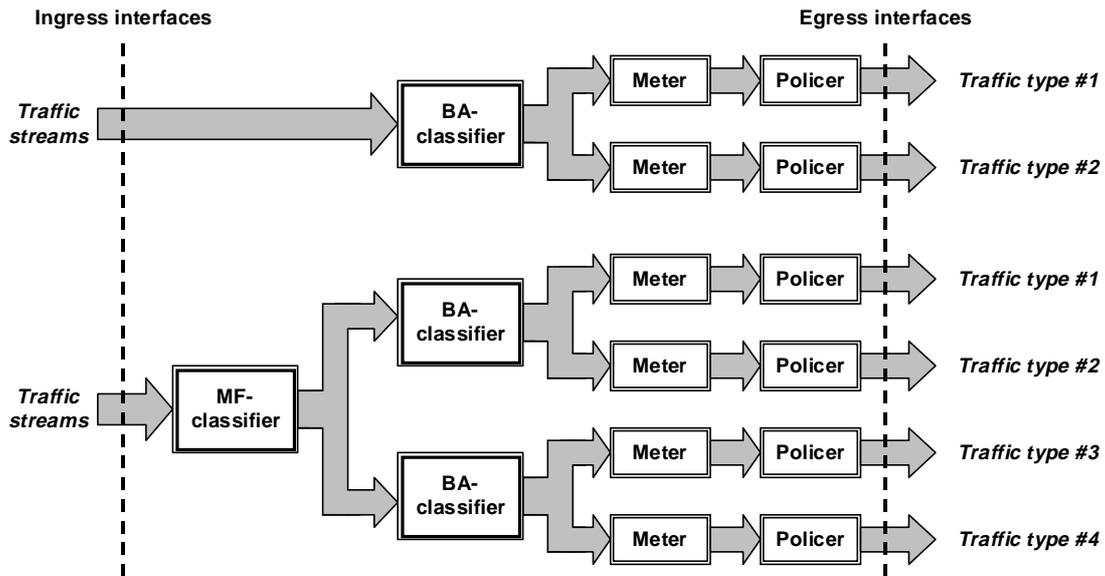


Fig. 5. Two “TCA-enabling” configuration of which one based on a BA-classifier and another on MF- and BA-classifiers

Incoming packets are subject to the TCA, which is a part of the SLA that was formed statically or dynamically between the originator of the traffic and the owner of the boundary node. To say it in other words, the TCA is a per-customer entity. The TCA has two sub-components, a constraint TCA and a fine-grain TCA [BROUT]. The constraint TCA is essential as it serves to protect the provider’s resources at each DiffServ service level. The fine-grain TCA defines per-flow value-added functionality that the provider may offer to the customer. The latter is unlikely to be used at boundaries between providers where enforcement of aggregates is the primary concern. A TCA implementation in a network node may either have both TCA’s sub-components or only the constraint TCA. In case of both, the fine-grain TCA is applied first.

Where Fig. 5 shows what are the functional components needed for implementing TCAs, Fig. 6 shows the logical idea behind them. In Fig. 6, two service levels are shown that are applied to packets with a specified DSCP on a given transfer rate. Constraint and fine-grain TCAs that are shown for the “Better Best Effort” service, are merely two sets of independent rules that the network administrator has configured as wished for. The sets have the presented five (four if AB- or MF-classifiers not used) categories, of which ‘BA Filter’ and ‘PHB’ specify the DSCP and the mapping of

corresponding packets to the given PHB. The ‘MF Filter’ specifies rules by which traffic is separated to different BA-classifiers, as shown in Fig. 5. ‘Profile’ specifies the configuration of meters that are used to determine the conformance of traffic submitted for each service instance. Non-conforming traffic is targeted to treatment specified in ‘Disposition’, which can be e.g. remarking, discarding or shaping into profile.

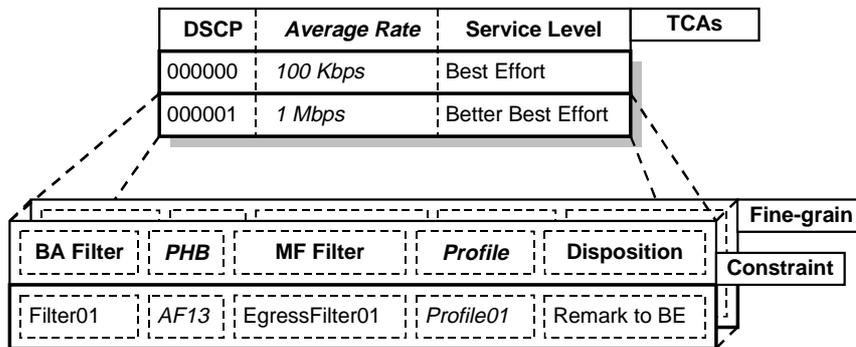


Fig. 6. An example of a TCA and its sub-components, fine-grain and constraint TCAs.

The TCA and the rules therein constitute an essential part of the SLA. Another fundamental part to the service offering is the pricing and billing mechanism. In addition to these, several general service characteristics are specified in SLA, such as authentication mechanisms, support capabilities in case of failure and encryption services. The negotiation of the SLA is either static or dynamic, depending on how much human interaction is involved. Dynamic negotiation presents challenging problems in shape of requirements for resource provisioning mechanisms, customer equipment compatibility and users’ reactions to dynamically changing SLA. Static negotiation is the current norm.

5.2.3. Network Provisioning

The negotiation of service level agreement doesn’t solely guarantee correct allocation of network’s resources to different classes of services. In addition to the negotiations, network’s resources have to be provisioned so that the results of the agreements can be met. By **network provisioning** is understood the determination and allocation of the resources, both physical and logical, needed at various points in a network. A network is physically provisioned when physical resources are added or removed at some points in the network. Logical provisioning comprehends the modification of operating parameters within existing physical network equipment. [FRM/DS]

Provisioning is not only required on network's boundary but in interior also. A differentiated service provider can't sell a contract (SLA) on the boundary of his network if the standards can't be met in the interior. On boundary the provider's minimum responsibility is to offer sufficient physical resources. The standards of the negotiated TCAs can be met by logically provisioning the physical resources at each service level according to agreed aggregated profiles. Configuring internal provisioning is trickier as understanding is required of the composition and volume of traffic at different parts of network. At that point qualitative and quantitative traffic need to have different priorities.

Quantitative traffic agreement contains concrete assurances for traffic. A quantitative service could be e.g. "90% of in profile traffic delivered at service level C will experience no more than 50 ms latency". Instead, qualitative agreement would in that case be e.g. "Most traffic offered at service level C will be delivered with low latency". The volumes of quantitative services can be fairly well approximated whereas qualitative traffic can't. Therefore making quantitative commitments in the SLA for qualitative services should be avoided and these services should have lower priority. In practice this means provisioning most of the network's resources to quantitative services, while leaving sufficient capacity remaining to accommodate some amount of qualitative traffic.

5.3. Per-Hop-Behaviours

A Per-Hop-Behaviour (PHB) is a description of the externally observable forwarding behaviour of a DS node applied to a particular DS behaviour aggregate. PHBs provide the means by which a node allocates resources to behaviour aggregates. Similar requirements for packet loss, delay and jitter, for example, are the factors that make up a behaviour aggregate to which uniform forwarding behaviour is applied. [ARCH/DS]

A provider of differentiated services decides what are the available services in his network. The services probably have some common characteristics so that they can be divided into few manageable groups. The provider commits himself to providing physical resources so that his customers have the services they pay for. Based on these decisions, the provider selects the Per-Hop-Behaviours that are required for

implementing the services as well as negotiates the Service Level Agreements with his customers. Finally the PHBs and SLAs are configured into the network's nodes.

PHBs are gathered into PHB groups for manageability and consistency. PHBs are grouped based on their similar properties such as bandwidth requirements or behavioural characteristics mentioned above. A single PHB is a special case of a PHB group. The relationship between PHBs in a group may be in terms of absolute or relative priority. PHBs are implemented in nodes by means of some buffer management and packet scheduling mechanisms that are configured into traffic conditioners.

PHBs that are intended for wide usage are standardised within IETF. Standardised PHBs have a recommended codepoint value, which is set to packets that are wished to receive treatment specified within the PHB. Multiple codepoint values may be mapped to a single PHB. Every codepoint in use must be mapped to some PHB as the treatment for packets without specified local policy is either mapping to the default PHB (best-effort service) or discarding. In addition to standardised ones, there may exist only locally defined PHBs in a network. These PHBs may be e.g. experimental or they may use a local service, and traffic mapped to them is kept within the originating network. As the available space in the TOS-field (see Fig. 3) is limited and room is left for its usage in the future, the codepoint mappings can freely (an exception to this specified in sec. 5.3.2) be made by network administrators within independent networks. However, this may require re-marking in the network boundary.

5.3.1. Network Resource Usage

DS-capable network's resources are provisioned to services by allocating suitable share of resources to each PHB group. Proper allocation necessitates knowledge of impacts of one group's requirements to the others. Resource allocation for individual PHBs within a group can be based on e.g. prioritising one over another. In PHB specifications themselves behavioural characteristics are given instead of implementation guidelines. This leaves room for different implementation mechanisms for a particular PHB group.

Traffic conditioners control the usage of resources based on the administratively configured PHB groups and through enforcement of negotiated TCAs, possibly in interaction with domain's other nodes and TCs. Special protocols, a control entity and administrative actions may be needed for interaction with TCs. This is outside the scope of DiffServ.

5.3.2. Proposed PHBs

The simplest example of a PHB is the one which guarantees a minimal bandwidth allocation of X% of a link to a behaviour aggregate over some reasonable time interval. Although realisable, a PHB is ought to provide wider ground for services that use it. There are currently few proposed PHBs [AF/DS, EF/DS, HDR/DS], which are briefly presented in the following. These PHBs provide the basis for service examples presented in Sec. 5.4. As explained above, none of the PHBs are mandatory for a node to be considered DS-compliant but when implemented the specifications must be met.

5.3.2.1. *Default and Class Selector PHBs*

DiffServ can't be taken into use if it doesn't provide backward compatibility. The Default PHB exists for that. Its purpose is to provide the best-effort behaviour that the current routers perform. Default PHB is the one that is used for packets for which no other agreement exists. The Default PHB (i.e. best-effort service) gets the lowest priority compared to all other PHBs. Therefore any traffic that doesn't conform to its profile can easily either be remarked to the Default PHB or be discarded.

Traffic that is subjected to the Default PHB can be described to achieve the following kind of service: *the network will deliver as many of these packets as possible and as soon as possible, depending on the prevailing network load and state.* A reasonable implementation of this PHB would be a queuing discipline that sends packets of this aggregate whenever the output link is not required to satisfy another PHB. However, to ensure at least some bandwidth for hosts, which don't employ DiffServ, some resources may need to be reserved for Default behaviour aggregates. [HDR/DS]

The recommended codepoint for the Default PHB is the bit pattern '000000'; the value '000000' must map to a PHB that meets the specifications of Default PHB. The

codepoint chosen for Default behaviour is compatible with existing practice [RFC791]. When a codepoint is not mapped to a standardised or local use PHB, it should be mapped to the Default PHB.

The mere Default PHB is not enough to provide sufficient backward compatibility. The precedence bits (see Fig. 3) of IPv4 TOS-field are widely used [RFC1122] in existing networks' equipment and must therefore be supported by DiffServ. The greater the value of the precedence bits is, the higher is the priority of the packet. This same method is applied in the Class Selector PHB so that the bit patterns 'xxx000', eight in all, are reserved as a set of Class Selector Codepoints (see Fig. 7). Compatibility for 'D', 'T' and 'R' bits is not provided. The Class Selector PHB Requirements on codepoint '000000' are compatible with those listed for the Default PHB above. Forwarding of each of the eight priority classes is done separately.

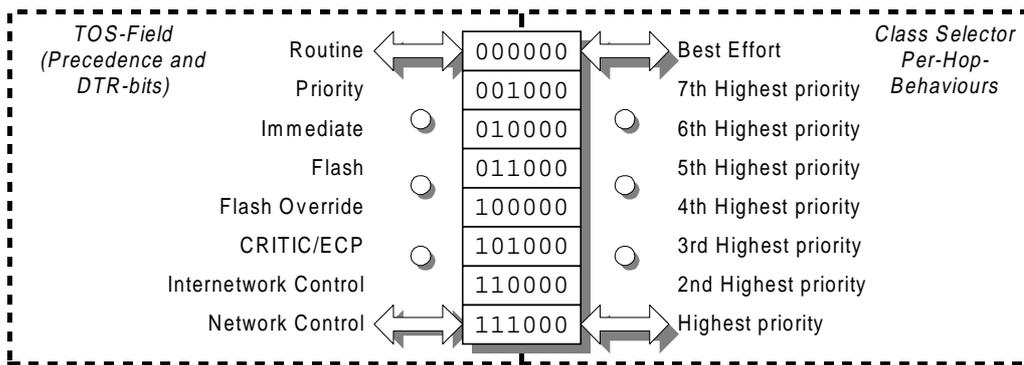


Fig. 7. Usage of DS-field in Class Selector PHB.

5.3.2.2. Assured Forwarding PHB Group

The motivation behind the AF PHB is the need for fixed bandwidth lines that especially companies use extensively. In a typical application, a company uses the Internet to interconnect its geographically distributed sites and wants an assurance that IP packets within this intranet are forwarded with high probability as long as the aggregate traffic from each site does not exceed the subscribed information rate (profile). Packets that are out of profile are forwarded with lower probability. Irrespective of whether packets belonging to a same microflow are in or out of profile, it is important they are not reordered.

AF PHB group provides four classes of different levels of forwarding assurances and resources (buffer space and bandwidth) for IP packets received from a customer DS domain. Packets from each of the classes are marked with one of three drop precedence values. The customer or the originating DS domain does the marking in accordance with the customer’s subscription. In a DS node, the level of forwarding assurance of an IP packet thus depends on: [AF/DS]

- How much forwarding resources has been allocated to the AF class that the packet belongs to? Packets in one class are forwarded independently from the others and within the service rate (bandwidth) that has been configured for the class.
- What is the current load of the AF class?
- In case of congestion within the class, what is the drop precedence of the packet? Packets with a higher drop precedence value are more preferably discarded.

Table 1. *The four AF classes and the related three drop precedence values.*

	Class 1	Class 2	Class 3	Class 4
Low Drop Precedence (1)	001010	010010	011010	100010
Medium Drop Precedence (2)	001100	010100	011100	100100
High Drop Precedence (3)	001110	010110	011110	100110

AF service classes presented in Table 1 are referred to as AF_nm, where ‘n’ marks the number of the class (1-4 currently) and ‘m’ the precedence value (1-3 currently). Implementation of the AF PHB group sets certain requirements for packet queuing in network’s nodes. More on this in Sec. 7.1.

5.3.2.3. Expedited Forwarding PHB

One major group of services is the one requiring assured bandwidth with low loss, low latency and low jitter in an end-to-end connection. Such services are e.g. a point-to-point connection or a virtual leased line, carrying time-sensitive data, speech, video or some combination of them. EF PHB offers this type of service through DS domains. [EF/DS]

Loss, latency and jitter are all due to the queues traffic experiences while transiting the network. By configuring nodes to meet the specifications of EF PHB, traffic aggregates see no (or very small) queues and therefore have well-defined minimum departure rate.

In other words, aggregates are independent of the intensity of other traffic at the node. Queues arise when (short-term) traffic arrival rate exceeds departure rate at some node. It is the network boundary traffic conditioners' task to bind the rates for traffic aggregates such that, at every transit node, the aggregate's maximum arrival rate is less than that aggregate's minimum departure rate.

The departure rate of an aggregate's packets from any DiffServ node must equal or exceed a rate that has been specified for the service the packets belong to. It is also required in the EF PHB specification that a network administrator must be able to configure the rate into DiffServ nodes. Codepoint '101110' has been recommended for the EF PHB. Sec. 7.1 provides some information on queue management in EF.

5.4. Service Examples

In the following two example services are presented one of which is built on the AF PHB group and the other on the EF PHB. It must be born in mind that, in general, DiffServ services are all for unidirectional traffic only and they are for traffic aggregates, not individual microflows. Another important aspect is the scope of a service, which refers to the topological extent over which the service is offered. For example, a provider offers a service to one of its customers and the traffic from the customer enters the provider's network at ingress interface A. The service may then be applied to all traffic in one of the following ways:

- a) from ingress interface A to any egress interface.
- b) from ingress interface A to a set of egress interfaces.
- c) between the interface A and some egress interface B.

There are some common things for both the examples. First of all, policers need to be configured at traffic ingress points. Secondly, the used PHBs need to be implemented at core network equipment.

5.4.1. Better Than Best-Effort

Better than best-effort (BBE) traffic has a higher priority over the competing best-effort traffic and thus provides reduced latency. Quantitative performance measures can't be

given for BBE as it is clearly a qualitative service (see Sec. 5.2.3) and depends on how resources are provisioned. The scope of the service applies from a given ingress interface to any egress interface. [FRM/DS]

Table 2. TCA specification for BBE service.

Codepoint	Service rate	Applicable egress IF	Handling of excess traffic
AF13 Mark	1 Mbps	Any	Re-marked with AF11 mark

BBE can be constructed as specified in Table 2. A provider offers the service with 1 Mbps aggregate rate. Traffic within the 1 Mbps limit is directed to the AF13 PHB and excess traffic is re-marked to AF11 PHB. One of the prerequisites for AF, preserving the original order of packets, is met when only one queue is used for implementing the both AF11 and AF13 PHBs. The provisioning of the PHBs and how prioritising AF13 over AF11 is done is up to the provider.

5.4.2. Leased Line

This is a quantitative service, which emulates traditional leased line service. It promises to deliver traffic with very low latency and very low drop probability, up to a negotiated rate. Above this rate, traffic is dropped. Corporate VPN's and IP telephony are two likely applications to use this service. [FRM/DS]

Table 3. TCA specifications for Leased Line service.

Codepoint	Service rate	Applicable egress IF	Handling of excess traffic
EF-Mark	500 Kbps	Egress point B	Discard
EF-Mark	250 Kbps	Egress point C	Discard

This example considers a customer with three geographically dispersed networks interconnected via a single provider network. Customer attachment points are represented as A, B and C. Table 3 shows the TCAs for attachment point A, which are included in a single SLA. Customer has two leased lines established for interconnecting point A to point B, and respectively point A to point C. EF PHB is used for both of them with service rates 500 Kbps and 250 Kbps. Excess traffic will be discarded.

The provider needs to configure the policers at ingress point A for both the egress points B and C. The policers are of MF-type, as classification of packets is based on the codepoint value and the destination. The routers along the ways from A to B and A to C need to be provisioned to carry up to 750 Kbps of traffic in case both of the leased lines cross the same router.

6. Traffic Provisioning between Providers and Customers

6.1. Customers' Responsibilities

It is in every customer's preference that his traffic receives just the kind of treatment that best satisfies the traffic's needs. The customer wouldn't want to pay for or he wouldn't need better service, worse service would in turn be unsatisfactory. Therefore it is within the customer's interest that interdomain traffic is shaped and only those who are entitled to some service, receive it. Interior routers encounter the heaviest load while boundary routers are less loaded. However, boundary routers are the ones who finally enable the services.

A customer domain's responsibility is to enforce the SLAs that has been formed between peering domains. Inability to do so results in rejection of traffic in boundary routers within peering domains and consequently in service level degradation in connections which cross the customer domain.

6.2. Interoperability

The term interoperability refers to the ability of two different networks to work together. Interoperability describes how the networks e.g. can communicate and share data with each other, regardless of whether the networks use the same network architecture. With DiffServ, two distinct interoperability issues arise. First, the case with two DS-capable domains with at least to some extent differing network provisioning. Secondly, the case with two networks one of which is partly or fully non-DS-capable.

The first case has been dealt with earlier. The major issue is how well service level agreements between domains have been negotiated. Also, what are the provisioned shares of available resources that have been committed to different behaviour aggregates. Interoperability is more or less static in the early days of DiffServ, but as

time passes and development takes place the amount of human interaction decreases and operations become dynamic.

In the second case is assumed that the non-DS-capable domain deploys no traffic conditioning functions on domain boundary nodes. It is therefore difficult to keep up with the service level requirements, even if DS-compliant nodes existed in the domain's interior. To overcome this restriction there may exist an agreement between the two domains which describes how the egress traffic from the DS-capable domain should be marked before entry into the non-DS-capable domain. Alternatively, if the non-DS-capable domain consists of legacy nodes and that is known by the other domain, then the egress traffic may be re-marked with the Class Selector PHB codepoints. Where there is no knowledge of the traffic management capabilities of the downstream domain, and no agreement in place, a DS domain egress node may choose to re-mark DS codepoints to zero, under the assumption that the non-DS-capable domain will treat the traffic uniformly with best-effort service. [ARCH/DS]

7. Requirements for Network Elements

7.1. Queue Management

Utilising some packet queuing mechanisms in network's nodes does the actual implementation of PHBs. There are several mechanisms each of which with different characteristics. What they all have in common is that the incoming packets are written into a queue, then the packets are read from the queue in some order and placed in the egress interfaces, possibly discarding excess packets. The mechanisms that are applicable for implementing PHBs include for instance: [CISCO_Queue]

- *Priority queuing (PQ)*. Network managers define how they wish traffic to be prioritised in the network. By defining a series of filters based on packet characteristics, traffic is placed into a number of queues; the queue with the highest priority is serviced first, then the second highest and so on. If the highest PQ is always full, then this queue will continually be serviced and packets from the other queues will queue up and be dropped.
- *Weighted Fair Queuing (WFQ)*. Traffic is classified into conversations and priority (or weight) is applied to identified traffic to determine how much bandwidth each conversation is allowed relative to other conversations. Conversations are broken into two categories: those requiring large or small amounts of bandwidth. The goal is to always have bandwidth available for the small bandwidth conversations and allow the large bandwidth conversations to split the rest proportionally to their weights. Packets in the queue are reordered so that low-volume conversations are moved forward and high-volume conversations toward the tail of the queue.
- *Class-Based Queuing (CBQ), a.k.a. Custom Queuing (CQ)*. Uses the same classification facility as PQ. The difference is that from each class (or queue) only certain maximum amount of packets can be read at a time. Thus the blockages due to heavy load of higher priority traffic are not experienced in CPQ.
- *Random Early Drop (RED)*. RED is a high-speed congestion avoidance mechanism rather than strictly a congestion management mechanism, such as PQ, CBQ or WFQ. RED aims to control the average queue size by indicating to the end hosts

when they should temporarily slow down transmission of packets. RED does this with TCP's congestion control mechanism by randomly dropping packets prior to periods of high congestion, this way telling the packet source to decrease its transmission rate.

The standardised PHB groups may require that certain queuing mechanism must be used for the implementation. It is therefore important that different mechanisms are available and selectable by a network administrator in a particular vendor's equipment.

For the implementation of AF PHB group minimisation of long-term congestion within each class is required, while short-term congestion resulting from bursts is allowed. The utilised packet-dropping algorithm must treat all packets within a class equally, thus allowing consistent end-to-end service semantics. An active queue management algorithm, such as RED, is therefore required for the AF PHB group. [AF/DS]

Several types of queue scheduling mechanisms may be employed to implement the EF PHB. A simple PQ is adequate as long as there is no higher prioritised queue that could delay EF packets for more than a packet time. Another possible implementation is a CBQ-scheduler that gives the EF queue priority up to the configured rate. [EF/DS]

7.2. Boundary Routers

Boundary routers are the essential part of DiffServ. They are in charge of ingress and egress traffic's compliance to agreements. A DS boundary router provides the traffic conditioning section prior to routing core. After the routing core lies the PHB section that enforces the PHB configuration. The combination of traffic conditioning at ingress interfaces and PHB treatment at egress interfaces results in a DiffServ service. [BROUT]

Boundary routers are likely to provide a monitoring interface that enables collection of statistics regarding traffic carried at various DiffServ service levels. These statistics are important for accounting purposes and for tracking compliance to service level agreements (SLAs) negotiated with customers.

Other parts of a boundary router are the SLA and PHB configuration tables that are configured through a DiffServ provisioning interface. The provisioning interface can be arranged via one of a number of management protocols, such as SNMP. A boundary router may also contain optional RSVP capabilities.

7.3. Interoperability of IntServ/RSVP and DiffServ

A network utilising RSVP resource reservation protocol can be thought to be a customer of a network that utilises DiffServ. As RSVP is currently rather lightly employed and it is expected to remain as such, the RSVP-employing networks are considered stub networks that reserve bandwidth between each other. DiffServ mechanisms are used within larger transit networks in that model. The issues raised by this are how RSVP bandwidth requests could be understood by DiffServ and vice versa, and how requested service level could be translated into a DiffServ service [RSVP/DS]

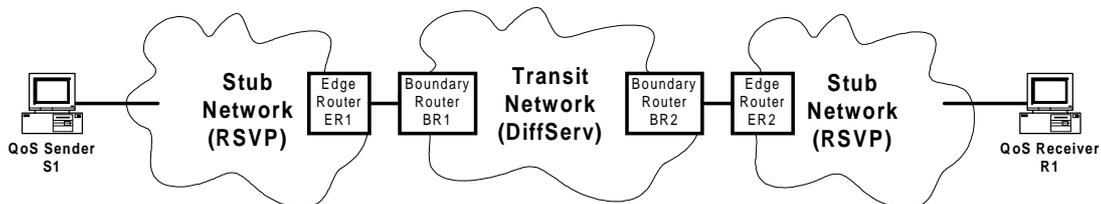


Fig. 8. Two RSVP-capable stub networks interconnected with a DS-capable network.

In the Fig. 8 above is shown the situation where RSVP- and DS-capable networks interact. The edge routers handle RSVP's requests and they provide an interface to the admission control for the DiffServ transit network. The presented DS boundary routers are not required to run RSVP. The operators of stub (customer) networks negotiate the available service levels with the operator of the transit network. If static provisioning is used, then the admission control functions at edge routers are configured accordingly. With dynamic provisioning the admission control is required to communicate with counterparts within the DiffServ transit network.

It is assumed that there are two different schemes how the mapping from IntServ service types to DiffServ service levels can be done. In the first "default mapping" scheme there is a well known mapping from IntServ service type to a PHB that will

invoke the appropriate behaviour in the DiffServ network. These mappings are not necessarily one-to-one and they are configured to edge routers. In an alternate "customer-specified mapping" scheme an edge router (ER1 in Fig. 8) determines the PHB that should be used to obtain the corresponding DiffServ service level. The edge router sets the PHB in an arriving RSVP reservation request (RESV) message and forwards the message to the sending host (S1 in Fig. 8). The sending host from then on marks outgoing packets with the indicated PHBs.

In Table 4 is presented the procedure how RSVP and DiffServ interact while reserving resources. One of the two mapping schemes above is applied at the step number 8.

Table 4. Resource reservation request in stages using RSVP and DiffServ.

Step	Src	Dest	Description
1	S1		An RSVP PATH message is generated describing the traffic offered by the sending application.
2	S1	R1	The PATH message is forwarded towards R1. Standard RSVP processing is applied at RSVP-compliant nodes within the sending stub network.
3	ER1	R1	According to standard RSVP processing, the PATH state is installed in the edge router. The PATH message is sent onward to the transit network
4	ER1	R1	The PATH message is carried transparently through the transit network. It is processed in the receiving stub network according to standard RSVP processing rules.
5	R1		At R1 an RSVP RESV message is generated, indicating interest in the offered traffic at a certain IntServ service level.
6	R1	S1	The RESV message is forwarded towards S1. It may be rejected at any RSVP node in the receiving stub network if resources are noticed insufficient to carry the traffic requested.
7	ER2	S1	At ER2 the RESV message is subjected to standard RSVP processing. It is carried transparently through the transit network if it is not rejected due to inadequate resources.
8	ER1		The admission control service running at ER1 compares the resources requested to the resources available at the corresponding DiffServ service level, in the DiffServ enabled transit network. If the RESV message is admitted, the available capacity for the service class is updated, by subtracting the approved resources from the available capacity.
9	ER1	S1	If the RESV message was not rejected, it is forwarded towards S1.
10	ER1	S1	The RESV message proceeds through the sending stub network. RSVP nodes in the sending stub network may reject it.
11	S1	R1	S1 begins to set the DS-field in the headers of transmitted packets to the value that maps to the IntServ service type specified in the RESV message.

8. Other Aspects

8.1. *Security and Tunnelling Considerations*

The major security issue caused by the introduction of DiffServ to a network is a consequence of the possibility for hosts to request for certain service level. The issues are the potential for denial-of-service attacks and the related potential for theft-of-service by unauthorised traffic, which are dealt with in the following.

Different services obtain different (i.e. better or worse) levels of QoS. Therefore it becomes tempting to modify the DS field to codepoints indicating behaviours used for enhanced services or by injecting packets with the DS-field set to such codepoints. Possibility for such modifications results in service degradation, i.e. denial-of-service, which depletes the resources available to forward traffic streams. This is the natural consequence, as finally all packets would request for the best service available. The defence against such theft- and denial-of-service attacks consists of the combination of traffic conditioning at DS boundary nodes along with security and integrity of the network infrastructure within a DS domain.

Conditioning must be done in each ingress node so that such service attacks don't happen. Monitoring incoming packets and checking that the originator is entitled to the requested service at the specified level does this. This means ensuring that traffic conforms to the applicable TCA(s) and the domain's service provisioning policy. In practice every node must ensure that all traffic originated from it carries acceptable DS codepoints. Traffic authentication may be required to validate the use of some DS codepoints (e.g., those corresponding to high-quality services). Such authentication may be performed by technical means (e.g. IPSec) or by non-technical means from knowledge of from which inbound link the packets came.

IPsec is a foundation for security protocols that is designed to provide interoperable and high quality security for IPv4 and IPv6 that is based on cryptography [RFC2401]. IPsec functions either in transport mode, which is a security association between two hosts, or in tunnel mode in which the security method is applied to an IP tunnel. In

tunnel there are two DS-fields, that of the protected encapsulated packet's header and that of the outer packet's header. The outer header's DS-field is not included in the cryptographic calculations and can thus be changed when the packet traverses the tunnel. However, the outer DS-field cannot be copied to the inner DS-field when the packet is decapsulated because the current IPsec requires that the inner header cannot be modified in a tunnel egress node. IPsec therefore provides protection against theft-of-service for the tunnel endpoints but not for the intermediate route itself.

8.2. Multicast in DiffServ

Multicast packets consume more resources than unicast packets as they may take multiple paths across a network due to packet replication. Each replicated packet heads towards a member of the multicast group. Group membership may be static or dynamic depending on whether one can participate the multicast tree on the fly.

Dynamic multicast group membership poses a problem for DiffServ, as it is difficult to predict in advance the amount of required network resources. Therefore it can be difficult to provide quantitative service guarantees to multicast senders. A not-so-elegant solution to this could be the reservation of codepoints and PHBs for exclusive use by multicast traffic only, thus separating unicast traffic. Another issue is that multicast traffic should not cause any SLA violations with downstream domains when packets traverse multiple routes towards separate egress nodes and domains. To overcome this separate peering SLA for multicast traffic probably need be established.
[ARCH/DS]

8.3. Effects on Network Management

There are several issues that require network management's control when DiffServ is employed. At least at the first stages network provisioning is done statically. This requires extensive network administration depending on the size of the network and how often the configuration changes. A designated Bandwidth Broker [BB] with a policy database may be used for configuring leaf routers within the local domain.

Service management is another big issue. Security, flexibility and reliability are the keywords in bandwidth sharing, customer billing and capacity planning, to name a few. Centralised service configuration and policy administration features are likely to evolve. Maintaining QoS requires thorough SLA configurations between peering domains.

9. Conclusions

Differentiated Services promises a lot but yet there are many questions unanswered. Can appropriate end-to-end QoS be achieved on connections spanning over multiple domains? How well PHBs correspond to each other on peering domains? Even though modifications to hosts are not required, will there be any and of what kinds? Will security aspects become a major issue if they yet are not? What will be the burden on network and service management like?

Despite of the question marks, DiffServ has certain benefits all of which other service differentiation approaches don't have. It is scalable, its implementation is not tightly tied and it does not require some specific types of expensive hardware. Also, ISPs don't need to use services that are predefined in some standardisation organisation. Instead they can construct services of their own.

All in all, Differentiated Services seems to have the potential to become the long awaited universal service differentiation approach to Internet. The steps towards that include thorough field-tests, convincing router manufacturers on the architecture, implementation of router software and hardware updates and finally, convincing the service providers on the possibilities. After that it is merely up to bill-payer's approval.

10. References

- [802.1] ANSI/IEEE, *Information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks – Common specifications - Part 3: Media Access Control (MAC) Bridges*, Standard 802.1D, 1998.
- [AF/DS] IETF, *Assured Forwarding PHB Group*, Internet draft, draft-ietf-diffserv-af-06.txt, Feb. 1999.
- [ARCH/DS] IETF, *An Architecture for Differentiated Services*, RFC 2475, Dec. 1998.
- [ATM] ATM Forum, *ATM Traffic Management Specification Version 4.0*, 1996.
- [BB] IETF, *A Two-bit Differentiated Services Architecture for the Internet*, Internet draft, draft-nichols-diff-svc-arch-01.txt, April 1999.
- [BROUT] IETF, *Requirements of DiffServ Boundary Routers*, Internet draft, draft-bernet-diffedge-01.txt, Nov. 1998.
- [CISCO_QoS] QoS documentation in the homepage of Cisco Systems, http://www.cisco.com/warp/public/cc/cisco/mkt/ios/qos/tech/qos_wp.htm.
- [CISCO_Queue] Queuing documentation in the homepage of Cisco Systems, http://www.cisco.com/warp/public/cc/cisco/mkt/iworks/protocol/dlsw/prodlit/toc_rg.htm.
- [Co95] Comer, D., *Internetworking with TCP/IP Volume I: Principles, Protocols and Architecture*, Prentice Hall, 1995.
- [EF/DS] IETF, *An Expedited Forwarding PHB*, Internet draft, draft-ietf-diffserv-phb-ef-02.txt, Feb. 1999.
- [FRELAY] ANSI, *DSSI Core Aspects of Frame Relay*, 1990.
- [FRM/DS] IETF, *A Framework for Differentiated Services*, Internet draft, draft-ietf-diffserv-framework-01.txt, Oct. 1998.

-
- [HDR/DS] IETF, *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*, RFC 2474, Dec. 1998.
- [MPLS] IETF, *Multiprotocol Label Switching Architecture*, Internet draft, draft-ietf-mpls-arch-05.txt, April 1999.
- [RFC1349] IETF, *Type of Service in the Internet Protocol Suite*, RFC 1349, July 1992.
- [RFC1812] IETF, *Requirements for IP Version 4 Routers*, RFC 1812, June 1995.
- [RFC2401] IETF, *Security Architecture for the Internet Protocol*, RFC 2401, Nov. 1998.
- [RFC2460] IETF, *Internet Protocol, Version 6 (IPv6) Specification*, RFC 2460, Dec. 1998.
- [RFC791] IETF, *Internet Protocol*, RFC 791, Sep. 1981.
- [RSVP/DS] IETF, *A Framework for Use of RSVP with DiffServ Networks*, Internet draft, draft-ietf-diffserv-rsvp-01.txt, Nov. 1998.
- [TR] ANSI/IEEE, *Information technology – Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 5: Token ring access method and physical layer specifications*, Standard 802.5, 1998.
- [WG/DS] DiffServ Mail Index, <http://www-nrg.ee.lbl.gov/diff-serv-arch/>.

A.1 Definitions

Full notion	Explanation
Classifier	An entity, which selects packets based on the content of packet headers according to defined rules.
Dropper	A device that performs dropping, i.e. discards packets based on specified rules.
DS behaviour aggregate	A collection of packets with the same DS codepoint crossing a link in a particular direction.
DS boundary node	A DS node that connects one DS domain to a node in another (possibly DS-capable) domain.
DS-capable	Capable of implementing DS as described in [ARCH/DS]. Usually used in reference to a domain consisting of DS-compliant nodes.
DS codepoint, DSCP	A specific value of the DS field, used to select a PHB.
DS-compliant	Capable of implementing DS as described in [ARCH/DS] and [HDR/DS]. Usually used in reference to a node or device.
DS egress node	A DS boundary node that handles traffic leaving a DS domain.
DS field	The IPv4 header TOS octet or the IPv6 Traffic Class octet when interpreted as defined in [HDR/DS].
DS ingress node	A DS boundary node that handles traffic entering a DS domain.
Hop	The link between two directly connected NE's and the associated cost.
Legacy node	A node that interprets IPv4 Precedence as specified in [RFC791, RFC1812] but which is otherwise not DS-compliant.
Marker	A device that performs marking, i.e. sets the DS codepoint in a packet based on defined rules.
Meter	A device that performs metering, i.e. measures the temporal properties (e.g. rate) of traffic selected by a classifier.
Microflow	A single instance of an application-to-application flow of packets which is identified by source address, source port, destination address, destination port and protocol id. A single microflow is of the finest granularity for service differentiation.
Network provisioning	The determination and allocation of the resources, both physical and logical, needed at various points in a network
Per-Hop-Behaviour, PHB	The externally observable forwarding behaviour applied at a DS-compliant node to a DS behaviour aggregate.
PHB group	A set of one or more PHBs that can only be meaningfully

	specified and implemented simultaneously, due to a common constraint such as a queue servicing or queue management policy.
Policing	The process of discarding packets (by a dropper) within a traffic stream in accordance with the state of a corresponding meter enforcing a traffic profile.
Service	The overall treatment of a defined subset of a customer's traffic within a DS domain or end-to-end.
Service Level Agreement, SLA	A service contract between a customer (organization (source domain) or another DS domain (upstream domain)) and a service provider that specifies the forwarding service a customer should receive.
Service provisioning policy	A policy which defines how traffic conditioners are configured on DS boundary nodes and how traffic streams are mapped to DS behaviour aggregates to achieve a range of services.
Traffic conditioner	An entity which performs traffic conditioning functions and which may contain meters, markers, droppers, and shapers. Traffic conditioners are typically deployed in DS boundary nodes only.
Traffic conditioning	Control functions performed to enforce rules specified in a TCA, including metering, marking, shaping, and policing.
Traffic Conditioning Agreement, TCA	An agreement specifying classifier rules and any corresponding traffic profiles and metering, marking, discarding and/or shaping rules which are to apply to the traffic streams selected by the classifier. A part of SLA.
Traffic profile	A description of the temporal properties of a traffic stream such as rate and burst size.